



UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS

FACULTAD DE NEGOCIOS

PROGRAMA ACADÉMICO DE ADMINISTRACIÓN DE EMPRESAS

**PROGRAMA ACADÉMICO DE ADMINISTRACIÓN DE BANCA Y
FINANZAS**

PROGRAMA ACADÉMICO DE NEGOCIOS INTERNACIONALES

Aplicación de Data Science en la empresa Partners Technology S. A. C.

TRABAJO DE INVESTIGACIÓN

Para optar el grado de bachiller en Administración de Empresas

Para optar el grado de bachiller en Administración de Banca y Finanzas

Para optar el grado de bachiller en Negocios Internacionales

AUTORES

Angeles Rubiños, Gianpierre Alberto (0000-0003-4497-4264)

Aspilcueta Mantari, Carlos Nicanor (0000-0003-2247-1121)

Jara Cayturo, Nélica Eliza (0000-0003-2065-1874)

Turpo Chayña, Luis Antonio (0000-0002-7536-8454)

ASESOR

Palacios Ruiz, Julio César (0000-0003-1610-5205)

Lima, 13 de diciembre de 2020

Dedicatoria

El presente trabajo de investigación está dedicado a nuestras familias y a todas las personas que nos ayudaron a crecer profesionalmente.

Agradecimientos

Queremos agradecer de manera especial a todo el equipo de la empresa Partners Technology S. A. C. por brindarnos todas las facilidades para el desarrollo del presente trabajo de investigación.

Resumen

El objeto de estudio del presente trabajo es la empresa Partners Technology S. A. C., donde se realizó un exhaustivo análisis interno y externo con el propósito de identificar el problema principal de la organización. Así pues, se determinó que su mayor dificultad es el inadecuado conocimiento de sus clientes. Por esta razón, la investigación se centra en reconocer las variables que influyen en el comportamiento de compra de los usuarios. Para lograr dicho objetivo se empleó la metodología de ciencia de datos mediante un análisis híbrido, es decir, exploratorio y explicativo. Los resultados obtenidos permitirán responder a las preguntas de *Data Science* formuladas. Es importante mencionar que para la creación de la base de datos utilizada se recurrió a fuentes internas y externas a la compañía, de las cuales se lograron dilucidar las variables que muestran las características de sus clientes. También cabe destacar que se realizó un análisis estadístico de todas las variables cuantitativas con el fin de encontrar correlaciones. De esta forma se hallaron importantes descubrimientos que aportarán valor a la compañía y permitirán el logro de sus objetivos, dado que en el presente trabajo se proponen soluciones al problema señalado y se sugieren las acciones que se deberán adoptar a corto plazo.

Palabras clave: objetivos, metodología, *Data Science*, Partners Technology.

Abstract

The object of study of this work is the company Partners Technology S.A.C. Therefore, an exhaustive internal and external analysis of the company was carried out in order to identify their main problem. In this sense, it was determined that inadequate knowledge of customers is the greatest difficulty that the organization goes through. For this reason, the objective of this research is to identify the variables that influence the purchasing behavior of Partners Technology. To achieve the proposed objective, the Data Science methodology was developed through a hybrid analysis, that is, exploratory and explanatory. Likewise, the results obtained will allow answering the Data Science questions asked. It is important to mention that, for the creation of the database used in this work, internal and external sources were used from the company Partners Technology S.A.C. Likewise, the selected variables provide characteristics of their clients. It should be noted that a statistical analysis of all quantitative variables was carried out in order to find correlations. In this sense, by Data Science techniques, results were verified are the main ones. Likewise, important discoveries were found that will add value to the company and enable it to achieve its objectives. This paper proposes solutions to the research problem and suggests actions to be taken in the short term.

Keywords: objectives; methodology; Data Science; Partners Technology

Índice

1	CAPÍTULO I. PRESENTACIÓN DE LA EMPRESA	1
1.1	Presentación de la Empresa para el Proyecto	1
1.2	Modelo de Negocio.....	2
1.3	Objetivos de la Empresa	4
1.4	Análisis PESTEL	5
1.5	Análisis de Porter	7
2	CAPÍTULO II. PROYECTO DE CIENCIA DE DATOS	9
2.1	Identificación del Problema	9
2.2	Formulación de Preguntas para <i>Data Science</i>	9
2.3	Justificación	9
2.4	Objetivos del Proyecto	10
2.5	Hipótesis	10
3	CAPÍTULO III. MARCO TEÓRICO	11
3.1	Metodología de Ciencia de Datos	11
4	CAPÍTULO IV. ARQUITECTURA DE DATOS.....	13
4.1	Arquitectura de Datos de la Empresa Partners Technology S.A.C.....	13
4.2	Arquitectura Tecnológica.....	13
4.3	Insumos	14
4.4	Procesamiento	15
4.5	Almacenamiento	15
4.6	Destino	16

4.7	Tecnología, Herramientas e Infraestructura.....	16
5	CAPÍTULO V. REQUISITOS Y RECOPIACIÓN DE DATOS.....	177
5.1	Determinación de Datos Necesarios para Responder las Preguntas.....	177
5.2	Ubicación de Fuentes de Datos.....	177
5.3	Uso de Herramientas para Determinar Posibles Relaciones entre los Datos....	18
5.4	Revisión y Depuración de Datos que se Usaran en el Estudio	19
5.5	Calidad de Datos	20
5.6	Gobierno de Datos	22
5.7	Diccionario de Datos.....	25
5.8	Visualización de Datos	25
6	CAPÍTULO VI. PRODUCCIÓN Y ANÁLISIS EXPLORATORIO DE DATOS	
	27
6.1	Plan de Muestreo y Tamaño de la Muestra.....	27
6.2	Determinación de la Población Objetivo (Elemento, Unidad, Alcance y Tiempo).....	27
6.3	Diseño del Estudio de Muestreo	28
6.4	Ejecución del Muestreo.....	29
6.5	Resumen y Examen de la Distribución de Variables.....	30
6.6	Distribución Estadística de las Variables.....	33
6.7	Resultado del Análisis de Relaciones	35
6.8	Análisis de Regresión Lineal Múltiple	36
6.9	Estudio Probabilístico de los Datos	38

7	CAPÍTULO VII. TÉCNICAS DE CIENCIA DE DATOS	40
7.1	Técnica de Ciencia de Datos Supervisada	40
7.2	Técnica de Ciencia de Datos no Supervisada	40
8	CAPÍTULO VIII. EL MODELO Y LA HISTORIA.....	46
8.1	Determinación del Modelo más Apropriado para la Descripción o Predicción.	46
8.2	Descripción de los Hallazgos.....	46
8.3	Definición y Justificación de los Criterios para la Visualización de los resultados.....	46
8.4	Estrategias de Implementación	47
9	CAPÍTULO IX. CALIBRADO E IMPLEMENTACIÓN	49
9.1	Métodos de Ajuste	49
9.2	Retroalimentación.....	56
10	CONCLUSIONES.....	57
11	FACTORES CRÍTICOS DEL ÉXITO DEL PROYECTO	58
12	DECISIONES POR TOMAR A PARTIR DE LOS DATOS ESTUDIADOS ..	59
13	REFERENCIAS BIBLIOGRÁFICAS.....	60
14	ANEXOS.....	61

Índice de Tablas

Tabla 1 Ubicación de fuentes de datos.....	18
Tabla 2 Clasificación de variables según tipo y escala	28
Tabla 3 Estadísticas de la regresión	36
Tabla 4 Prueba F para varianzas de dos muestras.....	39
Tabla 5 Distribución de clientes por grupos.	42
Tabla 6 Estadísticas de la regresión.	49
Tabla 7 Estadísticas de la regresión	51
Tabla 8 Estadísticas de la regresión	53

Índice de Figuras

Figura 1	Arquitectura de datos del Partners Technology S.A.C.....	13
Figura 2	Arquitectura de datos tecnológica	14
Figura 3	Insumos del proyecto.....	14
Figura 4	Almacenamiento del proyecto	16
Figura 5	Metodología para la calidad de datos	22
Figura 6	Gobierno de datos.....	24
Figura 7	Los diez sectores principales de la empresa Partners Technology en el 2019 ..	26
Figura 8	Histograma con curva normal de cantidad de años en el mercado.....	30
Figura 9	Histograma con curva normal de cantidad de trabajadores.....	31
Figura 10	Histograma con curva normal de cantidad de monto invertido en proyectos de implementación tecnológica	32
Figura 11	Diagrama de caja y bigotes.....	33
Figura 12	Cuadro de medidas de resumen de las variables	34
Figura 13	Mapa de correlación de variables	35
Figura 14	Gráfica de la correlación lineal múltiple entre las variables cantidad de años en el mercado y monto invertido en proyectos de implementación tecnológica.	37
Figura 15	Gráfica de la correlación lineal múltiple entre las variables cantidad de trabajadores y monto invertido en proyectos de implementación tecnológica.	38
Figura 16	El método Elbow con K-means++	41
Figura 17	Tabulación de clústeres.	42
Figura 18	Gráfico de cantidad de empresas por clúster.	43
Figura 19	Gráfico de curva de comportamiento de los clústeres.	43

Figura 20 Grupo 0.....	44
Figura 21 Grupo 1	44
Figura 22 Grupo 2.....	45
Figura 23 Gráfico comparativo entre modelo de regresión planteado y el monto invertido en proyectos de implementación tecnológica.	48
Figura 24 Gráfico de relación cantidad de años vs. monto invertido de proyectos de implementación tecnológica.	50
Figura 25 Gráfico de relación cantidad de trabajadores vs. monto invertido en proyectos de implementación tecnológica.....	50
Figura 26 Gráfico de relación cantidad de años vs. monto invertido en proyectos tecnológicos.....	52
Figura 27 Gráfico de relación cantidad de cantidad de trabajadores vs. monto invertido en proyectos.	52
Figura 28 Gráfico de relación cantidad de años vs. monto invertido en proyectos tecnológicos.....	54
Figura 29 Gráfico de relación cantidad de trabajadores vs. monto invertido en proyectos de implementación tecnológica.....	54
Figura 30 Gráfico de testeó del modelo.	56

Capítulo I. Presentación de la Empresa

1.1 Presentación de la Empresa para el Proyecto

Partners Technology S. A. C. es una empresa peruana identificada con el RUC 20556739693 fundada el 1 de febrero de 2015. Desde el inicio de sus operaciones se enfocó en brindar soluciones tecnológicas a compañías del sector corporativo, educativo y gubernamental. La organización se ubica en la calle Las Golondrinas 393, oficina 902 en San Isidro. Su estructura está dividida en tres áreas: comercial, administrativa y tecnología de la información. El portafolio de productos y servicios que ofrece Partners Technology S. A. C. está enfocado en la calidad total a través de la integración de actividades innovadoras y de transformación digital. Las soluciones que brinda son de equipamiento de cómputo, impresión, digitalización, infraestructura tecnológica, multimedia, puntos de venta, educación, servicios y financieras. Estas tienen el objetivo de potenciar el negocio de sus clientes.

Sus principales clientes son el Ministerio de la Mujer, la Marina de Guerra del Perú, el Ministerio Público, el Ministerio de Salud, entre otras. Cabe destacar que también atiende a importantes universidades y colegios privados, como la Universidad Norbert Wiener, el Grupo Lauréate, el SENATI o la Universidad Científica del Sur. Además, en menor porcentaje, presta servicios a empresas privadas, como Laive, Hermes, la Financiera Compartamos, la Clínica Internacional, y otras prestigiosas organizaciones. Desde el inicio de sus operaciones Partners Technology S. A. C. ha demostrado cumplir con sus obligaciones hacia sus colaboradores y clientes apoyada la práctica de valores como la responsabilidad, la competitividad y el trabajo en equipo. Su misión es acercar la tecnología e innovación para acelerar los procesos a sus clientes.

Asimismo, su visión es ser un referente de transformación digital y líder en el rubro tecnológico mediante soluciones globales que generen valor a sus socios de negocio.

1.2 Modelo de Negocio

En 2008 Day y Gastel sostuvieron que la metodología Canvas “fue creada con el fin de establecer una relación lógica entre cada uno de los componentes de la organización y todos los factores que influyen para que tenga o no éxito, y que se plasma en un lienzo” (p. 1). Estos componentes son los socios clave, las actividades clave, la propuesta de valor, la relación con los clientes, los canales de distribución, los segmentos de clientes, la estructura de costes y las fuentes de ingresos, que para el caso particular de Partners Technology S. A. C. se describen a continuación:

Socios Clave

Los socios claves de la organización son los canales mayoristas y las marcas de tecnología. En primer lugar, la compañía trabaja con canales mayoristas como Deltron, Intcomex, Ingram Micro, Compudisket, Igarachi, PC Link, Nexys, Nexus, entre otros, que se encargan de importar equipamiento tecnológico. También, la empresa cuenta con alianzas con las principales marcas de tecnología como HP, Lenovo, Dell, ViewSonic, Cisco y otras.

Actividades Clave

La actividad clave de la empresa es la venta de productos y servicios tecnológicos. Por ello, su propuesta de valor busca aumentar la productividad de sus clientes, priorizando el uso de tecnologías innovadoras que mejoren sus procesos. En este contexto, la capacitación constante del personal del área comercial es fundamental para lograr el éxito de sus operaciones. Igualmente, las acciones realizadas contribuyen al logro de objetivos de la organización.

Propuesta de Valor

La propuesta de valor de Partners Technology es garantizar el acompañamiento a sus clientes desde la detección de necesidades hasta la atención posventa. Asimismo, su enfoque busca resolver problemas cotidianos como la optimización de la eficiencia del trabajo. Para lograrlo la compañía tiene como ventaja competitiva la diferenciación de precios que obtiene por parte de las principales marcas de tecnología, lo cual se complementa con la atención personalizada que ofrece.

Relación con los Clientes

La relación con los clientes es el pilar de la empresa puesto que busca fidelizarlos mediante la atención personalizada. Las acciones comerciales se basan en negociaciones efectivas y de colaboración mutua que ponen en práctica valores como la responsabilidad, la confianza, la honestidad y el trabajo en equipo. Su objetivo es fortalecer los vínculos comerciales y lograr relaciones a largo plazo que sean beneficiosas para ambas partes.

Canales de Distribución

El canal de distribución de la empresa es su único almacén ubicado en el distrito de San Isidro. La mayoría de entregas se realizan en la dirección que indica el cliente, por lo cual la compañía dispone de movilidades propias y el apoyo de los canales mayoristas para el despacho de mercadería. Es importante resaltar que se tiene presupuestado habilitar un canal *e-commerce* para la compra mediante la página web y las redes sociales.

Segmentos de Clientes

Los segmentos de clientes de Partners Technology son el sector público, el privado y el de la educación. Estos se encuentran divididos en rubros como el agrícola, transportes, finanzas, comunicaciones, etc. Cabe señalar que a raíz de la pandemia por el COVID-19 agregaron un

cuarto grupo denominado usuarios finales, debido a la demanda de equipamiento tecnológico por parte de personas naturales.

Estructura de Costes

El costo más importante e inherente al modelo de negocio de la compañía es la compra de productos tecnológicos. Otros gastos relacionados son el pago de alquiler de la oficina, la luz, el agua, el internet, la seguridad, las planillas, el transporte, el mantenimiento, los gastos de representación, las comisiones de venta, entre otros. El costo más elevado es el pago de planillas a sus trabajadores. Estos servicios pueden ser fijos o variables.

Fuentes de Ingresos

La principal fuente de ingresos de la empresa es la venta de equipamientos y servicios tecnológicos dirigidos al segmento corporativo, educativo y gubernamental. Las soluciones de oficina, *outsourcing* de impresión y cableado estructurado son las que generan mayores ganancias. Estas actividades se llevan a cabo mediante licitaciones, subastas electrónicas, concursos públicos y privados, adjudicaciones simplificadas, contrataciones directas, y comparaciones de precios.

1.3 Objetivos de la Empresa

Objetivo General

- Dar a conocer nuestra marca a nivel nacional como principal proveedor de equipos y servicios tecnológicos a finales del año 2020.

Objetivos Específicos

- Desplegar campañas de fidelización a los clientes *target*.

- Incrementar en 45 % el porcentaje de negocios que ofrecen servicios.
- Incrementar en 40 % la cartera de clientes del sector privado.

1.4 **Análisis PESTEL**

En palabras de la ONU Mujeres (2012) “el análisis PESTEL es un instrumento de planificación estratégica para definir el contexto de una campaña. Analiza factores externos políticos, económicos, sociales, tecnológicos, ambientales y jurídicos” (párr. 1). Por lo tanto, realizar este análisis permitirá conocer la situación de la empresa respecto a diversos factores internos y externos.

Análisis Político

La inestabilidad política originada por los tres cambios de gobierno y las fuertes fricciones entre el poder ejecutivo y legislativo es el principal problema que atraviesa el país en la actualidad. El antecedente a este escenario es el resquebrajamiento de los partidos políticos. En la práctica el sistema de Gobierno peruano es semiparlamentario dado que solo se necesita el 66 % de votos congresales para vacar al presidente por incapacidad moral. Esto genera incertidumbre en la población.

Análisis Económico

La pandemia COVID-19 provocó el cierre de muchas empresas y dejó sin trabajo a casi la mitad de la población urbana del país. Perú ha sido uno de los países más afectados por el coronavirus debido a la reducción del PBI en 30.2 % con respecto al año 2019. Esta caída pone en recesión la economía y paraliza muchos sectores productivos, como el turismo, los servicios y el comercio. En consecuencia, el número de personas con trabajo activo se contrajo en un 40 % y la pobreza se incrementó en un 27.5 %.

Análisis Sociocultural

El aislamiento social obligatorio ha modificado los hábitos de consumo y el día a día de millones de peruanos. En este escenario el *e-commerce* se viene desarrollando como una alternativa que garantiza el desarrollo de protocolos de bioseguridad. Una de las principales causas de este incremento es el aumento del uso de internet durante la cuarentena. Sin embargo, en el Perú solo el 40.1 % de los hogares cuentan con este servicio, en su mayoría por medio de celulares y computadoras de escritorio.

Análisis Tecnológico

El trabajo remoto incrementó la demanda de *laptops*, computadoras y otros artículos tecnológicos, cuya escasez tuvo como consecuencia un aumento de los precios. Muchas empresas han optado por invertir en seguridad, debido al incremento de ciberataques. Sumado a esto, uno de los principales hechos que se registró en Perú durante el 2020 fue la aprobación del uso de la señal 5G, el cual beneficiará a empresas del rubro comunicaciones y servicios.

Análisis Ecológico

Perú firmó el 10 de octubre de 2013 el Convenio de Minamata. Este acuerdo expresa el compromiso del país por minimizar los riesgos de la salud humana a causa de las emisiones de mercurio y sus compuestos. Por esto, muchos fabricantes de tecnología se han visto obligados a retirar progresivamente sus productos del mercado y a diseñar otros que no contengan este elemento. También, han tenido que reducir las importaciones de equipamiento, por ejemplo, proyectores y lámparas de vapor.

Análisis Legal

El decreto supremo que declaró el estado de emergencia por las graves circunstancias que afectan la vida de la nación a consecuencia del brote del COVID-19 redujo las operaciones de las

empresas. Asimismo, limitó el ejercicio del derecho a la libertad de tránsito. Posteriormente, se promulgó el Decreto N.º 110-2020-PCM que hace referencia a las cuatro fases de reactivación económica y la reanudación progresiva de actividades. Esta norma se desarrolló conforme a las recomendaciones del Ministerio de Salud (MINSA).

1.5 Análisis de Porter

El análisis de las fuerzas competitivas en los negocios es parte fundamental del ciclo de vida de un negocio.

La competitividad y la rentabilidad determinan la estructura de un sector; ésta no depende de si el sector ofrece productos o servicios, si es emergente o maduro, de alta tecnología o baja tecnología, si está regulado o no lo está. Aun cuando una multitud de factores..... pueden afectar la rentabilidad de un sector en el corto plazo, es la estructura del sector, la cual se manifiesta en las fuerzas competitivas, la que determina la rentabilidad del sector en el mediano y largo plazo. (Porter, 2008, p. 1)

Rivalidad de la Industria

La rivalidad entre competidores existentes es alta debido a la gran cantidad de empresas que se dedican al rubro de tecnologías de la información. Según el portal Perú Compras 1782 empresas de este sector lograron adjudicar órdenes con el Estado, de las cuales el 81.2 % eran pymes. Igualmente, INEI señaló que hasta el primer trimestre del 2020 existían 57 701 empresas dedicadas a brindar productos y servicios tecnológicos. La mayoría de estas compañías dispone de una vertical de negocio definida.

Amenaza de Ingreso de Nuevos Competidores

La amenaza de nuevos competidores es alta. Según datos brindados por INEI, en promedio, 140 pymes inician actividades cada mes en la industria prestación de productos y servicios

tecnológicos a empresas. Pese a la coyuntura del COVID-19, el sector demostró contar con una demanda estable que despertó el interés de muchos emprendedores. Cabe señalar que en el Perú es relativamente fácil crear una empresa, pues este proceso solo demora entre 10 y 26 días.

Poder de Negociación de Compradores

El poder de negociación de los compradores es medio dado que, como se dijo, muchas compañías pertenecen a un mismo conglomerado. Por un lado, el sector público centraliza sus compras por la plataforma Perú Compras. Este programa escoge a la organización que oferta el menor precio para el equipo solicitado. La situación es igual con respecto a licitaciones y concursos públicos. Por otro lado, en las empresas privadas el poder de negociación es alto y depende en la mayoría de casos de la calidad y el precio.

Poder de Negociación de Proveedores

El poder de negociación con los proveedores es medio en vista de que las empresas tienen que cumplir ciertos requisitos para comprar a un mayorista. Las regulaciones son iguales para acceder a líneas de crédito. En algunos casos la cadena de intermediación puede desaparecer cuando el suministrador tiene la experiencia y la facturación que necesita el cliente. Los fabricantes de tecnología regulan los precios de los productos. Por esta razón, la mayoría de abastecedores manejan costos similares.

Amenazas de Productos Sustitutos

La amenaza de productos o servicios sustitutos es alta, dado que los clientes requieren innovación constante. La competencia entre los fabricantes de tecnología propone transformación y desarrollo de mejoras que impulsen la productividad de las empresas. Por ese motivo, la principal ventaja del rubro es la renovación periódica que incrementa la demanda.

Capítulo II. Proyecto de Ciencia de Datos

2.1 Identificación del Problema

Durante el trabajo de investigación se identificó que el principal problema de la empresa Partners Technology S. A. C. es el inadecuado conocimiento de sus clientes. Las consecuencias de la cuestión mencionada son la reducción de la participación de mercado, el incumplimiento de los objetivos trazados y, por ende, diversas pérdidas económicas.

2.2 Formulación de Preguntas para *Data Science*

A continuación, se exponen las preguntas de *Data Science* que se formularon para el presente proyecto:

Principal

1. ¿Cuáles son las principales variables que influyen en el comportamiento de los clientes de la empresa Partners Technology?

Secundarias

2. ¿Cómo predecir mediante la ciencia de datos cuáles serán los clientes con mayor inversión en proyectos de implementación tecnológica?

3. ¿Cómo segmentar a los clientes de la empresa Partners Technology mediante el uso de ciencia de datos?

2.3 Justificación

Las preguntas de *Data Science* se centran en incrementar la productividad de las pequeñas y medianas empresas mediante el uso de la ciencia de datos. En el Perú alrededor del 50 % de las pymes tienen el riesgo de desaparecer del mercado antes de los tres años de iniciadas sus

actividades. Por ello, en el contexto de la pandemia COVID-19 se considera importante proponer una solución a esta problemática.

2.4 **Objetivos del Proyecto**

Objetivo General

- Identificar los diez clientes *target* de la empresa Partners Technology S. A. C. mediante el uso de técnicas de ciencia de datos.

Objetivos Específicos

- Identificar las tres variables principales que influyen en el comportamiento de los clientes de la empresa Partners Technology S. A. C. mediante el uso de la metodología de ciencia de datos de IBM.
- Segmentar en grupos a los clientes de la empresa Partners Technology S. A. C. según las principales variables que determinan su comportamiento mediante el uso de ciencia de datos.
- Crear un modelo de regresión múltiple para conocer el comportamiento de los clientes de la empresa Partners Technology S. A. C.

2.5 **Hipótesis**

Para el presente trabajo se propone validar o rechazar la siguiente hipótesis:

El análisis de regresión múltiple predecirá el comportamiento de los clientes de la empresa Partners Technology.

Capítulo III. Marco Teórico

3.1 Metodología de Ciencia de Datos

La metodología de ciencia de datos brinda al científico de datos un marco sobre cómo proceder con métodos, procesos y argumentos con el objetivo de obtener resultados. En el presente trabajo de investigación se consideró el modelo propuesto por Rollins J. (2015) que consta de diez etapas y forma parte de un proceso iterativo para descubrir *insights*.

Comprensión del Negocio

En esta etapa se define el problema, los objetivos del proyecto y los requisitos para su solución desde un punto de vista empresarial.

Enfoque Analítico

Se expresa el problema bajo el contexto de técnicas estadísticas y aprendizaje automático para que la empresa consiga el resultado deseado.

Requisitos de Datos

Una vez elegido el enfoque analítico se determinan los requisitos de datos. Estos requieren unos contenidos, formatos y representaciones específicos.

Recopilación de Datos

Los científicos identifican y recogen los datos disponibles y relevantes para el dominio del problema.

Comprensión de Datos

Esta etapa generalmente utiliza estadísticas descriptivas y técnicas de visualización a fin de comprender el contenido de los datos, evaluar su calidad y descubrir *insights*.

Preparación de Datos

Aquí se incluyen todas las actividades para construir un conjunto de datos, como la limpieza, la combinación y la transformación de estos.

Modelado de Datos

En esta etapa se utiliza la primera versión del conjunto de datos y se desarrollan modelos predictivos o descriptivos, según el enfoque analítico definido.

Evaluación del Modelo

La evaluación permite comprender la calidad del modelo y garantizar que se aborde de manera adecuada el problema empresarial.

Implementación del Modelo

Cuando el modelo ha sido desarrollado y aprobado satisfactoriamente por los promotores del negocio se lleva a cabo su implementación.

Retroalimentación

En la fase final se obtiene información sobre el rendimiento del modelo y su impacto en el entorno donde se implementó. Esto permite a los científicos de datos analizar y ajustar el modelo hasta mejorar su precisión y utilidad.

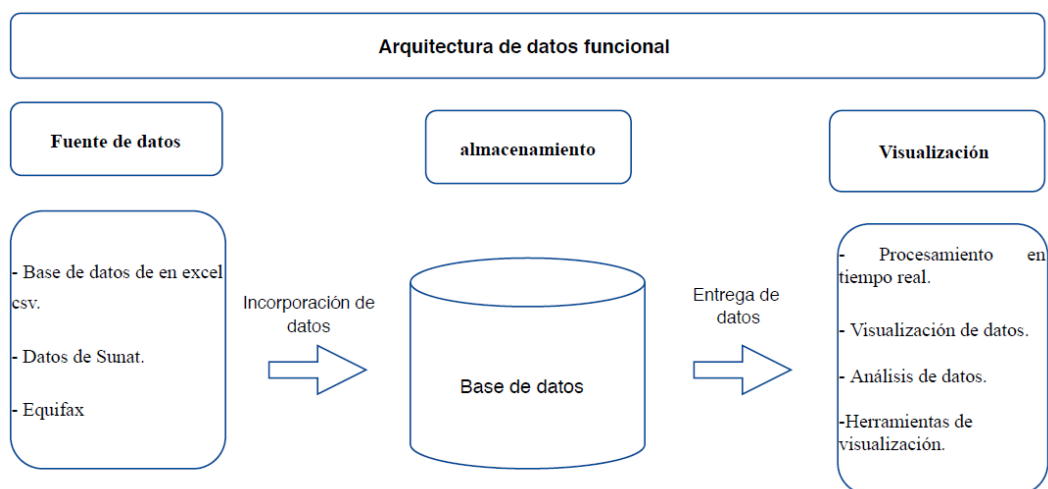
Capítulo IV. Arquitectura de Datos

4.1 Arquitectura de Datos de la Empresa Partners Technology S.A.C

En esta parte del proyecto se analiza la composición de los datos de la base de datos, las reglas que los rigen y cómo se recopilan, almacenan, organizan, integran y utilizan en los diferentes sistemas que tiene la compañía. Cabe mencionar que la empresa no cuenta con una arquitectura de datos 100 % diseñada y orientada a los objetivos comerciales; por ejemplo, no dispone de un arquitecto de datos, quien facilitaría a cada miembro de la compañía una base de datos propia.

Figura 1

Arquitectura de datos del Partners Technology S.A.C.



Nota. Elaboración propia

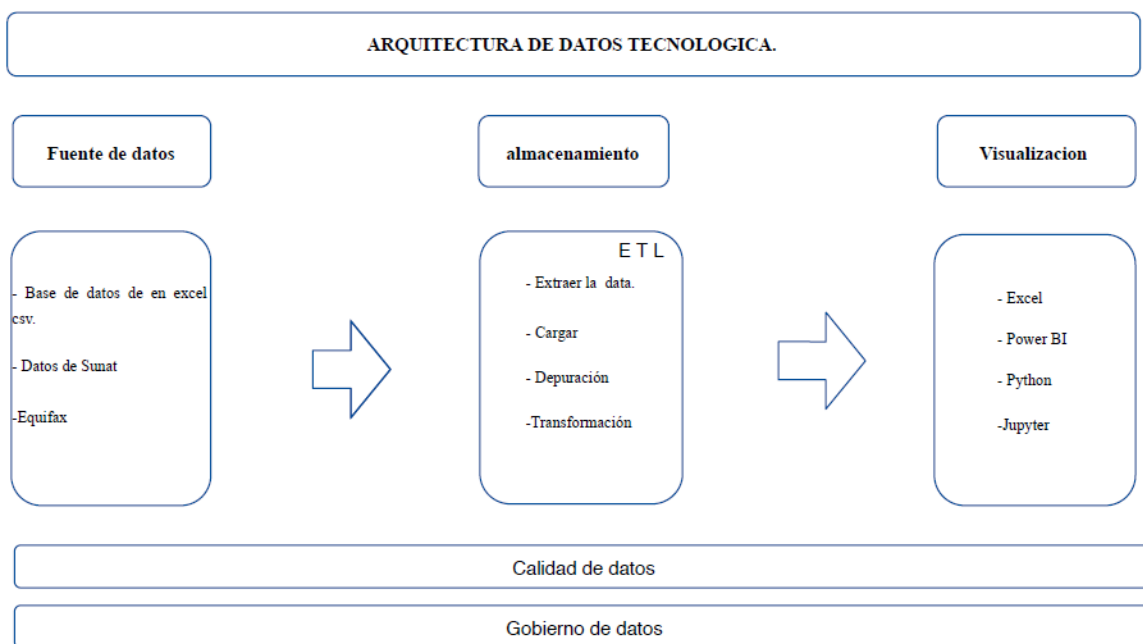
4.2 Arquitectura Tecnológica

Esta fase se centra en revisar la composición de la solución tecnológica, es decir, los datos que componen la base de datos, cuáles son las reglas que la rigen, qué datos se recopilan y cómo

estos se almacenan, organizan, integran y utilizan en los diferentes sistemas de los que dispone la compañía.

Figura 2

Arquitectura de datos tecnológica



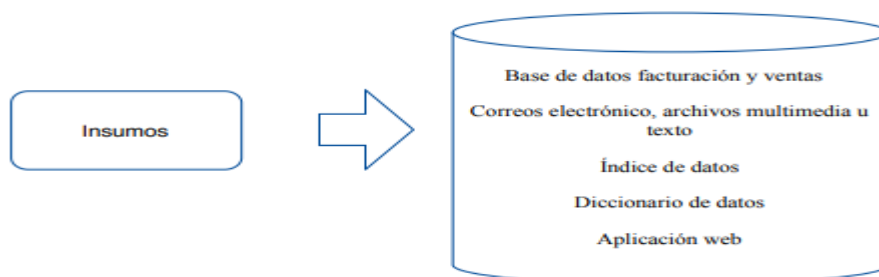
Nota. Elaboración propia

4.3 Insumos

La Figura 3 hace referencia a los insumos que se utilizaron en el presente trabajo de investigación.

Figura 3

Insumos del proyecto



Nota. Elaboración propia

4.4 Procesamiento

Actualmente la empresa no cuenta con un sistema que procese los datos recolectados. Por eso, se basa en el análisis de sus expertos para leerlos y así sacar las conclusiones o recomendaciones que sean útiles para el cumplimiento de sus objetivos empresariales. Para el procesamiento de los datos e insumos se deben realizar operaciones de registro, duplicación, verificación, separación, clasificación, intercalación, cálculo y recuperación. Estas acciones se efectúan a través de métodos manuales y electrónicos.

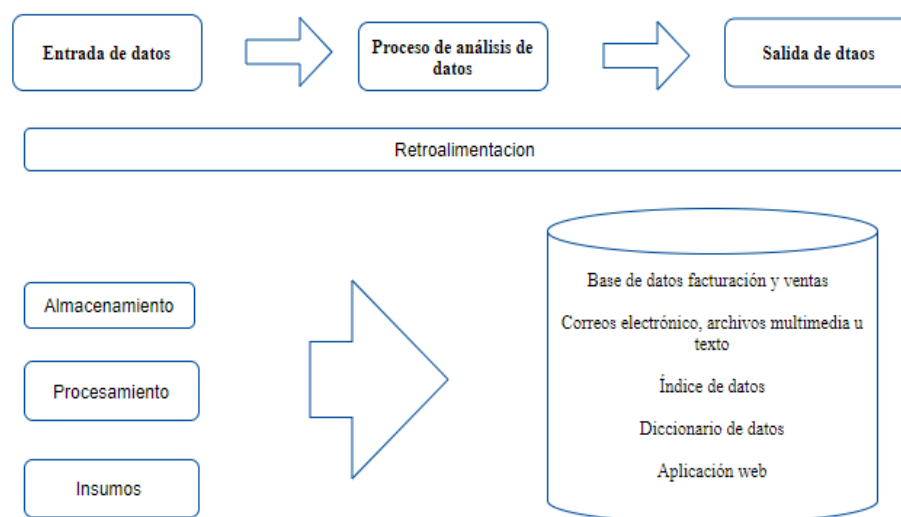
4.5 Almacenamiento

En Partners Technology S. A. C. el almacenamiento de sus bases de datos se realiza en un servidor a nivel *hardware* que permite gestionar y consultar los sus clientes remotos (Departamento de TI y Gerencia General, Financiera y Comercial). La compañía debe contar con servidores físicos y digitales (en la nube) que permitan la interacción más adecuada para consultas o proveeduría de información a sus principales clientes o *hosters*. El almacenamiento es una pieza clave dentro de la arquitectura TI empresarial, cuya calidad y eficacia impactará directamente en la capacidad de los usuarios (empleados, socios y clientes) para acceder de forma rápida y ágil a las aplicaciones que permitan satisfacer sus demandas de información.

Además, los sistemas de almacenamiento son críticos a la hora de desplegar estrategias de *backup* y continuidad de negocio. En la Figura 5 se detalla el proceso en las etapas de entrada, análisis y salida de datos.

Figura 4

Almacenamiento del proyecto



Nota. Elaboración propia

4.6 Destino

El destino será empleado en las herramientas Power BI, Excel y Python.

4.7 Tecnología, Herramientas e Infraestructura

Se trabajó con Google Collaborate, una iniciativa diseñada para construir y entregar rápidamente herramientas de ciencia e ingeniería de datos de clase mundial. Google Collaborate incluye cuadernos como Jupyter y Zeppelin, R Studio IDE y Open Refine.

Capítulo V. Requisitos y Recopilación de Datos

5.1 Determinación de Datos Necesarios para Responder las Preguntas

La recopilación de datos es fundamental para tomar buenas decisiones comerciales y para desarrollar productos valiosos y atractivos. Un sistema de recopilación de datos apropiado es de gran utilidad para identificar los problemas que puede experimentar una empresa e implementar las soluciones correctas (Melo, 2018). El punto de partida del presente trabajo de investigación es buscar los datos necesarios para conocer el comportamiento de los clientes de la empresa Partners Technology S. A. C., que corresponderían al sector, los años en el mercado, la cantidad de trabajadores, la facturación anual, el monto invertido en proyectos de implementación tecnológica, los gastos administrativos, el monto adeudado en Equifax, el monto de compra, los días de línea de crédito y el promedio de días de atraso en los pagos.

5.2 Ubicación de Fuentes de Datos

La siguiente tabla detalla la ubicación de los datos utilizados durante la investigación, los cuales provienen de fuentes internas y externas a la empresa.

Tabla 1*Ubicación de fuentes de datos*

Fuente	Tipo de fuente	Variable	Enlace
SUNAT	Externa	Razón social, RUC, sector, cantidad de años en el mercado, cantidad de trabajadores	https://e-consultaruc.sunat.gob.pe/cl-ti-itmrconsruc/jcrS00Alias
Base de datos propia de Partners Technology	Interna	Monto facturado, monto invertido en proyectos de implementación tecnológica, gastos administrativos,	Base de datos propia empleada para estudio de mercado
Equifax	Externa	Monto adeudado en Equifax.	https://soluciones.equifax.com.pe/
Base de datos propia de Partners Technology	Interna	Monto de compra, días de línea de crédito, días de atraso en pagos	Base de datos de clientes Partners Technology

Nota. Elaboración propia

5.3 Uso de Herramientas para Determinar Posibles Relaciones entre los Datos

Para visualizar correctamente la correlación entre las variables se identificó que el modelo más conveniente es el predictivo. En consecuencia, se emplearon las herramientas Microsoft Excel, Python y Power BI; la primera se utilizó para el análisis estadístico y la correlación lineal simple y múltiple. Python permitió elaborar un análisis de K-medias y el mapa de correlación y en Power BI se realizaron las visualizaciones que responden a las preguntas de investigación.

5.4 Revisión y Depuración de Datos que se Usaran en el Estudio

En primer lugar, se recopilaron 6343 metadatos obtenidos de la base de datos denominada Facturación Partners Technology 2019, los cuales estaban distribuidos en 23 columnas y 535 filas. Luego, se realizaron acciones de revisión y depuración en Excel para detectar datos incorrectos, incompletos, inexactos o irrelevantes. Por último, se empleó la metodología de *data cleansing* para mejorar la confianza de la organización en sus datos y modificarlos, sustituirlos o eliminarlos. Así, se determinó que el área de Tecnología de la Información es la más apta para realizar la depuración de datos dentro de la organización, y que la base de datos proporcionada por la empresa no contaba con los metadatos adecuados para alcanzar el propósito de la investigación.

Se encontraron 101 datos inválidos que fueron modificados y 120 nulos, los cuales fueron completados. Sumado a esto se realizaron otras acciones con el fin de evitar la generación de valores atípicos que no permitirían que la información se presentara de manera clara, como se describe a continuación:

- Se eliminaron las columnas *fecha de emisión*, *fecha de vencimiento*, *número de operación*, *porcentaje de descuento del banco*, *pago de Izzipay*, *retención del 3 %*, *detracción del 10 %* y *penalidad de la base de datos de facturación*, pues no contienen datos relevantes para el estudio.
- Se convirtieron las ventas en dólares a moneda nacional usando el tipo de cambio SUNAT como referencia (TC 3.51) y se agruparon las columnas *monto facturado en soles* y *monto facturado en dólares* en una misma.
- Se eliminó la columna *sector* porque solo consideraba los sectores Gobierno, privado y educación.

- Se eliminaron errores semánticos de los metadatos, es decir, signos de puntuación como comas, puntos, guiones, entre otros.
- Se agregaron las columnas *cantidad de años en el mercado*, *cantidad de trabajadores*, *facturación del cliente en el año 2019*, *monto invertido en gastos administrativos* y *monto adeudado en Equifax*.
- Se transformaron los metadatos que eran incongruentes y se verificó la veracidad de estos, al corroborarlos con las fuentes mencionadas. En esta fase se corrigieron, principalmente, errores semánticos, como la correcta escritura de la razón social y su relación con las variables RUC y sector. Esto se hizo con el objetivo de uniformizar y revisar la heterogeneidad de la fuente de datos.

En conclusión, se depuraron 4203 datos que representan el 66.26 % de total presentado. Además, se filtraron 250 metadatos que presentaban errores semánticos y se agregaron 4186 datos que corresponden a seis nuevas variables que se obtuvieron de diversas fuentes de información. Luego de verificar la veracidad de las fuentes estos datos fueron agregados a la nueva base de datos, la cual se renombró como *base de datos clientes Partners Technology*, que contiene 7176 datos divididos en 12 columnas y 598 filas.

5.5 Calidad de Datos

La metodología empleada en la presente investigación garantiza la exactitud, la completitud, la integridad, la actualidad, la coherencia, la relevancia, la accesibilidad y la confiabilidad de los datos. Estas características permitirán la mejora continua del proyecto y darán una mayor precisión en las conclusiones. El intervalo de confianza de los metadatos es del 95 %. Se consideró pertinente manejar esta estadística dado que es la primera vez que se implementará un proyecto de ciencia de datos en la compañía.

Según la revista *PowerData* (2014) esta data debe cumplir el grado en el que un conjunto de características inherentes se desempeñe con los requisitos, esto es, con la necesidad o expectativa establecida, generalmente implícita u obligatoria. Por ello, se realizaron las siguientes acciones de metodología de calidad de datos:

- Perfilado. Se analizaron los datos desde un punto de vista estructural, es decir, se validó la correcta sintaxis de estos.
- Limpieza. Se empleó la metodología de *data cleansing*.
- *Match*. Se corroboró la data con las estadísticas brindadas de las diversas fuentes de datos recopiladas en la investigación.
- Consolidado. Se consolidaron en una sola base de datos los metadatos limpios y los nuevos que fueron obtenidos de fuentes externas a la organización.
- Monitoreo. Se monitoreó y verificó que los datos mostrados en la nueva data brindaran información relevante acerca del objetivo de la investigación.

De esta forma, se propuso que el área administrativa sea la encargada de realizar las acciones de calidad de datos dentro de la organización, al ser la que cuenta con información relevante acerca de los clientes. Asimismo, se recomendó que este proceso cíclico sea llevado a cabo de manera periódica, pues se requiere que los metadatos estén actualizados para elaborar análisis y evaluar los avances del objetivo. La integración de las áreas de la empresa permitirá que este proceso se ejecute de manera más rápida y fácil.

Figura 5

Metodología para la calidad de datos



Nota. Elaboración propia

5.6 Gobierno de Datos

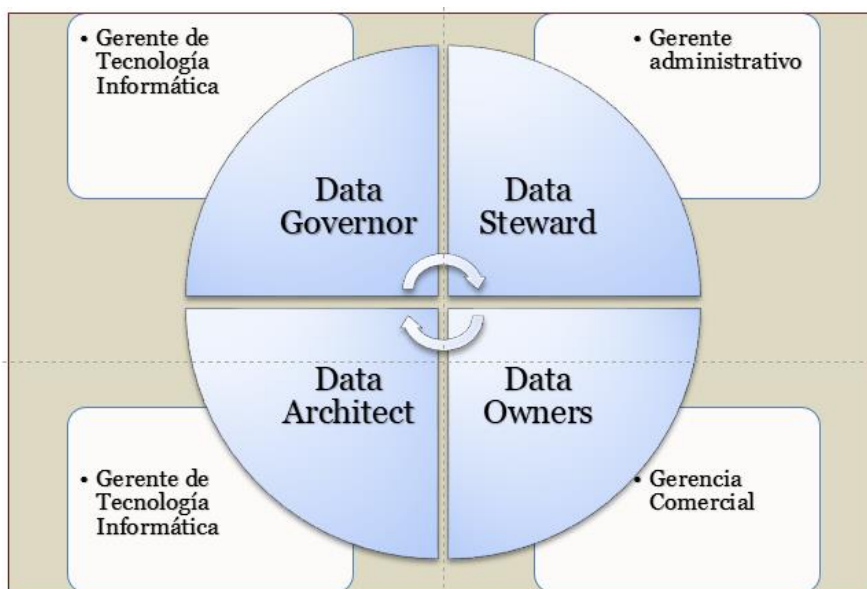
El gobierno de datos tiene como finalidad gestionar la información de la empresa Partners Technology S. A. C. mediante la identificación de su procedencia. Por eso, se llegó a la conclusión de que el área de tecnología de la información es la más apta para controlar, usar, gestionar y resguardar la información contenida en la data. Este departamento se encargará de confirmar la procedencia de los datos, así como de evaluar si estos son consecuentes con las políticas de la empresa y su importancia dentro de la organización.

Una definición clara es que el gobierno de datos es una disciplina encargada de la orquestación de gente, procesos y tecnología que permite habilitar a una compañía a impulsar la información como un recurso de valor empresarial, y al mismo tiempo, es la encargada de mantener a los usuarios, auditores y reguladores satisfechos, usando la mejora de la calidad de los datos para retener clientes, constituyendo y guiando a nuevas oportunidades en el mercado. (IBM, 2012, párr. 4)

Para un correcto gobierno de datos es necesario reconocer los roles y las funciones de los trabajadores de las tres áreas de la empresa. El objetivo de este proceso es establecer políticas y estándares para el uso y gestión de los datos dentro de la organización. Así las cosas, los roles necesarios son: *data governor*, *data steward*, *data architect* y *data owners*.

- *Data governor*: el gerente de tecnología de la información proporcionó los datos necesarios realizar la investigación.
- *Data steward*: el gerente administrativo de la compañía se encarga de brindar la información a todas las áreas. Para el presente proyecto se le solicitó su permiso para acceder a la base de datos de facturación del año 2019.
- *Data architect*: es el gerente de tecnología de información.
- *Data owners*: es el gerente comercial.

El proceso de recopilación, interpretación y análisis de los datos mostrados en el presente trabajo de investigación se llevó a cabo teniendo en cuenta el perfil de los usuarios. Esto quiere decir que se priorizó su fácil entendimiento y accesibilidad, por lo cual los datos fueron almacenados en lugares simples para su posterior ubicación. Sin embargo, esto no significa que se encuentren a simple vista de personas externas de la organización, sino que prima la seguridad mediante un adecuado gobierno de datos.

Figura 6*Gobierno de datos*

Nota. Elaboración propia

La metodología de gobierno de datos empleada en la presente investigación priorizó los siguientes objetivos:

- Establecer metas para guiar y desarrollar las actividades de gobierno de datos efectuadas en este y en futuros proyectos de investigación.
- Definir métricas que permitirán evaluar la efectividad del proyecto.
- Tomar decisiones adecuadas mediante políticas enfocadas en el cumplimiento de las métricas y objetivos de la empresa Partners Technology S. A. C.
- Comunicar las políticas a toda la organización para que se implementen y estandaricen los procesos.
- Medir los resultados a través de la retroalimentación y verificación del proceso de gobierno de datos.

- Auditar permanentemente para corroborar la funcionalidad del proyecto e identificar errores

5.7 Diccionario de Datos

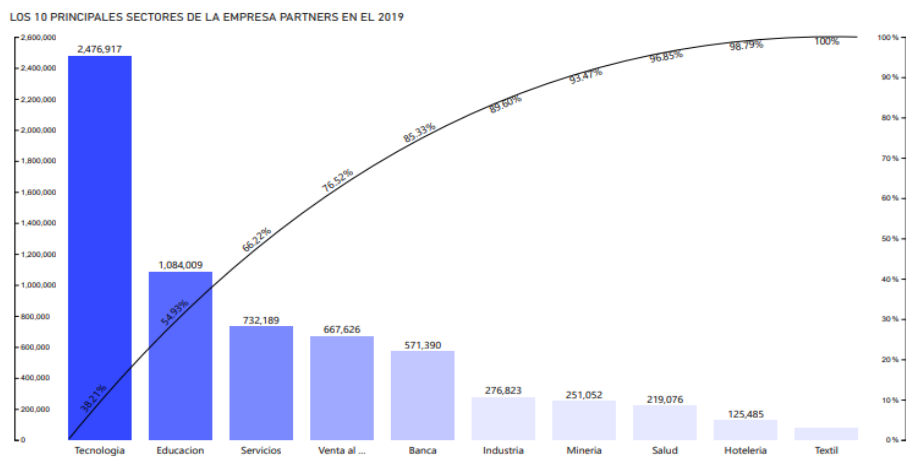
El diccionario que se realizó en la presente investigación se encuentra detallado en el Anexo 7. Este contribuirá a fortalecer la gestión de datos y orientar a los usuarios sobre el tipo de contenido que encontrarán. Con su implementación se logrará un ordenamiento funcional y conceptual de los datos, mayor control de la información, la construcción de modelos de datos y la consolidación de conceptos claros, unificados y consistentes de uso diario para la entidad.

5.8 Visualización de Datos

Con el objetivo de responder a las preguntas propuestas y contar la historia que se pretende se elaboraron ciertas visualizaciones. Como bien se dijo, los gráficos desarrollados durante el proceso se realizaron con la ayuda de Excel, Power BI y Google Collaborate, herramientas que facilitarán la toma de decisiones en el área comercial de la empresa Partners Technology S. A. C. al ayudar a entender mejor las principales variables que determinan el comportamiento de sus clientes. Esta información se elaboró teniendo en cuenta a los usuarios, es decir, a la gerencia comercial de la empresa. Por esta razón, se piensa que el área de Tecnología de la Información sea la encargada de realizar las visualizaciones. Es importante reiterar que estas representaciones permitirán que la información sea comprendida de manera rápida y concisa. Por tanto, con base en el objetivo que se quiere conseguir y los resultados que se desean comunicar, se priorizaron la estética y la accesibilidad.

Figura 7

Los diez sectores principales de la empresa Partners Technology en el 2019



Nota. Elaboración propia

Entonces, se hace evidente que los sectores tecnología, educación y servicios representan un 66.22 % de las ventas totales de Partners Technology S. A. C.

Capítulo VI. Producción y Análisis Exploratorio de Datos

6.1 Plan de Muestreo y Tamaño de la Muestra

De acuerdo con Kleeberg y Ramos (2009) la estadística busca hacer inferencias acerca de una población con base en la información contenida en una muestra. Por consiguiente, el muestreo se define como la disciplina que trata con el conjunto de técnicas para tomar u obtener una muestra. El objetivo del plan de muestreo de la presente investigación es seleccionar un grupo de clientes de la empresa Partners Technology S. A. C, los cuales representan la población del estudio, para analizar sus datos e inferir correlaciones. Para ello se emplearon 598 registros sobre sus principales características con la finalidad de responder a las preguntas de *Data Science* planteadas. Mediante el uso del método aleatorio simple se determinó que la muestra está compuesta por 79 clientes. Sin embargo, para tener mayor certeza y exactitud se trabajó con la totalidad de los datos. Durante el desarrollo de la investigación se empleó un nivel de confianza del 95 %, siendo el coeficiente de certeza de 1.96. Asimismo, se definió que la probabilidad de éxito del proyecto será del 50 %.

6.2 Determinación de la Población objetivo (Elemento, Unidad, Alcance y Tiempo)

- Elemento muestral: clientes del sector privado de Partners Technology S. A. C.
- Unidad muestral: facturación total de Partners Technology S. A. C.
- Alcance: Lima y provincias de Perú
- Tiempo: la investigación registra a los clientes que realizaron compras desde el 1 de febrero del 2015 hasta el 31 de diciembre de 2019.

- Marco muestral: compras de equipos tecnológicos, cantidad de trabajadores, años en el mercado, cantidad de trabajadores, inversión en proyectos de implementación tecnológica, línea de crédito, días de atraso de pagos.

6.3 Diseño del Estudio de Muestreo

Antes de realizar el estudio de muestreo se identificó el tipo y la escala de las variables empleadas en el estudio, como se muestra en la Tabla 2.

Tabla 2

Clasificación de variables según tipo y escala

Variable	Tipo de variable	Escala
Razón social	Cualitativa	Nominal
RUC	Cuantitativa	Discreta
Sector	Cualitativa	Nominal
Cantidad de años en el mercado	Cuantitativa	Discreta
Cantidad de trabajadores	Cuantitativa	Discreta
Monto facturado	Cuantitativa	Continua
Monto invertido en proyectos de implementación tecnológica	Cuantitativa	Continua
Gastos administrativos	Cuantitativa	Continua
Monto adeudado en Equifax	Cuantitativa	Continua
Monto de compra	Cuantitativa	Continua
Días de línea de crédito	Cuantitativa	Discreta
Días de atraso en pagos	Cuantitativa	Discreta

Nota. Elaboración propia

En esta investigación cada cliente representa un parámetro dado que se obtuvo de una población total, que son todas las compañías que atendió Partners Technology S. A. C. hasta el 31 de diciembre de 2019. Por ello, solo se utilizaron las variables cuantitativas para el diseño del estudio de muestreo, así:

$f(x) =$ [Cantidad de años en el mercado, cantidad de trabajadores, facturación del cliente, monto invertido en proyectos de implementación tecnológica, monto invertido en gastos administrativos, monto adeudado en Equifax, monto de compra a Partners Technology, días de línea de crédito, promedio de días de atraso].

6.4 Ejecución del Muestreo

Se utilizó el método aleatorio simple para la ejecución del plan de muestreo, empleando una población finita, con base en la siguiente fórmula:

Donde los elementos del muestreo son:

- n = tamaño de la muestra
- Z = nivel de confianza deseado
- p = probabilidad de éxito
- q = probabilidad de fracaso
- e = margen de error
- N = tamaño de la población

Para hallar la muestra se utilizó la herramienta Google Collaborate mediante el lenguaje de programación Python. Se obtuvo como resultado 78.53 clientes. Es importante mencionar que cada uno de ellos está representado por un número entero y, debido a que el margen de error es del 5 %, se decidió que el tamaño muestral es un total de 79 clientes. Para fines didácticos se trabajó con el total de la población para contar con mayor exactitud.

6.5 Resumen y Examen de la Distribución de Variables

Interpretación de la Distribución Normal de las Variables Cuantitativas

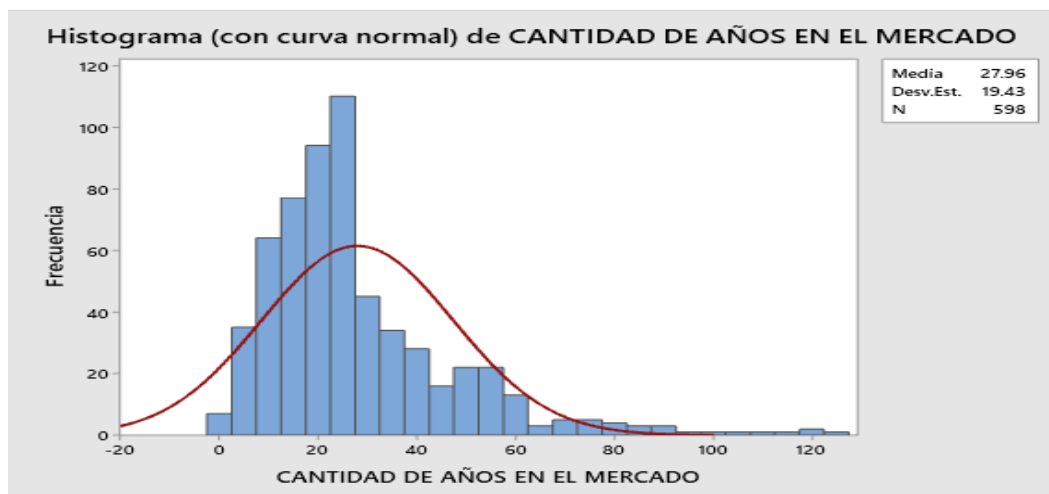
A continuación, se detalla el análisis estadístico y la distribución de las variables que permiten alcanzar los objetivos de la investigación.

Interpretación de la Distribución Normal de la Variable Cantidad de Años en el Mercado

Los resultados obtenidos señalan que para una muestra de 598 datos, la distribución de la variable *cantidad de años en el mercado* oscila entre 1 y 124, con un promedio de 27.96 y una desviación estándar de 19.43, como se plasma en la Figura 9.

Figura 8

Histograma con curva normal de cantidad de años en el mercado



Nota. Elaboración propia

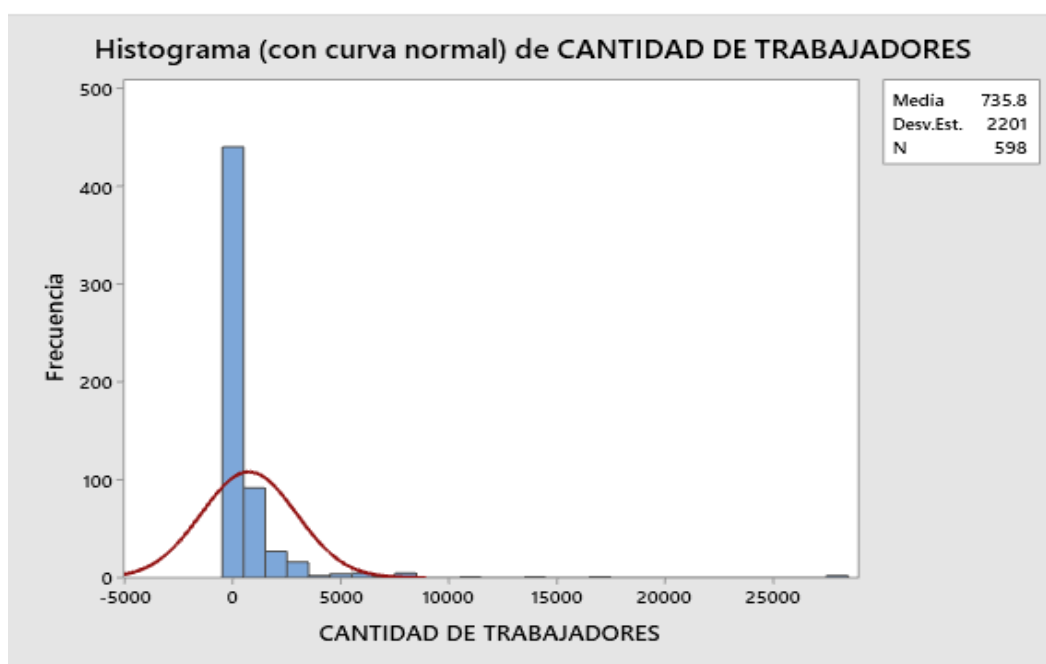
La distribución tiene asimetría positiva y se alarga a valores mayores que la media. Así, se verifica que el 65.72 % de los clientes no supera los 27 años de experiencia en el mercado.

Interpretación de la Distribución Normal de la Variable Cantidad de Trabajadores

Para la misma muestra la distribución de la variable *cantidad de trabajadores* oscila entre 1 y 28 300, con un promedio de 735.83 y una desviación estándar de 2201.42, como se muestra en la Figura 10.

Figura 9

Histograma con curva normal de cantidad de trabajadores



Nota. Elaboración propia

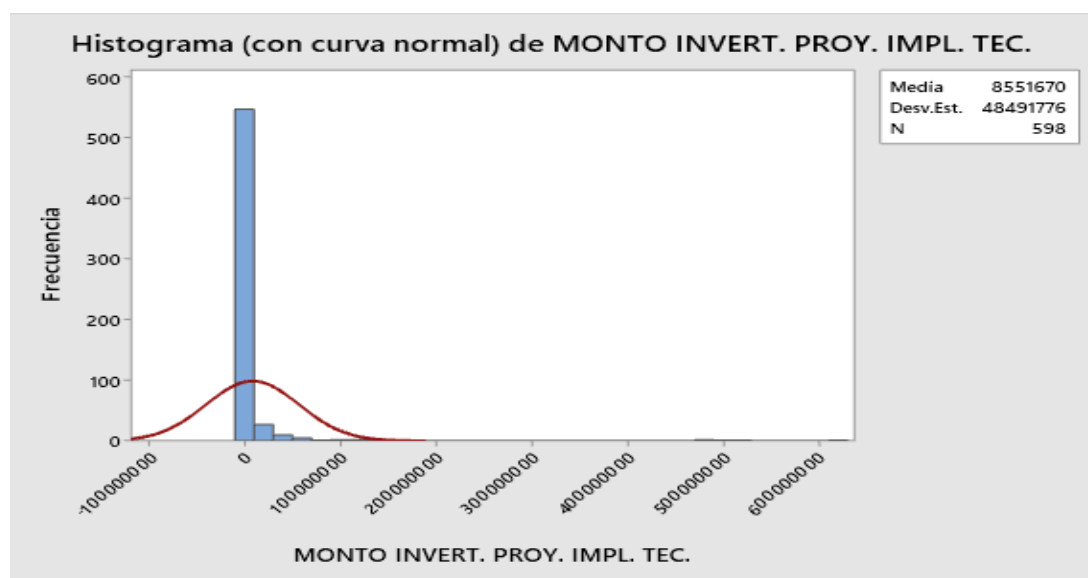
La distribución tiene asimetría positiva y se alarga a valores mayores que la media. Entonces, el 80.43 % de los clientes posee menos de 735 trabajadores.

Interpretación de la Distribución Normal de la Variable Monto Invertido en Proyectos de Implementación Tecnológica

La distribución de la variable *monto invertido en proyectos de implementación tecnológica* en 2019 oscila entre 50 000 y 615 000 000, con un promedio de 8 551 669.656 y una desviación estándar de 48 491 776.21 (ver Figura 11).

Figura 10

Histograma con curva normal de cantidad de monto invertido en proyectos de implementación tecnológica



Nota. Elaboración propia

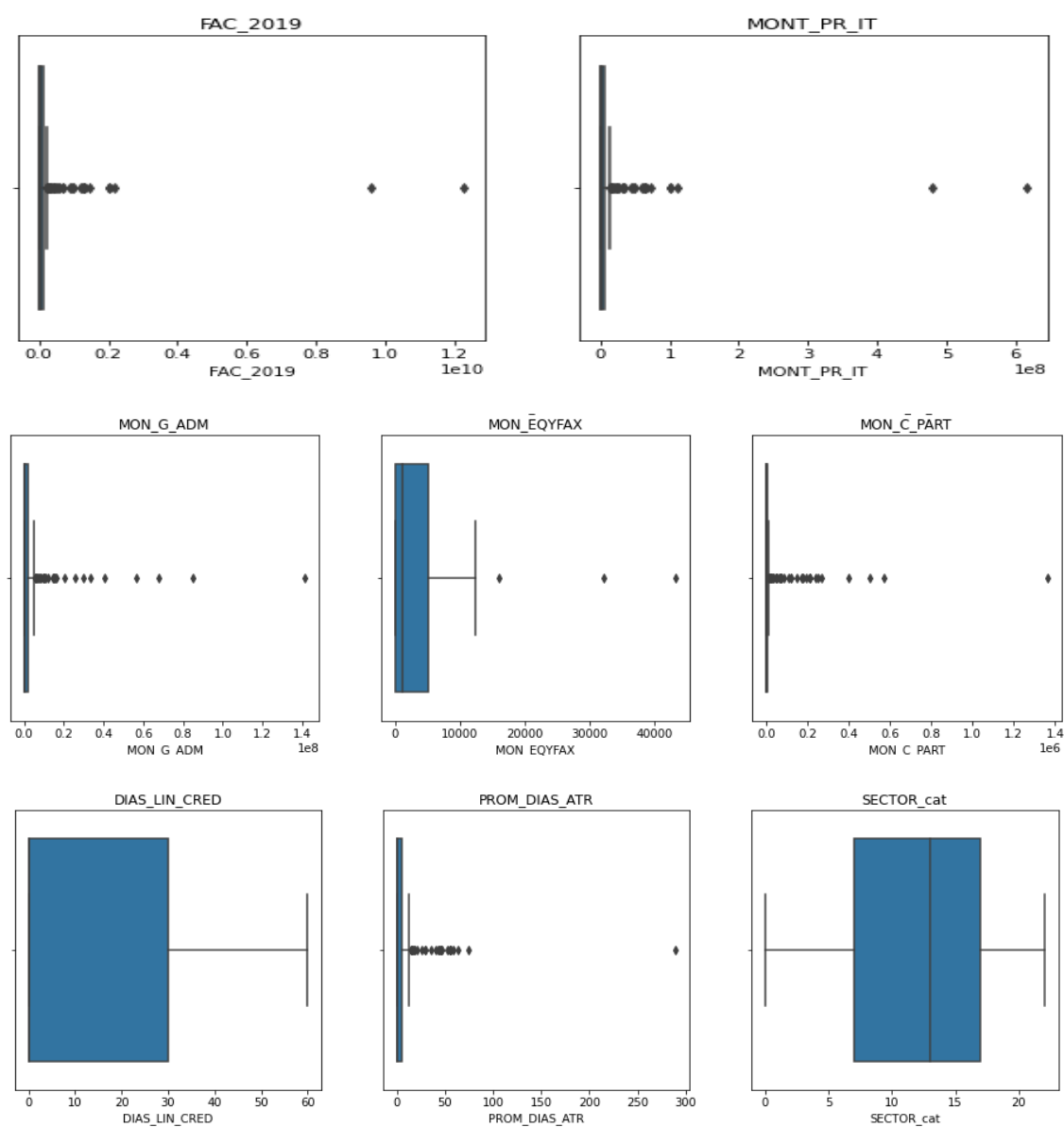
La distribución tiene asimetría positiva y se alarga a valores mayores que la media. Se evidencia existe una mayor proporción de empresas con un monto invertido en proyectos de implementación tecnológica que no supera los S/ 8 551 670.

6.6 Distribución Estadística de las Variables

A continuación, se presentan los diagramas de caja y bigotes que muestran la distribución de las variables.

Figura 11

Diagrama de caja y bigotes



Nota. Elaboración propia

Es notorio que existe una alta concentración de los diferentes montos en cero, pero a la vez hay clientes que poseen valores más elevados, quienes se presentan como *outliers*. Los valores atípicos son representados por los clientes grandes que posee la empresa y que realizan una mayor inversión en el rubro tecnológico. Esto se comprueba al momento de realizar el modelo predictivo.

Figura 12*Cuadro de medidas de resumen de las variables*

Estadísticas	Cantidad de años en el mercado	Cantidad de trabajadores	Facturación del cliente en el 2019	Monto invertido en proyectos de implementación	Monto invertido en gastos administrativos	Monto adeudado en Equifax	Monto de compra a <i>PartnersTechnology</i> en el 2019	Días de línea de crédito	Promedio días de atraso
Media	27.96	735.83	171,033,393.12	8,551,669.66	3,679,155.52	4,912.20	11,079.99	4.88	2.45
Error típico	0.79	90.02	39,659,523.86	1,982,976.19	450,113.88	291.86	2,948.36	0.48	0.59
Mediana	23.00	155.50	22,025,000.00	1,101,250.00	777,500.00	3,823.00	-	-	-
Moda	25.00	2.00	45,000,000.00	2,250,000.00	10,000.00	-	-	-	-
Desviación estándar	19.43	2,201.42	969,835,524.30	48,491,776.21	11,007,102.12	7,137.16	72,099.32	11.73	14.49
Varianza de la muestra	377.36	4,846,251.88	940,580,944,192,288,000.00	2,351,452,360,480,720.00	121,156,296,982,474.00	50,938,985.14	5,198,312,045.17	137.55	209.98
Curtosis	4.28	89.42	105.93	105.93	89.42	26.59	217.58	4.62	263.27
Coefficiente de asimetría	1.76	8.32	10.05	10.05	8.32	4.50	13.05	2.31	14.19
Rango	123.00	28,299.00	12,299,000,000.00	614,950,000.00	141,495,000.00	70,255.00	1,361,343.21	60.00	290.00
Mínimo	1.00	1.00	1,000,000.00	50,000.00	5,000.00	-	-	-	-
Máximo	124.00	28,300.00	12,300,000,000.00	615,000,000.00	141,500,000.00	70,255.00	1,361,343.21	60.00	290.00
Suma	16,722.00	440,027.00	102,277,969,085.00	5,113,898,454.25	2,200,135,000.00	2,937,497.60	6,625,834.49	2,917.00	1,467.00
Cuenta	598.00	598.00	598.00	598.00	598.00	598.00	598.00	598.00	598.00
Percentil 25	15.00	44.50	8,080,160.00	404,008.00	222,500.00	1,121.00	-	-	-
Percentil 50	23.00	155.50	22,025,000.00	1,101,250.00	777,500.00	3,823.00	-	-	-
Percentil 75	35.00	534.00	56,001,750.00	2,800,087.50	2,670,000.00	6,126.75	-	-	-

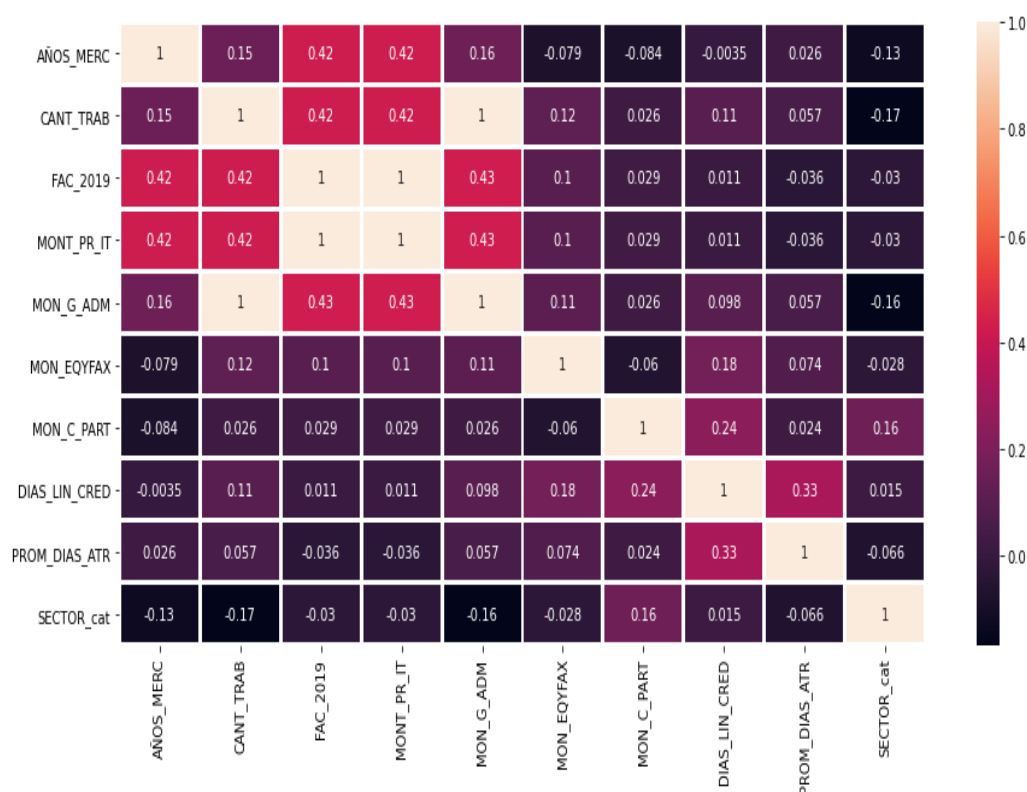
Nota. Elaboración propia

6.7 Resultado del Análisis de Relaciones

Como ya se mencionó, se utilizó la herramienta Google Collaborate mediante el lenguaje de programación Python para determinar las relaciones entre las variables. Se encontraron dos correlaciones interesantes que permitirán alcanzar los objetivos de la investigación. En la Figura 13 se representan los coeficientes de correlación entre las variables utilizadas.

Figura 13

Mapa de correlación de variables



Nota. Elaboración propia

Como se puede observar existe una fuerte correlación entre las variables *años en el mercado*, *cantidad de trabajadores* y *monto invertido en proyectos de implementación tecnológica*.

Evaluar esta última resulta útil para alcanzar los objetivos trazados. Así las cosas, resulta

pertinente pasar a explicar la ecuación de la regresión lineal simple para las variables mencionadas.

6.8 Análisis de Regresión Lineal Múltiple

El estudio de regresión lineal múltiple tiene como objetivo predecir el monto de inversión en proyectos de implementación tecnológica de los clientes de la empresa *Partners Technology S. A. C.* La elección de las variables se basa en su grado de correlación. Posteriormente, la ecuación resultante de la regresión lineal múltiple se ejecutará en el modelado. Es importante mencionar que para el estudio se identificaron las variables dependientes e independientes.

- Variable dependiente (Y): monto de invertido en proyectos de implementación tecnológica
- Variable independiente 1 (X_1): cantidad de años en el mercado
- Variable independiente 2 (X_2): cantidad de trabajadores

A continuación, se detallan las estadísticas de regresión de las variables mencionadas.

Tabla 3

Estadísticas de la regresión

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.42
Coefficiente de determinación R^2	0.17
R^2 ajustado	0.17
Error típico	44227002.3
Observaciones	598

Nota. Elaboración propia

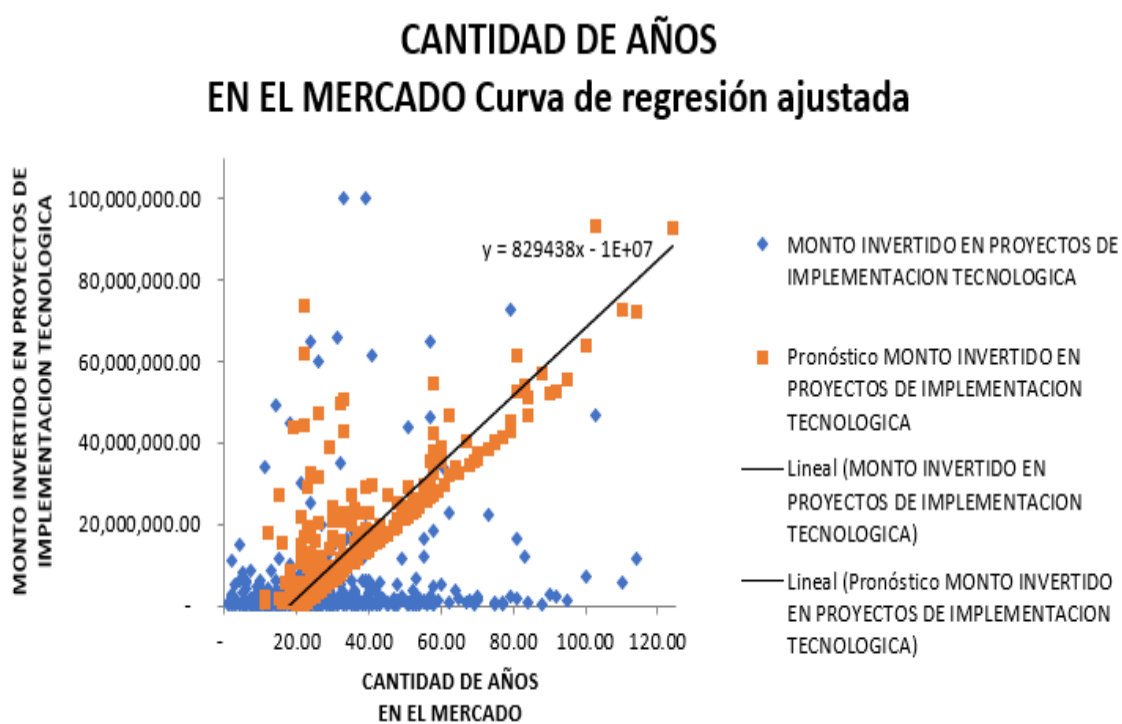
A partir de las estadísticas anteriores se llega a la siguiente ecuación de regresión lineal múltiple:

$$y = 746983.8X_1 + 5468.65X_2 - 16360394.92.$$

El coeficiente de correlación 0.42 es un valor óptimo que indica una alta correlación entre las variables mencionadas, lo cual lleva a lograr los objetivos de la investigación. Es importante mencionar que las empresas destinan un presupuesto especial a este rubro con el propósito de reducir costes, incrementar sus ganancias y diseñar estrategias para aumentar la productividad de sus trabajadores.

Figura 14

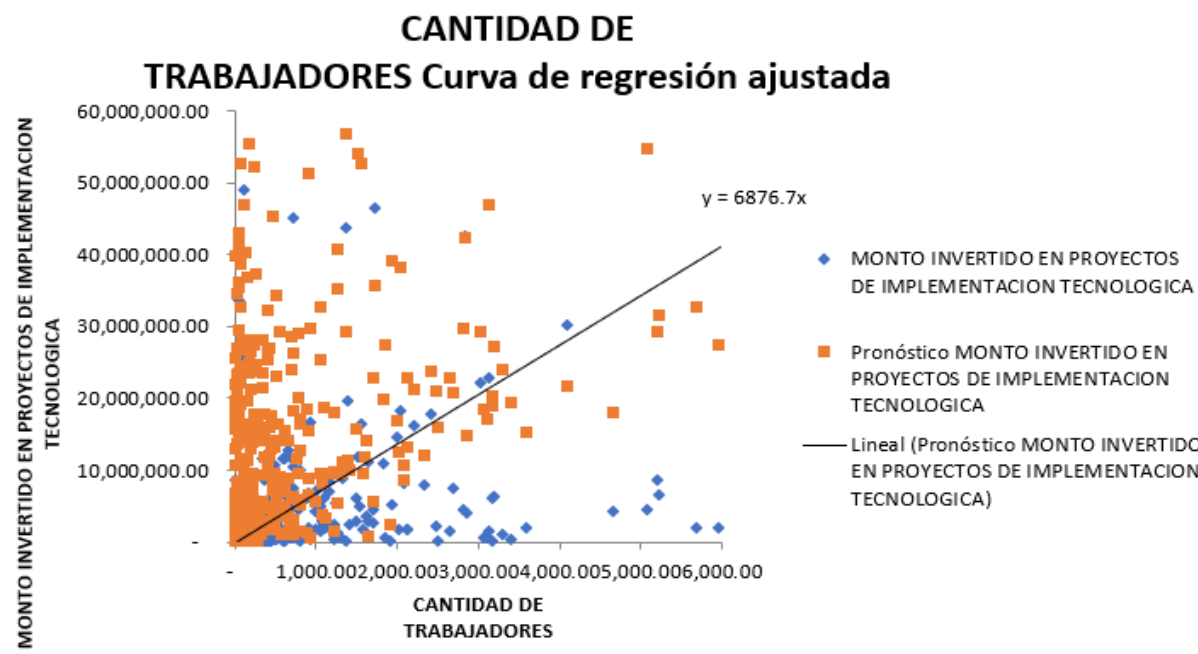
Gráfica de la correlación lineal múltiple entre las variables cantidad de años en el mercado y monto invertido en proyectos de implementación tecnológica.



Nota. Elaboración propia

Figura 15

Gráfica de la correlación lineal múltiple entre las variables cantidad de trabajadores y monto invertido en proyectos de implementación tecnológica.



Nota. Elaboración propia

6.9 Estudio Probabilístico de los Datos

Para la ejecución del estudio probabilístico se emplearon los datos de la variable *monto invertido en proyectos de implementación tecnológica* de los sectores educación y servicios, los cuales comprenden 51 y 54 registros, respectivamente. Se escogieron dichos sectores porque se encuentran dentro de los cinco con mayor cantidad de clientes y, según la información recopilada, tienen tendencia favorable a desarrollarse.

Hipótesis: existe alguna diferencia entre la variación del monto invertido en proyectos de implementación tecnológica en el sector educativo y el de servicios.

Hipótesis Nula: $H_0: \sigma_1^2 = \sigma_2^2$

Hipótesis alterna: $H_0: \sigma_1^2 \neq \sigma_2^2$

Tabla 4

Prueba F para varianzas de dos muestras.

	<i>Educación</i>	<i>Servicios</i>
Media	7717324.76	14039453.78
Varianza	2.13574E+14	4.49732E+15
Observaciones	51	54
Grados de libertad	50	53
F	0.047489203	
P(F<=f) una cola	0	
Valor crítico para F (una cola)	0.628338185	

Nota. Elaboración propia

Al ser $F <$ valor crítico se acepta la hipótesis nula y, en consecuencia, se rechaza la hipótesis alterna. Esto quiere decir que sí existe una diferencia entre la variación de proyectos de implementación tecnológica en el sector educativo y el de servicios.

Capítulo VII. Técnicas de Ciencia de Datos

7.1 Técnica de Ciencia de Datos Supervisada

La técnica de ciencia de datos supervisada fue empleada en este trabajo para predecir el monto de inversión en proyectos de implementación tecnológica de los clientes de la empresa Partners Technology S. A. C. Para ello se utilizó la ecuación detallada en el capítulo VI, esta es:

$$y = 746983.8X_1 + 5468.65X_2 - 16360394.92.$$

Donde

y = monto invertido en proyectos de implementación tecnológica

X_1 = cantidad de años en el mercado

X_2 = cantidad de trabajadores

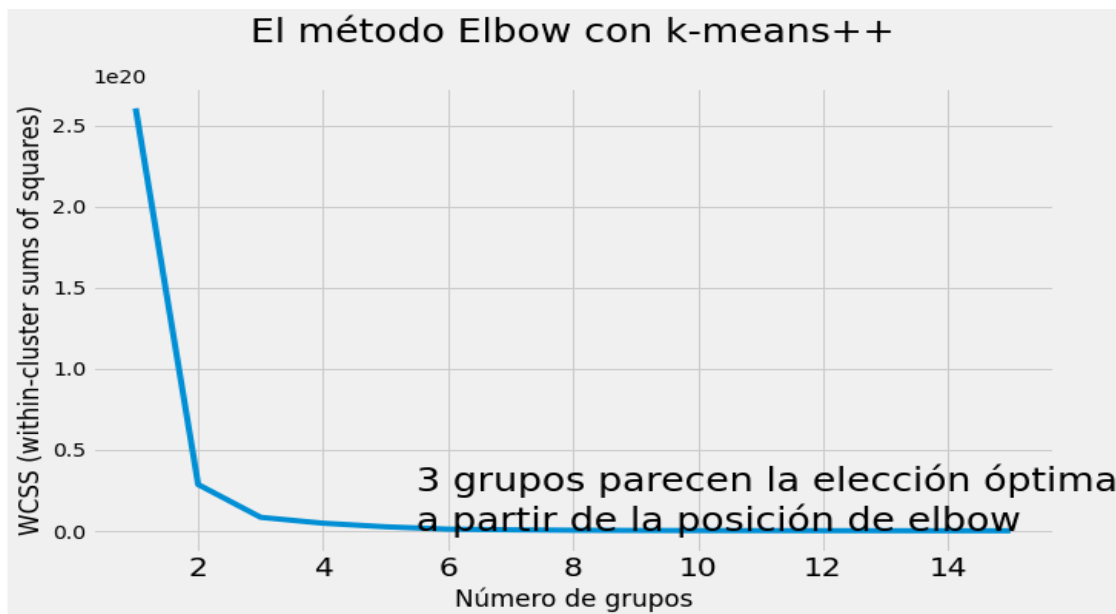
Como resultado de la ejecución de esta técnica a la población total del estudio, se obtuvo que las diez empresas que más invertirán en proyectos de implementación tecnológica son Rímac Seguros y Reaseguros, Teleatento del Perú S. A. C., Laive S. A., Pontificia Universidad Católica del Perú, Banco de Crédito del Perú, Interbank, Camposol S. A., Sociedad Minera Austria Duvaz S. A. C, British American Hospital S. A. y El Pedregal.

7.2 Técnica de Ciencia de Datos no Supervisada

También se empleó la técnica de ciencia de datos no supervisada de agrupamiento con el objetivo de segmentar a los clientes de la empresa y responder a las preguntas de *Data Science*. Para lograrlo se utilizó la herramienta K-means, con la que se eliminaron los datos nulos y se obtuvieron 205 registros de clientes, los cuales fueron segmentados en tres clústeres según su proximidad a la media. De igual forma, como se observa en la figura siguiente, se utilizó el método del codo.

Figura 16

El método Elbow con K-means++



Nota. Elaboración propia

Este método fue empleado porque permite verificar que un valor de $k = 3$ es adecuado, al ser el punto en el cual el grupo de datos tiene el menor error posible. Es importante mencionar que se seleccionaron únicamente las variables cuantitativas. En la Figura 19 se detalla la creación de tres clústeres, que se agruparon por su proximidad a la media de cada variable; los datos que se encuentren cerca a dichos parámetros quedarán agrupados en 0.1 o 2.

Figura 17

Tabulación de clústeres.

	AÑOS_MERC	CANT_TRAB	FAC_2019	MONT_PR_IT	MON_G_ADM	MON_EQYFAX	MON_C_PART	DIAS_LIN_CRED	PROM_DIAS_ATR	SECTOR_cat
0	28.136126	581.513089	6.569322e+07	3.024056e+06	2.633099e+06	2884.698953	31502.917225	14.015707	7.492147	11.455497
1	122.000000	9995.500000	1.095269e+10	5.476347e+08	4.997750e+07	7906.500000	21926.435000	15.000000	0.000000	11.000000
2	40.333333	3873.666667	1.402024e+09	7.010120e+07	1.936833e+07	3216.258333	37859.665000	15.000000	3.000000	9.916667

Nota. Elaboración propia

En la Tabla 8 se muestra la distribución de clientes por grupos.

Tabla 5

Distribución de clientes por grupos.

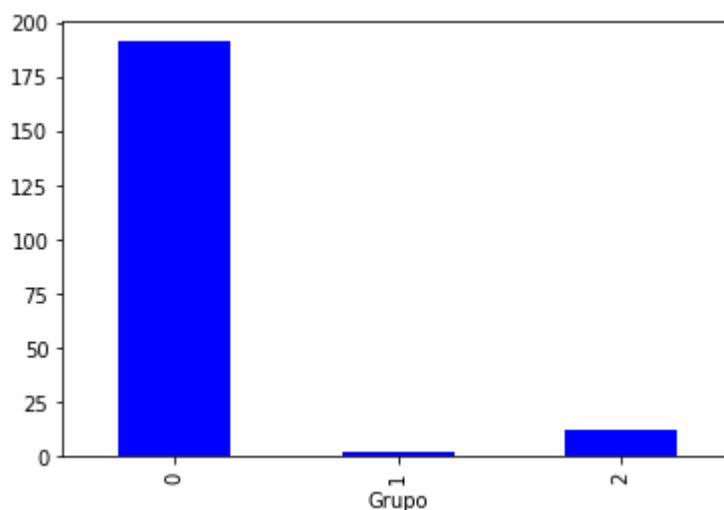
Grupo	Cantidad de clientes
0	191
1	2
2	12

Nota. Elaboración propia

Así las cosas, se observa que existen tres grupos bien definidos de clientes, que se diferencian por las variables *monto invertido en proyectos de implementación tecnológica*, *cantidad de años en el mercado*, *cantidad de trabajadores*. A su vez, se identifica que estas últimas variables son directamente proporcionales e influyen en el tamaño de una compañía, que puede ser pequeña, mediana o grande. La misma situación se presenta con la variable *monto invertido en proyectos de implementación tecnológica*. Es posible decir que los grupos de clientes 1 y 2 son los que poseen mayor inversión en proyectos de implementación tecnológica. Entonces, con base en esta información se concluye que estas empresas son grandes, debido a que poseen mayor cantidad de años en el mercado y de trabajadores. Igualmente, se determinó la cantidad de clientes para cada grupo como se muestra en las siguientes figuras.

Figura 18

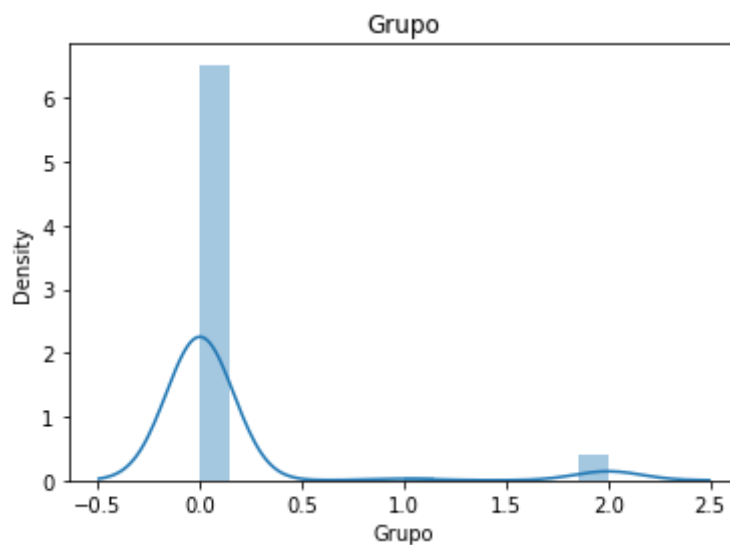
Gráfico de cantidad de empresas por clúster.



Nota. Elaboración propia

Figura 19

Gráfico de curva de comportamiento de los clústeres.



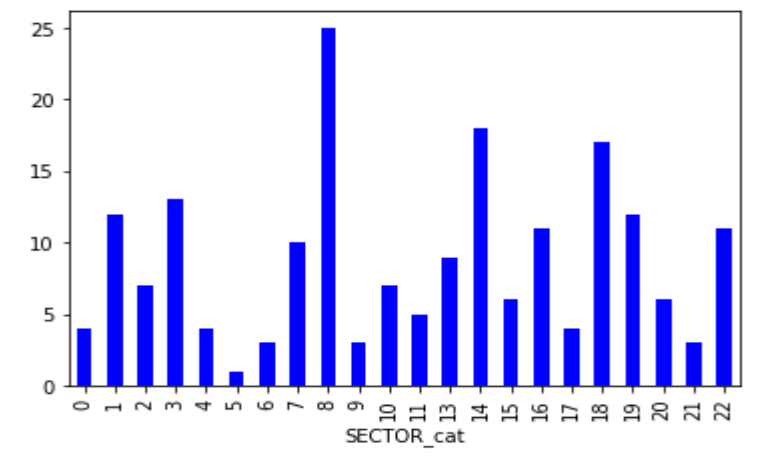
Nota. Elaboración propia

En estas figuras se observa que los clientes del grupo 0 son 189 más que en el grupo 1, y 179 más que el grupo 2. Sin embargo, los más convenientes son el 1 y 2 puesto que invierten más en proyectos de implementación tecnológica. Asimismo, se concluye que los clientes tipo 0 son en

su mayoría pymes dado que la media de sus variables es baja. Tras este análisis, se dio una etiqueta a cada sector para definir la composición de los grupos y se descubrió la siguiente composición:

Figura 20

Grupo 0.

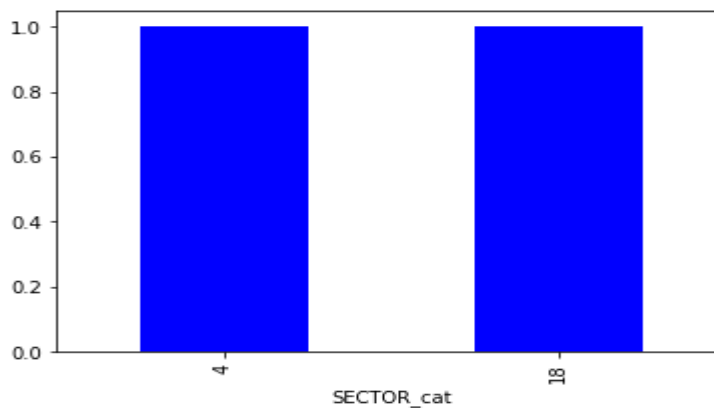


Nota. Elaboración propia

Se identifica que la mayoría de los clientes del grupo 0 pertenecen a los sectores educación, manufactura y servicios.

Figura 21

Grupo 1

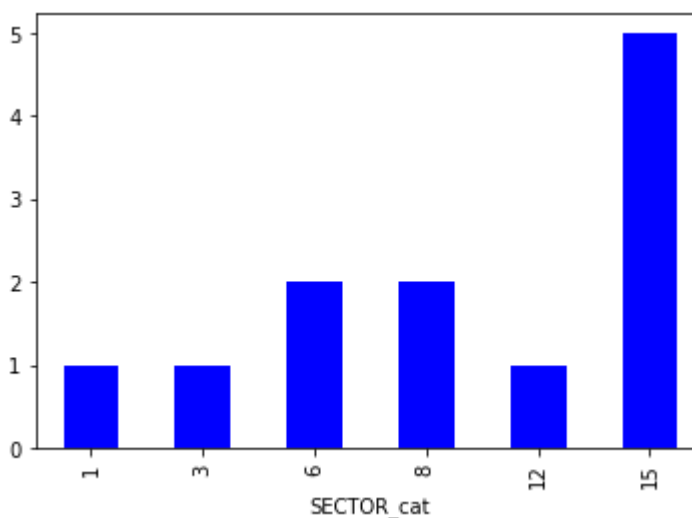


Nota. Elaboración propia

Por su parte, los clientes del grupo 1 pertenecen a los sectores automotriz y servicios.

Figura 22

Grupo 2



Nota. Elaboración propia

Finalmente, los clientes del grupo 2 pertenecen en su mayoría al sector manufactura. De forma específica se tiene que los clientes de cada grupo son:

- Grupo 1. Automotores Gildemeister-Perú S. A. y Rimac Seguros y Reaseguros.
- Grupo 2. Banco de Crédito del Perú, Compañía Minera Ponderosa S. A., Compañía Minera Santa Luisa S. A., American Airlines, Camposol S. A., Compañía Minera Morococha S. A., Compañía Minera Coimolache S. A., Compañía Minera Casapalca, Lima Airport Partners S. A., Mi Banco-Banco de la Microempresa S. A., Universidad Peruana de Ciencias Aplicadas S. A. y Adecco Consulting S. A.

Capítulo VIII. El Modelo y la Historia

8.1 Determinación del Modelo más Apropriado para la Descripción o Predicción

Se concluye que el método más apropiado para este proyecto es el predictivo en vista de que permite conocer a los clientes *target* de la compañía Partners Technology S. A. C. a partir de los datos etiquetados.

8.2 Descripción de los Hallazgos

Según lo detallado en el análisis se reconoció que el enfoque analítico predictivo conduce a la identificación de las variables más importantes que influyen en el comportamiento de compra de los clientes de la empresa Partners Technology S. A. C. Estas son: monto invertido en proyectos de implementación tecnológica, cantidad de años en el mercado y cantidad de trabajadores.

8.3 Definición y Justificación de los Criterios para la Visualización de los Resultados

Para definir y justificar los datos la visualización de los resultados se optó por utilizar la herramienta Google Collaborate mediante el lenguaje de programación Python por su fácil uso y accesibilidad. Se consideraron cuatro criterios fundamentales para su Elaboración, la conveniencia, la disrupción, la estética y la jerarquía.

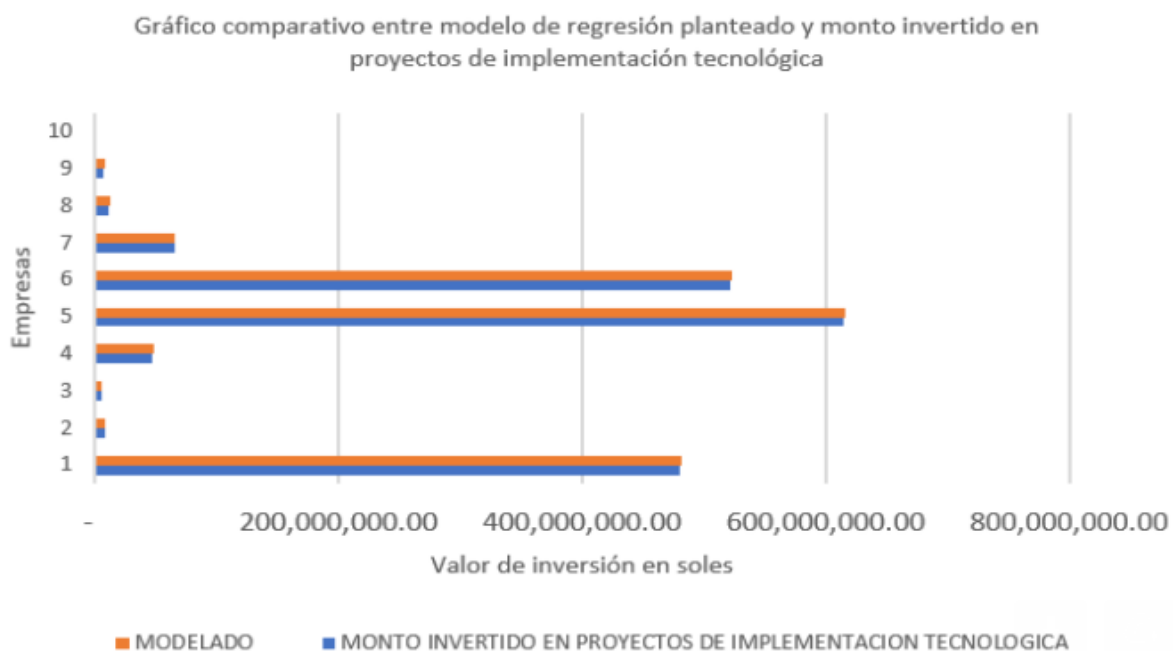
- Conveniencia: se decidió emplear Python por su ser una herramienta de fácil acceso y uso.
- Disrupción: se decidió utilizar este lenguaje de programación porque no es convencional y representó un reto para el equipo de trabajo que elaboró la presente investigación.
- Estética: se procuró que las visualizaciones fueran armoniosas y tuvieran colores sobrios para su fácil lectura.

8.4 Estrategias de Implementación

El objetivo de la estrategia de implementación del presente estudio es incrementar en un 20 % el monto de compra de la empresa Partners Technology S. A. C mediante la atención a sus clientes. Esto se logrará mediante la implementación de programas que busquen motivar la compra de los clientes registrados en su base de datos. En este punto es importante mencionar que la pandemia COVID-19 obligó a muchas compañías del sector privado a detener sus proyectos de implementación tecnológica e, incluso, sus operaciones. Por lo mismo, se requiere tomar acciones que contrarresten los efectos de la competencia y de la situación actual. La predicción del monto invertido en proyectos de implementación tecnológica permitirá a Partners Technology disminuir costos y recursos al atender clientes que no le generen un volumen de compra importante. Por eso, se propone aplicar la técnica de ciencia de datos en el área de Tecnologías de la Información para filtrar a los clientes *target* y ejecutar una estrategia de ventas en torno a ellos.

Figura 23

Gráfico comparativo entre modelo de regresión planteado y el monto invertido en proyectos de implementación tecnológica.



N°	RAZON SOCIAL
1	RIMAC SEGUROS Y REASEGUROS
2	TELEATENTO DEL PERU S.A.C.
3	LAIVE S A
4	PONTIFICIA UNIVERSIDAD CATOLICA DEL PERU
5	BANCO DE CREDITO BCP
6	INTERBANK
7	CAMPOSOL S.A.
8	SOCIEDAD MINERA AUSTRIA DUVAZ S.A.C.
9	BRITISH AMERICAN HOSPITAL S.A.
10	EL PEDREGAL S.A

Nota. Elaboración propia

Como se observa en la figura anterior, mediante el modelo de regresión lineal múltiple se predice que las empresas que más invertirán en proyectos de implementación tecnológica son el Banco de Crédito del Perú e Interbank, ambas pertenecientes al sector financiero.

Capítulo IX. Calibrado e Implementación

9.1 Métodos de Ajuste

El método de ajuste para la regresión lineal múltiple se hizo a través de la evaluación del error típico; es decir, la variación producida por valores distorsionantes o poco conocidos. Por esta razón, mientras más pequeño sea el error típico, más certeras serán las predicciones que se realicen a partir de dicho modelo. Es preciso mencionar que se utilizó la herramienta de *Excel* para realizar el método

Calibrado al 40 %

En la primera fase del calibrado se utilizaron 239 registros de clientes, los cuales representan el 40 % de los datos. Aquí, se identificó que el coeficiente de correlación múltiple es 0.56 y el error típico es 61741926.

Tabla 6

Estadísticas de la regresión.

<i>Estadísticas de la regresión</i>	
Coeficiente de correlación múltiple	0.56
Coeficiente de determinación R ²	0.32
R ² ajustado	0.31
Error típico	61741926
Observaciones	239

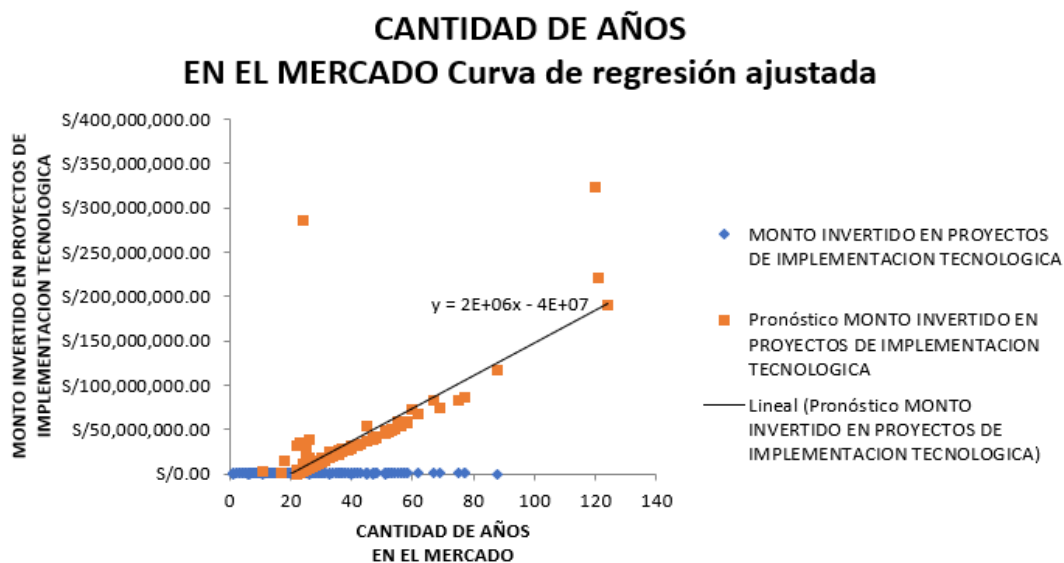
Nota. Elaboración propia

Análisis de variables dependientes

- X_1 = cantidad de años en el mercado

Figura 24

Gráfico de relación cantidad de años vs. monto invertido de proyectos de implementación tecnológica.



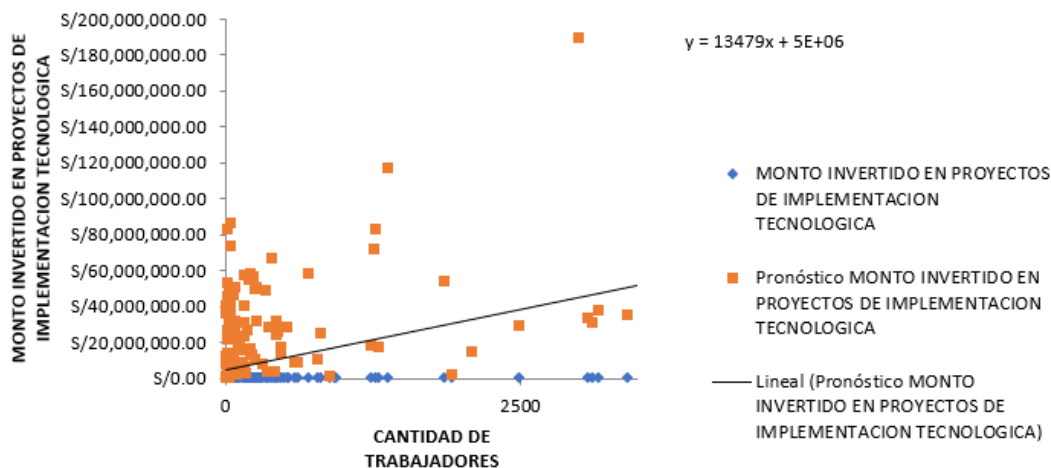
Nota. Elaboración propia

- X_2 = cantidad de trabajadores.

Figura 25

Gráfico de relación cantidad de trabajadores vs. monto invertido en proyectos de implementación tecnológica.

CANTIDAD DE TRABAJADORES Curva de regresión ajustada



Nota. Elaboración propia

Calibrado al 60 %

En la segunda fase del calibrado se utilizaron 359 registros de clientes, que representan el 60 % de los datos. Se identificó que el coeficiente de correlación aumentó en 0.02 y el error típico disminuyó en 47 225 123, con respecto al calibrado al 40 %.

Tabla 7

Estadísticas de la regresión

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0.58
Coefficiente de determinación R ²	0.09
R ² ajustado	0.09
Error típico	14516803
Observaciones	359

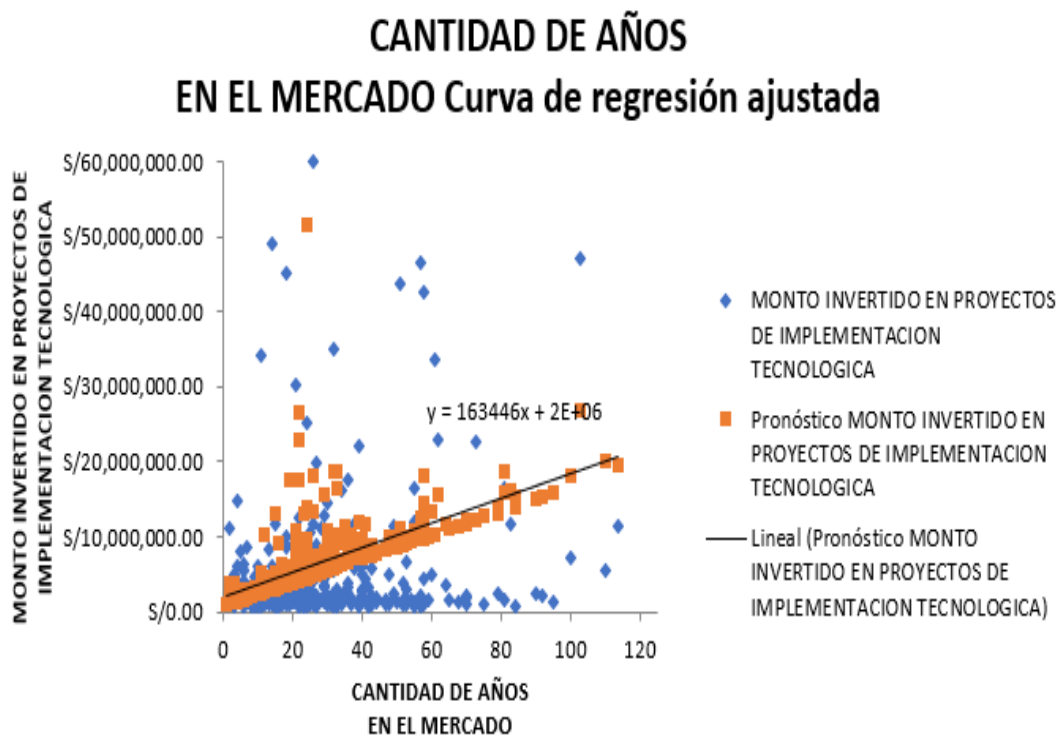
Nota. Elaboración propia

Análisis de variables dependientes

- X₁ = cantidad de años en el mercado.

Figura 26

Gráfico de relación cantidad de años vs. monto invertido en proyectos tecnológicos.

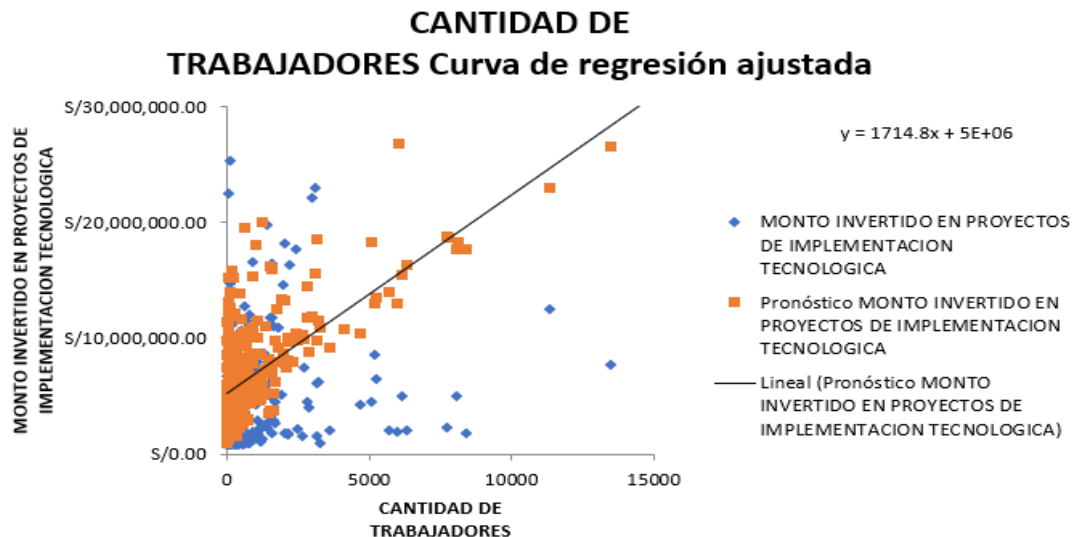


Nota. Elaboración propia.

- X_2 = cantidad de trabajadores.

Figura 27

Gráfico de relación cantidad de cantidad de trabajadores vs. monto invertido en proyectos.



Nota. Elaboración propia

Calibrado al 87 %

En esta etapa del calibrado se demostró que el modelo de regresión múltiple permitirá elaborar predicciones certeras. Lo mencionado se fundamenta en la disminución del error típico y el aumento del coeficiente de correlación en cada una de las etapas del proceso.

Tabla 8

Estadísticas de la regresión

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.61
Coefficiente de determinación R^2	0.20
R^2 ajustado	0.19
Error típico	45361883
Observaciones	521

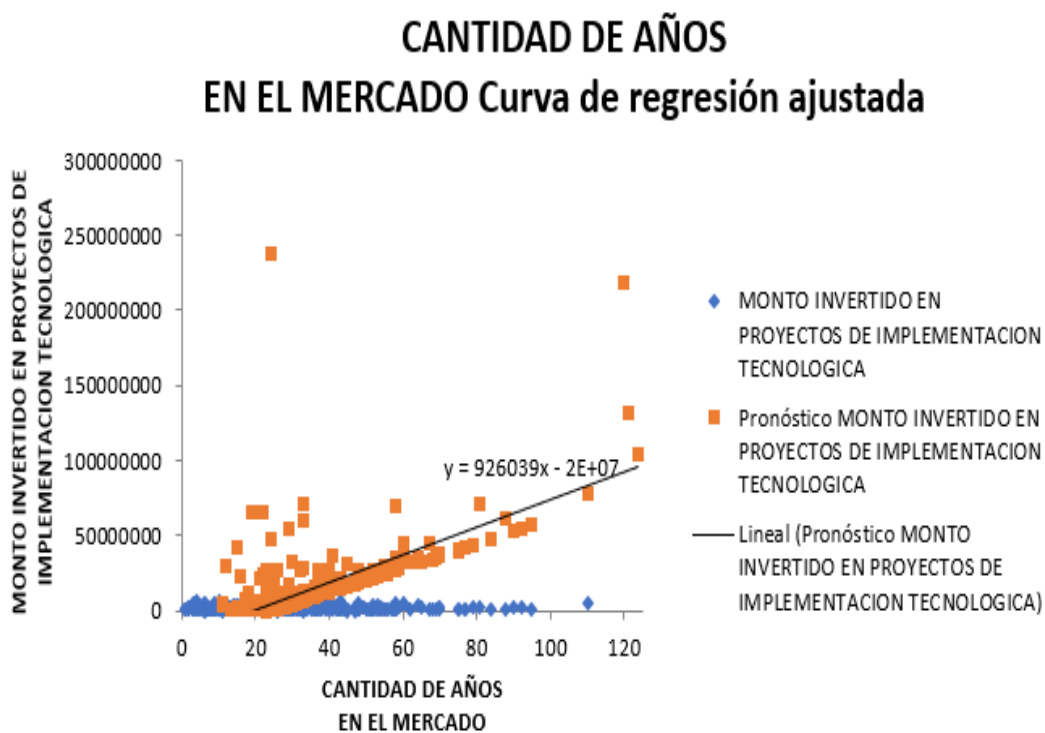
Nota. Elaboración propia

Análisis de variables dependientes

- X_1 = cantidad de años en el mercado.

Figura 28

Gráfico de relación cantidad de años vs. monto invertido en proyectos tecnológicos.

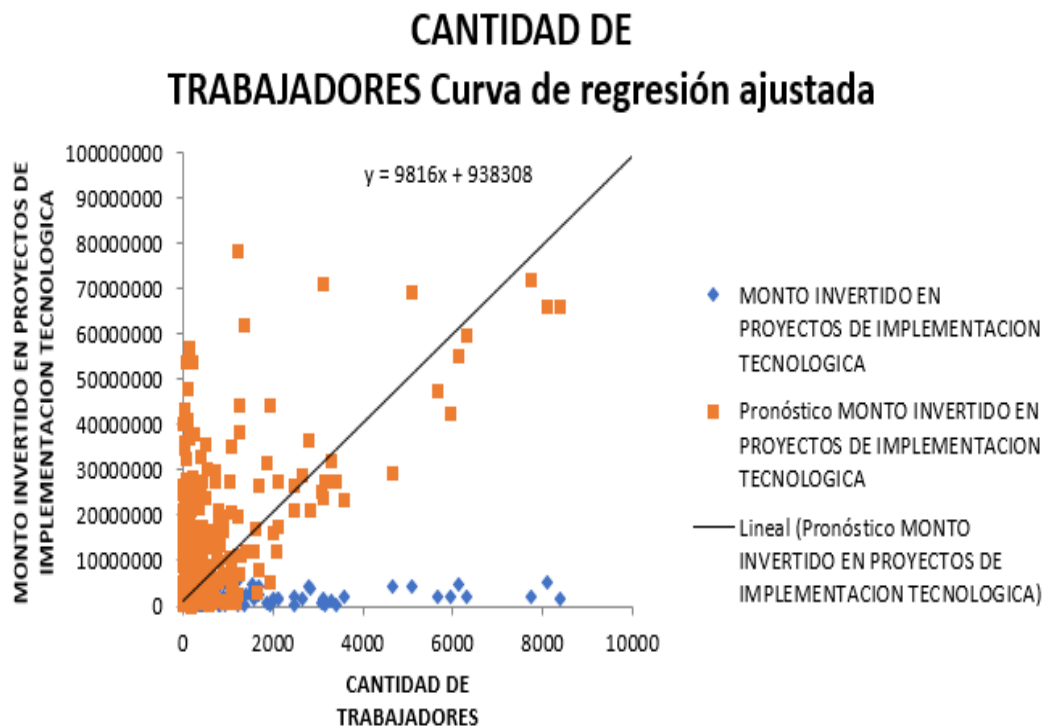


Nota. Elaboración propia

- X_2 = cantidad de trabajadores

Figura 29

Gráfico de relación cantidad de trabajadores vs. monto invertido en proyectos de implementación tecnológica.



Nota. Elaboración propia

Finalmente se testeó el modelo en base a la última calibración de 87% de la población y el 13% restante.

Figura 30

Gráfico de testeo del modelo.



Nota. Elaboración propia

9.2 Retroalimentación

Del resumen de estadísticas se identificó que el método supervisado de regresión lineal múltiple es el modelo más óptimo para predecir los montos de compra de los clientes de la empresa Partners Technology S. A. C. Esto se sustenta en que el valor del coeficiente R^2 ajustado es 0.19, lo cual indica que el modelo es confiable. Así también se comprueba la hipótesis propuesta en el estudio.

Conclusiones

Los 10 clientes target de la empresa Partners Technology S. A. C son: Rímac Seguros y Reaseguros, Teletanto del Perú S. A. C., Laive S.A, Pontificia Universidad Católica Perú, Banco de Crédito del Perú, Interbank, Camposol S. A., Sociedad Minera Austria Duvaz S. A. C., British American Hospital S. A., El Pedregal S. A. Las tres principales variables para crear el modelo de regresión lineal múltiple, el cual se reconoce como el más óptimo, son: monto invertido en proyectos de implementación tecnológica, cantidad de trabajadores y cantidad de años en el mercado.

Por consiguiente, se concluye que los segmentos de clientes 1 y 2 son los que más invertirán en proyectos de implementación tecnológica. Asimismo, se encuentra viable elaborar un modelo de regresión lineal múltiple que prediga el monto invertido en proyectos de implementación tecnológica de los clientes de la empresa Partners Technology S. A. C y que permita conocer el comportamiento de los clientes.

Factores Críticos del Éxito del Proyecto

El presente trabajo de investigación se enfocó en encontrar soluciones que ayuden a la gerencia a tomar mejores decisiones corporativas por medio de la arquitectura tecnológica propuesta. Por ende, los factores críticos de su éxito fueron aquellos que brindaron una visión integral de la gestión de datos de la empresa *Partners Technology S. A. C.* y que pueden llevar a cumplir los siguientes puntos:

- Incrementar la participación de mercado de la empresa por medio de la mejora de la gestión de los datos de sus clientes.
- Incrementar la facturación mediante una adecuada selección de sus clientes.
- Clasificar los clientes de la compañía según su rentabilidad esperada.
- Generar nuevas visualizaciones que mejoren la toma de decisiones a nivel gerencial.
- Determinar variables que ayuden al análisis de los clientes.

Así las cosas, los factores críticos que favorecieron al éxito del presente estudio fueron: el hallazgo de las variables determinantes, el esclarecimiento de los objetivos propuestos, el análisis minucioso de los datos y de sus resultados, y la ejecución de los procesos de la metodología de la ciencia de datos.

Decisiones por Tomar a partir de los datos estudiados

1. Se nombrarán *data owners* y se le brindará adecuado acceso a la información de los datos a cada uno, de acuerdo con el gobierno de datos recomendado.
2. El departamento comercial incorporará el uso de Power Bi para un mejor análisis de sus clientes.
3. La empresa utilizará una plataforma en la nube para la administración de la información, así como para su correcta distribución.

Referencias Bibliográficas

- Day, R., & Gastell, B. (2008). *Publicación Científica y Técnica N.º 621*. Organización Panamericana de la Salud. Washington.
- IBM. (2012). *Seis pasos para el gobierno de datos*.
<https://developer.ibm.com/es/technologies/data-management/articles/gobierno-datos/>
- Kleeberg, F., & Ramos, J. (2009). Aplicación de las técnicas de muestreo en los negocios y la industria. *Ingeniería Industria*, 27, 11-40.
- Melo, S. (2018). *Conoce los diferentes métodos de recopilación de datos*.
<https://mydatascope.com/blog/es/conoce-los-diferentes-metodos-de-recopilacion-de-datos/>
- ONU Mujeres. (2012). *Análisis PESTEL*. <https://www.endvawnow.org/es/articles/1182-analisis-pestel.html>
- Porter, M. (2008). *Las cinco fuerzas competitivas que le dan forma a la estrategia*. Harvard Business Review.
- PowerData. (2014). *Introducción a la calidad de datos: definición, control y beneficios*.
redacción. <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/368784/introduccion-a-la-calidad-de-datos-definicion-control-y-beneficios>
- Rollins, J. (2015). *Metodología fundamental para la ciencia de datos*. IBM:
<https://www.ibm.com/downloads/cas/WKK9DX51>

ANEXOS

Anexo 1.

Aliados clave	Actividades clave	Propuesta de Valor	Relación con el cliente	Segmentos de clientes
Marcas de tecnología: HP, Lenovo, Dell, otros. Canales mayoristas: Ingram Micro, Deltron, Máxima Internacional, Nexys, otros.	Venta al por mayor y menor de equipamiento y soluciones tecnológicas	Soluciones tecnológicas integrales que incluyen equipamiento tecnológico y servicios a cargo de profesionales especializados en cada área. Acompañamiento a los clientes en la propuesta, implementación, puesta en marcha, capacitación y seguimiento a las soluciones ofrecidas de la mano de los fabricantes de tecnología.	Atención personalizada mediante canales post venta y reuniones periódicas.	Gobierno Privado Educación Usuarios Finales
	Recursos clave		Canales de distribución	
	Recursos humanos: personal altamente calificado Infraestructura: oficinas para atención al público, almacenes y operaciones Tecnología: equipamiento y <i>softwares</i>		Oficinas y almacén ubicados en San Isidro Entrega: Dirección que indique el cliente	
Estructura de costes		Estructura de ingresos		
Costos fijos: salarios de personal, alquiler de oficinas, pago de servicios Costos variables: transporte interno de productos, costo de equipamiento, mano de obra externa		Venta de equipamiento y servicios tecnológicos al contado o a crédito, arrendamiento tecnológico, financiamiento bancario, servicios de outsourcing		

Anexo N° 2 :Análisis FESTEL

Análisis PESTEL		
P – Políticos	E – Económicos	S - Social
<p>1.- Inestabilidad política que pone en riesgo las inversiones y la continuidad de proyectos.</p> <p>2.- Promoción de proyectos de financiamiento por parte del poder ejecutivo como Activa Perú.</p> <p>3.- Proximidad de las elecciones presidenciales del 2021 que podría aumentar la inestabilidad.</p> <p>4.- Deficiente manejo político de la pandemia que afecta a todos los sectores económicos del Perú.</p>	<p>1.-Recesión económica en el Perú producto de la reducción del PBI y el incremento del desempleo.</p> <p>2.- Descenso de los proyectos de inversión a causa de la pandemia COVID-19.</p> <p>3.- Reducción del presupuesto tecnológico de las entidades públicas debido al COVID-19.</p> <p>4.- Incremento del tipo de cambio que afecta en los costos de compra y venta de la compañía.</p>	<p>1.- El nivel de ingresos de los hogares ha disminuido afectando a las empresas.</p> <p>2.- El cambio en los hábitos de consumo ha incrementado el canal <i>e-commerce</i> en el país.</p> <p>3.- Incremento del home para prevenir contagios de COVID-19.</p> <p>4.- Incremento de la demanda de servicios educativos a nivel nacional.</p>
T – Tecnológicos	E – Ecológicos	L - Legal
<p>1.- La velocidad de las innovaciones tecnológicas beneficia a las compañías de TI.</p> <p>2.- El costo de acceso a nuevas tecnologías es cada vez menor debido al incremento de la oferta.</p> <p>3.- Incremento de la demanda de equipamiento tecnológico de consumo debido al <i>home office</i>.</p> <p>4.- Tendencia por usar herramientas virtuales de colaboración para realizar reuniones.</p>	<p>1.- Disminución de la contaminación debido a la tendencia por el trabajo remoto.</p> <p>2.-El tratado de Minamata reduce la importación y exportación de productos con mercurio.</p> <p>3.- Regulaciones que promueven el uso eficiente de la energía.</p> <p>4.- Conciencia social ecológica mediante el incremento del uso de bicicleta.</p>	<p>1.- Decreto supremo de estado de emergencia nacional que afecta a las industrias.</p> <p>2.- Regulaciones legales que reducen aforos e instalan protocolos sanitarios en las empresas.</p> <p>3.- Normativas legales que se deben cumplir para contrataciones con el Estado peruano.</p> <p>4.- Decreto Legislativo que crea la central de Compras Públicas- Perú Compras.</p>

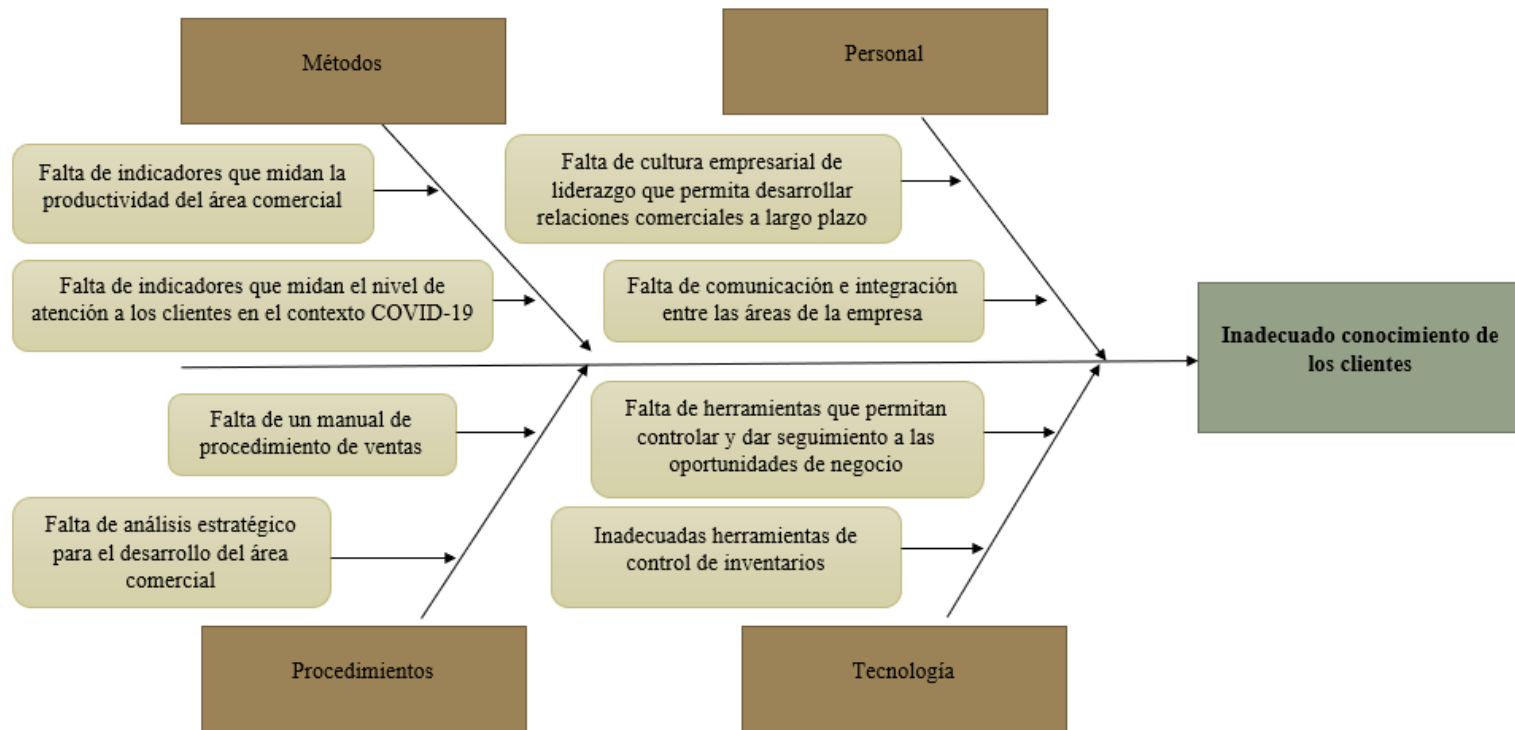
Anexo 3. Análisis FODA

Análisis FODA	
F- Fortalezas	D- Debilidades
<p>1.- Alianzas estratégicas con las principales marcas de tecnología y socios mayoristas que le permiten financiar grandes proyectos.</p> <p>2.- Portafolio amplio de soluciones tecnológicas que satisfacen las necesidades de tres grandes segmentos de clientes.</p> <p>3.- Personal altamente calificado y capacitado para brindar una óptima atención durante todo el proceso de venta.</p> <p>4.- Conocimiento y experiencia en el sector Gobierno, lo cual se refleja en su participación de mercado.</p>	<p>1.- La empresa no tiene una estrategia clara que segmente de manera adecuada a sus clientes.</p> <p>2.- La propuesta de valor de la empresa no está claramente diferenciada de su competencia.</p> <p>3.- Deficiente gestión comercial que impide la oportuna detección de nuevas oportunidades de negocio y seguimiento a las ventas.</p> <p>4.- Exceso de problemas operativos internos que se reflejan en el deficiente manejo de inventarios y tiempos de despacho.</p>
O- Oportunidades	A- Amenazas
<p>1.- Innovación tecnológica constante que, en consecuencia, demanda renovación periódica de las empresas.</p> <p>2.- El incremento del trabajo remoto aumentará la compra de equipamiento tecnológico en los hogares.</p> <p>3.- Incremento del canal <i>e-commerce</i> generará nuevas oportunidades de negocio y la posibilidad de atender nuevos mercados.</p> <p>4.- Desarrollo de empresas del rubro salud, comunicaciones, entre otros que requieren invertir en tecnología debido al COVID-19.</p>	<p>1.- Crisis política y económica que genera inestabilidad y disminuye las inversiones a nivel nacional.</p> <p>2.- La pandemia COVID-19 ha originado la reducción del PBI y el incremento del desempleo, afectando a las MYPE`S.</p> <p>3.- La constante variación del tipo de cambio impacta en la rentabilidad de los negocios.</p> <p>4.- Reducción de los proyectos tecnológicos en las empresas públicas y privadas como consecuencia de la recesión económica.</p>

Anexo 4. FODA Cruzado

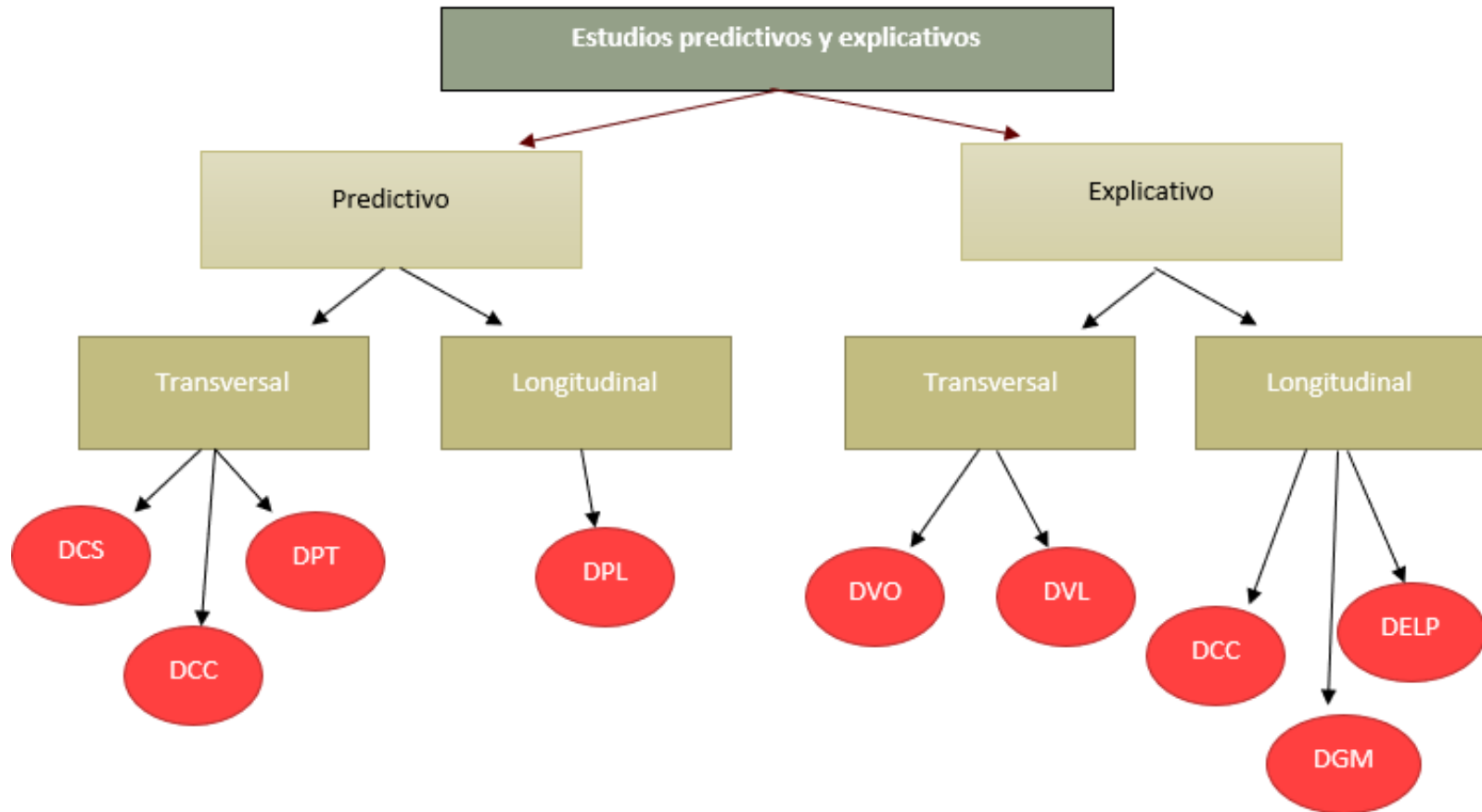
FODA cruzado	Fortalezas	Debilidades
<p>Oportunidades</p>	<p style="text-align: center;"><u>Estrategias ofensivas</u></p> <p>F.O (1.1): Creación de un programa de fidelización a clientes de la mano de sus principales aliados. F.O (1.2): Compra de productos tecnológicos para <i>stock</i> que ayuden a impulsar la venta a usuarios finales. F.O (2.3): Inversión en publicidad que ofrezca el catálogo de productos e incremente la presencia en el mercado.</p>	<p style="text-align: center;"><u>Estrategias de reorientación</u></p> <p>D.O (3.3): Impulsar la venta por redes sociales mediante el incremento del canal <i>e-commerce</i>. D.O (3.4): Potenciar los servicios enfocados a los sectores con mayor demanda en el contexto del COVID-19. D.O (4.1) Adquisición de un software de inventarios que permita controlar y mejorar los procesos existentes.</p>
<p>Amenazas</p>	<p style="text-align: center;"><u>Estrategias defensivas</u></p> <p>F.A (1.1): Impulsar programas junto a los principales socios para financiar los proyectos de los clientes. F.A (2.3): Fomentar programas que impulsen el arrendamiento tecnológico en los clientes. F.A (4.4): Programa para atender mayor cantidad de entidades públicas, las cuales cuentan con mayor solidez.</p>	<p style="text-align: center;"><u>Estrategias de supervivencia</u></p> <p>D.A (1.4): Enfocarse en un solo sector a fin de captar mayor cantidad de clientes y mejorar la atención. D.A (4.1): Modificar los procedimientos logísticos a fin reducir costos internos. D.A (3.2): Programas de capacitación para dar seguimiento a las oportunidades de negocio de los clientes.</p>

Anexo 5. Diagrama de Ishikawa



Anexo 6. Construcción de modelos explicativos y predictivos

Construcción de modelos explicativos y predictivos



Anexo 7. Diccionario de datos

Variable	Tipo	Longitud	Nulo	Índice	Único	Formato	Descripción	Regla de negocio
Razón Social	Cadena	150	No	No	No		Razón Social	Ninguna
RUC	Decimal	11	No	No	No	Entero 10, decimal 2	RUC	Ninguna
Sector	Cadena	50	No	No	No		Sector	Ninguna
Cantidad de años en el mercado	Decimal	2	No	No	No	Entero 10, decimal 2	Cantidad de años en el mercado	Ninguna
Cantidad de trabajadores	Decimal	5	No	No	No	Entero 10, decimal 2	Cantidad de trabajadores	Ninguna
Facturación del cliente en el 2019	Decimal	20	No	No	No	Entero 10, decimal 2	Facturación del cliente en el 2019	Ninguna

Monto invertido en proyectos de implementación tecnológica	Decimal	20	No	No	No	Entero 10, decimal 2	Monto invertido en proyectos de implementación tecnológica	Ninguna
Monto invertido en gastos administrativos	Decimal	20	No	No	No	Entero 10, decimal 2	Monto invertido en gastos administrativos	Ninguna
Monto adeudado en Equifax	Decimal	20	No	No	No	Entero 10, decimal 2	Monto adeudado en Equifax	Ninguna
Monto de compra a <i>Partners Technology</i> en 2019	Decimal	20	No	No	No	Entero 10, decimal 2	Monto de compra a <i>Partners Technology</i> en 2019	Ninguna
Días de línea de crédito	Decimal	3	No	No	No	Entero 10, decimal 2	Días de línea de crédito	Ninguna
Promedio de días de atraso	Decimal	3	No	No	No	Entero 10, decimal 2	Promedio de días de atraso	Ninguna

Anexo 8. Dashboard Sector Servicios

Servicios

732,19 mil

MONTO DE COMPRA A PARTNERS
EN EL 2019

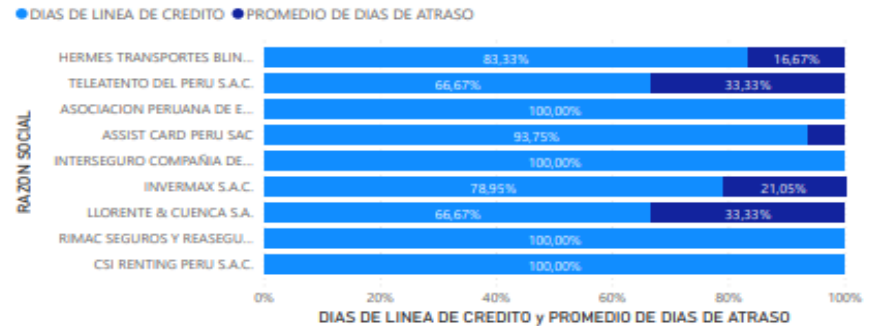
31,50

Promedio de DIAS DE LINEA DE
CREDITO

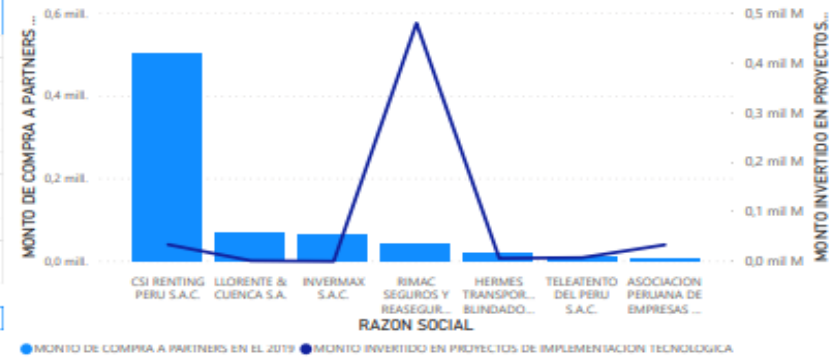
LAS 10 PRINCIPALES EMPRESAS DEL SECTOR DE TECNOLOGÍA

RAZON SOCIAL	FACTURACION DEL CLIENTE EN 2019	MONTO DE COMPRA A PARTNERS EN EL 2019
CSI RENTING PERU S.A.C.	680000000	502.560,21
LLORENTE & CUENCA S.A.	29922394	70.509,60
INVERMAX S.A.C.	5004326	65.512,72
RIMAC SEGUROS Y REASEGUROS	9605386292	43.852,87
HERMES TRANSPORTES BLINDADOS S A	125181868	18.295,73
TELEATENTO DEL PERU S.A.C.	153268781	12.123,47
ASOCIACION PERUANA DE EMPRESAS DE SEGUROS	670576990	5.564,03
INTERSEGURO COMPAÑIA DE SEGUROS S.A.	122979901	4.682,64
KIN ANALYTICS PERU SOCIEDAD ANONIMA CERRADA - KIN ANALYTICS PERU S.A.C.	4360000	4.594,92
ASSIST CARD PERU SAC	1800000	4.492,58
Total	11398480552	732.188,77

DIAS DE LINEA DE CREDITO y PROMEDIO DE DIAS DE ATRASO por RAZON SOCIAL



MONTO DE COMPRA A PARTNERS EN EL 2019 y MONTO INVERTIDO EN PROYECTOS DE IMPLEMENTACION TECNOLOGICA por RAZON SOCIAL



Anexo 9. Dashboard Sector Tecnología

Tecnología

2,47 mill.

MONTO DE COMPRA A PARTNERS EN EL 2019

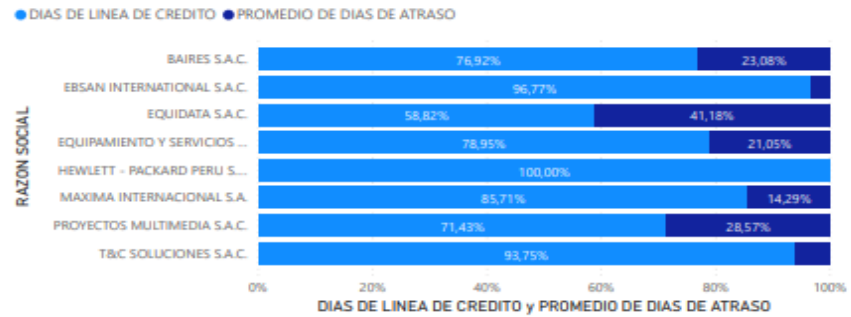
24,00

Promedio de DIAS DE LINEA DE CREDITO

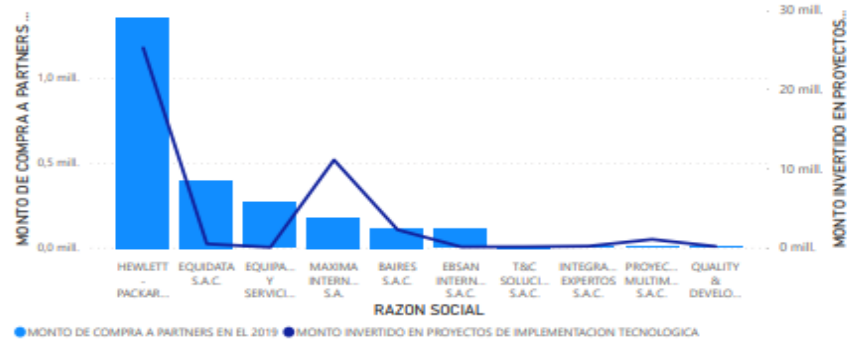
LAS 10 PRINCIPALES EMPRESAS DEL SECTOR DE TECNOLOGÍA

RAZON SOCIAL	FACTURACION DEL CLIENTE EN 2019	MONTO DE COMPRA A PARTNERS EN EL 2019
HEWLETT - PACKARD PERU S.R.L.	505554805	1.361.343,21
MAXIMA INTERNACIONAL S.A.	222580425	179.164,10
BAIRES S.A.C.	46090000	119.614,12
PROYECTOS MULTIMEDIA S.A.C.	21930000	6.452,92
EQUIDATA S.A.C.	10584000	400.858,56
INTEGRADORES EXPERTOS S.A.C.	48000000	8.960,09
QUALITY & DEVELOPMENT SOFTWARE S.A.C.	4189864	4.248,00
T&C SOLUCIONES S.A.C.	3299556	12.301,71
EBSAN INTERNACIONAL S.A.C.	29400000	110.433,80
EQUIPAMIENTO Y SERVICIOS TECNOLOGICOS S.A.C.	1963000	269.500,83
Total	823931650	2.472.877,34

DIAS DE LINEA DE CREDITO y PROMEDIO DE DIAS DE ATRASO por RAZON SOCIAL



MONTO DE COMPRA A PARTNERS EN EL 2019 y MONTO INVERTIDO EN PROYECTOS DE IMPLEMENTACION TECNOLÓGICA por RAZON SOCIAL



Anexo 10. Dashboard Sector Educación

Educación

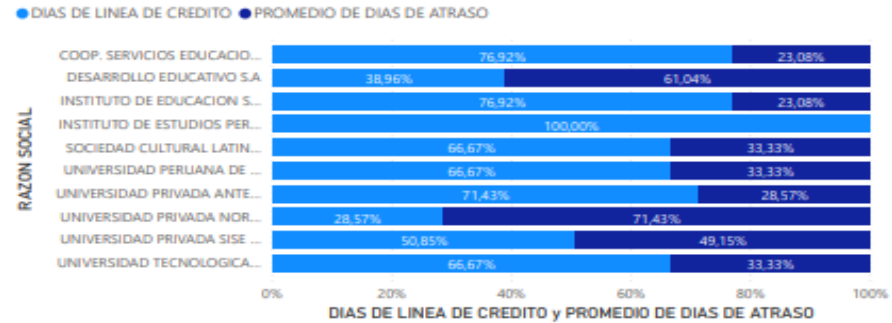
1,08 mill.

MONTO DE COMPRA A PARTNERS EN EL 2019

30,00

Promedio de DIAS DE LINEA DE CREDITO

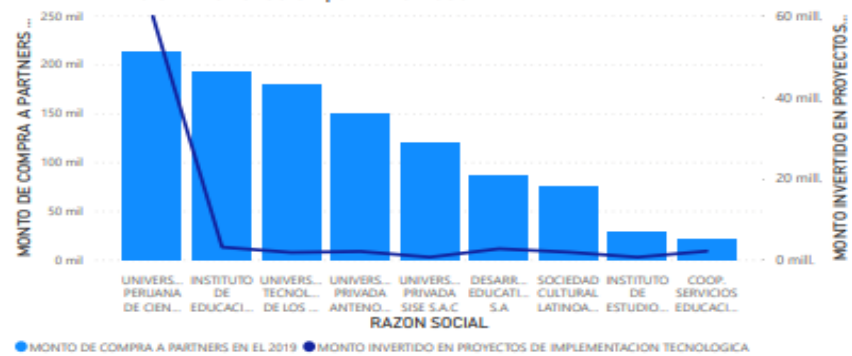
DIAS DE LINEA DE CREDITO y PROMEDIO DE DIAS DE ATRASO por RAZON SOCIAL



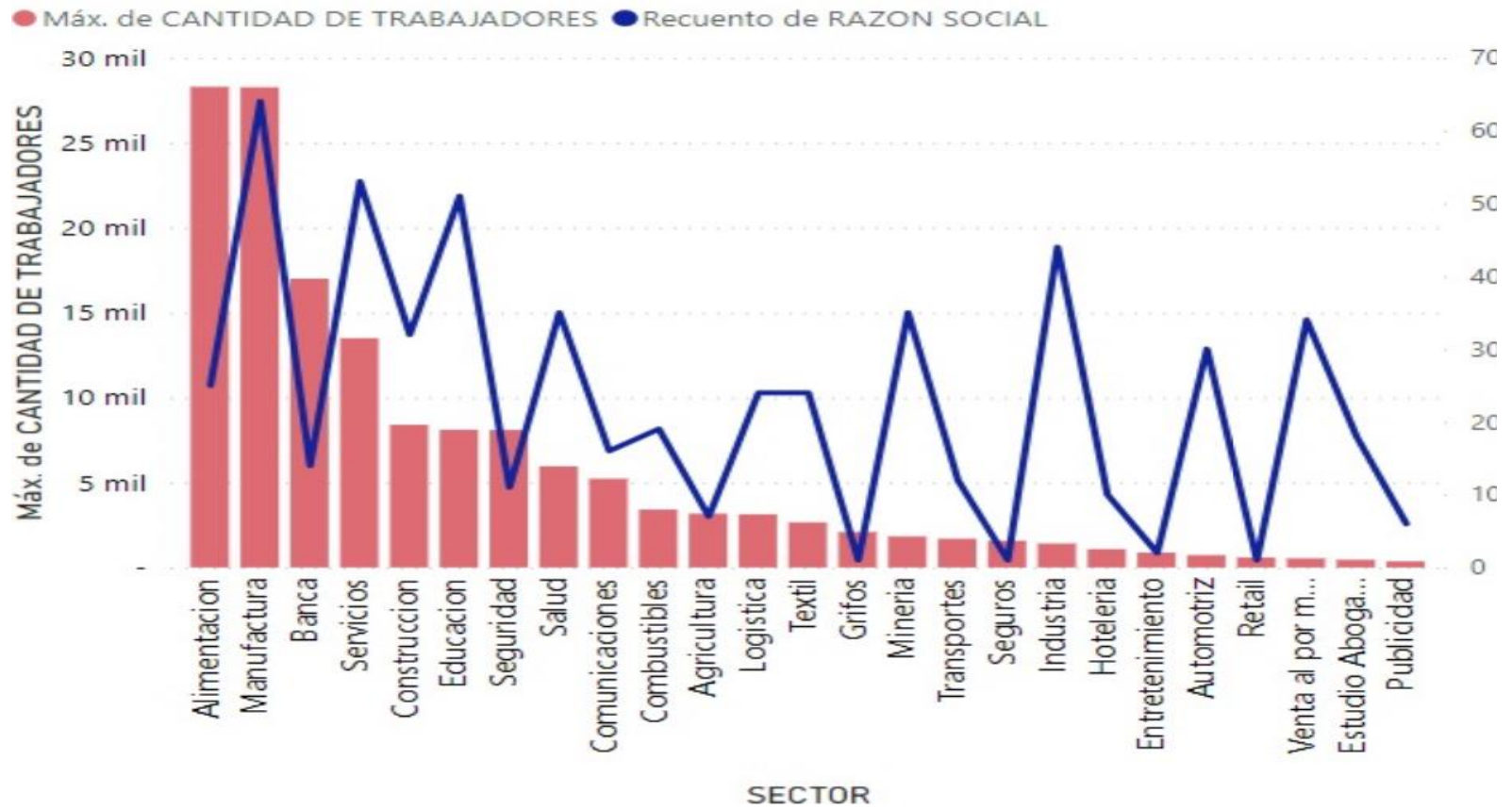
LAS 10 PRINCIPALES EMPRESAS DEL SECTOR DE TECNOLOGÍA

RAZON SOCIAL	FACTURACION DEL CLIENTE EN 2019	MONTO DE COMPRA A PARTNERS EN EL 2019
▲ COOP. SERVICIOS EDUCACIONALES ABRAHAM LI	45000000	22.353,92
DESARROLLO EDUCATIVO S.A	56510000	87.494,00
INSTITUTO DE EDUCACION SUPERIOR TECNOLÓGICO PRIVADO DANIEL ALCIDES CARRION SOCIEDAD ANONIMA CERRADA	64536000	193.274,56
INSTITUTO DE ESTUDIOS PERUANOS	15119971	28.401,42
SOCIEDAD CULTURAL LATINOAMERICANA S.A.C	40000000	75.437,40
UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS S.A.C	1200000000	213.815,71
UNIVERSIDAD PRIVADA ANTONOR ORREGO	43360000	150.476,24
UNIVERSIDAD PRIVADA NORBERT WIENER S.A	127254000	6.814,50
UNIVERSIDAD PRIVADA SISE S.A.C	15176000	119.995,09
UNIVERSIDAD TECNOLÓGICA DE LOS ANDES	38500000	178.902,90
Total	1645455971	1.076.965,74

MONTO DE COMPRA A PARTNERS EN EL 2019 y MONTO INVERTIDO EN PROYECTOS DE IMPLEMENTACION TECNOLÓGICA por RAZON SOCIAL



Anexo 11. *Dashboard* Cantidad de Trabajadores por Cantidad de Empresas por Sector



Anexo 12. Dashboard Monto de Inversión en Proyectos de Investigación Según Sector.