

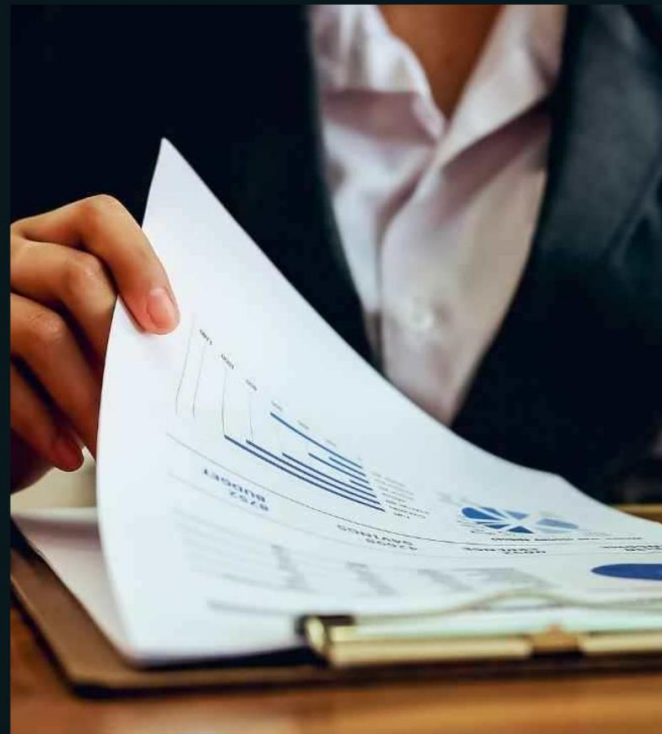


Universidad Nacional de Tucumán  
Facultad de Ciencias Económicas  
Instituto de Administración  
**XII Muestra Académica de Trabajos de  
Investigación de la Licenciatura en  
Administración**



# People Analytics y Ausentismo

Análisis Cuantitativo de Negocios II  
Instituto de Administración



**OMAR A. BRANDÁN**



## ÍNDICE

<b>RESUMEN.....</b>	<b>3</b>
<b>INTRODUCCIÓN .....</b>	<b>3</b>
<b>PRESENTACIÓN DEL PROBLEMA.....</b>	<b>4</b>
<b>Preguntas de Investigación .....</b>	<b>5</b>
<b>OBJETIVOS.....</b>	<b>5</b>
<b>Objetivo General .....</b>	<b>5</b>
<b>Objetivos Específicos.....</b>	<b>5</b>
<b>MARCO TEÓRICO .....</b>	<b>5</b>
<b>People Analytics .....</b>	<b>5</b>
<b>Ausentismo.....</b>	<b>6</b>
<b>Análisis Multivariante .....</b>	<b>6</b>
<b>Lenguaje de programación R .....</b>	<b>8</b>
<b>MARCO METODOLÓGICO .....</b>	<b>8</b>
<b>APLICACIÓN .....</b>	<b>9</b>
<b>Análisis de Clústers.....</b>	<b>9</b>
<b>Análisis de Correspondencias Múltiple.....</b>	<b>12</b>
<b>ANÁLISIS DE RESULTADOS.....</b>	<b>18</b>
<b>CONCLUSIONES .....</b>	<b>18</b>
<b>REFERENCIAS.....</b>	<b>19</b>
<b>APÉNDICE .....</b>	<b>19</b>



---

## **PEOPLE ANALYTICS Y AUSENTISMO**

### **Estudio de caso en empresa de electrodomésticos**

**OMAR AGUSTÍN BRANDÁN**

Instituto de Administración – Facultad de Ciencias Económicas – Universidad Nacional de Tucumán  
[omaraugustinbrandan@gmail.com](mailto:omaraugustinbrandan@gmail.com)

### **RESUMEN**

Con la evolución de Big Data, varias tendencias están cambiando el mundo del trabajo y abonando el camino para el auge de People Analytics, su uso surge como una respuesta a las necesidades que se presentan en el área de Talento Humano para mejorar sus procesos y generar soluciones prácticas y eficientes. People Analytics permite eliminar la incertidumbre en la Gestión del Talento Humano y facilita la toma de decisiones desde esta perspectiva y con una clara alineación con los objetivos de la Organización.

El presente trabajo se realizará en una empresa comercial dedicada a la comercialización de electrodomésticos y productos para el hogar con sede en la provincia de Tucumán. Por cuestiones de confidencialidad será denominada como “La Mentora S.A.”.

La Gerencia de Recursos Humanos, en búsqueda de potenciar la productividad de una nómina que supera las 240 personas distribuidas en diferentes sucursales y el centro logístico, detectó un alto índice de ausentismo que imposibilitaba la misma.

Esta investigación es una aplicación de conceptos de People Analytics por medio de la técnica de Análisis de Conglomerados o Clúster y Análisis de Correspondencias Múltiple, que tendrá como principal objetivo brindar información oportuna para detectar causas y tendencias de faltas de los empleados en el trabajo e identificar aspectos a mejorar que contribuyan a una mejor gestión.

Para ello se recopilaron y depuraron datos, se identificaron y se caracterizaron grupos de personal en las distintas sucursales y en los diversos puestos de trabajo mediante el software estadístico R.

**Palabras Clave:** Análisis de Clúster - Análisis de Correspondencias Múltiple - People Analytics - Software R

### **INTRODUCCIÓN**

Con la evolución de Big Data, varias tendencias están cambiando el mundo del trabajo y abonando el camino para el auge de People Analytics. El impacto de la tecnología digital en el mercado laboral, el uso de las redes sociales, la influencia de los Millennials, la globalización en la economía laboral, la disminución de la lealtad de los empleados y la consiguiente necesidad de que los empleadores involucren talento competente en la organización, han generado presión interna sobre la Gestión de Talento Humano, es decir, se presenta la necesidad de conseguir empleados con mejores competencias que tenga buenas habilidades y que agreguen valor a la compañía. Teniendo en cuenta lo anterior, el uso de



People Analytics surge como una respuesta a las necesidades que se presentan en el área de Talento Humano para mejorar sus procesos y generar soluciones prácticas y eficientes.

People Analytics permite eliminar la incertidumbre en la Gestión del Talento Humano y facilita la toma de decisiones desde esta perspectiva y con una clara alineación con los objetivos de la Organización, aplicándose de 6 formas:

1. Identificación de departamentos o personas que necesitan mayor atención.
2. Determinación de las acciones que tienen mayor impacto en los resultados
3. Pronóstico de los niveles de fuerza laboral
4. Conocer las razones que llevan a las personas a dejar la organización
5. Adaptación de la plantilla a los cambios del entorno empresarial.

La empresa que se tomará de referencia para la realización de este trabajo se denominará "La Mentora S.A." como nombre de fantasía para preservar su identidad e información. La misma se dedica a la venta por mayor y menor de electrodomésticos para el hogar (heladeras, cocinas, computadoras, equipos de audio, etc.), cuenta con once sucursales distribuidas en la provincia de Tucumán, dos en la provincia de Salta y una en Catamarca. Además, dispone de un centro logístico de operaciones, ubicado en el parque industrial de Tucumán, el cual funciona como almacén central que se encarga del abastecimiento de mercadería a todas las sucursales, y por último un área e-commerce que se dedica a la comercialización de los productos a través de la página web de la empresa, ubicada en las inmediaciones del centro logístico.

Al día de la fecha cuenta con una nómina que supera las 240 personas distribuidas en las diferentes sucursales y el centro logístico. Debido al gran volumen de personal que integran la empresa es que resulta indispensable una Gerencia de Recursos Humanos, que actualmente lleva el nombre de "Gerencia de Talento y Desarrollo Organizacional", que se encargue de atender a la productividad del personal generando el clima laboral adecuado.

Se consideró oportuno que para ello las decisiones a tomarse se basen no sólo en intuición por parte de expertos en el área sino también mediante técnicas analíticas son el análisis de clúster y el análisis de correspondencias múltiple.

Estas técnicas multivariantes buscan clasificar objetos, en este caso colaboradores, logrando un enfoque de gestión del talento humano basado en datos, que serán recolectados de las planillas de la propia empresa que contienen motivos y días de ausencia, edad, antigüedad, puesto y tipo de contrato.

## **PRESENTACIÓN DEL PROBLEMA**

La empresa dispone de diferentes planillas de Excel para la carga de novedades, cronograma de vacaciones, fechas de vencimientos de contratos, etc. Al contar con una diversidad de planillas, analizar la información de manera integral se dificulta.

A su vez, la gerencia le da poca relevancia al análisis de los datos que se recolectan del personal, los cuales son una fuente de información valiosa que está siendo desaprovechada y en la que podría otorgar puntos de partida para



contribuir a la mejora de la gestión de recursos humanos, donde se detectó cierto patrón habitual de ausencia por parte del personal.

### **Preguntas de Investigación**

- ¿Cómo debe representarse la información para comprender el patrón de ausentismo?
- ¿De qué manera pueden clasificarse los empleados?
- ¿De qué manera se relacionan los motivos de ausencia del personal, sus puestos de trabajo y el tipo de contrato?

### **OBJETIVOS**

#### **Objetivo General**

Brindar a La Mentora S.A. información que permita comprender el patrón de ausentismo presente en su plantilla de empleados.

#### **Objetivos Específicos**

- Clasificar, mediante Análisis de Clúster, a los empleados en grupos según su edad, antigüedad y días de ausentismo.
- Representar, mediante Análisis de Correspondencias Múltiple, la relación entre los motivos de ausencia, los puestos de trabajo y el tipo de contrato del personal.

### **MARCO TEÓRICO**

#### **People Analytics**

Según Jaramillo Toledo, Rincón Rodríguez, Sánchez Jaimes y Sierra Meneses (2021), People Analytics es la intersección de las Ciencias del Comportamiento, Matemáticas, Tecnologías y Gestión de los Recursos Humanos. Surge de Business Intelligence y Big Data, y consiste en la identificación y cuantificación sistemática de la influencia de los factores humanos en la actividad de la organización. Es decir, es un enfoque de Gestión de Talento Humano basado en datos que permite responder preguntas como: ¿cuál es el daño a la organización debido a la rotación de empleados? ¿Cuáles son los factores que determinan la rotación de empleados? ¿Cuáles son los mayores riesgos en la fuerza laboral? En general, People Analytics permite eliminar la incertidumbre en la Gestión del Talento Humano y facilita la toma de decisiones desde esta perspectiva y con una clara alineación con los objetivos de la organización.

Por su parte, Cravino (2017) concluye que People Analytics no se basa en la mera obtención de datos, sino que también será necesario detectar cuáles son las variables más importantes que desea conocer la organización y proceder luego, a la medición de las mismas para la toma de decisiones estratégicas.



## Ausentismo

Chiavenato (2019) asegura que el ausentismo designa las faltas o ausencias de los empleados en el trabajo. En un sentido más amplio, es la suma de los periodos en que los empleados se encuentran ausentes en el trabajo, ya sea por falta, retardo o algún otro motivo. Las causas del ausentismo son diversas y se carece de estudios acerca de su análisis, valores reales, identificación de tipos, calidad, tamaño y repercusión del ausentismo en la productividad.

No siempre las causas del ausentismo se deben al propio empleado, muchas veces se deben a la organización, a una supervisión deficiente, al empobrecimiento de las tareas, a la falta de motivación y estímulo, a las condiciones desagradables de trabajo, a la precaria provisión del empleado en la organización y a los impactos psicológicos de una dirección deficiente. En la práctica, las causas principales de ausentismo son:

1. Enfermedad comprobada.
2. Enfermedad no comprobada.
3. Razones diversas de carácter familiar.
4. Retardos involuntarios por causas de fuerza mayor.
5. Faltas voluntarias por motivos personales.
6. Dificultades y problemas financieros.
7. Problemas de transporte.
8. Poca motivación para el trabajo.
9. Supervisión precaria de los jefes.
10. Políticas inadecuadas de la organización.

Algunos especialistas consideran causas de ausentismo los accidentes de trabajo, lo que crea confusión al comparar los índices de ausentismo de diversas organizaciones. Las causas de ausentismo deben diagnosticarse, y así emprender una acción coordinada en el nivel de la supervisión con el debido apoyo de las políticas de la organización y de la dirección para reducir los niveles de ausencia y retardos del personal.

## Análisis Multivariante

Hair, Anderson, Tatham y Black (1999) definen al análisis multivariante como un conjunto de técnicas de análisis de datos en expansión. Entre las técnicas más conocidas tenemos regresión múltiple y correlación múltiple; análisis discriminante múltiple; componentes principales y análisis factorial común; análisis multivariante de varianza y covarianza; correlación canónica; (1) análisis clúster; análisis multidimensional y análisis conjunto. Entre las técnicas emergentes también incluidas están (2) análisis de correspondencias; modelos de probabilidad lineal como logit y probit; y modelos de ecuaciones simultáneas/estructurales.

Los autores recién mencionados afirman que el examen previo de los datos es un paso necesario, que lleva tiempo, y que habitualmente se descuida por parte de los analistas de datos. Un análisis cuidadoso de los datos conduce a una mejor predicción y a una evaluación más precisa de la dimensionalidad.



Si bien las técnicas multivariantes suponen un tremendo poder analítico en manos del investigador, también crean una gran carga para éste, ya que tiene que asegurarse de que se mantengan los cimientos teóricos y estadísticos sobre las que se basan. Examinando con anterioridad los datos de la aplicación de una técnica multivariante, el investigador gana una comprensión básica de los datos y las relaciones entre las variables. En primer lugar, el investigador obtiene un conocimiento básico de los datos y las relaciones entre las variables. Las técnicas multivariantes plantean grandes demandas al analista en la comprensión, interpretación y articulación de resultados basados en relaciones cuya complejidad está en continuo aumento. El conocimiento de las interrelaciones de variables puede ayudar enormemente en la especificación y refinamiento del modelo multivariante, así como proporcionar una perspectiva razonable para la interpretación de los resultados. En segundo lugar, las técnicas multivariantes demandan mucho más de los datos que se van a analizar.

Proceso de aplicación de una técnica multivariante:

Paso 1. Definición del problema que se está investigando, sus objetivos y decisión sobre la técnica multivariante que se debe utilizar. Primero hay que centrarse en el tema que se investiga, y no en las técnicas que se van a emplear, lo que evitaremos que dejemos fuera del análisis conceptos importantes.

Paso 2. Desarrollo del plan de análisis. Una vez establecido el modelo conceptual, el énfasis se centra en aplicar adecuadamente la técnica elegida.

Paso 3. Análisis de las condiciones de aplicabilidad de la técnica elegida. Una vez recogidos los datos, es necesario conocer cuáles son las hipótesis en que se basan las técnicas multivariantes y, que si no se cumplen, hace que carezca de sentido aplicarlas.

Paso 4. Estimación del modelo multivariante y ajuste de global del mismo. Verificar si el nivel de bondad del ajuste es adecuado.

Paso 5. Interpretación de los resultados. Poner atención en los efectos de las variables individuales examinando sus coeficientes, cargas factoriales, utilidades.

Paso 6. Validación del modelo. Antes de aceptar los resultados, se debe aplicar una serie de técnicas de diagnóstico que aseguren que estos resultados son generalizables al conjunto de la población.

- (1) El Análisis Clúster es la denominación de un grupo de técnicas multivariantes cuyo principal propósito es agrupar objetos basándose en las características que poseen. Clasifica objetos de tal forma que cada objeto es muy parecido a los que hay en el conglomerado con respecto a algún criterio de selección predeterminado. Los conglomerados de objetos resultantes deberían mostrar un alto grado de homogeneidad interna (dentro del conglomerado) y un alto grado de heterogeneidad externa (entre conglomerados). Por tanto, si la clasificación es acertada, los objetos dentro de los conglomerados estarán muy próximos cuando se representen gráficamente, y los diferentes grupos estarán muy alejados. Desde la derivación de taxonomías en biología para la agrupación de todos los organismos vivos a clasificaciones psicológicas basadas en la personalidad y otros rasgos personales,



pasando por los análisis de segmentación de los mercados, el análisis cluster ha tenido siempre una fuerte tradición en la agrupación de individuos.

- (2) El Análisis de Correspondencias es una técnica de interdependencia recientemente desarrollada que facilita tanto la reducción dimensional de una clasificación de objetos sobre un conjunto de atributos y el mapa perceptual de objetos relativos a estos atributos. Proporciona una representación multivariante de la interdependencia de datos no métricos que no es posible realizar con otros métodos.

Según Aldás y Uriel (2017), a diferencia de añadir variables filas o columna a la solución de un análisis de correspondencias simple, en el Análisis de Correspondencias Múltiple (ACM) todas las variables contribuyen a conformar los ejes. Los datos pueden aparecer presentados de distinta forma. Podemos tenerlos con las distintas variables categóricas en columnas y los individuos en filas o como un conjunto anidado de tablas cruzadas donde todas las variables se cruzan con todas, lo que se conoce como Matriz de Burt. También pueden tenerse como matriz de indicadores, que tendría tantas filas como casos y las columnas no serían las variables, sino las categorías de las mismas que tomarían el valor de 1 en el caso en que ese individuo tenga el atributo que representa esa categoría de la variable en cuestión, y cero en caso contrario.

### **Lenguaje de programación R**

Jaramillo, Rincón, Sánchez y Meneses (2021) resaltan la importancia de los lenguajes de programación, que en la actualidad se adaptan a las necesidades de las personas y, por ende, a las de las empresas que quieren mantenerse en una carrera tecnológica sin precedentes. Tener información sobre ellos y mantenerse actualizado es una responsabilidad de aquellos que forman parte del tejido de la empresa. En la industria, los dos lenguajes de programación más populares en la actualidad para el trabajo de ciencia de datos son R y Python. Ambos son hoy en día los lenguajes de programación más utilizados como interfaz de herramientas estadísticas para análisis de datos, siendo muy populares en el campo de la minería de datos. Su espectro de aplicabilidad cubre desde la investigación biomédica, el análisis web, la economía y las ciencias sociales, hasta implementaciones en bioinformática, actuaría y matemáticas financieras.

R es un entorno enfocado al análisis estadístico que puede ser considerado también como una suite integrada de facilidades de software para manipulación de datos, cálculos y herramientas gráficas.

### **MARCO METODOLÓGICO**

La metodología elegida para llevar a cabo la investigación se sustenta a través del libro “Metodología de la Investigación” desarrollado por Hernández Sampieri (2018).





Las herramientas de análisis estadístico multivariado que se utilizarán, mediante la ayuda del software estadístico R, serán Análisis de Clústers y Análisis de Correspondencias Múltiple, ambas técnicas exploratorias y con relaciones entre las variables de interdependencia.

Clustering se llevará a cabo mediante la aplicación del método no jerárquico de k-medias y ACM mediante el resumen de un espacio de propiedades generando nuevas variables-resumen.

El enfoque de investigación es, por ende, cuantitativo, con un diseño no experimental de corte transversal con un alcance exploratorio y descriptivo.

La muestra seleccionada es de tipo no probabilístico, ya que el procedimiento no es mecánico ni se basa en fórmulas de probabilidad, sino que dependerá del proceso de toma de decisiones.

La técnica de recolección de datos será a través de revisión de las bases de datos con las que cuenta la empresa. Para el tratamiento de los mismos se procedió realizar actividades de data cleaning que consistieron en limpiar y depurar los datos al detectar, corregir y eliminar registros corruptos o imprecisos.

## **APLICACIÓN**

En esta sección se desarrolla la aplicación de las técnicas propuestas.

### **Análisis de Clústers**

El análisis comienza con la recopilación de los datos iniciales, referidos al personal de la empresa La Mentora S.A. entre los años 2018 y 2022.

La base de datos original cuenta con 10 variables medidas para 1645 registros, procedentes de la actividad de la empresa en el periodo bajo análisis.

En una primera inspección, se detectan variables que tienen medición 0, otras que surgen de operaciones matemáticas entre las mismas, y registros incompletos o mal cargados. Se procede en segunda instancia a eliminar los registros incompletos o incongruentes y a quitar del análisis las variables innecesarias. La base de datos depurada es la siguiente (se muestran los 5 primeros registros y las variables codificadas):

Tabla 1: base de datos depurada para análisis de clústers

<b>Edad</b>	<b>Antigüedad</b>	<b>Días</b>
30	1	0,5
30	1	0,5
31	2	1
32	3	1
32	3	12

Fuente: elaboración propia



A continuación, se cargó la base de datos en R y se procedió a explorar la misma. La base de datos de trabajo contiene 3 variables y 1632 registros. Se procede a realizar un análisis para la detección de observaciones atípicas y así lograr una mayor exactitud en los resultados. Se utiliza la distancia de Mahalanobis (detección multivariantes de casos atípicos).

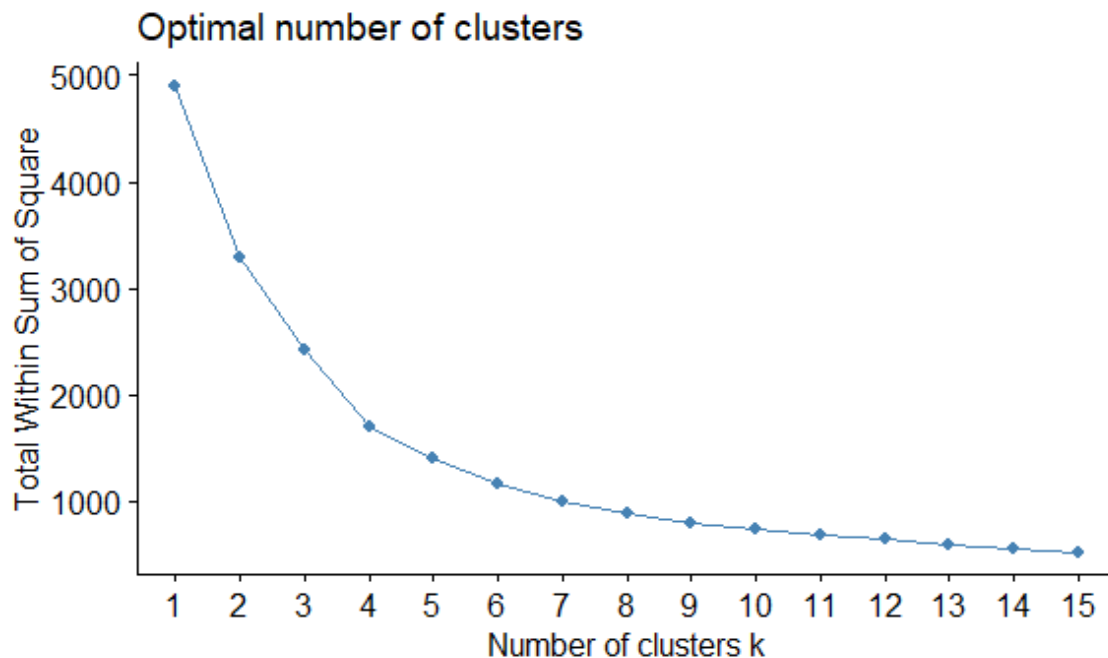
Se obtuvo como valor crítico del test de Mahalanobis, 13,2767, por lo cual se procede a depurar la base de datos previa, quitando los registros que están por encima de este valor (por considerarlas valores atípicos. No se eliminaron observaciones.

Como los valores de las variables difieren significativamente entre ellos, y para eliminar un posible efecto negativo por dicho motivo a la hora de realizar el análisis de clúster, se procedió a escalar la base de datos (ver Apéndice – Scripts 1).

Se procede a continuación a realizar el análisis de clúster (ver Apéndice – Scripts 2). Como paso siguiente, se obtiene la matriz de distancias. Para la misma, se elige trabajar con la distancia euclídea (ver Apéndice – Scripts 3).

Luego de esto, se procede a buscar el número de clústers para realizar el análisis (ver Apéndice – Scripts 4). Se obtienen las siguientes gráficas:

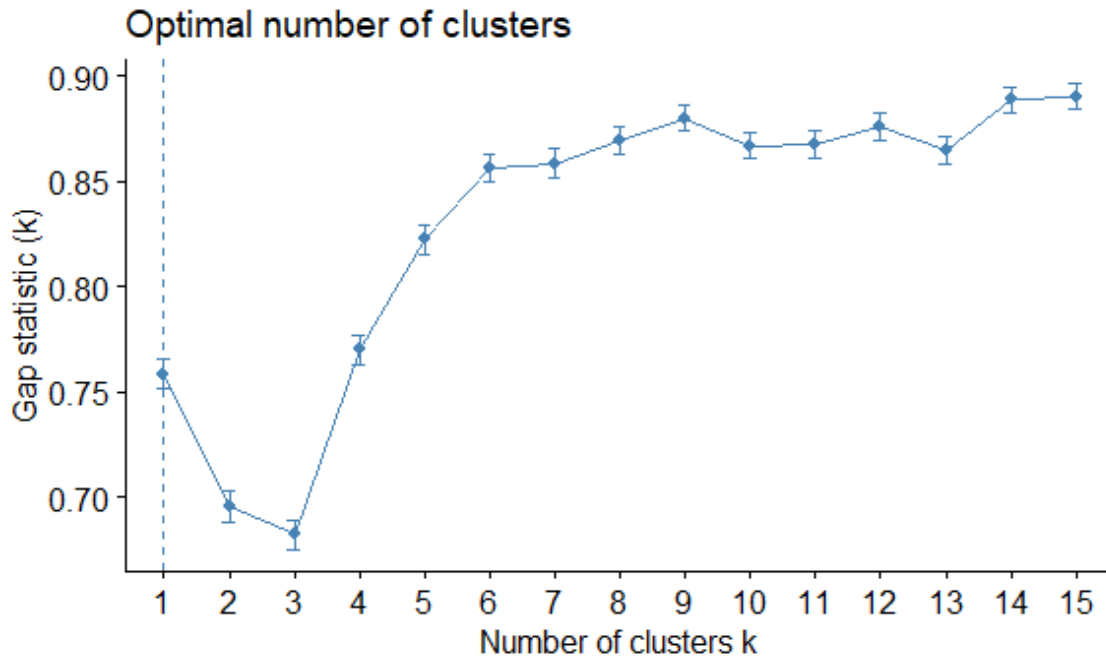
Gráfico 1: número óptimo de clusters según el método Within Sum Square



Fuente: elaboración propia en RStudio

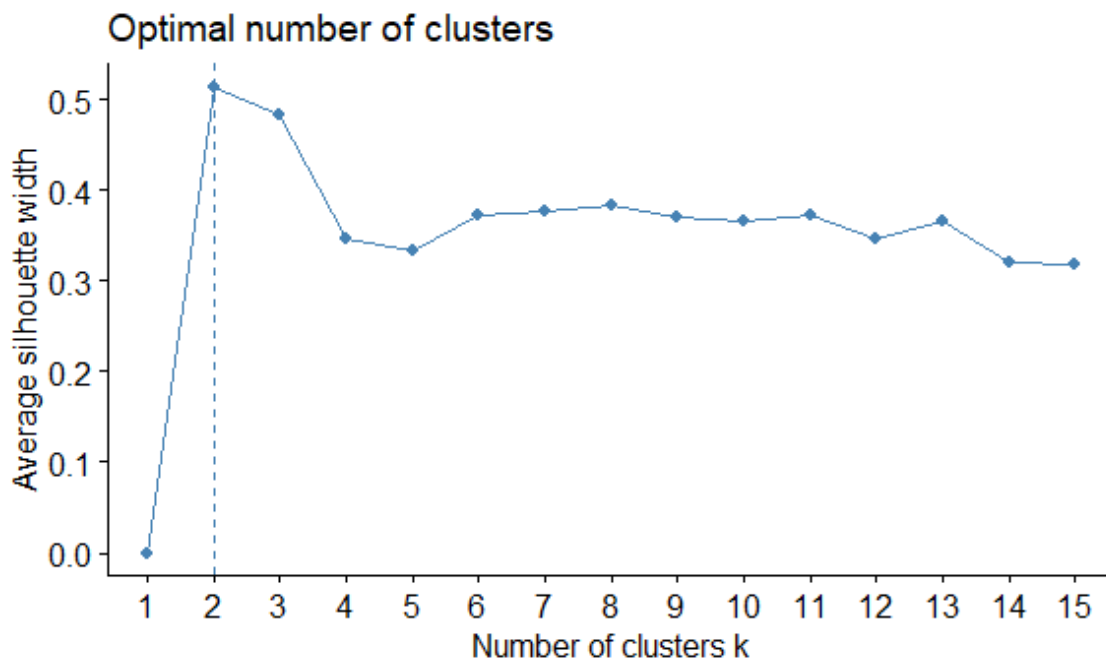


Gráfico 2: número óptimo de clusters según el método Gap Statistic



Fuente: elaboración propia en RStudio

Gráfico 3: número óptimo de clústers según el método Average Silhouette

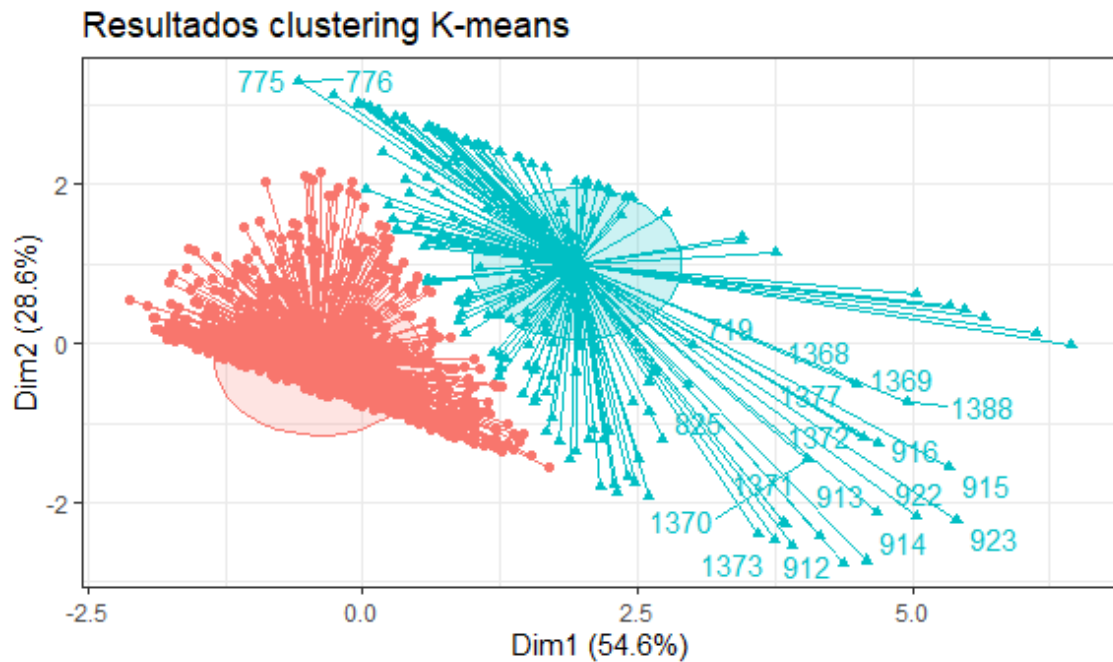


Fuente: elaboración propia en RStudio



En base a los métodos Within Sum Square y Average Silhouette, se eligió un número óptimo de clústers (k) de 2. Se decide utilizar el método k-means (ver Apéndice – Scripts 5), obteniendo los siguientes 2 grupos:

Gráfico 4: clústers definidos (k=2) para el personal de La Mentora S.A. entre 2018 y 2022



Fuente: elaboración propia en RStudio

Finalmente, se procede a obtener los datos de los grupos (ver Apéndice – Scripts 6). Resulta relevante conocer las medias de cada grupo:

Tabla 2: medias de cada uno de los clústers k-means para el personal de La Mentora S.A. entre 2018 y 2022

Media de Cada Variable por Conglomerado, Utilizando los datos originales

cluster	Edad	Antigüedad	Días
1	38.43	5.55	3.59
2	50.21	13.17	21.71

Fuente: elaboración propia en RStudio

### Análisis de Correspondencias Múltiple

Los primeros pasos para aplicar esta técnica se realizaron de igual manera que en el análisis de clústers, con la diferencia de que también se tuvo que hacer una modificación en las categorías de las variables bajo estudio.



En “Puesto” se agruparon 6 categorías en 3: Administración, Auxiliar, Ventas; en “Contrato”, 8 categorías en 3: Eventual, Permanente, Tiempo determinado; y en “Motivos” (de ausencia), 6 en 4: ART, Ausente, Enfermedad, Lic. Legal. Dejando como resultado 3 variables, 1645 registros y la siguiente base depurada:

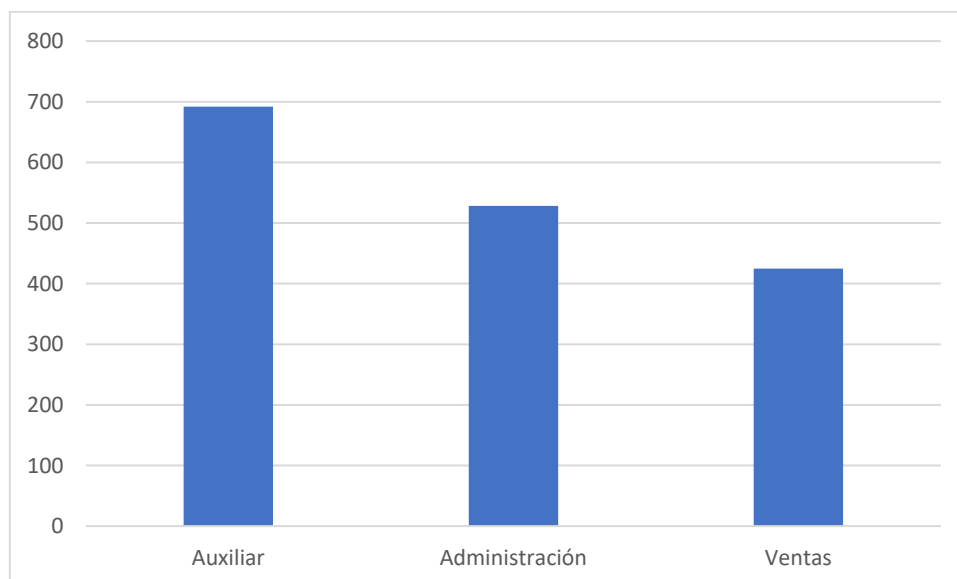
Tabla 3: base de datos depurada para ACM

PUESTO	CONTRATO	MOTIVOS
Administración	Permanente	Enfermedad
Administración	Permanente	Enfermedad
Administración	Permanente	Ausente
Administración	Permanente	Lic. Legal
Administración	Permanente	Lic. Legal

Fuente: elaboración propia

Como paso siguiente se cargaron los paquetes necesarios (ver Apéndice – Scripts 7) y se procedió a un análisis descriptivo de las tres variables categóricas en cuestión, donde se observa que la mayoría de los empleados que tienen faltas son auxiliares y con tipo de contrato permanente, y que la razón de ausencia más frecuente es la de enfermedad.

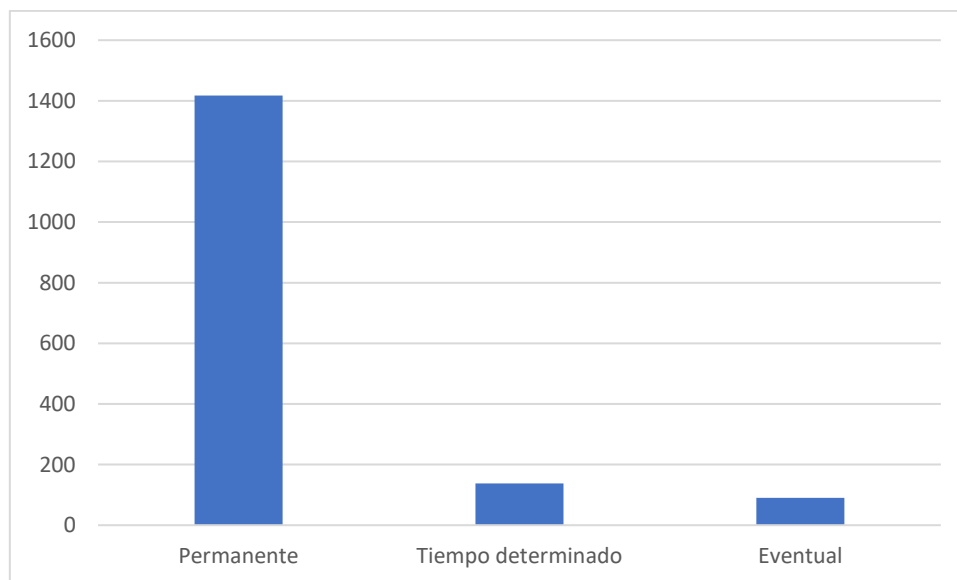
Gráfico 5: observaciones por categoría en Puesto



Fuente: elaboración propia

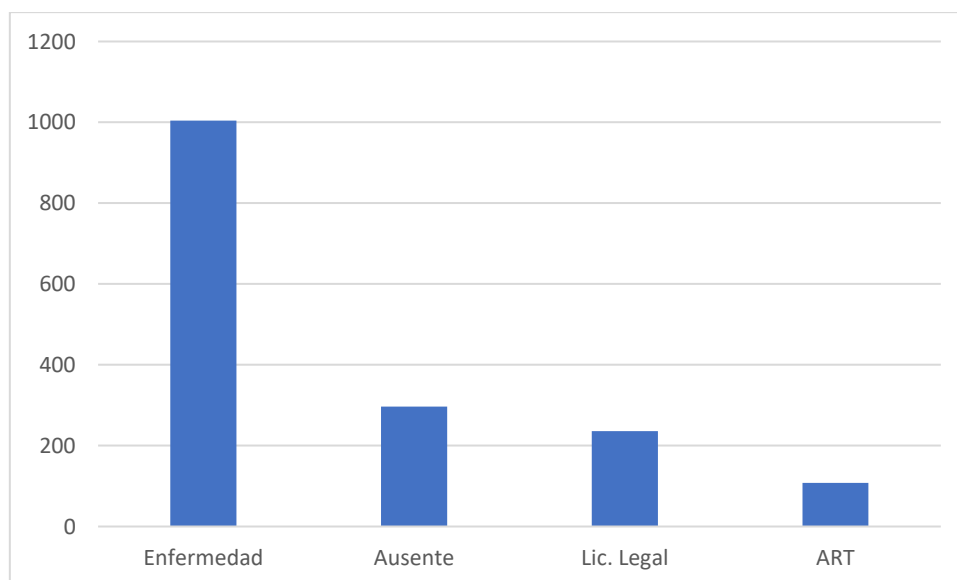


Gráfico 6: observaciones por categoría en Contrato



Fuente: elaboración propia

Gráfico 7: observaciones por categoría en Motivos

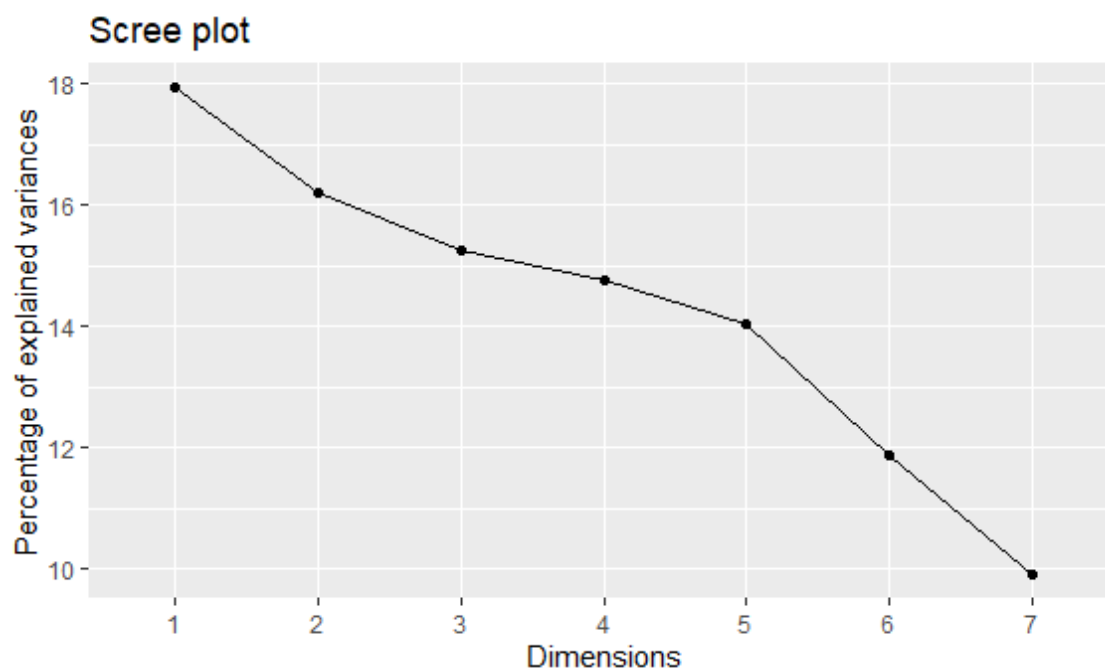


Fuente: elaboración propia



Se obtuvo el porcentaje de varianza explicada por dimensión (ver Apéndice – Scripts 8).

Gráfico 8: varianza explicada por dimensión 1 a 7

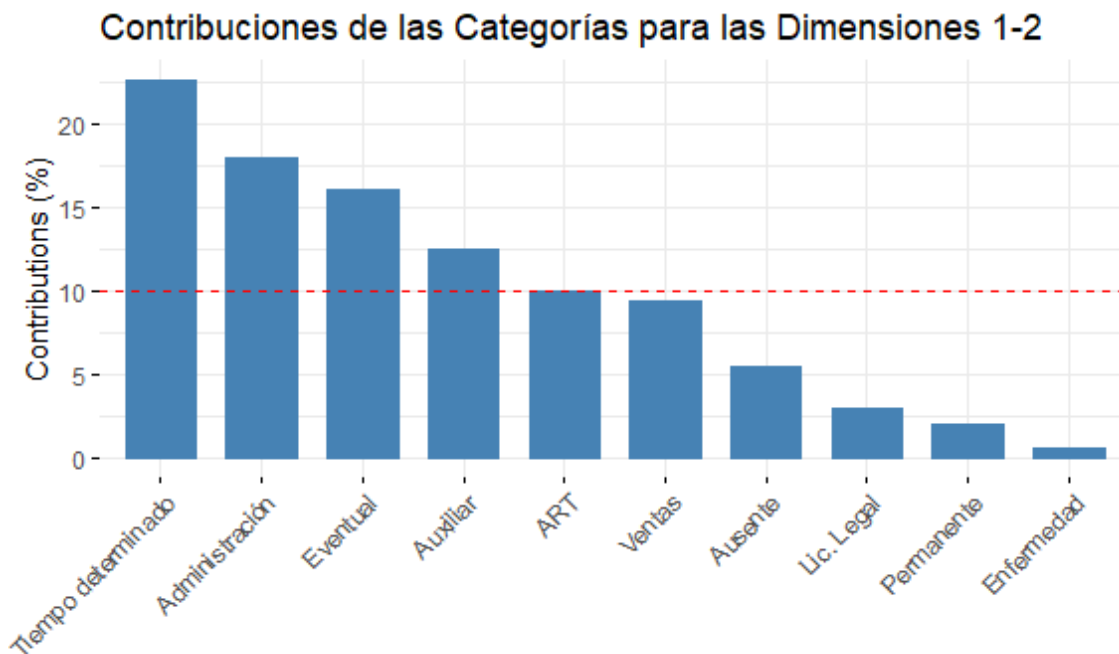


Fuente: elaboración propia en RStudio

Una vez analizado el gráfico anterior, con el cual podemos afirmar que con 2 dimensiones se explica un gran porcentaje de la varianza, se continuó con la contribución de las categorías para la dimensión 1 y la dimensión 2 (ver Apéndice – Scripts 9).



Gráfico 9: contribuciones de las categorías para las dimensiones 1 y 2



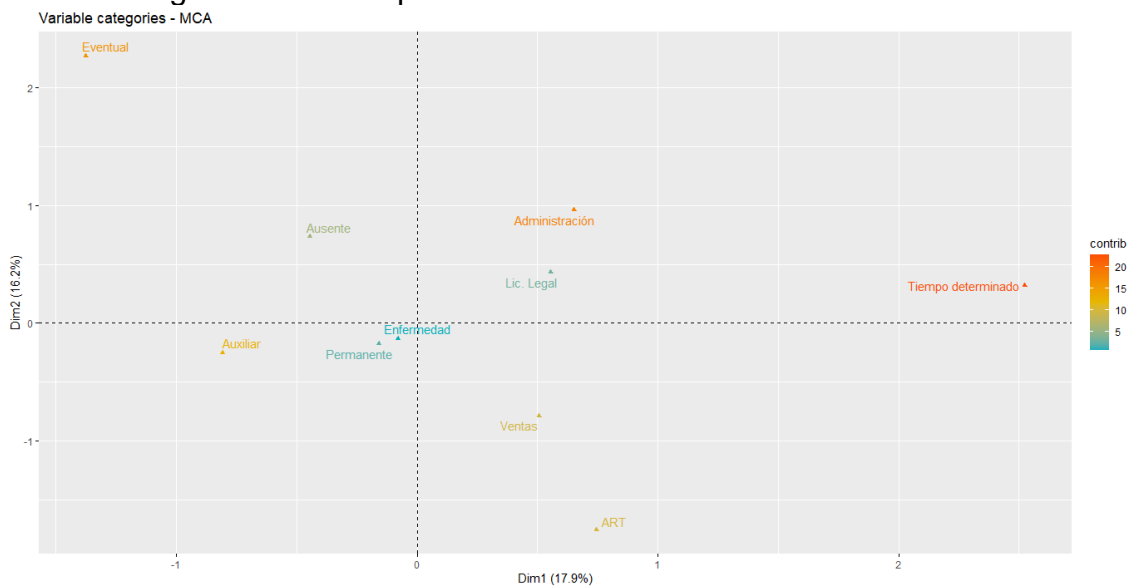
Fuente: elaboración propia en RStudio

Por último, se corren los scripts para conseguir un gráfico de calor por contribuciones (ver Apéndice – Scripts 10) con el objetivo de visualizar las correspondencias de las categorías de las variables bajo estudio y un gráfico de calor por  $\cos^2$  (ver Apéndice – Scripts 11) para representar la calidad de representación de las mismas.



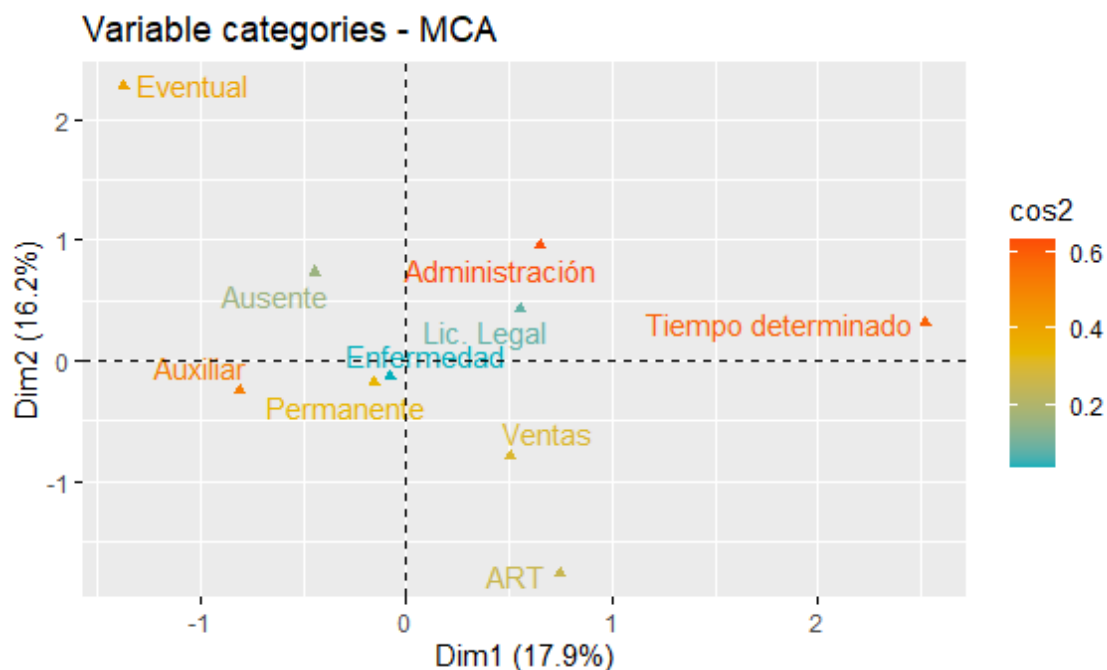


Gráfico 10: gráfico de calor por contribuciones



Fuente: elaboración propia en RStudio

Gráfico 11: gráfico de calor por cos<sup>2</sup>



Fuente: elaboración propia en RStudio



## ANÁLISIS DE RESULTADOS

Respecto a los resultados obtenidos mediante el Análisis de Conglomerados, podemos afirmar que los empleados de La Mentora pueden clasificarse en dos grupos:

El grupo 1, que se caracteriza por, en promedio, un ausentismo 22 días, un personal de 50 años de edad y una antigüedad de 13 años.

El grupo 2 contiene empleados que tienen en promedio 4 días de ausentismo, unos 39 años de edad y antigüedad de 6 años.

Se destaca la diferencia tanto de edad como de antigüedad entre los grupos, así como también de ausentismo, que es lo óptimo en un análisis de clúster y que fue lo buscado en ese trabajo de investigación, poder clasificar grupos objetivamente distintos.

Del Análisis de Correspondencia Múltiple se observan ciertas relaciones entre el ausentismo debido a enfermedad y los empleados auxiliares y en planta permanente. La posición de los puntos en el diagrama de dispersión de la categoría enfermedad debe ser interpretada con precaución al tener un  $\cos^2$  bajo, el cual mide el grado de asociación entre las categorías de las variables y un eje en particular, es decir, la calidad de la representación.

También encontramos correspondencias entre las ausencias debido a licencia legal y el personal administrativo, y, por último, entre ausentismo por accidentes de trabajo y vendedores.

El tipo de contrato eventual y con plazo determinado no tienen relación con ningún motivo de ausencia, lo cual cobra sentido al ser un trabajo con tiempo acortado.

Con respecto a otras categorías no podemos inferir lo mismo por la distancia entre ellas.

## CONCLUSIONES

Chiavenato (2019) afirma que la tendencia actual es atacar las causas de la rotación de personal o el ausentismo, y no sus efectos. Así, es fundamental diagnosticar las causas y los determinantes.

El trabajo de investigación realizado brinda a La Mentora información valiosa para comprender de una manera objetiva el fenómeno que se presenta actualmente en la organización y alcanzar sus objetivos de gestión de recursos humanos orientados a la productividad.

Dada la información obtenida por las dos técnicas de análisis multivariante, se concluye que la Gerencia del Talento y Desarrollo Organizacional debe poner foco en los empleados auxiliares en planta permanente que tienen una edad y antigüedad mayor, contribuyendo a emprender acciones coordinadas en el nivel de la supervisión con el debido apoyo de las políticas de la organización y de la dirección para reducir los niveles de ausencia del personal, ya sea modificando sus políticas de personal, rediseñando los puestos para hacerlos más atractivos y estimulantes, redefiniendo la gerencia para volverla más democrática y participativa, reconsiderando la remuneración para hacerla de



ganancia variable en función del desempeño y de las metas alcanzadas, entre otras.

## REFERENCIAS

- Aldás, J., Uriel, E. (2017). *Análisis multivariante aplicado con R*. Paraninfo.
- Chiavenato, I. (2019). *Administración de Recursos Humanos. El capital humano de las organizaciones*. Mc Graw Hill Interamericana.
- Cravino, L. (2017). *People Analytics es mucho más que Big Data*.  
<https://www.linkedin.com/pulse/people-analytics-es-mucho-m%C3%A1s-que-big-data-luis-maria-cravino/?originalSubdomain=es>
- D'Arterio, H. (2021). *Porqué Gestión de Recursos Humanos*. Ficha de cátedra de administración de recursos humanos I.
- García, J., Molina, J., Berlanga, A., Patricio M., Bustamante, L., Padilla, W. (2018). *Ciencia de Datos. Técnicas analíticas y aprendizaje estadístico*. Publicaciones Altaria, S.L.
- Hair, J., Anderson, R., Tatham, R., Black, W. (1999). *Análisis Multivariante*, 5° ed. Prentice Hall Iberia.
- Hernández Sampieri, R., Mendoza, C. (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill.
- Jaramillo Toledo, J., Rincón Rodríguez, D., Sánchez Jaimes, J., Sierra Meneses, J. (2021). *Beneficios de People Analytics en procesos de gestión de talento humano*. Universidad Sergio Arboleda.
- Lind, D., Marchal, W., Wathen, S. (2012). *Estadística aplicada a los negocios y la economía*. McGraw-Hill.
- Pereda Marín, S. (1997). *La Dirección de Recursos Humanos*. Fundación General Pontificia Universidad Politécnica de Madrid.

## APÉNDICE

### Scripts 1

```
RHC.esc <- scale(RHC, center = TRUE, scale = TRUE)
```



## Scripts 2

```
library(NbClust)  
library(cluster)  
library(ggplot2)  
library(factoextra)
```

## Scripts 3

```
dist <- hclust(d = dist(x = RHC.esc, method = "euclidean"), method = "complete")  
res.dist <- get_dist(RHC.esc, stand = TRUE, method = "euclidean")  
fviz_dist(res.dist, gradient = list(low = "#00AFBB", mid = "white", high =  
"#FC4E07"))
```

```
matriz.dis.euclid <- dist(RHC.esc, method = "euclidean", diag = TRUE)  
round(print(matriz.dis.euclid), 2)
```

## Scripts 4

```
fviz_nbclust(x = RHC.esc, FUNcluster = kmeans, method = "wss", k.max = 15,  
diss = get_dist(RHC.esc, method = "euclidean"), nstart = 50)
```

```
fviz_nbclust(x = RHC.esc, FUNcluster = kmeans, method = "gap_stat", k.max =  
15,  
diss = get_dist(RHC.esc, method = "euclidean"), nstart = 50)
```

```
fviz_nbclust(x = RHC.esc, FUNcluster = kmeans, method = "silhouette", k.max =  
15,  
diss = get_dist(RHC.esc, method = "euclidean"), nstart = 50)
```

## Scripts 5

```
set.seed(123)  
km.res <- kmeans(RHC.esc, 2, nstart = 25)  
print(km.res)  
fviz_cluster(object = km.res, data = RHC.esc, show.clust.cent = TRUE,  
ellipse.type = "euclid", star.plot = TRUE, repel = TRUE) +  
labs(title = "Resultados clustering K-means") +  
theme_bw() +  
theme(legend.position = "none")
```



## Scripts 6

```
library(kableExtra)
aggregate(RHC, by = list(cluster = km.res$cluster), mean) %>%
  kable(caption = "Media de Cada Variable por Conglomerado, Utilizando los
datos originales",
        align = "c",
        digits = 2) %>%
  kable_classic_2(html_font = "sans-serif",
                  lightable_options = c("hover", "striped")) %>%
  row_spec(0,
           bold = T,
           color = "white",
           background = "#219B6D")
```

## Scripts 7

```
library(ca)
library("ggplot2")
library("factoextra")
library("FactoMineR")
library("gridExtra")
```

## Scripts 8

```
res.mca <- MCA(RHA, graph = FALSE)
print(res.mca)

print(res.mca$var)

fviz_screplot(res.mca,geom="line")+
  theme_grey()
```

## Scripts 9

```
fviz_contrib(res.mca, choice = "var", axes = 1:2, top = 15)+labs(title = "
Contribuciones de las Categorías para las Dimensiones 1-2")
```

## Scripts 10

```
fviz_mca_var(res.mca, col.var = "contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             ggtheme = theme_grey(),
             , repel = TRUE)
```



---

## Scripts 11

```
fviz_mca_var(res.mca, col.var = "cos2",  
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
  ggtheme = theme_grey()  
  , repel = TRUE)
```