



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

**Metodología CRISP-DM para el análisis de asociación y  
Clustering para la diversificación de cursos de acuerdo al perfil  
de los clientes de la empresa ACCOM Perú SAC Piura, 2021**

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

**AUTORES:**

Br. Alburqueque Guaylupo, Hugo Enrique (ORCID: 0000-0003-0103-5099)

Br. Dongo Gutiérrez, Gian Carlos (ORCID: 0000-0002-1058-464X)

**ASESOR:**

Mgtr. More Valencia, Rubén Alexander (ORCID: 0000-0002-7496-3702)

**LÍNEA DE INVESTIGACIÓN:**

Sistema de Información y Comunicaciones

PIURA - PERÚ

2021

**Dedicatoria:**

El presente trabajo lo dedicamos a Dios, a nuestros padres por su amor y apoyo incondicional, a nuestras esposas que siempre estuvieron con nosotros y no nos dejaron retroceder y a nuestras hijas que son el motor principal para seguir adelante.

**Agradecimiento:**

En agradecimiento a los profesores que nos guiaron en todos estos años de aprendizaje y que hicieron posible la realización de este trabajo, a los amigos y familiares que nos apoyaron al logro de nuestra meta, la culminación de la carrera de Ingeniería de Sistemas.

## Índice de contenido

<b>Índice de Tablas</b> .....	v
<b>Índice de ilustraciones</b> .....	vi
<b>Resumen</b> .....	viii
<b>Abstract</b> .....	ix
<b>I. INTRODUCCIÓN</b> .....	1
<b>II. MARCO TEÓRICO</b> .....	4
<b>III. METODOLOGÍA</b> .....	18
3.1. Tipo y diseño de investigación .....	18
3.2. Variables y Operacionalización .....	18
3.3. Población, muestra y muestreo .....	21
3.4. Técnicas e instrumentos de recolección de datos .....	22
3.5. Procedimientos .....	23
3.6. Método de análisis de los datos .....	23
3.7. Aspectos éticos .....	24
<b>IV. RESULTADOS</b> .....	25
<b>V. DISCUSIÓN</b> .....	44
<b>VI. CONCLUSIONES</b> .....	49
<b>VII. RECOMENDACIONES</b> .....	50
<b>REFERENCIAS</b> .....	51

## Índice de Tablas

<b>Tabla 1</b> <i>Variables e indicadores</i> .....	18
<b>Tabla 2</b> <i>Descripción de indicadores</i> .....	19
<b>Tabla 3</b> <i>Población</i> .....	21
<b>Tabla 4</b> <i>Instrumentos de Metodología CRISP-DM</i> .....	22
<b>Tabla 5</b> <i>Instrumentos de diversificación de cursos</i> .....	23

## Índice de ilustraciones

Ilustración 1 Modelo Copo de Nieve.....	10
Ilustración 2 : Esquema de 4 niveles de CRISP-DM.....	12
Ilustración 3 Modelo de proceso CRISP-DM.....	13
Ilustración 4 Inscripciones por año.....	25
Ilustración 5 diagrama de la base datos.....	26
Ilustración 6 Tabla curso.....	26
Ilustración 7 Tablas de datos.....	27
Ilustración 8 Tablas de datos y tiempo.....	28
Ilustración 9 Modelado de la base datos.....	29
Ilustración 10 mapa de dimensiones.....	30
Ilustración 11 Variación de los datos por dimensiones.....	31
Ilustración 12 Mapa de clúster.....	31
Ilustración 13 Codificación en R.....	32
Ilustración 14 Codificación en R.....	32
Ilustración 15 Cauterización a través de K-means.....	33
Ilustración 16 Correspondencia por individuos (Meses del año).....	33
Ilustración 17 Definición de asociación y clúster.....	34
Ilustración 18 Mapa de Dimensiones.....	35
Ilustración 19 Mapa de Clúster.....	35
Ilustración 20 Variación de los datos por dimensiones.....	36
Ilustración 21 Codificación en R.....	36
Ilustración 22 Codificación en R.....	37
Ilustración 23 Cauterización a través de K-means.....	37
Ilustración 24 Correspondencia por individuos (Meses del año).....	38
Ilustración 25 Definición de asociación y clúster.....	38
Ilustración 26 Mapa de Cluster.....	39
Ilustración 27 Mapa de Dimensiones.....	40
Ilustración 28 Variación de los datos por dimensiones.....	40
Ilustración 29 Definición de asociación y clúster.....	41
Ilustración 30 Estrategia corporativa.....	42
Ilustración 31 Estrategia competitiva.....	42
Ilustración 32 Estrategia Funcional.....	43

.....

## Resumen

En las organizaciones, se vislumbra un crecimiento exponencial de grandes volúmenes de información, rebasando la capacidad de manejo, ante ello, con ayuda de la tecnología, se posibilita el encontrar información que no está al alcance a simple vista con sistemas transaccionales, por tanto es necesario la utilización de técnicas de búsqueda de información, que consigne patrones, relaciones significativas, tendencias o perfiles, que ayude en la toma de decisiones estratégicas en las organizaciones.

La investigación tiene como objetivo Aplicar la Metodología CRISP-DM para el análisis de asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes en la empresa ACCOM Perú SAC. La investigación es de tipo aplicada, diseño no experimental, la población estuvo compuesta por 5222 registros de captación de clientes y los estudiantes de los últimos ciclos.

En los resultados se definió una estrategia de análisis empresarial fundamentado en la metodología CRISP-D, describiendo las variables basado en la comprensión del negocio y de los datos seleccionando los atributos de un datawarehouse de 57 tablas; se preparó los datos considerando el perfil de los clientes, separándolos en un datamart conformada por 12 tablas con 52 atributos; el modelo los datos mediante la metodología copo de nieve generó una tabla de hechos con 5322 registros, obteniéndose 03 modelos tanto para la diversificación de cursos como para el cargo laboral del cliente. El primero conformado por 02 clúster con una correlación parcialmente negativa y el segundo con 04 clúster con una correlación parcialmente positiva y el tercero conformado por 03 clúster con una correlación parcialmente negativa, comprobada por la correlación dimensional. Finalmente, en la encuesta, el 44% están de acuerdo con las estrategias corporativas, el 50% están de acuerdo con las estrategias competitivas y finalmente el 39% están de acuerdo con la estrategia funcional; concluyendo que mediante la metodología CRISP-DM para el análisis de asociación y Clustering se realiza la diversificación de cursos de acuerdo al perfil de los clientes en la empresa ACCOM Perú SAC.

**Palabras clave:** Diversificación de cursos, perfil de los clientes con clustering, CRISP-DM diversificación de cursos

## **Abstract**

In organizations, an exponential growth of large volumes of information is envisaged, exceeding the management capacity, before this, with the help of technology, it is possible to find information that is not available to the naked eye with transactional systems, therefore it is necessary to use information search techniques, which record patterns, significant relationships, trends or profiles, which help in making strategic decisions in organizations.

The research aims to Apply the CRISP-DM Methodology for the analysis of association and Clustering for the diversification of courses according to the profile of the clients in the company ACCOM Peru SAC. The research is of an applicative type, non-experimental design, the population was made up of 5222 records of customer acquisition and students from the last cycles.

In the results, a business analysis strategy based on the CRISP-D methodology was defined, describing the variables based on the understanding of the business and the data, selecting the attributes of a datawarehouse of 57 tables; The data was prepared considering the profile of the clients, separating them into a datamart made up of 12 tables with 52 attributes; The data model using the snowflake methodology generated a table of facts with 5322 records, obtaining 02 models for both course diversification and the client's job title. The first made up of 02 clusters with a partially negative correlation and the second with 04 clusters with a partially positive correlation and the third made up of 03 clusters with a partially negative correlation, verified by dimensional correlation. Finally, in the survey, 44% agree with the corporate strategies, 50% agree with the competitive strategies and finally 39% agree with the functional strategy; concluding that by means of the CRISP-DM methodology for the analysis of association and Clustering, the diversification of courses is carried out according to the profile of the clients in the company ACCOM Peru SAC.

**Keywords:** Course diversification, client profile with clustering, CRISP-DM  
coursediversification



## I. INTRODUCCIÓN

En las últimas décadas, se vislumbra un crecimiento exponencial de grandes volúmenes de información, en las organizaciones se ha rebasado la capacidad de manejo, pues diariamente se generan millones de transacciones en las oficinas, centros comerciales, bancos, redes sociales y otros. Las personas al pasar su tarjeta, los mensajes de celulares, los mensajes de correo, la navegación por el Internet, los diversos hábitos, las preferencias; todo ello implica un registro constante de datos, y resultado de esto un cumulo de datos disperso como derrotero de información, historial de cada individuo integrado a esta sociedad, que dispersa cuan información se requiere para el desarrollo de cada actividad propia del ser humano (Acuña Gaitán, y otros, 2015).

Y ante esta vorágine de información acumulada en diferentes organizaciones, con el avance de la tecnología, se posibilita el encontrar información que no está al alcance a simple vista con sistemas transaccionales de manejo de datos, por lo se hace necesario la utilización de técnicas de búsqueda de información, que consigne patrones, relaciones significativas, tendencias o perfiles en grandes volúmenes de datos, apoyado con el uso de herramientas informáticas que nos conducen al uso de técnicas de minería de datos, reglas de asociación, clustering para el análisis de datos y que sea el insumo para la toma de decisiones estratégicas en las organizaciones (Apolaya Torres, y otros, 2018).

En este contexto, se han dado muchas aplicaciones mediante la minería de datos, con el fin de solucionar, detectar patrones de fraudes en instituciones financieras, detección de placas falsas, y otras soluciones (Arteaga Tigua, 2017). En Ecuador, tenemos su aplicabilidad en meteorología, en elecciones presidenciales, educación, donde sin duda se denota el desfase que se tiene entre la capacidad para almacenar datos que ha crecido exponencialmente pero no ha sucedido lo mismo con la capacidad de su procesamiento (Contreras Ortiz, y otros, 2018).

Por ello la minería de datos se ha convertido en una alternativa como soporte de los procesos, generando información para la toma de decisiones. Así en México, mediante técnicas de minería de datos, se crearon reglas de tendencia de precipitaciones pluviales, mediante técnicas como K-means y arboles de decisión C4.5 (Vázquez Zavaleta, 2018). Asimismo, en Colombia se han creado prototipos

de información crítica, utilizando la metodología CRISP-DM, construyendo modelos basado en clustering y en arboles de decisión (Estévez Cárdenas, 2016).

En el Perú, mediante la clasificación de diferentes técnicas de minería de datos, se propone un modelo en el campo educativo de nivel superior, con el propósito de mejorar la toma de decisiones en la escuela de ingeniería de la UPC (Apolaya Torres y Espinosa Diaz, 2018). Por ello ante su aplicabilidad en las diferentes disciplinas del quehacer humano se plantea contextualizar en empresas del rubro de capacitación, que ofrecen una gama de cursos virtuales ante este entorno de pandemia que estamos atravesando a nivel mundial.

La empresa ACCOM Perú S.A.C. se dedica a la capacitación de profesionales y empresas en la ciudad de Piura, con más de 11 años de experiencia, se ha convertido en una de las empresas líderes en el rubro de la capacitación. En todo este tiempo se han venido dictando cursos de capacitación y diplomados de especialización tales como, Excel, AutoCAD, AutoCAD Civil 3D, Metrados en Edificaciones, Diseño Gráfico, S10 Costos y Presupuestos, Seguridad Industrial y Salud en el Trabajo, Supervisor HSEQ y otros. La empresa ha vendido sus cursos durante estos 11 años en base al uso de redes sociales, en especial de la plataforma FACEBOOK, el uso de correo electrónico y llamadas, esto le ha permitido por cierto tiempo tener éxito al momento de vender sus cursos, pero ya hace un tiempo se viene observando una disminución en las inscripciones. La empresa solo cuenta con la experiencia de marketing obtenida de años anteriores de forma empírica, más no cuenta con estrategias de marketing bien definidas y documentadas, esto genera que tenga que recurrir constantemente al uso de las redes sociales de pago para obtener resultados positivos. Debido a la coyuntura de la COVID-19, se ha hecho más evidente la disminución gradual de la cantidad de inscripciones en cursos que anteriormente se inscribían hasta 20 participantes, ahora no llegan ni a las 10 participantes, asimismo en la cantidad de cursos en desarrollo, en 1 mes se podían realizar un total de 10 cursos de capacitación, en la actualidad, solo se logran desarrollar 4, con metas menores a las 10 inscripciones de participantes. Esto se debe a que, en la modalidad virtual, existe una diversidad de cursos con la misma temática, e incluso gratis, lo que determina que se escoja por otras opciones para su aprendizaje o actualización; además la modalidad de enseñanza virtual que propone la empresa no es tan atractiva en este mercado.

Esta situación se hace más claro la necesidad de plantear estrategias para poder superar la situación actual que está perjudicando severamente a la empresa.

Ante esta realidad se plantea el siguiente problema de investigación: ¿Cómo se aplica el análisis de asociación y clustering mediante la Metodología CRISP-DM para la diversificación de cursos de acuerdo al perfil de los clientes en la empresa ACCOM Perú SAC? y como problemas específicos: ¿Qué comprensión del negocio y los datos se podrá definir en el análisis de asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes?, ¿Como se prepara los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes?, ¿Como se modela los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes? y ¿Cómo es la evaluación de los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes?.

La investigación se justifica metodológicamente, debido a que se aplicará una metodología para realizar el análisis de asociación y clustering, para obtener una solución, que ayude a diversificarlos nuevos cursos considerando el perfil de los clientes en la empresa ACCOM Perú SAC. Asimismo, se justifica institucionalmente, debido al aporte que implica para la empresa ACCOM Perú SAC, la caracterización del perfil del cliente que permite la diversificación de los cursos existentes o la posibilidad de plantear nuevos en función de los hallazgos de la metodología aplicada. Y en cuanto a su justificación tecnológica, la aplicación de nuevas tecnologías conllevará sin lugar a dudas a plantear novedosas soluciones, nuevos modelos, haciendo uso para ello de tecnología emergente.

Ante ello se plantea el objetivo de la investigación: Aplicar la Metodología CRISP-DM para el análisis de asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes en la empresa ACCOM Perú SAC y como objetivos específicos: Definir la comprensión del negocio y de los datos en el análisis de asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes, preparar los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes, modelar los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes y evaluar los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes.

## II. MARCO TEÓRICO

Se ha considerado estudios de diferentes investigadores a fin a la temática de la investigación, en un contexto internacional, nacional y local.

En el plano internacional De Lucca (2017), presento la investigación “Data mining y turismo: Un enfoque integral para potenciar la imagen de marca Argentina”, cuyo objetivo general fue el analizar cuan útil es el uso de la minería de datos como herramienta inteligente para llevar a un nivel más alto la imagen de la marca. El Tipo de investigación fue aplicada y el diseño preexperimental, la población está compuesta por 2.074.703 personas, que se constituye en su público potencial, cuya muestra fue de 105 divulgaciones de turistas en el mes de setiembre del 2016. Entre los resultados, se clasifico en atributos y valores posibles, asimismo relaciones entre sustantivos, adjetivos y ciudades. El algoritmo J48 clasificó a cada una de las ciudades con adjetivos a partir de sus nombres y recursos tanto culturales como naturales. Entre sus conclusiones se tiene que pueden plantearse aplicaciones como marketing direccionado, soporte para la toma de decisiones de alto nivel, articulación con otras bases de datos y adecuación de circuitos turísticos en función del perfil del visitante extranjero.

Walid Ghobar (2017), realizó un sistema de recomendación basado en perfiles generados a través de asociación y agrupamiento, cuyo objetivo es crear un sistema capaz de realizar sugerencias basadas en el perfil de los usuarios. El Tipo de investigación es aplicada y el diseño preexperimental, teniendo como dato la población y la muestra un total de 17632 artistas y 1892 usuarios. En los resultados se puede obtiene que cuando K es igual a 6, emerge un subgrupo con 3 elementos, en el “Perfil 2”, el agrupamiento por genero de los artistas ha sido probado con k desde 1 hasta 10 con k-means, y los resultados indicaron que la  $k = 8$  es un valor suficiente para la realización del experimento. En el perfil principal la mayoría de recomendaciones ha sido de 03 items o más de casi 85% de los casos.y en el perfil secundario a traves de subgrupos de 29 casos.

Casariego Sarasquete (2019), realizó una investigación denominada Metodología de análisis y segmentación de clientes usando secuencias de comportamiento, cuyo objetivo general es aplicar la metodología en diversas industrias y sectores de

negocio para la segmentación y análisis de los clientes, pacientes y ciudadanos. Cuyo tipo de investigación fue aplicada y el diseño cuasi experimental, la población está compuesta por las transacciones de 12 meses en una tienda de moda y en un servicio sanitario ambas ficticias. En los resultados se tiene que la nueva forma de agrupación propuesta no sustituye a la anterior sino la complementa, debido a que descubre la existencia de micro segmentos dentro de los anteriores sobre los que se podría accionar considerándolo desde un punto de negocio, además de los pilotos realizados se consideró los algoritmos de visualización de secuencias y patrones los que mejor se adaptaron al contexto.

Sigüenza Canale (2018), realizó una investigación denominada Propuesta de una estrategia de diversificación de servicios para la microempresa pequeños genios, cuyo objetivo general fue elaborar una estrategia para diversificar los servicios y que a la vez permita incrementar los ingresos. Cuyo tipo de investigación es no aplicada y el diseño no experimental, la población y muestra estuvo conformada por los clientes que forman parte de la Empresa "Pequeños Genios", quienes consumen los servicios educativos. Los resultados demuestran en base a los ingresos de la empresa la necesidad de implementar estrategias de diversificación de servicios en la Microempresa "Pequeños Genios" de la ciudad de Milagro, debido que a partir del 2015 el decremento financiero se ve reflejado en un -20% en base al 2016; mientras que en el 2017 el porcentaje fue de -30%. La microempresa, proporcionara varios y mejores servicios, que logrará un mayor volumen de ventas, ya que les servirá para ampliar la cartera de clientes, el incremento de sus ingresos y utilidades, porque la percepción de los encuestados (padres de familia, estudiantes) determinan una necesidad de diversificar los servicios de la microempresa.

Arévalo Rodríguez, y otros (2016), realizó una investigación mediante minería de datos para la identificación de patrones de consumo de servicios turísticos para orientar la oferta en el sector hotelero, cuyo objetivo general fue identificar patrones de consumo de servicios turístico para el sector hotelero de la ciudad de Trujillo. El Tipo de investigación es aplicada y el diseño preexperimental, la población está compuesta Información de las Fanpage de las empresas hoteleras en la ciudad de Trujillo, y la muestra Fanpage del hotel Kallpa Hotel. Los resultados de la

investigación, se comparó diversas técnicas de minería de datos siendo arboles de decisión y clustering las más utilizadas. Además, mediante la técnica de clustering se obtuvo 03 clúster, identificándose características en cada uno de ellos, como servicio de alojamiento, turismo, festividad y restaurant. Además, en el cluster0 el servicio de alojamiento es de mayor consumo en los meses de febrero y agosto y el mes más bajo en enero. En sus conclusiones se tuvo que el patrón de comportamiento de los clientes tiende a incrementar a partir de la segunda mitad de año.

Riquelme Laferte (2017), realizó una investigación para la mejora del proceso de evaluación de nuevos convenios mediante modelos de predicción, cuyo objetivo general fue desarrollar un apoyo que permita predecir la rentabilidad de nuevos convenios de pago de remuneraciones mediante el uso de modelos de predicción. El Tipo de investigación es aplicada y el diseño preexperimental, la población y muestra estratificada de un 25% de los datos del universo total, se efectúa una evaluación del modelo final. La cantidad total de registros es de 1.408, de los cuales 1091 son convenios con rentabilidad positiva y 317 con rentabilidad negativa. Los resultados fueron que con el modelo permite una disminución de pérdida, con los casos revisados de un 19,66%. La disminución de la rentabilidad positiva es de un 0,83%. Se produce una pérdida de rentabilidad de 0,0436%. En sus conclusiones se tuvo que los modelos desarrollados permitirán generar un mayor conocimiento de los convenios y como es su comportamiento en el tiempo. Además de permitir una mejor toma de decisiones al disponer de información enfocada en el rubro del convenio y de los clientes que lo componen, disminuyendo la incertidumbre a la hora de evaluar.

Laura Ochoa (2016) realizó un estudio comparativo de técnicas no supervisadas de minería de datos, como objetivo fue la realización de un estudio, haciendo uso de técnicas no supervisadas para segmentación de los alumnos. El Tipo de investigación es aplicada y el diseño preexperimental, la población y muestra está compuesta por el tercer, cuarto y quinto semestre 2015 e impar de 2016. En los resultados se tuvo en consideración las distancias intra-cluster e inter-cluster y el coeficiente de silueta. Las distancias intra-cluster del método jerárquico aglomerativo utilizadas fueron Ward, Single, Complete, Average, Mcquitty, median,

centroid; representando mayor cohesión el método de agregación de Ward. En el caso de las distancias intra-cluster e inter-cluster, se utilizó el algoritmo de agrupación K-Means y PAM, obteniéndose mayor cohesión y similitud con el K-Means. Con en el algoritmo silueta, se midió la calidad de los cluster encontrados. En sus conclusiones se realizó el estudio comparativo de agrupaciones correspondientes a la siguiente escalara valorativa: BAJO, MEDIO y ALTO, obteniendo grupos de mejor calidad, con menor distancia intra-cluster y mayor distancia inter-cluster, que representa la mayor homogeneidad dentro del grupo.

Mendoza Bernedo, y otros (2018), realizó una investigación para la determinación de patrones de ventas en boticas independientes para mejorar las ventas, cuyo objetivo general fue establecer los patrones de ventas para Boticas Amberfarma E.I.R.L. El Tipo de investigación es aplicada y el diseño preexperimental, la población es de 465240, formada por los registros de ventas de 5 sucursales que pertenecen a la empresa entre los periodos de enero a diciembre del año 2017, con una muestra de 384 registros. En los resultados se tiene que Chorrillos tiene doce patrones identificados entre análisis exploratorio y reglas de asociación. Como conclusiones se tiene que la determinación de estos patrones influye de manera positivamente en las ventas, permitiendo la elaboración de estrategias comerciales en la dispensación y compra de productos en base a la identificación de patrones que permitirán aumentar la rentabilidad.

Rioja Curo (2020) realizó en su investigación una aplicación web para la elaboración de perfiles de consumidor basada en minería de datos y arquitectura cloud para el apoyo al proceso de conversión de Leads en la Asociación AIESEC en Perú, cuyo objetivo fue elaborar perfiles del consumidor basado en minería de datos. El Tipo de investigación es aplicada y el diseño cuantitativo preexperimental, la población1 estuvo conformada por 9245 leads, población2 409 asociados y muestra de 369 individuos y muestra 282 personas. Entre los resultados se tiene que se hizo una subclasificación en base a perfiles seleccionados previamente mediante un análisis combinatorio, validando la exactitud de 06 clústeres, corroborando el índice de Davies–Bouldin. En el listado de perfiles se contiene 20 perfiles de consumidor. Además, se reduce el tiempo de respuesta de preprueba y posprueba de 24 horas a menos de 10 segundos; el grado de aceptación de

preprueba y posprueba de la información por los usuarios de 6 a 8.5; además el grado de aceptación de preprueba y posprueba de la información brindada al consumidor de 3 a 8.1. Como conclusiones se plasma la ejecución de la aplicación web basada en la segmentación por los perfiles de consumidor, así también generando una mejora en el proceso de conversión de leads. Asimismo, se redujo el tiempo de respuesta de los leads al mínimo, generando una ventaja en función al tiempo empleado para la elaboración del programa y paquete que se ajuste a las necesidades del consumidor.

Contreras Arteaga, y Otros (2019), realizó la investigación sobre Analítica predictiva para conocer el patrón de consumo de los clientes utilizando IBM Spss Modeler y la metodología CRISP-DM, cuyo objetivo general fue de realizar una analítica predictiva para determinar patrones de consumo de los clientes en la empresa Cienpharma S.A.C. utilizando técnicas de minería de datos. El Tipo de investigación es aplicada y el diseño preexperimental, la población son los registros de las ventas de la base de datos transaccional de la empresa y como muestra registros de las ventas de los años 2014-2015 de la empresa. En los resultados se tiene que el algoritmo k-medias con 07 entradas produce 05 clústeres, siendo los tamaños mínimos de 1526 y máximos de 9960, pero mediante el algoritmo kohonen con 07 entradas produce 09 clústeres con tamaños mínimos de 318 y máximos de 4357. Se concluye que se logran construir 03 modelos de minería para evaluar y entrenar los datos: K-medias, Kohonen y Árbol de decisión (C5.0), obteniendo por modelo diferente cantidad de clústeres, de acuerdo a las entradas de datos evaluada, todo este proceso se realizó utilizando IBM SPSS Modeler.

Arcos Méndez (2020), realizó mediante minería de datos una aplicación para pronóstico de ventas, cuyo objetivo general fue implementar un modelo de minería de datos para el pronóstico de ventas, que cumpla con las necesidades de las empresas para reducir el tiempo y margen de error en el análisis de datos con fines predictivos. El Tipo de investigación es aplicada y el diseño cuantitativo preexperimental, la población y muestra fue 04 datasets con 115064 registros. Entre sus resultados se utilizó 04 algoritmos de aprendizaje, haciendo uso de la validación cruzada K-Fold, para obtener un pronostico de ventas con menos tasa de error, se obtuvo 04 datasets con 115064 registros; la tasa de error más baja y la



mejor puntuación en ejecución se obtuvo con el algoritmo ETR (Extra Trees Regressor). Entre sus conclusiones el análisis de datos para pronósticos de ventas permitió optimizar el tiempo y margen de error humano.

Rojas Gómez (2020) realizó la investigación sobre patrones de comportamiento de los usuarios de la empresa Adonai Employment utilizando minería de datos, cuyo objetivo general fue caracterizar los patrones de comportamiento de los usuarios de la empresa Adonai Employment extraídos de un log del servidor por medio de técnicas de minería de datos. El Tipo de investigación es aplicada, nivel descriptivo y diseño cuantitativo preexperimental, la población y muestra fue 3387 registros. En los resultados de los datos alojados en el sitio en el log del servidor de la empresa, se determinaron 11 reglas de asociación, con grados de confianza de 100%. Además los días martes, miércoles y jueves se presentaron la mayor cantidad de peticiones, acumulando un flujo de 52.32%. En sus conclusiones se tiene que por medio del descubrimiento de conocimientos en bases de datos, obtenidos del log del servidor de Adonai Employment, se logró identificar patrones útiles para la organización, a partir de los datos de navegación dejados en el log del servidor de una manera automática utilizando grandes cantidades de datos.

Herrera Meza (2017), realizó la investigación relacionada con la preferencia de los turistas extranjeros Millennials por AIRBNB, cuyo objetivo general fue determinar los factores vinculados al hecho que los turistas extranjeros millennials prefieren Airbnb en Lima, Perú en 2017. El Tipo de investigación es aplicada, nivel descriptivo y diseño cuantitativo preexperimental, la población fueron turistas extranjeros millennials que contrataron servicios de hospedajes a través del portal Airbnb en Lima en el año 2017 y cuya muestra fue 384 comentarios. Según los resultados del software utilizado, el Voyant Tool, se extraen las palabras asociadas a las palabras claves con mayor valoración y que más se repiten en los comentarios de los turistas. La conclusión de esta investigación es que se puede inferir que las preferencias por Airbnb guardan relación dado que son alojamientos de calidad, por contar con espacios según la necesidad particular de cada turista extranjero millennial y en donde se sienten tratados con amabilidad y educación por sus anfitriones. Por otro lado, en otros servicios como hoteles donde los procesos son más estandarizados no ocurre lo mismo.

Para lograr el entendimiento, es importante teorizar aquellos términos que están en relación con la temática de la investigación, respecto a la variable Metodología CRISP-DM para el análisis de asociación y clustering:

Minería de datos, se refiere a un conjunto de técnicas y herramientas que se aplican a un proceso no trivial de extraer y presentar conocimiento implícito, previamente desconocido, potencialmente útil y comprensible para los humanos, a partir de grandes conjuntos de datos, con la finalidad de predecir de forma automatizada tendencias y comportamientos (Galarreta Vásquez, 2016). En este proceso de descubrimiento de patrones inmersos en estos conjuntos de datos, se utiliza técnicas de análisis de datos que se basan en algoritmos y dependiendo de este, será la calidad del conocimiento obtenido Betancur Ospina (2017).

La Inteligencia de Negocios BI (Business Intelligence) es una herramienta, que le permite a diferentes tipos de organizaciones, que puedan soportar la toma de decisiones basadas en información precisa y oportuna; garantizando la generación del conocimiento necesario que permita escoger la alternativa que sea más conveniente para el éxito de la empresa (Rouhiainen, 2021)

En los modelados de inteligencia de negocios, el esquema de copo de nieve es una variante del modelo estrella, con la diferencia de que este modelo está conformado de una tabla de hechos que está conectada a diversas tablas de dimensiones, que pueden estar a su vez conectadas con otras tablas de dimensiones. Las tablas de dimensiones describen los componentes o entidades que se van a modelar en la organización. El modelo copo de nieve tiene la ventaja de que las tablas de las dimensiones están normalizadas, de esta manera se evita redundancia y con ello se ahorra espacio.

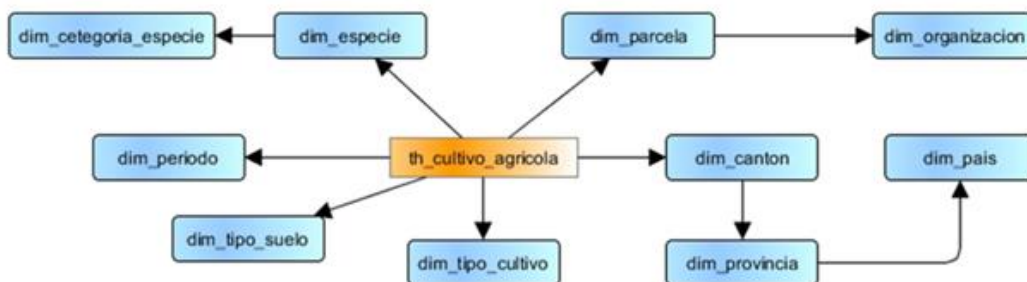


Ilustración 1 Modelo Copo de Nieve

Asimismo, este tipo de modelo ofrece la posibilidad de segregar los datos de las tablas de dimensiones y proveer un esquema que sustente los requerimientos de diseño (Goleman, 2018).

Entre las principales técnicas, se clasifican según su tarea de descubrimiento en: Agrupación o clustering, Clasificación, Asociación y otros. Clasificación: Consiste en clasificar los objetos, entre un rango de categorías. Un ejemplo son los llamados Mapas Asociativos de Memoria o Mapas Autoorganizados, entre los que se encuentran las redes de Kohonen, redes neuronales que clasifican objetos mediante aprendizaje por refuerzo. Dentro de este apartado también se encuentran los algoritmos que generan reglas de clasificación para los objetos (Schiatti Sisó, 2017). En el Clustering se agrupan los objetos con el mayor parentesco o similitud de características, estableciendo clases. La diferencia con la clasificación es que no hay una previa división del espacio en que se encuentran las categorías; otra diferencia es que los algoritmos de clasificación no agrupan, sino que los clasifican de uno a uno. Existen varios tipos de clustering, según la metodología empleada y muchos algoritmos de cada tipo (Puga Reyes, 2020). Asociación: Cuyo único objetivo es descubrir las relaciones tácitas que existen entre los objetos, o incluso entre los propios atributos de estos últimos, de los cuales se puede extraer una base de reglas, con estructura condicional (Si A es B y C es D, Entonces E es F) (Vazquez Zavaleta, 2018); Dependency modeling: Describen relaciones significativas de dependencia entre variables. Se establecen dos tipos de dependencias: la estructural y la cuantitativa. La dependencia estructural especifica las variables que son localmente dependientes. La dependencia cuantitativa da una medida numérica de como de fuertes son las dependencias (Venkat Ram, 2017); Predicción: Obtención (o entrenamiento) de modelos de predicción, con el fin de validar hipótesis de comportamiento preconcebidas (Satria, y otros, 2018). Regresión: Se toma a partir de modelos de datos, con la finalidad de evaluar una función o modelo que pueda generar una dependencia tanto positiva como negativa de ciertas variables respecto de otras, con el fin de poder predecir valores a partir de nuevos datos (Quispe Durán, 2018). Sumarización: Tiene como objetivo generar descripciones globales de conjuntos de datos. En algunos casos estas descripciones son cualitativas. Se usa, por ejemplo, para la extracción de información en textos (Arteaga Navarrete, y otros, 2018).

Sinaga, y otros (2020), indica que el Clustering consiste en la división de datos en grupos de objetos similares llamados Clúster. Dichos objetos son agrupados basándose en el principio de maximización de similitud dentro de los clúster y minimización de similitud entre clusters diferentes. No existe conocimiento previo a cerca de como deben conformarse los grupos, es por ese motivo que al clustering tambien se le considera como un técnica de aprendizaje no supervisado.

Lago Portela (2020), el método de las k-medias, es un método que permite asignar a cada observación el cluster que se encuentra más próximo en términos del centroide (media). En general, la distancia empleada es la euclídea. Se considera los siguientes pasos: 1. Se toman al azar k clúster iniciales. 2. Para el conjunto de observaciones, se vuelve a calcular las distancias a los centroides de los clusters y se resignan a los que estén más próximos. Se vuelven a recalcular los centroides de los k clusters después de las reasignaciones de los elementos. 3. Hasta que no se produzca ninguna reasignación, es decir, hasta que los elementos se estabilicen en algún grupo, se repiten los dos pasos anteriores. Usualmente, se especifican k centroides iniciales y se procede al paso (2) y, en la práctica, se observan la mayor parte de reasignaciones en las primeras iteraciones.

Respecto a la metodología CRISP-DM Según SNGULAR (2021) CRISP-DM, está dividida en 4 niveles de abstracción organizados de forma jerárquica (Figura 1) en tareas que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo de un proyecto de Data Mining (DM), en una serie de seis fases.

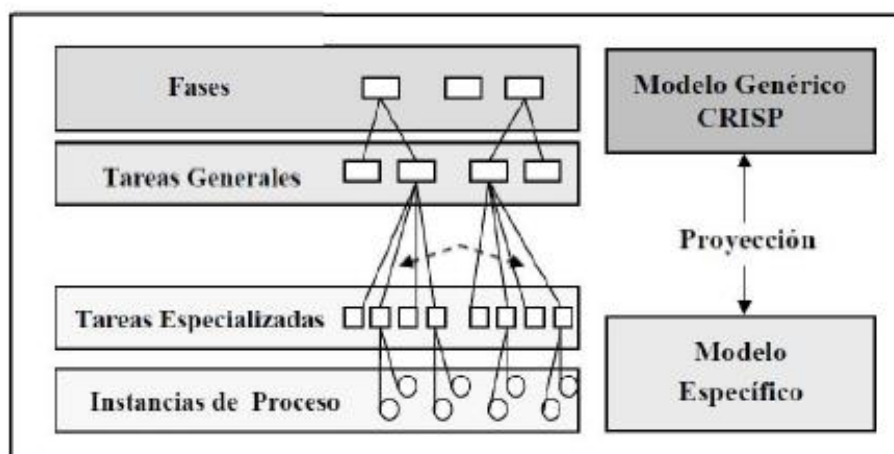


Ilustración 2 : Esquema de 4 niveles de CRISP-DM

Fuente: SNGULAR (2021)

a sucesión de fases no es necesariamente rígida. Cada fase es estructurada en varias tareas generales de segundo nivel. Las tareas generales se proyectan a tareas específicas, donde finalmente se describen las acciones que deben ser desarrolladas para situaciones específicas, pero en ningún momento se propone como realizarlas.

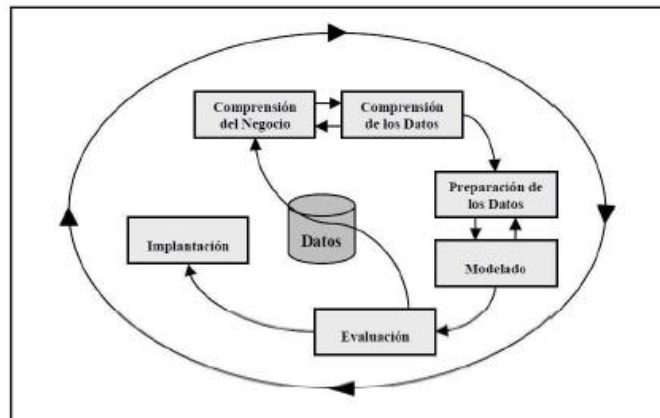


Ilustración 3 Modelo de proceso CRISP-DM

Según (Angarita García, y otros, 2016), se divide en 6 fases: Primera fase; Trata de la comprensión del negocio en sí, sus características para conocer sus fortalezas y carencias. Segunda fase; Comprensión de los datos, no es más que la comprensión de cómo están constituidos los datos en cada uno de los registros ya sean campos, columnas, filas, etc. Tercera fase; La preparación de los datos, ya se nota la diferencia y el camino a seguir para la estandarización de los datos haciendo uso de técnicas tales como la uniformidad de datos, la limpieza de datos o también llamado datacleaning, eliminación de redundancia, espacios en blanco, etc. Cuarta fase; El Modelamiento, en esta fase va tomando sentido lo expuesto desde la primera fase del proyecto, se hace uso de herramientas relacionadas a la minería de datos. Quinta fase; La Evaluación, no es más que la comparación de resultados obtenidos con los objetivos para analizar si conciden o no, o recomendaciones a realizar a futuro. Sexta fase; El Despliegue, se obtienen 2 tipos de resultados aplicando a minería de datos, estos son los modelos finales y las conclusiones de los modelos Fuente: SNGULAR (2021).

En el análisis de componentes principales (PC), un valor propio mayor a 1 indica que los PC representan más varianza que la contabilizada por una de las variables originales en los datos estandarizados. Esto se usa comúnmente como un punto

de corte para el cual se retienen las PC. Esto es cierto solo cuando los datos están estandarizados. También se puede limitar el número de componentes a ese número que representa una cierta fracción de la varianza total. Por ejemplo, si el análisis quiere el 80% de la varianza total explicada, se puede utilizar la cantidad de componentes para lograrlo. Un método alternativo para determinar el número de componentes principales consiste en observar un gráfico Scree, que es el gráfico de valores propios ordenados de mayor a menor. El número de componentes se determina en el punto, más allá del cual los valores propios restantes son todos relativamente pequeños y de tamaño comparable (Jolliffe, y otros, 2020).

La correlación entre una variable (Tipo de Curso) y un componente principal (PC) se utiliza como coordenadas de la variable en la PC. La representación de las variables difiere del gráfico de las observaciones: las observaciones están representadas por sus proyecciones, pero las variables están representadas por sus correlaciones (Abdi, y otros, 2010).

Con respecto a la teorización de la variable diversificación de cursos se tiene:

Entre las estrategias de ventas según Yturregui Valdivia (2018) que refiere Albrecht (1994) es mejorar la calidad de servicio, ante los cambios de los tiempos, por ellos las relaciones llegan a ser más importantes que los productos físicos. También se considera una práctica que tiene como objetivo, elevar los niveles de satisfacción de los clientes. El bajar precios, se utiliza para incrementar las ventas, pero reduce el margen de ganancia y podría implicar una reducción de la calidad del servicio. El mejorar el servicio al cliente, considera que conseguir un cliente es más difícil y sale más caro que retener un cliente antiguo. Los clientes esperan un buen trato, amable, personalizado y con esmero; por ello es necesario para la organización conocer las expectativas de los clientes. La segmentación del mercado, es la forma como la organización decide agrupar a los clientes más importantes conociendo sus necesidades y preferencias para lograr una ventaja competitiva. Para una adecuada segmentación se debe llegar a los clientes de una manera óptima y accesible con toda la mezcla de mercadotecnia y debe poseer segmentos que marquen una diferencia, esto permite tener una estrategia más adecuada al ajustar precios, productos y programas de acorde a las necesidades de los programas. El trabajar con referidos de bases de datos, es más que poseer un nombre, estos

datos en conjunto forman un contacto. El incrementar las promociones, diseñadas para estimular al consumidor; tales como descuentos por convenios, descuentos corporativos, descuentos por puntualidad y otros. Finalmente, el Perfeccionar el Merchandising, el cual son acciones que se llevan a cabo para presentar de forma adecuada a los productos, gestionando su rotación e incrementando su rentabilidad. Se distinguen 03 tipos: de presentación, de seducción y de gestión. Es fundamental para su funcionamiento contar con un portafolio de productos promocionales que se adecuen a los diferentes tipos de clientes.

La diversificación, según INCP (2021), asegura que la diversificación es una de las estrategias más implementadas en las organizaciones, cuando se trata de ampliar los ámbitos del negocio, por tanto, se pueden clasificar en estrategias relacionadas y no relacionadas. Las relacionadas son las que poseen relación con productos o servicios existentes en la organización, aprovechando sus recursos. El éxito de los nuevos servicios; el resultado y el proceso de un servicio suelen combinarse para crear la experiencia y los beneficios de los clientes, y se deben considerar ambos aspectos en la creación de nuevos servicios considerando la información que se posee. En el caso de las no relacionadas, son nuevas actividades por las que apuestan las empresas son completamente nuevas para su filosofía y no guardan ninguna relación con los productos o servicios que ha ofrecido hasta el momento.

Para la diversificación de cursos se debe considerar la estrategia corporativa, estrategia competitiva y estrategia funcional. Según Cevera Negueruela (2019), la estrategia corporativa va ligada a la creación de valor a la organización, define los negocios o unidades de negocio en los que se va operar, así como los recursos que se asignarán. Por ello coordina actividades, recursos humanos y económicos para la satisfacción de propietarios, stakeholders para alcanzar ventajas competitivas, considerando los siguientes elementos: Crecimiento, como globalización o diversificación de negocios o integración vertical a lo largo de la cadena de valor. Asimismo, se puede considerar adquisiciones, fusiones o alianzas estratégicas mediante el desarrollo interno o también denominado crecimiento orgánico (Rodríguez García, 2018). Diferenciación, se basa en la oferta de servicios específicos o exclusivos para un segmento del público (Turiano, 2019). Entorno, son relaciones con lo que rodea a la organización, lo que podría considerarse

estrategias relacionales, tomando en cuenta las relaciones con los stakeholders. Planeamiento: en este modelo el o los directivos aportan su experiencia y conocimientos especializados a los distintos negocios, de esta forma se promueve la coordinación e interrelación entre negocios creando compromiso compartido por toda la empresa y se puede apuntar a la consecución de objetivos a largo plazo (Bernal Payares, 2018)

Según Morales Ospina (2019), la estrategia competitiva, consiste en crear un valor para el cliente que la competencia no ofrece, explotando sus fortalezas desarrollando nuevas capacidades aprovechando las oportunidades; estos valores pueden estar ligados al precio o podría estar ligado a la innovación como métodos de aprendizaje más rápidos y efectivos para mejorar su posición competitiva respecto a las empresas del mismo rubro. Se consideran los siguientes elementos: Imagen de marca, es una gama de elementos tangibles e intangibles que representa los valores que se desean transmitir a los clientes. Estos elementos pueden ser el logo, la marca, color o algún diseño corporativo característico. Además, se consideran las emociones, valores o sentimientos que se logran transmitir mediante esta imagen (Folgado-Fernández, y otros, 2019). Posicionamiento, es un espacio o lugar que ocupa una determinada marca en la mente del cliente o consumidor y que hace una diferencia de sus competidores. Define lo que el cliente espera del servicio o producto y lo que representa para este (Carpio Maraza, y otros, 2019). Innovación, se considera un factor prioritario y relevante cuyo propósito conlleva a mejorar la competitividad de las empresas, haciendo frente a un entorno complejo y global, sin importar el tamaño, el tipo de actividad y el mercado al que van dirigido sus productos y servicios (Barrios Zarta, 2019). Tecnología, en este caso las Tecnologías de la Información y Comunicación es uno de los elementos que más se implementan como fuente de innovación en las organizaciones, pues es una plataforma para la presentación de nuevos productos o servicios al mercado u optimizar los existentes (Bernal-Jiménez, y otros, 2019).

Pérez Gorostegui (2017), teoriza con respecto a la estrategia funcional, representan la realización de la estrategia corporativa de la organización, y se deben considerar los elementos. Finanzas, es buscar estrategias para la fuente de financiamiento



para realizar la diversificación, determinar costes asociados; asimismo tener en claro las ofertas, promociones y descuentos que se deben ofrecer con relación al nuevo servicio (Palacios Rodríguez, 2020). Marketing, consiste en determinar de forma precisa las acciones que deberá llevar a cabo la empresa para lograr insertar sus productos o servicios en el mercado, la fijación de la caracterización del servicio, precios, técnicas de comunicación, canales de distribución más adecuados de acuerdo al contexto; considerando los accesos que posee cliente para el conocimiento de su servicio (Toledo, y otros, 2017). Dirección general, representa el área involucrada y soporte de la ejecución de las estrategias y debe estar alineadas a los objetivos estratégicos de la organización. La organización participa en el desarrollo de políticas, las cuales serán las directrices generales que se implementarán en el curso de acción que se elija y por tanto debes estar involucrado en el mismo (Alvarez Pablo, y otros, 2019).

Según Solano Torrenegra (Solano Torrenegra, 2018) El servicio al cliente tiene una fuerte relación entre la empresa y este último, ya que es el alma del negocio. En cualquiera de sus presentaciones ya sea de manera particular, de nivel bajo, medio o alto, empresas micro comercializadoras, mediana o macro empresas multinacionales. Es importante definir qué se entiende por "cliente". Este último se ve representado a través de bajas y altas expectativas ante lo que pudiera ofrecer su organización con relación a lo que existe en el mercado, exigiendo cada vez más la mejora en los servicios. Por otra parte, existen dos herramientas fundamentales como son la segmentación y la cartera de clientes. La segmentación de cartera, consiste en clasificar la base de clientes en función a características muy semejantes o patrones de consumo, con la única finalidad satisfacer sus necesidades a través de una gran variedad de productos y servicios. La cartera de clientes conlleva al seguimiento y tutoría personalizada de un grupo de clientes con un alto índice de potencial, ya sean éstos personas con un rubro natural o jurídica a quienes se les brinda la mejor atención para así lograr incrementar el volumen de negocios y la venta los servicios o productos.

### III. METODOLOGÍA

#### 3.1. Tipo y diseño de investigación

##### 3.1.1 Tipo de investigación

La investigación es de tipo aplicada, ya que se realizará una solución mediante la Metodología CRISP-DM, que permitirá la diversificación de cursos de acuerdo al perfil de los clientes en la empresa ACCOM Perú SAC (Roberto, Hernández Sampieri, Mendez Valencia Sergio, Mendoza Torres Christian Paulina, 2017).

##### 3.1.2 Diseño de investigación

La investigación es de tipo no experimental, tomando como base el paradigma cuantitativo, ya que los datos se examinarán de manera numérica, y así demostrar la eficacia de la metodología a emplear para la diversificación de cursos de acuerdo al perfil de los clientes en la empresa ACCOM Perú SAC

#### 3.2 Variables y Operacionalización

##### 3.2.1. Variables

Metodología CRISP-DM para el análisis de asociación y clustering

**Tabla 1** Variables e indicadores

VARIABLES	INDICADORES
Metodología CRISP-DM para el análisis de asociación y Clustering	<ul style="list-style-type: none"><li>• Descriptor de variables</li><li>• Calidad de los datos</li><li>• Modelado de los datos</li><li>• Técnica de modelado</li><li>• Evaluación óptima de clúster</li><li>• Patrones de asociación</li></ul>

---

Diversificación de cursos	<ul style="list-style-type: none"> <li>• Crecimiento</li> <li>• Diferenciación</li> <li>• Entorno</li> <li>• Planeamiento</li> <li>• Imagen de marca</li> <li>• Posicionamiento</li> <li>• Innovación</li> <li>• Tecnología</li> <li>• Finanzas</li> <li>• Marketing</li> <li>• Dirección general</li> </ul>
---------------------------	--

---

### 3.2.2. Indicadores

**Tabla 2** Descripción de indicadores

Variable Independiente	
Indicador	Descripción
Descriptor de variables	Se considera los atributos que forman parte del conjunto de datos que se analizará.
Calidad de los datos	Se considera el detalle y el entendimiento de los datos del proyecto
Modelado de los datos	Se considera el filtrado o limpieza de los datos que se disponen para conservar consistencia.
Técnica de modelado	Es el modelo de los datos, mediante técnica de minería de datos
Evaluación óptima de clúster	Comparación de los datos si corresponden a los objetivos propuestos

Patrones de asociación	de Inferencia del conocimiento mediante los modelos extraídos.
Variable dependiente	
Crecimiento	Es la diversificación de negocios o integración vertical a lo largo de la cadena de valor.
Diferenciación	Es la oferta de servicios específicos o exclusivos para un segmento del público
Entorno	Relaciones con lo que rodea a la organización, considerando los stakeholders
Planeamiento	Es la relación en base a una coordinación entre los negocios, creando compromiso entre todas las partes que componen la empresa
Imagen de marca	Son los Elementos palpables e impalpables que representa los valores que se desean transmitir a los clientes
Posicionamiento	Se considera un espacio ocupa una determinada marca en el subconsciente del cliente o consumidor
Innovación	Factor prioritario y relevante cuyo propósito conlleva a mejorar la competitividad de las empresas.
Tecnología	Es uno de los elementos, pues se puede utilizar una o varias plataformas para la presentación de nuevos productos o servicios u optimizar los existentes
Finanzas	Es determinar costes asociados; como tener en claro las ofertas, promociones y descuentos que se deben ofrecer con relación al nuevo servicio
Marketing	Consiste en determinar de forma precisa las acciones que deberá llevar a cabo la empresa para lograr insertar sus propios productos o servicios, considerando los accesos que posee cliente para el conocimiento de su servicio
Dirección general	Dirección general, representa el área involucrada y soporte de la ejecución de las

	estrategias, las mismas que serán las directrices generales que se implementarán en el curso de acción que se elija y por tanto debes estar involucrado en el mismo
--	---

### 3.3 Población, muestra y muestreo

#### 3.3.1 Población

Cruz, y otros (2015), lo describe como un conjunto de elementos, individuos, entidades con singularidades que se utilizan como unidades de muestreo, también conocido como Universo (Cruz, y otros, 2015). La población para el desarrollo de la Investigación son la población A, constituida por los registros de estudiantes de las bases de datos de ACCOM y la población B, conformada por los estudiantes de los últimos ciclos universitarios y profesionales que componen el mercado de la capacitación en la ciudad de Piura.

**Tabla 3** Población

Población	Unidad de Análisis	TOTAL (2011-2020)
A	Registros de captación de clientes	5222
B	Profesionales y estudiantes de los últimos ciclos	384

#### 3.3.2 Muestra

Para la muestra A, aplicando técnicas de minería de datos, se requiere una extensa cantidad de datos a procesar para un mejor ajuste del modelo, por ello en la investigación se trabajará con una muestra por conveniencia de esta manera no se requerirá realizar el muestreo.

Para la muestra B, cálculo de la muestra se consideró la siguiente fórmula.

$$n = \frac{z^2 * p * q}{e^2}$$

n = Representa el tamaño de la muestra.

z = Es el valor de confiabilidad y su nivel de significancia (95%).

e = Es el margen de error (5%).

p = Probabilidad de éxito, representada en porcentaje de la con características o atributos (se asume 50% para este caso).

q = Probabilidad de fracaso, representada en porcentaje de la con características o atributos (se asume 50% para este caso).

Reemplazamos en la fórmula obtenemos:

$$n = \frac{1.96^2 * 0.5 * 0,5}{0.05^2}$$

La muestra estará compuesta por 384 encuestas.

### 3.4 Técnicas e instrumentos de recolección de datos

#### 3.4.1 Técnicas

Como instrumento de la variable Metodología CRISP-DM para el análisis de asociación y clustering se utilizó fichas de registros para los indicadores propuestos para la recolección de información de las mismas y la técnica que se empleará será la observación directa y para la diversificación de cursos se utilizó la técnica de la encuesta.

#### 3.4.2 Instrumentos

**Tabla 4** Instrumentos de Metodología CRISP-DM

Indicador	Técnica	Instrumento
Descriptor de variables	Observación	Fichas de registro
Calidad de los datos		
Modelado de los datos		
Técnica de modelado		
Evaluación óptima de clúster		
Patrones de asociación		

**Tabla 5** Instrumentos de diversificación de cursos

Indicador	Técnica	Instrumento
Crecimiento	Encuesta	Cuestionario
Diferenciación		
Entorno		
Planeamiento		
Imagen de marca		
Posicionamiento		
Innovación		
Tecnología		
Finanzas		
Marketing		
Dirección general		

### 3.5 Procedimientos

En primer lugar, se realizó una entrevista con el representante máximo de la empresa ACCOM Perú SAC, Ing. Misael Henry Julca García, para el consentimiento de la extracción de la información de los clientes que se han capacitados desde el año 2011 hasta la actualidad de sus sistemas de información web. Posteriormente se elaboró los instrumentos que serán aplicados considerando la metodología CRISP-DM, además de verificar con la información que se encuentra disponible para su extracción y posterior procesamiento.

En cuanto a la diversificación se aplicó como instrumento la encuesta y el instrumento la encuesta a los profesionales y estudiantes de los últimos ciclos universitarios y profesionales que componen el mercado de la capacitación en la ciudad de Piura, mediante formularios de Google Forms.

### 3.6 Método de análisis de los datos

La información es obtenida y registrada a través de las técnicas e instrumentos, será procesada mediante el software R y el uso de sus

respectivas librerías para el análisis de minería de datos. Para la elaboración de los instrumentos de recolección de la información, serán validados por expertos para su aplicación en el desarrollo de la investigación. Además, se utilizará algoritmos como referente principal entre los diversos métodos para encontrar patrones o tendencias y mediante técnicas no supervisadas.

En cuanto a la diversificación de cursos, aplicó una prueba piloto de los resultados de la aplicación de la encuesta a los estudiantes de los últimos ciclos universitarios y profesionales que componen el mercado de la capacitación en la ciudad de Piura. Según (González González, y otros, 2017)

### 3.7 Aspectos éticos

Teniendo en consideración el respeto por los derechos de autor y sus fuentes de información, se realiza la apropiada cita según la norma ISO 690:2010, el cual se muestra de manera estructurada la conformación de las citas y referencias.

Según lo establecido por el Colegio de Ingenieros del Perú y sus artículos, basados en la ética, normas y reglas que por obligación se debe seguir. Sin excepción alguna para no incurrir en una falta, de nivel leve o grave (CIP, 1987).

También se considera lo expresado en el artículo N°43 del código de ética profesional del Colegio de Ingenieros del Perú, referente a la inviolabilidad del trabajo ajeno, asimismo se respetó el artículo N° 41 considerándose la idea expresada por los autores. Asimismo, lo expresado por la Universidad César Vallejo en lo que respecta a faltas de ética y sanciones. Además del artículo N° 1 del código de ética de esta última, hace mención en cuanto a que la investigación se desarrolló con el más alto rigor en el aspecto científico, de responsabilidad y honestidad en cuanto a la elaboración del informe basado estrictamente con el manejo de la información recopilada. Finalmente se considera el artículo N°15 con respecto al plagio y el artículo N°16 con respecto a la autoría del trabajo.



## IV. RESULTADOS

4.1. En la dimensión de comprensión del negocio, la preparación de datos y la comprensión de estos. Cuya función es de comprender el negocio desde una perspectiva empresarial y la realidad en que se encuentran actualmente con respecto a estrategias empleadas de manera empírica durante todos estos años, obteniendo como resultado una disminución considerable en las inscripciones considerando los años 2020 y 2021 viéndose afectados por la coyuntura actual que atraviesa el mundo debido al COVID-19, considerando la metodología CRISP-D, entendiendo el negocio y plasmando objetivos técnicos considerando las inscripciones como el proceso más relevante donde se mide efectividad de las estrategias, los ingresos de dinero y la productividad de los colaboradores.

Para plasmar los objetivos del negocio con los datos, se describió las variables más relevantes para alcanzar lo planteado en la investigación con la variedad de registros existentes en el datawarehouse de la institución. Luego se realizó un estudio de las bases de datos, la calidad de los datos para comprobar su validez de análisis, el tipo de datos o sub conjunto de datos que almacena con el fin de diversificar los cursos de acorde el perfil de los clientes.

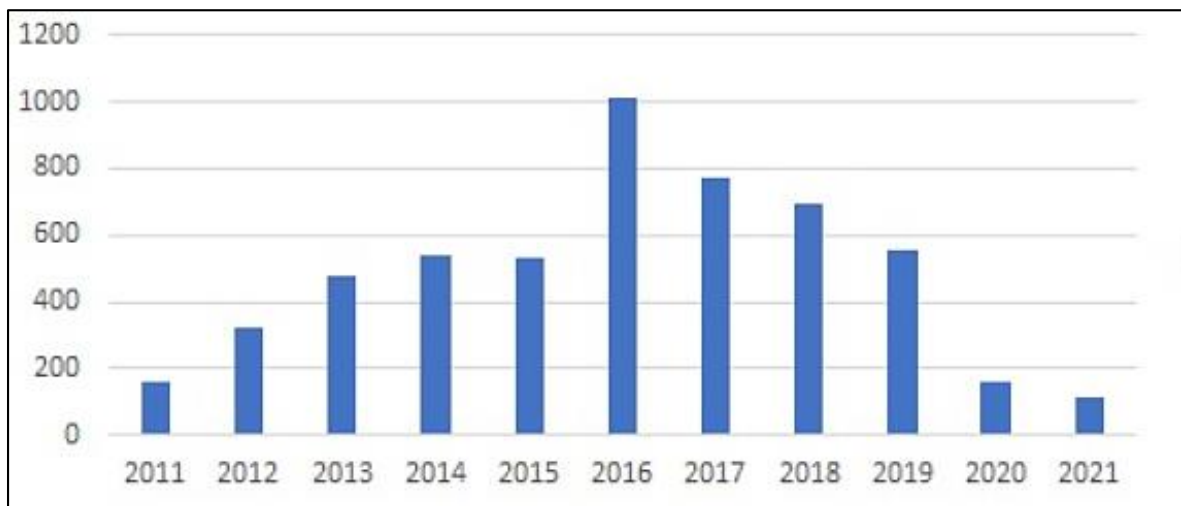


Ilustración 4 *Inscripciones por año*

4.2 Con respecto a preparación de los datos, la calidad de datos se tienen el diagrama de la base datos con la cual está trabajando la empresa ACCOM Perú S.A.C, aplicando la estrategia de modelamiento de base datos para poder definir modelos de asociación a través de clúster aplicando inteligencia de negocio.

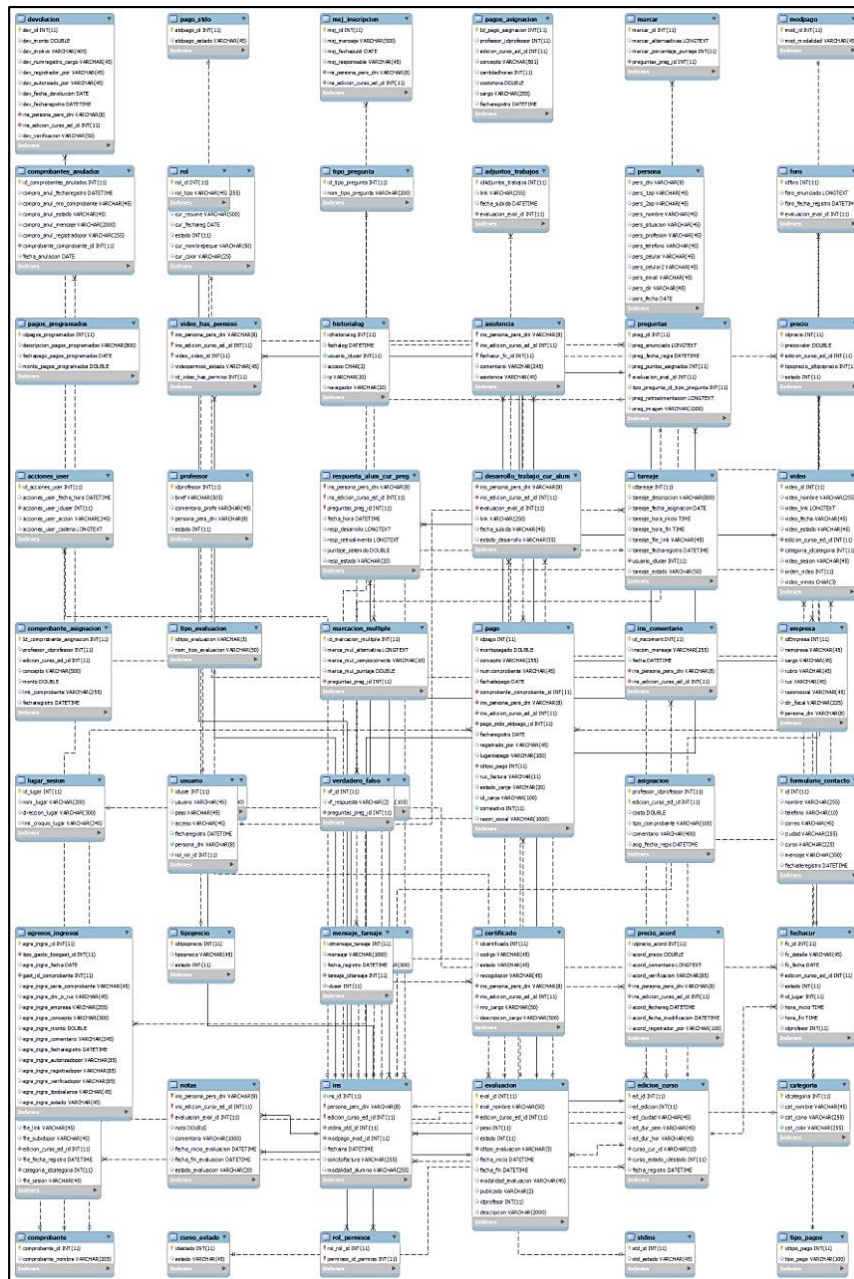


Ilustración 5 diagrama de la base datos

Nombre de Archivo: Tabla curso

Fecha de creación:

Descripción: Tabla de datos que contiene los cursos que dicta la empresa ACCOM PERÚ SAC

Campo	Tamaño	Tipo de Date	Descripción
cod_curso	7	Varchar	Codigo del curso
curso	45	Varchar	Nombre del curso
cod_tcurso	7	Varchar	Codigo del tipo de curso

Ilustración 6 Tabla curso

**Nombre de Archivo:** Tabla Hechos ACCOM Perú

**Fecha de creación:** 6/11/2021

**Descripción:** Tabla de datos que contiene la datamart de las inscripciones de la empresa.

Campo	Tamaño	Tipo de Dato	Descripción
cod_inscripcion	8	Varchar	Codigo de inscripción al curso
pers_dni	8	Varchar	DNI de la persona
persona.cod_genero		Integer	Código del genero de la persona
persona.cod_profesior	7	Varchar	Código de la profesión de la persona
persona.cod_cargo	7	Varchar	Código del cargo que ocupa la persona en la empresa
persona.cod_area	7	Varchar	Código del área que ocupa la persona en la empresa
cod_curso	7	Varchar	Código del curso
cod_Einscripcion	7	Varchar	Código del estado de la inscripción al curso
cod_modalidad	7	Varchar	Código de la modalidad del curso
fechains	dd/mm/yyyy	Date	Fecha de inscripción

**Relaciones:**

Tabla hechos con tabla inscripción, tabla persona, tabla genero, tabla profesión, tabla cargo, tabla área, tabla curso, tabla estado inscripción y tabla tiempo

**Nombre de Archivo:** Tabla persona

**Fecha de creación:** 1/10/2021

**Descripción:** Tabla de datos que contiene los datos personales de la persona que se registro en la empresa

Campo	Tamaño	Tipo de Dato	Descripción
pers_dni	8	Varchar	DNI de la persona
pers_edad		Integer	Edad de la persona
cod_genero		Integer	Código del genero de la persona
cod_profesion	7	Varchar	Código de la profesión de la persona
cod_cargo	7	Varchar	Código del cargo que ocupa la persona en la empresa
cod_area	7	Varchar	Código del área que ocupa la persona en la empresa

**Relaciones:**

Tabla persona con tabla genero, tabla profesión, tabla cargo, tabla área

**Nombre de Archivo:** Tabla genero

**Fecha de creación:** 1/10/2021

**Descripción:** Tabla de datos que contiene el genero de la persona

Campo	Tamaño	Tipo de Dato	Descripción
cod_genero	7	Varchar	Codigo del genero
genero		Intenger	Descripción del genero

**Relaciones:**

Tabla persona con tabla genero, tabla profesión, tabla cargo, tabla área

**Nombre de Archivo:** Tabla profesión

**Fecha de creación:** 1/10/2021

**Descripción:** Tabla de datos que contiene las profesiones que se han registrado en la empresa

Campo	Tamaño	Tipo de Dato	Descripción
cod_profesion	7	Varchar	Codigo de la profesión
profesion	45	Varchar	Nombre de la profesión

**Nombre de Archivo:** Tabla cargo

**Fecha de creación:** 1/10/2021

**Descripción:** Tabla de datos que contiene los cargos que se han registrado en la empresa

Campo	Tamaño	Tipo de Dato	Descripción
cod_cargo	7	Varchar	Codigo del cargo
cargo	45	Varchar	Nombre del cargo

**Nombre de Archivo:** Tabla área

**Fecha de creación:** 1/10/2021

**Descripción:** Tabla de datos que contiene las áreas que se han registrado en la empresa

Campo	Tamaño	Tipo de Dato	Descripción
cod_area	7	Varchar	Codigo del área
area	45	Varchar	Nombre del área

Ilustración 7 Tablas de datos

Nombre de Archivo: Tabla curso

Fecha de creación: 1/10/2021

Descripción: Tabla de datos que contiene los cursos que dicta la empresa ACCOM PERÚ SAC

Campo	Tamaño	Tipo de Dato	Descripción
cod_curso	7	Varchar	Codigo del curso
curso	45	Varchar	Nombre del curso
cod_tcurso	7	Varchar	Codigo del tipo de curso

Nombre de Archivo: Tabla curso tipo

Fecha de creación: 1/10/2021

Descripción: Tabla de datos que contiene los tipos de curso de la empresa ACCOM PERÚ SAC

Campo	Tamaño	Tipo de Dato	Descripción
cod_tcurso	7	Varchar	Codigo del área
tipo	45	Varchar	Descripción de los tipos de cursos

Nombre de Archivo: Tabla inscripciones

Fecha de creación: 1/10/2021

Descripción: Tabla de datos que contiene la información de las inscripciones que se han realizado

Campo	Tamaño	Tipo de Dato	Descripción
cod_inscripcion	7	Varchar	Código de la inscripción
pers_dni	8	Varchar	DNI de la persona
fechains	dd/mm/yyyy	Date	Fecha de inscripción al curso
cod_curso	7	Varchar	Código de la profesión de la persona
cod_Einscripcion	7	Varchar	Código del estado de la inscripción al curso
cod_modalidad	7	Varchar	Código de la modalidad del curso

Nombre de Archivo: Tabla estado inscripción

Fecha de creación: 1/10/2021

Descripción: Tabla de datos que contiene el estado de las inscripciones

Campo	Tamaño	Tipo de Dato	Descripción
cod_Einscripcion	7	Varchar	Codigo del estado de inscripción al curso
estado	45	Varchar	Descripción de los tipos de cursos

Nombre de Archivo: Tabla modalidad

Fecha de creación: 1/10/2021

Descripción: Tabla de datos que contiene el estado de las inscripciones

Campo	Tamaño	Tipo de Dato	Descripción
cod_modalidad	7	Varchar	Codigo de la modalidad del curso
descripcion	45	Varchar	Descripción de la modalidad

Nombre de Archivo: Tabla tiempo

Fecha de creación: 1/10/2021

Descripción: Tabla de datos que contiene diferentes formas de mostrar las fechas del 2011 al 2021

Campo	Tamaño	Tipo de Dato	Descripción
fecha	dd/mm/yyyy	Date	Muestra la fecha
anio	yyyy	Date	Muestra la fecha en función al número del año
num_mes	mm	Date	Muestra la fecha en función al número del mes
mes	mmmm	Date	Muestra la fecha en función al nombre del mes
mes_anio	mmm-yy	Date	Muestra la fecha en función a las primeras 3 letras del nombre del mes y el año
num_semana		Date	Muestra la fecha en función al número de la semana
num_día_semana		Date	Muestra la fecha en función al número del día de la semana
día_semana	dddd	Date	Muestra la fecha en función al nombre del día de la semana
día_mes	dd	Date	Muestra la fecha en función al número del día del mes
trimestre		Date	Muestra la fecha en función al número de trimestres del año
semestre		Date	Muestra la fecha en función al número de semestres del año
trimestre_anio		Date	Muestra la fecha en función acada uno de los años y su número de trimestres
semestre_anio		Date	Muestra la fecha en función acada uno de los años y su número de semestres

Ilustración 8 Tablas de datos y tiempo

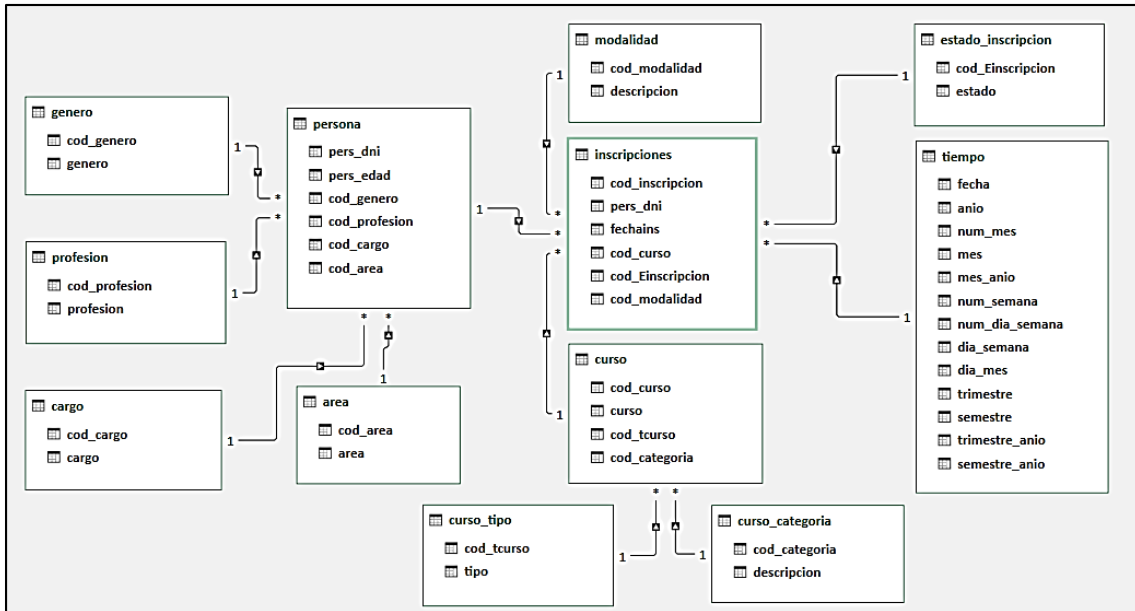


Ilustración 9 Modelado de la base datos

En la ilustración 9, el modelado de la base datos con las tablas hechos y la dimensión de tiempo, así como la reducción de las 53 tablas de origen a tan solo 12 tablas.

En lo que corresponde al Modelado de los datos, se opta por la preparación de los datos para obtener una matriz de dos entradas de estos últimos, con la finalidad de obtener 2 modelos haciendo uso del Análisis por componentes, así conocer las asociaciones y clústeres de las categorías por curso en los meses del año, así como los cargos por personas.

**Tabla 6** Formato de modelado por categoría de curso

	Direccion	DisenioDibujo	OfimaticaCalculo	ProcesoAdmin
Abril	119	70	188	80
Agosto	103	115	246	115
Diciembre	101	69	116	33
Enero	90	212	284	118
Febrero	73	64	125	45
Junio	147	75	192	59
Julio	92	74	233	47
Marzo	116	69	118	58
Mayo	74	59	176	66
Noviembre	87	88	139	75
Octubre	63	81	216	65
Setiembre	102	54	157	73

**Tabla 7** Formato de modelado por cargo de persona

	Gerente	JefeArea	Operario	Supervisor	Estudiante	Otros
Enero	3	37	130	48	15	471
Febrero	0	16	66	29	9	187
Marzo	2	20	90	33	8	208
Abril	4	42	128	59	8	216
Mayo	2	16	91	33	13	221
Junio	1	33	139	46	9	245
Julio	0	23	137	42	2	242
Agosto	0	23	145	59	19	333
Setiembre	1	19	95	38	11	222
Octubre	0	24	113	41	2	245
Noviembre	0	17	84	42	6	240
Diciembre	2	14	61	26	6	210

Con respecto al primer modelo, se necesita conocer la asociación entre las categorías de los cursos, para ello se utilizó el software R, utilizando el PCA junto con la librería “Factoextra” representado a través del mapa de dimensiones como se observa en la ilustración 10, así mismo se utiliza el algoritmo de K-Means, que es un método que actualiza dinámicamente los k prototipos de los datos, con la finalidad de maximizar la similitud dentro del clúster; logrando 2 clúster con una

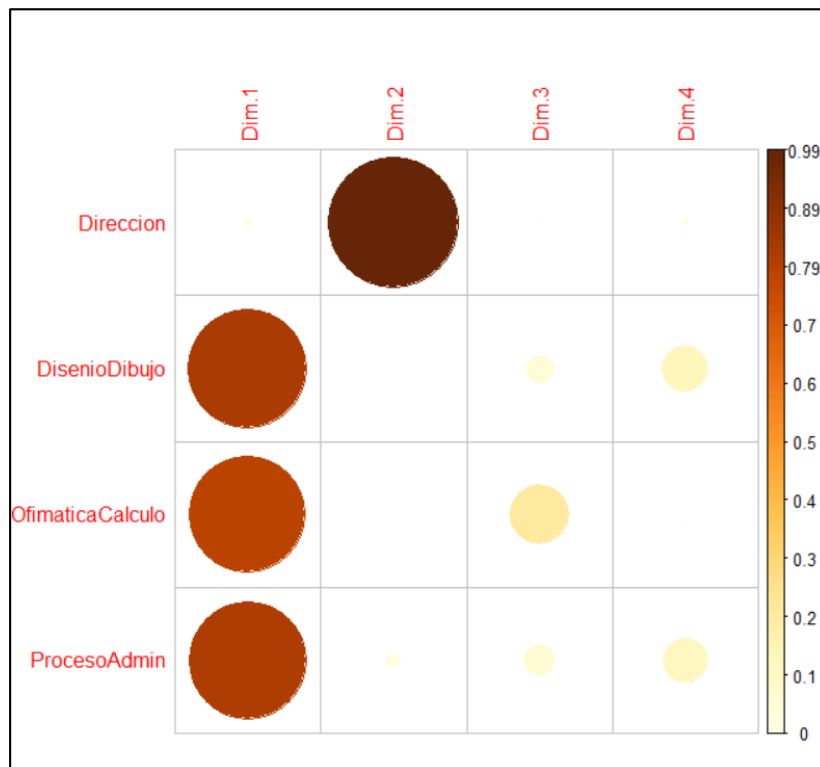


Ilustración 10 mapa de dimensiones

correlación parcialmente negativa entre ellos como se aprecia en la ilustración 12, la cual se ve respaldada por la correlación dimensional expresada en la ilustración 9. Con esto demuestra que mientras las categorías de cursos: “Diseño y Dibujo, Procesos administrativos y Ofimática de cálculo avanzan en un mismo sentido o dirección independiente mente de su fuerza, a diferencia de “Dirección” que tiende a ir en sentido contrario (jalando dimensionalmente) ubicado en primer cuadrante.

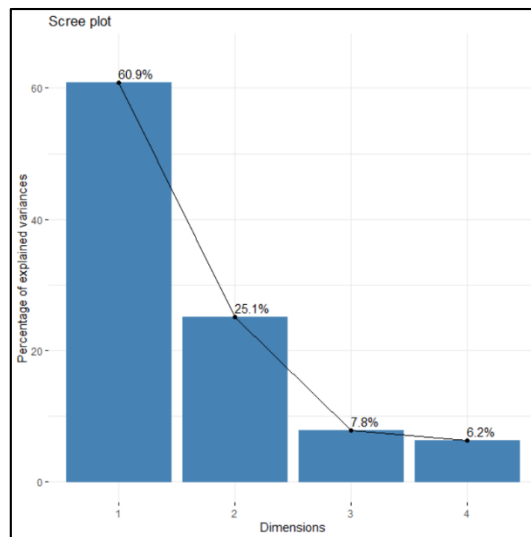


Ilustración 11 Variación de los datos por dimensiones

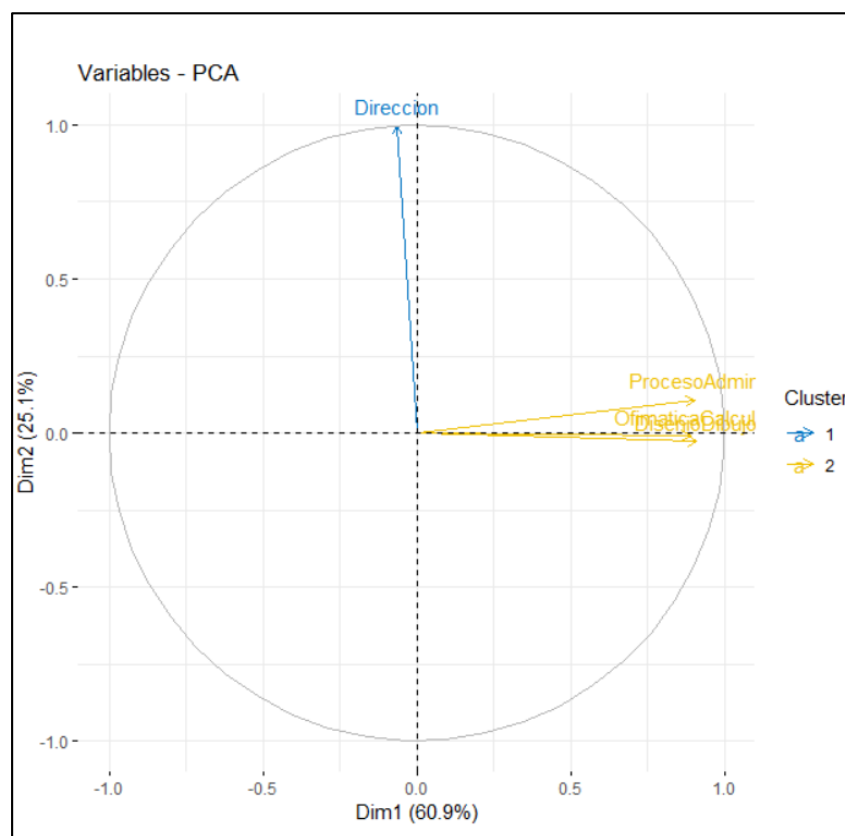


Ilustración 12 Mapa de clúster

En la ilustración 11, se muestra el aporte de cada una de las dimensiones expresada de manera porcentual y de manera acumulativa como también se observa en las ilustraciones 13 y 14, como son los resultados por variables del análisis de componentes principales.

```
> eig.val <- get_eigenvalue(racp)
> eig.val
      eigenvalue variance.percent cumulative.variance.percent
Dim.1  2.4353925      60.884812      60.88481
Dim.2  1.0052509      25.131273      86.01608
Dim.3  0.3115631       7.789078      93.80516
Dim.4  0.2477935       6.194837     100.00000
```

Ilustración 13 Codificación en R

```
> fviz_eig(racp, addlabels = TRUE, ylim = c(0, 65))
> var <- get_pca_var(racp)
> var
Principal Component Analysis Results for variables
=====
      Name      Description
1 "$coord"    "Coordinates for the variables"
2 "$cor"      "Correlations between variables and dimensions"
3 "$cos2"     "Cos2 for the variables"
4 "$contrib"  "contributions of the variables"
> head(var$coord)
      Dim.1      Dim.2      Dim.3      Dim.4
Direccion  -0.06522282  0.996522414  0.02381516  0.04606409
DiseñoDibujo  0.90917120 -0.026115374 -0.21205200  0.35743483
OfimaticaCalculo  0.88849548 -0.008784302  0.45859603 -0.01372248
ProcesoAdmin  0.90284105  0.106933732 -0.23605001 -0.34310875
> head(var$cos2)
      Dim.1      Dim.2      Dim.3      Dim.4
Direccion  0.004254017  0.99305692109  0.0005671617  0.0021219006
DiseñoDibujo  0.826592279  0.00068201273  0.0449660493  0.1277596589
OfimaticaCalculo  0.789424214  0.00007716396  0.2103103159  0.0001883065
ProcesoAdmin  0.815121958  0.01143482309  0.0557196054  0.1177236136
> head(var$contrib)
      Dim.1      Dim.2      Dim.3      Dim.4
Direccion  0.1746748  98.786969549  0.1820375  0.85631819
DiseñoDibujo  33.9408243  0.067845025  14.4324038  51.55892685
OfimaticaCalculo  32.4146611  0.007676089  67.5016695  0.07599334
ProcesoAdmin  33.4698398  1.137509337  17.8838892  47.50876163
```

Ilustración 14 Codificación en R



```

> set.seed(123)
> res.km <- kmeans(var$coord, centers = 2, nstart = 25)
> res.km
K-means clustering with 2 clusters of sizes 1, 3

Cluster means:
      Dim.1      Dim.2      Dim.3      Dim.4
1 -0.06522282 0.99652241 0.023815156 0.04606409
2  0.90016924 0.02401135 0.003498008 0.00020120

Clustering vector:
      Dirección  DiseñoDibujo OfimaticaCalculo  ProcesoAdmin
1              1              2              2              2
2              1              2              2              2

Within cluster sum of squares by cluster:
[1] 0.0000000 0.5673195
(between_SS / total_SS = 71.3 %)

Available components:
[1] "cluster"      "centers"      "totss"      "withinss"    "tot.withinss" "betweenss"   "size"      "iter"
[9] "ifault"

```

Ilustración 15 Cauterización a través de K-means

En la ilustración 15, se genera los 2 centroides con un tamaño de 1 y 3, los cuales se encuentran expresados de manera tácita, para así poder general la parcialización de los datos, generando un total de variabilidad del 71.3 % del total de la base de datos.

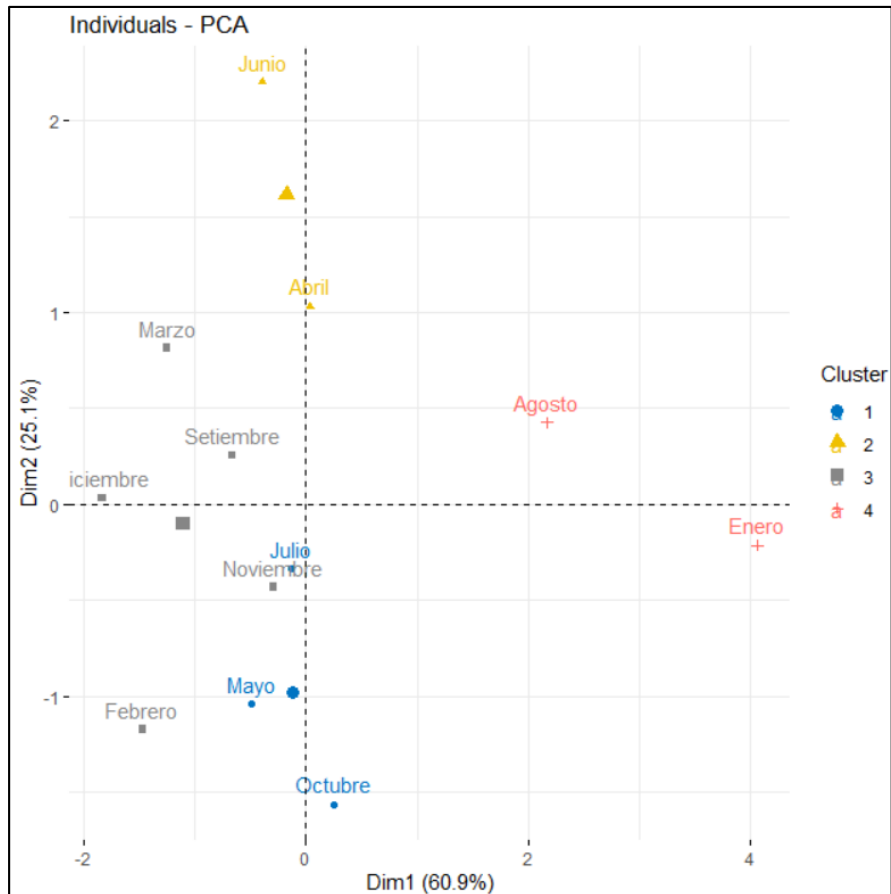


Ilustración 16 Correspondencia por individuos (Meses del año)

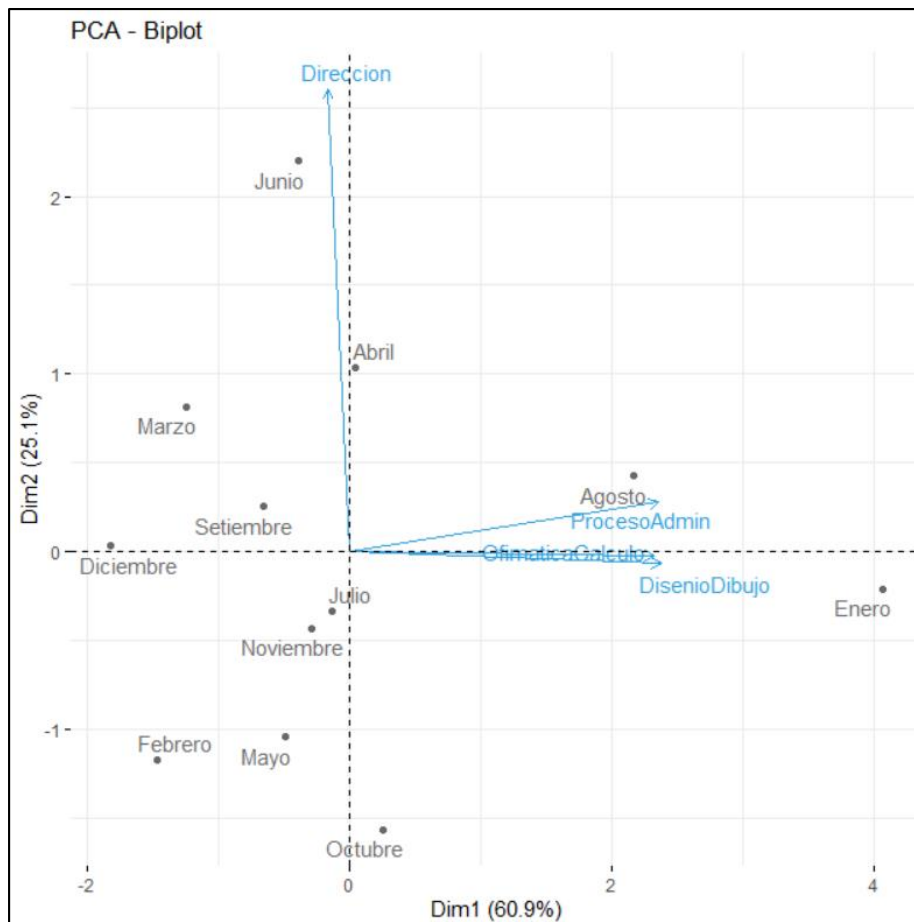


Ilustración 17 Definición de asociación y clúster

Con respecto al segundo modelo se necesita conocer la asociación entre los cargos de sus clientes con la finalidad de establecer un análisis del estado actual de la base según el perfil. Se utiliza las mismas librerías anteriormente mencionada en el primer modelo. Logrando 4 clúster con una correlación parcialmente positiva como se demuestra por la correlación dimensional expresada en la ilustración 18.

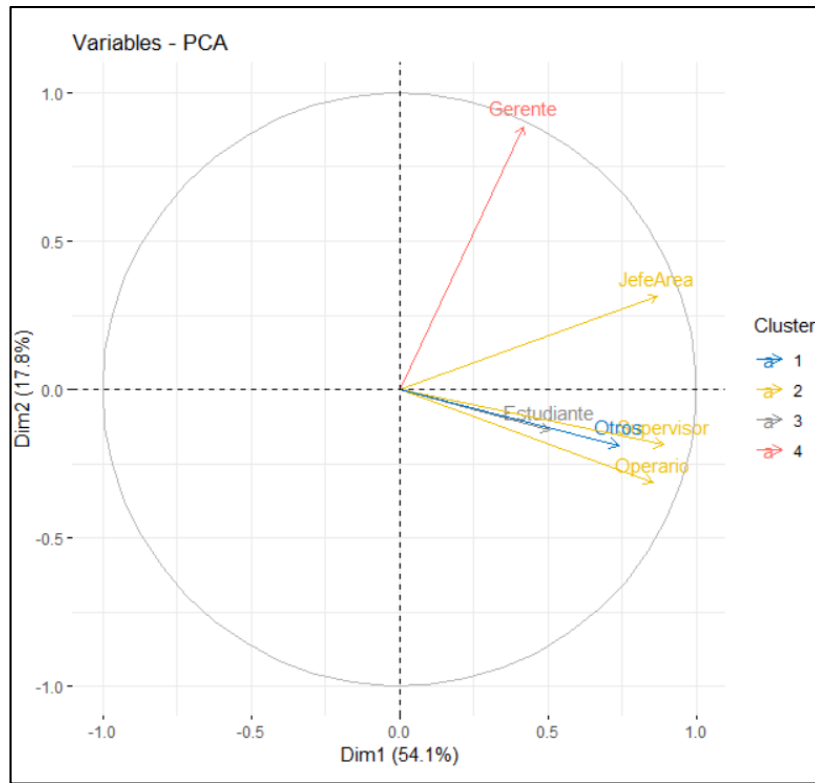


Ilustración 19 Mapa de Clúster

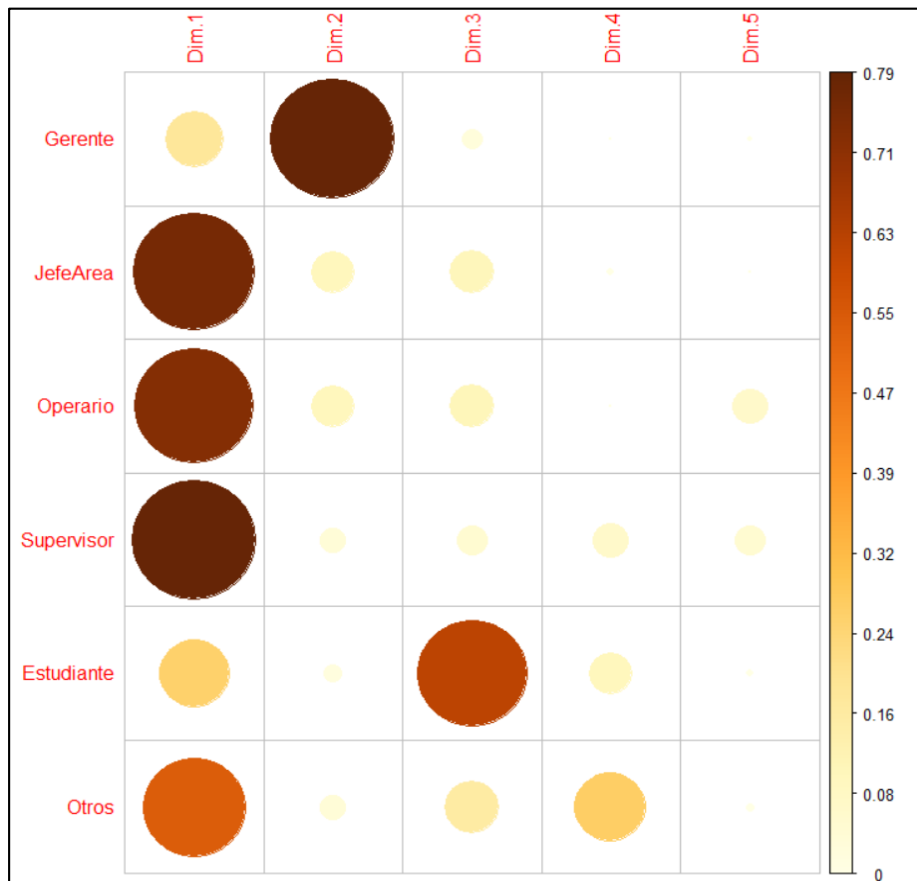


Ilustración 18 Mapa de Dimensiones

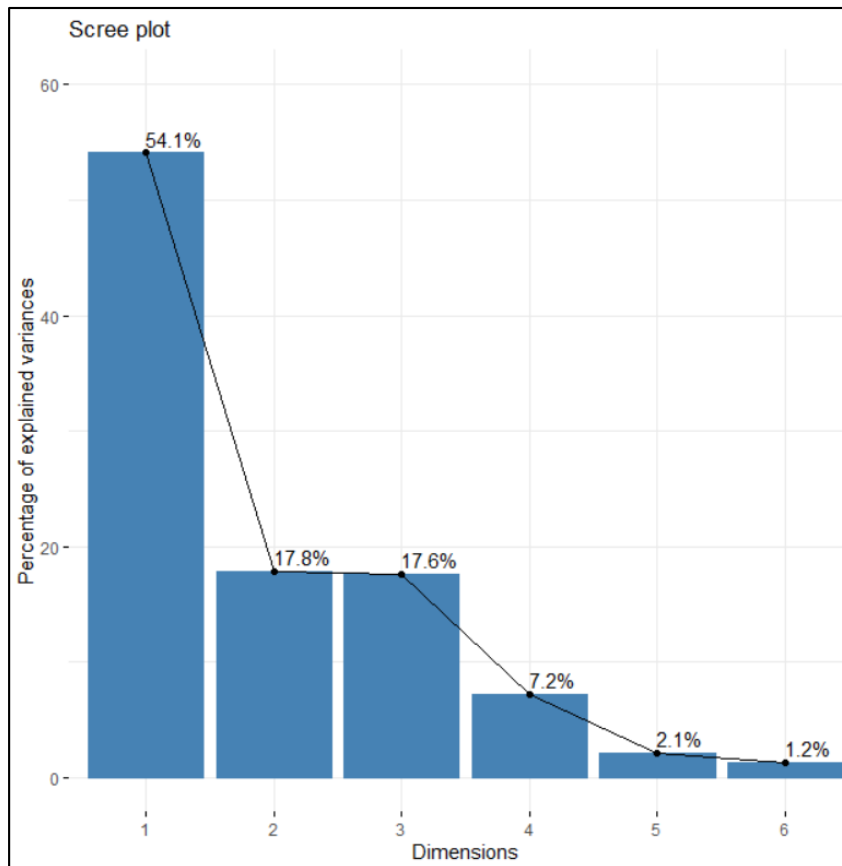


Ilustración 20 Variación de los datos por dimensiones

En la ilustración 20, se muestra el aporte de cada una de las dimensiones expresada de manera porcentual y de manera acumulativa como también se observa en las ilustraciones 21 y 22, como son los resultados por variables del análisis de componentes principales.

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	3.24616924	54.102821	54.10282
Dim.2	1.07024999	17.837500	71.94032
Dim.3	1.05478844	17.579807	89.52013
Dim.4	0.43107502	7.184584	96.70471
Dim.5	0.12313557	2.052259	98.75697
Dim.6	0.07458175	1.243029	100.00000

Ilustración 21 Codificación en R

```

> racp$var
$coord
      Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
Gerente  0.4187871  0.8860141  0.1556058  0.02362302  0.04466735
JefeArea 0.8689404  0.3121059 -0.3189861 -0.05348917 -0.02033227
Operario  0.8530336 -0.3137282 -0.3230278  0.02857594  0.25310100
Supervisor 0.8880147 -0.1873287 -0.2250970  0.25915586 -0.22346903
Estudiante 0.5052449 -0.1368895  0.7870841  0.31438723  0.05379843
Otros     0.7377129 -0.1885810  0.3928181 -0.51072322 -0.06192089

$cor
      Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
Gerente  0.4187871  0.8860141  0.1556058  0.02362302  0.04466735
JefeArea 0.8689404  0.3121059 -0.3189861 -0.05348917 -0.02033227
Operario  0.8530336 -0.3137282 -0.3230278  0.02857594  0.25310100
Supervisor 0.8880147 -0.1873287 -0.2250970  0.25915586 -0.22346903
Estudiante 0.5052449 -0.1368895  0.7870841  0.31438723  0.05379843
Otros     0.7377129 -0.1885810  0.3928181 -0.51072322 -0.06192089

$cos2
      Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
Gerente  0.1753826  0.78502098  0.02421316  0.0005580470  0.0019951721
JefeArea 0.7550574  0.09741008  0.10175212  0.0028610911  0.0004134013
Operario  0.7276663  0.09842536  0.10434697  0.0008165841  0.0640601166
Supervisor 0.7885701  0.03509206  0.05066866  0.0671617607  0.0499384079
Estudiante 0.2552724  0.01873873  0.61950143  0.0988393283  0.0028942716
Otros     0.5442203  0.03556278  0.15430609  0.2608382078  0.0038341967

$contrib
      Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
Gerente  5.402756  73.349310  2.295547  0.1294547  1.6203053
JefeArea 23.259953  9.101619  9.646686  0.6637107  0.3357286
Operario 22.416156  9.196484  9.892692  0.1894297  52.0240566
Supervisor 24.292328  3.278865  4.803680  15.5800633  40.5556326
Estudiante 7.863805  1.750874  58.732292  22.9285679  2.3504757
Otros    16.765001  3.322848  14.629103  60.5087737  3.1138012

```

Ilustración 22 Codificación en R

```

> res.km <- kmeans(racp$var$coord, centers = 4, nstart = 15)
> res.km
K-means clustering with 4 clusters of sizes 1, 3, 1, 1

Cluster means:
      Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
1 0.7377129 -0.18858096  0.3928181 -0.51072322 -0.061920891
2 0.8699962 -0.06298368 -0.2890370  0.07808088  0.003099899
3 0.5052449 -0.13688948  0.7870841  0.31438723  0.053798435
4 0.4187871  0.88601410  0.1556058  0.02362302  0.044667350

Clustering vector:
      Gerente  JefeArea  Operario Supervisor Estudiante  Otros
           4           2           2           2           3           1

Within cluster sum of squares by cluster:
[1] 0.0000000 0.3927135 0.0000000 0.0000000
      (between_SS / total_SS = 86.1 %)

Available components:

[1] "cluster"      "centers"      "totss"      "withinss"      "tot.withinss" "betweenss"
[7] "size"         "iter"         "ifault"

```

Ilustración 23 Cauterización a través de K-means

En la ilustración 23, se genera los 4 centroides con un tamaño de 1, 3, 1 y 1. Analizado para así poder general la parcialización de los datos, generando un total de variabilidad del 86.1 % del total de la base de datos.

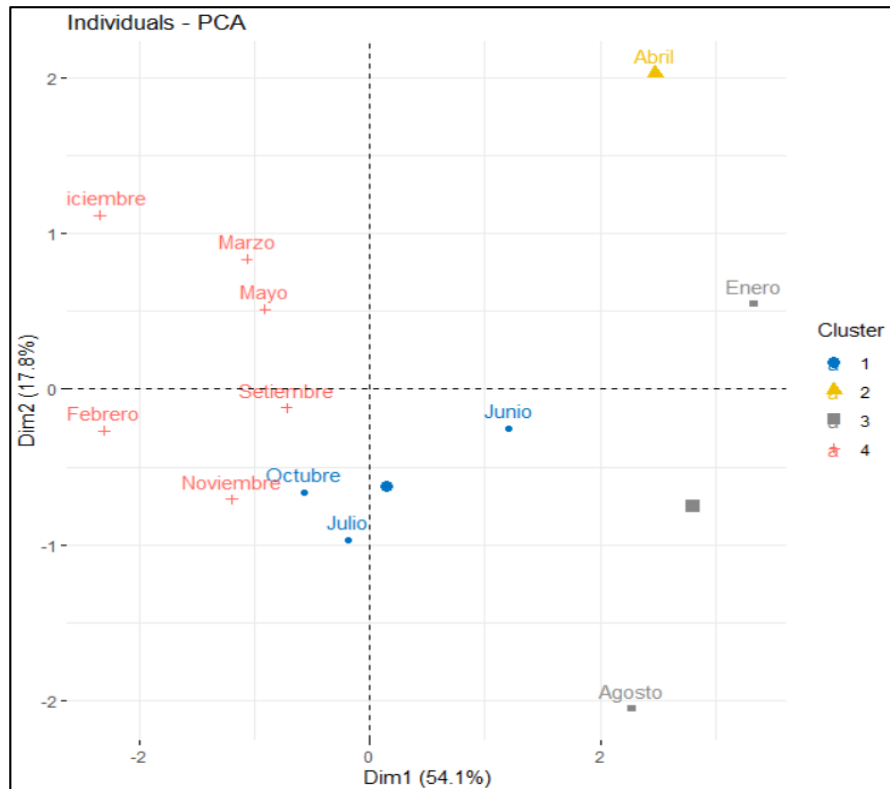


Ilustración 24 Correspondencia por individuos (Meses del año)

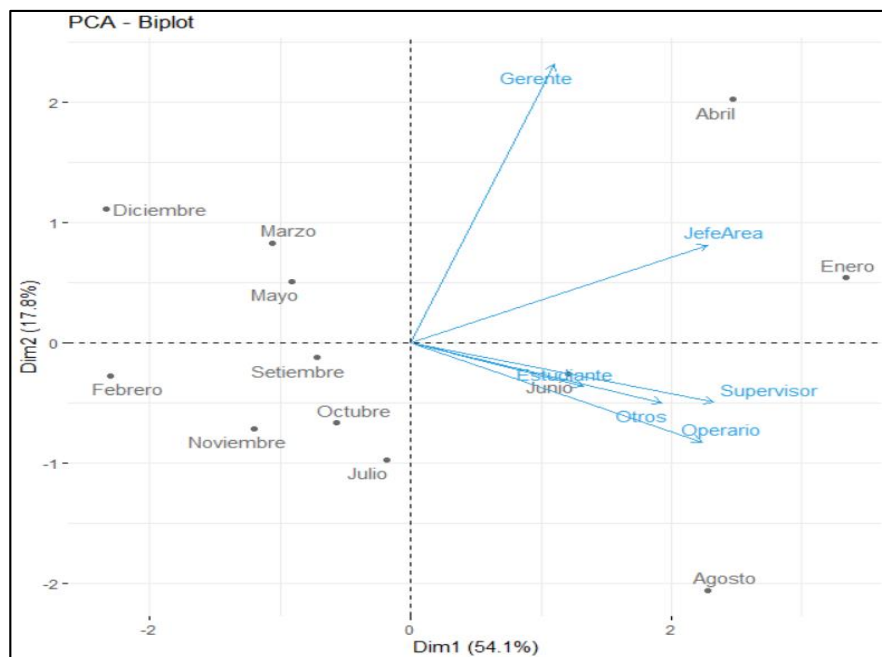


Ilustración 25 Definición de asociación y clúster

Con respecto al tercer modelo se necesita conocer la asociación entre los cargos de sus clientes con la categoría de los cursos con la finalidad de establecer un análisis del estado actual de la base de datos. Se utiliza las mismas librerías anteriormente mencionada en el primer y segundo modelo. Logrando 3 clúster con una correlación parcialmente negativa como se demuestra en la ilustración 26.

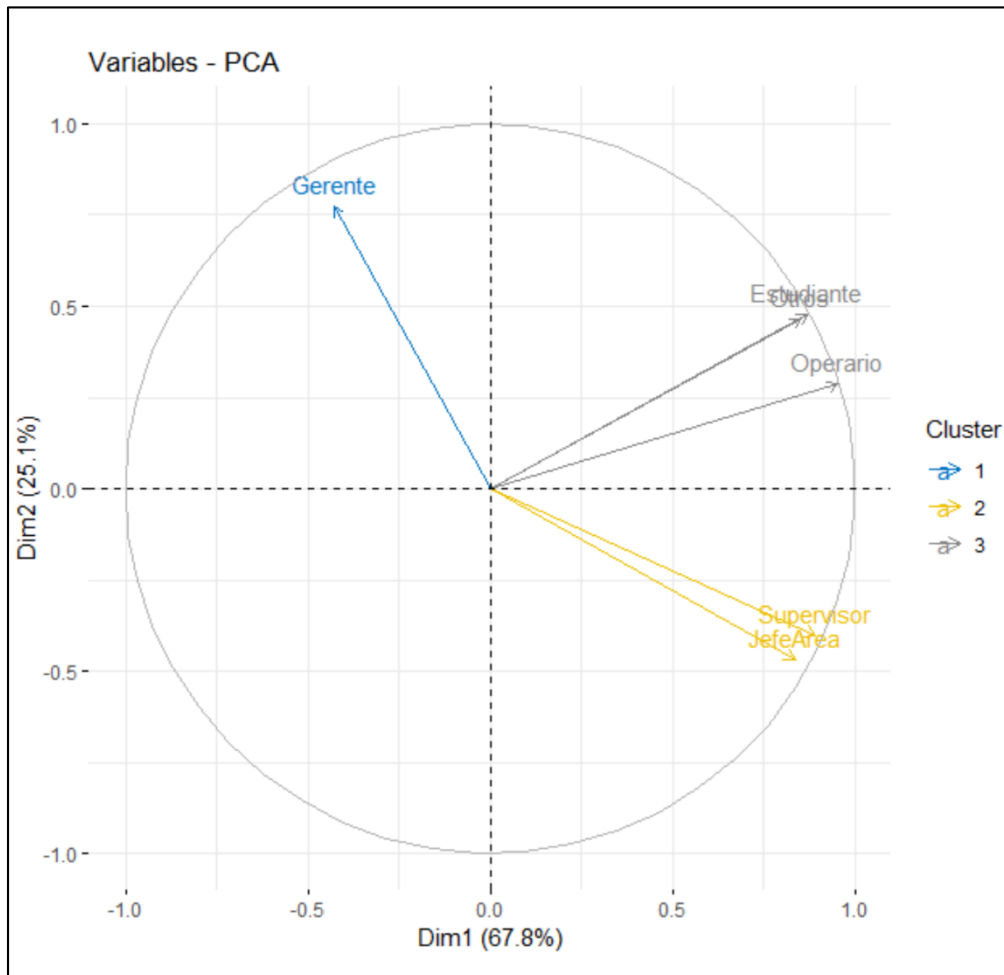


Ilustración 26 Mapa de Cluster

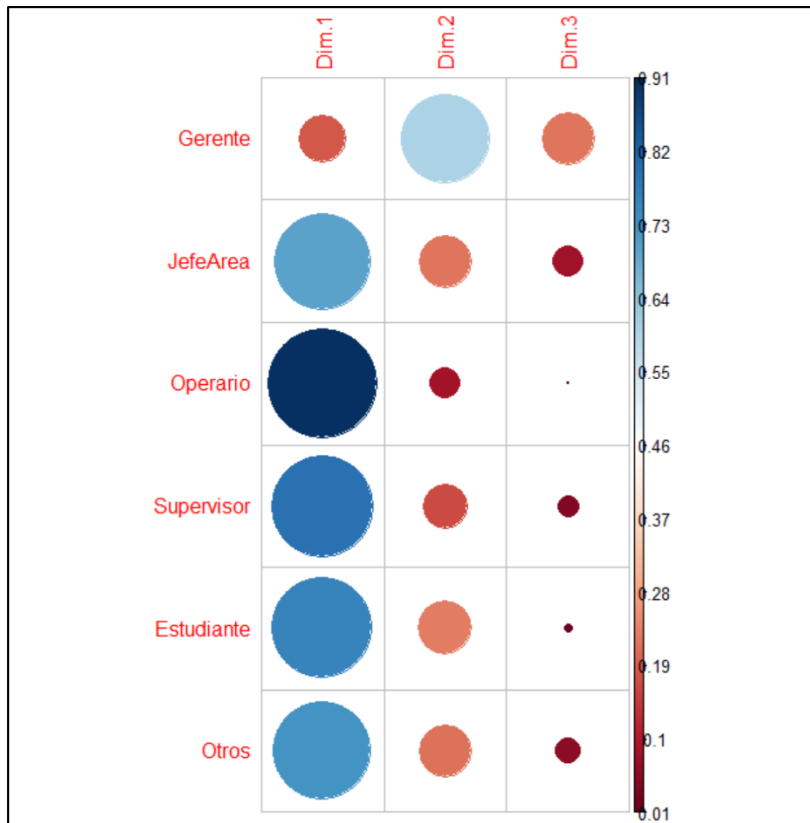


Ilustración 27 Mapa de Dimensiones

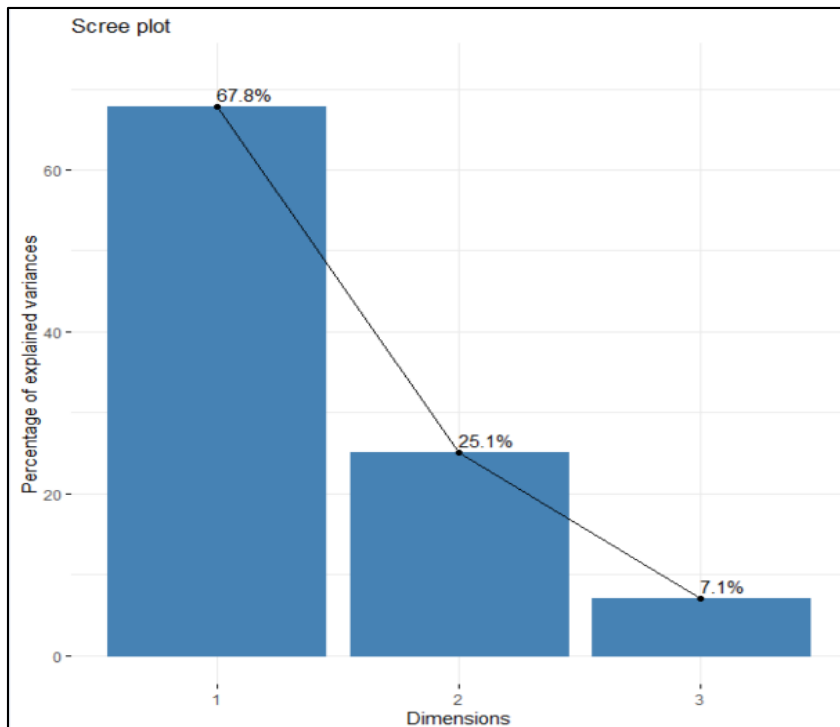


Ilustración 28 Variación de los datos por dimensiones



En la ilustración 29, se establece la asociación por cursos y cargos por persona, la cual muestra las inclinaciones de los clientes hacia los cursos que han llevado en función a los cargos que desempeñan.

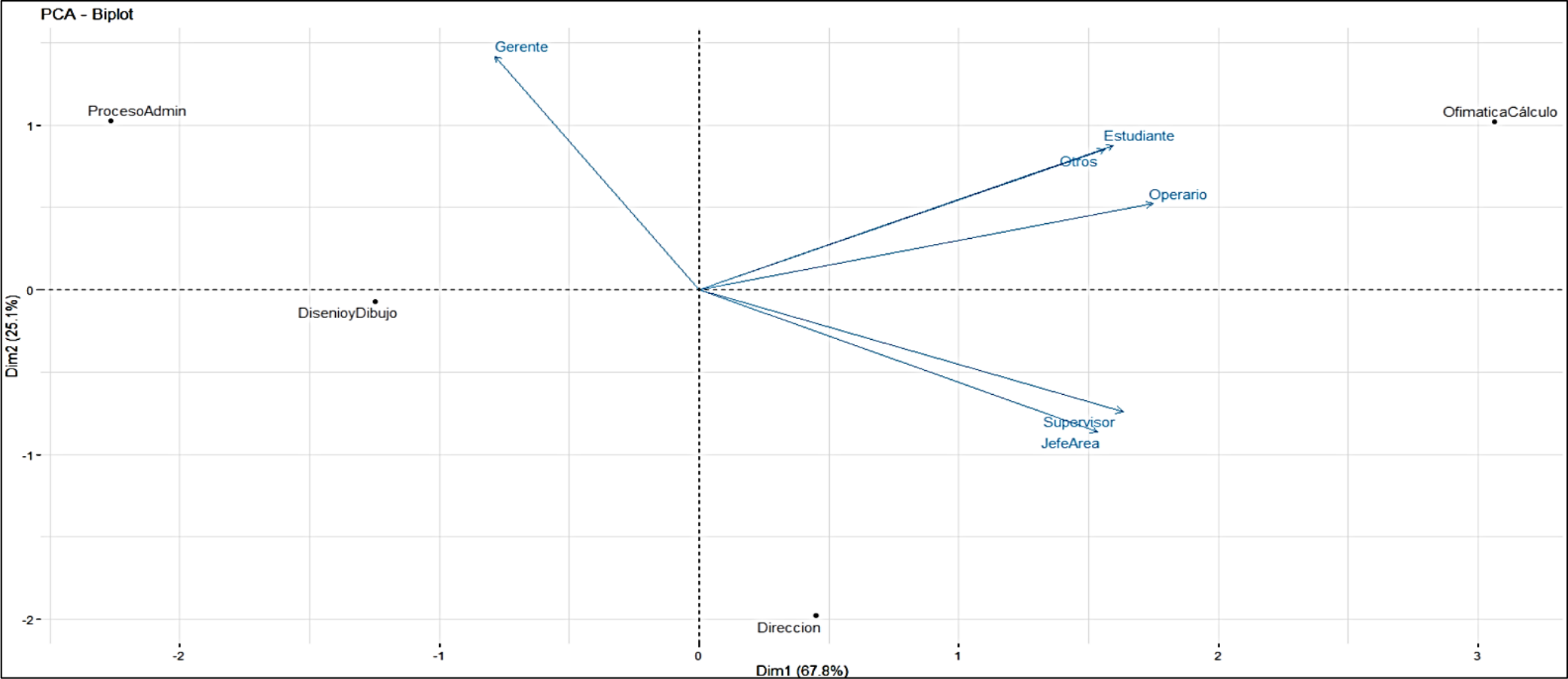


Ilustración 29 Definición de asociación y clúster

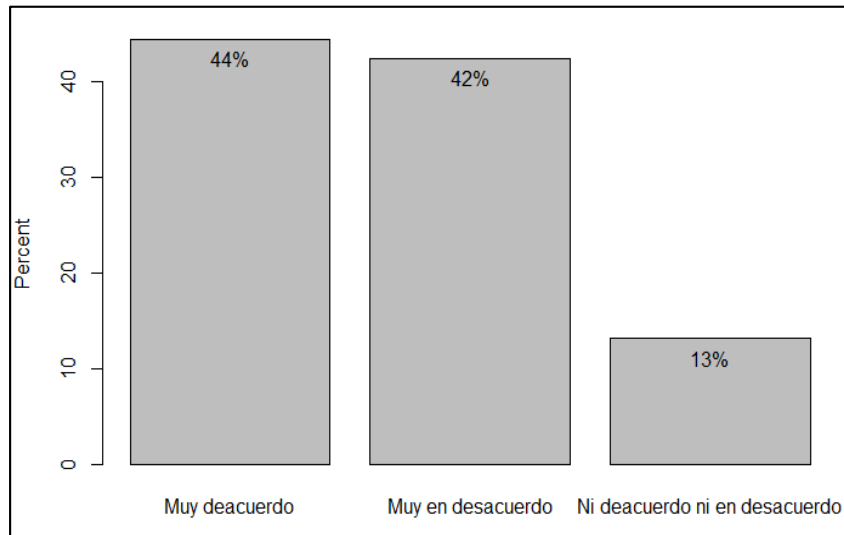


Ilustración 30 Estrategia corporativa

Con respecto a los resultados obtenidos en la encuesta, se hizo un análisis de percentiles utilizando tres grupos como escala valorativa. En la ilustración 26 se obtuvo que el 44% de los encuestados establecen que están de acuerdo con las estrategias corporativas que tiene la empresa en base a cada uno de sus indicadores (Crecimiento, Diferenciación, Entorno y planeamiento), mientras que el 42% indica que está en desacuerdo y un 13% se encuentra en un punto medio, ni de acuerdo ni en desacuerdo.

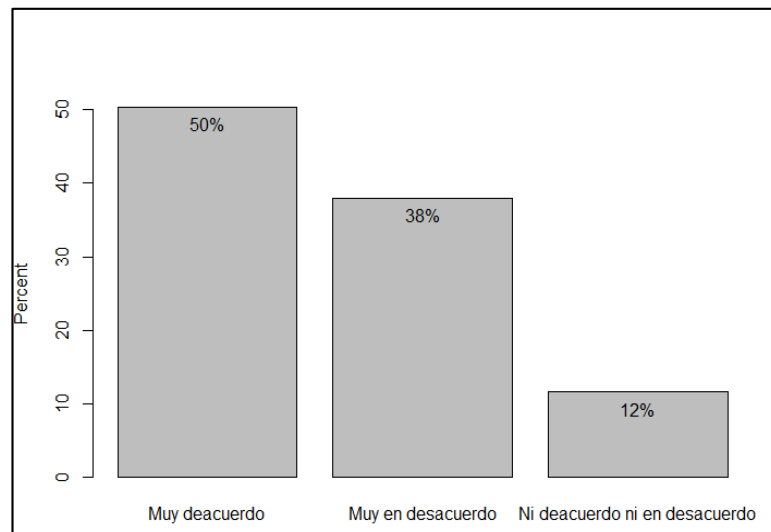


Ilustración 31 Estrategia competitiva

En la ilustración 27 se obtuvo que el 50% de los encuestados establecen que están de acuerdo con las estrategias competitivas que tiene la empresa en base

a cada uno de sus indicadores (Imagen de marca, Posicionamiento, Innovación y Tecnología), mientras que el 38% indica que está en desacuerdo y un 12%

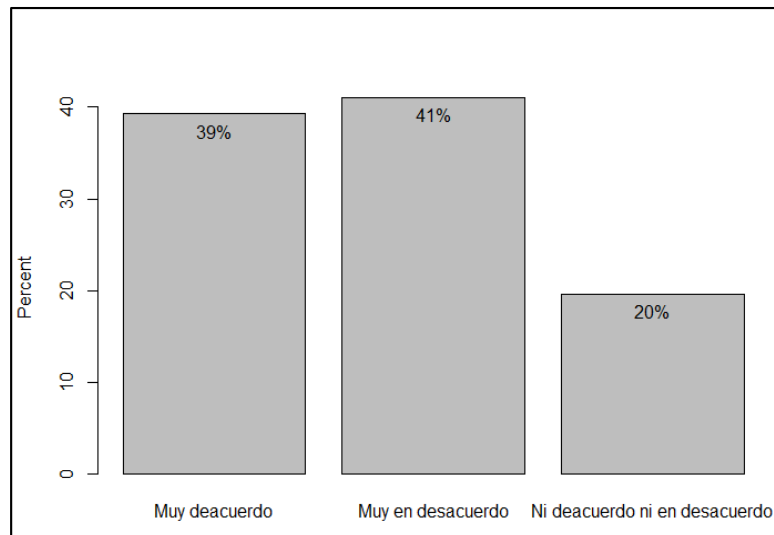


Ilustración 32 Estrategia Funcional

se encuentra en un punto medio, ni de acuerdo ni en desacuerdo.

En la ilustración 28 se obtuvo que el 39% de los encuestados establecen que están de acuerdo con las estrategias funcionales que tiene la empresa en base a cada uno de sus indicadores (Finanzas, Marketing y Dirección General), mientras que el 41% indica que está en desacuerdo y un 20% se encuentra en un punto medio, ni de acuerdo ni en desacuerdo.

## V. DISCUSIÓN

Con referencia al objetivo definir la comprensión del negocio y de los datos en el análisis de asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes, se definió la comprensión desde una perspectiva empresarial en el contexto de pandemia debido al COVID-19 en los años 2020 y 2021, considerando la metodología CRISP-D, entendiendo el negocio y plasmando objetivos técnicos considerando las inscripciones como el proceso más relevante donde se mide efectividad de las estrategias, los ingresos de dinero y la productividad de los colaboradores. Para plasmar los objetivos del negocio con los datos, se describió las variables más relevantes para alcanzar lo planteado en la investigación con la variedad de registros existentes en el datawarehouse de la institución. Luego se realizó un estudio de las bases de datos, la calidad de los datos para comprobar su validez de análisis, el tipo de datos o sub conjunto de datos que almacena con el fin de diversificar los cursos de acorde el perfil de los clientes.

Estos resultados son coherentes con lo planteado por De Lucca (2017), que basado en las reglas de asociación y clasificación; debió crear una base de datos para aplicación de data mining al igual que la presente investigación. La información fue seleccionada y se caracterizó los datos en atributos. Esto le sirvió para poder recolectar 105 publicaciones que cumplieran con los requisitos para su posterior limpieza de los datos. Asimismo, el autor describe los 20 atributos y define cada uno de ellos en los resultados en coherencia con lo desarrollado en la investigación actual. Asimismo, concuerda con lo realizado por Casariego Sarasquete (2019) en la comprensión del negocio, considerando al nuevo consumidor (cliente, ciudadano, paciente) y en la comprensión de los datos se remite al concepto de perfil del cliente 360, considerada como una visión integral 360 grados del perfil del cliente. También con los resultados de Arévalo Rodríguez, y otros (2016), pues en sus resultados afirma que la información estaba en una forma compleja, con muchas columnas que indicaban los valores de consumo de acuerdo al tema de publicación, ciudad y país y tuvo que rediseñarse de acuerdo a las necesidades de la investigación.

Asimismo, se relacionan con los resultados de Casariego Sarasquete (2019) que plantea una novedosa metodología de segmentación de clientes denominada

MSC2, que consiste en aplicación selectiva de las entidades de comportamiento que se generan con los pasos relacionados con la investigación comprensión del negocio, los datos, preparación de los datos, modelado analizando en ellos su capacidad de clasificación y agrupamiento. Compuestos por uno o más de los componentes: secuencias de comportamiento, tópicos de secuencias, reglas predictivas de secuencia, patrones de comportamiento, tópicos de patrones y reglas predictivas de patrones; que considera como relevantes fases tomadas en cuenta en la actual investigación.

En cuanto a la preparación los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes, se empezó con la selección de las tablas, considerando un datawarehouse de 57 tablas que contiene la información de los diferentes procesos que se ejecutan en la empresa, se decidió separar en un datamart de las inscripciones de la empresa, dicho datamart consta de 11 tablas. Se procedió a eliminar registros vacíos, registros datos que no tenían concordancia con los campos que los contenían, así como registros duplicados. Al momento de integrar los datos se aplicó inteligencia de negocios mediante la metodología del copo de nieve.

Estos resultados son coherentes con el estudio de De Lucca (2017), pues para la limpieza y el pre-procesamiento de los datos procedió a eliminar aquellos hashtags que no aportan ninguna información y que no pueden ser categorizados en los atributos. Como segundo paso unificó los valores que hacían referencia a un mismo atributo y así hasta obtener una base de datos uniforme. Por otro lado, en publicaciones donde no contenía un valor para cada atributo, se completó con el valor "not-avail", siendo necesario pues todos los atributos debían contener un valor, no pudiendo estar en blanco. Sin embargo, es necesario destacar que al momento de analizar la base de datos con las técnicas de data mining, se excluirán aquellas relaciones que mencionen alguno de los valores no disponibles; tal como debió realizarse en la presente investigación.

En lo relacionado al modelado de los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes, al momento de integrar los datos se aplicó inteligencia de negocios mediante la metodología del copo de nieve, que mediante el análisis se obtuvo 12 tablas: Genero, Profesión,

Cargo, área, Persona, Curso\_tipo, Curso\_categoría, Curso, Estado\_inscripciones, Modalidad, Inscripciones y la tabla tiempo. Finalmente se obtuvo la Tabla Hechos, con la unión de todas las tablas y a la vez se construyó una tabla temporal, reflejo de la tabla Hechos con un total de 5322 registros. Después de aplicar el modelamiento de datos, analizar la tabla hechos, se definen 2 modelos, los cuales están compuestos por una tabla de 2 por 2 (filas y columnas) O también llamada una matriz de doble entrada, donde los valores a estudiar en cada uno de los modelos están enfocados al registro de inscripciones para un análisis por categoría de cursos/meses y otro por el cargo laboral de la persona/meses

Estos resultados se relacionan con los de la investigación de Casariego Sarasquete (2019) quien representa un modelo que cumpla con propiedades como generalización, cronología y portabilidad; así como incluir capítulos de información estratégica y táctica del cliente como transacciones, interacciones, comunicaciones, métricas, segmento, contexto y semántica. Para el modelo de datos considera 03 tablas de hechos: eventos y acciones, métricas y segmentos y contexto.

En cuanto al objetivo evaluar los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes para asociar los valores a cada registro, se consideraron 02 modelos, en el primero para conocer la asociación entre las categorías de los cursos, con el software R, utilizando el PCA representado a través del mapa de dimensiones, con el algoritmo de K-Means, que es un método que actualiza dinámicamente los k prototipos de los datos, logrando maximizar la similitud dentro del clúster; con 02 clúster con una correlación parcialmente negativa, comprobada mediante correlación dimensional. Esto demostró que mientras las categorías de cursos: “Diseño y Dibujo, Procesos administrativos y Ofimática de cálculo avanzan en un mismo sentido o dirección independiente de su fuerza, la “Dirección” tiende a ir en sentido contrario (jalando dimensionalmente) ubicado en primer cuadrante, formado por las categorías “DisenioDibujo”, “OfimaticaCalculo”, “ProcesoAdmin” y “Dirección” en una segunda dimensión. En el segundo modelo se necesita conocer la asociación entre los cargos de sus clientes con la finalidad de establecer un análisis del estado actual de la base según el perfil, se logró 04 clúster con una correlación parcialmente

positiva, comprobada por la correlación dimensional formado por las categorías “jefe de área”, “operario”, “supervisor” y “Gerente” en una segunda dimensión.

Asimismo, en los datos de la encuesta a los clientes, se hizo un análisis de percentiles utilizando tres grupos como escala valorativa, donde el 44% están de acuerdo con las estrategias corporativas de los indicadores (Crecimiento, Diferenciación, Entorno y planeamiento), mientras que el 42% indica que está en desacuerdo y un 13% se encuentra en un punto medio, ni de acuerdo ni en desacuerdo. En cuanto a estrategias competitivas, el 50% de los encuestados establecen que están de acuerdo de sus indicadores (Imagen de marca, Posicionamiento, Innovación y Tecnología), mientras que el 38% indica que está en desacuerdo y un 12% se encuentra en un punto medio, ni de acuerdo ni en desacuerdo. Finalmente, en cuanto a estrategias funcional, el 39% de los encuestados establecen que están de acuerdo con las estrategias funcional de sus indicadores (Finanzas, Marketing y Dirección General), mientras que el 41% indica que está en desacuerdo y un 20% se encuentra en un punto medio, ni de acuerdo ni en desacuerdo.

Estos hallazgos son coherentes con el estudio de Walid Ghobar (2017), donde en sus resultados utilizando k-means gestionó la cantidad de datos del dataset que contenía los datos, para ello se eliminó columnas que no eran necesarias. Se presentaron errores asociados al volumen de datos siendo necesario la eliminación de algunos registros que no aportaban algún peso en las recomendaciones del perfil. En el resultado con k-means, con un  $K=6$  para agrupación de perfiles de usuarios, utilizó como referencia los datos del grafo del método del codo, probando desde 1 hasta 10 clusters. Con  $K=2$  resulta aun variación en la distorsión considerando las  $K$  próximas; con  $K=6$  confianza entre 30% y 40% indican una gran oportunidad de ofrecer recomendaciones para todos los grupos. Y en ese mismo contexto se tiene 02 experimentos, el primero tiene 02 perfiles, el primero con 371 usuarios, es decir el 65.32%, el perfil 2 con con 197 usuarios es decir 34.68% con recomendaciones desde 1 a 5 items; y el segundo con un perfil 1 con 275 usuarios, al 48,42% y perfil 02 con 293 (51,58%) usuarios sin recomendaciones. Asimismo se relaciona con la investigación de Laura Ochoa (2016) quien en sus resultados plantea que en las técnicas no supervisadas, se debe seleccionar una técnica que

minimice la distancia intra-cluster(cohesión) y maximice la distancia inter-cluster (separación) y para evaluar la calidad de los clusters el coeficiente silueta (silhouette) el mismo que sus valores están entre 0 y 1, más cercano a 1 es mejor. Si es negativo representa una mala agrupación; y se puede obtener con el algoritmo kmeans, lo cual coincide en parte con el algoritmo utilizado en la presente investigación pues además manifiesta que la segmentación de alumnos puede mejorar el proceso de enseñanza-aprendizaje mediante cursos de nivelación especializada, tal como resulta la diversificación de cursos planteada en la presente investigación.

También guarda coherencia con la investigación de Arévalo Rodríguez, y otros (2016), según los resultados concluyen que los Clúster agrupan los distintos datos por la proximidad de sus valores o por las similitudes entre ellos, esto es distinto de un clúster a otro por lo que es necesario realizar una comparación entre ellos para determinar patrones de consumo. No basta con elegir el valor de K, sino analizar la variación en la distorsión considerando las K próximas al clúster en referencia. Y finalmente es coherente con la investigación de Sigüenza Canale (2018), quien concluye que del análisis realizado plantea que una alternativa para incrementar los ingresos económicos de la microempresa es implementar una estrategia de diversificación de servicios, tomando en cuenta que los servicios actuales son limitados y sus ingresos económicos dependen mucho de la asistencia de estudiantes. De la misma manera, los resultados evidencian que las empresas dedicadas a la prestación de servicios educativos deben adaptarse a las nuevas tendencias, donde la aplicación de una estrategia de diversificación óptima permitiría mejorar la captación de nuevos clientes, y por ende la generación de ingresos económicos, sustanciales para mantenerse en el mercado competitivo.



## VI. CONCLUSIONES

- a) Se definió una estrategia de análisis desde una perspectiva empresarial fundamentado en la metodología CRISP-D, describiendo las variables basado en el entendimiento y comprensión del negocio y de los datos seleccionando los atributos de un datawarehouse de 57 tablas de la información de la empresa acerca de los cursos de capacitación los clientes.
- b) Se preparó los datos mediante clustering para la diversificación de cursos de considerando el perfil de los clientes, separándolos en un datamart con las tablas de hechos ACCOM Perú, persona, genero, profesión, cargo, área, curso, tipocurso, inscripciones, tipoinscripcion, modalidad y tiempo con 52 atributos.
- c) Se modelo los datos para la diversificación de cursos de acuerdo al perfil de los clientes, mediante la metodología copo de nieve con 12 tablas para una tabla de hechos con 5322 registros, obteniéndose 03 modelos tanto para la diversificación de cursos como para el cargo laboral del cliente.
- d) Se evaluó los datos mediante clustering para la diversificación de cursos de acuerdo al perfil de los clientes con 03 modelos, el primero con 02 clúster con una correlación parcialmente negativa, el segundo con 04 clúster con una correlación parcialmente positiva y el tercero con 03 clúster con una correlación parcialmente negativa, comprobada por la correlación dimensional. Y en los datos de la encuesta, el 44% están de acuerdo con las estrategias corporativas (Crecimiento, Diferenciación, Entorno y planeamiento), el 50% están de acuerdo con las estrategias competitivas (Imagen de marca, Posicionamiento, Innovación y Tecnología) y finalmente el 39% están de acuerdo con la estrategia funcional (Finanzas, Marketing y Dirección General).
- e) Se concluye que aplicando la metodología CRISP-DM para el análisis de asociación y clustering se logra realizar la diversificación de cursos de acuerdo al perfil de los clientes en la empresa ACCOM Perú SA.

## VII. RECOMENDACIONES

- a) Para futuras investigaciones se recomienda plantear una reingeniería a la base de datos de la empresa, reorganizándolos con el fin de obtener información de calidad para generar mejores modelos.
- b) Se recomienda considerar nuevos modelos con atributos mas cohesionados que permitan optimizar el grado de confiabilidad en la asociación y clustering de los modelos obtenidos.
- c) Se recomienda en futuras investigaciones considerar nuevas estrategias de diversificación de cursos, considerando datos como el interés de los clientes que permita a las tecnicas de mineria adaptarse a la necesidades de la empresa.
- d) Se sugiere en otras investigaciones considerar articular con otras bases de datos como redes sociales u empresas que se posean convenios interinstitucionales para ampliar el potencial de busqueda de asociaciones que permitan ampliar la toma de decisiones.
- e) Se sugiere en investigaciones futuras considerar paneles de control(dashboard) relacionados con modelos, productos de investigación de esta naturaleza, para favorecer la toma de decisiones en tiempo real.
- f) Se recomienda integrar en nuevas investigaciones nuevos algoritmos que optimice la cantidad de grupos en la que se segmentará los datos para las diferentes técnicas de clustering.

## REFERENCIAS

**Abdi y Williams. 2010.** *Correlación entre variables.* 2010.

**Acuña Gaitán, Marbell Del Socorro, Tatiana Palma, Leila y Cesar Huete, Julio. 2015.** *Desarrollo de un modelo basado en minería de datos aplicando el algoritmo de clúster en el área de mercadeo y ventas de la empresa DISPROCEL S.A (distribuidora de productos móviles) en el II semestre del año 2015.* 2015.

**Alvarez Pablo, Jessica Patricia, Medina Torre, Luis Eduardo y Sifuentes Álvarez, Luz Tania. 2019.** *Estrategia de diversificación del servicio y capacidad de.* Lima : s.n., 2019.

**Angarita García, David Alberto y Muñoz Franco, Juan José. 2016.** *Búsqueda de patrones en el comportamiento de los visitantes de la plataforma "Oferto" de la cámara de comercio de Armenia y del Quindío, a través de la aplicación de minería web.* Manizales, Colombia : s.n., 2016.

**Apolaya Torres, Carlos Humberto y Espinosa Diaz, Adolfo. 2018.** *TÉCNICAS DE INFERENCIAS, PREDICCIÓN Y MINERÍA DE DATOS.* Lima, Perú : s.n., 2018.

**Arcos Méndez, Pedro. 2020.** *Aplicación de minería de datos para pronóstico de ventas.* México : s.n., 2020.

**Arévalo Rodríguez, Tania y Noriega Coronel, Mario. 2016.** *Minería de datos para la identificación de patrones de consumo de servicios turísticos para orientar la oferta en el sector hotelero de la ciudad de Trujillo.* Trujillo : s.n., 2016.

**Arias, Fidias G. 2016.** *El proyecto de investigación.* Brazil : Edotorial Episteme, 2016.

**Arteaga Navarrete, Diana, Remigio Mondragón, Rosalba y Calderón Nepamuceno, Dora María. 2018.** *Minería de datos aplicado al marketing.* México : s.n., 2018.

**Arteaga Tigua, Alexis Josue. 2017.** *Implementación de una aplicación web aplicando minería de datos para el fortalecimiento de las ofertas hoteleras en la parroquia Puerto Cayo del Cantón Jipijapa.* Ecuador : Universidad Estatal del srur de Manabi, 2017.

**Barrios Zarta, Jairo. 2019.** *Impacto de la innovación en la competitividad de las empresas del sector agroindustrial, en el Espinal-Tolima.* s.l. : Revista Innova, 2019.

**Bernal Payares, Omaira. 2018.** *Planeación estratégica y sostenibilidad corporativa.* 2018.

**Bernal-Jiménez, M. C. y Rodríguez-Ibarra, D. L. 2019.** *Las tecnologías de la información y comunicación como factor de innovación y competitividad empresarial.* 2019. ISSN 0122-1701.

**BETANCUR OSPINA, ANDRÉS. 2017.** *GENERACIÓN DE CONOCIMIENTO EN MERCADEO A TRAVÉS DE LA IDENTIFICACIÓN DE PATRONES DE CONSUMO APOYADO EN MINERÍA DE DATOS. CASO DE ESTUDIO: SOMOS JEANS SAS CON LA MARCA SQUARE LINE JEANS.* Manabi, Ecuador : s.n., 2017.

**Carpio Maraza, Amira, y otros. 2019.** *Estrategias del marketing viral y el posicionamiento de marca en los restaurantes turísticos de la Región de Puno.* 2019.

**Casariego Sarasquete, Nicolás Martín. 2019.** *Metodología de análisis y segmentación de clientes usando secuencias de comportamiento.* Madrid, España : UNIR, 2019.

**Cevera Negueruela, Diego. 2019.** *Estrategia Corporativa: Fusiones y Adquisiciones.* 2019.

- CIP, Colegio de Ingenieros del Perú. 1987.** CIP. [En línea] 1987.  
[http://www.cip.org.pe/publicaciones/reglamentosCNCD2018/codigo\\_de\\_etica\\_del\\_cip.pdf](http://www.cip.org.pe/publicaciones/reglamentosCNCD2018/codigo_de_etica_del_cip.pdf).
- CONTRERAS ARTEAGA, ABEL ISAIAS y SANCHEZ COTRINA, FRANK WILLIAM. 2019.** *Analítica predictiva para conocer el patrón de consumo de los clientes en la empresa CIENPHARMA S.A.C. utilizando Ibm Spss Modeler y la metodología CRISP-DM.* Ttrujillo : s.n., 2019.
- Contreras Ortiz, Pedro Sebastián y Mora Alvarez, Milton Xavier. 2018.** *Aplicación de técnicas de clustering y detección de anomalías para la clasificación de consumos de energía eléctrica.* Cuenca, Ecuador : s.n., 2018.
- Cruz, C., Olivares, S. y González, M. 2015.** *Metodología de la investigación.* . México D.F. : Grupo Editorial Patria, 2015.
- De Lucca, Magdalena. 2017.** *Data mining y turismo: Un enfoque integral para potenciar la imagen de marca Argentina*". Argentina : s.n., 2017.
- Esteban Nieto, Nicomedes Teodoro. 2015.** *Tipos de Investigación.* 2015.
- Estévez Cárdenas, Juan Camilo. 2016.** *Análisis y predicción del comportamiento de la operación de transporte de carga automoto usando técnicas de minería de datos.* Colombia : s.n., 2016.
- Fidias G., Arias. 2016.** *El proyecto de investigación.* 2016. 7ma Edición.
- Folgado-Fernández, José Antonio, Duarte, Paulo Alexandre y Hernández-Mogollón, José Manuel. 2019.** *Influencia de internet versus medios tradicionales sobre la imagen de marca ciudad.* 2019.
- Galarreta Vásquez, Josué. 2016.** *Inducción de reglas de asociación de minería de datos en base de datos de entidad RETAIL.* Lambayeque, Perú : s.n., 2016. ISSN 2313-1926.
- Gimenez, Yanina. 2010.** *Clasificación no supervisada: El método de k-medias.* Buenos Aires, Argentina : s.n., 2010.
- Goleman, Daniel. 2018.** *Inteligencia emocional en la empresa.* s.l. : Conecta, 2018.
- González González, José Leandro, Mar Cornelio, Omar y Puid Díaz, Pedro M. 2017.** *Sistema para la generación de reglas de clasificación utilizando algoritmo de programación genético.* s.l. : Científica de la Universidad de las Ciencias Informáticas, 2017. 2306-2495.
- Grández Márquez, Miguel Angel. 2017.** *Aplicación de minería de datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos nutricionales.* Lima, Perú : s.n., 2017.
- Hernández Escobar, Arturo Andrés, y otros. 2018.** *Metodología de la Investigación Científica.* 2018. 978-84-948257-0-5.
- Hernández Sampieri, Roberto, y otros. 2017.** *Fundamenstos de Investigación.* México : Mc Graw Hill Education, 2017. 978-607-15-1395-3.
- Hernández-Sampieri, R., Fernández-Collado, R. y Baptista-Lucio, P.,. 2017.** *Selección de la muestra.* 2017.
- Herrera Castrillo, Cliffor Jerry. 2019.** *Estadística y probabilidades.* s.l. : Universidad Nacional Autónoma de Nicaragua, 2019.

- Herrera Meza, María Del Pilar. 2017.** *Preferencia de los turistas extranjeros Millennials por AIRBNB en Lima, Perú.* Lima, Perú : s.n., 2017.
- INCP. 2021.** INCP. *Definición y principales tipos de diversificación empresarial.* [En línea] 2021. <https://incp.org.co/definicion-y-principales-tipos-de-diversificacion-empresarial/>.
- Jollife, Jackson y Sommers. 2020.** *Análisis de Componentes Principales.* 2020.
- Lago Portela, Iria. 2020.** *Técnicas de formación de grupos: Métodos de particionamiento.* 2020.
- Laura Ochoa, Leticia Marisol. 2016.** *ESTUDIO COMPARATIVO DE TÉCNICAS NO SUPERVISADAS DE MINERÍA DE DATOS PARA SEGMENTACIÓN DE ALUMNOS.* Arequipa, Perú : s.n., 2016.
- Malpica Castro, Francisco Luis. 2019.** *Módulo de detección de patrones de consumo a partir del análisis de clientes de servicios de facturación electrónica.* Veracruz, México : s.n., 2019.
- Melo Chura, Alcides Demetrio. 2018.** *Patrones para la estimación de consumo de medicamentos con minería de datos redes puno.* Puno, Perú : s.n., 2018.
- Mendoza Bernedo, Juan Francisco y Anchiraico Bernaola, Wilder Raúl. 2018.** *Determinación de patrones de ventas en boticas independientes para mejorar las ventas.* Lima, Perú : s.n., 2018.
- Morales Ospina, Alberto. 2019.** *El servicio al cliente como estrategia competitiva.* 2019.
- Núñez Blas, Pilar Sara. 2016.** *Utilización de técnicas de minería de datos para la identificación de rasgos de comportamiento en procesos de aprendizaje colaborativo en modelos de e-learning y b-learning.* Huaraz, Perú : s.n., 2016.
- Obando Velásquez, Daniel André. 2017.** *Estudio y análisis de entornos comerciales mediante la evaluación, comparación y experimentación de algoritmos de minería de datos.* Arequipa, Perú : s.n., 2017.
- Ordaz Sanz, José Antonio, Melgar Hiraldo, María del Carmen y Rubio Castaño, Carmen María. 2016.** *Métodos estadísticos y econométricos en la empresa y para finanzas.* s.l. : Universidad Pablo de Olavide, 2016.
- Ortiz Neira, Karina Alejandra. 2016.** *Análisis de agrupamiento de datos para el apoyo al diagnóstico de la tuberculosis.* Colombia : s.n., 2016.
- Palacios Rodríguez, Miguel Ángel. 2020.** *Planeación Estratégica, instrumento funcional al interior de las organizaciones.* 2020.
- Pérez Gorostegui, Eduardo. 2017.** *Curso de economía de la empresa: Introducción.* Madrid, España : Editorial Universitaria Ramón Areces, 2017.
- Puga Reyes, Eduardo Xavier. 2020.** *Análisis de algoritmos de Clustering basados en particionamiento y basados en Densidad para el procesamiento de trayectorias GPS.* Guayaquil, Ecuador : s.n., 2020.
- Quispe Durán, Juan Bautista. 2018.** *Impacto del modelo de minería de datos en el pronóstico de ventas de la empresa Cellservice Eirl. en el periodo 2012-2016.* Cajamarca, Perú : s.n., 2018.
- Reyes García, Hugo Ricardo. 2018.** *Diversificación e integración vertical en una empresa familiar.* 2018.

- Rioja Curo, Walter Miguel. 2020.** *Aplicación web para la elaboración de perfiles de consumidor basada en minería de datos y arquitectura cloud para el apoyo al proceso de conversión de Leads en la Asociación AIESEC en Perú.* Chiclayo, Perú : s.n., 2020.
- Riquelme Lafertte, Michael. 2017.** *Mejora del proceso de evaluación de nuevos convenios mediante modelos de predicción.* Santiago, Chile : s.n., 2017.
- Rodríguez García, José Antonio. 2018.** *Análisis de la aplicación de la estrategia de crecimiento por diversificación y el estado de resultados en el grupo empresarial.* Juliaca, Perú : s.n., 2018.
- Rojas Gómez, Jéssica Dayana. 2020.** *Patrones de comportamiento de los usuarios de la empresa Adonai Employment utilizando minería de datos.* Bucaramanga, Ecuador : s.n., 2020.
- Rouhiainen, Lasse. 2021.** *Inteligencia artificial para los negocios. 21 casos prácticos y opiniones de expertos.* 2021.
- Satria, Abadi, y otros. 2018.** *Application model of k-means clustering: insights into promotion strategy of vocational high school.* s.l. : International Journal of Engineering & Technology, 2018.
- Schiatti Sisó, Laura Cristina. 2017.** *Estudio comparativo de diferentes algoritmos de clustering para la estimación de grupos de evaluados que comparten debilidades conceptuales similares.* Carategena de Indias, Colombia : s.n., 2017.
- Sigüenza Canales, Kelly José. 2018.** *Propuesta de una estrategia de diversificación de servicios para la micro empresa pequeños genios.* Guayaquil, Ecuador : s.n., 2018.
- Sinaga, K. P. y Yang, M. 2020.** *Unsupervised K-Means Clustering Algorithm.* 2020.
- SNGULAR. 2021.** <https://www.sngular.com/>. [En línea] 2021. <https://www.sngular.com/es/data-science-crisp-dm-metodologia/>.
- Solano Torrenegra, Ernesto. 2018.** *¿Cuanta razón tiene el cliente?* 2018. 9781549993732.
- Suárez Aldunate, Paulina Teresa. 2017.** *Diseño de un sistema de pronóstico para una plataforma de marketing móvil geolocalizado.* Santiago, Chile : s.n., 2017.
- Toledo, Luciano, Augusto, Garber, Marcos, Fernando y de Farias Shiraishi, Guilherme. 2017.** *Notas críticas sobre las Estructuras Organizacionales y el Marketing.* 2017.
- Turiano, Julieta. 2019.** *Exclusividad en la diferenciación.* 2019.
- UCV, Universidad César Vallejo. 2020.** [ucv. ucv.edu.pe](https://www.ucv.edu.pe/wp-content/uploads/2020/11/RCUN%2%B00262-2020-UCV-Aprueba-Actualizaci%C3%B3n-del-C%C3%B3digo-%C3%89tica-en-Investigaci%C3%B3n-1-1.pdf). [En línea] 28 de Agosto de 2020. <https://www.ucv.edu.pe/wp-content/uploads/2020/11/RCUN%2%B00262-2020-UCV-Aprueba-Actualizaci%C3%B3n-del-C%C3%B3digo-%C3%89tica-en-Investigaci%C3%B3n-1-1.pdf>.
- Vazquez Zavaleta, Miguel Angel. 2018.** *Minería de datos para generación de reglas de tendencia de precipitación pluvial en el estado de Morelos.* México : s.n., 2018.
- Venkat Ram, Sudarshan. 2017.** *A Data Mining Approach to Modeling Customer Preference: A Case Study of Intel Corporation.* Arizona, Estados Unidos : s.n., 2017.
- Walid Ghobar, Enio. 2017.** *Un sistema de recomendación basado en perfiles generados por agrupamiento y asociaciones.* España : s.n., 2017.

**Yturregui Valdivia, Lesvi Marly. 2018.** *Estrategia de ventas para incrementar la captación de alumnos en la escuela de Posgrado Uss Chiclayo.* Pimentel, Piura : s.n., 2018.

## ANEXOS

### ANEXO 01: Matriz de operacionalización:

VARIABLE	DEFINICION CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIÓN	INDICADOR	ÍTEMS	ESCALA MEDICIONAL	TECNICA / INSTRUMENTOS
Metodología CRISP-DM para el análisis de asociación y Clustering	El análisis de información consiste en descubrir patrones, tendencias, características o comportamiento de los clientes en un repositorio de información	El análisis de información será realizado mediante la técnica de la observación, utilizando como instrumento las fichas de registro donde se medirá las dimensiones de comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implantación.	Comprensión del negocio y de datos	Descriptor de variables		Nominal	Observación / Fichas de registro
				Calidad de los datos			
			Preparación de los datos	Modelado de los datos			
			Modelado	Técnica de modelado			
			Evaluación	Evaluación óptima de clúster			
Diversificación de cursos	Es una estrategia que consiste en el lanzamiento de un nuevo producto o servicio o ingreso a un nuevo mercado alineados o no al rubro de la empresa.	Para la diversificación de cursos, la información se recolecta mediante la técnica de la encuesta dimensionada por la estrategia corporativa, competitiva y funcional	Estrategia corporativa	Crecimiento	1	Ordinal	Encuesta / Cuestionario
				Diferenciación	2		
				Entorno	3		
				Planeamiento	4		
			Estrategia competitiva	Imagen de marca	5		
				Posicionamiento	6		
				Innovación	7		
				Tecnología	8		
			Estrategia funcional	Finanzas	9		
					10		
				Marketing	11		
					12		
				Dirección general	13		



ANEXO 02: Matriz de consistencia:

Título: Metodología CRISP-DM para el análisis de asociación y Clustering para la diversificación de cursos de acuerdo al perfil de los clientes de la empresa ACCOM Perú SAC.					
Autores: Albuquerque Guaylupo, Hugo Enrique; Dongo Gutiérrez, Gian Carlos					
PROBLEMA	OBJETIVOS	VARIABLE	DIMENSIONES	INDICADORES	METODOLOGÍA
<b>General:</b> ¿Cómo se aplica el análisis de asociación y clustering mediante la Metodología CRISP-DM para la diversificación de cursos de acuerdo al perfil de los clientes en la empresa ACCOM Perú SAC?	<b>General:</b> Aplicar la Metodología CRISP-DM para el análisis de asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes en la empresa ACCOM Perú SAC.	<b>VD</b> Diversificación de cursos	Estrategia corporativa	Crecimiento	<b>METODOLOGÍA APLICADA:</b> <b>TIPO DE INVESTIGACION:</b> APLICADA  <b>DISEÑO DE INVESTIGACION:</b> NO EXPERIMENTAL  <b>POBLACIÓN, MUESTRA, MUESTREO</b> <b>Población:</b> La población está constituida por 5222 registros de capacitación de clientes y 384 profesionales y estudiantes de los últimos ciclos. <b>Muestra:</b> Aleatoria simple <b>Muestreo:</b> Probabilístico  <b>MÉTODO DE INVESTIGACIÓN</b> Descriptivo
<b>Específicos</b> ¿Qué comprensión del negocio y los datos se podrá definir en el análisis de asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes?	<b>Específicos</b> Definir la comprensión del negocio y de los datos en el análisis de asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes			Estrategia competitiva	
			Posicionamiento		
			Innovación		
¿Cómo se prepara los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes?	Preparar los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes		Estrategia funcional	Finanzas	
		Marketing			
¿Cómo se modela los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes?	Modelar los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes	<b>VI</b> Metodología CRISP-DM para el análisis de asociación y clustering	Comprensión del negocio y de datos	Descriptor de variables	
¿Cómo es la evaluación de los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes?	Evaluar los datos mediante asociación y clustering para la diversificación de cursos de acuerdo al perfil de los clientes			Preparación de los datos	
			Modelado		
			Evaluación	Técnica de modelado	
			Implantación	Evaluación óptima de clúster	
				Patrones de asociación	

ANEXO 03: FICHAS DE REGISTRO

Ficha de registro 01				
Investigador		Alburqueque Guaylupo, Hugo Enrique Dongo Gutiérrez, Gian Carlos		Tipo de prueba
Institución		ACCOM Perú SAC.		
Fecha de Inicio			Fecha final	
Variable		Indicador	Medida	Fórmula
<b>Metodología CRISP-DM para el análisis de asociación y Clustering</b>		<b>Descriptor de variables</b>		
#	Numero	Nombre	Tipo de dato	Descripción
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

Ficha de registro 02					
Investigador		Alburqueque Guaylupo, Hugo Enrique Dongo Gutiérrez, Gian Carlos		Tipo de prueba	
Institución		ACCOM Perú SAC			
Fecha de Inicio			Fecha final		
Variable		Indicador	Medida	Fórmula	
Metodología CRISP-DM para el análisis de asociación y Clustering		Calidad de los datos			
#	Registro	Valores nulos	Valores fuera de rango	Observación	
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Ficha de registro 03							
Investigador		Alburqueque Guaylupo, Hugo Enrique Dongo Gutiérrez, Gian Carlos			Tipo de prueba		
Institución		ACCOM Perú SAC					
Fecha de Inicio			Fecha final				
Variable		Indicador		Medida		Fórmula	
Metodología CRISP-DM para el análisis de asociación y Clustering		Modelado de los datos					
#	Código de estudiante	Campo1	Campo2	Campo3	Campo4	Campo5	Campo6
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

Ficha de registro 04						
Investigador		Alburqueque Guaylupo, Hugo Enrique Dongo Gutiérrez, Gian Carlos		Tipo de prueba		
Institución		ACCOM Perú SAC				
Fecha de Inicio			Fecha final			
Variable		Indicador	Medida	Fórmula		
<b>Metodología CRISP-DM para el análisis de asociación y Clustering</b>		Técnica de modelado				
#	Técnica	Algoritmo	Valor	Probabilidad		Acertividad
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						

Ficha de registro 05				
<b>Investigador</b>		Alburqueque Guaylupo, Hugo Enrique Dongo Gutiérrez, Gian Carlos		<b>Tipo de prueba</b>
<b>Institución</b>		ACCOM Perú SAC		
<b>Fecha de Inicio</b>		<b>Fecha final</b>		
<b>Variable</b>		<b>Indicador</b>	<b>Medida</b>	<b>Fórmula</b>
<b>Metodología CRISP-DM para el análisis de asociación y Clustering</b>		Evaluación óptima de clúster		
<b>#</b>	<b>Grupo</b>	<b>K</b>	<b>Centroide</b>	<b>Observación</b>
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

Donde:

**Grupo:** Es el clúster resultante del clustering

**K:** Es el número de veces (épocas) que el conjunto de entrenamiento es probado para encontrar valores óptimos de asertividad

**Centroide:** Es el valor central de cada uno de los clústeres

Ficha de registro 06					
Investigador		Alburqueque Guaylupo, Hugo Enrique Dongo Gutiérrez, Gian Carlos		Tipo de prueba	
Institución		ACCOM Perú SAC			
Fecha de Inicio			Fecha final		
Variable		Indicador	Medida	Fórmula	
Metodología CRISP-DM para el análisis de asociación y Clustering		Patrones de asociación			
#	Regla	Tipo de corridas	% Correctamente Clasificados	Tiempo	
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

ANEXO 04: CUESTIONARIO

**“Metodología CRISP-DM para el análisis de asociación y Clustering para la diversificación de cursos de acuerdo al perfil de los clientes de la empresa ACCOM Perú SAC”**

**OBJETIVO:** Aplicar la Metodología CRISP-DM para el análisis de asociación y Clustering para la diversificación de cursos de acuerdo al perfil de los clientes en la empresa ACCOM Perú SAC.

**INSTRUCCIONES:** Marque con una X la alternativa que usted considere valida de acuerdo al ítem en los casilleros siguientes:

MUY DE ACUERDO	ALGO DE ACUERDO	NI DE ACUERDO NI DESACUERDO	ALGO EN DESACUERDO	MUY EN DESACUERDO
MA	A	NN	D	MD

ITEM	PREGUNTA	VALORACION				
		MA	A	NN	D	MD
1	¿Está de acuerdo en que una empresa de capacitación debe implementar laboratorios de cómputo en sus instalaciones?	5	4	3	2	1
2	¿Está de acuerdo que una empresa de capacitación debería ofrecer asesorías a sus alumnos después de horas de clase?	5	4	3	2	1
3	¿Está de acuerdo que una empresa de capacitación debe brindar comodidad y seguridad al visitar sus instalaciones?	5	4	3	2	1
4	¿Está de acuerdo que una empresa de capacitación le brinde información de próximos cursos o eventos de capacitación?	5	4	3	2	1
5	¿Está de acuerdo que una empresa de capacitación debe proyectar variedad en sus cursos y otros servicios?	5	4	3	2	1
6	¿Está de acuerdo que para decidirse por una empresa de capacitación esta debe brindar cursos y servicios que otras empresas no tienen?	5	4	3	2	1



7	¿Está de acuerdo que una empresa de capacitación le brinde asesoramientos después de terminado el curso?	5	4	3	2	1
8	¿Está de acuerdo que una empresa de capacitación le brinde una plataforma virtual para fortalecer de los temas aprendidos en clase?	5	4	3	2	1
9	¿Está de acuerdo que una empresa de capacitación brinde facilidades de pago?	5	4	3	2	1
10	¿Está de acuerdo que una empresa de capacitación brinde descuentos especiales a ex alumnos?	5	4	3	2	1
11	¿Está de acuerdo que una empresa de capacitación realice lanzamientos de sus cursos por medio de las redes sociales?	5	4	3	2	1
12	¿Está de acuerdo que una empresa de capacitación utilice elementos humorísticos virales(mememes) en redes sociales para promocionar sus cursos?	5	4	3	2	1
13	¿Está de acuerdo que el director de capacitación interactúe con los clientes para solicitar críticas que ayuden a mejorar el servicio de capacitación?	5	4	3	2	1

ANEXO 05: Estadísticos de fiabilidad

Alfa de Cronbach      N de elementos  
0.88                      25

Donde:

Coeficiente de confiabilidad del cuestionario	0.88
Número de ítems del instrumento	13
Sumatoria de las varianzas de los ítems.	7.5744
Varianza total del instrumento.	40.9696

Fuente: Fuente propia

## ANEXO 06: Validación por evaluación de expertos

### CONSTANCIA DE VALIDACIÓN DE INSTRUMENTOS



### CONSTANCIA DE VALIDACIÓN DE FICHAS DE REGISTRO

Yo, Arnaldo Martín Alvarado con DNI N.º 43490242 con grado académico de Ingeniero de Sistemas, profesión Ing. de Sistemas desempeñándome actualmente como Analista de Seguridad de la Información y Confidencialidad del Negocio.

Por medio de la presente hago constatar que he revisado con fines de validación los instrumentos:

Fichas de registro del proyecto de tesis “**Metodología CRISP-DM para el análisis de asociación y Clustering para la diversificación de cursos de acuerdo al perfil de los clientes de la empresa ACCOM Perú SAC**”

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones

Ficha de registro N.º 01: Descriptor de variables	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad				✓
2. Objetividad				✓
3. Actualidad				✓
4. Organización				✓
5. Suficiencia				✓
6. Intencionalidad				✓
7. Consistencia				✓
8. Coherencia				✓
9. Metodología				✓

<b>Ficha de registro N.º 02: Calidad de los datos</b>	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad				✓
2. Objetividad				✓
3. Actualidad				✓
4. Organización				✓
5. Suficiencia				✓
6. Intencionalidad				✓
7. Consistencia				✓
8. Coherencia				✓
9. Metodología				✓

<b>Ficha de registro N.º 03: Modelado de los datos</b>	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad				✓
2. Objetividad				✓
3. Actualidad				✓
4. Organización				✓
5. Suficiencia				✓
6. Intencionalidad				✓
7. Consistencia				✓
8. Coherencia				✓
9. Metodología				✓

<b>Ficha de registro N.º 04: Técnica de modelado</b>	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad				✓
2. Objetividad				✓
3. Actualidad				✓
4. Organización				✓

5. Suficiencia				✓
6. Intencionalidad				✓
7. Consistencia				✓
8. Coherencia				✓
9. Metodología				✓

<b>Ficha de registro N.º 05:</b> Evaluación óptima de clúster	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad				✓
2. Objetividad				✓
3. Actualidad				✓
4. Organización				✓
5. Suficiencia				✓
6. Intencionalidad				✓
7. Consistencia				✓
8. Coherencia				✓
9. Metodología				✓

<b>Ficha de registro N.º 06:</b> Patrones de asociación	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad				✓
2. Objetividad				✓
3. Actualidad				✓
4. Organización				✓
5. Suficiencia				✓
6. Intencionalidad				✓
7. Consistencia				✓
8. Coherencia				✓
9. Metodología				✓

En señal de conformidad firmo la presente en la ciudad de Piura con fecha  
11-11-2021

**NOMBRE COMPLETO** : Arnaldo Martín Alvarado Jacliva  
**GRADO ACADÉMICO** : Ingeniero de Sistemas.  
**DNI** : 43490242  
**Especialidad** : Ingeniero de sistemas  
**E-mail** : amal@cajatrujillo.com.pe.

  
**FIRMA Y SELLO**

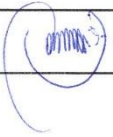
### CONSTANCIA DE VALIDACIÓN DE CUESTIONARIO

Yo, Arnaldo Martín Alvarado Jachín, con DNI N.º 43490242 con grado académico de Ingeniero de Sistemas, profesión Ing. de Sistemas desempeñándome actualmente como Analista de Seguridad de la Información y Continuidad del Negocio.

Por medio de la presente hago constatar que he revisado con fines de validación los instrumentos:

**Cuestionario de tesis “Metodología CRISP-DM para el análisis de asociación y Clustering para la diversificación de cursos de acuerdo al perfil de los clientes de la empresa ACCOM Perú SAC”**

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones:

Título de la investigación: "MEDOLOGIA CRISP-DM PARA LA DIVERSIFICACIÓN DE CURSOS DE ACUERDO AL PERFIL DE LOS CLIENTES DE LA EMPRESA ACCOM PERU SAC, PIURA 2021,"						
Apellidos y nombres del investigador: Dongo Gutiérrez Gian Carlos, Albuquerque Guaylupo Hugo Enrique						
Apellidos y nombres del experto:						
ASPECTO POR EVALUAR				OPINION DEL EXPERTO		
VARIABLES	DIMENSIONES	INDICADORES	ITEM DE LA PREGUNTA	SI CUMPLE	NO CUMPLE	OBSERVACIONES /SUGERENCIAS
DIVERSIFICACIÓN DE CURSOS	ESTRATEGIA CORPORATIVA	Crecimiento	¿Está de acuerdo en que una empresa de capacitación debe implementar laboratorios de cómputo en sus instalaciones?	✓		
		Diferenciación	¿Está de acuerdo que una empresa de capacitación debería ofrecer asesorías a sus alumnos después de horas de clase?	✓		
		Entorno	¿Está de acuerdo que una empresa de capacitación debe brindar comodidad y seguridad al visitar sus instalaciones?	✓		
		Planeamiento	¿Está de acuerdo que una empresa de capacitación le brinde información de próximos cursos o eventos de capacitación?	✓		
	ESTRATEGIA COMPETITIVA	Imagen de marca	¿Está de acuerdo que una empresa de capacitación debe proyectar variedad en sus cursos y otros servicios?	✓		
		Posicionamiento	¿Está de acuerdo que para decidirse por una empresa de capacitación esta debe brindar cursos y servicios que otras empresas no tienen?	✓		
		Innovación	¿Está de acuerdo que una empresa de capacitación le brinde asesoramientos después de terminado el curso?	✓		
		Tecnología	¿Está de acuerdo que una empresa de capacitación le brinde una plataforma virtual para fortalecer de los temas aprendidos en clase?	✓		
ESTRATEGIA FUNCIONAL	Finanzas		¿Está de acuerdo que una empresa de capacitación brinde facilidades de pago?	✓		
			¿Está de acuerdo que una empresa de capacitación brinde descuentos especiales a ex alumnos?	✓		
	Marketing		¿Está de acuerdo que una empresa de capacitación realice lanzamientos de sus cursos por medio de las redes sociales?	✓		
			¿Está de acuerdo que una empresa de capacitación utilice elementos humorísticos virales(memes) en redes sociales para promocionar sus cursos?	✓		
	Dirección general		¿Está de acuerdo que el director de capacitación interactúe con los clientes para solicitar críticas que ayuden a mejorar el servicio de capacitación?	✓		
Firma del experto: 			Fecha: 11 / 11 / 2021			



## CONSTANCIA DE VALIDACIÓN DE INSTRUMENTOS



## CONSTANCIA DE VALIDACIÓN DE FICHAS DE REGISTRO

Yo, LISSETH TEJADA ESPINOZA con DNI N.º 46120273 con grado académico de MAGISTER EN ADMINISTRACIÓN DE NEGOCIOS Y RELACIONES I., profesión INGENIERÍA DE SISTEMAS desempeñándome actualmente como ANALISTA DE INDICADORES DE MINISTERIO PÚBLICO.-SULLANA.

Por medio de la presente hago constatar que he revisado con fines de validación los instrumentos:

Fichas de registro del proyecto de tesis **“Metodología CRISP-DM para el análisis de asociación y Clustering para la diversificación de cursos de acuerdo al perfil de los clientes de la empresa ACCOM Perú SAC”**

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones

Ficha de registro N.º 01: Descriptor de variables	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			✓	
2. Objetividad			✓	
3. Actualidad				✓
4. Organización				✓
5. Suficiencia			✓	
6. Intencionalidad			✓	
7. Consistencia			✓	
8. Coherencia			✓	
9. Metodología				✓

<b>Ficha de registro N.º 02: Calidad de los datos</b>	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			✓	
2. Objetividad			✓	
3. Actualidad			✓	
4. Organización				✓
5. Suficiencia			✓	
6. Intencionalidad			✓	
7. Consistencia			✓	
8. Coherencia			✓	
9. Metodología				✓

<b>Ficha de registro N.º 03: Modelado de los datos</b>	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			✓	
2. Objetividad			✓	
3. Actualidad			✓	
4. Organización				✓
5. Suficiencia			✓	
6. Intencionalidad			✓	
7. Consistencia			✓	
8. Coherencia			✓	
9. Metodología				✓

<b>Ficha de registro N.º 04: Técnica de modelado</b>	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			✓	
2. Objetividad			✓	
3. Actualidad			✓	
4. Organización			✓	

5. Suficiencia			✓	
6. Intencionalidad			✓	
7. Consistencia			✓	
8. Coherencia			✓	
9. Metodología				✓

<b>Ficha de registro N.º 05:</b> Evaluación óptima de clúster	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			✓	
2. Objetividad			✓	
3. Actualidad			✓	
4. Organización			✓	
5. Suficiencia			✓	
6. Intencionalidad			✓	
7. Consistencia			✓	
8. Coherencia			✓	
9. Metodología				✓

<b>Ficha de registro N.º 06:</b> Patrones de asociación	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			✓	
2. Objetividad			✓	
3. Actualidad			✓	
4. Organización			✓	
5. Suficiencia			✓	
6. Intencionalidad			✓	
7. Consistencia			✓	
8. Coherencia			✓	
9. Metodología				✓

En señal de conformidad firmo la presente en la ciudad de Piura con fecha

---

**NOMBRE COMPLETO** : LISSETH ZULAY TEJADA ESPINOZA  
**GRADO ACADÉMICO** : MAGISTER EN ADMINISTRACIÓN DE  
NEGOCIOS Y RELACIONES INTERNACIONALES  
**DNI** : 46120273  
**Especialidad** : INGENIERÍA DE SISTEMAS  
**E-mail** : LTEJADAE@GMAIL.COM

  
**FIRMA Y SELLO**


### CONSTANCIA DE VALIDACIÓN DE CUESTIONARIO

Yo, LISSETH TEJADA ESPINORA con DNI N.º 46120273 con grado académico de MAGISTER EN ADMINISTRACIÓN DE NEGOCIOS Y RELACIONES F., profesión INGENIERÍA DE SISTEMAS desempeñándome actualmente como ANALISTA DE INDICADORES DE MINISTERIO PÚBLICO - SUYAMA.

Por medio de la presente hago constatar que he revisado con fines de validación los instrumentos:

Cuestionario de tesis **“Metodología CRISP-DM para el análisis de asociación y Clustering para la diversificación de cursos de acuerdo al perfil de los clientes de la empresa ACCOM Perú SAC”**

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones:

Título de la investigación: "MEDOLOGIA CRISP-DM PARA LA DIVERSIFICACIÓN DE CURSOS DE ACUERDO AL PERFIL DE LOS CLIENTES DE LA EMPRESA ACCOM PERU SAC, PIURA 2021,"						
Apellidos y nombres del investigador: Dongo Gutiérrez Gian Carlos, Alburqueque Guaylupo Hugo Enrique						
Apellidos y nombres del experto: TEJADA ESPINOZA, LISETH ZULAY						
ASPECTO POR EVALUAR				OPINION DEL EXPERTO		
VARIABLES	DIMENSIONES	INDICADORES	ITEM DE LA PREGUNTA	SI CUMPLE	NO CUMPLE	OBSERVACIONES /SUGERENCIAS
DIVERSIFICACIÓN DE CURSOS	ESTRATEGIA CORPORATIVA	Crecimiento	¿Está de acuerdo en que una empresa de capacitación debe implementar laboratorios de cómputo en sus instalaciones?	✓		
		Diferenciación	¿Está de acuerdo que una empresa de capacitación debería ofrecer asesorías a sus alumnos después de horas de clase?	✓		
		Entorno	¿Está de acuerdo que una empresa de capacitación debe brindar comodidad y seguridad al visitar sus instalaciones?	✓		
		Planeamiento	¿Está de acuerdo que una empresa de capacitación le brinde información de próximos cursos o eventos de capacitación?	✓		
	ESTRATEGIA COMPETITIVA	Imagen de marca	¿Está de acuerdo que una empresa de capacitación debe proyectar variedad en sus cursos y otros servicios?	✓		
		Posicionamiento	¿Está de acuerdo que para decidirse por una empresa de capacitación esta debe brindar cursos y servicios que otras empresas no tienen?	✓		
		Innovación	¿Está de acuerdo que una empresa de capacitación le brinde asesoramientos después de terminado el curso?	✓		
		Tecnología	¿Está de acuerdo que una empresa de capacitación le brinde una plataforma virtual para fortalecer de los temas aprendidos en clase?	✓		
ESTRATEGIA FUNCIONAL	Finanzas		¿Está de acuerdo que una empresa de capacitación brinde facilidades de pago?	✓		
			¿Está de acuerdo que una empresa de capacitación brinde descuentos especiales a ex alumnos?	✓		
	Marketing		¿Está de acuerdo que una empresa de capacitación realice lanzamientos de sus cursos por medio de las redes sociales?	✓		
			¿Está de acuerdo que una empresa de capacitación utilice elementos humorísticos virales(memes) en redes sociales para promocionar sus cursos?	✓		
	Dirección general		¿Está de acuerdo que el director de capacitación interactúe con los clientes para solicitar críticas que ayuden a mejorar el servicio de capacitación?	✓		
Firma del experto: 			Fecha: 11/11/2021			

**CONSTANCIA DE VALIDACIÓN DE INSTRUMENTOS**



**CONSTANCIA DE VALIDACIÓN DE FICHAS DE REGISTRO**

Yo, Teófilo P. Conza Calle con DNI N.º 028 20231 con grado académico de Maestría en Dirección y Gestión de TI, profesión Ingeniero en informática desempeñándome actualmente como Docente de la escuela de sistemas.

Por medio de la presente hago constatar que he revisado con fines de validación los instrumentos:

Fichas de registro del proyecto de tesis **“Metodología CRISP-DM para el análisis de asociación y Clustering para la diversificación de cursos de acuerdo al perfil de los clientes de la empresa ACCOM Perú SAC”**

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones

Ficha de registro N.º 01: Descriptor de variables	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			X	
2. Objetividad				X
3. Actualidad				X
4. Organización			X	
5. Suficiencia			X	
6. Intencionalidad				X
7. Consistencia			X	
8. Coherencia			X	
9. Metodología			X	

<b>Ficha de registro N.º 02: Calidad de los datos</b>	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			X	
2. Objetividad				X
3. Actualidad				X
4. Organización			X	
5. Suficiencia			X	
6. Intencionalidad			X	
7. Consistencia			X	
8. Coherencia			X	
9. Metodología			X	

<b>Ficha de registro N.º 03: Modelado de los datos</b>	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			X	
2. Objetividad				X
3. Actualidad				X
4. Organización			X	
5. Suficiencia			X	
6. Intencionalidad			X	
7. Consistencia			X	
8. Coherencia			X	
9. Metodología			X	

<b>Ficha de registro N.º 04: Técnica de modelado</b>	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			X	
2. Objetividad				X
3. Actualidad				X
4. Organización			X	



5. Suficiencia			X	
6. Intencionalidad			X	
7. Consistencia			X	
8. Coherencia			X	
9. Metodología			X	

<b>Ficha de registro N.º 05:</b> Evaluación óptima de clúster	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			X	
2. Objetividad			X	
3. Actualidad				X
4. Organización				X
5. Suficiencia			X	
6. Intencionalidad			X	
7. Consistencia			X	
8. Coherencia			X	
9. Metodología			X	

<b>Ficha de registro N.º 06:</b> Patrones de asociación	DEFICIENTE	ACEPTABLE	BUENO	MUY BUENO
1. Claridad			X	
2. Objetividad			X	
3. Actualidad			X	
4. Organización				X
5. Suficiencia				X
6. Intencionalidad			X	
7. Consistencia			X	
8. Coherencia			X	
9. Metodología			X	

En señal de conformidad firmo la presente en la ciudad de Piura con fecha 09/11/2021

**NOMBRE COMPLETO** : Teófilo R. Conea Calle  
**GRADO ACADÉMICO** : Maestría en Dirección y Gestión  
**DNI** : 028 20231  
**Especialidad** : Ingeniería de Sistemas  
**E-mail** : ferococa@gmail.com

**FIRMA Y SELLO**



REG. COPIA 14444  
Reg. COPIA 14444

**CONSTANCIA DE VALIDACIÓN DE CUESTIONARIO**

Yo, Vicente L. Conza Calle con DNI N.º 04820231 con grado académico de Maestría en Dirección y Gestión de TI, profesión Ingeniero en Informática desempeñándome actualmente como Docente de Ingeniería de Sistemas.


Por medio de la presente hago constatar que he revisado con fines de validación los instrumentos:

Cuestionario de tesis **“Metodología CRISP-DM para el análisis de asociación y Clustering para la diversificación de cursos de acuerdo al perfil de los clientes de la empresa ACCOM Perú SAC”**

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones:



Y. FLORENTINO  
C. M. S. S. S. S.  
SACOM PERU SAC  
REG. C. R. 14222

Título de la investigación: "MEDOLOGIA CRISP-DM PARA LA DIVERSIFICACIÓN DE CURSOS DE ACUERDO AL PERFIL DE LOS CLIENTES DE LA EMPRESA ACCOM PERU SAC, PIURA 2021,"						
Apellidos y nombres del investigador: Dongo Gutiérrez Gian Carlos, Albuquerque Guaylupo Hugo Enrique						
Apellidos y nombres del experto:						
ASPECTO POR EVALUAR				OPINION DEL EXPERTO		
VARIABLES	DIMENSIONES	INDICADORES	ITEM DE LA PREGUNTA	SI CUMPLE	NO CUMPLE	OBSERVACIONES /SUGERENCIAS
DIVERSIFICACIÓN DE CURSOS	ESTRATEGIA CORPORATIVA	Crecimiento	¿Está de acuerdo en que una empresa de capacitación debe implementar laboratorios de cómputo en sus instalaciones?	✓		
		Diferenciación	¿Está de acuerdo que una empresa de capacitación debería ofrecer asesorías a sus alumnos después de horas de clase?	✓		
		Entorno	¿Está de acuerdo que una empresa de capacitación debe brindar comodidad y seguridad al visitar sus instalaciones?	✓		
		Planeamiento	¿Está de acuerdo que una empresa de capacitación le brinde información de próximos cursos o eventos de capacitación?	✓		
	ESTRATEGIA COMPETITIVA	Imagen de marca	¿Está de acuerdo que una empresa de capacitación debe proyectar variedad en sus cursos y otros servicios?	✓		
		Posicionamiento	¿Está de acuerdo que para decidirse por una empresa de capacitación esta debe brindar cursos y servicios que otras empresas no tienen?	✓		
		Innovación	¿Está de acuerdo que una empresa de capacitación le brinde asesoramientos después de terminado el curso?	✓		
		Tecnología	¿Está de acuerdo que una empresa de capacitación le brinde una plataforma virtual para fortalecer de los temas aprendidos en clase?	✓		
ESTRATEGIA FUNCIONAL	Finanzas		¿Está de acuerdo que una empresa de capacitación brinde facilidades de pago?	✓		
			¿Está de acuerdo que una empresa de capacitación brinde descuentos especiales a ex alumnos?	✓		
	Marketing		¿Está de acuerdo que una empresa de capacitación realice lanzamientos de sus cursos por medio de las redes sociales?	✓		
			¿Está de acuerdo que una empresa de capacitación utilice elementos humorísticos virales(memes) en redes sociales para promocionar sus cursos?	✓		
	Dirección general		¿Está de acuerdo que el director de capacitación interactúe con los clientes para solicitar críticas que ayuden a mejorar el servicio de capacitación?	✓		
Firma del experto: 			Fecha: 09/11/21			

SECRETARÍA DE INVESTIGACIONES  
 INSTITUTO TECNOLÓGICO DE PIURA  
 INSTITUTO TECNOLÓGICO DE PIURA  
 INSTITUTO TECNOLÓGICO DE PIURA

## ANEXO 07: PROPUESTA PARA PLAN DE DIVERSIFICACIÓN DE CURSOS

Después de haber obtenido los resultados de la encuesta se representaron mediante gráficos de barras lo cual permitió analizar de mejor forma dichos resultados.

Para la dimensión estrategia corporativa se obtuvo un 44% de personas que están muy de acuerdo en implementar laboratorios de cómputo, ofrecer asesorías a los alumnos después de horas de clase, brindarles comodidad y seguridad cuando visiten sus instalaciones y mantenerlos informados de cursos que puedan ser de su interés, un 42% está muy en desacuerdo y un 13% no está de acuerdo ni en desacuerdo.

Para la dimensión estrategia competitiva se obtuvo un 50% de personas que están muy de acuerdo en que la empresa debe proyectar variedad en sus cursos y servicios, brindar cursos y servicios novedosos, brindar asesoramientos después de terminado el curso y brindarle una plataforma tecnológica que ayude al estudiante con los temas de clase, un 38% está muy en desacuerdo y un 12% no está de acuerdo ni en desacuerdo.

Para la dimensión estrategia funcional se obtuvo un 39% de personas que están muy de acuerdo en que la empresa debe brindar facilidades de pago, realizar los lanzamientos de sus cursos por redes sociales, usar elementos humorísticos (memes) en redes sociales para promocionar sus cursos en redes sociales y que el director de capacitación interactúe con los alumnos a fin de obtener críticas que ayuden a mejorar los cursos, un 41% está muy en desacuerdo y un 20% no está de acuerdo ni en desacuerdo.

De acuerdo al análisis de los gráficos y teniendo en cuenta que estamos en una situación no muy común debido a la pandemia de la COVID-19, podemos hacer las siguientes sugerencias para un posible plan de diversificación de cursos:

1. La empresa ACCOM PERU SAC deberá instalar laboratorios de cómputo en un futuro próximo, preocupándose de ofrecer la comodidad y seguridad de los estudiantes mediante la implementación de los protocolos de seguridad establecidos en ese periodo de tiempo para que puedan recibir sus clases sin temor al contagio. Debemos recordar que en estos tiempos de pandemia

la enseñanza es en modalidad virtual pero que poco a poco se busca regresar a la presencialidad, por eso la empresa debe estar preparada para recibir a sus estudiantes.

2. La empresa ACCOM PERU SAC deberá ofrecer asesorías después de clases haciendo las coordinaciones respectivas con los estudiantes para organizar en los días lunes, martes y miércoles sesiones de preguntas para resolverlas con ejercicios en sesiones en vivo y grabadas de media hora los días jueves y viernes.
3. La empresa ACCOM PERU SAC deberá brindar información de sus cursos a sus estudiantes al término de cada curso mediante sus redes sociales y ofreciendo promociones de descuento, pero no solo con la intención de obtener inscripciones si no también con la intención de fomentar en el estudiante la actualización continua de sus conocimientos.
4. La empresa ACCOM PERU SAC deberá brindar sus cursos con una introducción gratuita para todas las personas interesadas en inscribirse de tal forma que puedan ver como se desarrollarán dichos cursos para que los posibles estudiantes sepan que es lo que van a aprender, como es que el docente dicta el curso y vean la diferenciación con otras empresas que solo se preocupan en obtener la inscripción del estudiante.
5. La empresa ACCOM PERU SAC deberá brindar talleres gratuitos para fomentar el interés de los alumnos en capacitarse, esto servirá además para poder recolectar información de posibles estudiantes realizando consultas de que cursos serían de su interés y promover la cartera de cursos que tienen disponibles.
6. La empresa ACCOM PERU SAC deberá implementar talleres de acuerdo al perfil de los clientes que ya han llevado un curso anteriormente (de ACCOM PERU SAC o de otra empresa de capacitación) y que en la actualidad necesitan actualizar sus conocimientos sin tener que volver a inscribirse al mismo curso, esto permitirá diversificar los cursos generales a temas específicos.

7. La empresa ACCOM PERU SAC deberá mejorar su plataforma virtual para poder brindar mejores servicios como chat en vivo con el docente en horas en específico para que los alumnos puedan realizar preguntas que en la clase no pudieron hacer, salas de conversación para los estudiantes de un curso para que puedan interactuar entre ellos para realizar consultas y responderlas entre ellos mismo fomentando el trabajo en equipo, un repositorio donde los alumnos puedan subir consultas de algún tema en específico para que el docente las desarrolle en las horas de clase. Debemos tratar de que un curso no sea rígido y que solo se en desarrollar temas establecidos en un temario, el estudiante debe aportar a las clases y generar ideas para complementar los cursos.
8. La empresa ACCOM PERU SAC deberá ofrecer facilidades de pago a los estudiantes sin afectar sus ingresos por el dictado de los cursos, se recomienda brindar el pago en cuotas sin afectar al estudiante ya que se debe tener en cuenta la situación actual con la pandemia de la COVID-19 ya que un factor importante para las inscripciones es el apoyo que pueda sentir el estudiante por parte de la empresa.
9. La empresa ACCOM PERU SAC deberá fortalecer los lazos con los ex alumnos brindándoles ofertas que le resulten atractivas, talleres de complementación para cursos que ya hayan aprobado de tal manera que aliente a dichos ex alumnos a retomar sus capacitaciones teniendo en cuenta que se debe promover la capacitación continua y no solo los ingresos económicos.
10. La empresa ACCOM PERU SAC deberá usar sus redes sociales de manera más académica, brindar algunos consejos con respecto a temas en específico de los cursos de capacitación mediante transmisiones en vivo o videos grabados, no solo enfocarse en vender cursos por las redes sociales, el estudiante debe querer interactuar con las redes sociales de la empresa también por temas de aprendizaje.
11. La empresa ACCOM PERU SAC deberá evitar el uso de elementos humorísticos (memes) para la promoción de sus cursos en sus redes

sociales y centrarse más en temas de aprendizaje como tutoriales, ayudas visuales y ejercicios que le puedan servir a cualquier persona para practicar. Esto mostrará al posible estudiante el interés que la empresa tiene por ayudar en su aprendizaje este o no inscrito en un curso de la empresa.

12. La empresa ACCOM PERU SAC deberá interactuar más con los estudiantes mediante el director de capacitación, es importante conocer todas las impresiones que tienen los estudiantes no solo acerca de los cursos, sino también del docente, de la empresa, la atención que recibe por parte de los colaboradores. Se debe tratar de obtener toda la información posible para poder mejorar todas las debilidades que tenga la empresa.

13. La empresa ACCOM PERU SAC deberá constantemente estar a la expectativa mediante encuestas de cómo va evolucionando la aceptación de los cursos de manera virtual debido a que es cierto que por la realidad actual de la COVID-19 la mayoría de personas prefieren el desarrollar sus cursos de forma virtual, pero también es cierto que hay cursos que no tienen la suficiente aceptación en modalidad virtual y se necesita llevarlo de manera presencial. El regreso a la modalidad presencial debe ser consultado constantemente para poder entender cuál es el momento adecuado para ser aplicado.

14. La empresa ACCOM PERU SAC deberá aplicar estrategias de publicidad y marketing para.

Gracias a los resultados obtenidos de la aplicación de la metodología crisp-dm y de los procesos de clustering y asociación que nos dicen que los perfiles de clientes formados por estudiante, operario y otros tienen una mayor afinidad a inscribirse a cursos de ofimática cálculo siendo también este grupo de cursos los que más inscripciones generan en la empresa ACCOM PERU SAC. podemos sugerir que las recomendaciones mencionadas anteriormente se tomen en cuenta al momento de desarrollar un plan estratégico que como objetivo inicial sea la diversificación de cursos empezamos por los de ofimática cálculo para más adelante también diversificar sus servicios.



## ANEXO 08: AUTORIZACIÓN DEL TRABAJO DE INVESTIGACIÓN EN LA EMPRESA ACCOM PERÚ SAC



Acción Competitiva SAC  
Centro de Capacitación y Actualización Profesional

### CONSTANCIA

El Gerente General de Capacitación del Centro de Formación Profesional y Empresarial Acción Competitiva SAC, que suscribe;

#### HACE CONSTAR:

Que el Sr. HUGO ENRIQUE ALBURQUEQUE GUAYLUPO y el Sr. GIAN CARLOS DONGO GUTIÉRREZ, ambos estudiante de la carrera profesional de Ingeniería de Sistemas de la UNIVERSIDAD CESAR VALLEJO FILIAL PIURA DEL X CICLO, han desarrollado su trabajo de investigación: "Metodología CRISP-DM para el análisis de asociación y Clustering para la diversificación de cursos de acuerdo al perfil de los clientes de la empresa ACCOM Perú SAC Piura, 2021" en nuestra Empresa Acción Competitiva (ACCOM) y cuentan con la autorización del gerente general ING. M. HENRY JULCA GARCÍA.

Durante el desarrollo de su investigación en la empresa han demostrado en todo momento alto espíritu de colaboración, iniciativa, responsabilidad y puntualidad.

Se expide el presente a solicitud del interesado(a), para los fines que estime conveniente.

Piura, 29 de setiembre de 2021

CC. Archivo



  
Ing. M. Henry Julca García  
Gerente General  
Acción Competitiva SAC.



#### INFORMES:

Correo: [informes@accom.pe](mailto:informes@accom.pe) | [www.accom.pe](http://www.accom.pe)

Oficina: Urbanización Santa Isabel, Lote t-8. (Boulevard del Centro Comercial Santa Isabel). Oficina 12, 2do piso.

Teléfono: 073600347 C: 955009420 | 973910786 RPM: | \*0083715