

Ciencia de datos y decisiones en la gestión de recursos humanos

Marcelo Enrique Medina Galván¹

Resumen

La ciencia de datos posee un inmenso potencial para transformar la gestión de recursos humanos, pero el verdadero reto radica en la capacidad de las organizaciones para convertir grandes volúmenes de datos en información útil y conocimiento que impulse la toma de decisiones estratégicas, teniendo en cuenta al Análisis de Clustering, como técnica de agrupamiento de empleados de características similares, y el análisis de Correspondencias Múltiples como facilitador de la interpretación de resultados de encuestas.

Palabras claves: Ciencia de datos/ Análisis de Clustering/ Toma de decisiones

Introducción

Las organizaciones cuentan en la actualidad con una gran cantidad de datos, muchos de los cuales se encuentran recopilados en forma automática y sistemática. La mayoría de estos datos permanecen sin ser explotados. Por ejemplo, se recopilan datos demográficos de los empleados, pero no se utilizan para caracterizar a grupo e identificar segmentos de empleados para dirigir acciones. O se registran datos de la producción de los equipos de trabajo y no se relacionan con otras variables para explicar las razones de las diferencias entre los grupos de trabajo.

Definición del problema

El aumento tanto del volumen de datos

disponibles como de la velocidad y la capacidad para procesarlos posibilitó a las organizaciones y a las empresas un nuevo marco de opciones para el desarrollo de sus actividades y de sus negocios. Todas las áreas funcionales de las organizaciones y de las empresas buscan obtener provecho de la gran cantidad de datos, obtenida con gran facilidad, para mejorar las operaciones y los negocios. Pero para que los datos tengan algún sentido, para que se traduzcan en decisiones, es necesario su captura, administración, procesamiento y análisis. En el caso de una empresa industrial, el área de Recursos Humanos recolecta datos, se elaboran informes, pero es necesario una revisión de las dimensiones y variables que se estudian actualmente.

¹Correo de contacto: mmedina@face.unt.edu.ar. Instituto de Administración, Facultad de Ciencias Económicas, Universidad Nacional de Tucumán. Argentina.

Objetivo general

Evaluar de diferentes técnicas de análisis estadístico multivariado (análisis de clusters y análisis de correspondencias) para proponer mejoras en la toma de decisiones en la gestión de recursos humanos. Las técnicas estadísticas se aplican a los datos del área de recursos humanos de una empresa industrial de la provincia de Tucumán en el período 2023.

Ciencia de datos

García (2018) considera que la ciencia de datos es hoy en día la herramienta fundamental para la explotación de datos y la generación de conocimiento. Entre los objetivos que persigue se encuentra la búsqueda de modelos que describan patrones y comportamientos a partir de los datos con el fin de tomar decisiones o hacer predicciones. Es un área que ha experimentado un enorme crecimiento al extenderse el acceso a grandes volúmenes de datos e incluso su tratamiento en tiempo real, requiriendo de técnicas sofisticadas que puedan tratar con los problemas prácticos como escalabilidad, robustez ante errores, adaptabilidad con modelos dinámicos, etc. Elorraiga (2021) define a la ciencia de datos es la conjunción de ciencia computacional, estadísticas, matemáticas y conocimientos de la empresa o del negocio

Marco metodológico

Este trabajo de investigación responde a un diseño cuantitativo no experimental de corte transversal. En una investiga-

ción no experimental se observan fenómenos tal como se dan en su contexto natural, para posteriormente analizarlos. Los diseños de investigación transversal recolectan datos en un solo momento, en un tiempo único.

Se accederá a los registros del área de recursos humanos disponibles en la empresa referidos a los datos sociodemográficos de los operarios. La población bajo estudio es la nómina de operarios de una planta industrial ubicada en el Gran San Miguel de Tucumán durante el año 2023. Como técnicas de recolección de datos se recurrirá a entrevistas, observación directa y acceso a los datos en los diferentes registros y sistemas que cuenta la empresa. Como técnicas de análisis se utilizará análisis de clustering y análisis de correspondencias. Como software se utiliza el paquete R para el procesamiento de los datos.

Resultados

Análisis de Clustering

Como primera medida, se analizan los datos sociodemográficos obtenidos. Para esta primera aproximación, se trabaja con las variables cuantitativas disponibles. De esta manera, se conforma una base de datos de 197 registros con 5 variables:

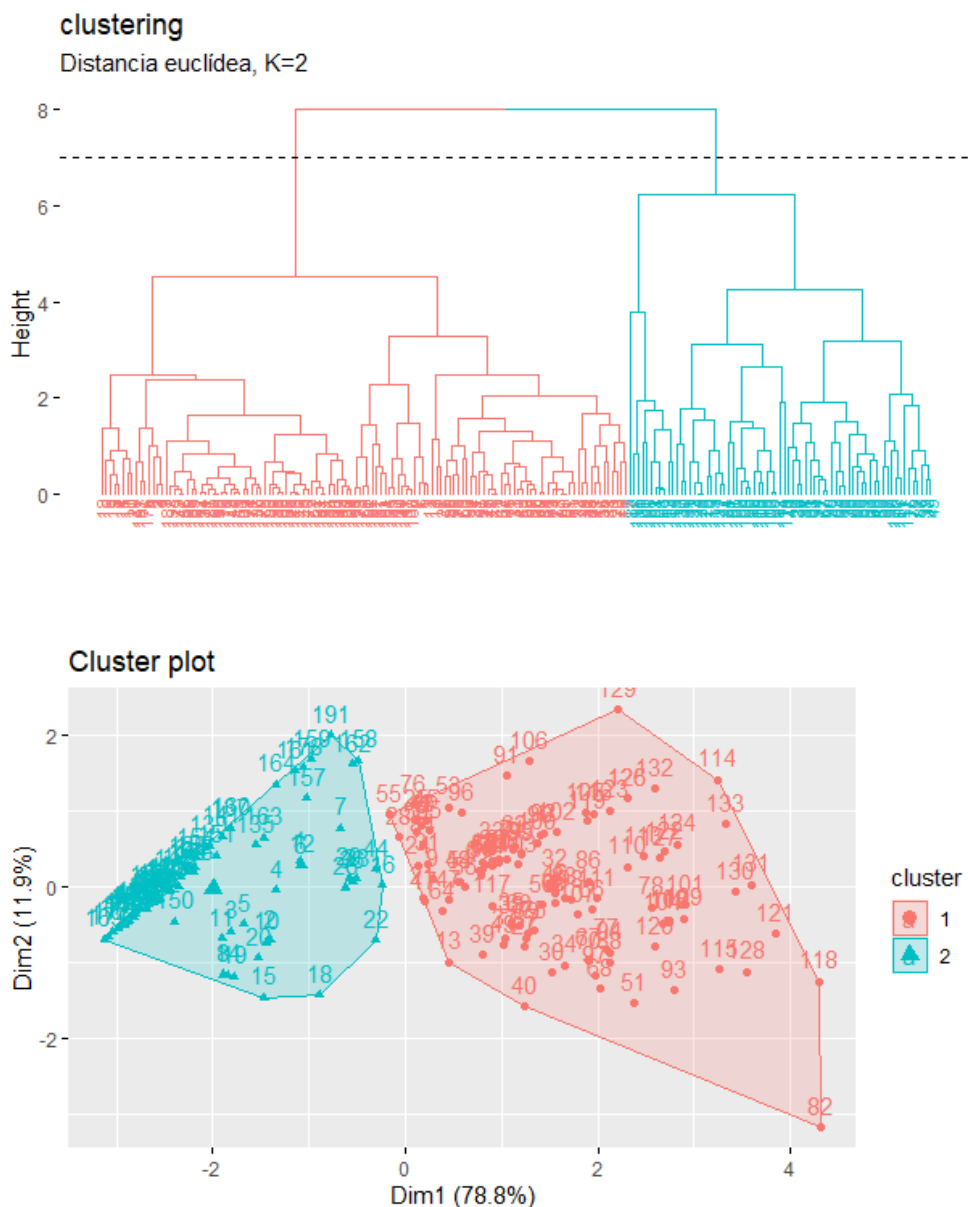
- Edad: años cumplidos a la fecha.
- Antigüedad: meses que la persona lleva trabajando en la empresa.
- Grupo familiar: cantidad de personas que conforman la familia del empleado.

- Total, de hijos: cantidad de hijas mujeres y de hijos varones del empleado.
- Promedio de edad de hijos: media de la edad de la totalidad de hijos del empleado.

Luego de confección de la base de datos con sus respectivos registros y variables, se procede a analizar la presencia de valores atípicos (outliers). Mediante un test de Mahalanobis, se determinan dos

outliers. Los registros corresponden a una persona de 23 años, con 21 meses de antigüedad y con dos familiares en su grupo; y a una persona de 56 años, con 15 meses de antigüedad y una persona en su grupo familiar.

Una vez realizada esta limpieza, la base final cuenta con 195 registros. El análisis de clúster muestra los siguientes resultados:



Como dato importante, entre las dos dimensiones se refleja un poco más del 90% de la variabilidad de los datos. Claramente se pueden diferenciar dos

clústers, uno con 132 observaciones y otro con 62. Las medias de cada grupo son las siguientes:

Clúster	Edad	Antigüedad	Grupo Familiar	Total Hijos	Promedio edad hijos
1	46,20455	215,87879	3,2045455	2,3030303	16,498258
2	28,90323	36,24194	0,4677419	0,2903226	1,532258

El grupo 1 tiene un promedio de 46 años, 215 meses de antigüedad, 3 familiares a cargo y 2 hijos con un promedio de 16 años. El grupo 2 tiene un promedio de 28 años, 36 meses de antigüedad, con uno o sin familiares a cargo y sin hijos o con hijos chicos (con un promedio de 2 años).

Análisis de Correspondencias Múltiples

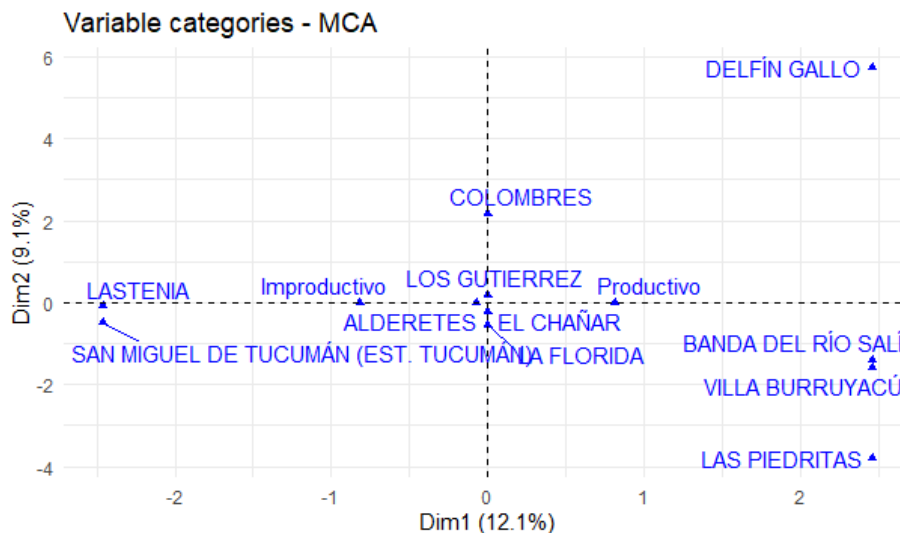
Para realizar este análisis exploratorio y descriptivo, se trabajó con algunas variables de la base de datos sociodemográficas. Como una primera gran limitación, se tomaron los datos de productividad obtenidos desde la nueva planilla para el registro de la misma, la cual contenía datos únicamente de 64 legajos (debido al poco tiempo de implementación). De esta manera, se consideraron diversas variables que a priori serían relevantes,

para explorar las relaciones entre las mismas. La base de datos consideró las siguientes variables:

- Legajo (únicamente utilizada como campo llave)
- Nivel de estudio
- Estado civil
- Partido
- Localidad
- Grupo familiar (codificada con sí o no)
- Hijos (codificada con sí o no)
- Productividad (este dato se tomó de la nueva planilla implementada, y se la codificó para que los legajos con productividad mayor a 0 sean productivos y los que son menos a 0, improductivos)

Con esta base de datos, se comenzaron a combinar las variables para efectuar diferentes análisis de correspondencias. Se presentan los gráficos de cada uno de ellos a continuación.

Gráfico 1: AC con localidad y productividad

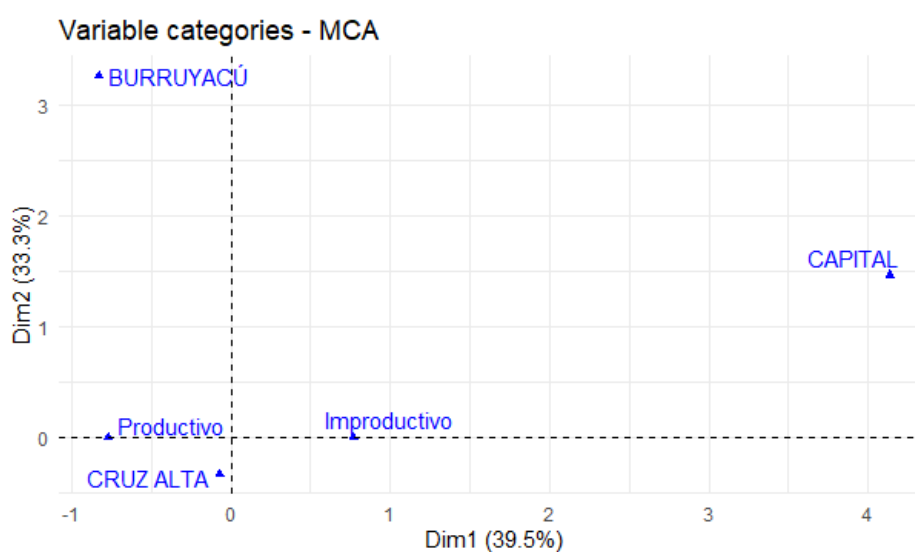


Fuente: elaboración propia mediante el software R

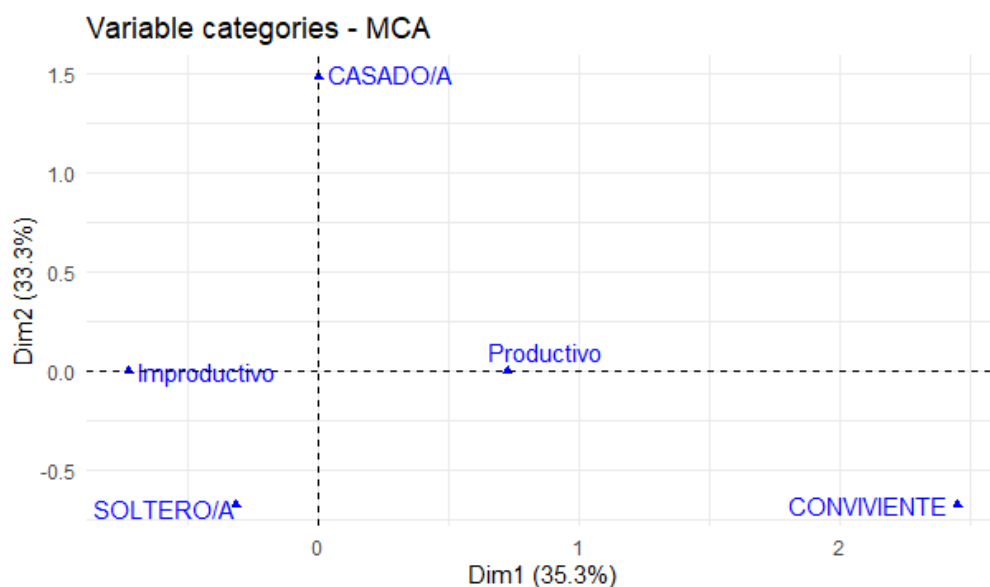
En el gráfico 1 no se aprecian correspondencias claras. Las dos dimensiones no explican una buena parte de la variabilidad de los datos.

En el gráfico 2 se puede pensar en una correspondencia entre la productividad y el partido de Cruz Alta.

Gráfico 2: AC con partido y productividad

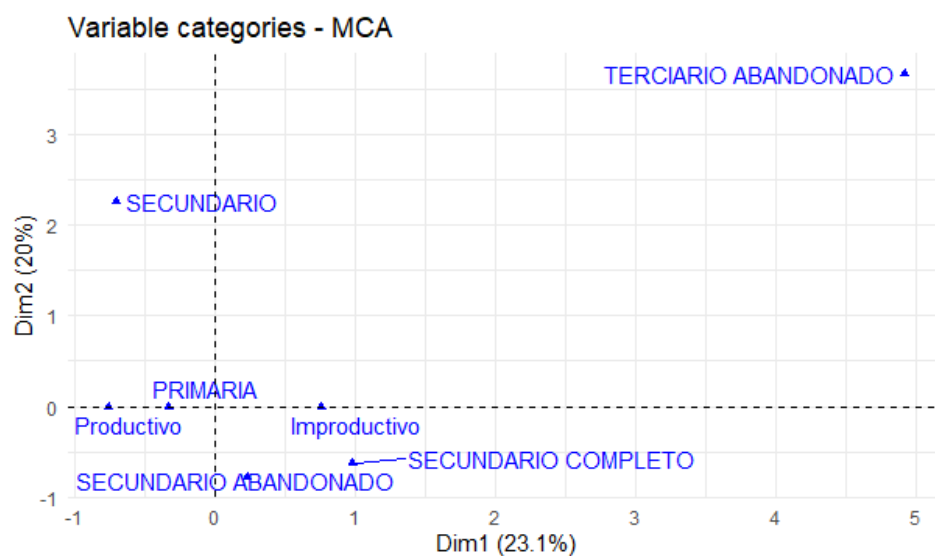


Fuente: elaboración propia mediante el software R

Gráfico 3: AC con estado civil y productividad

Fuente: elaboración propia mediante el software R

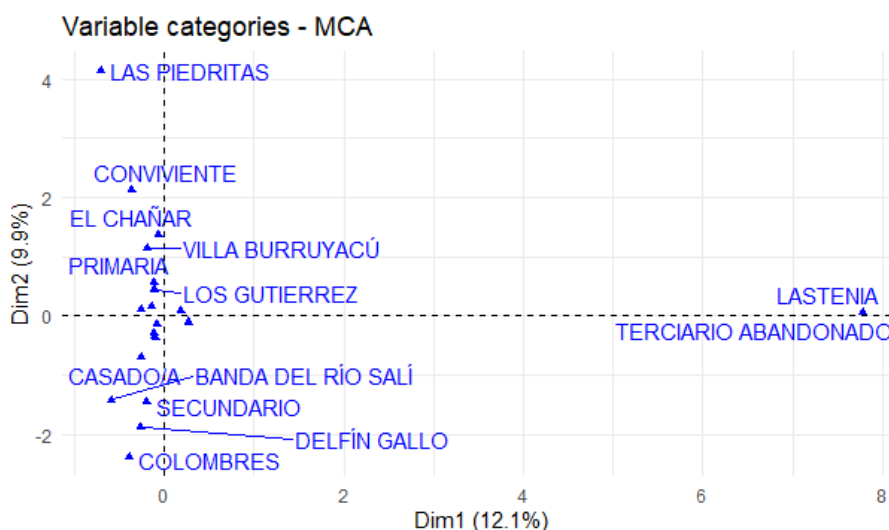
En el gráfico 3 no se aprecian correspondencias claras. Es posible pensar que los solteros son más improductivos.

Gráfico 4: AC con nivel de estudios y productividad

Fuente: elaboración propia mediante el software R

En el gráfico 4 no se aprecian correspondencias claras teniendo en cuenta la productividad. Los empleados productivos parecen ser personas con estudios primarios.

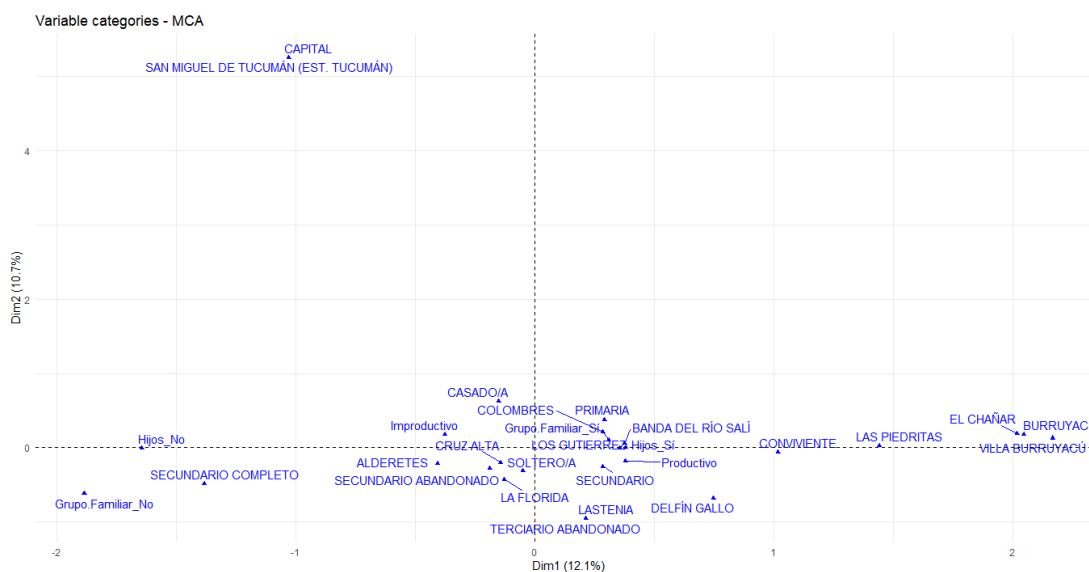
Gráfico 5: ACM con nivel de estudio, estado civil y localidad



Fuente: elaboración propia mediante el software R

En el gráfico 5 no se aprecian correspondencias claras. Las dos dimensiones no explican una buena parte de la variabilidad de los datos.

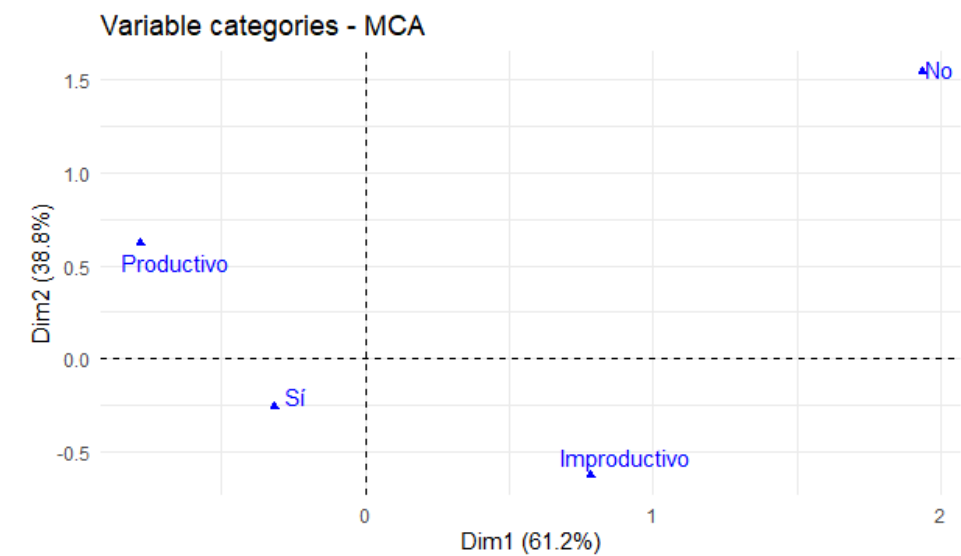
Gráfico 6: ACM con nivel de estudio, estado civil, grupo familiar, hijos, partido y productividad



Fuente: elaboración propia mediante el software R

En el gráfico 6 no se aprecian correspondencias claras. Las dos dimensiones no explican una buena parte de la variabilidad de los datos.

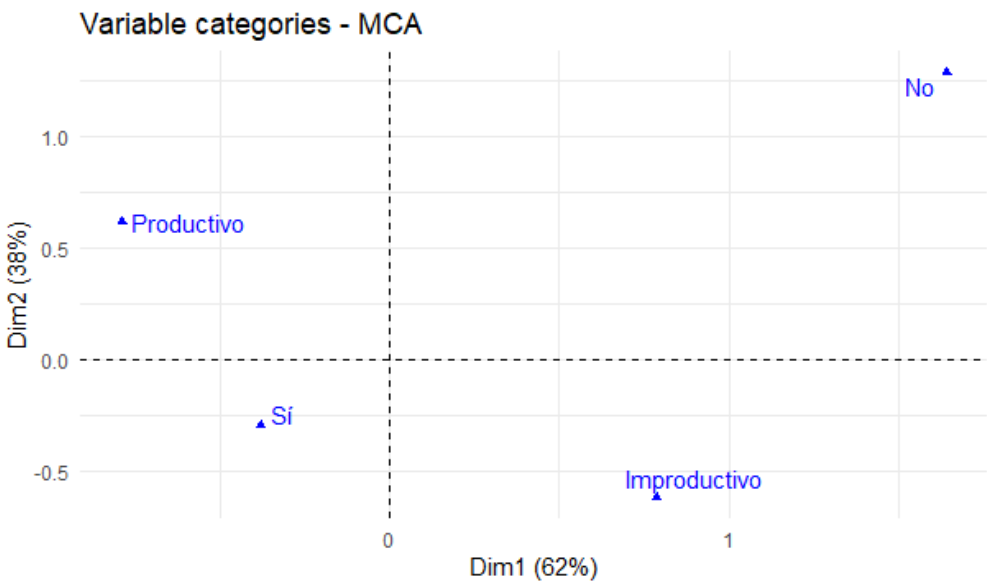
Gráfico 7: AC con grupo familiar y productividad



Fuente: elaboración propia mediante el software R

En el gráfico 7 no se aprecian correspondencias claras. Es posible pensar que los empleados productivos son aquellos que tienen grupo familiar.

Gráfico 8: AC con hijos y productividad



Fuente: elaboración propia mediante el software R

En el gráfico 8 no se aprecian correspondencias claras. Es posible pensar que los empleados productivos son aquellos que tienen hijos.

Conclusiones

La ciencia de datos apoya la toma de decisiones en la gestión de recursos humanos con herramientas como el análisis de clustering y el análisis de correspondencias múltiples.

El desafío de las empresas radica en transformar los datos en información y conocimiento.

Bibliografía de referencia:

Aballay, J. (2019). "Ciencia de Datos y la Supply Chain". Recuperado en agosto de 2022, de <https://ieec.edu.ar/ciencia-de-datos-y-la-supply-chain-1/>

Davenport, T. (2007). *Competing on Analytics: The New Science of Winning*, Harvard Business School Press, E.E.U.U.

Elorriaga, J. P.; Ochoa, O.; Di Pietro, N. (2021). "¿Qué es la ciencia de datos y cuáles son sus roles dentro de una empresa?". Recuperado en agosto de 2022, de: [https://www.gm2.dev/es/ciencia-de-datos-y-](https://www.gm2.dev/es/ciencia-de-datos-y-sus-roles-en-una-empresa/)

[sus-roles-en-una-empresa/](#)

García, J.; Molina, J.; Berlang, A.; P. Miguel; Bustamante, Á.; Padilla, W. (2018). *Ciencia de datos. Técnicas analíticas y aprendizaje estadístico*. Alfaomega. Chile.

Hernández Sampieri, R. (2018): *Metodología de la investigación. Las Rutas Cuantitativa, Cualitativa y Mixta*. McGraw-Hill Education. México.

Render, B.; Stair, R.; Hanna, M.; Hael, T. (2017). *Métodos Cuantitativos para los Negocios*. Pearson. México.

