

**UNIVERSIDAD PRIVADA ANTENOR ORREGO**  
**FACULTAD DE INGENIERÍA**  
PROGRAMA DE ESTUDIO DE INGENIERIA DE COMPUTACIÓN Y SISTEMAS



TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE  
COMPUTACIÓN Y SISTEMAS

---

Uso de técnicas de web scraping para el análisis de datos de jugadores profesionales del futbol  
peruano para el periodo 2021

---

Línea de investigación: Gestión de datos y de información

**Autores:**

Monzón Laiza, Reynaldierrri Freud y Eudes  
Vargas Ulloa, Italo Noel

**Jurado Evaluador:**

Presidente: Castillo Robles, Edward Fernando  
Secretario: Calderón Sedano, José Antonio  
Vocal: Infantes Quiroz, Freddy Henry

**Asesor:**

Ullón Ramírez, Agustín Eduardo  
Código Orcid: <https://orcid.org/0000-0003-1198-1855>

Trujillo-Perú

2023

Fecha de sustentación: 2023/02/13

**“USO DE TÉCNICAS DE WEB SCRAPING PARA EL ANÁLISIS DE DATOS  
DE JUGADORES PROFESIONALES DEL FUTBOL PERUANO PARA EL  
PERIODO 2021”**

**Elaborado por:**

Br. Monzón Laiza Reynaldierra Freud y Eudes

Br. Vargas Ulloa Italo Noel

**Aprobada por:**



**Ms. Edward Fernando Castillo Robles**

**Presidente**


**CIP: 192352**



**Ms. José Antonio Calderón Sedano**

**Secretario**

**CIP: 139198**



**Ms. Freddy Henry Infantes Quiroz**

**Vocal**

**CIP: 139578**



**Ms. Agustín Eduardo Ullón**

**Ramírez**

**Asesor**

**CIP: 137602**

## DEDICATORIA

A mis padres por su esfuerzo y motivación por ayudarme a enfocarme y realizar uno de mis logros personales y profesionales; y a mis docentes universitarios por la paciencia en la búsqueda del camino del conocimiento.

**Br. Reynaldierrri Freud y Eudes Monzón Laiza**

A mis padres, mi hermano, y mi esposa por su apoyo en el desarrollo de la tesis. Y por su paciencia, comprensión y solidaridad, sin el cual no habría sido posible. Por eso, muchas gracias a todos.

**Br. Italo Noel Vargas Ulloa**

## RESUMEN

### “USO DE TÉCNICAS DE WEB SCRAPING PARA EL ANÁLISIS DE DATOS DE JUGADORES PROFESIONALES DEL FUTBOL PERUANO PARA EL PERIODO 2021”

Por:

**Br. Monzón Laiza Reynaldierri Freud y Eudes**

**Br. Vargas Ulloa Italo Noel**

El poder recurrir a una fuente confiable y detallada de todos los encuentros deportivos de los clubes peruanos de la Liga 1; y poder consultar en el tiempo, información como resultado de los encuentros deportivos por temporada, jugadores destacados, clubes con más campeonatos, faltas deportivas, edad de los jugadores, nivel económico de los equipos, entre otras métricas, es sumamente escasa y generalmente la información no está relacionada, ni tiene una homogeneidad. Ese interés en poder consultar todo en una sola interfaz general, se trazó como objetivo del proyecto de esta tesis, la creación de un Dashboard en Power BI, y que como fuente de datos se utilizaría unas páginas online futbolísticas, que, aunque no contiene todas las métricas con respecto a otros países, se usó para la creación de dicho dashboard mencionado anteriormente. Para ello se recurrió en la fase de recolección de datos, el uso de una técnica de web scraping, mediante una extensión de Google Chrome, por ser de fácil uso y conveniencia del proyecto. En el desarrollo de este proyecto se usó la metodología SCRUM, basado en 2 fases: una fase de planificación y la otra fase de desarrollo para el desarrollo del Dashboard. Paso seguido se publicó de manera online, para que cualquier persona interesada, pueda interactuar de manera remota siempre que tenga un acceso privilegiado. Al término del desarrollo del proyecto se evaluó mediante un cuestionario dirigido a un grupo experto, la valoración de dashboard, y se concluyó que, a pesar del reto de la poca información detallada, se pudo generar un reporte general con diferentes vistas, siendo este una base para análisis más interesantes en el futuro.

Palabras clave: Encuentros deportivos, jugadores destacados, faltas deportivas, edad jugadores, dashboard, Power BI, web scraping

## **ABSTRACT**

“USE OF WEB SCRAPING TECHNICES FOR THE ANALYSIS OF PROFESSIONAL PLAYERS OF THE PERUAN FOOTBALL FOR THE PERIOD 2021”

Por:

**Br. Monzón Laiza Reynaldierrri Freud y Eudes**

**Br. Vargas Ulloa Italo Noel**

Being able to resort to a reliable and detailed source of all the sporting events of the Peruvian clubs of League 1; and to be able to consult in time, information as a result of seasonal sports matches, featured players, clubs with the most championships, sports fouls, age of the players, economic level of the teams, among other metrics, it is extremely scarce and generally the information is not related, nor is it homogeneity. That interest in being able to consult everything in a single general interface, was outlined as the objective of the project of this thesis, the creation of a Dashboard in Power BI, and that online soccer pages would be used as a data source, which, although it does not contain all the metrics like other countries, was used for the creation of said dashboard mentioned above. For this, the use of a web scraping technique was used in the data collection phase, using an extension of Google Chrome, for being easy to use and convenient to the project. In the development of this project, the SCRUM methodology was used, based on 2 phases: one planning phase and the other development phase for the development of the Dashboard. Step by step it was published online, so that anyone interested can interact remotely whenever they have privileged access. At the end of the development of the project, the evaluation of dashboard was evaluated through a questionnaire addressed to an expert group, and it was concluded that, despite the challenge of little detailed information, a general report could be generated with different views, this being a basis for more interesting analyzes in the future.

Keywords: sports matches, featured players, sports fouls, age of the players, dashboard, Power BI, web scraping

## PRESENTACIÓN

Señores Miembros del Jurado:

De acuerdo a los reglamentos de la presentación de trabajos de investigación, presentamos la tesis titulada: **“Uso de técnicas de web scraping para el análisis de datos de jugadores profesionales del fútbol peruano para el periodo 2021”** para poder obtener el Título Profesional de Ingeniero de Computación y Sistemas.

El trabajo de tesis se desarrolló dentro del marco de referencia de los reglamentos de la Facultad de Ingeniería y del programa de estudio de Ing. de Computación y Sistemas, así como basado en los saberes adquiridos en todo el tiempo estudiado en la universidad.

Los autores.

# ÍNDICE DE CONTENIDO

<b>DEDICATORIA.....</b>	<b>ii</b>
<b>RESUMEN.....</b>	<b>iii</b>
<b>ABSTRACT.....</b>	<b>iv</b>
<b>PRESENTACIÓN.....</b>	<b>v</b>
<b>ÍNDICE DE CONTENIDO .....</b>	<b>vi</b>
<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
1.1. Planteamiento del problema .....	1
1.2. Objetivos del estudio .....	10
1.3. Justificación del estudio .....	11
1.3.1. Importancia del trabajo.....	11
1.3.2. Viabilidad de la investigación .....	11
1.3.3. Aportes .....	11
1.4. Delimitación del problema.....	11
1.5. Características problemáticas .....	11
1.6. Definición del problema.....	12
1.7. Formulación del problema .....	12
1.8. Formulación de la hipótesis.....	12
<b>2. MARCO TEÓRICO.....</b>	<b>13</b>
2.1. Antecedentes .....	13
2.2. Definiciones .....	16
2.2.1. Web Scraping.....	16
2.2.2. Extensión Scraper .....	18
2.2.3. HTML .....	19
2.2.4. ¿Qué es el Dashboard? .....	19
2.2.5. ¿Qué es un dato? .....	20
2.2.6. ¿Qué es el análisis de datos?.....	20

2.2.6.1	Características del análisis de datos.....	21
2.2.6.2	Tipos de análisis de los datos .....	22
2.2.6.3	Obtención de los datos .....	23
2.2.6.4	Almacenamiento de los datos .....	23
2.2.6.5	Segmentación de los datos .....	24
2.2.6.6	Proceso del análisis de datos.....	24
2.2.6.7	Importancia del análisis de datos .....	25
2.2.7.	Campos de aplicación en el análisis de datos.....	26
2.2.7.1	En el sector Económico .....	26
2.2.7.2	En el campo de la medicina .....	27
2.2.7.3	En el manejo de recursos humanos .....	28
2.2.7.4	En el campo deportivo.....	29
2.2.8.	El futuro del análisis de datos en el fútbol.....	30
2.2.9.	Framework SCRUM .....	31
2.2.10.	La herramienta Power BI .....	33
2.2.11.	Otras Herramientas existentes .....	34
2.2.11.1	Pentaho .....	34
2.2.11.2	Tableau .....	34
2.2.11.3	IBM Watson Analytics.....	35
<b>3.</b>	<b>MATERIALES Y METODOS.....</b>	<b>36</b>
3.1.	Material .....	36
3.2.	Método .....	36
<b>4.</b>	<b>RESULTADOS.....</b>	<b>38</b>
4.1.	Planificación de la metodología SCRUM.....	38
4.1.1.	Lista de Producto o Product Backlog.....	38
4.1.2.	Sprint Backlog .....	39
4.1.3.	Incremento .....	42



4.2.	Desarrollo de la metodología SCRUM .....	43
4.2.1.	Definir el alcance del proyecto .....	43
4.2.2.	Búsqueda de la data.....	44
4.2.3.	Seleccionando la técnica de web scraping.....	45
4.2.4.	Raspado de los datos .....	46
4.2.5.	Obtener los datos (GD) .....	55
4.2.6.	Preparación de datos (DP) .....	58
4.2.7.	Modelado de datos (DM).....	75
4.2.8.	Reporte de datos (DR) .....	78
4.2.9.	Almacenamiento y publicación del Dashboard.....	82
5.	<b>DISCUSIÓN DE RESULTADOS.....</b>	<b>84</b>
5.1.	Formulación del problema .....	84
5.2.	Hipótesis.....	84
5.3.	Muestra a aplicar .....	84
5.4.	Validación de la solución .....	84
5.4.1.	Variable independiente .....	84
5.4.2.	Variable dependiente .....	88
6.	<b>CONCLUSIONES .....</b>	<b>103</b>
7.	<b>RECOMENDACIONES.....</b>	<b>106</b>
	<b>ANEXOS.....</b>	<b>112</b>

## ÍNDICE DE FIGURAS

Figura 1 Mapa mundial de la FIFA en el mundo.....	1
Figura 2 Población deportiva por País .....	2
Figura 3 Total de fondos disponibles de la FIFA del 2016 al 2020 .....	3
Figura 4 Total de fondos aprobados desde el 2016 al 2020 .....	3
Figura 5 Proyectos aprobados por País del 2016 al 2020 .....	4
Figura 6 Valor de dólares por País en Infraestructura.....	4
Figura 7 Creación de capacidad por país .....	5
Figura 8 Valor en dólares de los costes de las competiciones por país.....	5
Figura 9 Valor en dólares de subvenciones por País .....	6
Figura 10 Monto en dólares de otros proyectos por país.....	6
Figura 11 Estructura de funcionamiento de la extensión scraper de Google Chrome.....	18
Figura 12 El proceso de análisis y visualización de datos.....	25
Figura 13 Proceso del Framework Scrum .....	32
Figura 14 Ejemplo de un Dashboard de Segmentación de mercados .....	33
Figura 15 Interface de Power Bi Desktop .....	33
Figura 16 Ejemplo de la herramienta Pentaho .....	34
Figura 17 Ejemplo de la herramienta Tableau, en una campaña y retenciones	34
Figura 18 Ejemplo de la herramienta IBM Watson Analytics .....	35
Figura 19 Forma gráfica del diseño de investigación .....	36
Figura 20 Pagina de Transfermarkt .....	44
Figura 21 Pagina de FBREF .....	44
Figura 22 Pagina de Wikipedia.....	45
Figura 23 Vista histórica de cambios de la página Transfermarkt en Waybackmachine.....	45
Figura 24 Instalado extensión Scraper de Google Chrome.....	46
Figura 25 Uso de la extensión Scrape de Google parte 1 .....	46
Figura 26 Uso de la extensión Scrape de Google parte 2 .....	47
Figura 27 Relación de la información extraída de las páginas mencionadas....	47
Figura 28 Apertura 2014-2021_Liga1.....	48
Figura 29 Clasificación Internacional de equipos peruanos desde el 2015 hasta el 2021 .....	48

Figura 30 Clasificación_Liga_2015-2021.....	49
Figura 31 Clausura 2014-2021_Liga1 .....	49
Figura 32 Equipos de Países con sus continentes .....	50
Figura 33 Estadios Nacionales.....	50
Figura 34 Estadística_estándar_2015_2021.....	51
Figura 35 Estadísticas_adicionales_2015_2021 .....	51
Figura 36 Local_visitante_2014-2021 .....	52
Figura 37 Marcadores_Partidos_2014-2021.....	52
Figura 38 Porteria_equipo_2015_2021.....	53
Figura 39 Relación_EquiposGanadores_1912-2021.....	53
Figura 40 Relación_jugadores_2015-2021 .....	54
Figura 41 TiempoJuego_DelEquipo_2015_2021 .....	54
Figura 42 Top_goleadores_1928-2021.....	55
Figura 43 Selección de tipo de fuente en la carga en el Power BI.....	55
Figura 44 Seleccionado el archivo.....	56
Figura 45 Seleccionando la opción cargar .....	56
Figura 46 Visualización del archivo cargado.....	57
Figura 47 Importación de un archivo csv .....	57
Figura 48 Estructura del archivo Pais_Continentes.csv.....	58
Figura 49 Pais_Continentes.csv (Data para el modelado) .....	58
Figura 50 Estructura del archivo Paises.csv .....	59
Figura 51 Paises.csv (Data para el modelado).....	59
Figura 52 Estructura del archivo Estadios.xlsx .....	59
Figura 53 Estadios.xlsx (Data para el modelado).....	59
Figura 54 Estructura del archivo Top_Goleadores.....	60
Figura 55 Top_Goleadores.xlsx (Data para el modelado) .....	60
Figura 56 Estructura del archivo Campeón_Por_Eras.xlsx .....	60
Figura 57 Campeón_Por_Eras.xlsx (Data para el modelado).....	60
Figura 58 Estructura de los archivos Clasificados del 2015.csv al 2021.csv.....	60
Figura 59 Formato de agregar filas en Power BI.....	61
Figura 60 Combinando Clasificados del 2015 al 2021 .....	61
Figura 61 Anexando Clasificados del 2015 al 2021(Data para el modelado)....	62
Figura 62 Estructura de los archivos del Apertura y Clausura desde el 2014 al 2021 .....	63

Figura 63 Anexando archivos del Apertura y Clausura del 2014 al 2021 parte 01 .....	63
Figura 64 Anexando de archivos del Apertura y Clausura del 2014 al 2021 parte 02.....	64
Figura 65 Anexando de archivos del Apertura y Clausura del 2014 al 2021 parte 03.....	64
Figura 66 Estructura de los archivos de la Clasificación_Liga_2015.csv al 2021.csv .....	65
Figura 67 Estructura del archivo Local_visitante_2014 al 2021 parte 01 .....	65
Figura 68 Reemplazando valores con Power BI .....	66
Figura 69 Estructura del archivo Local_visitante_2014 al 2021 parte 02.....	67
Figura 70 Formato de combinar filas en Power BI .....	67
Figura 71 Combinando Clasificación_Liga_2015 con Local_visitante_2015.....	68
Figura 72 Resultado de combina Clasificación_Liga_Local_visitante_2015....	68
Figura 73 Presentación de datos de Estadística estándar, adicional y portería	69
Figura 74 Combinando Estadística estándar, adicional, portería, clasificación liga y local visitante.....	70
Figura 75 Estructura del archivo Anexando_Liga_estadísticas_2015-2021 .....	71
Figura 76 Estructura del archivo de Marcadores y partidos del 2014.....	72
Figura 77 Estructura del archivo Anexando Marcadores y partidos del 2015-2021 .....	73
Figura 78 Estructura de Detalle de jugadores del 2014 al 2021.....	73
Figura 79 Estructura de Detalle de jugadores del 2014 al 2021 .....	74
Figura 80 Modelado de países por continentes.....	75
Figura 81 Modelado de estadios peruanos .....	75
Figura 82 Modelo de top de goleadores de 1928 al 2021.....	75
Figura 83 Modelado de Equipo campeón por eras .....	76
Figura 84 Modelado de la Clasificación de equipos del Torneo Apertura y Clausura desde 2014-2021 .....	76
Figura 85 Modelado de Clasificación Internacional de los equipos peruanos del 2015 al 2021 .....	76
Figura 86 Anexando la combinación de datos estándar del 2015 al 2021 .....	77
Figura 87 Detalle de partidos del Apertura y Clausura del 2014 al 2021 .....	77
Figura 88 Detalle de Jugadores del 2014 al 2021 .....	77

Figura 89 Dashboard General.....	78
Figura 90 Reporte de la información deportiva de los jugadores peruanos del año 2014 - 2021 .....	78
Figura 91 Reporte de arqueros con más partidos ganados y penales detenidos del 2015 -2021 .....	79
Figura 92 Reporte de faltas deportivas por club deportivo del 2015 al 2021 ....	79
Figura 93 Reporte de rendimiento de los jugadores en la costa, sierra y selva del 2014 - 2021 .....	80
Figura 94 Reporte de Valor monetario promedio en soles, euros y dólares de los jugadores peruanos al 2021 .....	80
Figura 95 Reporte de influencia de los jugadores en el futbol peruano en los torneos del 2014 – 2021 .....	81
Figura 96 Reporte de cantidad de goles marcados por los jugadores en los clubes peruanos entre el 2014 al 2021 .....	81
Figura 97 Reporte de Top de Goleadores del 1928 al 2021 .....	82
Figura 98 Logueo en Power BI Desktop .....	82
Figura 99 Publicación en Power BI.....	83

## ÍNDICE DE TABLAS

Tabla 1 Datos económicos de los integrantes de la COMMEBOL Fuente: Página oficial de la FIFA.....	2
Tabla 2 Ranking Internacional de los primeros 100 clubes deportivos.....	9
Tabla 3 Ranking Internacional de clubes deportivos de Sudamérica .....	9
Tabla 4 Tabla de los 5 fichajes más caros de jugadores peruanos .....	10
Tabla 5 Principales Etiquetas HTML .....	19
Tabla 6 Operacionalización de las variables.....	37
Tabla 7 Lista de Producto.....	38
Tabla 8 Información del Sprint 01 .....	39
Tabla 9 Detalle de tareas del Sprint 01 .....	39
Tabla 10 Información del Sprint 02 .....	40
Tabla 11 Detalle de tareas del Sprint 02.....	40
Tabla 12 Información del Sprint 03 .....	40
Tabla 13 Detalle de tareas del Sprint 03.....	41
Tabla 14 Información del Sprint 04 .....	41
Tabla 15 Detalle de tareas del Sprint 04.....	41
Tabla 16 Información del Sprint 05 .....	42
Tabla 17 Detalle de tareas del Sprint 05.....	42
Tabla 18 Alcance del proyecto de tesis .....	43
Tabla 19 Tabla de datos de la variable independiente.....	84
Tabla 20 Tabla de datos de la variable dependiente registros.....	88
Tabla 21 Tabla de datos de la variable dependiente tiempo.....	92
Tabla 22 Tabla de datos de la variable dependiente utilidad .....	97

# 1. INTRODUCCIÓN

## 1.1. Planteamiento del problema

Una de los juegos más populares a nivel mundial es el fútbol; este deporte no solo es el rey en popularidad, sino en entretenimiento en diversos países del mundo.

El ente máximo representante de esta disciplina deportiva, es la FIFA (Federación Internacional de Asociación de Fútbol), ente deportivo que se encarga de modificar las reglas del juego, además de organizar campeonatos alrededor del mundo, así como la copa mundial de este año, que es Catar 2022.

La influencia del fútbol a nivel mundial, se ha vuelto un negocio que mueve alrededor de casi dos mil ochocientos millones de dólares a nivel mundial.



Figura 1 Mapa mundial de la FIFA en el mundo

Como se puede visualizar en la figura 1, hay siete organizaciones deportivas, sectorizadas que se encargan de campeonatos por zonas continentales y las cuales son:

1. CONCACAF (La Confederación de Norteamérica, Centroamérica y el Caribe de Fútbol)
2. CONMEBOL (La Confederación Sudamericana de Fútbol)
3. UEFA (la Unión de Federaciones Europeas de Fútbol)
4. CAF (La Confederación Africana de Fútbol)
5. AFC (La Confederación Asiática de Fútbol)
6. OFC (La Confederación de Fútbol de Oceanía).

Como se enumeró anteriormente los entes deportivos por zona continental, nos correspondería la COMMEBOL, siendo participes de este ente los países de: Venezuela, Colombia, Ecuador, Perú, Bolivia, Brasil, Uruguay, Paraguay, Argentina y Chile (Federación Internacional de Asociación de Fútbol, 2022)

Actualmente de los países mencionados, los dos países más representativos y de un buen nivel en las copas mundiales es: Brasil y Argentina, siendo reconocidos ese estatus futbolístico a nivel mundial, y ahí es donde; haremos una pequeña comparación con datos proporcionados en la página oficial de la FIFA, de cómo se encuentran actualmente no solo los dos países mencionados anteriormente, sino del resto de países sudamericanos, pertenecientes a la COMMEBOL.

1	Venezuela	32,286,226	\$ 10,350,000	\$ 7,950,000	2	\$ 2,250,000	0	0	0	0
2	Colombia	49,424,241	\$ 9,750,000	\$ 9,750,000	8	\$ 741,767	\$ 539,523	\$ 753,000	\$ 2,000,000	\$ 215,710
3	Ecuador	16,838,101	\$ 9,850,000	\$ 7,325,000	6	\$ 1,500,000	\$ 16,660	\$ 708,340	0	0
4	Perú	32,504,705	\$ 9,750,000	\$ 7,750,000	2	0	0	\$ 2,250,000	0	0
5	Bolivia	11,199,210	\$ 11,025,000	\$ 8,407,591	5	\$ 1,410,558	\$ 72,000	0	0	\$ 150,033
6	Brasil	212,691,030	\$ 9,750,000	\$ 7,750,000	4	\$ 892,567	\$ 807,942	0	\$ 546,449	\$ 3,043
7	Uruguay	3,468,268	\$ 9,750,000	\$ 7,450,000	12	\$ 520,000	\$ 55,000	\$ 160,000	\$ 645,000	\$ 870,000
8	Paraguay	6,886,250	\$ 10,275,000	\$ 10,275,000	6	0	0	0	0	0
9	Argentina	44,644,145	\$ 9,750,000	\$ 9,750,000	8	\$ 1,846,225	0	\$ 2,025,851	0	\$ 377,924
10	Chile	18,472,084	\$ 9,750,000	\$ 7,950,000	8	\$ 89,849	\$ 35,821	\$ 2,624,330	0	0

Tabla 1 Datos económicos de los integrantes de la COMMEBOL Fuente: Página oficial de la FIFA

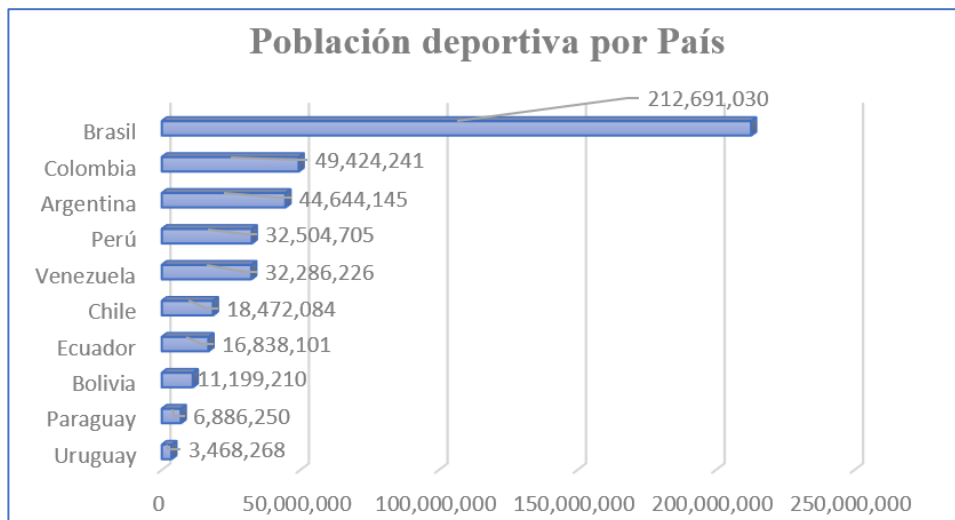
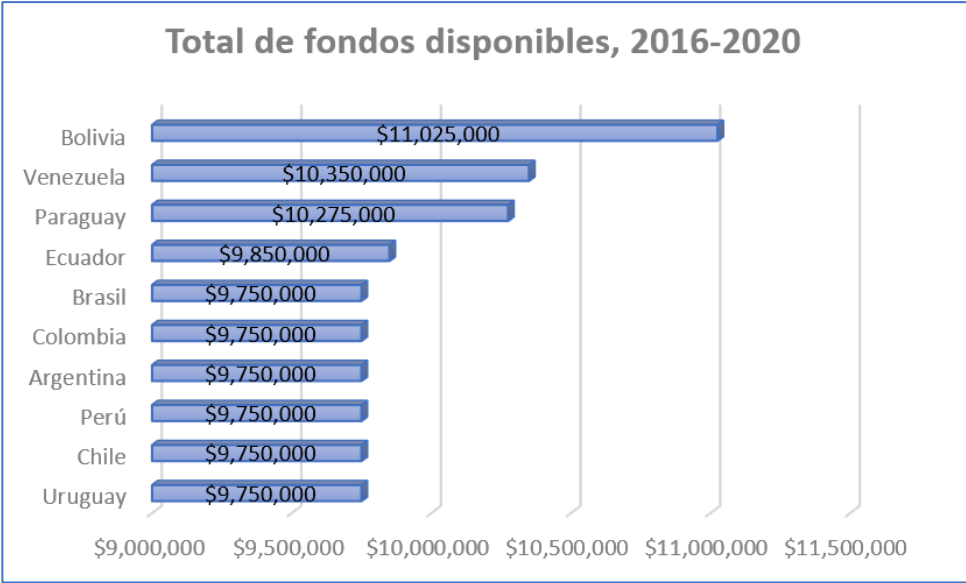
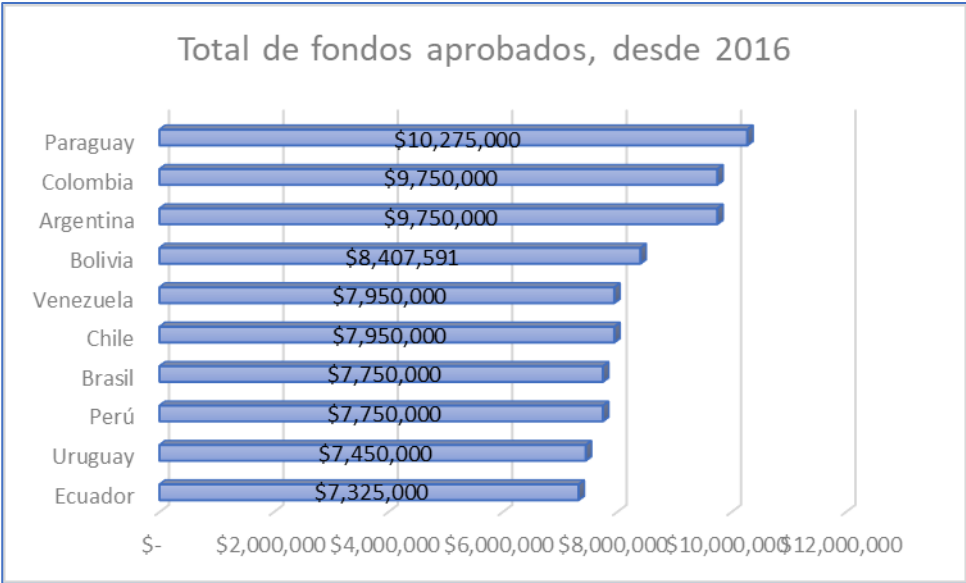


Figura 2 Población deportiva por País





*Figura 3 Total de fondos disponibles de la FIFA del 2016 al 2020*



*Figura 4 Total de fondos aprobados desde el 2016 al 2020*

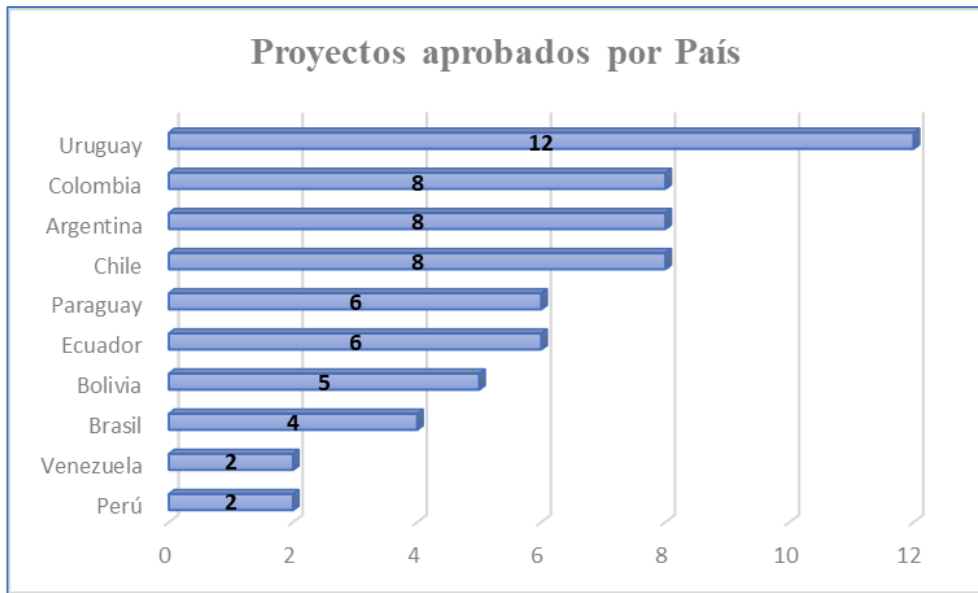


Figura 5 Proyectos aprobados por País del 2016 al 2020

### Valor de dólares por País en infraestructura

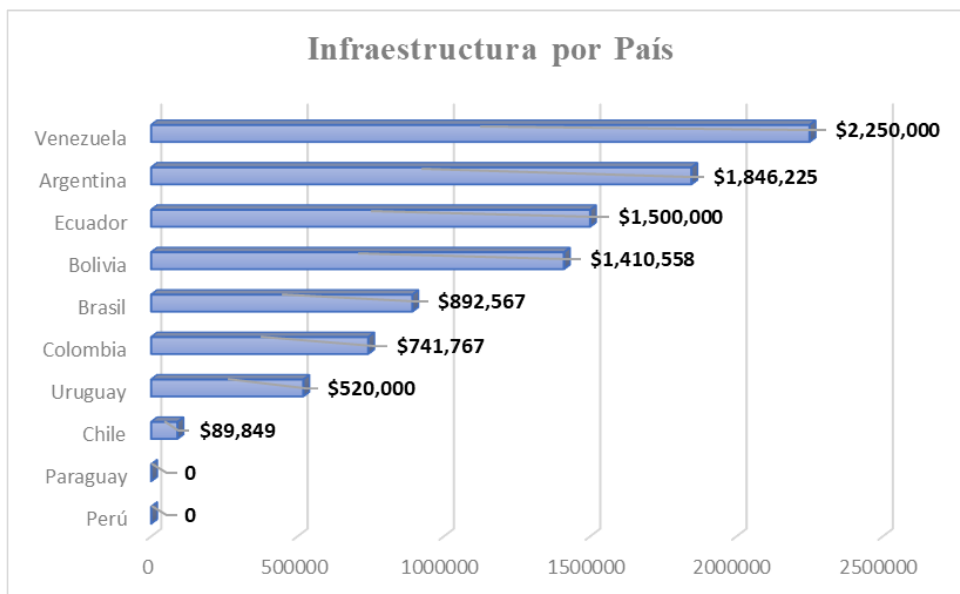


Figura 6 Valor de dólares por País en Infraestructura

## Valor de dólares por País en capacidad de construcción

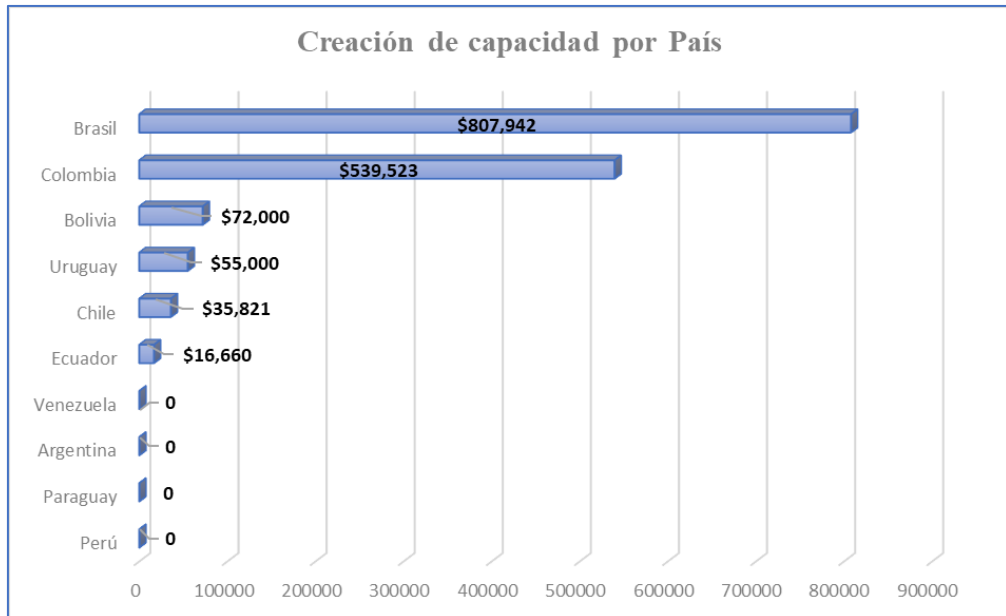


Figura 7 Creación de capacidad por país

## Valor en dólares de las competencias por País



Figura 8 Valor en dólares de los costes de las competencias por país

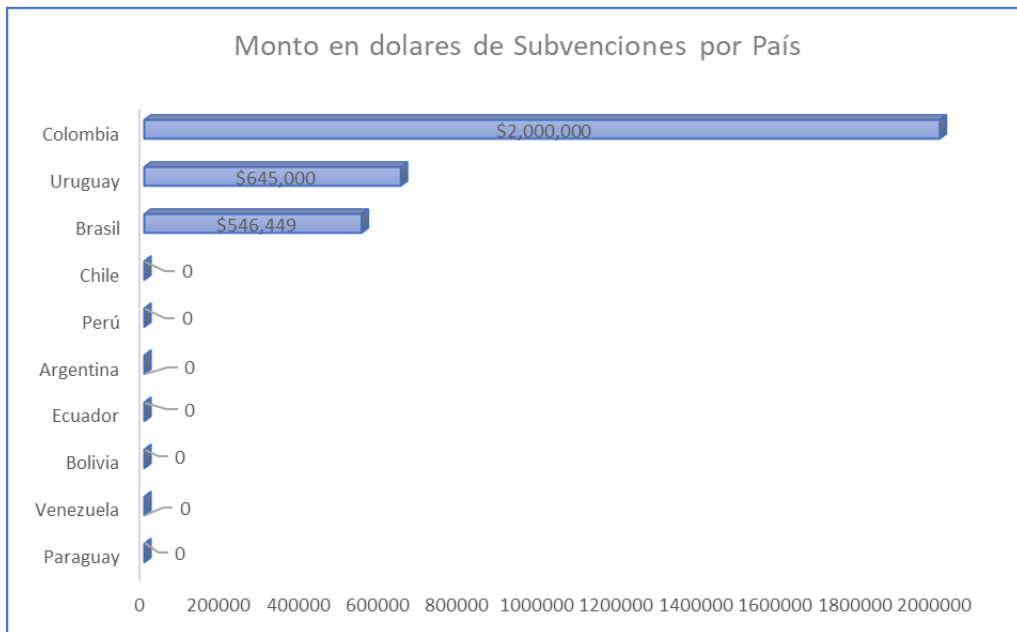


Figura 9 Valor en dólares de subvenciones por País

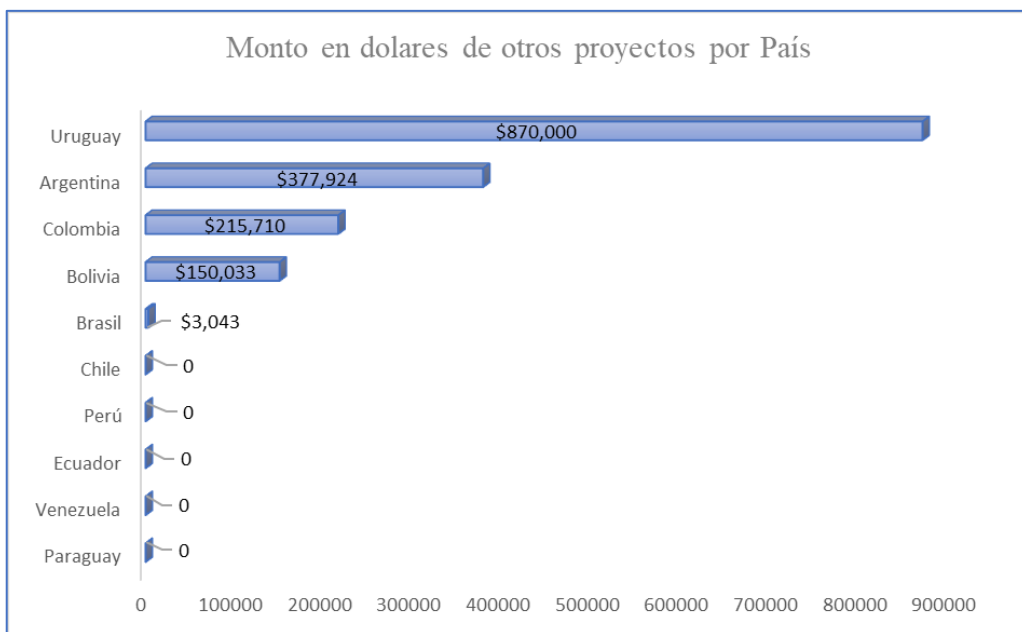


Figura 10 Monto en dólares de otros proyectos por país

Como se puede visualizar en los datos de la tabla 1, se puede ver en la figura 2, que el país de Brasil tiene la población deportiva más grande de la región sudamericana, con una población deportiva de apenas del 15%, si lo comparamos con ese país, a su vez con la figura 3, el Perú tiene una disponibilidad de fondos similar, como Ecuador, Colombia, Chile, Uruguay

Brasil y Argentina, siendo este un punto resaltable debido a la paridad económica frente a los otros países sudamericanos.

En la figura 4 no hay mucha diferencia con respecto a los fondos aprobados desde el 2016 al 2020.

En la figura 5, se destaca Uruguay como primer país con proyectos aprobados, siendo Perú y Venezuela los que ocupan los últimos lugares.

En la figura 6, el Perú no contaba con proyectos de infraestructura, ni capacidad de construcción como se muestra en la figura 7. Aunque ocupa un segundo lugar en el costo de competencias para el País del Perú, que se muestra en la figura 8.

Finalmente, en la figura 9 dentro de los proyectos aprobados por la FIFA, el Perú no subvenciona a ningún equipo y tampoco tiene un plan presupuestal para proyectos adicionales.

Entonces debería tomarse en cuenta según los datos de la FIFA, que el apoyo económico por la parte de esta entidad es similar entre los 10 integrantes de la COMMEBOL, y debería de tomarse en cuenta y generar proyectos de infraestructura y subvenciones de equipos para el desarrollo de nuevos talentos en el fútbol nacional.

Ahora si nos enfocamos en los clubes nacionales, se definen por dos ligas deportivas en el futbol peruano, los cuales son: la liga 2 o semiprofesional y la Liga 1 o profesional, pero solo nos enfocaremos en la liga 1.

En el desarrollo de la Liga 1, se realizan dos torneos que se denominan Torneo Apertura y Torneo Clausura, que son actividades deportivas para evaluar el desempeño futbolístico de los clubes deportivos en el Perú.

En el desarrollo y finalización de cada Torneo Clausura, no solo se conoce el ganador de cada Torneo. Sino que clubes extranjeros buscan jugadores potenciales para incluirlos en sus clubes, siendo así los torneos nacionales una ventana para ver el desempeño de los jugadores de los diferentes clubes y poder emigrar y desarrollar su potencial futbolístico y económico.

Otro de las aristas de los torneos nacionales, es escoger un club representativo del Perú y participe en campeonatos internacionales.

Otro punto importante en el desarrollo de los diferentes torneos nacionales, es conocer que jugadores pueden representarnos en el evento

futbolístico más grande del mundo una vez clasificados, que es el mundial que se realiza cada cuatro años.

Ahora si queremos clasificar los diferentes clubes nacionales a nivel internacional, no aparecemos ni en el puesto 100 de dicho ranking, como se muestra en la tabla 2.

Ranking Internacional	Club	Pais	Ranking Internacional	Club	Pais
1	Manchester City	Inglaterra	31	Brentford FC	Inglaterra
2	Liverpool FC	Inglaterra	32	Nottingham Forest	Inglaterra
3	París Saint-Germain FC	Francia	33	Ajax de Ámsterdam	Países Bajos
4	Bayem Múnich	Alemania	34	Stade Rennais FC	Francia
5	Chelsea FC	Inglaterra	35	Leeds United	Inglaterra
6	Real Madrid CF	España	36	Crystal Palace	Inglaterra
7	FC Barcelona	España	37	Brighton & Hove Albion	Inglaterra
8	Manchester United	Inglaterra	38	Sevilla FC	España
9	Tottenham Hotspur	Inglaterra	39	Olympique de Lyon	Francia
10	Arsenal FC	Inglaterra	40	SL Benfica	Portugal
11	Atlético de Madrid	España	41	Fiorentina	Italia
12	Inter de Milán	Italia	42	Real Betis Balompié	España
13	AC Milan	Italia	43	FC Oporto	Portugal
14	Aston Villa	Inglaterra	44	Olympique de Marsella	Francia
15	Juventus de Turín	Italia	45	SS Lazio	Italia
16	RB Leipzig	Alemania	46	Valencia CF	España
17	Borussia Dortmund	Alemania	47	Athletic Club	España
18	SSC Nápoles	Italia	48	OGC Niza	Francia
19	West Ham United	Inglaterra	49	Fulham FC	Inglaterra
20	Bayer 04 Leverkusen	Alemania	50	Sporting CP	Portugal
21	Leicester City	Inglaterra	51	Eintracht Fráncfort??	Alemania
22	Everton FC	Inglaterra	52	PSV Eindhoven	Países Bajos
23	Newcastle United	Inglaterra	53	VfL Wolfsburgo	Alemania
24	AS Roma	Italia	54	Borussia Mönchengladbach	Alemania
25	Wolverhampton Wanderers	Inglaterra	55	US Sassuolo	Italia
26	Real Sociedad	España	56	Red Bull Salzburgo	Austria
27	Villarreal CF	España	57	LOSC Lille Métropole	Francia
28	AS Mónaco	Mónaco	58	Fenerbahce	Turquía
29	Atalanta de Bérgamo	Italia	59	TSG 1899 Hoffenheim	Alemania
30	Southampton FC	Inglaterra	60	AFC Boumemouth	Inglaterra

Ranking Internacional	Club	Pais	Ranking Internacional	Club	Pais
61	SE Palmeiras	Brasil	81	Udinese	Italia
62	SC Friburgo	Alemania	82	SC Braga	Portugal
63	Club Brujas KV	Bélgica	83	FC Nantes	Francia
64	Zenit de San Petersburgo	Rusia	84	Watford FC	Inglaterra
65	Galatasaray	Turquía	85	Montpellier HSC	Francia
66	Trabzonspor	Turquía	86	FC Dinamo de Kiev	Ucrania
67	Torino FC	Italia	87	Racing Club de Estrasburgo	Francia
68	Clube de Regatas do Flamengo	Brasil	88	Celtic FC	Escocia
69	Getafe CF	España	89	1.FSV Mainz 05	Alemania
70	RC Lens	Francia	90	1.FC Unión Berlín	Alemania
71	Bolonia	Italia	91	Hertha Berlín	Alemania
72	Olympiakos El Pireo	Grecia	92	Sport Club Corinthians Paulista	Brasil
73	Feyenoord Rotterdam	Países Bajos	93	Hellas Verona	Italia
74	RC Celta de Vigo	España	94	Norwich City	Inglaterra
75	VfB Stuttgart	Alemania	95	US Salernitana 1919	Italia
76	Rangers FC	Escocia	96	RSC Anderlecht	Bélgica
77	AC Monza	Italia	97	GNK Dinamo Zagreb	Croacia
78	CA Osasuna	España	98	KRC Genk	Bélgica
79	Besiktas JK	Turquía	99	CA River Plate	Argentina
80	Clube Atlético Mineiro	Brasil	100	RCD Espanyol	España

*Tabla 2 Ranking Internacional de los primeros 100 clubes deportivos*

Ranking Internacional	Club	Pais
61	SE Palmeiras	Brasil
68	Clube de Regatas do Flamengo	Brasil
80	Clube Atlético Mineiro	Brasil
92	Sport Club Corinthians Paulista	Brasil
99	CA River Plate	Argentina

*Tabla 3 Ranking Internacional de clubes deportivos de Sudamérica*

Ante lo expuesto anteriormente, podríamos decir que ningún club peruano cuenta con un nivel internacional y eso podría ser una limitante, para poder tomar en cuenta y contratar un jugador de algún club peruano.

¿Pero cuales criterios se toman en cuenta para un fichaje de un jugador? y ¿Qué tiempo se necesita para evaluar esa decisión?

Para empezar ningún jugador es contratado en un plazo de veinticuatro horas, siempre puede ser en días, meses o inclusive años, ya que se maneja criterios como: edad, datos médicos sobre el jugador, desempeño futbolístico, posición de juego, conducta deportiva, pretensiones económicas, motivación personal, entre otros criterios que considere el club interesado.

A continuación, mostraremos un ejemplo de los 5 fichajes más caros de jugadores peruanos. (BOLAVIP, 2021)

Nacionalidad	Jugador	Club Proveniente	Club Destino	Edad	Posición	Goles anotados	Partidos jugados	Monto Fichaje €	Monto Fichaje S/.
Peruano-Italiano	Gianluca Lapadula	Pescara	AC Milan	25 años	Delantero	8	29	€ 9,500,000	S/ 38,744,483
Peruano-Italiano	Gianluca Lapadula	AC Milan	Génova	26 años	Delantero	7	36	€ 11,000,000	S/ 44,862,033
Peruano	Andre Carillo	Benfica	Al-Hilal	28 años	Extremo derecho	19	95	€ 9,600,000	S/ 39,152,319
Peruano	Jefferson Farfán	PSV Holandés	FC Schalke 04	24 años	Media punta - Extremo derecho	53	228	€ 10,000,000	S/ 40,783,666
Peruano	Juan Manuel Vargas	Catiana	Florentina	25 años	Lateral izquierdo	25	196	€ 11,000,000	S/ 48,940,399

Tabla 4 Tabla de los 5 fichajes más caros de jugadores peruanos

Esta información generalmente es manejada de manera interna entre clubes por intermedio de un agente, ya que más de un club puede estar interesado en un jugador. Estas negociaciones siempre pueden realizarse de manera muy reservada o muy abierta, dependiendo de la popularidad del jugador.

Ante lo descrito anteriormente, vamos a tratar de resolver esa duda y ver si el análisis de los datos basado en un modelo podría beneficiarlos o ser de interés.

Para ello se busca apuntar a un análisis de los datos futbolísticos, que pueda obtenerse de diferentes fuentes en la red. Se pretende tener un referente de los datos encontrados y si pueden servir de algún de modelo para interactuar con ellos, ya que la finalidad no es solo de desempeño futbolístico de los clubes peruanos, sino comercial de los jugadores.

## 1.2. Objetivos del estudio

El **objetivo general** es:

- Analizar la información disponible sobre los jugadores de la liga 1 en el fútbol peruano.

Los **objetivos específicos** son los siguientes:

- Establecer la búsqueda de sitios online, las cuales puedan ofrecen información futbolística sobre la liga 1 de clubes peruanos.
- Elaborar la planificación del spring basado en el framework de Scrum.
- Utilizar alguna técnica de web scraping.
- Construir un Dashboard en Power BI, que pueda mostrar el análisis de datos del fútbol peruano.



### **1.3. Justificación del estudio**

#### **1.3.1. Importancia del trabajo**

- Uno de los puntos importantes en la investigación, es analizar y presentar la información de los jugadores de los diferentes clubes deportivos peruanos, para el interés desde cualquier aficionado, o club deportivo en el fútbol peruano.
- La finalidad es darle un valor comercial, que ayude a la toma de decisión por personas interesadas en los diferentes clubes y sus jugadores, en el fútbol peruano.

#### **1.3.2. Viabilidad de la investigación**

- Sera viable, ya que los datos adquiridos serán obtenidos a través de alguna técnica de web scraping, hacia portales online, que puedan proporcionar información de equipos peruanos y sus jugadores.
- Sera viable porque se cuenta con las herramientas necesarias y los conocimientos para el desarrollo del proyecto de investigación.

#### **1.3.3. Aportes**

- Académico: Nos permitirá conocer los conocimientos adquiridos en la universidad, así como generar un punto de vista poco tratado y descuidado en nuestro territorio nacional a nivel deportivo, ya que esto podría dar origen a la investigación de otras bases deportivas o a empresas emergentes de cualquier campo empresarial.
- Sociedad: los interesados a nivel deportivo, podrán contar con una herramienta tecnológica para la toma de sus decisiones.

### **1.4. Delimitación del problema**

El proyecto de investigación a realizarse, estará delimitado en el análisis y presentación de los datos de las ligas deportivas de fútbol profesional en todo el territorio peruano, en un dashboard y usando web scraping.

### **1.5. Características problemáticas**

- La información deportiva no es un tema muy usado de forma tecnológica en el fútbol profesional peruano.

- La ausencia de un análisis, hace imposible manejar un uso futuro del Big Data en el fútbol profesional peruano.
- Al no haber un análisis nacional del fútbol, no podría existir una comparativa deportiva entre los diversos jugadores peruanos.

### **Análisis de características problemáticas**

- Las instituciones deportivas, no tienen alguna herramienta tecnológica que pueda obtener y procesar data, para general algún valor comercial.
- Al no haber algún análisis deportivo que se pueda registrar y almacenar, hará que esa información se pierda y no pueda utilizarse en un futuro cercano o lejano.
- La falta de comparativas entre jugadores en base a los datos, hace que la afinidad de las personas entre jugadores y entrenadores, pueda afectar en la toma de decisiones.

### **1.6. Definición del problema**

El problema se basa en desarrollar un dashboard, que pueda representar información histórica del nivel futbolístico de los diferentes clubes deportivos del Perú y sus jugadores.

### **1.7. Formulación del problema**

¿Cómo realizar un análisis significativo de datos sobre los jugadores de la liga 1 en fútbol peruano?

### **1.8. Formulación de la hipótesis**

$H_A$  = La técnica scraping del plugin webscraper permite recuperar información para el análisis de los datos de jugadores y clubes peruanos es significativa.

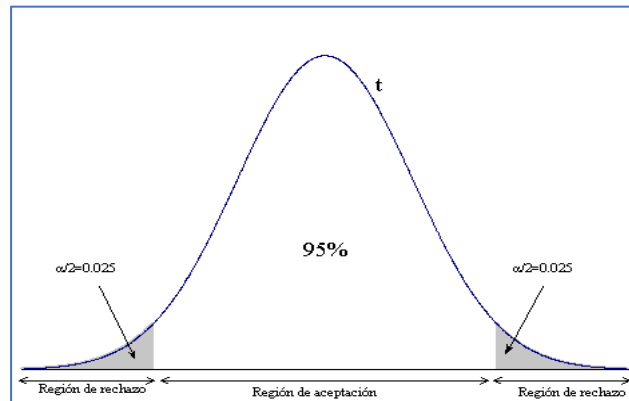
$H_0$  = La técnica scraping del plugin webscraper NO permite recuperar información para el análisis de los datos de jugadores y clubes peruanos es significativa

- ✓ Nivel de significancia

Si  $p \geq 0.05$ , Aceptamos la  $H_0$  y rechazamos la  $H_A$

Si  $p < 0.05$ , Rechazamos la  $H_0$  y aceptamos la  $H_A$

- ✓ Determinar la región de aceptación y rechazo de la hipótesis nula



## 2. MARCO TEÓRICO

### 2.1. Antecedentes

Antecedentes Nacionales:

- En esta tesis doctoral de la Universidad San Ignacio de Loyola, deja claro sobre los diversos problemas del fútbol peruano, como la crisis económica de los clubes deportivos, la mala gestión organizacional de los clubes deportivos, la falta de crecimiento institucional de los clubes del fútbol peruano a nivel local, nacional y por la tanto internacional, afectando de sobremanera cualquier iniciativa deportiva de cualquier club nuevo o de un pasado histórico, donde identifica las diversas variables que pueden afectar a los clubes deportivos. El tesista propone construir una estrategia de neuromarketing y posicionamiento de los clubes deportivos a nivel internacional y así generar interés de nuevos socios, hinchas y sponsor publicitarios, que puedan hacer crecer económicamente a un club deportivo. (YALE RIOS, 2021)

Antecedentes Internacionales:

- El siguiente trabajo corresponde a la Universidad del País Vasco, donde el autor propone recoger grandes datos, con el objetivo de proporcionar visualmente un panel de control que nos ayude a interpretar de manera

resumida, concisa y simple, y a la vez puedan lograr una mejor toma de decisiones. Una de las finalidades del autor es que el panel que se logre a desarrollar, se actualice constantemente de manera autónoma brindando información en tiempo real al momento de establecer conexión con la fuente de datos empleada, a su vez, también se desarrollaron habilidades creativas cuando relacionaron un conjunto de datos con el fin de conseguir que el mayor número de usuarios puedan interpretar sin mayor dificultad los indicadores del panel. Por último, proponen una mejora para la implementación de base de datos automatizado, utilizando procesos de extracción tales como los cubos OLAP o sistemas inteligentes como el Data Mining, para que así se pueda aplicar la analítica predictiva que les permita completar de manera satisfactoria un sistema de inteligencia de negocios. (Neva, 2018-2019)

- El siguiente trabajo es una memoria de la Universidad de las islas de Baleares, donde la investigación se centra en recolectar datos de forma masiva, con el fin de explorar las herramientas del web scraping y la implementación de la misma para el análisis de datos. La investigación da a conocer los diferentes framework para extraer dicha tarea, como la librería scrapy de Python. El proceso para obtener dicha información es obtenerlo mediante un código generado en el módulo scraper (Spider), y una vez finalizado este proceso, lo enviara los datos a una API para ser procesada, sirviendo como nexo entre el módulo scraper y el módulo de almacenamiento. Finalmente, una vez procesada dicha data y convertida data en información, será almacenada en una base de datos para su posterior uso, Adicional a ello hay un módulo de control, que se encargará de la ejecución periódica del scrapear de manera automática. (Vicente Stenhouse, 2018)
- Este artículo fue realizado por la empresa española Freedom & Flow Company, el cual basa su estudio en utilizar herramientas tecnológicas como wearable devices (pulseras de actividad, relojes para correr, medidores de estrés, etc.) para obtener datos, analizarlos y usarlos en la productividad individual de las personas en una empresa, la reducción de

bajas laborales y el retorno de inversión, generando una tasa positiva de productividad del 12%, que se debe a la gestión de las personas a través del Small Data, que no es más que es un conjunto de datos con un volumen y un formato que hacen que los datos sean accesibles, informativos y procesables. Al contrario del Big Data donde Gartner dice que un 80 % de la información almacenada mediante el Big Data, no se aprovecha de o es útil para la toma de decisiones para las organizaciones. El almacenamiento como tal, genera un gasto importante para las organizaciones, y que posteriormente esa información no pueda ser utilizada sería un problema. Ahora si tenemos en cuenta que el 60 % de la información, pierde su valor inmediatamente después de ser extraída, se transforma si no es utilizada en data oscura, y se refieren a “toda aquella información y activos que las organizaciones recolectan diariamente y procesan y almacenan; pero que generalmente no usan para otros fines” (Freedom & Flow Company, 2016)

- El siguiente trabajo es un proyecto de investigación de la Universidad Politécnica de Cataluña, el cual para su desarrollo se investiga sobre las diversas metodologías de Data Mining para selección de datos y también se estudia sobre las diferentes metodologías de machine learning las cuales nos permiten comparar los diferentes comportamientos y comprender mejor cómo funcionan internamente. Hacen uso de la metodología CRISP-DM y lo aplican al proyecto de investigación para establecer la dirección en que va el proyecto. Logran su objetivo con el uso de Python y las técnicas de web scraping para obtener información que solo está disponible en la página web de baloncesto de la Liga Endesa. Una vez obtenido la fuente de datos a través del raspado web y enlazarlo a una base de datos creada, se manipulan y almacenan en una estructura de datos. Finalmente, con la información obtenida se pudieron obtener máximas ventajas de un equipo local para que se lleve la victoria. El estudio logra con satisfacción la finalidad relacionada con las estadísticas del baloncesto y el estudio de las bases de datos, logrando pronosticas si el equipo a estudiar ganara o perderá un encuentro. (Moncosí Linares, 2020-2021)

- Este proyecto de investigación fue desarrollado en la Universidad de Valladolid, donde se estudia sobre el uso del Big Data y las semejanzas del análisis de datos y las diferencias que los identifican, para que así logren comprender la importancia del Big Data, así como también proponer algunas líneas de evolución como la medicina, el periodismo y la agricultura. Con esto se quiere mejorar la vida de las personas y el uso de las técnicas y herramientas de Big Data, para facilitar y mejorar las cosas en la vida de las personas. También se trata sobre conceptos de Business Intelligence o Inteligencia Empresarial, así como sus características, ventajas, relaciones y diferencias con el análisis de datos y el Big Data. Como ultimo abarcan la herramienta de análisis de datos de Power BI, Click Sense entre otras, para el análisis de la gestión deportiva, e implementando la visualización y representación de los resultados en informes generados de manera fácil de entender, donde lo visual destaca por encima del resto. (De Miguel Toquero, 2020)

## **2.2. Definiciones**

Para el desarrollo del trabajo, es conocer y describir el campo de conocimiento sobre el que se va a trabajar.

### **2.2.1. Web Scraping**

Según la web Aukera, “El web scraping consiste en navegar automáticamente una web y extraer de ella información. Esto puede ser muy útil para muchísimas cosas y beneficioso para casi cualquier negocio. A día de hoy, no creo que exista una sola empresa de éxito que no lo haga —o que no quiera hacerlo—. De hecho, la empresa reina del scrapeo es Google, que para que su buscador funcione así de bien tiene que estar constantemente scrapeando la red entera” (AUKERA, 2021)

#### **Uso común del web scraping:**

- Consolidar contenido o información comercial: Las diversas empresas en el internet, crean diferentes ofertas en la red, y una forma de utilizar el web scraping, es agrupar toda esa información de

las diferentes ofertas en la red, en un solo sitio comercial y funcional. Un ejemplo de ello es trivago.

- Popularidad en el internet: el web scraping permite usar, el análisis del uso de dato de las redes sociales como, por ejemplo, el análisis de sentimientos mediante técnicas de ciencia de datos o Data Science y usarlos en otros sistemas, para realizar un estudio social o tendencia en las redes sociales.
- Identificar tendencias digitales: Como se mencionó en el párrafo anterior, al usar el web scraping en las redes sociales y analizar esos datos, permite identificar tendencias de que le gusta a la gente en la red, como productos,
- influencers, marcas comerciales, en un tiempo determinado, información que puede ser usada para marketing y generar campañas comerciales.
- El mejor precio: las empresas pueden usar el web scraping, para identificar que empresas de la competencia están ofreciendo el mejor precio en tiempo real. Esto debería permitir a la empresa competidora ajustar sus campañas comerciales al cliente final.
- Monitorear la competencia: Como hablamos del punto anterior, no solo se puede identificar qué precios ofrece la competencia, sino que productos o servicios ofrecen a su clientela, un marco de referencia indispensable para tomar acciones que puedan beneficiar a una organización.
- Mejorar el comercio electrónico: Como se menciones en los puntos anteriores, con la optimización del mejor precio y los productos que se ofrecen, se puede identificar que productos o servicios, no son ofrecidos por la competencia.

- Análisis de la búsqueda de google: Mediante el web scraping, se puede tener un marco referencial de búsquedas o palabras claves conocidos como keyword y potenciar con herramientas SEO, el posicionamiento de contenido digital en el google, un buscador usado a nivel mundial.

### 2.2.2. Extensión Scaper

Existen diversas técnicas de hacer web scraping, como el por medio de un software especializado como la librería scrapy de Python, o cualquier otro lenguaje de programación, o a través de la plataforma de Google Chrome con Google Sheets, inclusive herramientas de pago como Screaming Frog scraping, pero el que utilizaremos para el desarrollo de esta tesis, será la extensión de Google Chrome llamada scraper. Aquí dejamos el link del plugin:

<https://chrome.google.com/webstore/detail/scraper/mbigbapnjcgaffohmbkdlecaccepngjd>

Scaper:

Es una extensión de Google Chrome que te permite hacer web scraping directamente desde la web.

Te sirve para extraer todo lo que esté similar dentro de una web, imágenes, textos, tablas, etc.

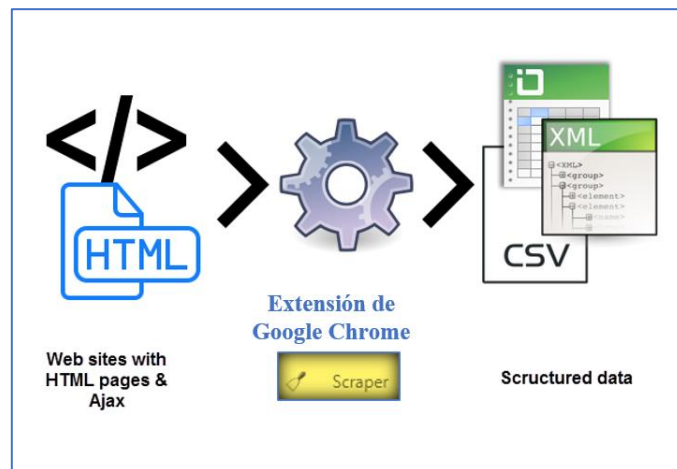


Figura 11 Estructura de funcionamiento de la extensión scraper de Google Chrome



### 2.2.3. HTML

“HTML es un lenguaje de marcado de hipertexto o “Hyper Text Markup Language” por el desarrollo de sus iniciales en inglés, básicamente este lenguaje se escribe en su totalidad con elementos, estos elementos están constituidos por etiquetas, contenido y atributos” (DevCode, 2021) HTML es un markup language, lo que significa que está escrito con códigos que puede leer una persona sin que sea necesario compilarlo primero.

En otras palabras, el texto en una página web está «marcado» con estos códigos para dar instrucciones al navegador web sobre cómo mostrar el texto. Estas etiquetas de marcado son las propias etiquetas HTML (IEBS, 2019)

Lista de etiquetas HTML:

En la creación de cualquier documento HTML, las etiquetas más usadas generalmente son:

<b>Etiqueta</b>	<b>Descripción</b>
<body>	Cuerpo o contenido del documento web
<head>	Cabecera o información de la página
<div>	Dividir contenido
<a>	Hipervínculos o enlaces
<strong>	Formato de texto en negrita
 	Saltos de línea
<H1> ... <H6>	Título del nivel 1 al 6
<img>	Añadir imágenes en un sitio web
<ol>	Sirve para ordenar listas
<ul>	Sirve para listas desordenadas
<li>	Para elementos en una lista

*Tabla 5 Principales Etiquetas HTML*

### 2.2.4. ¿Qué es el Dashboard?

*Un Dashboard es una herramienta de gestión de la información que monitoriza, analiza y muestra de manera visual los indicadores clave de desempeño (KPI), métricas y datos fundamentales para hacer un*

*seguimiento del estado de una empresa, un departamento, una campaña o un proceso específico. (Ortiz, 2021)*

*Es una herramienta que las empresas utilizan para ayudar a rastrear, analizar y visualizar datos, por lo general para obtener una visión más profunda en el bienestar general de la organización, un departamento, o incluso un proceso específico. Detrás de las escenas, cuadros de mando conectar todo tipo de diferentes indicadores, fuentes de datos, APIs y servicios—y ayudan a las empresas a extraer información relevante a partir de esas fuentes y mostrar maneras fáciles de usar. (Microsoft, 2021)*

Podemos decir que el Dashboard, es una herramienta de gestión de la información, que sirve para extraer información relevante de diversas fuentes, para hacer un seguimiento de las diversas actividades empresariales de una empresa y poder manejarlo fácilmente con las diversas fuentes de extracción de datos.

#### **2.2.5. ¿Qué es un dato?**

*Un dato no es otra cosa que una representación simbólica de alguna situación o conocimiento, sin ningún sentido semántico, describiendo situaciones y hechos sin transmitir mensaje alguno. Puede ser un número, una letra o un hecho. (Bulmaro, 2021)*

#### **2.2.6. ¿Qué es el análisis de datos?**

*Se trata de un proceso que consiste en analizar información para extraer datos significativos de un conjunto determinado. Esta técnica de análisis se lleva a cabo con Big data en la mayoría de los casos, aunque lo cierto es que puede aplicarse a cualquier conjunto de datos. (LISA INSURTECH, 2021)*

La aplicación sistemática de técnicas estadísticas y lógicas para describir el alcance de los datos, modularizar la estructura de datos, condensar la representación de datos, ilustrar a través de imágenes, tablas y gráficos, y evaluar inclinaciones estadísticas, datos de probabilidad y derivar conclusiones significativas conocidas como Análisis de datos (Data Analysis, 2022)

Podemos concluir que el análisis de datos, es un proceso para obtener datos significativos, ya sea de una gran fuente de datos o pequeña, pero que buscare condensar su utilidad de forma estructurada y ordenada, a través de tablas, gráficos o datos probabilísticos.

### **2.2.6.1 Características del análisis de datos**

Para una buena gestión centralizada de los datos, se debe contar con ciertas características fundamentales, para que cualquier herramienta pueda ser útil, como lo menciona en la revista byte (Byte TI, 2021)

#### **1. Escalabilidad:**

Lo que uno busca con esta característica es a medida que crezca el volumen de los datos y las fuentes de la misma, pueda aumentar las capacidades de ingesta, procesamiento, y almacenamiento.

#### **2. Colaborativa:**

Esta característica se da cuando hay gran cantidad de datos y profesionales capaces de trabajar juntos y tener un mejor desarrollo de los datos.

#### **3. Integración:**

Es fundamental que los nuevos recursos, puedan integrarse con las nuevas herramientas tecnológicas que aparezcan en el mercado, y puedan ayudar a adaptarse ante los cambios, evitando la obsolescencia.

#### **4. Usabilidad sencilla:**

Se busca mediante esta característica, proveer de una plataforma que pueda aprender y usarse de manera rápida y sencilla, siendo este su valor.

#### **5. Autonomía:**

Esta característica es necesario, ya que el valor comercial, es muy diferente al valor de producción, y por lo tanto debe crearse un espacio diferente para cada propósito, según se requiera en el análisis de datos.

### 2.2.6.2 Tipos de análisis de los datos

A partir de la categoría de análisis descriptivo hasta el prescriptivo, aumenta la complejidad y el esfuerzo de la evaluación de datos, pero también el valor añadido para la empresa.

A continuación, se enumera los tipos de análisis de datos:

#### 1. **Análisis descriptivo:** What happened (Lo que sucedió)

El método de análisis descriptivo es el punto de partida de cualquier reflexión analítica, y tiene como objetivo responder a la pregunta de ¿qué sucedió?

Lo hace ordenando, manipulando e interpretando datos sin procesar de diversas fuentes para convertirlos en información valiosa para su organización.

#### 2. **Análisis exploratorio:** How to explore data relationships (Cómo explorar las relaciones de datos)

Como su nombre indica, el objetivo principal del análisis exploratorio es explorar. Antes de eso, todavía no existía la noción de la relación entre los datos y las variables. Una vez que se investigan los datos, el análisis exploratorio le permite encontrar conexiones y generar hipótesis y soluciones para problemas específicos. Un área de aplicación típica de TI es la minería de datos.

#### 3. **Análisis de diagnóstico:** Why it happened (Por qué sucedió)

El análisis de datos de diagnóstico capacita a los analistas y ejecutivos al ayudarlos a obtener una comprensión contextual firme de por qué sucedió algo. Si sabe por qué sucedió algo y cómo sucedió, podrá identificar las formas exactas de abordar el problema o el desafío.

#### 4. **Análisis predictivo:** What will happen (Lo que sucederá)

El método predictivo le permite mirar hacia el futuro para responder a la pregunta: ¿qué sucederá? Para ello, utiliza los resultados de los análisis descriptivos, exploratorios y diagnósticos mencionados anteriormente, además del aprendizaje automático (ML) y la

inteligencia artificial (IA). De esta manera, puede descubrir tendencias futuras, problemas potenciales o ineficiencias, conexiones y bajas en sus datos.

#### **5. Análisis prescriptivo: How will it happen (cómo sucederá)**

Otro de los tipos de métodos de análisis más efectivos en la investigación. Las técnicas de datos prescriptivos se cruzan con el análisis predictivo en la forma en que gira en torno al uso de patrones o tendencias para desarrollar estrategias comerciales prácticas y receptivas.

#### **2.2.6.3 Obtención de los datos**

El proceso para la gestión de datos, se tiene que incluir una diversidad de fuentes como bases de datos, sistemas ERP, sensores, gadgets, sitios webs, archivos csv, excell, json, pdf, entre otros. La obtención de esos datos, puede estar contenida de manera estructura o no estructurada. Posteriormente se debe homogenizar los datos para la siguiente fase del análisis.

#### **2.2.6.4 Almacenamiento de los datos**

*“Aunque William Harvey Inmon es considerado como el padre del Data Warehouse, fueron unos investigadores del IBM quienes le dieron este nombre.*

*Fue creado en la década de los 90, y consiste en un conjunto de datos almacenados que pueden ser consultados por las empresas mediante tecnologías como la Minería de Datos. [...] Estos datos sirven de apoyo para tomar decisiones y se encuentran ordenados, organizados por temas, son temporales y no volátiles.” (Toquero, 2020, pág. 61)*

En el almacenamiento de datos, es necesario conocer de un mercado de datos, que se puede describir de la siguiente manera:

*Son un subconjunto de un Data Warehouse orientado al análisis, almacenamiento e integración de los datos en un departamento o área de la empresa. Posee la misma complejidad que el Data Warehouse pero se estructura de manera diferente, ya que cada departamento*

*tiene unas necesidades distintas y una forma de gestionar y aplicar la información.” (Toquero, 2020, pág. 62)*

#### **2.2.6.5 Segmentación de los datos**

La gran cantidad de información recopilada para una investigación, puede ser muy grande y por eso debemos llevar a cabo una segmentación de datos. A continuación, vamos a proponer un ejemplo en el ámbito futbolístico, el cual detallamos de la siguiente manera:

Al finalizar una temporada de fútbol de cualquier país, se comienza a negociar sobre los fichajes de comprar y vender jugadores según sus necesidades de los interesados. Para ello la dirección técnica deberá obtener datos deportivos, tanto de los que juegan en tu localidad como de los que juegan en otras ligas extranjeras.

Ahora una vez obtenidos los datos, se tardaría mucho tiempo en analizar todos los parámetros de todos los jugadores, por eso razón se tendría que segmentar el análisis de datos. Para ello, se filtrarán los datos, parámetros o indicadores que nos interesa valorar, siendo este un filtro con respecto al resto de datos.

*Si un equipo ha encajado muchos goles durante una temporada, seguramente les interesará fichar a buenos defensas. Teniendo esta necesidad, ¿les interesará valorar a todos los jugadores de la liga o, incluso, a todos los defensas de la liga? La respuesta es evidente; no. (Toquero, 2020, pág. 64)*

Como se citó en el párrafo anterior podemos decir que, usando herramientas adecuadas se podrá filtrar la información de todos los jugadores como, los que tengan buenos datos de despegue de balones, robos de balón, pases acertados, faltas deportivas, etc., De esta manera la dirección técnica, puede centrarse en valores mencionados anteriormente para después valorar aspectos económicos u algún criterio deportivo de cuál es el mejor jugador para fichar.

#### **2.2.6.6 Proceso del análisis de datos**

Para el proceso del análisis de datos, primeramente, se tendrá que obtener el dato en crudo, mediante una recolección, para posteriormente

procesarlo y normalizar los datos. En la siguiente figura se representa de una mejor manera.

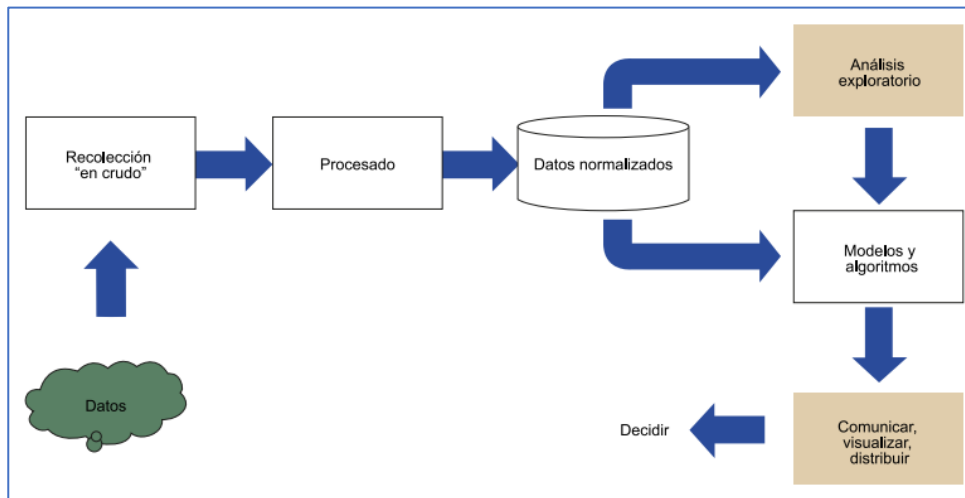


Figura 12 El proceso de análisis y visualización de datos

La figura 1 muestra el proceso de análisis y visualización del big data, y cómo el análisis exploratorio y la visualización (marcados como cuadros grises) se engloban en este. Por una parte, los datos a analizar proveen de diferentes fuentes, y están en diferentes formatos. A veces incluso serán incompletos o contradictorios. Es por ello que, tras la recolección de los datos “en crudo”, será necesario procesarlos y almacenarlos en formato normalizado.

(Macias LLoret, Julbe, & Nin, 2017)

### 2.2.6.7 Importancia del análisis de datos

Se debe entender el potencial, que un análisis de datos puede aportar en una organización. Siguiendo esta pauta nombraremos tres importantes:

#### 1. Toma de decisiones informada:

Desde una perspectiva de gestión, puede beneficiarse del análisis de sus datos, ya que le ayuda a tomar decisiones basadas en hechos y no en una simple intuición. Por ejemplo, puede comprender dónde invertir su capital, detectar oportunidades de crecimiento, predecir sus ingresos o abordar situaciones poco comunes antes de que se conviertan en problemas. De esta manera, puede extraer información relevante de todas las áreas de su organización y, con la ayuda del

software de tablero, presentar la información de una manera profesional e interactiva a las diferentes partes interesadas.

## **2. Reducción de costes:**

Otro gran beneficio es la reducción de costes. Con la ayuda de tecnologías avanzadas como el análisis predictivo, las empresas pueden detectar oportunidades de mejora, tendencias y patrones en sus datos y planificar sus estrategias en consecuencia. Con el tiempo, esto le ayudará a ahorrar dinero y recursos en la implementación de las estrategias equivocadas. Y no solo eso, al predecir diferentes escenarios, como las ventas y la demanda, también puede anticipar la producción y el suministro.

## **3. Dirigirse mejor a los clientes:**

Los clientes son posiblemente el elemento más crucial en cualquier negocio. Mediante el uso de análisis para obtener una visión de 360° de todos los aspectos relacionados con sus clientes, puede comprender qué canales utilizan para comunicarse con usted, sus datos demográficos, intereses, hábitos, comportamientos de compra y más. A la larga, impulsará el éxito de sus estrategias de marketing, le permitirá identificar nuevos clientes potenciales y evitará desperdiciar recursos en dirigirse a las personas equivocadas o enviar el mensaje equivocado. También puede realizar un seguimiento de la satisfacción del cliente analizando las reseñas de sus clientes o el rendimiento de su departamento de servicio al cliente.

### **2.2.7. Campos de aplicación en el análisis de datos**

#### **2.2.7.1 En el sector Económico**

*Hay 3 puntos importantes sobre los que se basa dicha relación: la democratización de los datos, las necesidades de los usuarios de personalización en las comunicaciones e interacciones con las marcas y las capacidades que están involucradas en el almacenamiento, procesamiento y análisis de datos que hacen que éstos puedan usarse*



*correctamente como insumo para la toma de decisiones. (López Vanegas, 2020)*

**La democratización de los datos**, el cual tomo fuerza debido a la crisis del covid-19, generando un cambio en las ventas y en las tendencias de consumo a nivel mundial. Generando así una liberación de datos al acceso público.

**Necesidades de los usuarios de personalización**, este concepto se a desarrollado de una manera abismal, debido a la explosión de las redes sociales. El digitalizar los datos, no solo ahorra tiempo, sino costes de transporte, ya que productos comercializados por las redes sociales o cualquier plataforma online, generan datos que pueden ser procesados y analizados, para generar una rápida segmentación del mercado, inclusive un perfil en función de las preferencias del consumidor.

**Capacidades alrededor de la Ciencia de Datos**, la conversión de los datos a información a través de un análisis descriptivo, se convierte en conocimiento para tomar decisiones con análisis predictivo, apoyado en la habilitación tecnológica para un monitoreo y evaluación de metas.

*Pero la evolución analítica de las empresas tiene que ir, no migrando porque pasar de un modelo de madurez analítica predictiva omite la descriptiva, pero si debe ir en una escala donde primero se madura lo descriptivo para pasar a lo predictivo, luego a lo prescriptivo (que por ejemplo son los simuladores de crédito) y finalmente el conjuntivo (procesamiento del lenguaje natural para integrarlo con los demás análisis). (López Vanegas, 2020)*

#### **2.2.7.2 En el campo de la medicina**

La generación de nuevas ideas para impulsar la investigación médica, se basa en la convergencia de la ciencia y la tecnología, y permitir un avance en la toma de decisiones, como fue la pandemia mundial. Aquí un ejemplo:

Derk Arts, fundador y CEO de Castor EDC, es una plataforma de captura de datos con sede en los Países Bajos, se dio cuenta de este problema mientras trabajaba en medicina.

Arts se sintió frustrado con los sistemas, por los procesos obsoletos y la falta de visión de futuro para los datos de los pacientes. Inventamos la plataforma Castor con una misión clara: "Queremos permitir que cada investigador capture datos para impulsar una mejor toma de decisiones."

*Más de 50.000 investigadores de 90 países utilizan el Castor en sus investigaciones. La plataforma de Castor ha apoyado más de 4.100 estudios comerciales y académicos que cubren una amplia gama de áreas de enfermedades, como diabetes, enfermedades cardiovasculares, enfermedades raras y oncología. Alteryx Analytic Process Automation Platform™ se utiliza para acceder, validar y transformar de forma segura grandes cantidades de datos de ensayos tradicionales y descentralizados. Los investigadores médicos obtienen acceso a los datos en paneles interactivos para que puedan compartir fácilmente información con el personal médico. (Arts, 2022)*

### **2.2.7.3 En el manejo de recursos humanos**

Haciendo el uso de un software de inteligencia artificial, se analiza miles de datos que ayudan a los directivos de recursos humanos a elaborar informes para un control de rendimiento general.

En el 2020, el teletrabajo se convirtió en una forma habitual de trabajo, lo que conlleva un nuevo desafío para la gestión del área de recursos humanos en las organizaciones. Este nuevo escenario laboral del llamado teletrabajo, se buscaba una comunicación y colaboración efectiva, un nuevo paradigma había nacido.

Un ejemplo de ello es la utilización de la herramienta Talentia HR Intelligence.

*Talentia Software ofrece Talentia HR Intelligence para la herramienta HCM, un software de gestión de RR. HH., que pone al alcance de los responsables de Recursos Humanos un sencillo sistema de análisis de datos en tiempo real. Es una herramienta perfecta para crear informes precisos. Así podremos orientar el futuro de la organización hacia el*

*destino óptimo, con la consulta y análisis de información histórica y puntual. Esto permite al director de RR. HH. determinar tendencias en resultados. (Sánchez, 2021)*

#### **2.2.7.4 En el campo deportivo**

Las diversas empresas de cualquier rubro o tamaño, no están exentas de utilizar el análisis de datos, para ayudarse en el proceso de toma de decisiones incluso en el sector deportivo. Aunque ahora se ha vuelto una fórmula ganadora en equipos mayormente europeos, al incluir la tecnología y el análisis de datos en las disciplinas deportivas.

Antes de los años 90, la fórmula para analizar un juego, era ver los videos de los partidos jugados y hacer sus anotaciones en pizarras acrílicas con plumón, o una hoja de papel y lápiz, todo eso fue cambiando en los años 90 el Club de Fútbol Manchester United, decidió usar la analítica para la toma de decisiones. Seguidamente para el 2010, los principales equipos de talla internacional comenzaron a usar la analítica de datos.

*Hablando de deporte y más concretamente del fútbol, este experimentó rápidos avances tecnológicos, con plataformas capaces de capturar y analizar datos de entrenamientos, partidos, dispositivos de Internet de las cosas (IoT) y wearables. Los entrenadores ahora dependen en gran medida de las métricas para guiar su toma de decisiones y ayudar a sus equipos a sobresalir. Los equipos de fútbol se basan en grandes cantidades de datos extraídos de innumerables fuentes para llevar su juego al siguiente nivel. Sensores de IoT y los dispositivos GPS rastrean los movimientos del jugador y la pelota en tiempo real. El seguimiento óptico puede incluso señalar la posición de los jugadores en el campo 25 veces por segundo, en relación con el balón, la oposición y los compañeros de equipo. (KIO NETWORKS, 2022)*

Un ejemplo del uso del análisis de datos es en los entrenamientos, ya que se utiliza dispositivos portátiles que miden la carga de trabajo, el movimiento y los niveles de fatiga de los jugadores para administrar su

condición física, posicionamiento y optimizar su rendimiento durante el juego. A su vez esos datos recopilados se utilizan para diseñar planes de entrenamiento personalizados.

Posteriormente con un almacenamiento en la nube de los datos, se combinan esos grandes volúmenes y usando la inteligencia artificial, permitía un análisis de datos de muchas fuentes para medir e incluso predecir el rendimiento del equipo.

Con la ayuda del análisis predictivo de los datos, se busca ayudar a los entrenadores comprender las diversas formaciones de equipos y las consecuencias, lo que puede ayudar al entrenador adaptar cual es el mejor entrenamiento.

*Una métrica que los equipos han descubierto que tiene poder predictivo son los goles esperados (xG), que mide la calidad de los tiros de los jugadores en el juego de ataque y la probabilidad de que los tiros resulten en goles. xG utiliza algoritmos que tienen en cuenta factores como la distancia desde el objetivo, los ángulos y más. (KIO NETWORKS, 2022)*

#### **2.2.8. El futuro del análisis de datos en el fútbol**

En el 2021 la FIFA en la copa árabe, realizó un análisis de cada jugador en cada segundo del partido, y que buscaba aumentar la competitividad a través de la tecnología.

Para esta copa árabe se asignó a cada jugador un analista, el cual supervisa, rastrea y codifica su rendimiento individual, incluyendo movimientos de un jugador con y sin balón. Este análisis intensivo recopila más de 15000 puntos de datos de cada partido, dando a los analistas de equipo, entrenadores, jugadores, aficionados y periodistas, nuevas perspectivas que ayuden a entender el juego.

*La FIFA quiere proporcionar a los directores técnicos, entrenadores y analistas de rendimiento las mejores perspectivas posibles que los ayuden a desarrollar el talento a nivel mundial, pero para ello, la FIFA necesita entender primero la situación actual del fútbol, tanto dentro como fuera del campo. (FIFA, 2022)*

Este estudio del análisis de los datos, podría ser valioso y darle una ventaja a los pequeños países que no tengan acceso a esta información.

*“A fin de cuentas, esa es la visión de la FIFA: hacer los Mundiales más competitivos”, continuó. “Si podemos dar información... y eso es lo que estamos haciendo en este momento con las naciones más pequeñas; y ayudarlas a desarrollar sus países para reducir la brecha... eso es lo que quiere la FIFA. Esa es su visión”. (FIFA, 2022)*

Para la copa mundial CATAR 2022, la FIFA proporcionara a cada equipo datos esenciales a cada una de las 32 selecciones, incluido las recomendaciones, recursos y perspectivas, a los expertos técnicos, entrenadores y jugadores.

### **2.2.9. Framework SCRUM**

“Scrum es un conjunto de buenas prácticas y reglas que proveen la estructura necesaria para ayudar al equipo a reducir la complejidad técnica, comercial e interpersonal del desarrollo del software en su conjunto y que, también capacita a equipos auto organizados para alcanzar sus objetivos” (RobertoTouza, 2021)

#### **Roles en el SCRUM:**

Los roles en Scrum, se encargan de la construcción de un software de calidad, y la gestión del mismo. Los roles son 3 y son:

- El Scrum Master: Es la persona encargada de guiar al equipo Scrum, para que cumpla el proceso de metodología Scrum y sus reglas. Debe gestionar adecuadamente las herramientas tecnológicas y crear un ambiente adecuado para el desarrollo y operatividad de los proyectos Scrum, y solucionar cualquier requerimiento para el producto. También debe trabajar de la mano con el dueño del proceso y asegurarse que hay una comprensión en el desarrollo del producto de software.
- El Product Owner: Representa a los clientes que usaran el software a construir, según los requerimientos del giro del negocio. Él también

está encargado de plasmar los requerimientos del proyecto al Scrum Master, para que él lo pueda dar a entender al equipo Scrum.

- Él es un puente entre el Scrum master y el equipo Scrum.
- El equipo Scrum: Profesionales con diversas habilidades y conocimientos técnicos, para desarrollar cualquier proyecto, que está plasmado en historias de usuario. Ellos se encargan de presentar avances del desarrollo, por cada sprint iniciado.

### El proceso SCRUM:

Para el desarrollo del proceso Scrum, empieza con la creación del Sprint, el cual mediante una reunión el dueño del proceso (Product Owner PO) presenta las historias en un elemento llamado Product Backlog en un orden prioritario de elaboración. Seguidamente el equipo Scrum selecciona la cantidad de historias que puede realizar de ese Sprint, para posteriormente el equipo Scrum, decidir cómo lo van a realizar. El siguiente evento sería el Sprint Planning, donde se inferirá como se desarrollará las historias de usuario del Product Backlog, donde también se creará el Daily Scrum, que no es más que reuniones de 15 minutos para retroalimentar las tareas realizadas del día anterior por el equipo Scrum, para saber si están siguiendo los requerimientos del Product Owner. Seguidamente se genera un Print Review que inspecciona los resultados, para generar el Print Retrospective, que se encarga de ver la calidad del producto. Todo esto es seguido por el Master Scrum, que fomenta un entorno adaptivo y de guía para el equipo Scrum.



Figura 13 Proceso del Framework Scrum

## 2.2.10. La herramienta Power BI

Power BI es un servicio gratuito de análisis de negocio basado en la nube y visualización de datos de negocio. Esta herramienta de Business Intelligence (BI), incorporada en la suite de productividad Microsoft Office 365, permite controlar la salud de un negocio mediante un dashboard en vivo, crear informes interactivos con Power BI Desktop y acceder a los datos en cualquier lugar con las aplicaciones nativas de móvil. (Armetrics, 2021)



Figura 14 Ejemplo de un Dashboard de Segmentación de mercados

Para un entendimiento de la herramienta de software en el manejo de la información a procesar, detallaremos de forma general sus elementos y las fases de la misma.



Figura 15 Interface de Power Bi Desktop

## 2.2.11. Otras Herramientas existentes

### 2.2.11.1 Pentaho

*Pentaho Data Integration (PDI), también conocido como Kettle, forma parte de la suite multiplataforma de inteligencia empresarial Open Source Pentaho. Incluye todo tipo de herramientas que están dirigidas principalmente a mejorar el proceso de toma de decisiones empresariales: software de gestión de almacenes de datos, herramientas de integración y análisis de datos, software para gestores y herramientas de minería de datos. (CE Entel, 2021)*



Figura 16 Ejemplo de la herramienta Pentaho

### 2.2.11.2 Tableau

*Es una herramienta de visualización de datos interactiva, es decir, el usuario tiene la posibilidad de interactuar con los datos: comparar, filtrar, conectar unas variables con otras, etc. Además, los informes y dashboards que se pueden crear con la herramienta son muy visuales lo que facilita la comprensión rápida de los datos. (Neteris, 2019)*

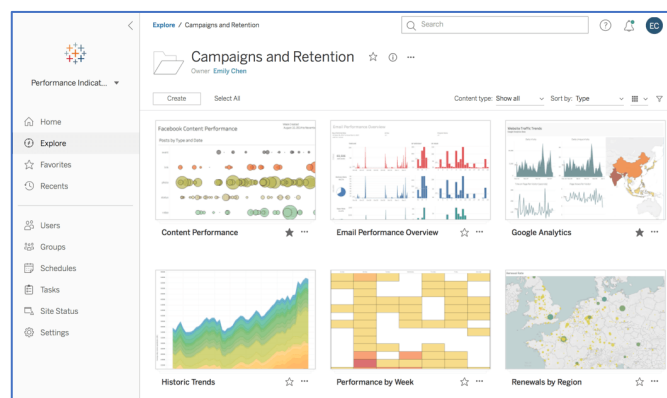


Figura 17 Ejemplo de la herramienta Tableau, en una campaña y retenciones



### 2.2.11.3 IBM Watson Analytics

IBM Watson Analytics es un servicio, basado en la nube, de inteligencia artificial. Su función es permitir a las empresas descubrir, analizar y visualizar, de una forma inteligente, datos e información relevante para descubrir patrones o tendencias y para ayudar a la toma de decisiones.

Dos de los principales puntos fuertes de IBM Watson Analytics son **el análisis**

**predictivo y la narración visual.** Abarca muchas categorías, como ventas, operaciones, finanzas, atención al cliente, recursos humanos, etc. En todas ellas

permite analizar y visualizar datos de forma interactiva, buscar de forma automática patrones de comportamiento y analizar las redes sociales. A mayores, permite la creación de cuadros de mando. (Toquero, 2020, pág. 79;80)

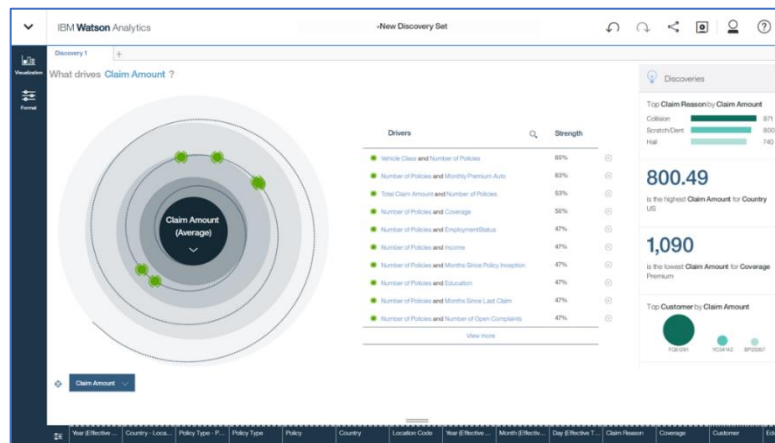


Figura 18 Ejemplo de la herramienta IBM Watson Analytics

### 3. MATERIALES Y METODOS

#### 3.1. Material

##### 3.1.1. Población

La investigación es realizada a los datos históricos de los clubes profesionales de la liga uno del futbol peruano

##### 3.1.2. Muestra

Portales deportivos online como transfermarkt y flashscore, las cuales contienen información detallada de los jugadores profesionales del futbol peruano.

##### 3.1.3. Unidad de análisis

Datos sobre los jugadores profesionales del futbol peruano.

#### 3.2. Método

##### 3.2.1. Tipo de investigación

Aplicativo

##### 3.2.2. Diseño de investigación

Se trata de una investigación transversal y por el diseño de contrastación es cuasi experimental

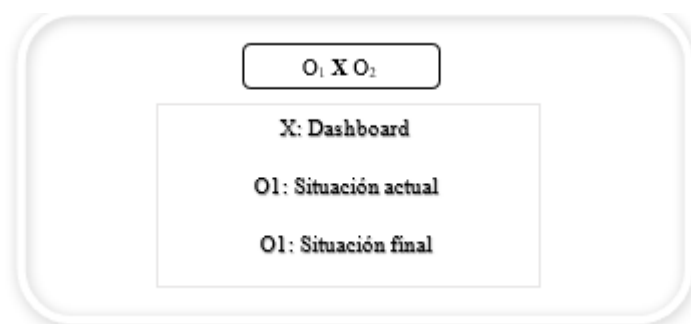


Figura 19 Forma gráfica del diseño de investigación

##### 3.2.3. Variables de estudio y Operacionalización

**VI:** Dashboard y el uso de técnicas de web scraping.

**VD:** Analizar la información deportiva de los jugadores profesionales en la Región La Libertad.

Variable	Dimensión	Indicador	Unidad de medida	Instrumento de Investigación
VI	Dashboard para generar un valor comercial	Nivel de facilidad de usabilidad Tiempo de accesibilidad	Númérico	Evaluación de resultados
VD	Registros	Número de registros que posee la información histórica	Nº Registros	Hoja de registros
	Tiempo de respuesta	Tiempo para analizar información	Intervalo de tiempo(minutos)	Hoja de medición de tiempos de respuesta
	Grado de utilidad	Grado de utilidad (sí o no)	% de acuerdo	Hoja resumen de grados de utilidad de los resultados

Tabla 6 Operacionalización de las variables

#### 3.2.4. Técnicas e instrumentos de recolección de datos

- Técnica de Web Scraping
- Aplicación de Power BI

#### 3.2.5. Técnicas de procesamiento y análisis de datos

Una vez recolectado toda la información proporcionada por los instrumentos de recolección, se analizará estadísticamente los resultados obtenidos de la información.

Según la investigación realizada usaremos la prueba paramétrica t, para corroborar si la muestra evaluada antes y después difieren. Aplicaremos el método descriptivo, el cual se encarga de la descripción de datos y características de una población.

## 4. RESULTADOS

En este capítulo mostraremos los resultados del desarrollo del proyecto de Tesis, en un Dashboard construido con el programa Power BI de Microsoft. Como primera etapa del desarrollo del proyecto, se basó en la **planificación** donde se detallará una lista de requerimientos del proyecto, prioridades, fechas de entrega, personal encargado de dicha tarea y el producto en si en el plazo acordado. Como segunda fase del proyecto se basó en el **desarrollo**, descrito en la etapa anterior de la planificación.

### 4.1. Planificación de la metodología SCRUM

Ahora siguiendo la metodología Scrum, nos ayudaremos de una plantilla en Excel, el cual detallaremos más adelante. A continuación, detallaremos los artefactos en SCRUM que son uno de los ejes principales en el desarrollo de esta metodología

#### 4.1.1. Lista de Producto o Product Backlog

Consta de una lista ordenada de requerimientos, las cuales se plasmarán como historias de usuario y donde se visualizará una información breve y detallada sobre el proyecto que se va a desarrollar.

Esta lista ira evolucionando ante cualquier mejora de ser necesario, para la adecuación del propósito del proyecto. A continuación, una lista de ello:

Prioridad	Id	Historia de Usuario	Valor	Tiempo estimado
1	HU01	Se necesita conocer información deportiva de los diversos jugadores peruanos, como el club que perteneció, su edad en el torneo, goles anotados, asistencias, goles por encuentros y penales ejecutados	15	5 semanas
2	HU02	Se necesita conocer los arqueros con más partidos ganados y penales detenidos	15	2 semanas
3	HU03	Se necesita conocer el club peruano con más faltas deportivas	15	2 semanas
4	HU04	Se necesita conocer el rendimiento de los jugadores en la costa sierra y selva	15	3 semanas
5	HU05	Se necesita conocer el valor monetario de los clubes peruanos en el 2021	10	3 semanas
6	HU06	Se necesita conocer la influencia de los jugadores en el fútbol peruano entre el 2014 y el 2021	10	4 semanas
7	HU07	Se necesita conocer la cantidad de goles marcados de los jugadores entre el 2014 y el 2021	10	3 semanas
8	HU08	Se necesita conocer el top de goleadores desde el 1928 hasta el 2021	10	6 semanas
			100	

Tabla 7 Lista de Producto

#### 4.1.2. Sprint Backlog

El Sprint Backlog es un artefacto que pretende ser una imagen muy visible y en tiempo real del trabajo que se planea realizar, fomentando la transparencia en el desarrollo del framework SCRUM.

Nuestra realización está hecha en una plantilla en Excel por cada uno de la lista del producto, que detallaremos a continuación:

#### Sprint 01:

INFORMACIÓN SOBRE EL SPRINT		1	INFORMACIÓN DE SPRINT ACTUAL	
Campo	Valor			
Fecha de comienzo	05/04/22		Fecha de fin de sprint	04/06/22
Longitud de sprint (bruto)	60		Longitud de sprint (neta)	44
Días de vacaciones	16		Horas disponibles totales	140.8
Tamaño del equipo	2		Suma de puntos de historia	45
Ocupación máxima del equipo	80%			
Jornada de trabajo diaria en horas	2.0		Número de tareas en el sprint	10
Número de Sprint actuales	1		Tareas completadas	10

Tabla 8 Información del Sprint 01

Comienzo del sprint		05-04-22	REPORTE DE RENDIMIENTO DE JUGADORES DESDE EL 2014 - 2021				
Duración del sprint en días		54	SPRINT 1				
Velocidad		80	SPRINT 1				
ID de sprint	ID de Historia de usuario	Puntos historia	Tareas	Asignado a	Estado	Completado en	Día de sprint
1	H01	2	Reunión con los interesados de la información	Vargas	Terminado	5/04/2022	0
1	H01	1	Reunión de equipo de trabajo	Freud	Terminado	6/04/2022	1
1	H01	1	Definir el alcance del proyecto	Vargas	Terminado	7/04/2022	2
1	H01	1	Búsqueda de data de interés	Freud	Terminado	8/04/2022	3
1	H01	1	Scrapeo de las páginas seleccionadas y guardado de la información	Vargas	Terminado	15/04/2022	10
1	H01	2	Modelamiento de la data	Freud	Terminado	18/04/2022	13
1	H01	1	Capacitación de Power BI	Vargas	Terminado	25/04/2022	20
1	H01	2	Creación de reporte de Información deportiva de los jugadores	Freud	Terminado	03/05/2022	28
1	H02	2	Creación de reporte de Información deportiva de los arqueros	Freud	Terminado	17/05/2022	42
1	H03	2	Creación de reporte de falta deportiva de los clubes peruanos	Vargas	Terminado	04/06/2022	60

Tabla 9 Detalle de tareas del Sprint 01

## Sprint 02:

INFORMACIÓN SOBRE EL SPRINT		INFORMACIÓN DE SPRINT ACTUAL	
	2		
Campo	Valor		
Fecha de comienzo	05/06/22	Fecha de fin de sprint	09/07/22
Longitud de sprint (bruto)	34	Longitud de sprint (neta)	24
Días de vacaciones	10	Horas disponibles totales	76.8
Tamaño del equipo	2	Suma de puntos de historia	15
Ocupación máxima del equipo	80%	Puntos historias abiertos	11
Jornada de trabajo diaria en horas	2.0	Número de tareas en el sprint	14
Número de Sprint actuales	1	Tareas completadas	7

Tabla 10 Información del Sprint 02

Comienzo del sprint	05-06-22	<b>Reporte de rendimiento de jugadores en la costa, sierra y selva del 2014 al 2021</b>					
Duración del sprint en días	24						
Velocidad	<b>80</b>						
		<b>SPRINT 2</b>					
ID de sprint	ID de Historia de usuario	Puntos historia	Tareas	Asignado a	Estado	Completado en	Día de sprint
2	H04	2	Reunión de equipo de trabajo	Vargas	Terminado	7/06/2022	2
2	H04	2	Avance de informe del proyecto	Freud	Terminado	14/06/2022	9
2	H04	4	Diseño de reporte de rendimiento de jugadores en la costa, sierra y selva	Vargas	Terminado	26/06/2022	21
2	H04	2	Elaboración de reporte	Freud	Terminado	02/07/2022	27
2	H04	2	Verificación de la información	Vargas	Terminado	05/07/2022	30
2	H04	1	Reunión con el asesor	Freud	Terminado	07/07/2022	32
2	H04	2	Avance de informe del proyecto	Freud	Terminado	09/07/2022	34

Tabla 11 Detalle de tareas del Sprint 02

## Sprint 03:

INFORMACIÓN SOBRE EL SPRINT		INFORMACIÓN DE SPRINT ACTUAL	
	3		
Campo	Valor		
Fecha de comienzo	09/07/22	Fecha de fin de sprint	31/07/22
Longitud de sprint (bruto)	22	Longitud de sprint (neta)	18
Días de vacaciones	4	Horas disponibles totales	57.6
Tamaño del equipo	2	Suma de puntos de historia	10
Ocupación máxima del equipo	80%		
Jornada de trabajo diaria en horas	2.0	Número de tareas en el sprint	6
Número de Sprint actuales	1	Tareas completadas	6

Tabla 12 Información del Sprint 03

Comienzo del sprint	09-07-22	<b>Reporte de valor monetario de los jugadores hasta el 2021</b>					
Duración del sprint en días	18						
Velocidad	<b>80</b>	<b>SPRINT 3</b>					
ID de sprint	ID de Historia de usuario	Puntos historia	Tareas	Asignado a	Estado	Completado en	Día de sprint
2	H05	1	Reunión de equipo de trabajo	Vargas	Terminado	12/07/2022	3
2	H05	1	Avance de informe del proyecto	Freud	Terminado	16/07/2022	7
2	H05	3	Diseño de reporte del valor monetario de los jugadores en los clubes peruanos	Vargas	Terminado	20/07/2022	11
2	H05	3	Verificación de la información	Freud	Terminado	25/07/2022	16
2	H05	1	Reunión con el asesor	Vargas	Terminado	26/07/2022	17
2	H05	1	Avance de informe del proyecto	Freud	Terminado	30/07/2022	21

Tabla 13 Detalle de tareas del Sprint 03

### Sprint 04:

INFORMACIÓN SOBRE EL SPRINT		4	INFORMACIÓN DE SPRINT ACTUAL	
<b>Campo</b>	<b>Valor</b>			
Fecha de comienzo	01/08/22		Fecha de fin de sprint	19/09/22
Longitud de sprint (bruto)	49		Longitud de sprint (neta)	34
Días de vacaciones	15		Horas disponibles totales	108.8
Tamaño del equipo	2		Suma de puntos de historia	20
Ocupación máxima del equipo	80%			
Jornada de trabajo diaria en horas	2.0		Número de tareas en el sprint	7
Número de Sprint actuales	1		Tareas completadas	7

Tabla 14 Información del Sprint 04

Comienzo del sprint	01-08-22	<b>Reporte de goles marcados por los jugadores en los torneos del 2014 al 2021</b>					
Duración del sprint en días	34						
Velocidad	<b>80</b>	<b>SPRINT 4</b>					
ID de sprint	ID de Historia de usuario	Puntos historia	Tareas	Asignado a	Estado	Completado en	Día de sprint
4	H06	2	Reunión de equipo de trabajo	Vargas	Terminado	1/08/2022	0
4	H06	2	Avance de informe del proyecto	Freud	Terminado	2/08/2022	1
4	H06	6	Diseño de reporte de la influencia de los jugadores en el futbol peruano	Vargas	Terminado	15/08/2022	14
4	H07	6	Diseño de reporte de la cantidad de goles marcados en los clubes peruanos	Freud	Terminado	31/08/2022	30
4	H07	1	Verificación de la información	Vargas	Terminado	01/09/2022	31
4	H07	1	Reunión con el asesor	Freud	Terminado	05/09/2022	35
4	H07	2	Avance de informe del proyecto	Freud	Terminado	19/09/2022	49

Tabla 15 Detalle de tareas del Sprint 04

## Sprint 05:

INFORMACIÓN SOBRE EL SPRINT		INFORMACIÓN DE SPRINT ACTUAL	
	5		
Campo	Valor		
Fecha de comienzo	20/09/22	Fecha de fin de sprint	17/10/22
Longitud de sprint (bruto)	27	Longitud de sprint (neta)	19
Días de vacaciones	8	Horas disponibles totales	60.8
Tamaño del equipo	2	Suma de puntos de historia	10
Ocupación máxima del equipo	80%		
Jornada de trabajo diaria en horas	2.0	Número de tareas en el sprint	8
Número de Sprint actuales	1	Tareas completadas	7

Tabla 16 Información del Sprint 05

Comienzo del sprint	20-09-22	Otros reportes de interés deportivo					
Duración del sprint en días	19						
Velocidad	80						
		SPRINT 5					
ID de sprint	ID de Historia de usuario	Puntos historia	Tareas	Asignado a	Estado	Completado en	Día de sprint
5	H08	1	Reunión de equipo de trabajo	Vargas	Terminado	21/09/2022	1
5	H08	1	Avance de informe del proyecto	Freud	Terminado	22/09/2022	2
5	H08	2	Diseño de reporte de los puestos de los clubes peruanos en los diversos torneos	Vargas	Terminado	23/09/2022	3
5	H08	2	Verificación de la información	Freud	Terminado	30/09/2022	10
5	H08	2	Reunión con el asesor	Freud	Terminado	15/10/2022	25
5	H08	1	Verificación de los requisitos del proyecto	Vargas	Terminado	16/10/2022	26
5	H08	1	Avance de informe del proyecto	Freud	Terminado	17/10/2022	27

Tabla 17 Detalle de tareas del Sprint 05

### 4.1.3. Incremento

Después de cada sprint, el equipo realizó una entrega de un incremento del avance del informe y de la funcionalidad del dashboard, este incremento debe estar terminado para poder ser publicado y utilizado al 100%.



## 4.2. Desarrollo de la metodología SCRUM

### 4.2.1. Definir el alcance del proyecto

PROYECTO DE TESIS		
ORGANIZACIÓN	Universidad Privada Antenor Orrego - UPAO	
NOMBRE DEL PROYECTO	Uso de técnicas de web scraping para el análisis de datos de jugadores profesionales del futbol peruano para el periodo 2021	
CLIENTE	Escuela de Ing. Computación y Sistemas	
NÚMERO DE VERSIÓN	0.0.1	
PUNTO DE CONTACTO DEL CLIENTE	044-604444	
INFORMACIÓN DE CONTACTO DE LA ORGANIZACIÓN	aullonr@upao.edu.pe	
NOMBRE	Agustín Eduardo Ullón Ramírez	
TELÉFONO	989744943	
CORREO ELECTRÓNICO	freud.monzon18@gmail.com	
FECHA DE PREPARACIÓN	23/08/2021	AUTOR: M. L.
GERENTE DE PROYECTO		
FECHA DE INICIO	FECHA FINAL	DURACIÓN DEL PROYECTO
03/11/2021	03/11/2022	1 año
EQUIPO DEL PROYECTO		
NOMBRE / TÍTULO	TELÉFONO	CORREO ELECTRÓNICO
Monzón Laiza Freud	996823384	rmonzon1@upao.edu.pe
Vargas Ulloa Italo	960287389	ivargasu@upao.edu.pe
PROYECTO		
INTRODUCCIÓN   Descripción sin detalles del requisito		
Se hace necesario ajustar un uso del análisis de manera menos descentralizada y colaborativa, que pueda beneficiar a cualquier negocio emergente, en un crecimiento de acuerdo a su realidad económica y experiencia digital en el procesamiento de sus datos		
ANTECEDENTES   ¿Qué llevó a la necesidad de este proyecto?		
Uno de los puntos importantes, es obtener información relevante de los jugadores de los diferentes clubes deportivos peruanos, usando técnicas de web scraping y plasmarlo en un dashboard que se pueda analizar		
NECESIDADES DE RECURSOS		
Se usará en hardware 2 laptops, 01 de la marca Lenovo y otra de la marca Asus. También se usará un dominio privado (cibiotec.com) en el cual se publicará el dashboard desde la plataforma de Power BI Desktop.		
RESUMEN DE TARIFAS		
FASE UNO	HORAS	TOTAL
Planificación	57	57
FASE DOS	HORAS	TOTAL
Desarrollo	1160	1217
ALMACENAMIENTO ESTIMADO DE DATOS	2 meses después de su publicación	

Tabla 18 Alcance del proyecto de tesis

## 4.2.2. Búsqueda de la data

Una de las partes fundamentales para el desarrollo del análisis deportivo de los diferentes clubes en el fútbol peruano, es contar con diversas fuentes de las cuales se van a fusionar para una exploración de los datos.

Uno de los problemas con respecto a ello, es la poca información de detalles técnicos, pero con el compromiso de generar una base de datos para sus posteriores usos y sea un referente para este tipo de investigaciones.

Aun así, se pudo encontrar una cantidad de páginas referentes a datos deportivos, los cuales son:

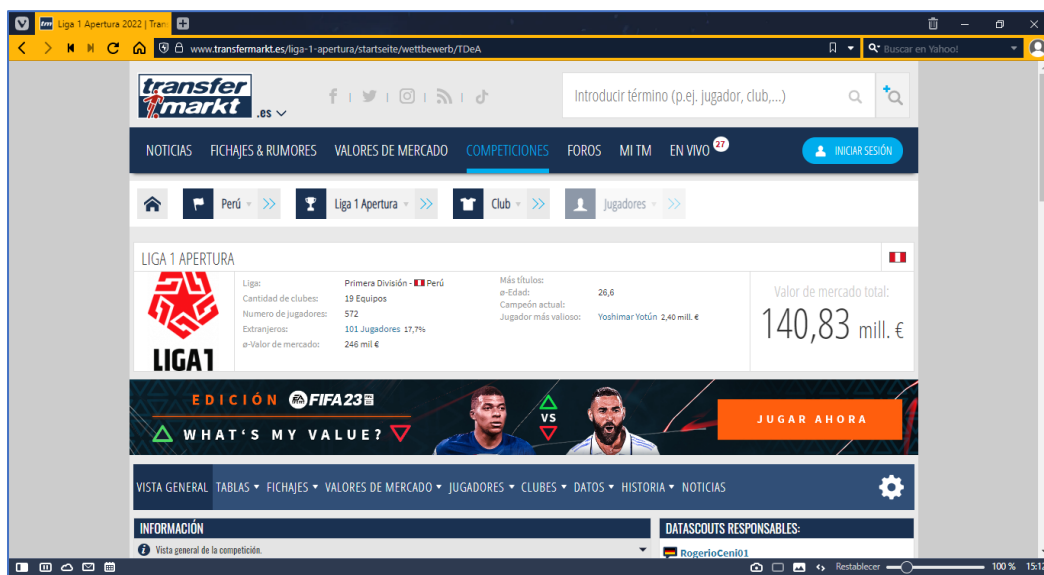


Figura 20 Pagina de Transfermarkt



Figura 21 Pagina de FBREF

Equipo	Títulos	Substitutos	Campeonatos Obtenidos		
			Liga Peruana de Fútbol Lima (1912-1921)	Era Amateur Lima & Callao (1926-1950)	Era Profesional Lima & Callao (1951-1965) Descentralizado (1966-Actualidad)
Universitario de Deportes	26	15		1929, 1934, 1939, 1941, 1945, 1949, 1949	1995, 1995, 1994, 1995, 1997, 1999, 1971, 1974, 1982, 1985, 1987, 1990, 1992, 1993, 1998, 1999, 2000, 2006, 2013
Alianza Lima	24	21	1919, 1919	1927, 1928, 1931, 1932, 1933, 1949	1992, 1994, 1995, 1992, 1983, 1988, 1978, 1977, 1978, 1997, 2001, 2003, 2004, 2009, 2017, 2021
Sporting Cristal	20	14			1998, 1991, 1988, 1970, 1972, 1979, 1980, 1983, 1988, 1991, 1994, 1995, 1999, 2002, 2005, 2012, 2014, 2016, 2018, 2020
Sport Boys	6	9		1935, 1937, 1942	1991, 1998, 1994
Deportivo Municipal	4	8		1938, 1940, 1943, 1950	
Universidad de San Martín	3	0			2007, 2008, 2010
Lima Cricket F. B. C.	2	1	1912, 1914		
Sport José Gálvez (D)	2	0	1919, 1919		
Sport Progreso (D)	2	1	1921, 1929		
Estrella Roja	2	4		1930, 1947	
X. Mercedes Suarez (D)	2	2		1944	1953
Unión Huaral	2	1			1976, 1999
F. B. C. Melgar	2	2			1981, 2015
Jorge Chávez N.º 1 (D)	1	2	1913		
Sport Juan Beroivoce (D)	1	0	1917		
Sport Inca (D)	1	0	1920		
Centro Iquillo	1	0			1957
Defensor Lima	1	0			1973
San Agustín	1	0			1988
Juan Aurich	1	2			2011
Emancipación	1	0			2019

Figura 22 Pagina de Wikipedia

### 4.2.3. Seleccionando la técnica de web scraping

Para esta parte del proyecto de tesis, se seleccionó hacer el web scraping con la extensión de Scraper de Google Chrome, por su facilidad de uso de cualquier usuario, por la rapidez de obtener información y por su costo cero para su utilización, por ser una herramienta gratuita de Google Chrome. Otra de las justificaciones de uso es porque una de las páginas como es Transfermarkt cambia continuamente su formato HTML, tendiendo como consecuencia la inutilización de cualquier código configurado para dicho scraping, aunque eso se podría solucionar con una herramienta de pago, la cual no está al alcance de la mayoría de personas.

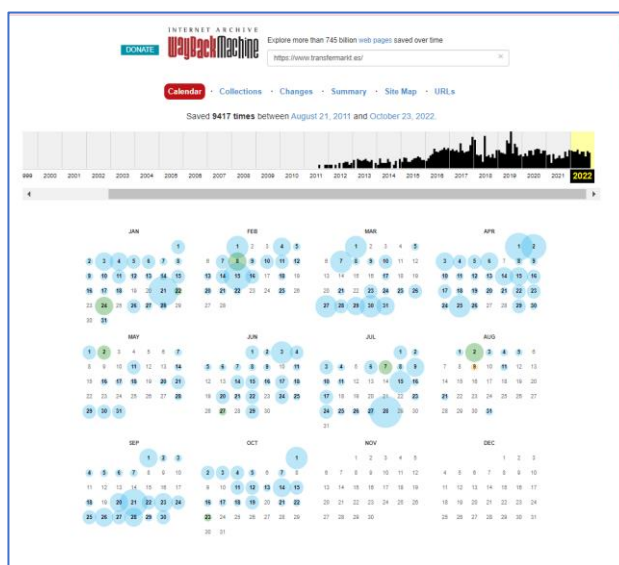


Figura 23 Vista histórica de cambios de la página Transfermarkt en Waybackmachine

#### 4.2.4. Raspado de los datos

Una vez identificadas las páginas anteriormente mencionadas, se procederá a realizar los siguientes pasos:

1. Abre Google Chrome y haz clic en tu Chrome Web Store
2. Busca “Scraper” en las extensiones
3. El primer resultado de búsqueda es la extensión “Scraper”
4. Haz clic en el botón Agregar a Chrome.

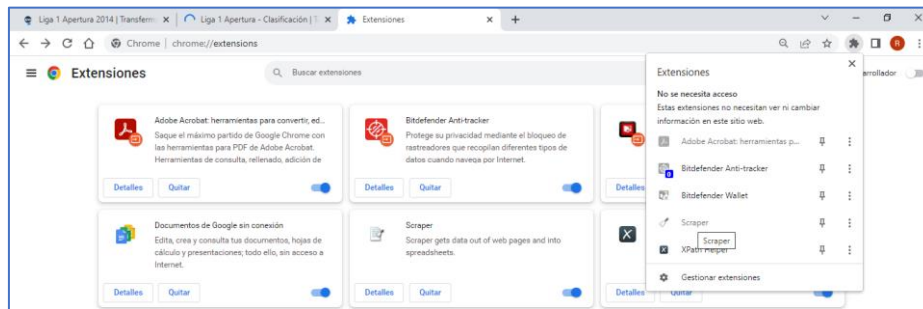


Figura 24 Instalado extensión Scraper de Google Chrome

Ahora, por ejemplo, vamos a abrir [https://www.transfermarkt.es/liga-1-apertura/torschuetzenliste/wettbewerb/TDeA/saison\\_id/2013/altersklass\\_e/alle/detailpos//plus/1](https://www.transfermarkt.es/liga-1-apertura/torschuetzenliste/wettbewerb/TDeA/saison_id/2013/altersklass_e/alle/detailpos//plus/1)

Ahora marca con el ratón seleccionando toda la parte interesada y buscas la opción Scrape similar.

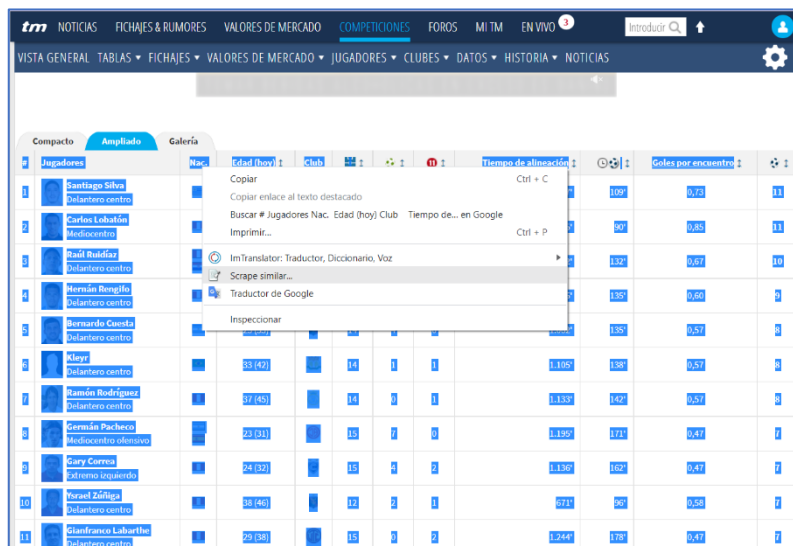


Figura 25 Uso de la extensión Scrape de Google parte 1

Posteriormente al ejecutar la extensión, procederá a captar en una ventana emergente los datos seleccionados. Uno puede copiarlo directamente a una hoja de cálculo, bloc de notas, word o exportarlo a

Google Docs. Adicionalmente puede modificar el XPath, dependiendo de la necesidad del usuario.

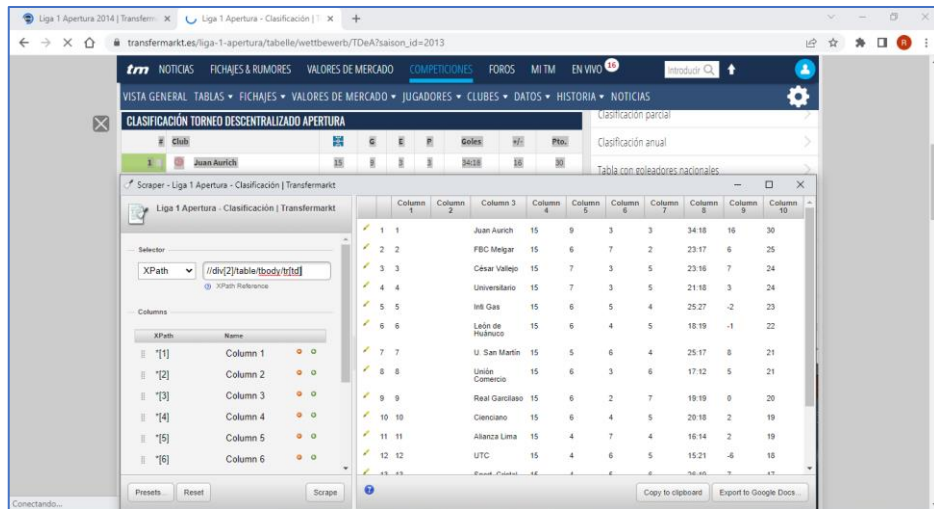


Figura 26 Uso de la extensión Scrape de Google parte 2

Una vez scrapeado la información se procederá a guardar la información en formato csv o extensión. xlsx, según sea más conveniente su obtención para seguir el siguiente proceso de limpieza de datos.

Se detalla a continuación los archivos separados por carpetas, una vez extraída dicha información. Se debe tener en cuenta que la data extraída y a procesar es desde el año 2014, hasta el año 2021 de la Liga 1, que es la liga profesional del fútbol peruano.

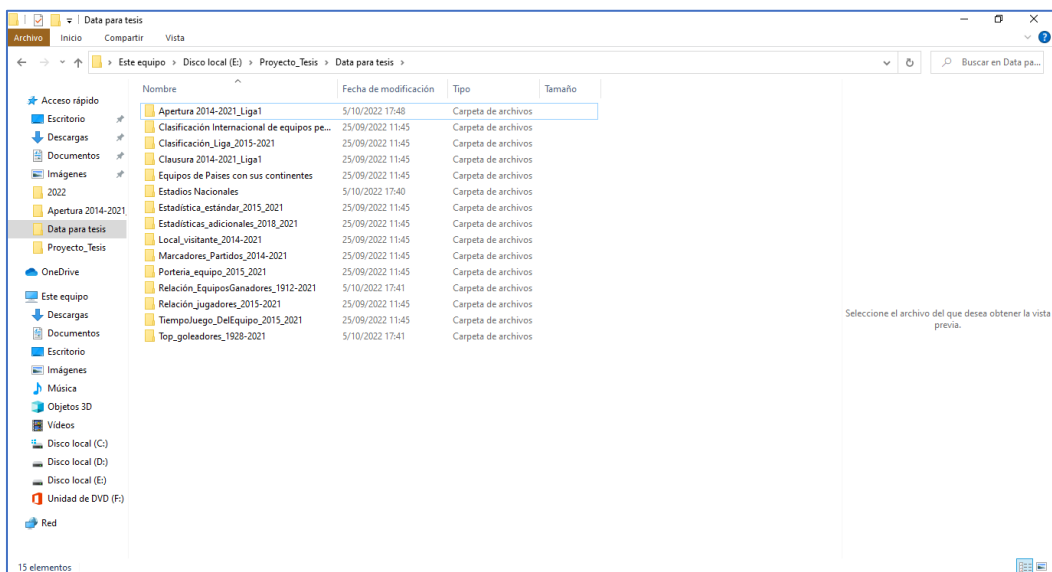


Figura 27 Relación de la información extraída de las páginas mencionadas

## Carpeta Apertura 2014-2021\_Liga1

### Detalle de los archivos obtenidos y la información de los mismos

The screenshot shows a file explorer window displaying a folder named 'Apertura 2014-2021\_Liga1'. It contains 12 Excel files, each representing a tournament stage from 2014 to 2021. An Excel spreadsheet is open, showing a table with columns for 'Torneo', 'Año', 'RL', 'Equipo', and various performance metrics like 'Puntos', 'Goles', 'Asistencias', 'Tarjetas', and 'Notas'.

Torneo	Año	RL	Equipo	PJ	PG	PE	PP	GF	GC	DG	Pts	PJ	Notas
1	Apertura	2021	1 Escudo del Club Sporting Cristal	9	8	0	1	18	6	12	24	2.67	? Final
2	Apertura	2021	2 Escudo del Club USMP	9	6	1	2	10	6	4	19	2.11	? Final
3	Apertura	2021	3 Escudo del Club U César Vallejo	9	5	3	1	17	8	9	18	2	
4	Apertura	2021	4 Escudo del Club Alianza Lima	9	4	4	1	12	6	6	16	1.78	
5	Apertura	2021	5 Escudo del Club Ayacucho	9	3	6	0	16	10	6	15	1.67	
6	Apertura	2021	6 Escudo del Club Universitario	9	4	3	2	12	11	1	15	1.67	
7	Apertura	2021	7 Escudo del Club Cienciano	9	4	2	3	14	12	2	14	1.56	
8	Apertura	2021	8 Escudo del Club Mannucci	9	3	3	3	13	14	-1	12	1.33	
9	Apertura	2021	9 Escudo del Club Sport Huancayo	9	3	3	3	8	9	-1	12	1.33	
10	Apertura	2021	10 Escudo del Club Alianza Atlético	9	3	1	5	10	9	1	10	1.11	
11	Apertura	2021	11 Escudo del Club UTC	9	3	1	5	9	12	-3	10	1.11	

Figura 28 Apertura 2014-2021\_Liga1

## Carpeta Clasificación Internacional de equipos peruanos desde el 2015 hasta el 2021

### Detalle de los archivos obtenidos y la información de los mismos

The screenshot shows a file explorer window displaying a folder named 'Clasificación Internacional de equipos peruanos desde el 2015 hasta el 2021'. It contains 7 CSV files for the years 2015 through 2021. An Excel spreadsheet is open, showing a table with columns for 'Equipo', 'Goleador', 'Resultado', and 'Resultado'. The table lists various international tournament results for Peruvian teams.

Equipo	Goleador	Resultado	Resultado
1	Juan Aurich, M	2015 Copa Libertadores (Second stage)	Por ganar la eliminación directa, 3er lugar en la Second stage
2	Sporting Cristal, M	2015 Copa Libertadores (Second stage)	Por el final de liga, 3er lugar en la Second stage
3	León Huánuco, M	2015 Copa Sudamericana (First stage)	Por el final de liga, Perdió en First stage contra CS Emelec
4	Melgar, M	2015 Copa Sudamericana (First stage)	Por el final de liga, Perdió en First stage contra Junior
5	Unión Comercio, M	2015 Copa Sudamericana (First stage)	Por el final de liga, Perdió en First stage contra Rionegro
6	Universitario, M	2015 Copa Sudamericana (First stage)	Por el final de liga, Perdió en Second stage contra Defensor Sporting

Figura 29 Clasificación Internacional de equipos peruanos desde el 2015 hasta el 2021

## Carpeta Clasificación\_Liga\_2015-2021

### Detalle de los archivos obtenidos y la información de los mismos

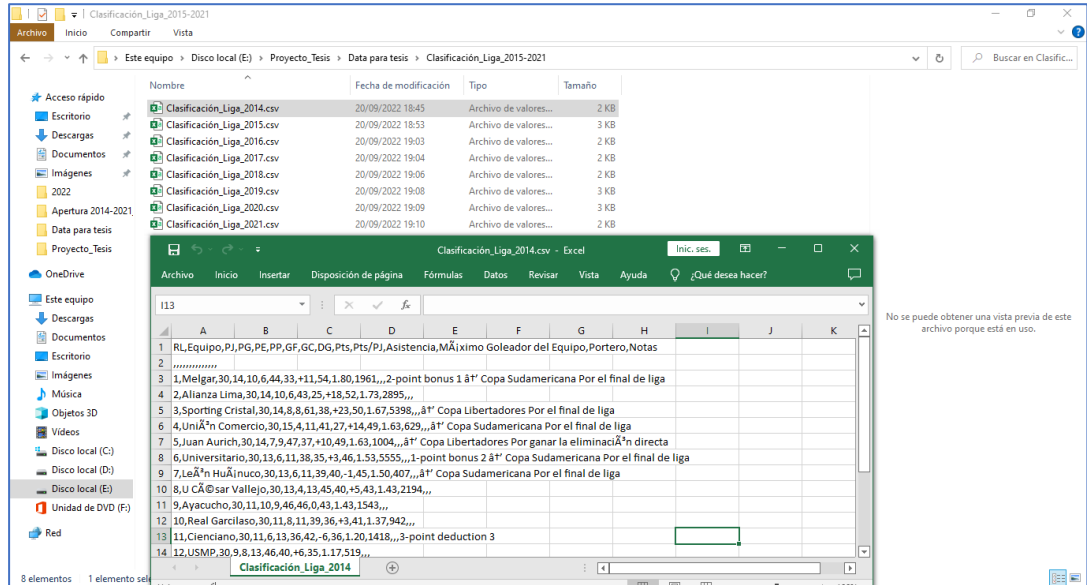


Figura 30 Clasificación\_Liga\_2015-2021

## Carpeta Clausura 2014-2021\_Liga1

### Detalle de los archivos obtenidos y la información de los mismos

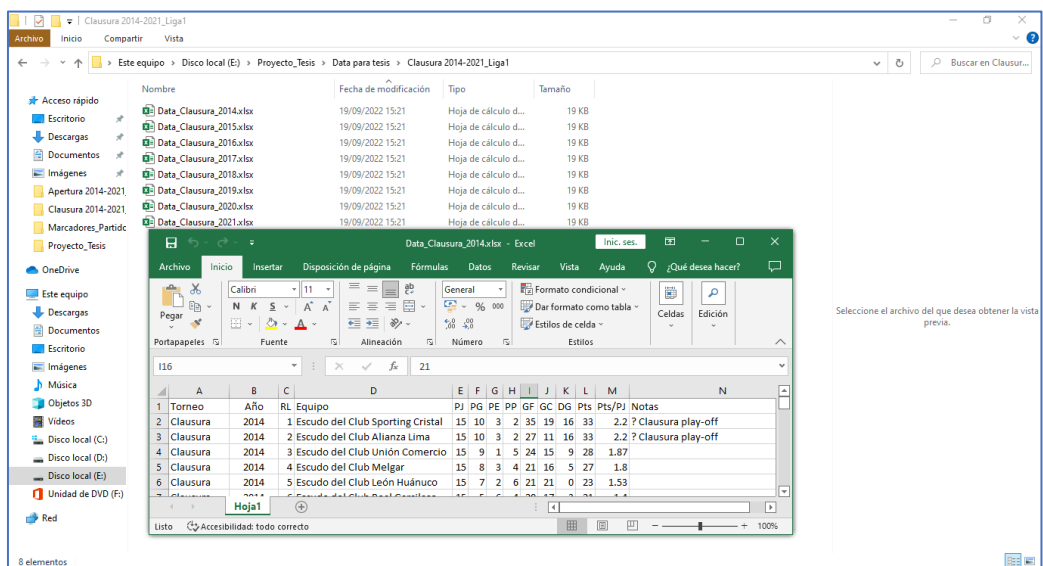


Figura 31 Clausura 2014-2021\_Liga1

## Carpeta Equipos de Países con sus continentes

### Detalle de los archivos obtenidos y la información de los mismos

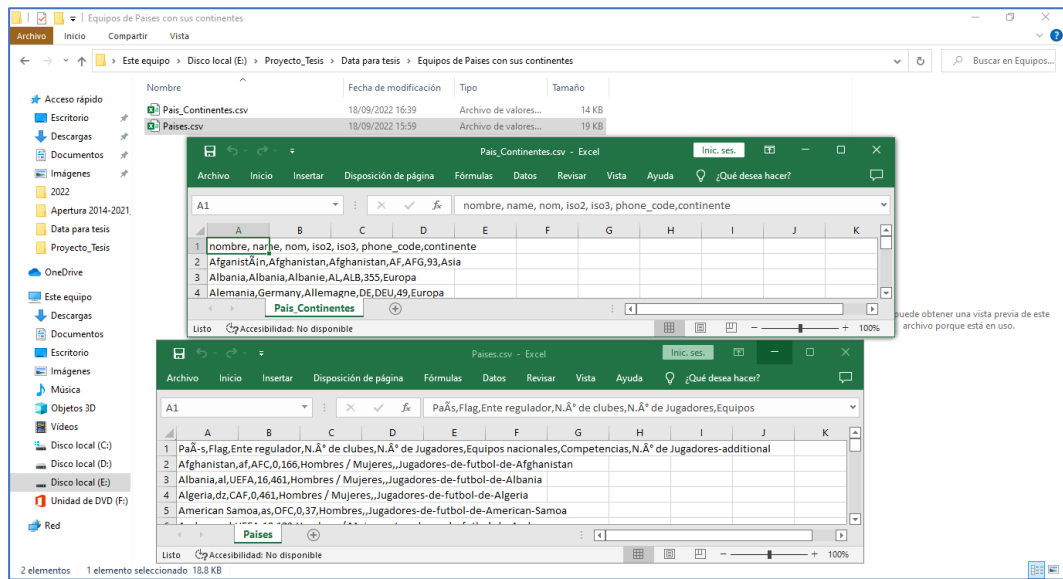


Figura 32 Equipos de Países con sus continentes

## Carpeta Estadios Nacionales

### Detalle de los archivos obtenidos y la información de los mismos

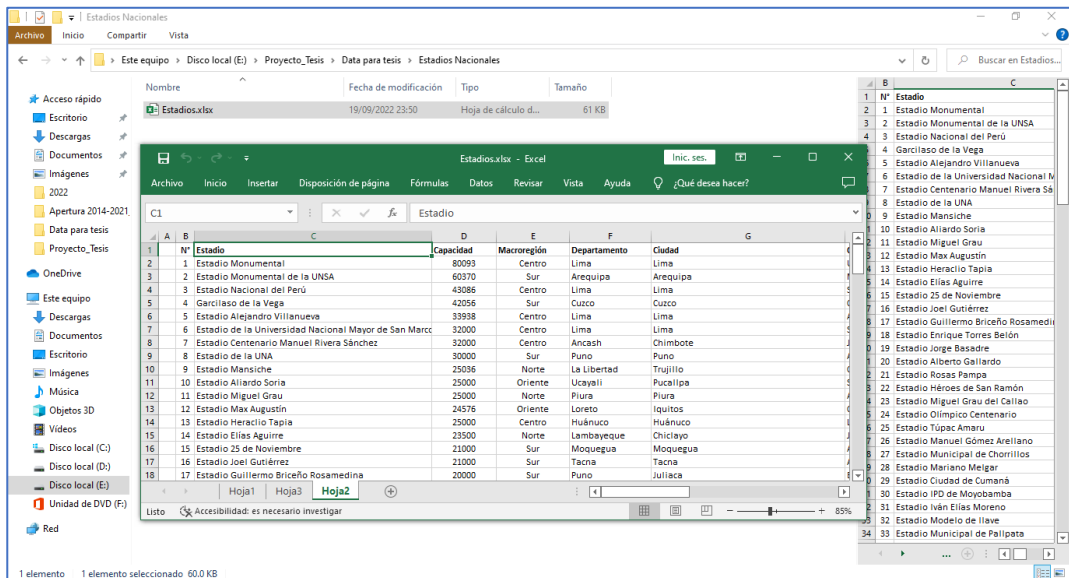


Figura 33 Estadios Nacionales



## Carpeta Estadística\_estándar\_2015\_2021

Detalle de los archivos obtenidos y la información de los mismos

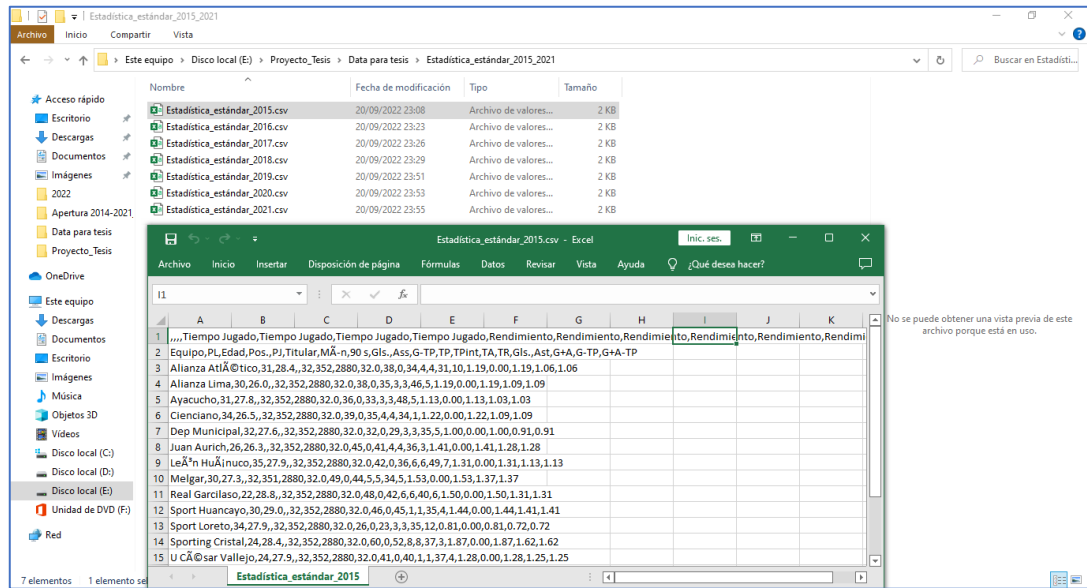


Figura 34 Estadística\_estándar\_2015\_2021

## Carpeta Estadísticas\_adicionales\_2015\_2021

Detalle de los archivos obtenidos y la información de los mismos

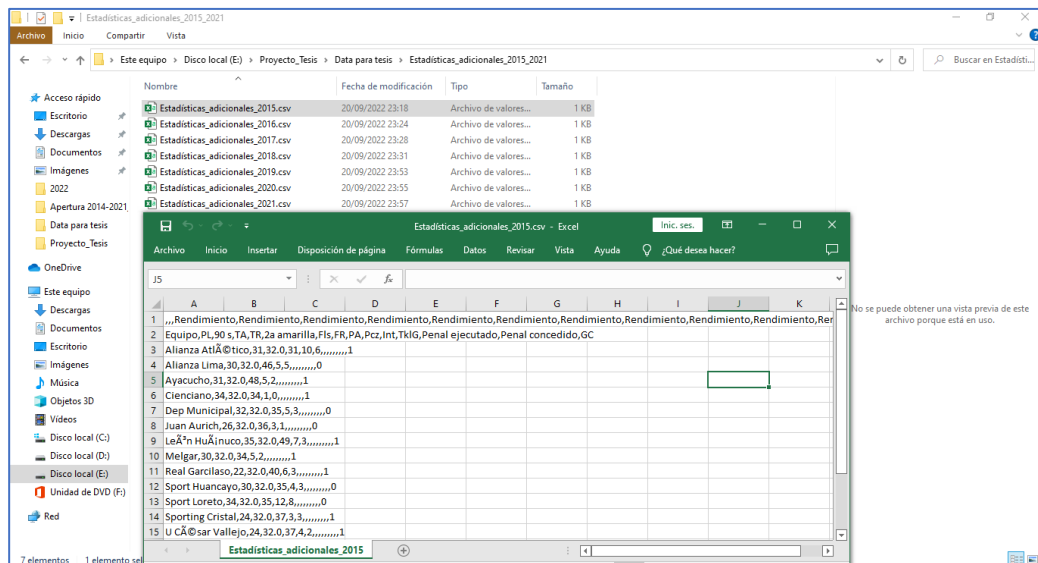


Figura 35 Estadísticas\_adicionales\_2015\_2021

## Carpeta Local\_visitante\_2014-2021

### Detalle de los archivos obtenidos y la información de los mismos

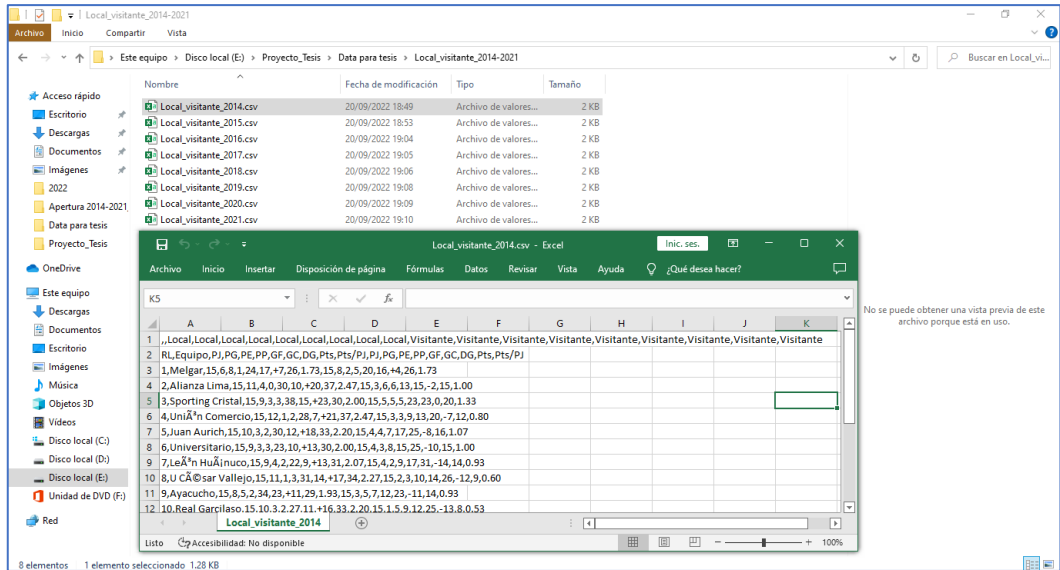


Figura 36 Local\_visitante\_2014-2021

## Carpeta Marcadores\_Partidos\_2014-2021

### Detalle de los archivos obtenidos y la información de los mismos

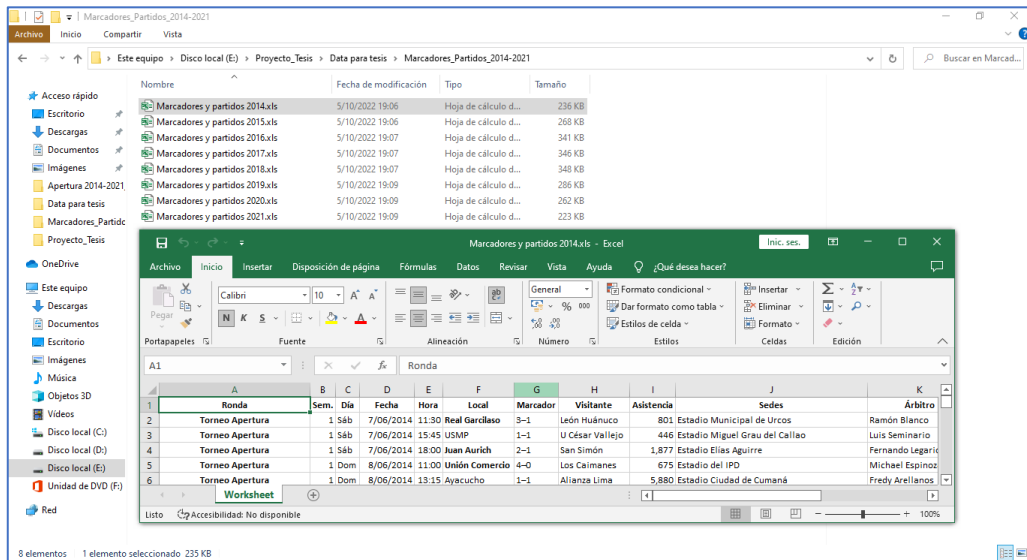


Figura 37 Marcadores\_Partidos\_2014-2021

## Carpeta Porteria\_equipo\_2015\_2021

### Detalle de los archivos obtenidos y la información de los mismos

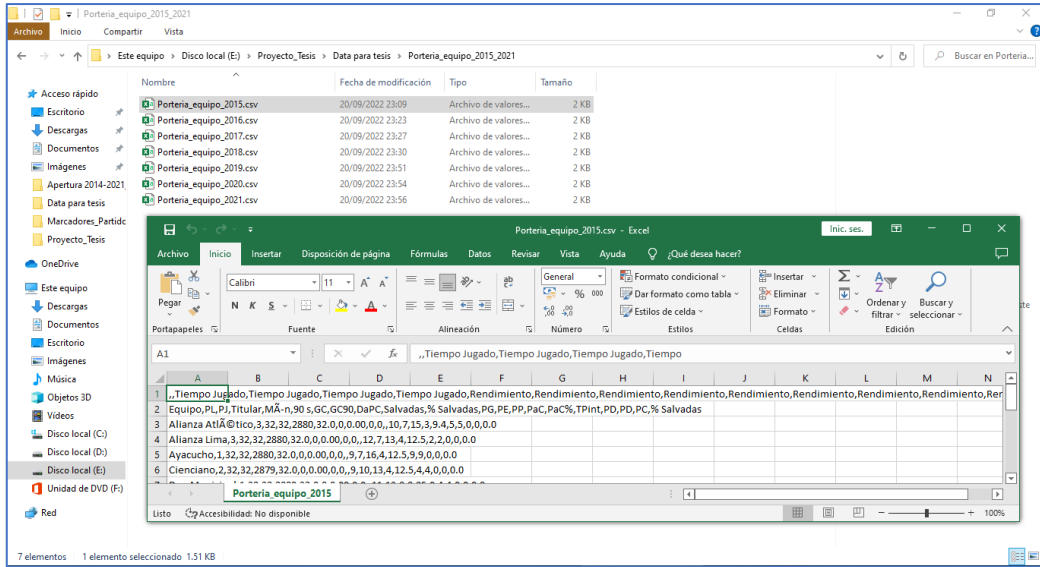


Figura 38 Porteria\_equipo\_2015\_2021

## Carpeta Relación\_EquiposGanadores\_1912-2021

### Detalle de los archivos obtenidos y la información de los mismos

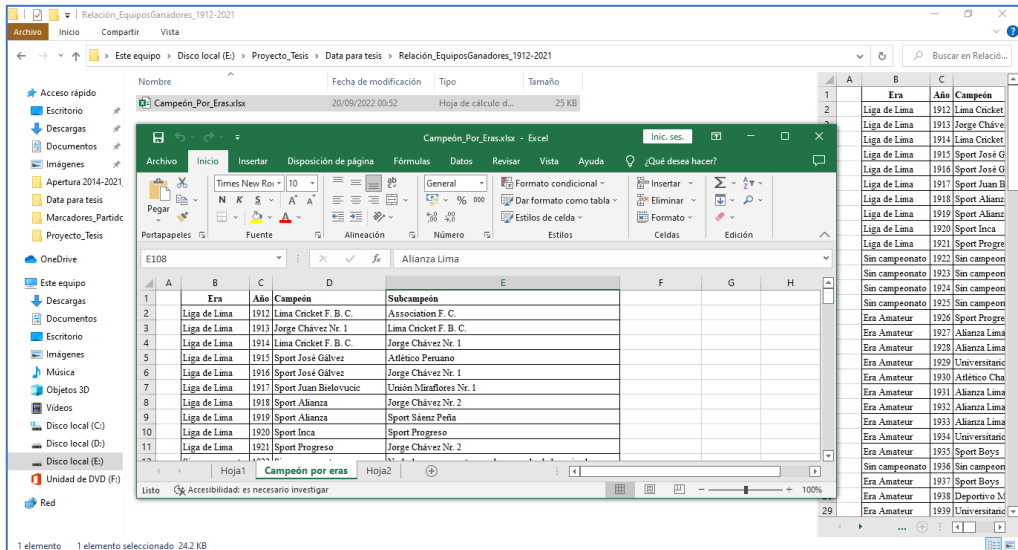


Figura 39 Relación\_EquiposGanadores\_1912-2021

## Carpeta Relación\_jugadores\_2015-2021

### Detalle de los archivos obtenidos y la información de los mismos

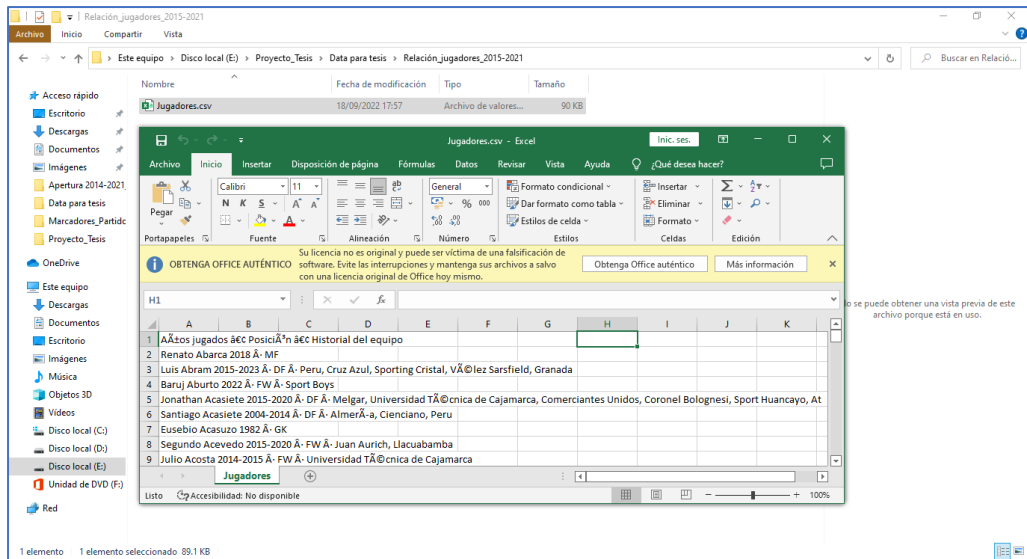


Figura 40 Relación\_jugadores\_2015-2021

## Carpeta TiempoJuego\_DelEquipo\_2015\_2021

### Detalle de los archivos obtenidos y la información de los mismos

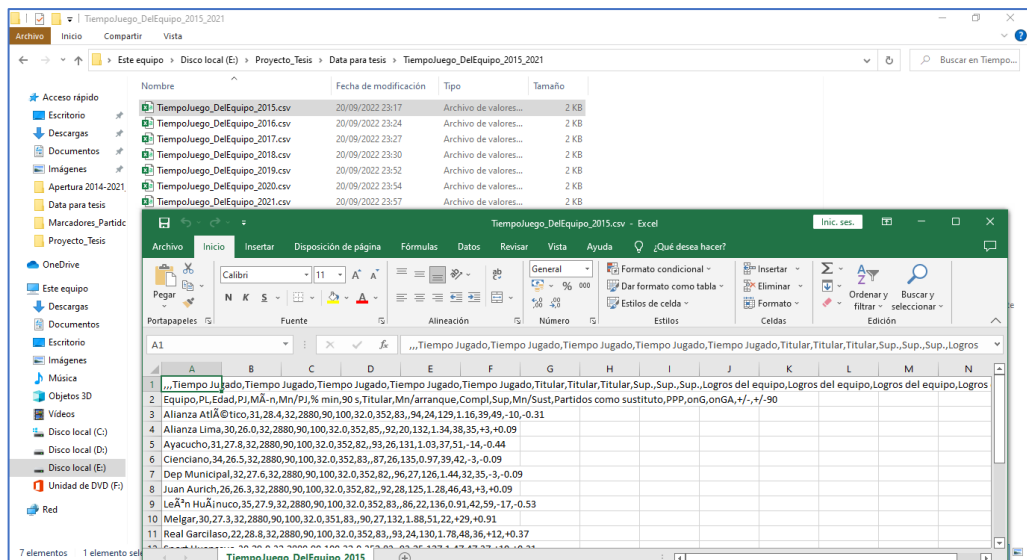
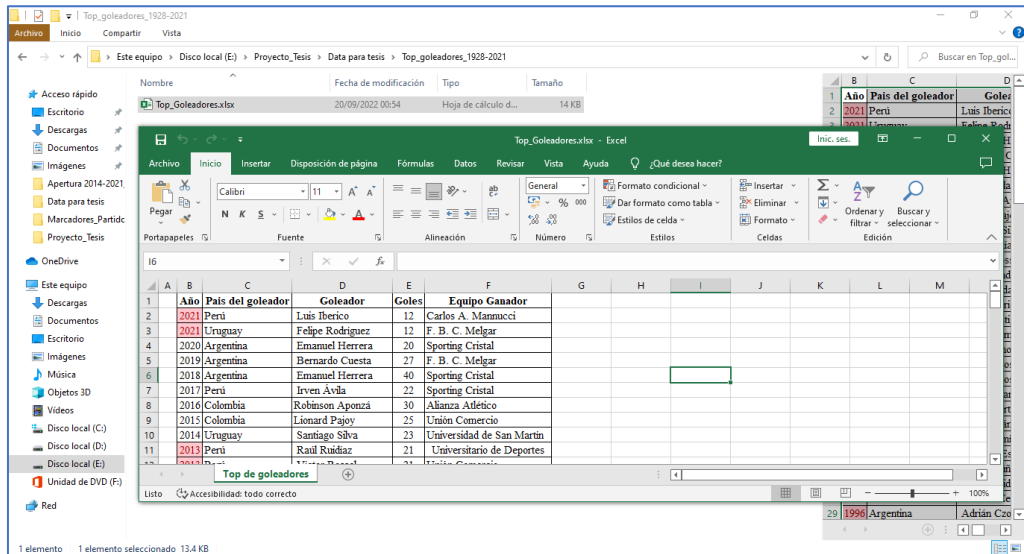


Figura 41 TiempoJuego\_DelEquipo\_2015\_2021

## Carpeta Top\_goleadores\_1928-2021

### Detalle de los archivos obtenidos y la información de los mismos



Año	Pais del goleador	Goleador	Goles	Equipo Ganador
2021	Peru	Luis Iberico	12	Carlos A. Mamucci
2021	Uruguay	Felipe Rodriguez	12	F. B. C. Melgar
2020	Argentina	Emanuel Herrera	20	Sporting Cristal
2019	Argentina	Bernardo Cuesta	27	F. B. C. Melgar
2018	Argentina	Emanuel Herrera	40	Sporting Cristal
2017	Peru	Irven Ayala	22	Sporting Cristal
2016	Colombia	Robinson Aponzá	30	Alianza Atlético
2015	Colombia	Lionard Pajov	25	Unión Comercio
2014	Uruguay	Santiago Silva	23	Universidad de San Martín
2013	Peru	Raul Ruidiaz	21	Universitario de Deportes
1996	Argentina	Adrian Cao		

Figura 42 Top\_goleadores\_1928-2021

#### 4.2.5. Obtener los datos (GD)

Importar la data al Power BI

- 1º. Abrimos el programa Power BI, y nos dirigimos al menú de **INICIO (1)** y de ahí a **Obtener datos (2)**, dentro de ella podemos seleccionar las diferentes fuentes de datos y sus respectivas extensiones. En nuestro caso vamos a utilizar dos extensiones, el csv (2.1) y el xlsx (2.2).

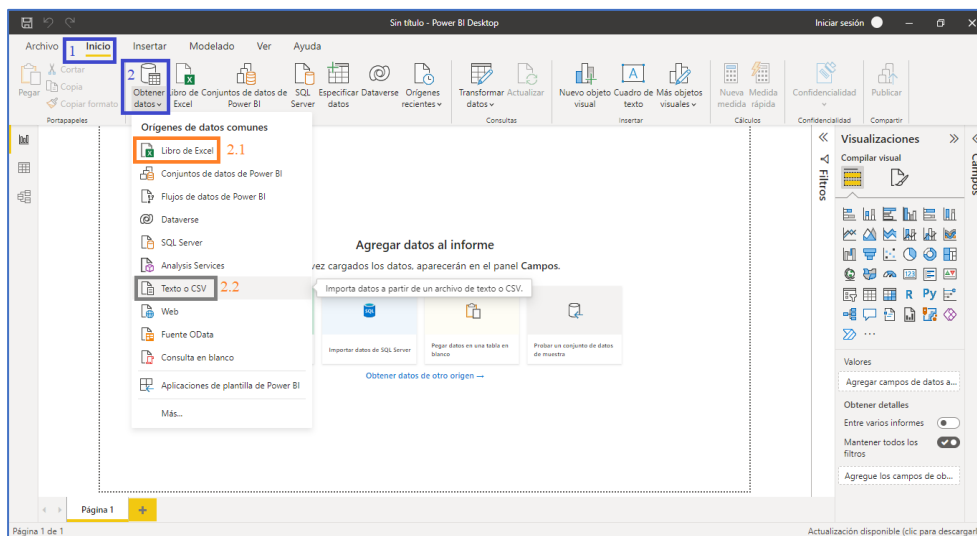


Figura 43 Selección de tipo de fuente en la carga en el Power BI

2º. Posteriormente después de seleccionar el tipo de fuente a importar, procederemos a seleccionar el archivo (Figura 27)

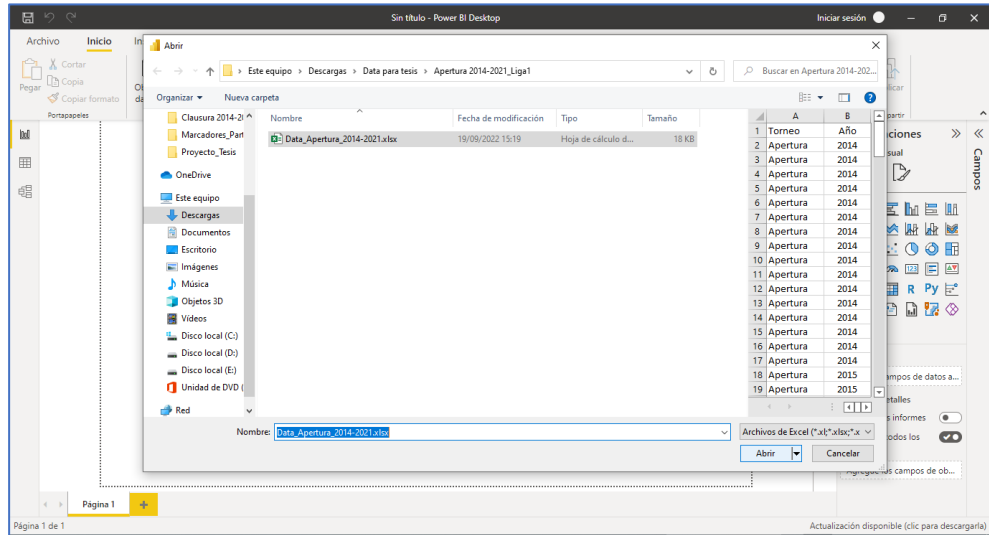


Figura 44 Seleccionado el archivo

3º. Seguidamente una ventana emergente aparecerá, con el nombre de **navegador** (1), de ahí se selecciona el **archivo a cargar** (2), para finalmente darle clic en **cargar** (3)

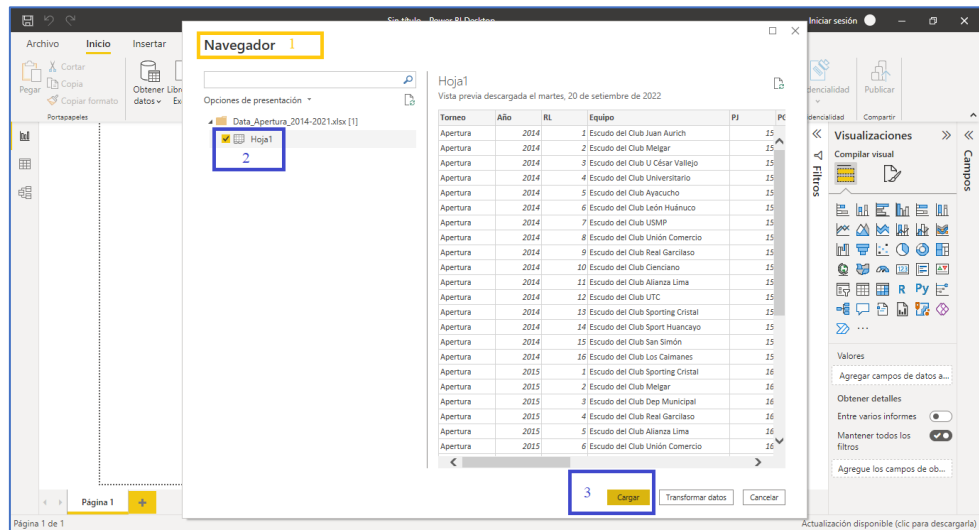


Figura 45 Seleccionando la opción cargar

4º. En este parte del proceso de carga se detalla lo siguiente: 1 (Modo Vista de los datos, 2 (Encabezado de la tabla de datos), 3 Nombre del archivo importado, ahí también dando clic derecho, hay otras opciones como eliminar, 4 (Vista modelo del archivo modelo), 5 (Espacio donde se

creará los dashboard al procesar la data) y finalmente 6 (Visualización de los registros del archivo importado. Para cualquier tipo de archivo la visualización va a ser la misma, al cargar la data.

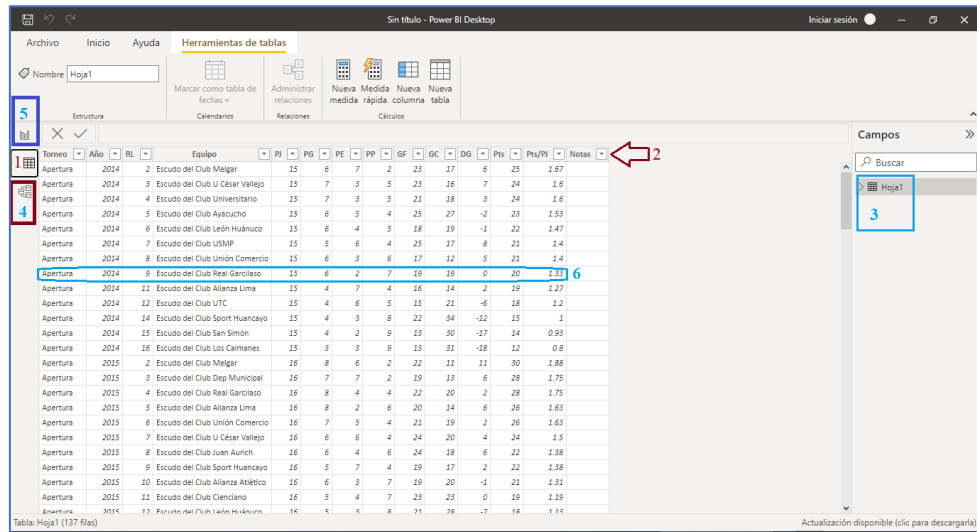


Figura 46 Visualización del archivo cargado

5º. Esta imagen distingue que al cargar un archivo csv, se tiene que tener en cuenta el delimitador, ya que nos proporciona una vista preliminar de cómo se cargara nuestro archivo al seleccionar el tipo de delimitador. Hay que tenerlo en cuenta porque los archivos csv, tienen diferentes delimitadores, dependiendo del archivo origen como haya sido estructurado. Su visualización debe ser igual que la figura anterior.

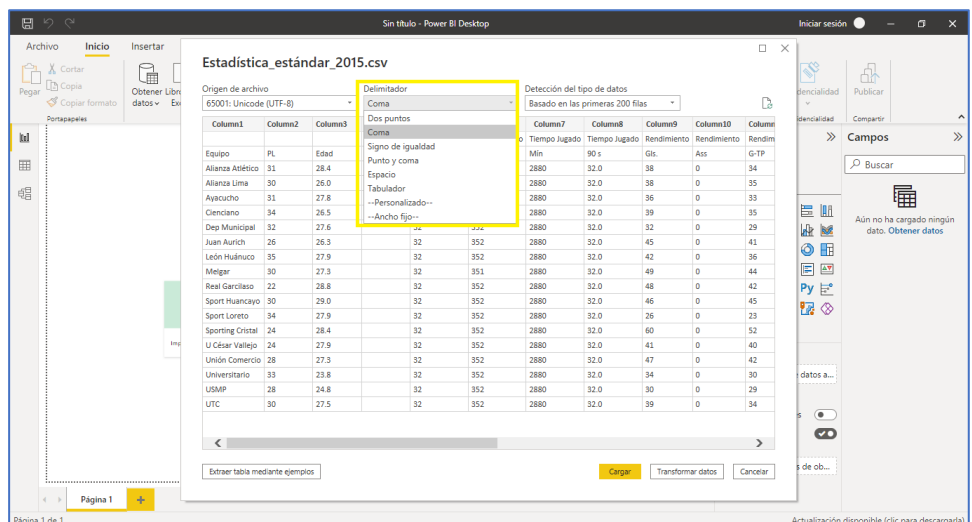


Figura 47 Importación de un archivo csv

6º. Una vez explicado la forma de cargar un archivo con extensión csv o xlsx, se procederá a cargar un total de **81 archivos**, de los cuales 54 archivos son csv y 27 archivos son xlsx.

#### 4.2.6. Preparación de datos (DP)

Una vez importado la data, se procederá a prepararla para su modelado posterior, dentro del cual iremos detallando uno por uno las transformaciones de los archivos csv y xlxs y que archivo utilizaremos. El orden de la preparación de los datos será desde el más básico, hasta el más elaborado.

Carpeta Equipos de Países con sus continentes

ARCHIVOS CONTENIDOS:

##### Pais\_Continentes.csv

	A <sup>B</sup> <sub>C</sub> Column1	A <sup>B</sup> <sub>C</sub> Column2	A <sup>B</sup> <sub>C</sub> Column3	A <sup>B</sup> <sub>C</sub> Column4	A <sup>B</sup> <sub>C</sub> Column5	A <sup>B</sup> <sub>C</sub> Column6	A <sup>B</sup> <sub>C</sub> Column7
1	nombre	name	nom	iso2	iso3	phone_code	continente
2	Afganistán	Afghanistan	Afghanistan	AF	AFG	93	Asia
3	Albania	Albania	Albanie	AL	ALB	355	Europa

Figura 48 Estructura del archivo Pais\_Continentes.csv

Columnas quitadas = `Table.RemoveColumns("#Tipo cambiado1", {"nom", " phone_code", " iso3"})`

Columnas reordenadas= `Table.ReorderColumns("#Columnas quitadas", {" name", "nombre", " iso2", "continente"})`

Columnas con nombres cambiados= `Table.RenameColumns("#Columnas reordenadas", {" name", "Pais"}, {"nombre", "Nombre del País"}, {" iso2", "FLag"}, {"continente", "Continente"})`

Resultado del archivo:

	A <sup>B</sup> <sub>C</sub> Pais	A <sup>B</sup> <sub>C</sub> Nombre del País	A <sup>B</sup> <sub>C</sub> FLag	A <sup>B</sup> <sub>C</sub> Continente
1	Afghanistan	Afganistán	AF	Asia
2	Albania	Albania	AL	Europa
3	Germany	Alemania	DE	Europa

Figura 49 Pais\_Continentes.csv (Data para el modelado)



## Paises.csv

A <sup>B</sup> <sub>C</sub> Column1	A <sup>B</sup> <sub>C</sub> Column2	A <sup>B</sup> <sub>C</sub> Column3	A <sup>B</sup> <sub>C</sub> Column4	A <sup>B</sup> <sub>C</sub> Column5	A <sup>B</sup> <sub>C</sub> Column6	A <sup>B</sup> <sub>C</sub> Column7	A <sup>B</sup> <sub>C</sub> Column8
1 País	Flag	Ente regulador	N.º de clubes	N.º de Jugadores	Equipos nacionales	Competencias	N.º de Jugadores-additional
2 Afghanistan	af	AFC	0	166	Hombres / Mujeres		Jugadores-de-futbol-de-Afgh...
3 Albania	al	UEFA	16	461	Hombres / Mujeres		Jugadores-de-futbol-de-Alba...

Figura 50 Estructura del archivo Paises.csv

Texto en mayúsculas = = Table.TransformColumns("#Columnas con nombre cambiado",{{"Flag", Text.Upper, type text}})

Columnas quitadas = Table.RemoveColumns("#Texto en mayúsculas",{"N.º de Jugadores-additional"})

Resultado del archivo:

A <sup>B</sup> <sub>C</sub> País	A <sup>B</sup> <sub>C</sub> Flag	A <sup>B</sup> <sub>C</sub> Ente regulador	1 <sup>2</sup> <sub>3</sub> N.º de clubes	1 <sup>2</sup> <sub>3</sub> N.º de Jugadores	A <sup>B</sup> <sub>C</sub> Equipos nacionales	A <sup>B</sup> <sub>C</sub> Competencias
1 Afghanistan	AF	AFC		0	166 Hombres / Mujeres	
2 Albania	AL	UEFA		16	461 Hombres / Mujeres	
3 Algeria	DZ	CAF		0	461 Hombres / Mujeres	

Figura 51 Paises.csv (Data para el modelado)

## Estadios.xlsx

A <sup>B</sup> <sub>C</sub> Nombre del Estadio	1 <sup>2</sup> <sub>3</sub> Capacidad de asistentes del estadio	A <sup>B</sup> <sub>C</sub> Zona	A <sup>B</sup> <sub>C</sub> Departamento	A <sup>B</sup> <sub>C</sub> Ciudad
1 Estadio Monumental		80093 Centro	Lima	Lima
2 Estadio Monumental de la UNSA		60370 Sur	Arequipa	Arequipa
3 Estadio Nacional del Perú		43086 Centro	Lima	Lima

ABC 123 Latitud	ABC 123 Longitud	A <sup>B</sup> <sub>C</sub> Club
-12.04318	-77.02824	Universitario de Deportes y Selección Peruana de Futbol
-16.39889	-71.535	Melgar
-12.04318	-77.02824	Selección Peruana de Futbol

Figura 52 Estructura del archivo Estadios.xlsx

Columnas quitadas = Table.RemoveColumns("#Tipo cambiado",{"N°"})

Columnas renombradas = Table.RenameColumns("#Columnas quitadas",{{"Capacidad", "Capacidad de asistentes del estadio"}})

Filas inferiores quitadas = Table.RemoveLastN("#Columnas con nombre cambiado",40)

Resultado del archivo:

A <sup>B</sup> <sub>C</sub> Nombre del Estadio	1 <sup>2</sup> <sub>3</sub> Capacidad de asistentes del estadio	A <sup>B</sup> <sub>C</sub> Zona	A <sup>B</sup> <sub>C</sub> Departamento	A <sup>B</sup> <sub>C</sub> Ciudad
1 Estadio Monumental		80093 Centro	Lima	Lima
2 Estadio Monumental de la UNSA		60370 Sur	Arequipa	Arequipa
3 Estadio Nacional del Perú		43086 Centro	Lima	Lima

A <sup>B</sup> <sub>C</sub> Departamento	A <sup>B</sup> <sub>C</sub> Ciudad	ABC 123 Latitud	ABC 123 Longitud	A <sup>B</sup> <sub>C</sub> Club
Lima	Lima	-12.04318	-77.02824	Universitario de Deportes y Selección Peruana de Futbol
Arequipa	Arequipa	-16.39889	-71.535	Melgar
Lima	Lima	-12.04318	-77.02824	Selección Peruana de Futbol

Figura 53 Estadios.xlsx (Data para el modelado)

## Top\_Goleadores.xlsx

123	Año	123	Pais del goleador	123	Goleador	123	Goles	123	Equipo Ganador
1	2021	Perú	Luis Iberico	12	Carlos A. Mannucci				
2	2021	Uruguay	Felipe Rodríguez	12	F. B. C. Melgar				
3	2020	Argentina	Emanuel Herrera	20	Sporting Cristal				

Figura 54 Estructura del archivo Top\_Goleadores

Encabezados promovidos = Table.PromoteHeaders("#Top de goleadores\_Sheet", [PromoteAllScalars=true])

Files inferiores quitadas = Table.RemoveLastN("#Tipo cambiado",1)

Resultado del archivo:

1-3	Año	A-C	Pais del goleador	A-C	Goleador	1-3	Goles	A-C	Equipo Ganador
1	2021	Perú	Luis Iberico	12	Carlos A. Mannucci				
2	2021	Uruguay	Felipe Rodríguez	12	F. B. C. Melgar				
3	2020	Argentina	Emanuel Herrera	20	Sporting Cristal				

Figura 55 Top\_Goleadores.xlsx (Data para el modelado)

## Campeón\_Por\_Eras.xlsx

A-C	Era	1-3	Año	A-C	Campeón	A-C	Subcampeón	ABC	Column5
1	Liga de Lima	1912	Lima Cricket F. B. C.	Association F. C.	null				
2	Liga de Lima	1913	Jorge Chávez Nr. 1	Lima Cricket F. B. C.	null				
3	Liga de Lima	1914	Lima Cricket F. B. C.	Jorge Chávez Nr. 1	null				

Figura 56 Estructura del archivo Campeón\_Por\_Eras.xlsx

Columnas quitadas = Table.RemoveColumns("#Tipo cambiado,{"Column5"})

Resultado del archivo:

A-C	Era	1-3	Año	A-C	Campeón	A-C	Subcampeón
1	Liga de Lima	1912	Lima Cricket F. B. C.	Association F. C.			
2	Liga de Lima	1913	Jorge Chávez Nr. 1	Lima Cricket F. B. C.			
3	Liga de Lima	1914	Lima Cricket F. B. C.	Jorge Chávez Nr. 1			

Figura 57 Campeón\_Por\_Eras.xlsx (Data para el modelado)

Clasificados2015.csv  
 Clasificados2016.csv  
 Clasificados2017.csv  
 Clasificados2018.csv

Clasificados2019.csv  
 Clasificados2020.csv  
 Clasificados2021.csv

A-C	Column1	A-C	Column2	A-C	Column3	A-C	Column4
1	Equipo	Género	Calificado para	Resultado			
2	Juan Aurich	M	2015 Copa Libertadores (Second stage) Por ganar la eliminación directa	3º lugar en la Second stage			
3	Sporting Cristal	M	2015 Copa Libertadores (Second stage) Por el final de liga	3º lugar en la Second stage			
4							
5	León Huánuco	M	2015 Copa Sudamericana (First stage) Por el final de liga	Perdió en First stage contra CS Emelec			
6	Melgar	M	2015 Copa Sudamericana (First stage) Por el final de liga	Perdió en First stage contra Junior			
7	Unión Comercio	M	2015 Copa Sudamericana (First stage) Por el final de liga	Perdió en First stage contra Rionegro			
8	Universitario	M	2015 Copa Sudamericana (First stage) Por el final de liga	Perdió en Second stage contra Defensor Sporting			

Figura 58 Estructura de los archivos Clasificados del 2015.csv al 2021.csv

El formato de estos 7 archivos es similar, por lo cual se tomará solo como referencia el primer csv de esta lista.

```

Filas en blanco eliminadas = Table.SelectRows("#Tipo cambiado1",
each
List.IsEmpty(List.RemoveMatchingItems(Record.FieldValues(_), {"", null})))
    
```

Posterior a ello, se van a juntar en un solo archivo llamado Clasificación\_Internacional\_2015\_2021 dentro de la interfaz del Power BI. Dando como resultado la siguiente estructura.

Para ello se utilizó la opción de anexar filas, como se visualiza en la imagen



Figura 59 Formato de agregar filas en Power BI

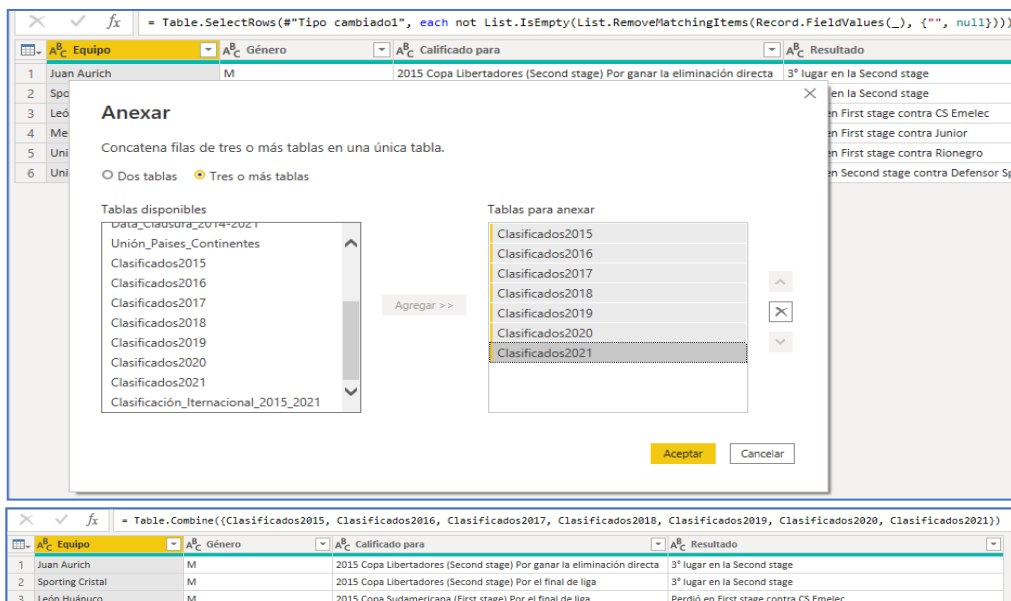


Figura 60 Combinando Clasificados del 2015 al 2021

Una vez combinado se procederá nuevamente a preparar los datos para el modelado del mismo.

Remover columnas = Table.RemoveColumns(Origen,{"Género"})  
 Valor reemplazado = Table.ReplaceValue("#Columnas quitadas",",",",",Replacer.ReplaceText,{"Calificado para"})

Dividir columna por posición = Table.SplitColumn("#Valor reemplazado", "Calificado para", Splitter.SplitTextByPositions({0, 4}, false), {"Calificado para.1", "Calificado para.2"})

Dividir columna por delimitador = Table.SplitColumn("#Columnas reordenadas", "Nombre del Campeonato - Copia - Copia", Splitter.SplitTextByEachDelimiter({"("}), QuoteStyle.Csv, false), {"Nombre del Campeonato - Copia - Copia.1", "Nombre del Campeonato - Copia - Copia.2"})

Dividir columna por delimitador = Table.SplitColumn("#Dividir columna por delimitador", "Nombre del Campeonato", Splitter.SplitTextByEachDelimiter({"("}), QuoteStyle.Csv, false), {"Nombre del Campeonato.1", "Nombre del Campeonato.2"})

Dividir columna por delimitador = Table.SplitColumn("#Tipo cambiado1", "Nombre del Campeonato.2", Splitter.SplitTextByEachDelimiter({"("}), QuoteStyle.Csv, false), {"Nombre del Campeonato.2.1", "Nombre del Campeonato.2.2"})

Columnas quitadas = Table.RemoveColumns("#Dividir columna por delimitador2",{"Nombre del Campeonato - Copia - Copia.1", "Nombre del Campeonato - Copia - Copia.2", "Nombre del Campeonato - Copia"})

Columnas con nombres cambiados = Table.RenameColumns("#Columnas quitadas1",{"Nombre del Campeonato.1", "Nombre del Campeonato"}, {"Nombre del Campeonato.2.1", "Puesto del Campeonato"}, {"Nombre del Campeonato.2.2", "Comentario uno"}, {"Resultado", "Comentario dos"})

Resultado del archivo:

	Año del campeonato	A <sup>B</sup> <sub>C</sub> Nombre del Equipo	A <sup>B</sup> <sub>C</sub> Nombre del Campeonato	A <sup>B</sup> <sub>C</sub> Puesto del Campeonato	A <sup>B</sup> <sub>C</sub> Comentario uno	A <sup>B</sup> <sub>C</sub> Comentario dos
1	2015	Juan Aurich	Copa Libertadores	Second stage	Por ganar la eliminación directa	3º lugar en la Second stage
2	2015	Sporting Cristal	Copa Libertadores	Second stage	Por el final de liga	3º lugar en la Second stage
3	2015	León Huánuco	Copa Sudamericana	First stage	Por el final de liga	Perdió en First stage contra CS Emelec

Figura 61 Anexando Clasificados del 2015 al 2021(Data para el modelado)

**Data\_Apertura\_2021.xlsx**  
**Data\_Apertura\_2020.xlsx**  
**Data\_Apertura\_2019.xlsx**  
**Data\_Apertura\_2018.xlsx**  
**Data\_Apertura\_2017.xlsx**  
**Data\_Apertura\_2016.xlsx**  
**Data\_Apertura\_2015.xlsx**  
**Data\_Apertura\_2014.xlsx**

**Data\_Clausura\_2014.xlsx**  
**Data\_Clausura\_2015.xlsx**  
**Data\_Clausura\_2016.xlsx**  
**Data\_Clausura\_2017.xlsx**  
**Data\_Clausura\_2018.xlsx**  
**Data\_Clausura\_2019.xlsx**  
**Data\_Clausura\_2020.xlsx**  
**Data\_Clausura\_2021.xlsx**

	A <sup>1</sup> Torneo	1 <sup>2</sup> Año	1 <sup>2</sup> RL	A <sup>1</sup> Equipo	1 <sup>2</sup> PJ	1 <sup>2</sup> PG
1	Apertura	2014		1 Escudo del Club Juan Aurich	15	9
2	Apertura	2014		2 Escudo del Club Melgar	15	6
3	Apertura	2014		3 Escudo del Club U César Vallejo	15	7

	1 <sup>2</sup> PE	1 <sup>2</sup> PP	1 <sup>2</sup> GF	1 <sup>2</sup> GC	1 <sup>2</sup> DG	1 <sup>2</sup> Pts	1.2 Pts/PJ
1		3	3	34	18	16	30
2		7	2	23	17	6	25
3		3	5	23	16	7	24

A <sup>1</sup> Notas
? Championship play-off
null
null

Figura 62 Estructura de los archivos del Apertura y Clausura desde el 2014 al 2021

No se lista los siguientes archivos por tener la misma estructura por cada año desde el 2014 hasta el 2021, ya sea en el torneo apertura como la clausura.

Posteriormente se unirá los 16 archivos en un solo archivo, para poder procesar la limpieza de los datos.

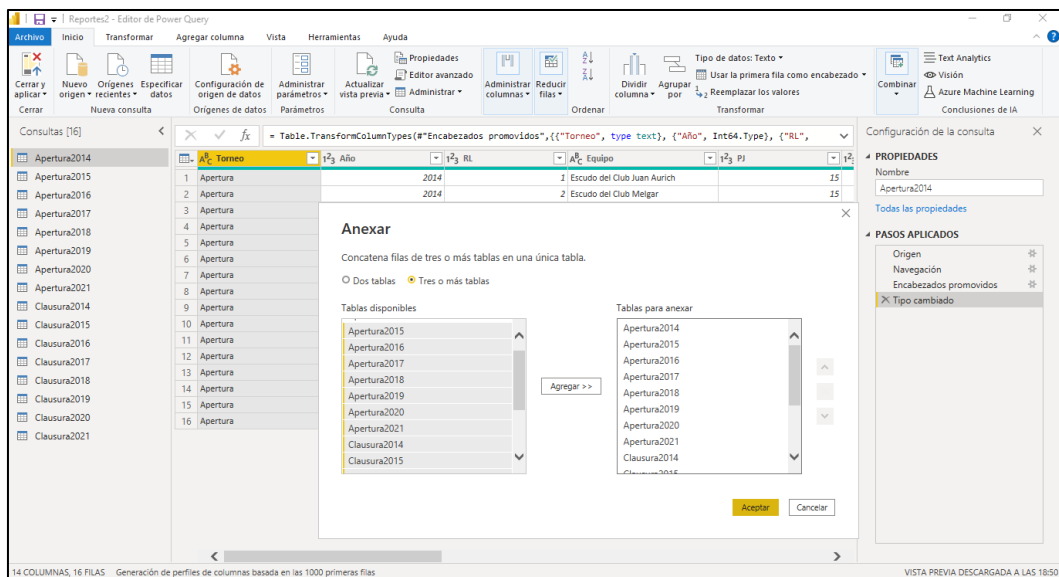


Figura 63 Anexando archivos del Apertura y Clausura del 2014 al 2021 parte 01

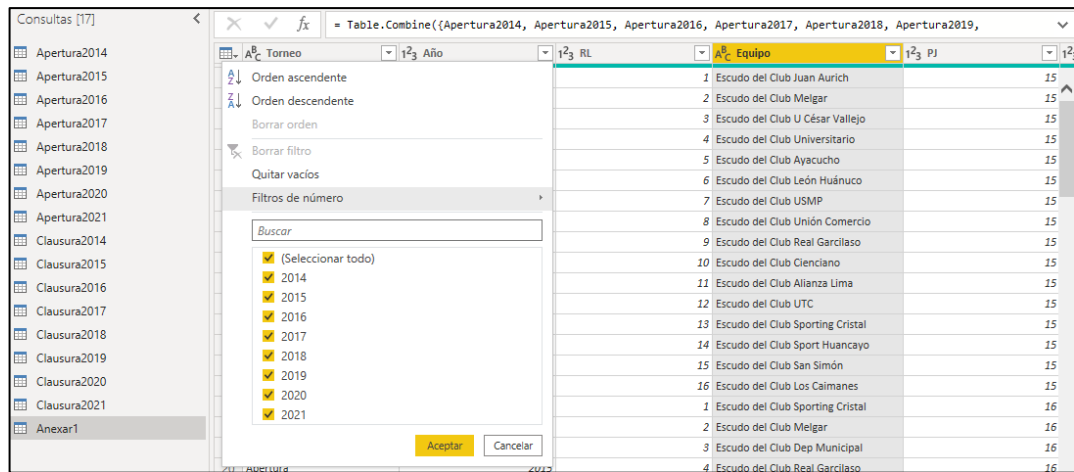


Figura 64 Anexando de archivos del Apertura y Clausura del 2014 al 2021 parte 02

Seguidamente se procederá a tratar la limpieza y homologación de los datos:

Valor reemplazado

=Table.ReplaceValue(Origen,"clausura","Clausura",Replacer.ReplaceText,{"Torneo"})

= Table.ReplaceValue("#"Valor reemplazado","?","",Replacer.ReplaceText,{"Notas"})

= Table.ReplaceValue("#"Valor reemplazado1",null,"",Replacer.ReplaceValue,{"Notas"})

= Table.ReplaceValue("#"Valor reemplazado2","Escudo del","",Replacer.ReplaceText,{"Equipo"})

Columnas con nombres cambiado = Table.RenameColumns("#"Valor reemplazado3",{"RL","Puesto"})

Resultado del archivo:

Torneo	Año	Puesto	Equipo	PJ	PG	PE	PP	GF	GC	DG	Pts	Pts/PJ
Clausura	2014	2	Club Alianza Lima	15	10	3	2	27	11	16	33	2.2
Clausura	2014	1	Club Sporting Cristal	15	10	3	2	35	19	16	33	2.2
Apertura	2014	13	Club Sporting Cristal	15	4	5	6	26	19	7	17	1.13
Apertura	2014	10	Club Cienciano	15	6	4	5	20	18	2	19	1.27
Apertura	2014	1	Club Juan Aurich	15	9	3	3	34	18	16	30	2
Clausura	2014	16	Club San Simón	15	2	3	10	13	31	-18	9	0.6

65 Anexando de archivos del Apertura y Clausura del 2014 al 2021 parte 03

**Clasificación\_Liga\_2014.csv**

**Clasificación\_Liga\_2015.csv**

**Clasificación\_Liga\_2016.csv**

**Clasificación\_Liga\_2017.csv**

**Clasificación\_Liga\_2018.csv**

**Clasificación\_Liga\_2019.csv**

**Clasificación\_Liga\_2020.csv**

**Clasificación\_Liga\_2021.csv**

1 <sup>2</sup> 3 RL	A <sup>B</sup> C Equipo	1 <sup>2</sup> 3 PJ	1 <sup>2</sup> 3 PG	1 <sup>2</sup> 3 PE	1 <sup>2</sup> 3 PP	1 <sup>2</sup> 3 GF	1 <sup>2</sup> 3 GC
1	1 Melgar		32	16	11	5	51
2	2 Sporting Cristal		32	16	10	6	60
3	3 Real Garcilaso		32	16	9	7	48

1 <sup>2</sup> 3 DG	1 <sup>2</sup> 3 Pts	1.2 Pts/PJ	1 <sup>2</sup> 3 Asistencia	A <sup>B</sup> C Máximo Goleador del Equipo	A <sup>B</sup> C Portero	A <sup>B</sup> C Notas
29	60		1.88	3367 Ysrael Zúñiga - 14	Daniel Ferreyra	1-point bonus 1 → Copa Libertadores Por ganar la elimi
20	58		1.81	1671 Carlos Lobatón - 12	Diego Penny	→ Copa Libertadores Por ganar la eliminación directa
12	57		1.78	808 Danilo Carando - 9	Diego Carranza	→ Copa Sudamericana Por perder la eliminación directa

Figura 66 Estructura de los archivos de la Clasificación\_Liga\_2015.csv al 2021.csv

Para estos archivos no se hizo ninguna modificación de la estructura de los datos por conveniencia, ya que va hacer unidos con otros archivos de información adicional por combinación. No se tomó en cuenta el archivo 2014, por tener información faltante para el resto de archivos.

Los siguientes archivos a unirse son:

**Local\_visitante\_2014.csv**

**Local\_visitante\_2015.csv**

**Local\_visitante\_2016.csv**

**Local\_visitante\_2017.csv**

**Local\_visitante\_2018.csv**

**Local\_visitante\_2019.csv**

**Local\_visitante\_2020.csv**

**Local\_visitante\_2021.csv**

A <sup>B</sup> C Column1	A <sup>B</sup> C Column2	A <sup>B</sup> C Column3	A <sup>B</sup> C Column4	A <sup>B</sup> C Column5	A <sup>B</sup> C Column6	A <sup>B</sup> C Column7	A <sup>B</sup> C Column8
1		Local	Local	Local	Local	Local	Local
2	RL	Equipo	PJ	PG	PE	PP	GF
3	1	Melgar	16	12	4	0	35

A <sup>B</sup> C Column9	A <sup>B</sup> C Column10	A <sup>B</sup> C Column11	A <sup>B</sup> C Column12	A <sup>B</sup> C Column13	A <sup>B</sup> C Column14	A <sup>B</sup> C Column15
Local	Local	Local	Visitante	Visitante	Visitante	Visitante
DG	Pts	Pts/PJ	PJ	PG	PE	PP
+27	40	2.50	16	4	7	5

A <sup>B</sup> C Column13	A <sup>B</sup> C Column14	A <sup>B</sup> C Column15	A <sup>B</sup> C Column16	A <sup>B</sup> C Column17	A <sup>B</sup> C Column18	A <sup>B</sup> C Column19	A <sup>B</sup> C Column20
Visitante	Visitante	Visitante	Visitante	Visitante	Visitante	Visitante	Visitante
PG	PE	PP	GF	GC	DG	Pts	Pts/PJ
4	7	5	16	14	+2	19	1.19
8	5	3	29	19	+10	29	1.81

Figura 67 Estructura del archivo Local\_visitante\_2014 al 2021 parte 01

Valor reemplazado = Table.ReplaceValue("#Tipo cambiado","PG","PG Local",Replacer.ReplaceText,{"Column4"})

A <sup>B</sup> <sub>C</sub> Column3	A <sup>B</sup> <sub>C</sub> Column4
Local	Local
PJ	PG Local
16	12
16	8

Figura 68 Reemplazando valores con Power BI

Eso sucederá con el resto de columnas:

```
Table.ReplaceValue("#Valor reemplazado", "PE", "PE Local", Replacer.ReplaceText, {"Column5"})
```

```
Table.ReplaceValue("#Valor reemplazado1", "PP", "PP Local", Replacer.ReplaceText, {"Column6"})
```

```
Table.ReplaceValue("#Valor reemplazado2", "GF", "GF Local", Replacer.ReplaceText, {"Column7"})
```

```
= Table.ReplaceValue("#Valor reemplazado4", "GC", "GC Local", Replacer.ReplaceText, {"Column8"})
```

```
= Table.ReplaceValue("#Valor reemplazado5", "DG", "DG Local", Replacer.ReplaceText, {"Column9"})
```

```
= Table.ReplaceValue("#Valor reemplazado6", "Pts", "Pts Local", Replacer.ReplaceText, {"Column10"})
```

```
= Table.ReplaceValue("#Valor reemplazado7", "Pts/PJ", "Pts/PJ Local", Replacer.ReplaceText, {"Column11"})
```

```
= Table.ReplaceValue("#Valor reemplazado8", "PJ", "PJ Visitante", Replacer.ReplaceText, {"Column12"})
```

```
= Table.ReplaceValue("#Valor reemplazado9", "PG", "PG Visitante", Replacer.ReplaceText, {"Column13"})
```

```
= Table.ReplaceValue("#Valor reemplazado10", "PE", "PE Visitante", Replacer.ReplaceText, {"Column14"})
```

```
= Table.ReplaceValue("#Valor reemplazado11", "PP", "PP Visitante", Replacer.ReplaceText, {"Column15"})
```

```
= Table.ReplaceValue("#Valor reemplazado12", "GF", "GF Visitante", Replacer.ReplaceText, {"Column16"})
```

```
= Table.ReplaceValue("#Valor reemplazado13", "GC", "GC Visitante", Replacer.ReplaceText, {"Column17"})
```

```
= Table.ReplaceValue("#Valor reemplazado14", "DG", "DG Visitante", Replacer.ReplaceText, {"Column18"})
```



= Table.ReplaceValue("#Valor reemplazado15","Pts","Pts Visitante",Replacer.ReplaceText,{"Column19"})

= Table.ReplaceValue("#Valor reemplazado16","Pts/PJ","Pts/PJ Visitante",Replacer.ReplaceText,{"Column20"})

Filas superiores quitadas = Table.Skip("#Valor reemplazado17",1)

Emcabzados promovidos = Table.PromoteHeaders("#Filas superiores quitadas1", [PromoteAllScalars=true])

Estos eventos se repetirán en el resto de csv, quedando después de la limpieza de la siguiente manera estructurada.

A <sup>o</sup>	RL	Equipo	PJ Local	PG Local	PE Local	PP Local	GF Local	GC Local
1	1	Melgar	16	12	4	0	35	8
2	2	Sporting Cristal	16	8	5	3	31	21
3	3	Real Garcilaso	16	11	3	2	29	15
A <sup>o</sup>	DG Local	Pts Local	Pts/PJ Local	PJ Visitante	PG Visitante	PE Visitante	PP Visitante	GF Visitante
+27	40	2.50	16	4	7	5	16	16
+10	29	1.81	16	8	5	3	29	19
+14	36	2.25	16	5	6	5	19	19
A <sup>o</sup>	PG Visitante	PE Visitante	PP Visitante	GF Visitante	GC Visitante	DG Visitante	Pts Visitante	Pts/PJ Visitante
4	7	5	16	14	+2	19	1.19	1.19
8	5	3	29	19	+10	29	1.81	1.81
5	6	5	19	21	-2	21	1.31	1.31

Figura 69 Estructura del archivo Local\_visitante\_2014 al 2021 parte 02

Posterior a ello se procederá a combinar

El archivo de Clasificación\_Liga\_2015 con Local\_visitante\_2015

El archivo de Clasificación\_Liga\_2016 con Local\_visitante\_2016

El archivo de Clasificación\_Liga\_2017 con Local\_visitante\_2018

El archivo de Clasificación\_Liga\_2019 con Local\_visitante\_2019

El archivo de Clasificación\_Liga\_2020 con Local\_visitante\_2020

El archivo de Clasificación\_Liga\_2021 con Local\_visitante\_2021



Figura 70 Formato de combinar filas en Power BI

### Combinar

Seleccione tablas y columnas coincidentes para crear una tabla combinada.

Clasificación\_Liga\_2015

RL	Equipo	PJ	PG	PE	PP	GF	GC	DG	Pts	Pts/PJ	Asistencia	Máximo Golea
1	Melgar	32	16	11	5	51	22	29	60	1.88	3367	Ysrael Zúñiga -
2	Sporting Cristal	32	16	10	6	60	40	20	58	1.81	1671	Carlos Lobatón
3	Real Garcilaso	32	16	9	7	48	36	12	57	1.78	808	Danilo Carandc
4	U César Vallejo	32	14	9	9	42	40	2	51	1.59	1182	Mauricio Mont

Local\_visitante\_2015

RL	Equipo	PJ Local	PG Local	PE Local	PP Local	GF Local	GC Local	DG Local	Pts Local
1	Melgar	16	12	4	0	35	8	27	
2	Sporting Cristal	16	8	5	3	31	21	10	
3	Real Garcilaso	16	11	3	2	29	15	14	
4	U César Vallejo	16	11	3	2	25	13	12	

Tipo de combinación

Externa izquierda (todas de la primera, coincidencias...)

Use las coincidencias aproximadas para comparar la combinación.

► Opciones de coincidencia aproximada

✓ La selección coincide con 17 de 17 filas de la primera tabla.

Aceptar Cancelar

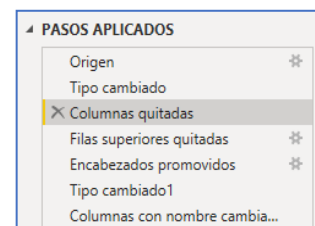
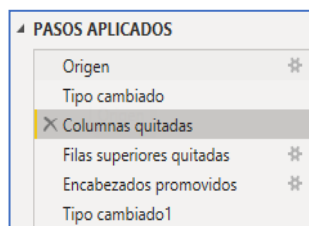
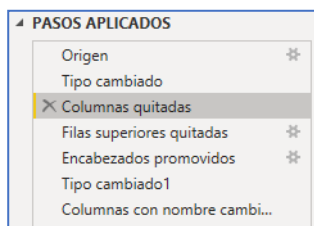
Figura 71 Combinando Clasificación\_Liga\_2015 con Local\_visitante\_2015

### Resultado:

RL	Equipo	PJ	PG	PE	PP	GF	GC	DG	Pts	Pts/PJ	Asistencia	Máximo Goleador del Equipo	Portero	Notas							
1	Melgar	32	16	11	5	51	22	29	60	1.88	3367	Ysrael Zúñiga - 14	Daniel Ferreyra	1-point bonus 1 → Copa Libertadores Por ganar la elimin							
2	Sporting Cristal	32	16	10	6	60	40	20	58	1.81	1671	Carlos Lobatón - 12	Diego Penny	→ Copa Libertadores Por ganar la eliminación directa							
3	Real Garcilaso	32	16	9	7	48	36	12	57	1.78	808	Danilo Carando - 9	Diego Carranza	→ Copa Sudamericana Por perder la eliminación directa							
Pts Local	1.2 Pts/PJ Local	PJ Visitante	PG Visitante	PE Visitante	PP Visitante	GF Visitante	GC Visitante	DG Visitante	Pts Visitante	1.2 Pts/PJ Visitante	Pts Local	1.2 Pts/PJ Local	PJ Visitante	PG Visitante	PE Visitante	PP Visitante	GF Visitante	GC Visitante	DG Visitante	Pts Visitante	1.2 Pts/PJ Visitante
40	2.5	16	4	7	5	16	14	19	29	1.81	16	2.25	16	8	5	3	29	15	19	21	1.59
36	2.25	16	5	6	5	16	19	21	29	1.81	16	2.25	16	8	5	3	29	15	19	21	1.59
Año	PJ Local	PG Local	PE Local	PP Local	GF Local	GC Local	DG Local	Pts Local	1.2 Pts/PJ Local	PJ Visitante	PG Visitante	PE Visitante	PP Visitante	GF Visitante	GC Visitante	DG Visitante	Pts Visitante	1.2 Pts/PJ Visitante			
2015	16	12	4	0	35	8	27	40	2.5	16	4	7	5	16	14	19	29	1.81			
2015	16	8	5	3	31	21	10	29	1.81	16	8	5	3	29	15	19	21	1.59			
2015	16	11	3	2	29	15	14	36	2.25	16	5	6	5	16	19	21	29	1.81			
Pts Local	1.2 Pts/PJ Local	PJ Visitante	PG Visitante	PE Visitante	PP Visitante	GF Visitante	GC Visitante	DG Visitante	Pts Visitante	1.2 Pts/PJ Visitante	PJ Local	PG Local	PE Local	PP Local	GF Local	GC Local	DG Local	Pts Local	1.2 Pts/PJ Local		
40	2.5	16	4	7	5	16	14	19	29	1.81	16	12	4	0	35	8	27	40	2.5		
36	2.25	16	5	6	5	16	19	21	29	1.81	16	8	5	3	31	21	10	29	1.81		
PJ Visitante	PG Visitante	PE Visitante	PP Visitante	GF Visitante	GC Visitante	DG Visitante	Pts Visitante	1.2 Pts/PJ Visitante													
4	7	5	16	14	19	21	1.19	1.19													
8	5	3	29	19	10	29	1.81	1.81													
5	6	5	19	21	-2	21	1.31	1.31													

Figura 72 Resultado de combina Clasificación\_Liga\_Local\_visitante\_2015

Estadística_estándar_2015		Estadísticas_adicionales_2015	
Porteria_equipo_2015 .csv			
Estadística_estándar_2016		Estadísticas_adicionales_2016	
Porteria_equipo_2016 .csv			
Estadística_estándar_2017		Estadísticas_adicionales_2017	
Porteria_equipo_2017 .csv			
Estadística_estándar_2018		Estadísticas_adicionales_2018	
Porteria_equipo_2018 .csv			
Estadística_estándar_2019		Estadísticas_adicionales_2019	
Porteria_equipo_2019 .csv			
Estadística_estándar_2020		Estadísticas_adicionales_2020	
Porteria_equipo_2020 .csv			
Estadística_estándar_2021		Estadísticas_adicionales_2021	
Porteria_equipo_2021 .csv			



Una vez tratado los datos se procederá a combinar los 3 archivos en una sola para combinarlo con el archivo Clasificación\_Liga con Local\_visitante.

Líneas más abajo se presenta de forma visual como quedo la estructura temporal de los archivos mencionados.

Equipo	N°P	TA	TR	2a amarilla
1 Alianza Atlético	31			6
2 Alianza Lima	30	46	5	5
3 Ayacucho	31	48	5	2

Figura 73 Presentación de datos de Estadística estándar, adicional y portería

Seguidamente se procederá a combinar los 3 archivos mencionado anteriormente de la siguiente manera:

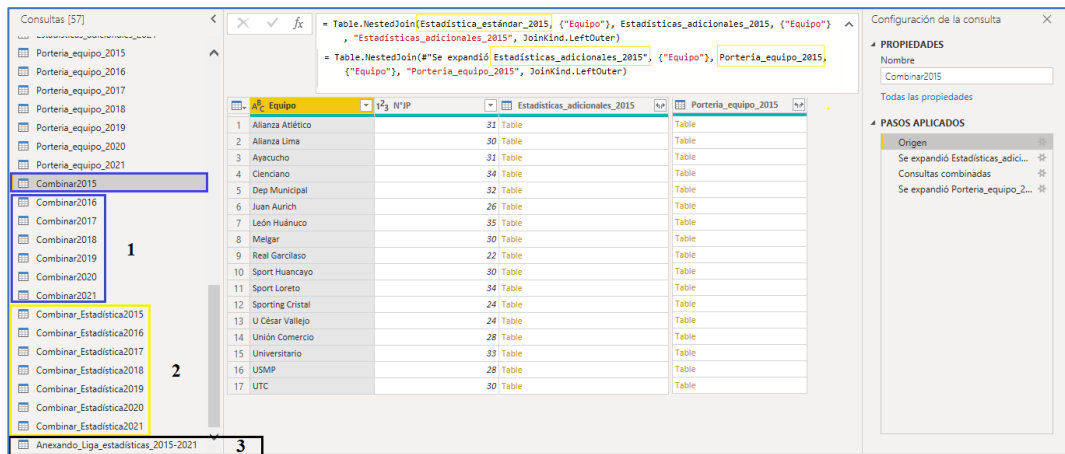
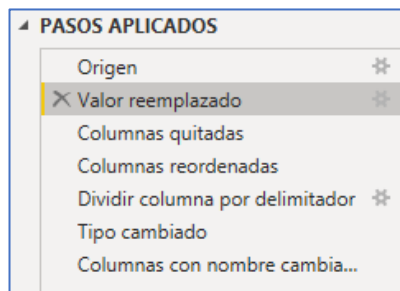


Figura 74 Combinando Estadística estándar, adicional, portería, clasificación liga y local visitante

1° En este paso se procedió a combinar Estadística estándar, adicional, portería por año.

2° En este paso se procedió a combinar Estadística estándar, adicional, portería, clasificación liga y local visitante por año.

3° En este paso se anexo los 5 archivos en un solo archivo, para transformar los datos, con los siguientes pasos:



= Table.ReplaceValue(Origen,"→"," ",Replacer.ReplaceText,{"Notas"})

= Table.RemoveColumns("#Valor reemplazado",{"Asistencia"})

= Table.ReorderColumns("#Columnas quitadas",{"Año", "PFC", "Equipo", "PJ", "PG", "PE", "PP", "GF", "GC", "DG", "Pts", "Pts/PJ", "Máximo Goleador del Equipo", "Portero", "Notas", "PJ Local", "PG Local", "PE Local", "PP Local", "GF Local", "GC Local", "DG Local", "Pts Local", "Pts/PJ Local", "PJ Visitante", "PG Visitante", "PE Visitante", "PP Visitante", "GF Visitante", "GC Visitante", "DG Visitante", "Pts Visitante",

"Pts/PJ Visitante", "N°JP", "TA", "TR", "2a amarilla", "N°P", "TPint", "PC", "PD", "PF", "% Salvadas}")

= Table.SplitColumn("#Columnas reordenadas", "Máximo Goleador del Equipo", Splitter.SplitTextByDelimiter("-", QuoteStyle.Csv), {"Máximo Goleador del Equipo.1", "Máximo Goleador del Equipo.2"})

=Table.TransformColumnTypes("#Dividir columna por delimitador",{"Máximo Goleador del Equipo.1", type text}, {"Máximo Goleador del Equipo.2", Int64.Type})

=Table.RenameColumns("#Tipo cambiado",{"Máximo Goleador del Equipo.1", "Goleador Equipo"}, {"Máximo Goleador del Equipo.2", "CGG"})

Resultado:

Año	PFC	Equipo	PJ	PG	PE	PP	GF	GC	DG	Pts	Pts/PJ	Goleador Equipo	CGG	Portero	Notas
1	2	Orden ascendente	32	16	11	5	5	22	29	60	1.88	Ysrael Zúñiga	14	Daniel Ferreyra	1-point bonus 1 Copa L
2	2	Orden descendente	32	10	7	15	3	49	-10	37	1.16	Roberto Jiménez	11	Cristian Pinzón	
3	2	Borrar orden	32	16	10	6	6	40	20	58	1.81	Carlos Lobatón	12	Diego Penny	Copa Libertadores Por
4	2	Borrar filtro	32	12	7	13	3								
5	2	Quitar vacíos	32	16	9	7	4								
6	2	Filtros de número	32	9	7	16	3								
7	2	Buscar	32	14	9	9	4								
8	2	Buscar	32	9	10	13	3								
9	2	Buscar	32	9	10	13	3								
10	2	Buscar	32	13	11	8	4								
11	2	Buscar	32	11	13	8	3								
12	2	Buscar	32	11	8	13	4								
13	2	Buscar	32	12	8	12	4								
14	2	Buscar	32	8	9	15	4								
15	2	Buscar	32	11	10	11	3								
16	2	Buscar	32	10	7	15	3								
17	2	Buscar	32	7	10	15	2								
18	2	Buscar	32	9	10	13	3								
		Aceptar	44	21	12	11	7								

Figura 75 Estructura del archivo Anexando\_Liga\_estadísticas\_2015-2021

**Marcadores y partidos 2014.xlsx**

**Marcadores y partidos 2015.xlsx**

**Marcadores y partidos 2016.xlsx**

**Marcadores y partidos 2017.xlsx**

**Marcadores y partidos 2018.xlsx**

## Marcadores y partidos 2019.xlsx

## Marcadores y partidos 2020.xlsx

## Marcadores y partidos 2021.xlsx

Una vez cargado estos archivos se tiene la siguiente estructura de datos, esto es para los archivos del 2014 al 2021

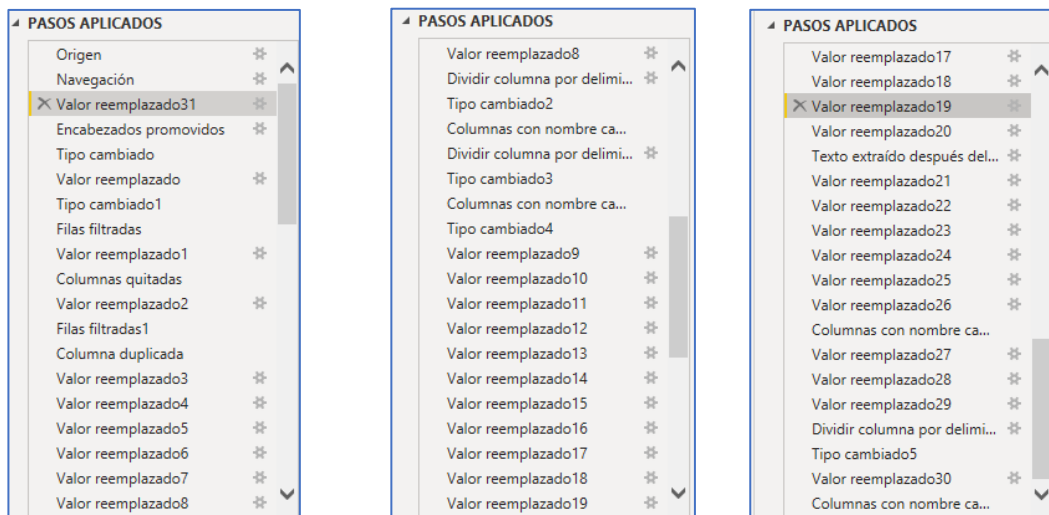
Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8
1	Ronda	Sem.	Día	Fecha	Hora	Local	Visitante
2	Torneo Apertura	1	Sáb	7/06/2014	31/12/1899 11:30:00	Real Garcilazo	León Huánuco
3	Torneo Apertura	1	Sáb	7/06/2014	31/12/1899 15:45:00	USMP	U César Vallejo

Column7	Column8	Column9	Column10	Column11	Column12	Column13
Marcador	Visitante	Asistencia	Sedes	Árbitro	Informe del partido	Notas
3-1	León Huánuco		801 Estadio Municipal de Urcos	Ramón Blanco	Informe del partido	null
1-1	U César Vallejo		446 Estadio Miguel Grau del Callao	Luis Seminario	Informe del partido	null

Figura 76 Estructura del archivo de Marcadores y partidos del 2014

Ahora se anexan los 8 archivos mencionados anteriormente y se procede a limpiar los datos.



Valor reemplazado = Table.ReplaceValue(Hoja1\_Sheet,null,"Sin comentarios",Replacer.ReplaceValue,{"Column13"})

Encabezados promovidos = Table.ReplaceValue(Hoja1\_Sheet,null,"Sin comentarios",Replacer.ReplaceValue,{"Column13"})

Columnas quitadas = Table.RemoveColumns("#Valor reemplazado1",{"Informe del partido"})

Columnas copiadas = Table.DuplicateColumn("#Filas filtradas1", "Marcador", "Marcador - Copia")

Dividir columnas por delimitador = Table.SplitColumn("#Valor reemplazado8", "Marcador", Splitter.SplitTextByDelimiter("-", QuoteStyle.Csv), {"Marcador.1", "Marcador.2"})

Columnas con nombre cambiado = Table.RenameColumns("#Tipo cambiado2",{"Marcador.1", "Marcador Local"}, {"Marcador.2", "Marcador Visitante"})

Texto extraído después del delimitador = Table.TransformColumns("#Valor reemplazado20", {"Marcador - Copia", each Text.AfterDelimiter(\_, "(", type text)})

Una vez procesado y limpiado los datos quedara de la siguiente manera:

### Resultado

1	2	3	4	5	6	7	8	9	10
Torneo Apertura	J	Sáb		7	Junio	2014	11:30:00	Real Garcilaso	
Torneo Apertura	J	Sáb		7	Junio	2014	15:45:00	USMP	
Torneo Apertura	J	Sáb		7	Junio	2014	18:00:00	Juan Aurich	

1	2	3	4	5	6	7	8	9	10
J	León Huánuco		0		0	801	Estadio Municipal de Urcos	Ramón Blanco	Sin
J	U César Vallejo		0		0	446	Estadio Miguel Grau del Callao	Luis Seminario	Sin
J	San Simón		0		0	2877	Estadio Elias Aguirre	Fernando Legario	Sin

Figura 77 Estructura del archivo Anexando Marcadores y partidos del 2015-2021

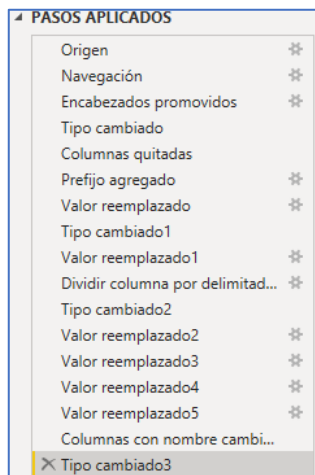
### Detalle\_jugadores.xlsx

1	2	3	4	5	6	7	8
TEMPORADA	N°J	Nac.	Jugador	Posición	Edad (hoy)	Club	Asistencias
APERTURA 2014	J	Uruguay	Santiago Silva	Delantero centro	24 (32)	Universidad San Martín de Porres	
APERTURA 2014	J	Perú	Carlos Lobatón	Mediocentro	34 (42)	Club Sporting Cristal	

1	2	3	4	5	6	7	8	9	10
Penalti	Tiempo de alineación	Minutos por gol	Goles por encuentro	Goles		null	null	null	null
0	1:197"	109'	0,73		11				
3	1:985"	90'	0,85		11				

Figura 78 Estructura de Detalle de jugadores del 2014 al 2021



Encabezados promovidos =

Table.PromoteHeaders(Detalle\_jugadores\_Sheet, [PromoteAllScalars=true])

Columnas quitadas = Table.RemoveColumns("#Tipo cambiado", {"Column14", "Column15", "Column16"})

Tipo cambiado = Table.TransformColumnTypes("#Encabezados promovidos", {"TEMPORADA", type text}, {"N°J", Int64.Type}, {"Nac.", type text}, {"Jugador", type text}, {"Posición", type text}, {"Edad (hoy)",

```
type text}, {"Club", type text}, {"Asistencias", Int64.Type}, {"Penalti",
Int64.Type}, {"Tiempo de alineación", type text}, {" Minutos por gol", type
text}, {"Goles por encuentro", Int64.Type}, {"Goles", Int64.Type},
{"Column14", type any}, {"Column15", type any}, {"Column16",
Int64.Type}})
```

```
Columnas quitadas = Table.RemoveColumns("#Tipo
cambiado",{"Column14", "Column15", "Column16"})
```

```
Prefijo agregado = Table.TransformColumns("#Columnas quitadas",
{{"Goles por encuentro", each "0." & Text.From(_, "es-PE"), type text}})
```

Valor reemplazado

```
= Table.ReplaceValue("#Prefijo
agregado", ".", "", Replacer.ReplaceText, {"Tiempo de alineación"})
```

```
= Table.ReplaceValue("#Tipo
cambiado1", "?", "0", Replacer.ReplaceText, {"Edad (hoy)"})
```

Tipo cambiado

```
= Table.TransformColumnTypes("#Valor reemplazado", {{"Goles por
encuentro", Percentage.Type}})
```

```
= Table.TransformColumnTypes("#Dividir columna por
delimitador", {{"Edad (hoy).1", Int64.Type}, {"Edad (hoy).2", type text}})
```

```
Dividir columna= Table.SplitColumn("#Valor reemplazado1", "Edad
(hoy)", Splitter.SplitTextByDelimiter("#(00A0)", QuoteStyle.Csv), {"Edad
(hoy).1", "Edad (hoy).2"})
```

Valor reemplazado

```
= Table.ReplaceValue("#Tipo cambiado2", "†
", "", Replacer.ReplaceText, {"Edad (hoy).2"})
```

```
= Table.ReplaceValue("#Valor
reemplazado2", null, "0", Replacer.ReplaceValue, {"Edad (hoy).2"})
```

```
= Table.ReplaceValue("#Valor
reemplazado3", "(", "", Replacer.ReplaceText, {"Edad (hoy).2"})
```

```
= Table.ReplaceValue("#Valor
reemplazado4", ")", "", Replacer.ReplaceText, {"Edad (hoy).2"})
```

```
Columnas con nombre cambiado = Table.RenameColumns("#Valor
reemplazado5", {"Edad (hoy).1", "Edad Jugador Torneo"}, {"Edad
(hoy).2", "Edad jugador Actual"})
```

```
Tipo cambiado= Table.TransformColumnTypes("#Columnas con nombre
cambiado", {"Edad jugador Actual", Int64.Type})
```

Resultado:

TEMPORADA	Nac.	Jugador	Posición	Club	Asistencias	Penalti	Tiempo de alineación	Minutos por gol	Goles por encuentro	Goles	Edad Jugador Torneo	Edad jugador Actual
APERTURA 2014	Perú	Gianfranco Labarti	Delantero centro	Universidad Técnica de C	0	2	1244'	176'	47%	7	29	38
APERTURA 2014	Perú	Horacio Calcaterra	Mediocentro	Club Sporting Cristal	1	0	989'	989'	80%	1	25	33
APERTURA 2014	Perú	Iván Santillán	Lateral izquierdo	Cusco FC	1	0	1017'	1.017'	80%	1	23	31

Figura 79 Estructura de Detalle de jugadores del 2014 al 2021



#### 4.2.7. Modelado de datos (DM)

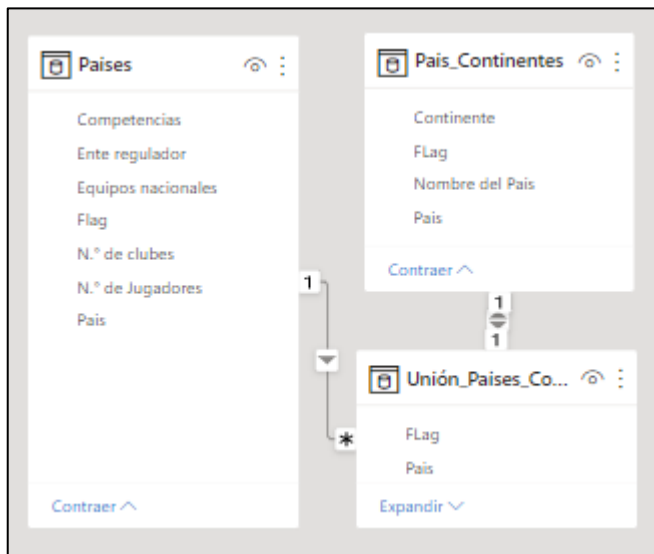


Figura 80 Modelado de países por continentes

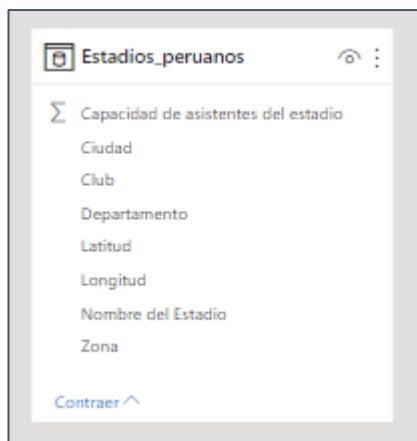


Figura 81 Modelado de estadios peruanos

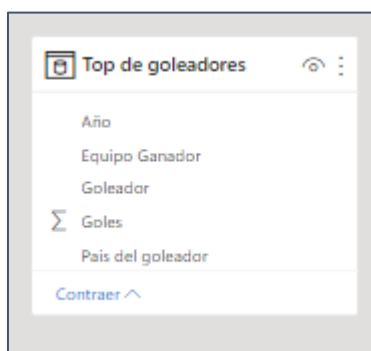


Figura 82 Modelo de top de goleadores de 1928 al 2021

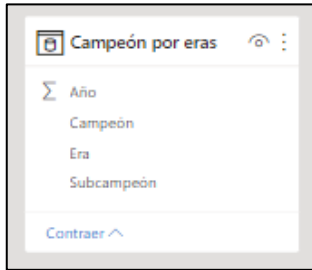


Figura 83 Modelado de Equipo campeón por eras

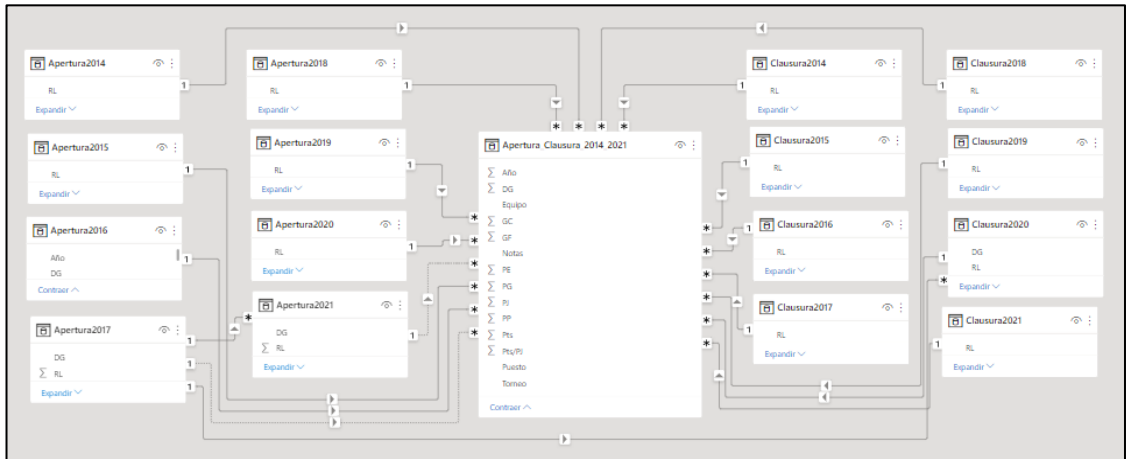


Figura 84 Modelado de la Clasificación de equipos del Torneo Apertura y Clausura desde 2014-2021

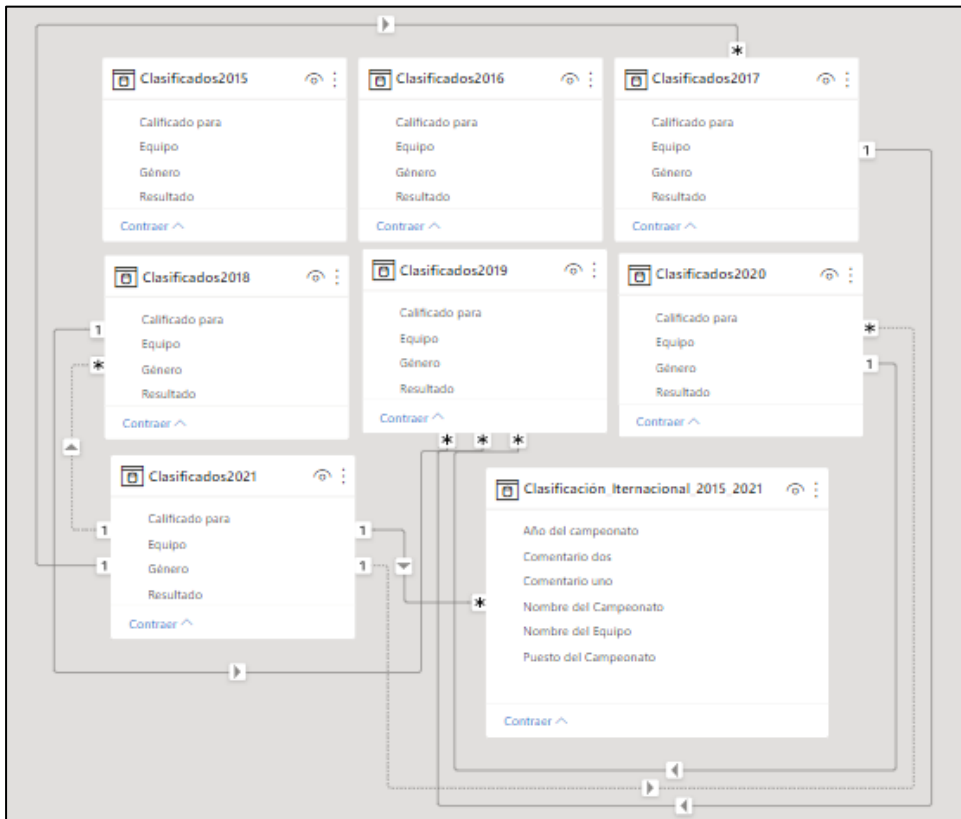


Figura 85 Modelado de Clasificación Internacional de los equipos peruanos del 2015 al 2021

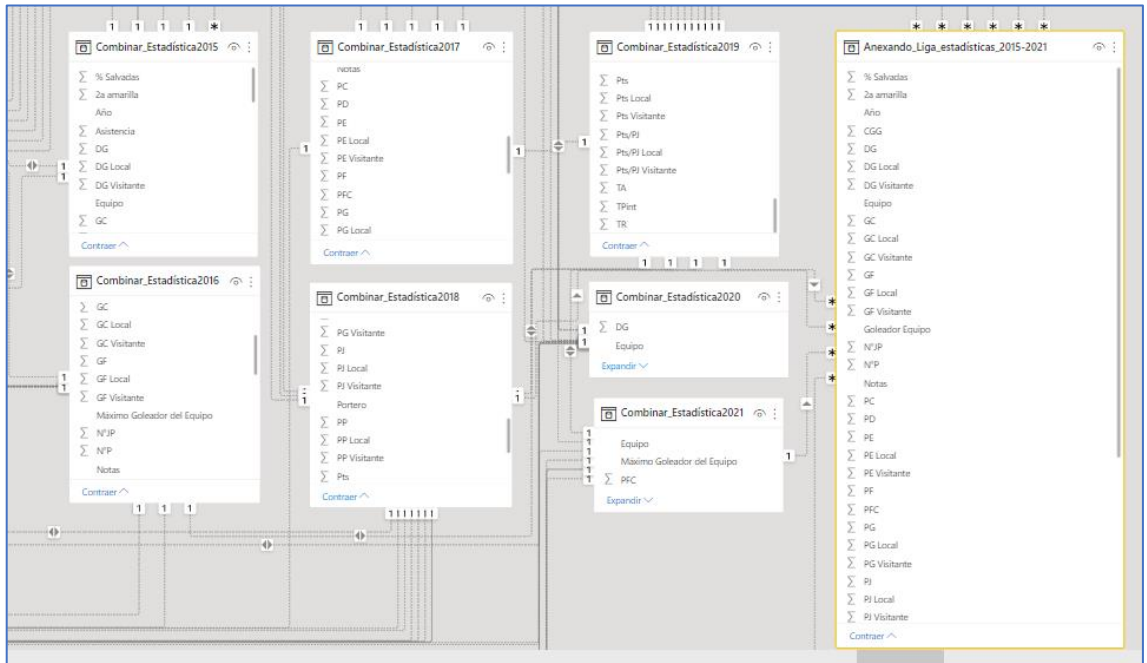


Figura 86 Anexando la combinación de datos estándar del 2015 al 2021

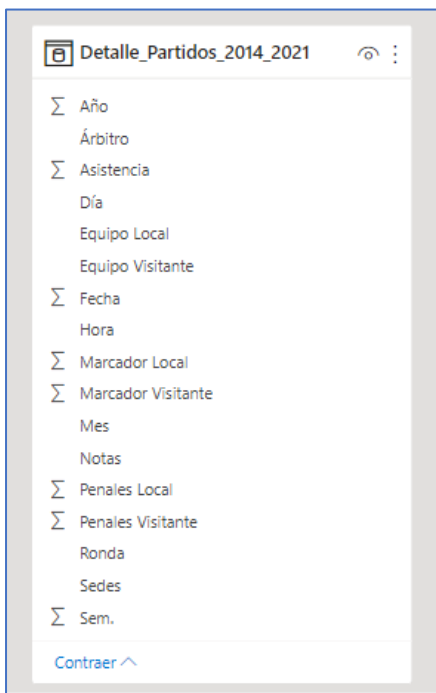


Figura 87 Detalle de partidos del Apertura y Clausura del 2014 al 2021

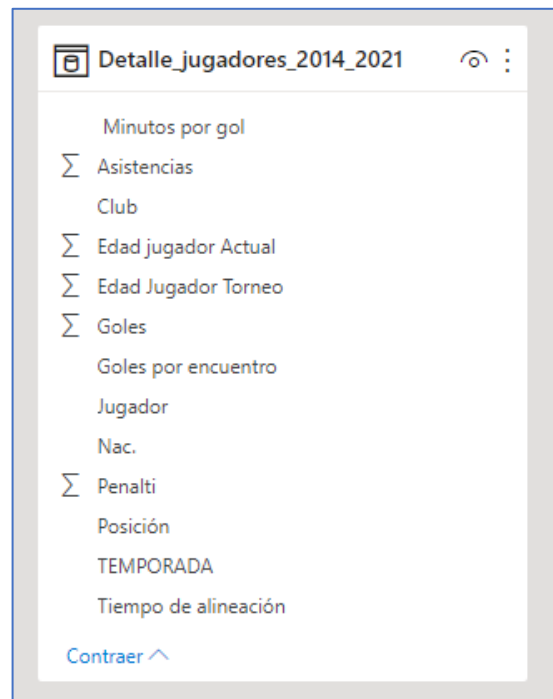


Figura 88 Detalle de Jugadores del 2014 al 2021

#### 4.2.8. Reporte de datos (DR)



Figura 89 Dashboard General.

#### Reporte de rendimiento de jugadores para fichaje del 2014 – 2021



Figura 90 Reporte de la información deportiva de los jugadores peruanos del año 2014 - 2021

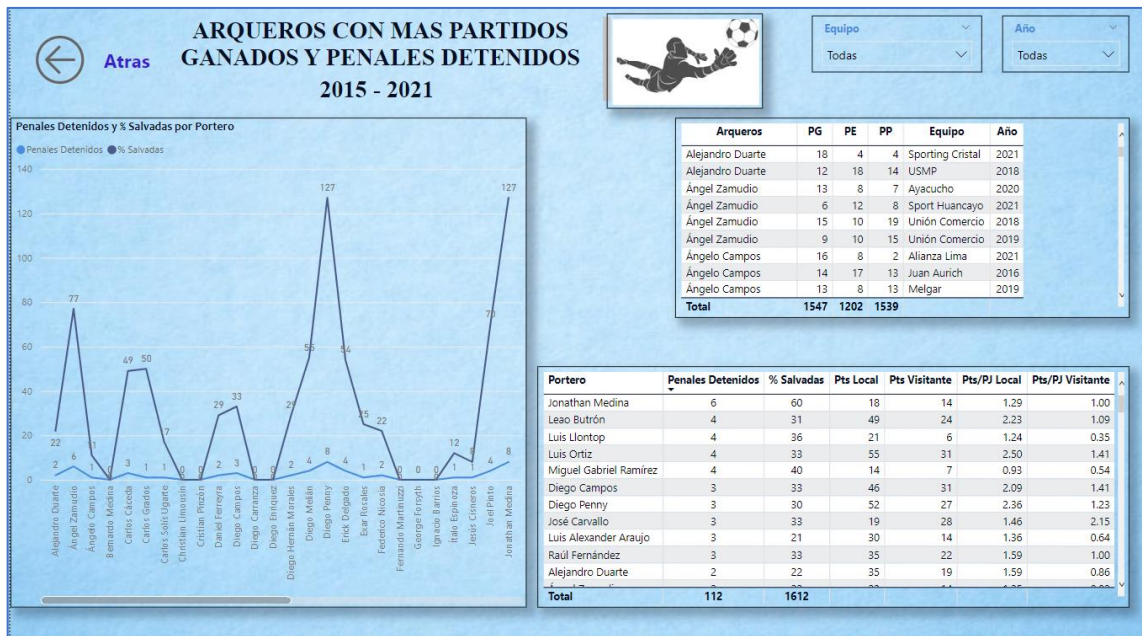


Figura 91 Reporte de arqueros con más partidos ganados y penales detenidos del 2015 -2021



Figura 92 Reporte de faltas deportivas por club deportivo del 2015 al 2021

## Reporte de rendimiento de los jugadores en la costa, sierra y selva del 2014 al 2021

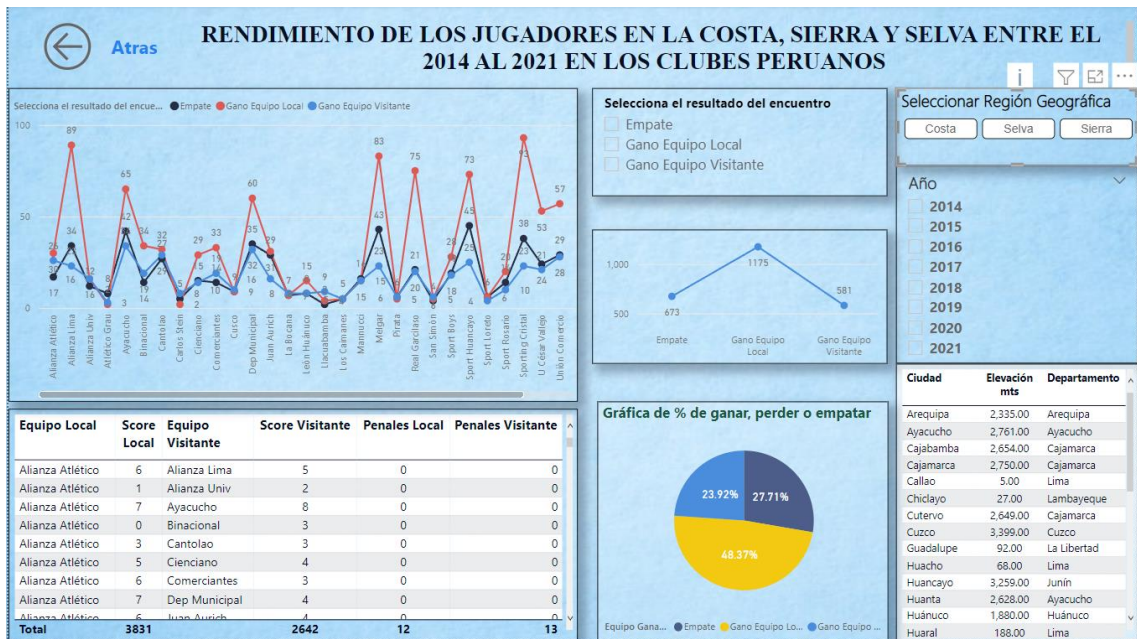


Figura 93 Reporte de rendimiento de los jugadores en la costa, sierra y selva del 2014 - 2021

## Reporte del valor monetario de los jugadores en clubes peruanos al 2021

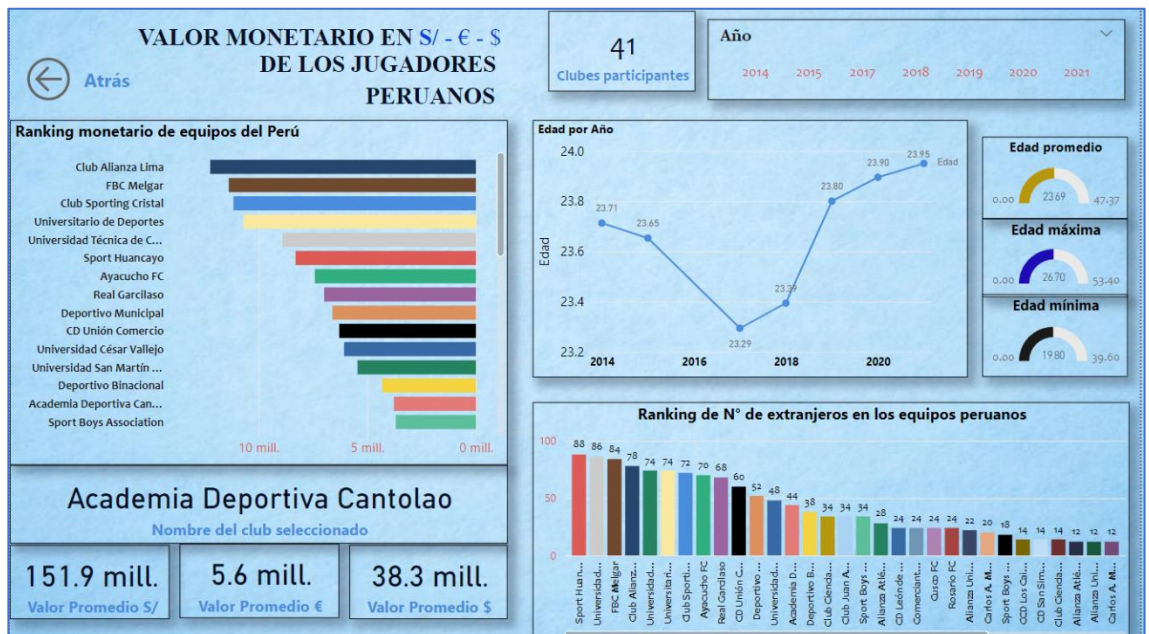


Figura 94 Reporte de Valor monetario promedio en soles, euros y dólares de los jugadores peruanos al 2021

# Reporte de goles marcados por los jugadores en los torneos del



Figura 95 Reporte de influencia de los jugadores en el fútbol peruano en los torneos del 2014 – 2021

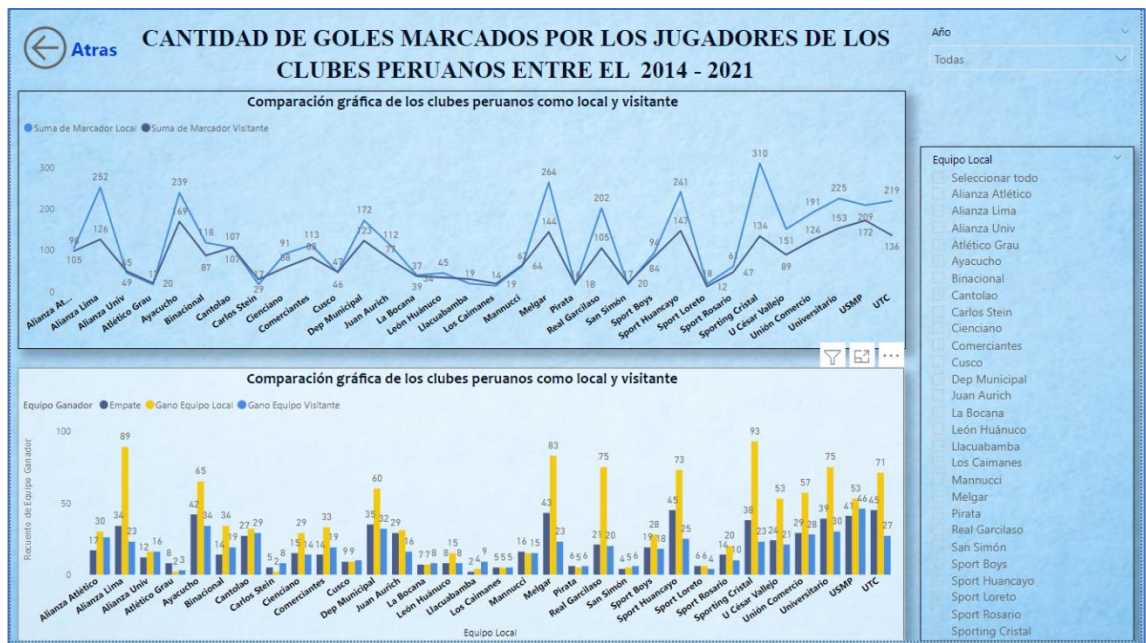


Figura 96 Reporte de cantidad de goles marcados por los jugadores en los clubes peruanos entre el 2014 al 2021

## Otros reportes de interés deportivo

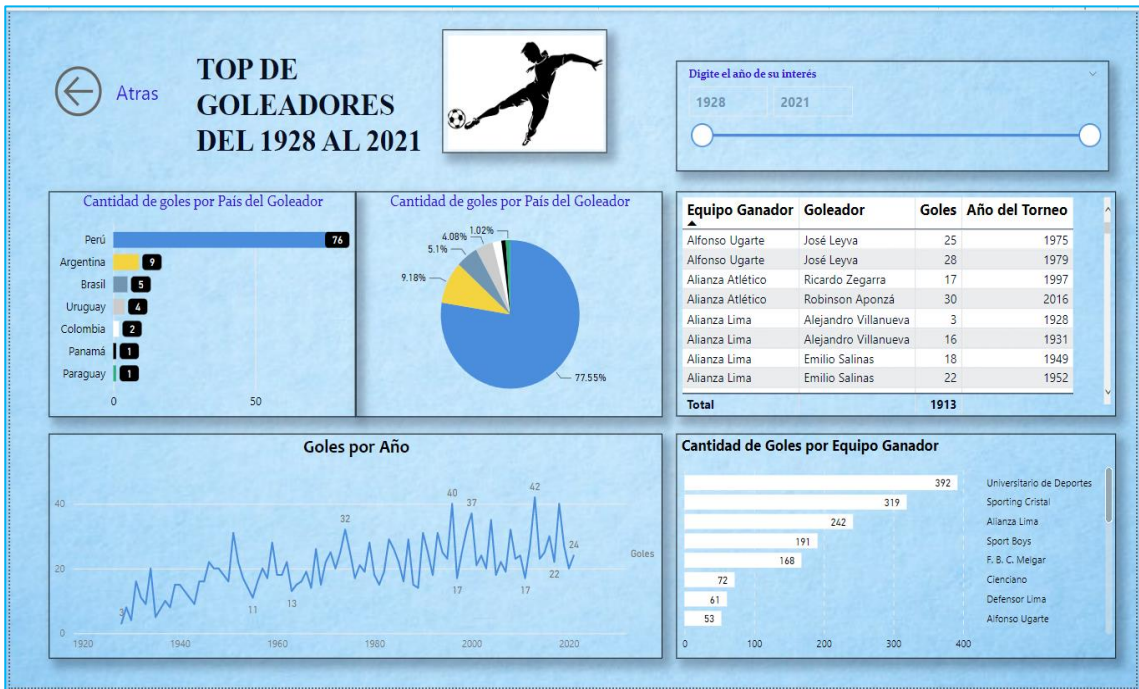


Figura 97 Reporte de Top de Goleadores del 1928 al 2021

### 4.2.9. Almacenamiento y publicación del Dashboard

Para esta parte del desarrollo se tiene que tener en cuenta lo siguiente:

Se debe tener una cuenta corporativa para poder publicar el dashboard a la red.

Una vez obtenido la cuenta se loguea desde la versión de Power BI desktop, como se muestra a continuación:

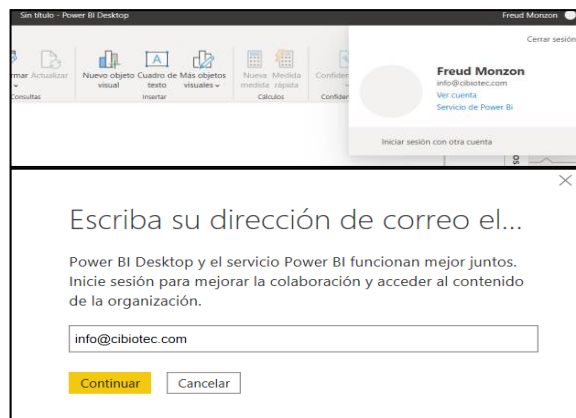


Figura 98 Logueo en Power BI Desktop



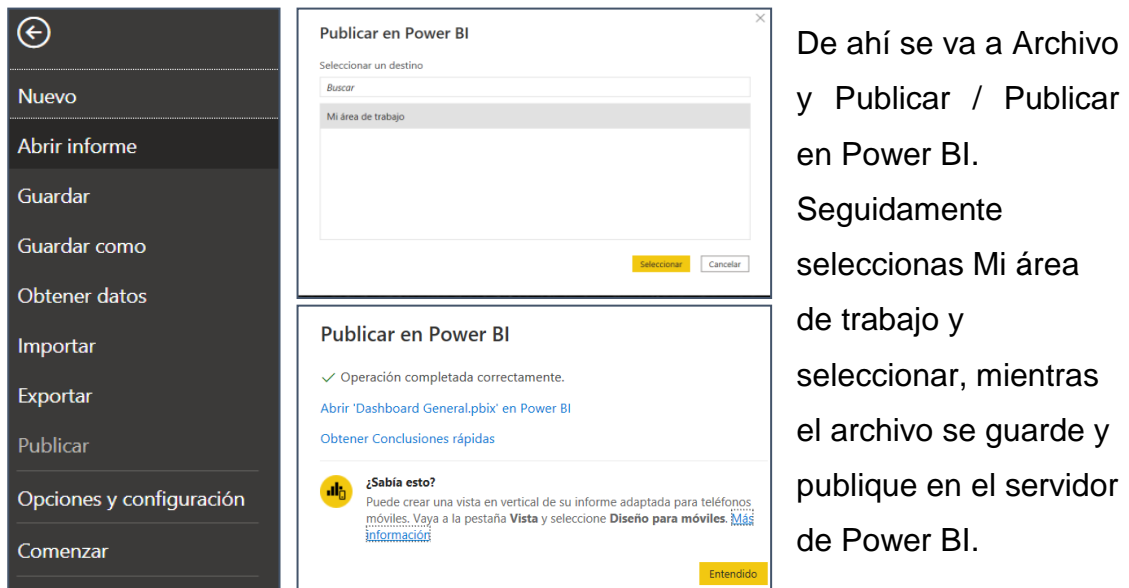


Figura 99 Publicación en Power BI

## 5. DISCUSIÓN DE RESULTADOS

### 5.1. Formulación del problema

¿Cómo realizar un análisis significativo de datos sobre los jugadores de la liga 1 en fútbol peruano?

### 5.2. Hipótesis

La técnica web scraping del plugin webscraper permite recuperar información para el análisis de los datos de jugadores y clubes peruanos es significativa.

### 5.3. Muestra a aplicar

Se aplicó una encuesta online por correo corporativo, a 10 docentes en forma aleatoria de la Universidad Privada Antenor Orrego.

Se evaluó que tan eficiente puede ser el Dashboard.

### 5.4. Validación de la solución

Para la contrastación de la hipótesis, se realizó una encuesta online para poder validar si la hipótesis se acepta o se rechaza. Los datos obtenidos fueron procesados en el programa IBM SPSS Statistics para mostrar sus resultados.

#### 5.4.1. Variable independiente

Para ello se tiene los siguientes datos de la siguiente muestra:

Eficiencia del uso del Dashboard										
Preguntas – Personas	Docente 1	Docente 2	Docente 3	Docente 4	Docente 5	Docente 6	Docente 7	Docente 8	Docente 9	Docente 10
Pregunta 1	5	4	4	4	3	5	4	3	4	4
Pregunta 2	4	4	3	5	4	4	4	3	4	5
Pregunta 3	4	5	5	5	5	5	5	3	4	3
Pregunta 4	5	5	4	5	4	4	3	4	5	4
Pregunta 5	5	4	4	5	3	5	4	5	3	4

Tabla 19 Tabla de datos de la variable independiente

DOCENTE	PRETEST	PROTEST	DIFERENCIA
1	2	4	-2
2	2.1	4	-1.9
3	2	4.8	-2.8
4	2.2	4.6	-2.4
5	3.5	4.2	-0.7
6	3.8	3.8	0
7	3	4	-1
8	2.4	4.2	-1.8
9	2.8	4	-1.2
10	1.3	4.2	-2.9

Una vez ingresados los datos al programa IBM SPSS Statistics, se arrojó los siguientes resultados:

	Válido		Casos Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
PRETEST	10	100.0%	0	0.0%	10	100.0%
POTEST	10	100.0%	0	0.0%	10	100.0%

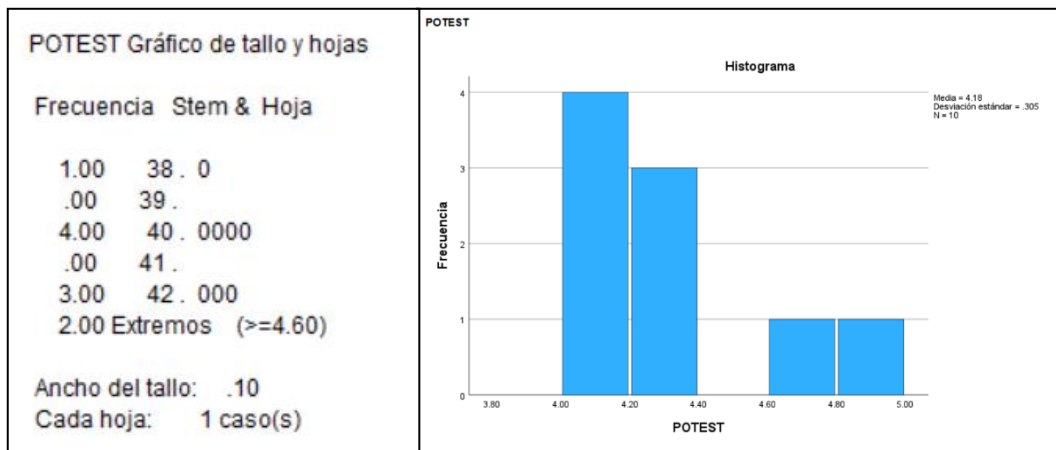
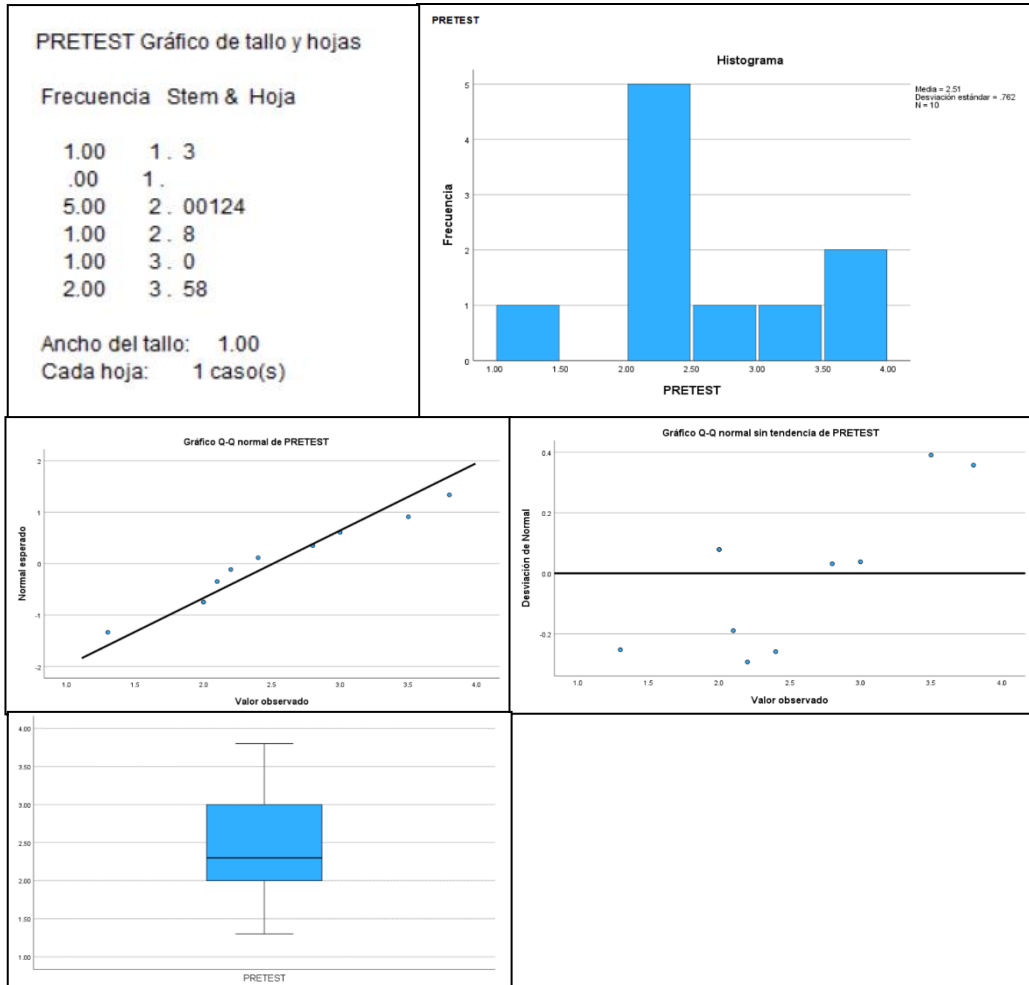
Aquí indica que se ha procesado todos los datos

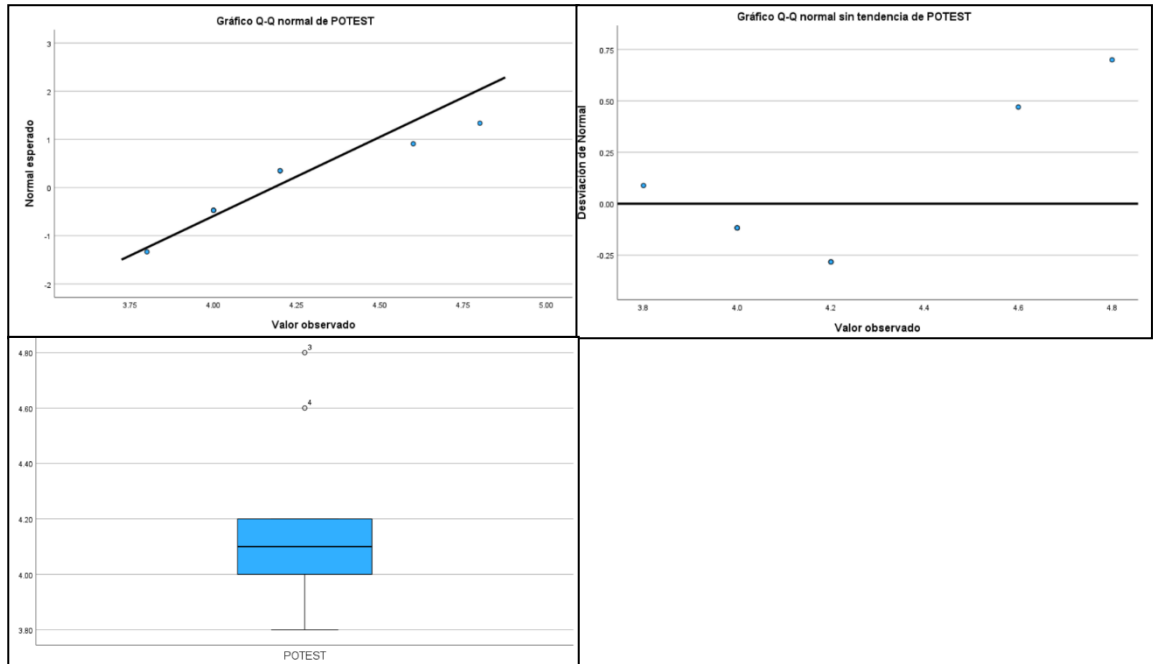
		Estadístico	Error estándar	
PRETEST	Media	2.5100	.24104	
	95% de intervalo de confianza para la media	Límite inferior	1.9647	
		Límite superior	3.0553	
	Media recortada al 5%	2.5056		
	Mediana	2.3000		
	Varianza	.581		
	Desv. estándar	.76223		
	Mínimo	1.30		
	Máximo	3.80		
	Rango	2.50		
	Rango intercuartil	1.13		
	Asimetría	.352	.687	
Curtosis	-.388	1.334		
POTEST	Media	4.1800	.09638	
	95% de intervalo de confianza para la media	Límite inferior	3.9620	
		Límite superior	4.3980	
	Media recortada al 5%	4.1667		
	Mediana	4.1000		
	Varianza	.093		
	Desv. estándar	.30478		
	Mínimo	3.80		
	Máximo	4.80		
	Rango	1.00		
	Rango intercuartil	.30		
	Asimetría	1.149	.687	

Al observar la Media en el pretest (2.3000) y el posttest(4.1), podemos entender que el programa de experimentación tuvo éxito.

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
PRETEST	.158	10	.200*	.956	10	.742
POTEST	.274	10	.032	.857	10	.071

\*. Esto es un límite inferior de la significación verdadera.  
a. Corrección de significación de Lilliefors





## Prueba T-Student

		Media	N	Desv. estándar	Media de error estándar
Par 1	PRETEST	2.5100	10	.76223	.24104
	PROTEST	4.1800	10	.30478	.09638

		N	Correlación	Significación	
				P de un factor	P de dos factores
Par 1	PRETEST & PROTEST	10	-.439	.102	.204

		Standardizer <sup>a</sup>	Estimación de puntos	Intervalo de confianza al 95%		
				Inferior	Superior	
Par 1	PRETEST - PROTEST	d de Cohen	.93696	-1.782	-2.785	-.745
		corrección de Hedges	1.02526	-1.629	-2.545	-.681

a. El denominador utilizado en la estimación de tamaños del efecto.  
 La d de Cohen utiliza la desviación estándar de muestra de la diferencia de medias.  
 La corrección de Hedges utiliza la desviación estándar de muestra de la diferencia de medias, más un factor de corrección.

		Diferencias emparejadas						Significación		
		Media	Desv. estándar	Media de error estándar	95% de intervalo de confianza de la diferencia		t	gl	P de un factor	P de dos factores
					Inferior	Superior				
Par 1	PRETEST - PROTEST	-1.67000	.93696	.29629	-2.34026	-.99974	-5.636	9	<.001	<.001

El grado de significación es: <.001

Si  $p \geq 0.05$ , Aceptamos la  $H_0$  y rechazamos la  $H_A$

Si  $p < 0.05$ , Rechazamos la  $H_0$  y aceptamos la  $H_A$

Entonces podemos decir que se rechaza la  $H_0$  y se acepta la  $H_A$

## 5.4.2. Variable dependiente

¿Consideras que la cantidad de años para el análisis ofrecida en el Dashboard, es la apropiada?										
Preguntas Personas	Docente 1	Docente 2	Docente 3	Docente 4	Docente 5	Docente 6	Docente 7	Docente 8	Docente 9	Docente 10
Pregunta 1	1	1	1	1	0	1	1	0	1	0
Pregunta 2	1	1	1	1	0	1	1	1	1	0
Pregunta 3	1	1	1	1	1	1	1	1	1	1
Pregunta 4	0	0	1	1	0	1	0	1	0	1
Pregunta 5	0	0	0	0	1	0	0	0	0	1

Tabla 20 Tabla de datos de la variable dependiente registros

DOCENTE	PRETESTINFORMACIONAP ROPIADA	PROTESTINFORMACIONAP ROPIADA	DIFERENCIA
1	0.1	0.6	-0.5
2	0.1	0.6	-0.5
3	0.1	0.8	-0.7
4	0.1	0.8	-0.7
5	0.1	0.4	-0.3
6	0.1	0.8	-0.7
7	0.1	0.6	-0.5
8	0.1	0.6	-0.5
9	0.1	0.6	-0.5
10	0.1	0.6	-0.5

Una vez ingresados los datos al programa IBM SPSS Estatistics, se arrojó los siguientes resultados:

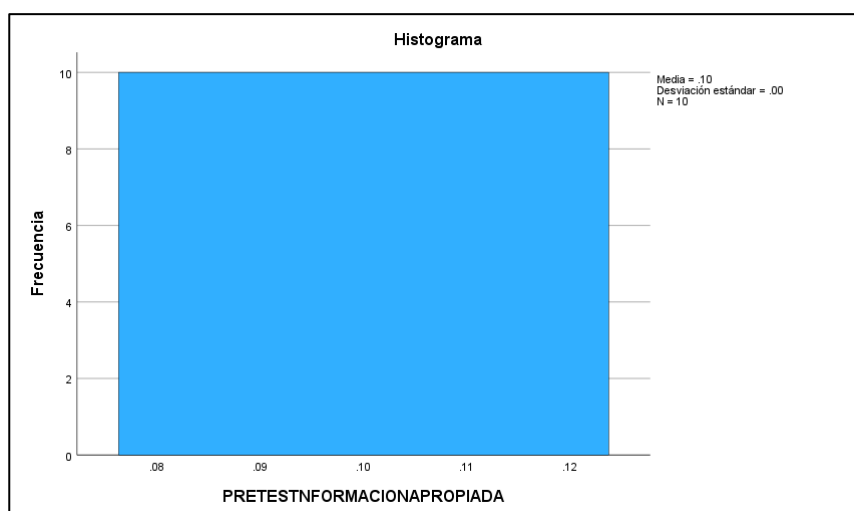
Resumen de procesamiento de casos						
	Válido		Casos Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
PRETESTINFORMACIONAP ROPIADA	10	100.0%	0	0.0%	10	100.0%
PROTESTINFORMACIONA PROPIADA	10	100.0%	0	0.0%	10	100.0%

Descriptivos			Estadístico	Error estándar
PRETESTINFORMACIONAP ROPIADA	Media		.1000	.00000
	95% de intervalo de confianza para la media	Límite inferior	.1000	
		Límite superior	.1000	
	Media recortada al 5%		.1000	
	Mediana		.1000	
	Varianza		.000	
	Desv. estándar		.00000	
	Mínimo		.10	
	Máximo		.10	
	Rango		.00	
	Rango intercuartil		.00	
	Asimetría		.	
	Curtosis		.	
PROTESTINFORMACIONA PROPIADA	Media		.6400	.04000
	95% de intervalo de confianza para la media	Límite inferior	.5495	
		Límite superior	.7305	
	Media recortada al 5%		.6444	
	Mediana		.6000	
	Varianza		.016	
	Desv. estándar		.12649	
	Mínimo		.40	
	Máximo		.80	
	Rango		.40	
	Rango intercuartil		.20	
	Asimetría		-.132	.687
	Curtosis		.179	1.334

Al observar la Media en el pretest (.1000) y el prostest(.6000), podemos entender que el programa de experimentación tuvo éxito

	Pruebas de normalidad					
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
PRETESTINFORMACIONAP ROPIADA	.	10	.	.	10	.
PROTESTINFORMACIONA PROPIADA	.324	10	.004	.794	10	.012

a. Corrección de significación de Lilliefors



PRETESTNFORMACIONAPROPIADA Gráfico de tallo y hojas

Frecuencia Stem & Hoja

10.00 1. 0000000000



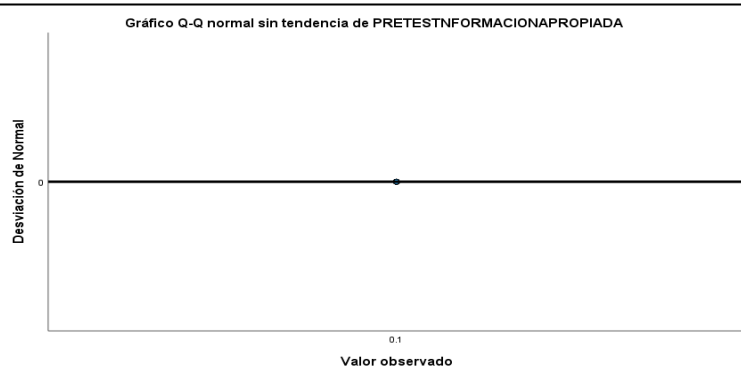
Ancho del tallo: .10

Cada hoja: 1 caso(s)

Gráfico Q-Q normal de PRETESTNFORMACIONAPROPIADA



Gráfico Q-Q normal sin tendencia de PRETESTNFORMACIONAPROPIADA



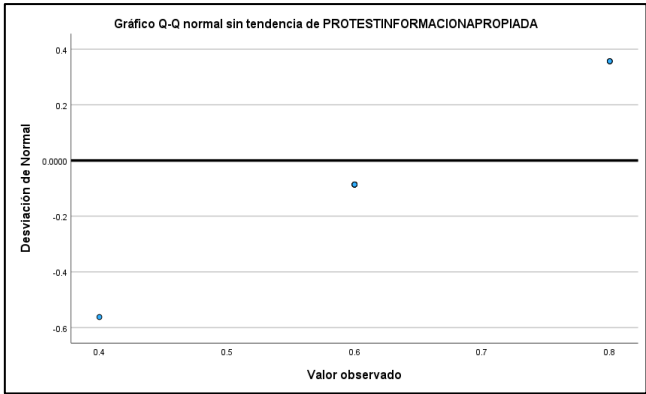
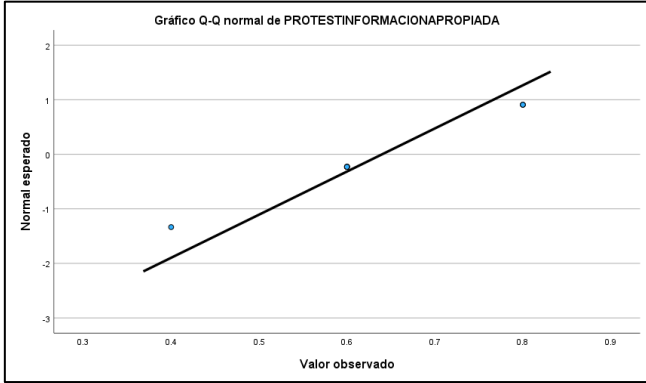
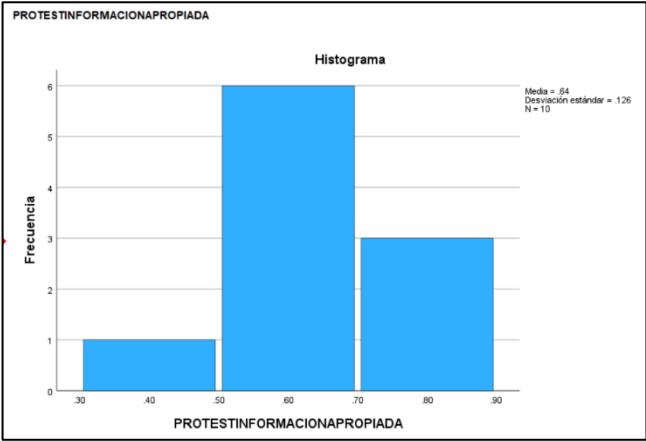


PROTESTINFORMACIONAPROPIADA Gráfico de tallo y hojas

Frecuencia Stem & Hoja

1.00	4 . 0
.00	5 .
6.00	6 . 000000
.00	7 .
3.00	8 . 000

Ancho del tallo: .10  
Cada hoja: 1 caso(s)



## Prueba T- Student

Estadísticas de muestras emparejadas					
		Media	N	Desv. estándar	Media de error estándar
Par 1	PRETESTINFORMACIONAP ROPIADA	.1000	10	.00000	.00000
	PROTESTINFORMACIONA PROPIADA	.6400	10	.12649	.04000

Correlaciones de muestras emparejadas					
		N	Correlación	Significación	
				P de un factor	P de dos factores
Par 1	PRETESTINFORMACIONAP ROPIADA & PROTESTINFORMACIONA PROPIADA	10	.	.	.

Tamaños de efecto de muestras emparejadas						
			Standardizer <sup>a</sup>	Estimación de puntos	Intervalo de confianza al 95%	
					Inferior	Superior
Par 1	PRETESTINFORMACIONAP ROPIADA - PROTESTINFORMACIONA PROPIADA	d de Cohen	.12649	-4.269	-6.294	-2.229
		corrección de Hedges	.13841	-3.901	-5.752	-2.037

a. El denominador utilizado en la estimación de tamaños del efecto.  
La d de Cohen utiliza la desviación estándar de muestra de la diferencia de medias.  
La corrección de Hedges utiliza la desviación estándar de muestra de la diferencia de medias, más un factor de corrección.

Prueba de muestras emparejadas										
		Diferencias emparejadas						Significación		
		Media	Desv. estándar	Media de error estándar	95% de intervalo de confianza de la diferencia		t	gl	P de un factor	P de dos factores
					Inferior	Superior				
Par 1	PRETESTINFORMACIONAP ROPIADA - PROTESTINFORMACIONA PROPIADA	-.54000	.12649	.04000	-.63049	-.44951	-13.500	9	<.001	<.001

El grado de significación es: <.001

Si  $p \geq 0.05$ , Aceptamos la  $H_0$  y rechazamos la  $H_A$

Si  $p < 0.05$ , Rechazamos la  $H_0$  y aceptamos la  $H_A$

Entonces podemos decir que se rechaza la  $H_0$  y se acepta la  $H_A$

¿Consideras el tiempo para el análisis ofrecida en el Dashboard, es la apropiada?										
Preguntas	Docente 1	Docente 2	Docente 3	Docente 4	Docente 5	Docente 6	Docente 7	Docente 8	Docente 9	Docente 10
Pregunta 1	1	1	0	1	1	0	1	0	1	0
Pregunta 2	1	0	0	1	0	1	1	1	1	1

Tabla 21 Tabla de datos de la variable dependiente tiempo

DOCENTE	PRETESTTIEMPOANALISIS	PROTESTTIEMPOANALISIS	DIFERENCIA
1	0.5	1	-0.5
2	0.5	0.5	0
3	0.5	0	0.5
4	0.5	1	-0.5
5	0	0.5	-0.5
6	0.5	0.5	0
7	0.5	1	-0.5
8	0.5	0.5	0
9	0.5	1	-0.5
10	0	0.5	-0.5

Una vez ingresados los datos al programa IBM SPSS Estatistics, se arrojó los siguientes resultados:

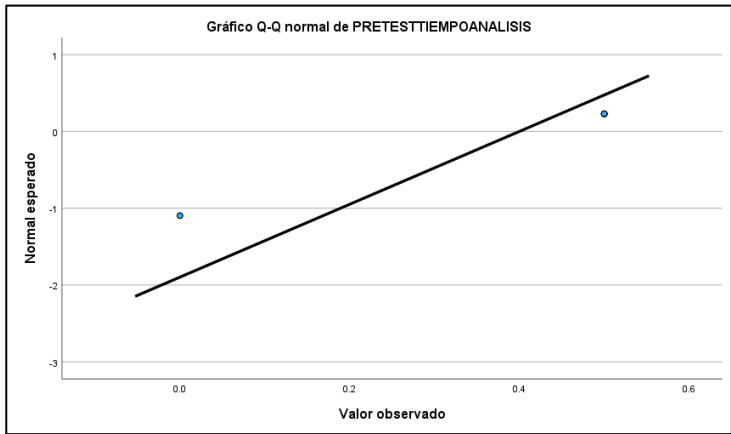
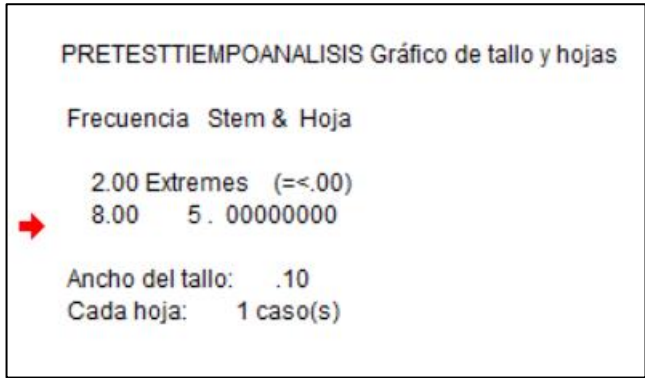
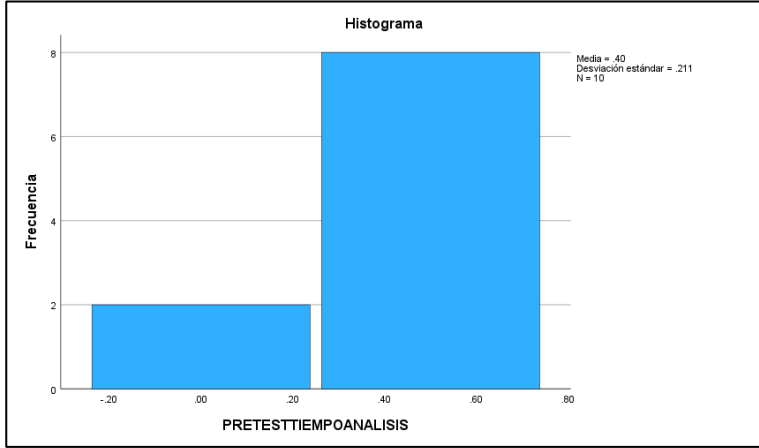
Resumen de procesamiento de casos						
	Válido		Casos Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
PRETESTTIEMPOANALISIS	10	100.0%	0	0.0%	10	100.0%
PROTESTTIEMPOANALISIS	10	100.0%	0	0.0%	10	100.0%

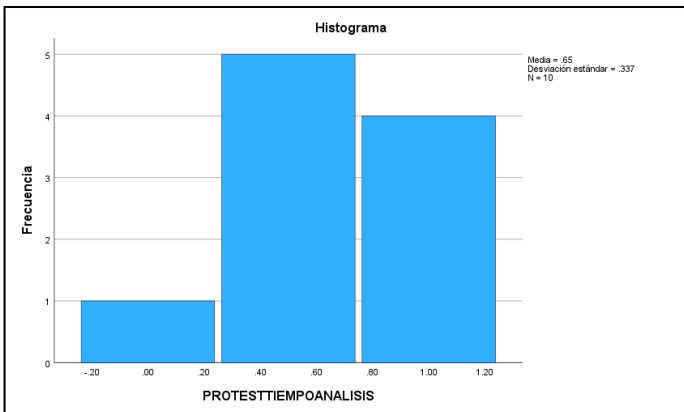
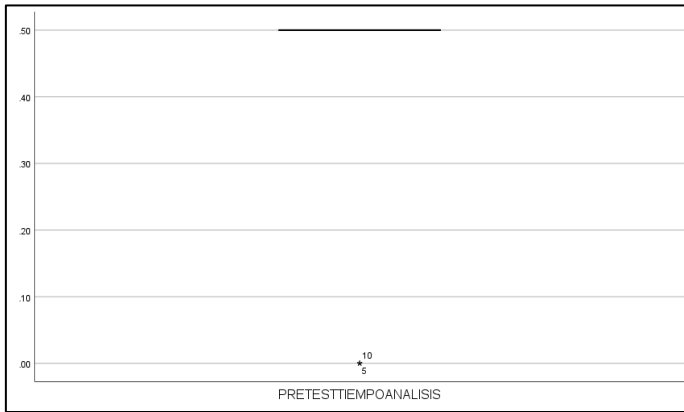
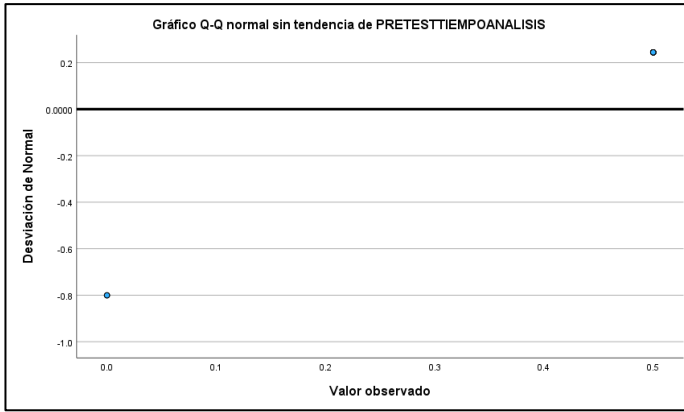
Descriptivos			Estadístico	Error estándar
PRETESTTIEMPOANALISIS	Media		.4000	.06667
	95% de intervalo de confianza para la media	Límite inferior	.2492	
		Límite superior	.5508	
	Media recortada al 5%		.4167	
	Mediana		.5000	
	Varianza		.044	
	Desv. estándar		.21082	
	Mínimo		.00	
	Máximo		.50	
	Rango		.50	
	Rango intercuartil		.13	
	Asimetría		-1.779	.687
	Curtosis		1.406	1.334
	PROTESTTIEMPOANALISIS	Media		.6500
95% de intervalo de confianza para la media		Límite inferior	.4086	
		Límite superior	.8914	
Media recortada al 5%			.6667	
Mediana			.5000	
Varianza			.114	
Desv. estándar			.33747	
Mínimo			.00	
Máximo			1.00	
Rango			1.00	
Rango intercuartil			.50	
Asimetría			-.434	.687
Curtosis			-.283	1.334

Al observar la Media en el pretest (5.000) y el posttest(5.000), podemos entender que el programa de experimentación no tuvo éxito.

Pruebas de normalidad						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
PRETESTTIEMPOANALISIS	.482	10	<.001	.509	10	<.001
PROTESTTIEMPOANALISIS	.272	10	.035	.802	10	.015

a. Corrección de significación de Lilliefors



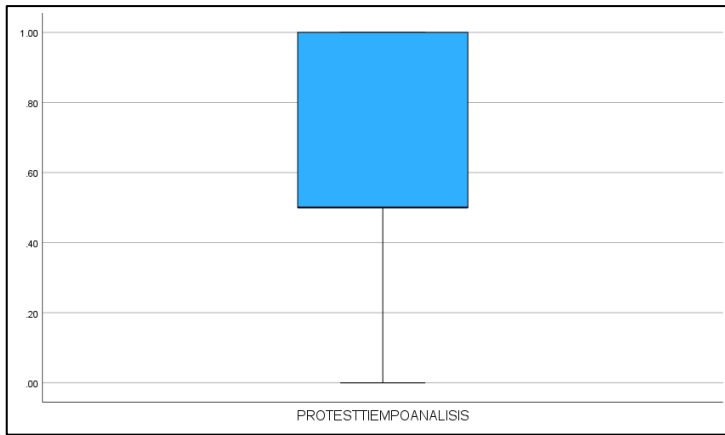
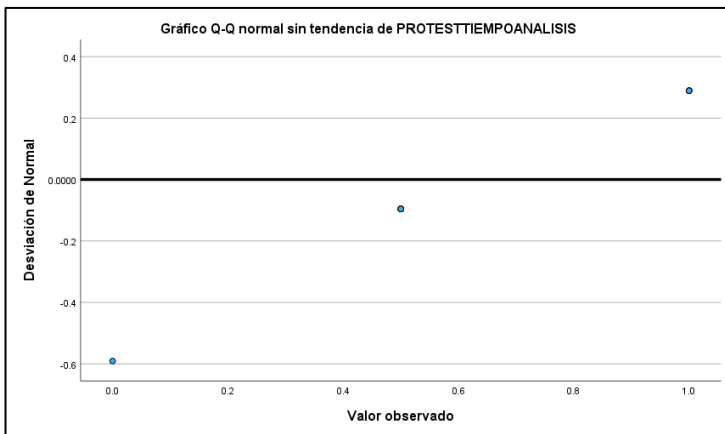
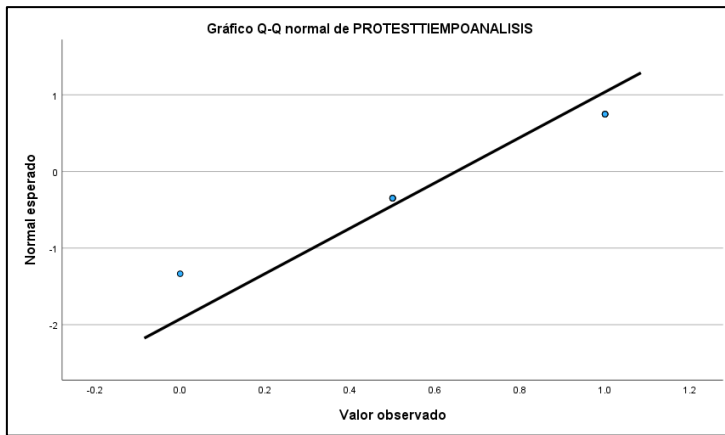


PRETESTTIEMPOANALISIS Gráfico de tallo y hojas

Frecuencia Stem & Hoja

1.00	0 . 0
5.00	0 . 55555
4.00	1 . 0000

Ancho del tallo: 1.00  
Cada hoja: 1 caso(s)



## Prueba T-Student

Estadísticas de muestras emparejadas					
		Media	N	Desv. estándar	Media de error estándar
Par 1	PRETESTTIEMPOANALISIS	.4000	10	.21082	.06667
	S				
	PROTESTTIEMPOANALISIS	.6500	10	.33747	.10672
	S				

Correlaciones de muestras emparejadas					
		N	Correlación	Significación	
				P de un factor	P de dos factores
Par 1	PRETESTTIEMPOANALISIS & PROTESTTIEMPOANALISIS	10	.234	.257	.515

Tamaños de efecto de muestras emparejadas						
		d de Cohen	Standardizer <sup>a</sup>	Estimación de puntos	Intervalo de confianza al 95%	
					Inferior	Superior
Par 1	PRETESTTIEMPOANALISIS - PROTESTTIEMPOANALISIS		.35355	-.707	-1.390	.006
		corrección de Hedges	.38687	-.646	-1.271	.006

a. El denominador utilizado en la estimación de tamaños del efecto. La d de Cohen utiliza la desviación estándar de muestra de la diferencia de medias. La corrección de Hedges utiliza la desviación estándar de muestra de la diferencia de medias, más un factor de corrección.

Prueba de muestras emparejadas										
		Diferencias emparejadas					Significación			
		Media	Desv. estándar	Media de error estándar	95% de intervalo de confianza de la diferencia		t	gl	P de un factor	P de dos factores
					Inferior	Superior				
Par 1	PRETESTTIEMPOANALISIS - PROTESTTIEMPOANALISIS	-.25000	.35355	.11180	-.50292	.00292	-2.236	9	.026	.052

El grado de significación es:  $>.001$

Si  $p \geq 0.05$ , Aceptamos la  $H_0$  y rechazamos la  $H_A$   
 Si  $p < 0.05$ , Rechazamos la  $H_0$  y aceptamos la  $H_A$

Entonces podemos decir que se rechaza la  $H_A$  y se acepta la  $H_0$

¿Consideras que la información ofrecida en el Dashboard es útil ?											
Preguntas Personas	Docentes										
	Docente 1	Docente 2	Docente 3	Docente 4	Docente 5	Docente 6	Docente 7	Docente 8	Docente 9	Docente 10	
Pregunta 8	1	1	1	1	1	1	1	1	1	1	
Pregunta 9	1	1	1	1	0	1	1	1	1	1	
Pregunta 10	1	0	1	1	0	1	0	1	1	1	

Tabla 22 Tabla de datos de la variable dependiente utilidad

DOCENTE	PRETESTUTIL	PROTESTUTIL	DIFERENCIA
1	0.67	1	-0.33
2	0.67	0.67	0.00
3	0.67	1	-0.33
4	0.67	1	-0.33
5	0.67	0.33	0.34
6	0.67	1	-0.33
7	0.67	0.67	0.00
8	0.67	1	-0.33
9	0.67	1	-0.33
10	0.67	1	-0.33

Resumen de procesamiento de casos						
	Válido		Casos Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
PRETESTUTIL	10	100.0%	0	0.0%	10	100.0%
PROTESTUTIL	10	100.0%	0	0.0%	10	100.0%

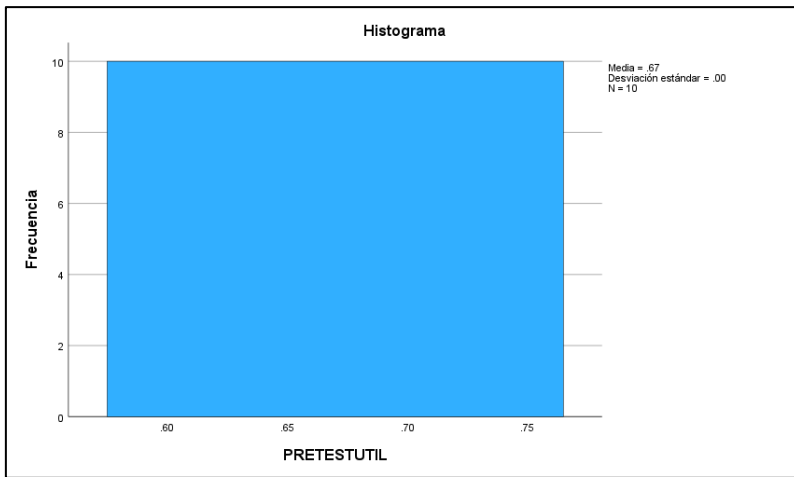
Descriptivos					
		Estadístico	Error estándar		
PRETESTUTIL	Media	.6700	.00000		
	95% de intervalo de confianza para la media	Límite inferior	.6700		
		Límite superior	.6700		
	Media recortada al 5%	.6700			
	Mediana	.6700			
	Varianza	.000			
	Desv. estándar	.00000			
	Mínimo	.67			
	Máximo	.67			
	Rango	.00			
	Rango intercuartil	.00			
	Asimetría	.			
	Curtosis	.			
	PROTESTUTIL	Media	.8670	.07377	
95% de intervalo de confianza para la media		Límite inferior	.7001		
		Límite superior	1.0339		
Media recortada al 5%		.8894			
Mediana		1.0000			
Varianza		.054			
Desv. estándar		.23329			
Mínimo		.33			
Máximo		1.00			
Rango		.67			
Rango intercuartil		.33			
Asimetría		-1.681	.687		
Curtosis		2.172	1.334		

Al observar la Media en el pretest (6.700) y el posttest(1.0000), podemos entender que el programa de experimentación tuvo éxito.

Pruebas de normalidad						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
PRETESTUTIL	.	10	.	.	10	.
PROTESTUTIL	.416	10	<.001	.649	10	<.001

a. Corrección de significación de Lilliefors

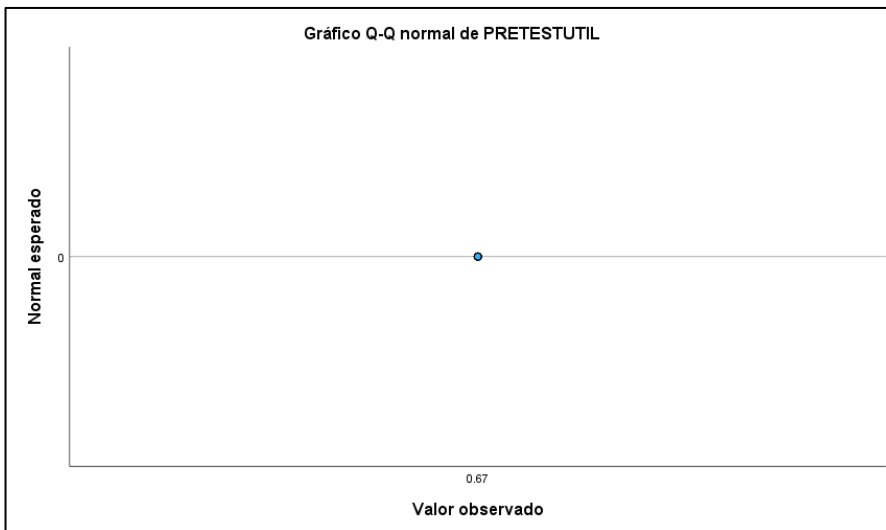


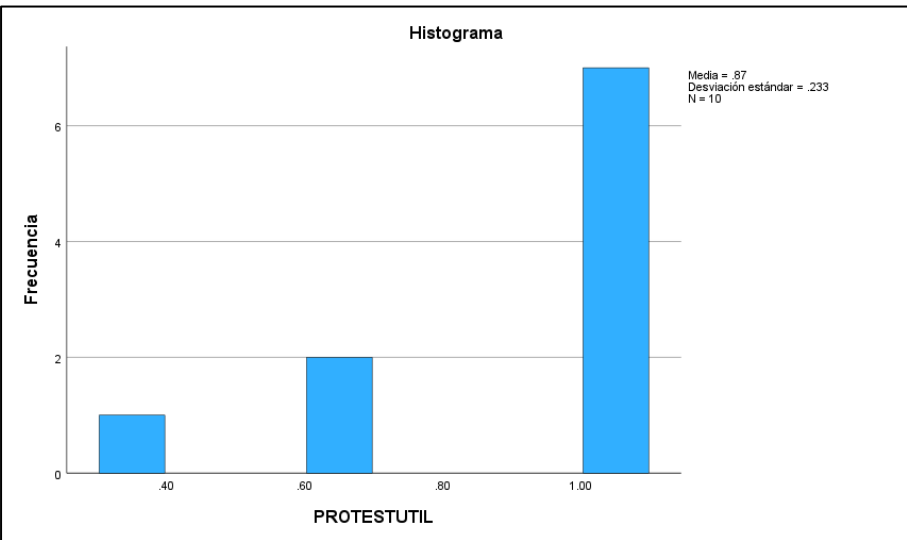
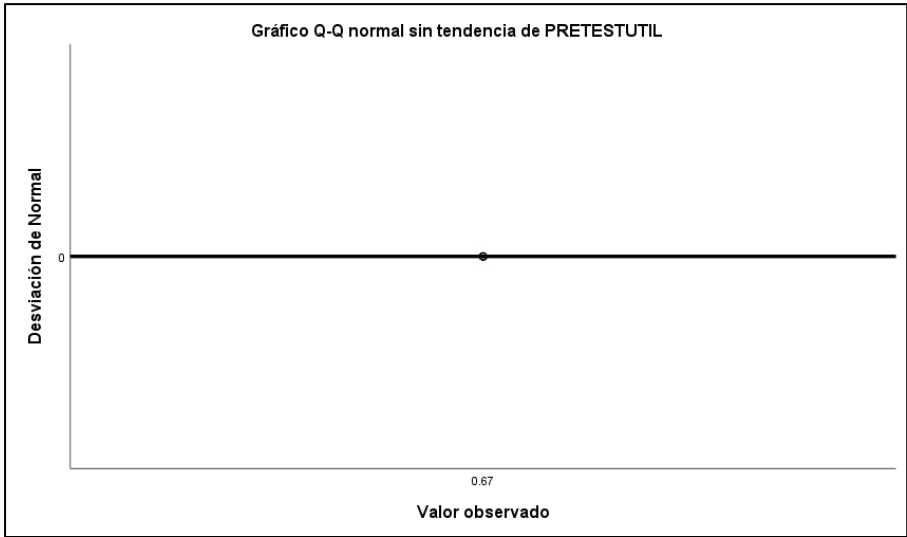


**PRETESTUTIL Gráfico de tallo y hojas**

Frecuencia	Stem & Hoja
10.00	6 . 7777777777

Ancho del tallo: .10  
Cada hoja: 1 caso(s)



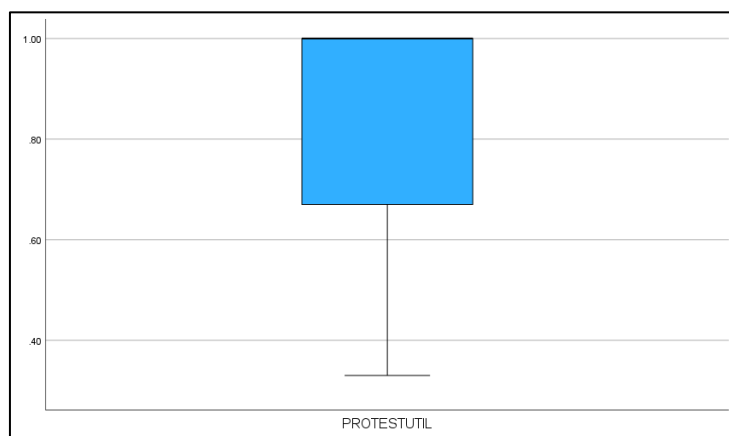
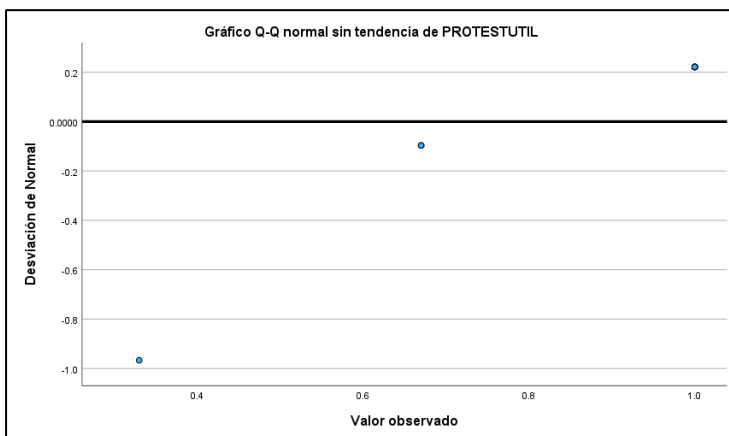
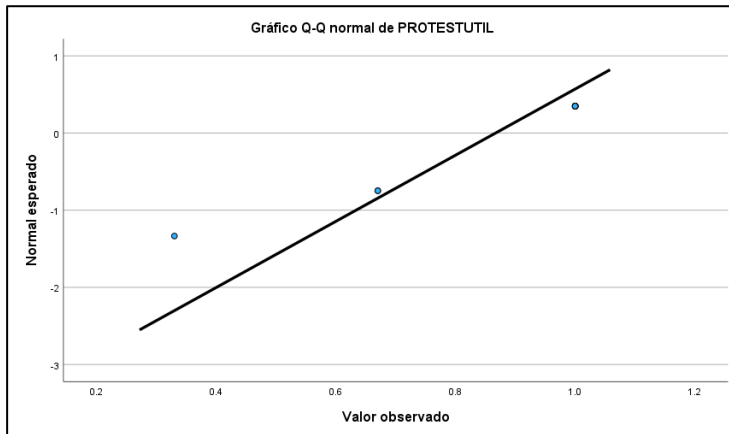


PROTESTUTIL Gráfico de tallo y hojas

Frecuencia Stem & Hoja

1.00	0. 3
2.00	0. 66
7.00	1. 0000000

Ancho del tallo: 1.00  
Cada hoja: 1 caso(s)



## Prueba T- Student

Estadísticas de muestras emparejadas					
		Media	N	Desv. estándar	Media de error estándar
Par 1	PRETESTUTIL	.6700	10	.00000	.00000
	PROTESTUTIL	.8670	10	.23329	.07377

Correlaciones de muestras emparejadas					
		N	Correlación	Significación	
				P de un factor	P de dos factores
Par 1	PRETESTUTIL & PROTESTUTIL	10	.	.	.

Tamaños de efecto de muestras emparejadas						
			Standardizer <sup>a</sup>	Estimación de puntos	Intervalo de confianza al 95%	
					Inferior	Superior
Par 1	PRETESTUTIL - PROTESTUTIL	d de Cohen	.23329	-.844	-1.557	-.099
		corrección de Hedges	.25527	-.772	-1.423	-.090

a. El denominador utilizado en la estimación de tamaños del efecto.  
 La d de Cohen utiliza la desviación estándar de muestra de la diferencia de medias.  
 La corrección de Hedges utiliza la desviación estándar de muestra de la diferencia de medias, más un factor de corrección.

Prueba de muestras emparejadas										
		Diferencias emparejadas						Significación		
		Media	Desv. estándar	Media de error estándar	95% de intervalo de confianza de la diferencia		t	gl	P de un factor	P de dos factores
					Inferior	Superior				
Par 1	PRETESTUTIL - PROTESTUTIL	-.19700	.23329	.07377	-.36388	-.03012	-2.670	9	.013	.026

El grado de significación es: <.026

Si  $p \geq 0.05$ , Aceptamos la  $H_0$  y rechazamos la  $H_A$

Si  $p < 0.05$ , Rechazamos la  $H_0$  y aceptamos la  $H_A$

Entonces podemos decir que se rechaza la  $H_0$  y se acepta la  $H_A$

## 6. CONCLUSIONES

- Según el objetivo general, el análisis de la información futbolística de la liga 1, es un reto por la poca información y la homogeneidad de los datos.

Pero aun así se pudo determinar las 5 primeras posiciones preferidas por los futbolistas peruanos, donde la primera posición de juego es ser delantero central, seguidamente de mediocampista ofensivo, defensa central, los extremos izquierdo y derecho, siendo relegadas el resto de posiciones, siendo la posición de defensa la menos preferida por los futbolistas peruanos.

Ahora si hablamos de la posición de arqueros, se puede visualizar a cuatro entre los tops, los cuales son: Diego Penny (Melgar/Sporting Cristal/USMP), Jonathan Medina (Sport Boys), Jesús Cisneros (Juan Aurich/León Huánuco), José Carvallo (UTC/Universitario) con sus respectivos clubes, desde el 2015 hasta el 2021.

Otro dato interesante es que los jugadores del club deportivo de la Universidad de San Martín de Porres (USMP), tienen el top de tarjetas rojas con 59 amonestaciones y de 36 amonestaciones de segundas amarillas, y el club centro deportivo Municipal con 373 amarillas, entre los años 2015 al 2021.

Otro punto de análisis es la región geográfica (Altitud) donde se realizan los encuentros deportivos entre los futbolistas peruanos, y con que frecuencia se desarrollan, entre la costa, la sierra y la selva. Para empezar de las 3 regiones geográficas. Para empezar en la región geográfica de la costa, se determinó que Sporting Cristal ganó 92 partidos, seguido de Alianza Lima con 89 partidos y finalmente con Universitario de deportes con 75 partidos. Ahora si nos vamos a la región geográfica de la sierra, como local gana en primer lugar el club Real Garcilaso con 75 partidos, en segundo lugar, con el club Melgar con 74 partidos y finalmente el club UTC con 64 partidos ganados como local. Finalmente, en la región geográfica de la selva, el club

ganador representate fue el club Unión comercio 57 partidos ganados como local.

Ante este panorama podemos decir que el porcentaje de que un equipo gane como local es del 48.37%, de un 23.92% de que pierda, y de un 27.71 de que empaten ambos equipos, sin importar la región geográfica.

Ahora si nos vamos a la cantidad de encuentros deportivos por región geográfica, podemos encontrar a 5 ciudades en el top, las cuales son Lima con 783 encuentros, el Callao con 254 encuentros, Cuzco con 151 encuentros, Arequipa con 139 encuentros y finalmente Huancayo con 126 encuentros deportivos. Se puede entender que hay una preferencia de encuentros deportivos en la costa, con valor porcentual del 57.17%, seguida por la sierra de un 37.41% de encuentros deportivos, para finalmente relegar a la selva con apenas un 5.44 % como valor porcentual.

Otro dato encontrado es que el club peruano más caro desde el año 2014 al 2021, es el Alianza Lima con un valor de 5.3 mil millones de soles y el Club Deportivo San Simón con un valor comercial en soles de 12.8 millones de soles.

Otro dato interesante es que se encontró que los primeros 5 clubes de Sporting Cristal, Club Manucci, Club Los caimanes, Club Melgar y el Sport Boys, tienen los mejores puestos en los torneos del apertura y clausura, entre los años 2014 al 2021. Otro punto importante es la cantidad de goles marcados como local y visitante, siendo los clubes el Sporting Cristal, Melgar, Alianza, Sport Huancayo, y Ayacucho.

Hay otros datos de interés es la clasificación internacional de los clubes peruanos, donde resalta el Melgar con 8 participaciones, 4 en la copa libertadores y 4 en la copa sudamericana, así como el campeón y subcampeón desde el año 1921 hasta el 2021.

- Según el objetivo específico número uno, se estableció que la página web de transfermarkt y fbref, proporcionaron la información futbolística de la liga 1, entre los años 2014 al 2021.

- Según el objetivo específico número dos, se elaboró en dos fases el desarrollo del dashboard mediante la metodología SCRUM. La primera fase fue la planificación de los Sprint, y la segunda fase fue el desarrollo de la creación de reportes, todo en un periodo de 7 meses.
- Según el objetivo específico número tres, se utilizó la extensión scraper de Google Chrome como técnica de web Scraping en los sitios webs seleccionados, para la obtención de los datos a procesar.
- Finalmente, según el objetivo específico número cuatro, se construyó un Dashboard en Power BI con los datos obtenidos de páginas online mencionadas e interactuar con ellos.

Posteriormente se basó la estructura de la información en cinco puntos principales, como la de un club en conocer los datos de un jugador para su fichaje, en el desempeño de los clubes deportivos por región geográfica (Altitud), así como de conocer el valor promedio de un equipo peruano. También se tomó en cuenta los detalles de los diversos encuentro deportivos, como puesto en el torneo de interés, así como las veces que un club gano como local o visitante, el top de goleadores en el tiempo, la clasificación internacional de los clubes peruanos, así como el campeón y subcampeón de los diversos encuentros deportivos en el tiempo.

Ante lo expuesto se quiere tener una base referente de datos, para el análisis de los datos y plasmarlos en un reporte general.

## 7. RECOMENDACIONES

- Para la implementación más avanzada del dashboard, se debe tener información más detallada para dicho análisis, como formaciones deportivas usadas por los clubes, métricas físicas de jugador, valores económicos de cada jugador, métricas de tiempos de los partidos, entre otras. Solo así se podría hacer una línea de tiempo de la valoración física y económica de cada jugador y sus clubes, pudieran elevar el uso más óptimo del dashboard y pueda sustentarse en el tiempo, ya que la información histórica es un respaldo inamovible para el análisis de datos.
- Tener más web online futbolísticas, más estructuradas y detalladas con información de las diferentes ligas, puede ayudar con la obtención de datos, pero también hay que tener en cuenta la estructura de datos y la homogeneidad de la misma. Si no se tiene la misma estructura en los datos, puede volverse contradictorio y no ser una información verificable y confiable de consultar.
- La metodología Scrum es de gran ayuda para el control de tareas, pero se tiene que tener compromiso para la realización de tareas, siendo su uso diario una necesidad para concluir objetivos granulares. Si no puede costearse alguna herramienta online gratuita, puede descargar plantillas online gratuitas de Excel, ya que los servicios de las páginas para usar la metodología SCRUM, solo te dan por tiempo limitado y los proyectos de este tipo pueden demorar hasta más de 1 año.
- Hay muchas técnicas de scraping, desde las más complejas o simples en su desarrollo, o desde las pagadas hasta las gratuitas. Uno tiene que resolver cual se adapta a sus necesidades. Otro punto a tener en cuenta en este tipo de páginas, es su modificación continua, siendo una desventaja una herramienta de scraping en páginas de estructura dinámica, por ser más difíciles de implementar.
- La línea de aprendizaje de Power BI en sus inicios es alta, debido a su fácil uso mediante botones y acciones. Incluso tiene botones de pago según sus necesidades. También tiene un uso de soluciones mediante códigos para



desarrolladores especializados o presentaciones muy complejas, por eso se aconseja llevar una información lo más estructura y limpia para una rápida generación de reportes, así no se recarga la plataforma de Power BI.

## REFERENCIAS BIBLIOGRAFICAS

- Arimetrics. (25 de Febrero de 2021). Recuperado el 29 de Mayo de 2022, de <https://www.arimetrics.com/glosario-digital/power-bi>
- Arts, D. (20 de Mayo de 2022). *Alteryx*. Recuperado el 28 de Mayo de 2022, de Castor + The Information Lab, Netherlands Analyzes COVID-19 Medical Research to Saves Lives with Alteryx: <https://www.alteryx.com/es-419/customer-center/castor-the-information-lab-netherlands-analyzes-medical-research-to-saves-lives-with-alteryx>
- Bulmaro, N. (25 de Febrero de 2021). *Culturación*. Recuperado el 28 de Mayo de 2022, de <https://culturacion.com/cual-es-la-diferencia-entre-dato-e-informacion/>
- Byte TI. (15 de Septiembre de 2021). *Revista Byte*. Recuperado el 28 de Mayo de 2022, de <https://revistabyte.es/actualidad-it/big-data/plataforma-analisis-de-datos/>
- CE Entel. (14 de Junio de 2021). Recuperado el 29 de Mayo de 2022, de <https://ce.entel.cl/grandes-empresas/articulos/pentaho-data-integration/>
- Data Analysis. (26 de Mayo de 2022). *hackr.io*. Recuperado el 28 de Mayo de 2022, de <https://hackr.io/blog/what-is-data-analysis-methods-techniques-tools>
- De Miguel Toquero, A. (27 de Marzo de 2020). *IMPLEMENTACIÓN DE UNA HERRAMIENTA DE ANÁLISIS EN LA GESTIÓN DEPORTIVA MEDIANTE POWER BI*. Proyecto de investigación, UNIVERSIDAD DE VALLADOLID, ESCUELA DE INGENIERIAS INDUSTRIALES, Valladolid. Recuperado el 05 de Octubre de 2021, de <https://uvadoc.uva.es/bitstream/handle/10324/45062/TFG-I-1752.pdf?sequence=1&isAllowed=y>
- FIFA. (10 de Marzo de 2022). *fifa*. Recuperado el 28 de Mayo de 2022, de <https://www.fifa.com/es/tournaments/mens/arabcup/arabcup2021/media-releases/el-futuro-es-ahora-la-fifa-lleva-el-analisis-del-rendimiento-a-una-dimension>
- Freedom & Flow Company. (2016). *El Small Data de la Gestión de Personas: Impacto de la Innovación en Ejercicio Físico, Salud y Tecnología, en las Competencias Profesionales*. (F. & Company, Ed.) 44. Recuperado el

04 de Junio de 2021, de <https://munideporte.com/imagenes/documentacion/ficheros/02AED20F.pdf>

KIO NETWORKS. (20 de Mayo de 2022). *KIO*. Recuperado el 28 de Mayo de 2022, de <https://www.kionetworks.com/frontierpost/cloud-evolucion-del-analisis-de-datos-en-el-deporte>

LISA INSURTECH. (12 de Diciembre de 2021). *LISA*. Recuperado el Sábado de Junio de 2022, de <https://www.lisainsurtech.com/es/2021/12/22/aprende-que-es-el-analisis-de-datos-y-su-importancia/>

López Vanegas, M. (01 de Diciembre de 2020). *Academia Pragma*. Recuperado el 28 de Mayo de 2022, de <https://www.pragma.com.co/blog/los-datos-como-base-de-la-economia>

Macias LLoret, M., Julbe, F., & Nin, J. (2017). *Gestión de bigdata: Tecnologías*. Material Docente de la UOC, Universitat Oberta de Catalunya. Recuperado el 28 de Mayo de 2022, de [http://openaccess.uoc.edu/webapps/o2/bitstream/10609/136328/2/Gestio%20de%20big%20data%3B%20tecnologi%20as\\_Mo%20dulo3\\_Ana%20lisis%20y%20visualizacio%20de%20datos.pdf](http://openaccess.uoc.edu/webapps/o2/bitstream/10609/136328/2/Gestio%20de%20big%20data%3B%20tecnologi%20as_Mo%20dulo3_Ana%20lisis%20y%20visualizacio%20de%20datos.pdf)

Microsoft. (04 de Junio de 2021). Recuperado el 28 de Mayo de 2022, de [https://www.google.com/search?q=dashboard+define&biw=1366&bih=638&tbs=qdr%3Ay&sxsrf=ALiCzsai82Pib\\_TYJEC8lxx7mWM6Tf7rMw%3A1654101554134&ei=MpaXYo\\_uB\\_KR5OUPivqfgA0&ved=0ahUKEwiPz7m82lz4AhXyCLkGHQr9B9AQ4dUDCA4&uact=5&oq=dashboard+define&gs\\_lcp=Cgdnd3Mtd2l6EA](https://www.google.com/search?q=dashboard+define&biw=1366&bih=638&tbs=qdr%3Ay&sxsrf=ALiCzsai82Pib_TYJEC8lxx7mWM6Tf7rMw%3A1654101554134&ei=MpaXYo_uB_KR5OUPivqfgA0&ved=0ahUKEwiPz7m82lz4AhXyCLkGHQr9B9AQ4dUDCA4&uact=5&oq=dashboard+define&gs_lcp=Cgdnd3Mtd2l6EA)

Moncosí Linares, L. (2020-2021). *ANÁLISIS, MODELIZACIÓN Y PREDICCIÓN DE PARTIDOS*. Trabajo Final de grado, Universidad Politécnica de Cataluña, Barcelona. Recuperado el 22 de Octubre de 2021, de [https://upcommons.upc.edu/bitstream/handle/2117/343836/Análisis%20%20modelización%20y%20predicción%20de%20partidos%20de%20baloncesto%20%28tfg\\_eebe%29%20Luís%20Moncosí.pdf?sequence=1&isAllowed=y](https://upcommons.upc.edu/bitstream/handle/2117/343836/Análisis%20%20modelización%20y%20predicción%20de%20partidos%20de%20baloncesto%20%28tfg_eebe%29%20Luís%20Moncosí.pdf?sequence=1&isAllowed=y)

- Neteris. (02 de Julio de 2019). Recuperado el 29 de Mayo de 2022, de Analytics: <https://neteris.com/software/tableau-software-visualizacion-datos/>
- Neva, S. (2018-2019). *DISEÑO E IMPLEMENTACIÓN DE UN CUADRO DE MANDO PARA EL SUBÁREA DE DEPORTES DE UN AYUNTAMIENTO*. Trabajo Fin de grado, Universidad del País Vasco. Recuperado el 21 de Octubre de 2021, de [https://addi.ehu.es/bitstream/handle/10810/42111/TFG\\_NEVA.pdf?sequence=1](https://addi.ehu.es/bitstream/handle/10810/42111/TFG_NEVA.pdf?sequence=1)
- Ortiz, D. (20 de Octubre de 2021). *Cyberclick*. Recuperado el 28 de Mayo de 2022, de <https://www.cyberclick.es/numerical-blog/que-es-un-dashboard>
- REVISTA DIGITAL INESEM. (08 de Febrero de 2017). *Gestion Empresarial*. (E. Chicano Tejada, Editor) Recuperado el Lunes de Junio de 2021, de Small data vs. big data. ¿Cuál es la mejor alternativa? ¿En qué se diferencian?: <https://revistadigital.inesem.es/gestion-empresarial/small-data/>
- Sánchez, A. (Junio de 2021). *Harvard Deusto*. Recuperado el 28 de Mayo de 2022, de <https://www.harvard-deusto.com/la-analitica-de-datos-en-rrhh#:~:text=La%20analítica%20de%20datos%20es,tomar%20las%20decisiones%20más%20acertadas.>
- Toquero, M. A. (2020). *Implementación de una herramienta de análisis en la gestión deportiva mediante Power BI*. Tesis, Universidad de Valladolid, Departamento de organización de empresas y comercialización e investigación de mercados, Valladolid. Recuperado el 28 de Mayo de 2022, de <https://uvadoc.uva.es/handle/10324/45062>
- UNIVERSIDAD CATOLICA SAN PABLO. (16 de Noviembre de 2020). *Big data: definición, tipos, características y beneficios*. Recuperado el 28 de Junio de 2021, de <https://postgrado.ucsp.edu.pe/articulos/que-es-big-data/>
- Vicente Stenhouse, N. (2018). *HabScraper: herramienta automatizada para la Memoria*, Universidad de las Islas Baleares. Recuperado el 15 de Julio de 2021, de [https://dspace.uib.es/xmlui/bitstream/handle/11201/151095/Memoria\\_EP\\_SU1195.pdf?sequence=1&isAllowed=y](https://dspace.uib.es/xmlui/bitstream/handle/11201/151095/Memoria_EP_SU1195.pdf?sequence=1&isAllowed=y)

YALE RIOS, J. A. (2021). *EFFECTOS DEL POSICIONAMIENTO INTERNACIONAL EN*. Tesis, Lima. Recuperado el 21 de Octubre de 2021, de [http://repositorio.usil.edu.pe/bitstream/USIL/11308/1/2021\\_Yale%20Rios.pdf](http://repositorio.usil.edu.pe/bitstream/USIL/11308/1/2021_Yale%20Rios.pdf)

# ANEXOS

Preguntas Respuestas **10** Configuración

Puntos totales:

Sección 1 de 3

## Cuestionario aplicado para la evaluación del Dashboard ✕ ⋮

Evaluación de Dashboard sobre Fútbol Peruano 2014 - 2021

Correo \*

Correo válido

Este formulario registra los correos. [Cambiar configuración](#)

### CUESTIONARIO

Descripción (opcional)

**El siguiente cuestionario trata sobre el funcionamiento y valoración de la información del Dashboard.**

**La evaluación se dará por una escala del 1 al 5, siendo 5 Muy bueno y 1 Muy malo.**

**También cuenta con respuestas de opciones, con la Opción Si o No**

VALORACIÓN	ESCALA
Muy bueno	5
Bueno	4
Regular	3
Malo	2
Muy malo	1

 Si  
 No

### FACILIDAD CON QUE SE USA EL DASHBOARD

Descripción (opcional)

1. Para usted es fácil navegar dentro del Dashboard.

- Muy Bueno
- Bueno
- Regular
- Malo
- Muy malo

2. La apariencia del Dashboard y fácil de entender.


- Muy bueno
- Bueno
- Regular
- Malo
- Muy malo

3. Cuando se solicita información al Dashboard, este despliega la información en el tiempo esperado.

- Muy bueno
- Bueno
- Regular
- Malo
- Muy malo

4. La búsqueda de información es sencilla y rápida.


- Muy bueno
- Bueno
- Regular
- Malo
- Muy malo

5. Como calificarías la interacción con el Dashboard   Varias opciones

Muy bueno



Bueno

Regular

Malo 

Muy malo

Añadir opción o [añadir respuesta "Otro"](#)

Clave de respuestas (0 puntos)   Obligatorio

---


**CANTIDAD DE REGISTROS**  
Descripción (opcional)

---

1. ¿Consideras que la cantidad de años para el análisis ofrecida en el Dashboard, es la apropiada

Si

No

2. ¿Esta bien los años del 2014 al 2021? \* 

Si

No

---

3. Consideras que la información histórica es una información relevante

Si

No

---

4. Si fuera el tiempo de 6 meses consideras una data histórica mínima? \*

Si

No

---

5. ¿Conoces alguna fuente de registros deportivos?

Si

No



Sección 2 de 3

TIEMPO PARA ANALIZAR LA INFORMACIÓN



Descripción (opcional)

6. ¿Crees que se puede analizar de forma rápida, la información brindada en el Dashboard ?

- Si
- No

7. Consideras el tiempo de media hora analizar datos deportivos en un dashboard \*

- Si
- No

Sección 3 de 3

UTILIDAD DEL DASHBOARD



Descripción (opcional)

8. Consideras la información útil \*

- Si
- No

9. Te apporto algo útil, la información prestada en el dashboard \*

- Si
- No

10. Para alguien que conozca, le seria útil la información? \*

- Si
- No

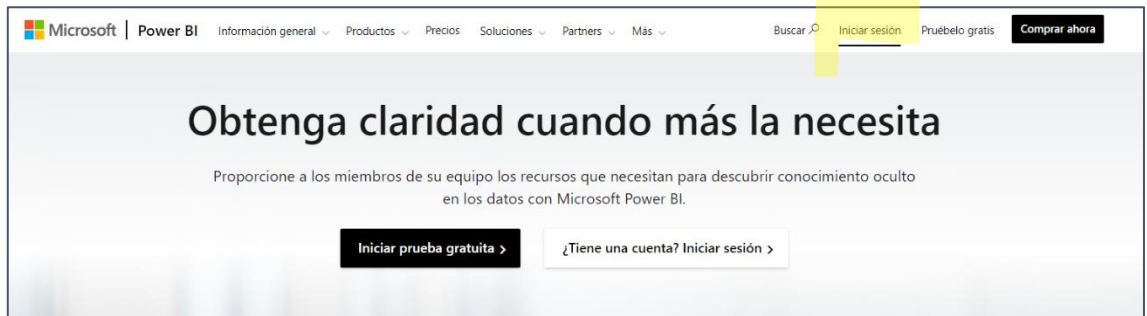
## INSTRUCTIVO PARA ENTRAR AL DASHBOARD GENERAL

Para entrar al Dashboard, se seguirá los siguientes pasos:

1. Copie el siguiente link a través del navegador de google chrome:

<https://powerbi.microsoft.com/es-es/>

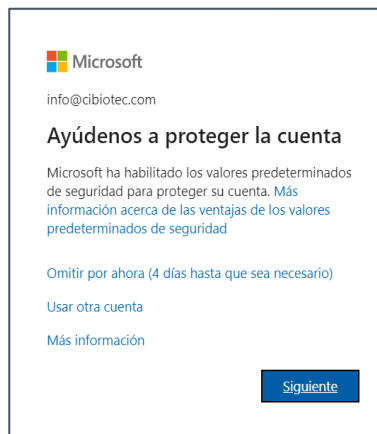
2. D



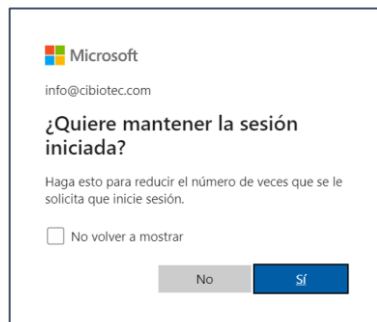
se a Iniciar Sesión, como se muestra en la siguiente figura

### ACCESO CON USUARIO INFO

3. Al ingresar se abrirá otra ventana, donde se ingresará el siguiente usuario [info@cibiotec.com](mailto:info@cibiotec.com) y la siguiente contraseña XXXXXXXXXX



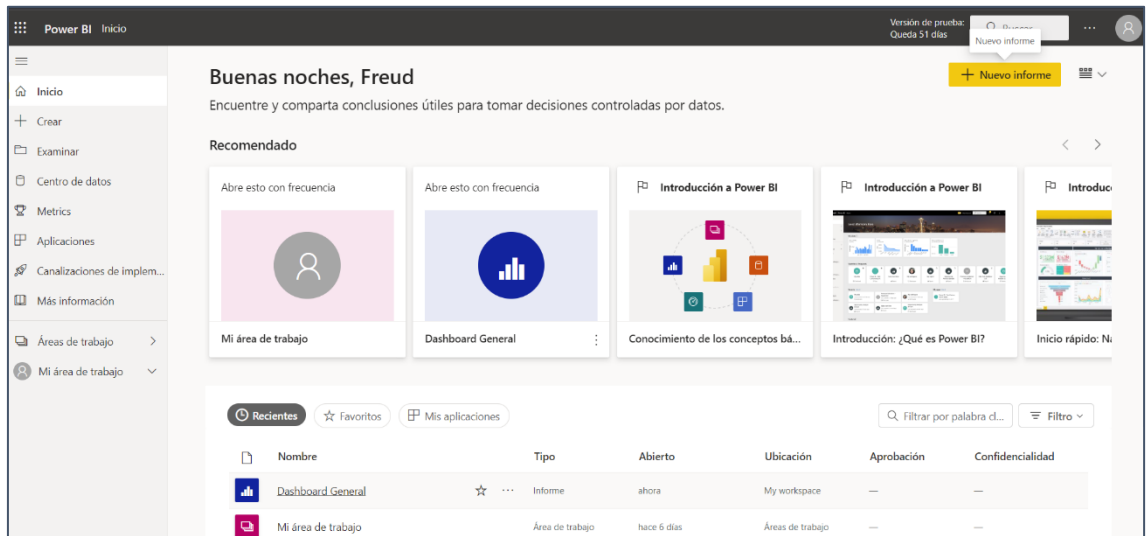
a. Seguidamente le aparecerá el siguiente cuadro, y le da en omitir por ahora:



b. Se selecciona la opción

NO

c. Se cargará la plataforma de Power BI y se dirige a Dashboard General, como se muestra en la siguiente imagen:



d. Finalmente cargue la **pantalla completa** como se muestra en la siguiente figura:

Desglosa para pantalla



e. Finalmente puede interactuar con el dashboard, para salir de la vista presione el botón Esc y cerrar sesión.