



ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES APLICADO A UNA EMPRESA DE SERVICIOS



Autor Facundo Paz

DNI 42.076.511

Carrera Licenciatura en Administración (Plan 2014)

Modalidad Trabajo de aplicación de conceptos y técnicas de administración en situación laboral de revista o ambiente real

Tutor Esp. Marcelo Medina Galván



ÍNDICE

RESUMEN	3
1. INTRODUCCIÓN	4
1.1. SOBRE LA EMPRESA.....	4
2. SITUACIÓN PROBLEMÁTICA.....	5
2.1. PREGUNTAS DE INVESTIGACIÓN.....	5
3. OBJETIVOS.....	5
3.1. OBJETIVO GENERAL	5
3.2. OBJETIVOS ESPECÍFICOS.....	5
4. MARCO METODOLÓGICO	6
5. MARCO TEÓRICO.....	6
6. DESARROLLO	11
6.1. PRIMERA ETAPA: COMPRESIÓN DE LA BASE DE DATOS.....	11
6.2. SEGUNDA ETAPA: ANÁLISIS PRELIMINAR DE LOS DATOS SELECCIONADOS	12
6.3. TERCERA ETAPA: APLICACIÓN DEL ACM Y ANÁLISIS DE LOS RESULTADOS.....	20
7. CONCLUSIONES FINALES Y REFLEXIONES.....	23
8. BIBLIOGRAFÍA	25
9. ANEXO - SCRIPTS UTILIZADOS	26



RESUMEN

Uno de los principales desafíos de toda empresa es la optimización de los recursos utilizados de tal forma que los objetivos se logren de una manera cada vez más eficiente. Para ello, la práctica del análisis de datos y el aprovechamiento de los mismos resulta una herramienta práctica muy recomendada ya que los mismos permiten realizar diagnósticos con un elevado nivel de claridad y profundidad, fundamentales a la hora de visualizar el contexto de la empresa y tomar decisiones fundadas en evidencia objetiva.

El presente trabajo tiene como objeto de estudio a PH, una empresa de servicios ubicada en la ciudad de San Miguel de Tucumán, con el objetivo de generar conocimiento que permita cualificar el desempeño en la gestión de cobranzas. A partir del estudio de la base de datos brindadas por el sistema de gestión de la empresa, y luego de un trabajo minucioso de limpieza y decodificación de las tablas, se emplea la técnica estadística del “análisis de correspondencias múltiples”, el cual es una herramienta apropiada a la hora de trabajar con bases de datos de naturaleza cualitativa de tres o más variables. Para ello será necesario la utilización de distintos softwares de análisis tales como Excel, Power BI y R studio. Finalmente se obtiene un mapa de calor de dos dimensiones, el cual representa gráficamente distintos “perfiles” de clientes según su rubro, tamaño de la empresa y nivel de deuda.

El trabajo cuenta con un enfoque mixto de diseño exploratorio secuencial (DEXPLOS). La recolección de datos se sustenta mediante la realización de entrevistas con el gerente y acceso a la base de datos de la empresa.

Palabras Clave

Analítica de negocios, estadística descriptiva, análisis de correspondencias múltiples



1. INTRODUCCIÓN

En un mundo cada vez más orientado a los datos, la capacidad de recopilar, analizar y utilizar los mismos se ha convertido en una importante ventaja competitiva para las empresas de servicios. En particular, la práctica del análisis de datos puede proporcionar información valiosa sobre la eficiencia, eficacia y efectividad en el desempeño de las distintas áreas de la empresa, detectando patrones, puntos críticos de mejora y caracterizando clientes para cualificar el servicio brindado y diferenciarse de la competencia.

En este trabajo se analiza el uso de datos en una empresa de servicios para caracterizar los clientes y relacionarlos con su situación de deudas, a fin de extraer información relevante a la gestión de cobranza que pueda transformarse en políticas y estrategias de mejora para la empresa.

Se aplican técnicas de análisis de correspondencias múltiples para variables cualitativas categóricas, a través del software estadístico R studio.

1.1. SOBRE LA EMPRESA

El presente trabajo se enfoca sobre una empresa de servicios ubicada en la capital de la provincia de Tucumán. La misma se encuentra en el rubro de la Higiene y Limpieza, y tiene como clientes a otras empresas, por lo que posee un enfoque comercial B2B. A partir de este momento, procederemos a identificar a la empresa con el nombre ficticio “PH”.

PH ofrece “sistemas de higiene”, los cuales se enfocan en la solución de problemas tales como la desodorización de baños, aromatización, purificación de ambientes, etc. La empresa funciona con 14 empleados distribuidos en 3 áreas de trabajo: vendedores, técnicos y ejecutivos de cuentas. Los primeros son los encargados de contactar a los clientes y ofrecer la cartera de productos disponibles. Cuando el cliente confirma la contratación, se programa la visita de un técnico, encargado de la instalación y mantenimiento de los sistemas. Finalmente, los ejecutivos de cuenta son los responsables de mantener la comunicación con el cliente y fortalecer la relación comercial con el mismo. El ejecutivo atiende reclamos, programa nuevas visitas, hace ventas cross-selling, y se ocupa de la gestión de cobranzas, por enunciar algunas de sus tareas.

Todas las operaciones, documentos y transacciones de la empresa se organizan y coordinan desde el software de gestión “ProKey”, el cual almacena una gran cantidad de información disponible. Sin embargo, ningún empleado dentro de la empresa comprende entre sus funciones el análisis de dichos datos ni tiene las competencias y conocimientos requeridos para ello.



2. SITUACIÓN PROBLEMÁTICA

“PH” es una empresa que requiere un alto nivel de coordinación logística y atención al detalle en los sistemas que ofrece, ya que su propuesta de valor es brindar un servicio de calidad y excelencia que permita a sus clientes presentar una imagen empresarial profesional y bien cuidada a sus respectivos usuarios. Esto los obliga a tener procesos internos muy ágiles, precisos y eficaces. A su vez, PH tiene a disposición una gran cantidad de datos registrados en su sistema de gestión ProKey que, de ser analizados, se podría detectar importantes aspectos a mejorar y fallas a corregir. Teniendo en cuenta estas dos características, se plantea la oportunidad de observar el contenido de la base de datos, explorar los campos que dispone y analizar a profundidad los mismos, con el fin de extraer información valiosa que permita tomar las decisiones necesarias para cualificar los procesos y desempeño de PH en su gestión empresarial.

2.1. PREGUNTAS DE INVESTIGACIÓN

1. ¿Qué datos tiene a disposición PH en su sistema de gestión?
2. ¿Qué técnicas de análisis pueden ser de utilidad para analizar la base de PH?
3. ¿Qué información se puede extraer de la base de datos para cualificar la gestión empresarial?

3. OBJETIVOS

3.1. OBJETIVO GENERAL

A partir del análisis de datos, generar conocimiento que permita cualificar el desempeño en la gestión de cobranzas de PH.

3.2. OBJETIVOS ESPECÍFICOS

- Explorar la base de datos de PH e identificar los campos de datos más relevantes.



- Seleccionar las técnicas y herramientas de análisis que puedan ser aplicables a la base de datos disponible.
- Aplicar las técnicas y herramientas seleccionadas para extraer información valiosa para la empresa.

4. MARCO METODOLÓGICO

Para el desarrollo del trabajo se utiliza un enfoque mixto de diseño exploratorio secuencial (DEXPLOS), el cual indaga la incidencia de las modalidades, categorías o niveles de una o más variables en una población; son estudios puramente descriptivos.

La recolección de datos se sustenta mediante la realización de entrevistas con el gerente y acceso a la base de datos de la empresa. El análisis se vale de herramientas tales como la hoja de cálculo Excel, el software de análisis estadístico R studio y el programa Power BI para la visualización de datos.

5. MARCO TEÓRICO

Analítica de negocios

La Analítica de Negocios (Business Analytics) es el estudio de los datos generados por la empresa a través del análisis estadístico. Este estudio implica la elaboración de modelos predictivos, la aplicación de técnicas de optimización y la comunicación de estos resultados a ejecutivos de la empresa, socios de negocios e incluso a clientes.

El Business Analytics es un conjunto de habilidades, tecnologías y prácticas que permite identificar tendencias y comportamientos a partir de datos empíricos, predecir escenarios y tomar decisiones óptimas para alcanzar metas de negocios.

El desafío en áreas de negocios es entender las enormes capacidades actuales y potenciales de la Analítica de Negocios de modo de aprovecharlas en el diseño de productos, servicios y procesos. En este devenir se transformarán formas de gestión ya existentes y se crearán muchas otras que aún no se prevé.

Render (2016) afirma que es un enfoque basado en datos que permite a las empresas tomar mejores decisiones. La analítica de negocios se puede dividir en las siguientes categorías:

- Analítica descriptiva: implica el estudio y la consolidación de los datos y permite describir el desempeño;



- Analítica diagnóstica: evalúa el desempeño pasado y cómo se efectúa en el momento presente;
- Analítica predictiva: está dirigida a pronosticar resultados futuros con base a los patrones de los datos históricos;
- Analítica prescriptiva: posibilita el uso de métodos de optimización para ofrecer nuevas y mejores formas de operar.

En la mejor práctica del management existe una clara sinergia entre las áreas tácticas y estratégicas de identificación de oportunidades y esbozo de respuestas ("¿qué podemos hacer?"), y las áreas técnicas donde se implementan estas respuestas ("¿cómo ponemos en práctica la solución?"). Lograr esta sinergia implica un desafío fundamental: por un lado, las áreas "blandas" suelen desconocer el potencial de infraestructura y calidad de información con que cuenta las áreas "duras"; por otro, estas últimas desconocen el mundo de los negocios y el potencial que encierran los datos con los que lidian permanentemente. Articular estas capacidades es un objetivo central del management actual y la demanda de profesionales con esta formación crece año tras año.

Análisis Multivariante

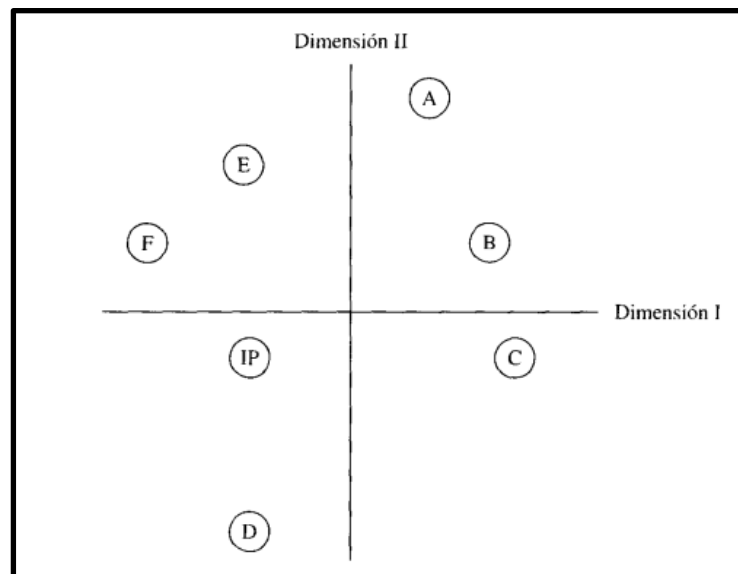
Según afirma Hair (1999), el análisis multidimensional tiene por objetivo transformar las características cualitativas de determinadas variables en distancias representadas en un espacio multidimensional. Por ejemplo, analizando el caso de preferencia de los consumidores por tiendas o marcas comerciales, si las marcas "A y B" son en opinión de los encuestados más similares que el resto de los pares posibles de objetos, las técnicas de análisis multidimensional situarán a los objetos "A y B" de tal forma que la distancia entre ellos en un espacio multidimensional sea menor que la distancia entre cualquier otro par de objetos. Los mapas perceptuales resultantes muestran el posicionamiento relativo entre los objetos.

Mapas perceptuales

Un mapa perceptual consiste en un modelo gráfico de cuatro cuadrantes, donde son representados los puntos de posicionamiento correspondiente de distintos atributos de las variables observables. El mapa perceptual de posicionamiento permite representar gráficamente y de manera sencilla el lugar que ocupa una característica cualitativa determinada (pueden ser productos, servicios, empresas, personas, etc) dentro del conjunto de datos observables, teniendo en cuenta la frecuencia y relación que tiene una variable con las demás.

El mapa perceptual resulta ser una herramienta valiosa para la toma de decisiones en cuestión de posicionamiento, ya que la fácil visualización es un soporte fundamental al análisis de datos.

Figura 01: Mapa perceptual teórico

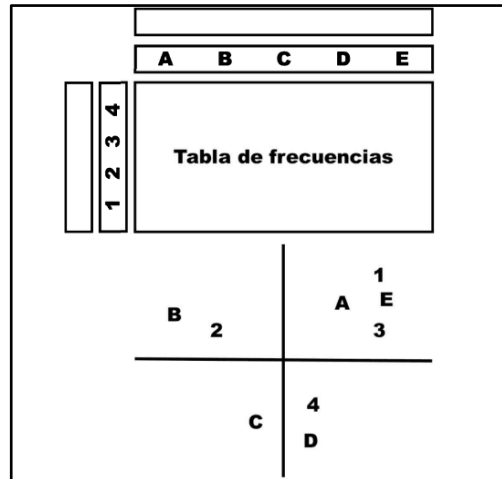


Fuente: Análisis Multivariante. Capítulo 10 (Hair, Anderson, Tatham , Black)

Análisis de Correspondencias

El análisis de correspondencias es una técnica de análisis estadístico para datos cualitativos categóricos . Afirma Aldás y Uriel (2017) que el análisis de correspondencias pretende representar en un espacio multidimensional reducido la relación y distancias entre las categorías de dos variables no métricas, por lo que se suele decir que el análisis de correspondencias sirve para la visualización de tablas de contingencias.

Figura 02: Finalidad básica del análisis de correspondencias



Fuente: Análisis multivariante aplicado con R. Cap 5 (Aldás y Uriel)

Análisis de correspondencias múltiples

El análisis de correspondencias múltiples es una técnica de interdependencia recientemente desarrollada que facilita tanto la reducción dimensional de una clasificación de objetos (por ejemplo, productos, personas, etc.) sobre un conjunto de atributos y el mapa perceptual de objetos relativos a estos atributos. Los investigadores se enfrentan constantemente a la necesidad de «cuantificar datos cualitativos» que encuentran en variables nominales. El análisis de correspondencias difiere de otras técnicas de interdependencia discutidas antes en su capacidad para acomodar tanto datos no métricos como relaciones no lineales. En su forma más básica, el análisis de correspondencias emplea una tabla de contingencia, que es la tabulación cruzada de dos variables categóricas. A continuación transforma los datos no métricos en un nivel métrico y realiza una reducción dimensional (similar al análisis factorial) y un mapa perceptual (similar al análisis multidimensional). El análisis de las correspondencias proporciona una representación multivariante de la interdependencia de datos no métricos que no es posible realizar con otros métodos.

Eigenvalores en el ACM

En el análisis de correspondencias múltiples (ACM), los eigenvalores son una medida importante para evaluar la importancia de los ejes principales en la representación de los datos.

En el ACM, se realiza una descomposición de valores propios (eigenvalores) y vectores propios (eigenvectores) de una matriz de contingencia. Los eigenvalores



representan la varianza explicada por cada eje principal en el análisis. Cuanto mayor sea el eigenvalor asociado a un eje, mayor será la cantidad de varianza de los datos que se explica por ese eje.

Los eigenvalores también se utilizan para determinar la cantidad de ejes principales a considerar en el análisis. Por lo general, se seleccionan los ejes principales con los eigenvalores más altos, ya que estos explican la mayor parte de la varianza en los datos. Los ejes con eigenvalores más bajos se consideran menos importantes y pueden ser descartados para simplificar la representación gráfica o el análisis posterior.

Medidas de similitud

Según Black (1999) la similitud entre objetos es una medida de correspondencia, o parecido, entre objetos que van a ser agrupados. La similitud entre objetos puede medirse de varias formas, pero tres métodos dominan las aplicaciones del análisis multivariante: medidas de correlación, medidas de distancia y medidas de asociación. Cada uno de los métodos representa una perspectiva particular de similitud, dependiendo tanto de sus objetivos como del tipo de datos. Tanto las medidas de distancia como la correlación exigen datos métricos, mientras que las medidas de asociación son para datos no métricos.

El Coseno Cuadrado

El cálculo del coseno al cuadrado está relacionado con el análisis de correspondencias múltiples (ACM) a través de la medida de asociación conocida como el coeficiente de contingencia. El coeficiente de contingencia es una medida de la asociación entre dos variables categóricas en un análisis de correspondencias.

En el ACM, se utiliza el coseno al cuadrado del ángulo entre dos puntos en el espacio de representación para calcular el coeficiente de contingencia. Este cálculo se basa en la idea de que cuanto más cercanos estén dos puntos en el espacio de representación, mayor será la asociación entre las categorías que representan esos puntos.

El coseno al cuadrado se utiliza para medir la similitud o la distancia entre dos puntos en el espacio de representación. Cuanto más cercano a 1 sea el valor del coseno al cuadrado, mayor será la similitud entre los puntos y, por lo tanto, mayor será la asociación entre las categorías que representan esos puntos.

6. DESARROLLO

6.1. PRIMERA ETAPA: COMPRENSIÓN DE LA BASE DE DATOS

El presente trabajo comienza con el análisis preliminar de 19 archivos de extensión .XML (extensive markup language), el cual requiere una transformación a través de Power Query para obtener las tablas visualizables en hojas de cálculo de Excel y a partir de esto tener acceso a la manipulación de dichos datos.

Figura 03: Transformación de archivos XML a Xmls a través de Power Query

	Idfranquicia	Periodo	Fecha	Nivel	UN	Idcliente	CteNombre
1	202304	1019	202304	6/4/2023	0 LOC	10001	JUANA
2	202304	1019	202304	6/4/2023	0 LOC	10001	JUANA
3	202304	1019	202304	6/4/2023	0 LOC	10001	JUANA
4	202304	1019	202304	6/4/2023	0 LOC	10001	JUANA
5	202304	1019	202304	6/4/2023	0 LOC	10001	JUANA
6	202304	1019	202304	6/4/2023	0 LOC	10005	DON TORIBIC
7	202304	1019	202304	6/4/2023	0 LOC	10005	DON TORIBIC
8	202304	1019	202304	6/4/2023	0 LOC	10005	DON TORIBIC
9	202304	1019	202304	6/4/2023	0 LOC	10005	DON TORIBIC
10	202304	1019	202304	6/4/2023	0 LOC	10008	SANTA MART
11	202304	1019	202304	6/4/2023	0 LOC	10008	SANTA MART
12	202304	1019	202304	6/4/2023	0 LOC	10008	SANTA MART
13	202304	1019	202304	6/4/2023	0 LOC	10011	PREMIER
14	202304	1019	202304	6/4/2023	0 LOC	10011	PREMIER
15	202304	1019	202304	6/4/2023	0 LOC	10011	PREMIER
16	202304	1019	202304	6/4/2023	0 LOC	10012	CIM
17	202304	1019	202304	6/4/2023	0 LOC	10012	CIM
18	202304	1019	202304	6/4/2023	0 LOC	10012	CIM
19	202304	1019	202304	6/4/2023	0 LOC	10012	CIM
20	202304	1019	202304	6/4/2023	0 LOC	10021	SANATORIO F
21	202304	1019	202304	6/4/2023	0 LOC	10021	SANATORIO F
22	202304	1019	202304	6/4/2023	0 LOC	10026	SANATORIO F
23	202304	1019	202304	6/4/2023	0 LOC	10026	SANATORIO F
24	202304	1019	202304	6/4/2023	0 LOC	10026	SANATORIO F
25	202304	1019	202304	6/4/2023	0 LOC	10040	LA PIZZADA Y
26	202304	1019	202304	6/4/2023	0 LOC	10040	LA PIZZADA Y

Fuente: Captura de pantalla del proceso de transformación de datos.

Una vez obtenida las tablas en formato legible por el procesador Excel, se observa que muchos de los archivos presentan campos codificados, y otros no contienen datos de relevancia para el análisis. Ante esto, se procede a coordinar reuniones con el gerente para consultar y comprender el significado e importancia de cada una de las bases, para luego efectuar una limpieza de todos los campos cuyos registros no contienen un aporte significativo para este trabajo.



Figura 04: Bases de datos en bruto antes de la limpieza

L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
cliedeu	fultfac	cultfac	ffimper	fproxfac	ruta	rutaP	grupofac	frecuenciaFac	IntervaloFac	dias	frecArt	frecuenciaArt	articulo	orden
15706	21/4/2023	21/4/2023	[Table]	[Table]	RF01	RF01	MA	M		1	-15	10	0 PDESS3002	1
15706	21/4/2023	21/4/2023	[Table]	[Table]	RF01	RF01	MA	M		1	-15	10	28 desc001	2
15706	21/4/2023	21/4/2023	[Table]	[Table]	RF01	RF01	MA	M		1	-15	10	28 PDPP002	3
15706	21/4/2023	21/4/2023	[Table]	[Table]	RF01	RF01	MA	M		1	-15	10	28 PACT001	4
15706	21/4/2023	21/4/2023	[Table]	[Table]	RF01	RF01	MA	M		1	-15	10	0 PAROP3001	5
15706	21/4/2023	21/4/2023	[Table]	[Table]	RF01	RF01	MA	M		1	-15	129	217 PPAAX2	6
15706	21/4/2023	21/4/2023	[Table]	[Table]	RF01	RF01	MA	M		1	-15	10	0 PARADA	7
15143	1/4/2023	1/4/2023	[Table]	[Table]	RF19	RF19	MA	M		1	5	10	28 PALS001	1
15143	1/4/2023	1/4/2023	[Table]	[Table]	RF19	RF19	MA	M		1	5	10	0 PAROP3001	2
15143	1/4/2023	1/4/2023	[Table]	[Table]	RF19	RF19	MA	M		1	5	129	217 PPAAX2	3
15143	1/4/2023	1/4/2023	[Table]	[Table]	RF19	RF19	MA	M		1	5	10	28 PASB001	4
15143	1/4/2023	1/4/2023	[Table]	[Table]	RF19	RF19	MA	M		1	5	10	0 PAROP3001	5
15143	1/4/2023	1/4/2023	[Table]	[Table]	RF19	RF19	MA	M		1	5	129	217 PPAAX2	6
15143	1/4/2023	1/4/2023	[Table]	[Table]	RF19	RF19	MA	M		1	5	10	0 PARADA	7
15143	1/4/2023	1/4/2023	[Table]	[Table]	RF19	RF19	MA	M		1	5	5	56 PAST021C	8
15143	1/4/2023	1/4/2023	[Table]	[Table]	RF19	RF19	MA	M		1	5	10	0 PAST001C	9
15143	1/4/2023	1/4/2023	[Table]	[Table]	RF19	RF19	MA	M		1	5	5	56 PAST002C	10

Fuente: Bases de datos brindadas por PH

Luego de la limpieza, se consideran útiles los siguientes campos de datos de los clientes en el Maestro de Clientes:

Datos Cuantitativos:

- Número de productos contratados
- Número de artículos por tipo de producto
- Número de familias de producto
- Facturación mensual por cliente
- Saldo adeudado por cliente

Datos Cualitativos

- Ciudad
- Localidad
- Rubro
- Sub-rubro
- Tamaño
- Tránsito

Dada las características de los datos disponibles y ante las distintas alternativas de herramientas de estudio existentes, se decide realizar un análisis de correspondencias múltiples.

6.2. SEGUNDA ETAPA: ANÁLISIS PRELIMINAR DE LOS DATOS SELECCIONADOS

Para aplicar el ACM se debe seleccionar las variables cualitativas categóricas que presentan a priori algún tipo de relación, de tal forma que su



análisis técnico pueda precisar con refuerzo estadístico cierta deducción empírica. Para el presente trabajo se decide utilizar los campos “Rubro”, “Tamaño” y “Deuda”. Nótese que esta última variable presenta una naturaleza cuantitativa, por lo que para su utilización se debe transformar o clasificar de alguna forma los datos de este campo para que puedan ser tomados como variables cualitativas. Más adelante se detalla la clasificación que se ve conveniente para dicha variable.

El propósito de la selección de estas variables es poder detectar si existe un determinado perfil de empresas cumplidoras con las facturas emitidas y un perfil de clientes conflictivos en la gestión de cobranzas.

Hipótesis inicial

El presente trabajo presenta una hipótesis inicial, el cual la aplicación del ACM al finalizar el desarrollo del mismo deberá afirmar o negar:

H1: Las empresas “grandes” tienen una probabilidad mayor a no tener deuda pendiente respecto a las “medianas” y “chicas”.

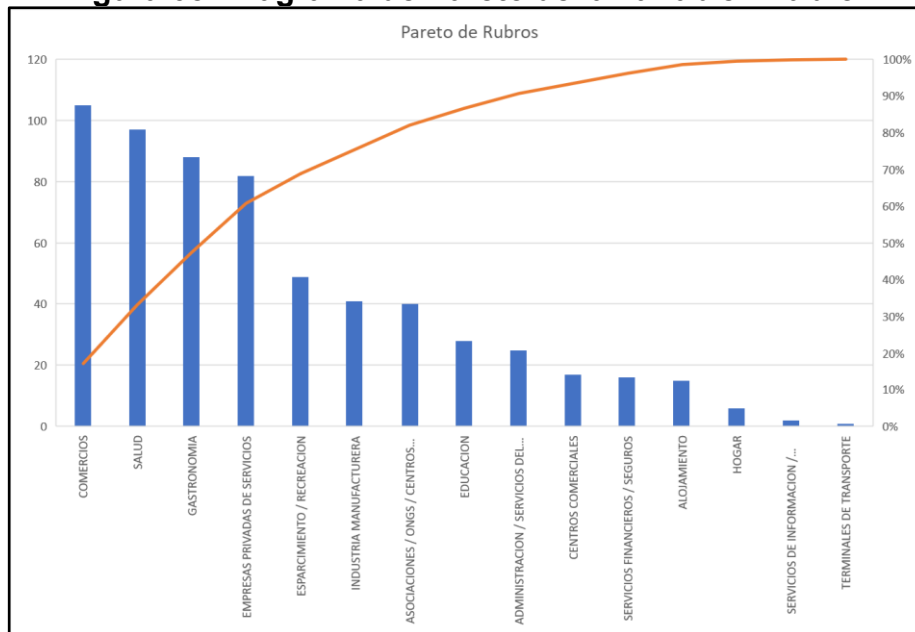
Esta deducción preliminar resulta del supuesto de que las grandes compañías suelen tener una estructura organizacional que permiten un mayor orden y prolijidad en sus operaciones.

Descripción de los datos seleccionados

En total se presentan 485 observaciones de clientes, las cuales están clasificados según los siguientes campos:

Rubro: Clasifica el tipo de actividad del cliente y en qué sector del mercado opera. Presenta 15 variables, de las cuales no todas tienen la misma representatividad respecto al total de variables.

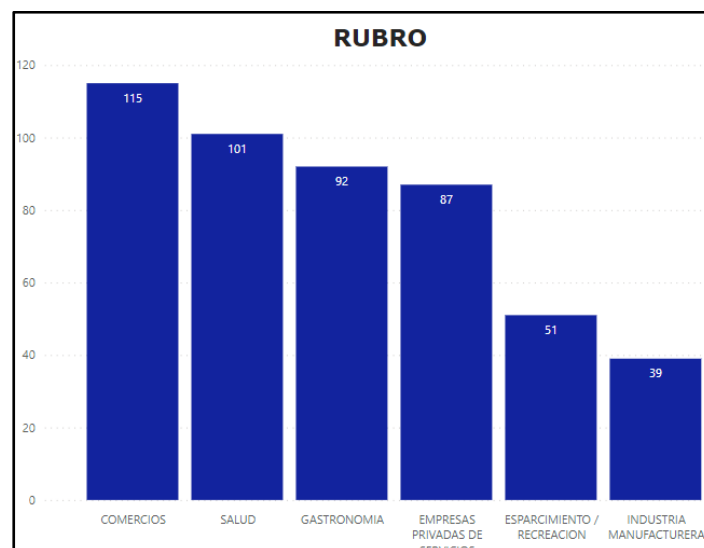
Figura 05: Diagrama de Pareto de la variable “Rubro”



Fuente: Elaboración propia en Excel

A fines prácticos y de mejorar el nivel de análisis, se decide seleccionar las variables que explican el 75,49% de participación de la base de datos. Los mismos son: Comercio, Salud, Gastronomía, Empresas privadas de servicios, Esparcimiento/Recreación e Industria Manufacturera.

Figura 06: Distribución de frecuencias de la variable “Rubro” seleccionadas

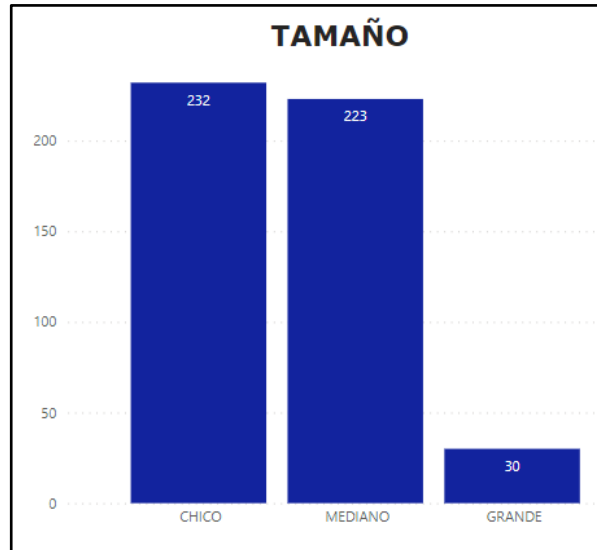


Fuente: Elaboración propia en Power BI



Tamaño: Clasifica el tamaño de la empresa según volumen de operaciones y cantidad de empleados.

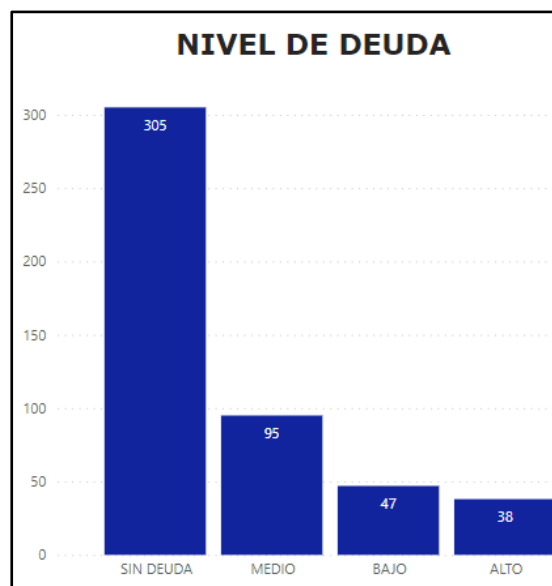
Figura 07: Distribución de frecuencias de la variable “Tamaño”



Fuente: Elaboración propia en Power BI

Deuda: Clasifica el tamaño de la empresa según la cantidad de meses adeudados.

Figura 08: Distribución de frecuencias de la variable “Deuda”



Fuente: Elaboración propia en Power BI



A continuación en la figura 09 se muestra una captura de las primeras 25 observaciones de la base de datos en limpio que actúa como input en el software estadístico R studio. En el anexo final del trabajo se comparten los scripts utilizados en el software para el presente trabajo.

Figura 09: Base de datos final

RUBRO	TAMAÑO	DEUDA
GASTRONOMIA	CHICO	SIN DEUDA
GASTRONOMIA	MEDIANO	MEDIO
GASTRONOMIA	CHICO	SIN DEUDA
SALUD	CHICO	SIN DEUDA
SALUD	CHICO	SIN DEUDA
SALUD	MEDIANO	BAJO
GASTRONOMIA	MEDIANO	ALTO
SALUD	CHICO	SIN DEUDA
GASTRONOMIA	CHICO	SIN DEUDA
SALUD	MEDIANO	MEDIO
GASTRONOMIA	CHICO	BAJO
ESPARCIAMIENTO / RECREACION	MEDIANO	MEDIO
GASTRONOMIA	CHICO	SIN DEUDA
GASTRONOMIA	CHICO	SIN DEUDA
GASTRONOMIA	CHICO	SIN DEUDA
GASTRONOMIA	CHICO	SIN DEUDA
GASTRONOMIA	MEDIANO	SIN DEUDA
EMPRESAS PRIVADAS DE SERVICIOS	GRANDE	BAJO
EMPRESAS PRIVADAS DE SERVICIOS	CHICO	SIN DEUDA
COMERCIOS	MEDIANO	MEDIO
SALUD	MEDIANO	BAJO
INDUSTRIA MANUFACTURERA	MEDIANO	BAJO
COMERCIOS	CHICO	SIN DEUDA
COMERCIOS	CHICO	SIN DEUDA
INDUSTRIA MANUFACTURERA	GRANDE	MEDIO

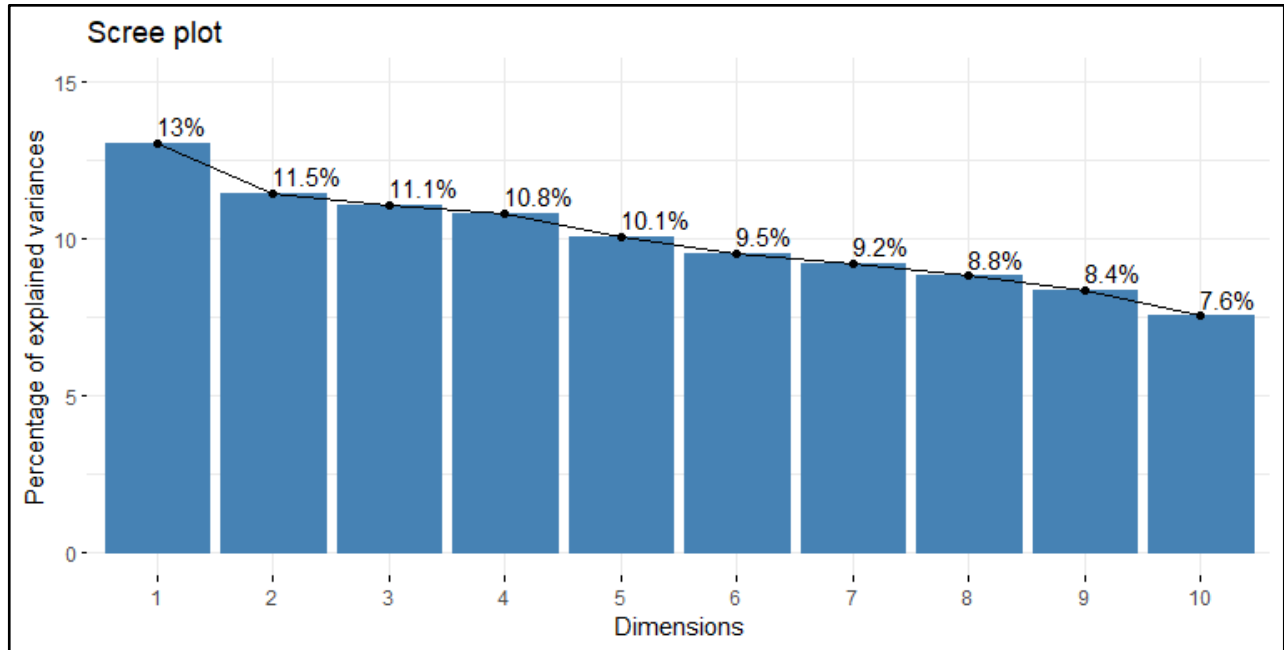
Fuente: Captura de Pantalla de Excel

A partir de los datos seleccionados y cargados en R studio se procede a realizar los siguientes análisis:

Análisis de los Eigenvalores y los componentes principales

Para determinar el número de componentes principales se puede mirar un Scree Plot, o también llamado gráfico de sedimentación, que es un plot de los eigenvalores ordenados de mayor a menor. El número de componentes es determinado en el punto más allá del cual los eigenvalores restantes son todos relativamente pequeños y de tamaño comparable.

Figura 10: Gráfico de sedimentación con los componentes principales del ACM



Fuente: Elaboración propia en R studio

Figura 11: Eigenvalores y Varianza porcentual de las primeras 6 dimensiones

 	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	0.4349	13.05	13.05
Dim.2	0.3821	11.46	24.51
Dim.3	0.369	11.07	35.58
Dim.4	0.3604	10.81	46.39
Dim.5	0.3359	10.08	56.47
Dim.6	0.3179	9.536	66

Fuente: Elaboración propia en R studio

A partir del análisis de los eigenvalores, podemos conocer las características de las dos principales dimensiones, las cuales serán utilizadas como ejes



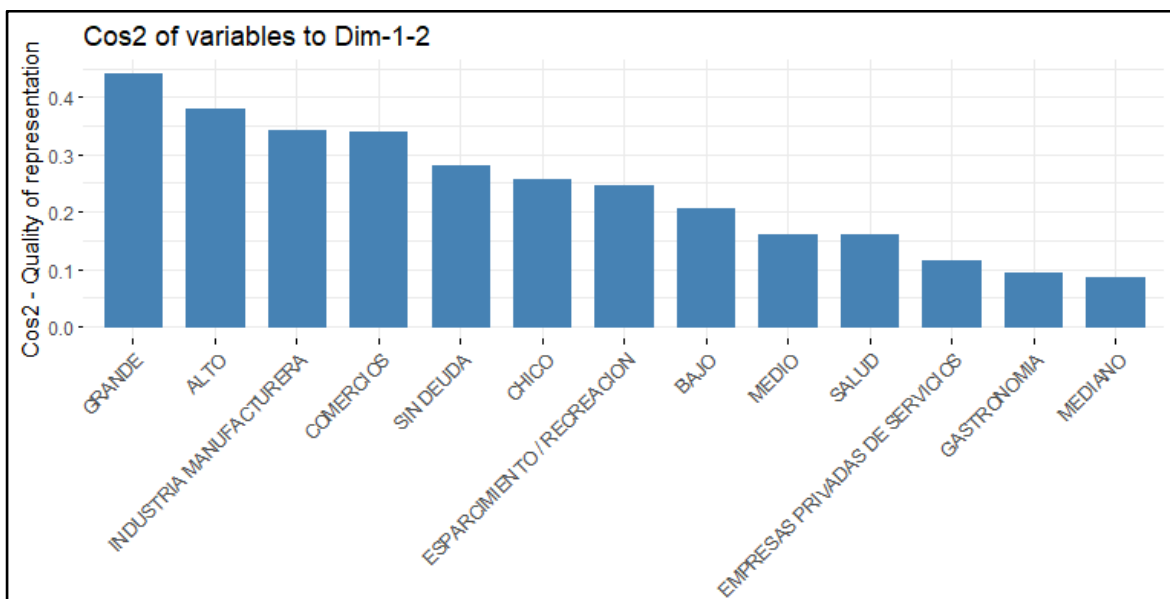
principales del mapa, y es entonces donde se ubicarán las demás variables permitiendo el estudio de las distancias entre cada una. Juntando ambas dimensiones, se logra explicar el 24,51% de la varianza total de la matriz.

Análisis del Coseno al Cuadrado

A la hora de examinar la calidad de la representación de cada una de las variables, resulta de mucha utilidad la prueba del Coseno al cuadrado (Cos^2), el cual tiene por función justamente calcular la similitud o distancia entre las dos dimensiones de los ejes. Mientras mayor sea la representación, la suma del Cos^2 será más cercana a 1, y mientras peor sea mayor se aproxima a 0.

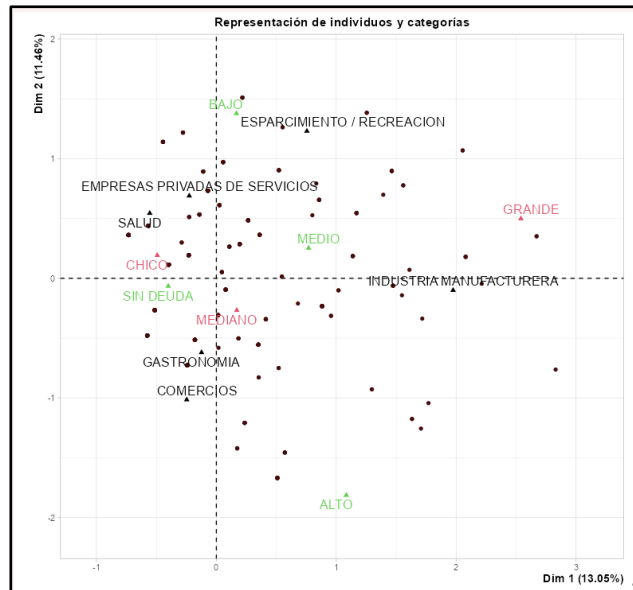
A partir de la siguiente figura podemos observar que para las dimensiones 1 y 2, las variables que presentan una mayor calidad de representación son “Grande”, “Alto” e “Industrias Manufactureras”.

Figura 12: Cos^2 de Categorías para las Dimensiones 1-2



Fuente: Elaboración propia en R studio

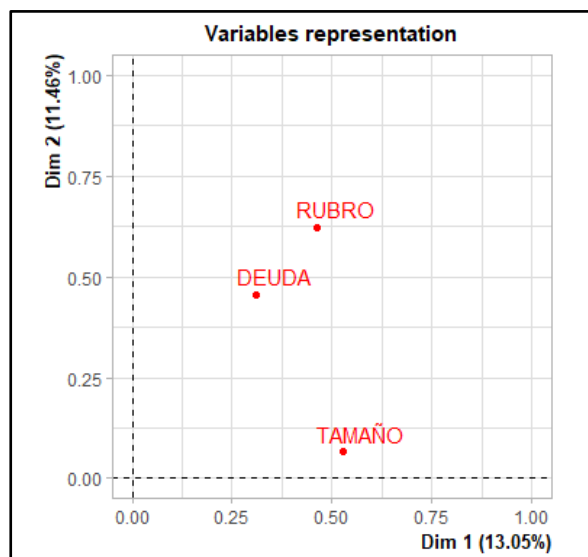
Figura 13: Gráfico de representación de individuos y categorías



Fuente: Elaboración propia en R studio

Este gráfico muestra la distribución de los individuos con mayor contribución en el plano, en conjunto con la ubicación de cada una de las categorías observadas en la base de datos.

Figura 14: Relación entre categorías y dimensiones



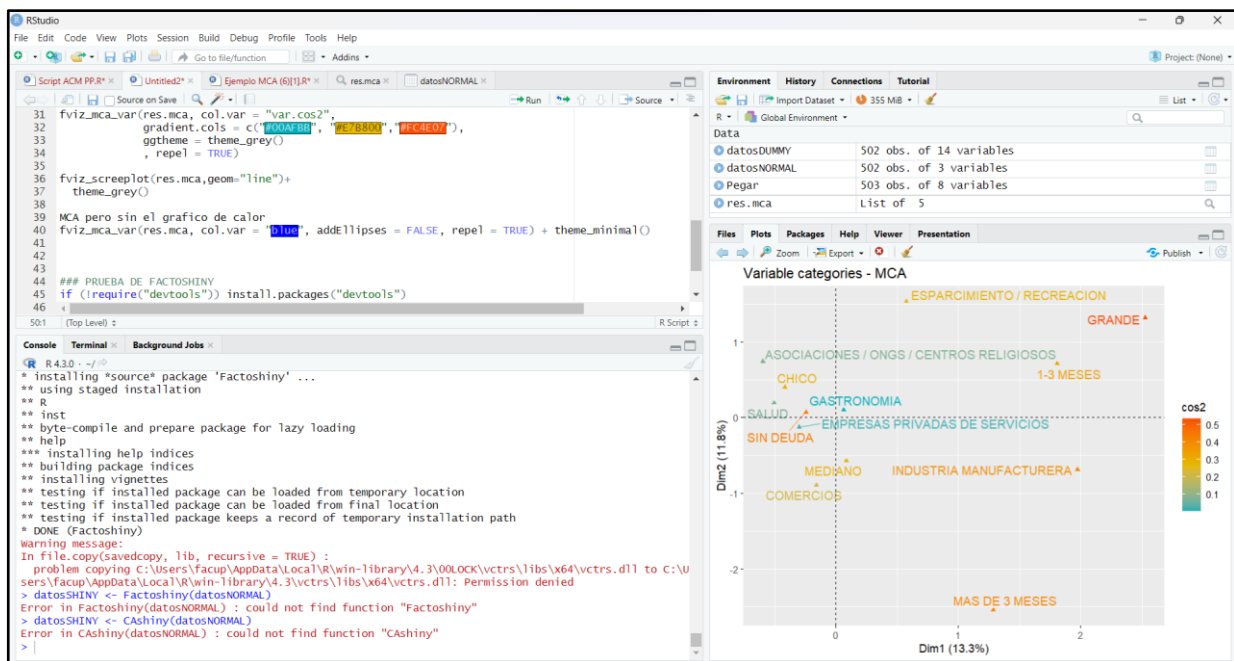
Fuente: Elaboración propia en R studio

Este gráfico ayuda a identificar las variables que están más correlacionadas con cada dimensión. Las correlaciones cuadradas entre variables y las dimensiones son usadas como coordenadas. En este caso se tiene que la variable “Tamaño” es la que presenta mayor correlación con la dimensión 1.

6.3. TERCERA ETAPA: APLICACIÓN DEL ACM Y ANÁLISIS DE LOS RESULTADOS

Finalmente se procede a programar la función que procesa los datos de la matriz, y que brinda como output un mapa de calor donde se ven ubicadas todas las categorías de la base de datos.

Figura 15: Software de análisis estadístico R studio



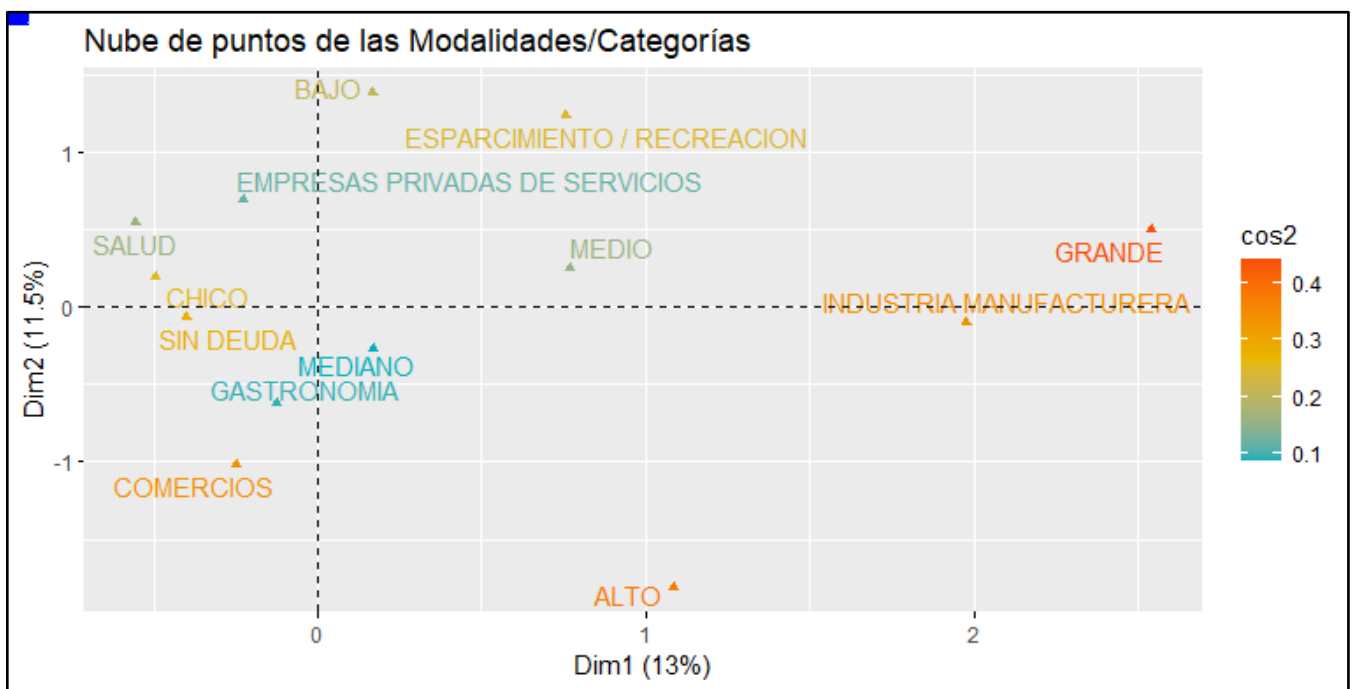
Fuente: Elaboración propia. Captura de pantalla de software R studio

En el siguiente gráfico se ven situadas todas las variables de nuestra base de datos. Para interpretar el mismo hay que realizar una doble lectura según los siguientes criterios:

- 1) Proximidad entre las categorías: Mientras más cercanos sean los puntos, mayor correlación tendrá el comportamiento de los mismos.
- 2) La calidad de la representación: Esto se mide a través del cálculo matemático del “coseno cuadrado” (Cos2), el cual permite saber qué

tan bien representado está el punto sobre el eje factorial. La calidad de un punto será tanto mayor cuanto más próximo a 1 sea el coseno cuadrado, representado en el mapa de calor con un color mas rojizo. Por otro lado, mientras menor sea el valor, el mismo se aproxima a 0 y tendrá una leyenda mas azulada.

Figura 16: Mapa de calor de la relación entre variables



Fuente: Elaboración propia en R studio

Respuesta a la hipótesis:

- Contrario a lo que plantea la hipótesis, se puede asegurar por la ubicación en el mapa que las empresas “grandes” están muy alejadas de una situación “Sin deuda”. Una posible explicación a esto es el burocrático proceso interno que muchas de ellas llevan a cabo, lo cual conlleva un tiempo mínimo de procesamiento en el que terminan pagando las facturas de manera sistemática a mes vencido, ocasionando la deuda de, al menos, 1 mes de atraso. Por lo tanto, la hipótesis se ve refutada, proponiendo entonces la siguiente afirmación:

“Las empresas “grandes” suelen tener, al menos, 1 mes de deuda”.



Más interpretaciones del análisis:

- Si bien las variables Gastronomía y Salud se encuentran cercanas a la leyenda “Sin deuda”, no se puede garantizar dicha afirmación ya que el nivel de Cos2 que presentan es demasiado bajo.
- Las empresas más grandes de la selección son las industrias manufactureras.
- No hay ningún rubro ni tamaño de cliente en particular que se caracterice por tener un nivel de deuda “alto”.
- A priori, no se puede concluir que el tamaño de la empresa tenga una relación directa con el nivel de deuda, ya que no presentan niveles de Cos2 significativos o una proximidad con alguna categoría de la variable “Deuda”.
- Contrario a lo que se planteaba inicialmente, los clientes “chicos” son los que menos meses pendientes tienen de deuda.

Vale aclarar que si bien esta herramienta nos permite formar “perfiles” según la correlación de las variables entre sí, no hay que caer en el error de la generalización, ya que cada empresa es única y puede tener su propia realidad particular. Esta técnica estadística nos sirve para tener un pantallazo general de ciertos “patrones” que pueden darse de manera natural entre los distintos perfiles.

Retroalimentación a PH:

Se recomienda al área de cobranzas pensar un incentivo para recompensar a los clientes que llevan una cuenta “sin deuda” y evitar el pago a mes vencido. Esto podría ser de utilidad sobre todo a las empresas grandes y del rubro industria manufacturera.

También se recomienda revisar el estado de las cuentas del rubro “Esparcimiento / recreación”, ya que es un rubro considerablemente lejano a la categoría Sin Deuda, y que se encuentra entre las variables de nivel de deuda “Bajo” y “Medio”. Posiblemente sea un rubro sutilmente más complicado que los demás en cuanto a la gestión de cobranzas.

Otro caso a revisar es el de las empresas del rubro “Industria Manufacturera”, no por su cercanía a alguna categoría en específico sino mas bien por lo contrario, su gran distancia respecto a las variables de nivel de deuda “Bajo” y “Sin deuda”.

7. CONCLUSIONES FINALES Y REFLEXIONES

En plena revolución digital y la era del Big Data, las empresas se enfrentan a un volumen de datos cada vez mayor, generados día a día en cada una de sus operaciones. Estos datos son una valiosa fuente de información que puede



impulsar el crecimiento y la toma de decisiones estratégicas. Es aquí donde entra en juego el Business Analytics, un proceso esencial para recopilar, organizar, analizar y visualizar datos con el objetivo de obtener información significativa y basar las decisiones en evidencia sólida.

La importancia del tratamiento de los datos radica en su potencial para revelar patrones, tendencias y relaciones ocultas. Los datos pueden proporcionar información valiosa sobre el comportamiento de los clientes, las preferencias del mercado, las eficiencias operativas y mucho más. Al aprovechar esta información, las empresas pueden tomar decisiones más informadas y estratégicas, lo que les permite adaptarse rápidamente a los cambios del mercado y obtener una ventaja competitiva. Año a año la tecnología nos va empujando a una realidad donde los datos son cada vez más abundantes y complejos, y aquellas empresas que logren estar a la vanguardia de la generación del conocimiento y la información son los que tendrán una verdadera ventaja competitiva a la hora de ser más eficientes en los procesos y más astutos en la oferta de valor.

Un elemento fundamental a la hora del Business Analytics son los sistemas de información. Estos permiten recopilar y almacenar datos de manera eficiente, asegurando su integridad y disponibilidad. Además, los sistemas de información facilitan el análisis y la visualización de los datos, lo que ayuda a los gerentes y tomadores de decisiones a comprender mejor la información y actuar en consecuencia. Es realmente necesario para una empresa la inversión en un buen sistema, que pueda compilar todos estos datos de manera fiable, de tal forma que sirvan de input para todos los análisis que posteriormente se pueden realizar.

Todo este ejercicio de recopilación y transformación de los datos va generando una suerte de “conocimiento organizacional sistematizado”, el cual también es un aspecto crucial en el desempeño de una empresa. Al tener un enfoque estructurado para capturar y organizar el conocimiento interno de la empresa, se puede aprovechar al máximo la información generada. Esto implica documentar las mejores prácticas, los procesos eficientes y las lecciones aprendidas, lo que permite una toma de decisiones más informada y basada en la experiencia acumulada.

En definitiva, considero que para ser competitivos en esta nueva década las empresas deben obligadamente plantearse la necesidad de incorporar el análisis de datos en el proceso operativo de las mismas. Para ello se debe contar con un sistema de gestión de calidad, que recopile todos los datos que la empresa genera con sus operaciones diarias y que luego serán los inputs de las técnicas de análisis. También se debe invertir en personal capacitado que tenga las competencias y conocimientos necesarios para poder transformar esos datos en información y conocimiento relevante, para poder realizar los informes necesarios para cualificar los procesos internos o la propuesta de valor que se esta ofreciendo al mercado. Y por último, pero no menos importante, la empresa debe estar dispuesta a prestar especial atención a las conclusiones que dichos informes contengan, ya que esa retroalimentación continua y el constante perfeccionamiento lo que hace que una empresa sea grande, ordenada, eficiente y exitosa.



8. BIBLIOGRAFÍA

- Aldás, J., Uriel, E. (2017). Análisis multivariante aplicado con R. Madrid, España. Editorial: Paraninfo.
- Anthony R. y Govindarajan V. (2008). Sistemas de Control de Gestión. Madrid, España. Editorial: McGraw Hill.
- Gómez Bravo L., Rodríguez Sánchez F. y Garmendia Suárez I. (2011). Indicadores de Gestión en la Empresa. Qué Medir y Cómo Utilizarlos. Caracas, Venezuela. Editorial: FIM Productividad.
- Hair, Anderson, Tatham , Black (1999) Análisis Multivariante. Madrid, España. Editorial: Prentice Hall Iberia.
- Hansen R. y Mowen M. (2007). Administración de Costos, Contabilidad y Control. México DF, México. Editorial: CENGAGE Learning.
- Hernández-Sampieri, R., & Mendoza, C. (2018). Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta. México. McGraw Hill.
- Horngren, Charles T, Srinkant M. y Rajan Madhav V (2012). Contabilidad de Costos. España. Editorial: PEARSON.
- Perez Carballo Veiga J. (2013). Control de Gestión Empresarial. Madrid, España. Editorial: ESIC.
- Render, B., Stair, R., & Hanna, M. (2012). Métodos cuantitativos para los negocios. México D.F. Editorial: Pearson.
- Kaplan, Robert S. y Norton, David P. (1996). El Cuadro de Mando Integral. Barcelona, España. Editorial: Gestión 2000.
- Simons, R.t (1998). Palancas de Control. Boston, EEUU. Editorial: TEMAS.



9. ANEXO - SCRIPTS UTILIZADOS

A continuación se detallan los Scripts utilizados en R studio para la aplicación del ACM. Donde diga “datos” en negrita, corresponde a las bases de datos que se deben cargar previamente en el software.

- `install.packages("FactoMineR")`
- `library(FactoMineR)`
- `MCAview<- MCA(datos)`
- `ggplot(datos, aes(x=RUBRO))+ geom_bar(fill= "#DDB4EB")`
- `ggplot(datos, aes(x=TAMAÑO))+ geom_bar(fill= "#FFD4A5")`
- `ggplot(datos, aes(x=DEUDA))+ geom_bar(fill= "#41894A")`
- `res.mca <- MCA(datos, graph = FALSE)`
- `print(res.mca)`
- `print(res.mcavarcos2)`
- `fviz_mca_var(res.mca, col.var = "var.cos2", gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"), ggtheme = theme_grey(), repel = TRUE)`
- `fviz_screplot(res.mca,geom="line")+theme_grey()`
- `fviz_mca_var(res.mca, col.var = "blue", addEllipses = FALSE, repel = TRUE) + theme_minimal()`