



Reto | Segmentación de datos empleados

Duración: 2. 5 h

Objetivos

El objetivo de este reto es poner en práctica los contenidos revisados durante el módulo realizando los siguientes pasos:

- Segmentarás datos que influyen en la tasa de deserción de empleados.
- Seleccionarás el número adecuado de grupos de empleados.
- Realizarás distintas visualizaciones sobre los resultados obtenidos.
- Analizarás los resultados para apoyar la toma de decisiones.

Introducción:

Para el reto, la [base de datos](#) tiene datos de empleados como la educación, el estado civil, el salario y otros factores que influyen en la tasa de deserción.


Los empleados son la parte más importante de una organización. Los empleados exitosos cumplen con los plazos, realizan ventas y construyen la marca a través de interacciones positivas con los clientes. La deserción de empleados es un costo importante para una organización y predecir tales deserciones es el requisito más importante del departamento de Recursos Humanos en muchas organizaciones.

En este reto, tu tarea es agrupar a los empleados y brindar una interpretación de los grupos obtenidos, tomando como una de las variables más relevantes para la descripción de los resultados la ratio de deserción (Attrition_rate).

Instrucciones:

Para poder solucionar exitosamente este reto, necesitarás llevar a cabo los siguientes pasos:

1. Genera una libreta en Google Colab para resolver el reto con scripts de Python.
2. Descarga el archivo en la plataforma **kaggle.com**
3. Abre el archivo *Train.csv* para conocer su estructura y familiarizarse con la información contenida en él.

4. Crea un DataFrame en pandas para el archivo.
5. Transforma los datos según las características de las variables:
 - Puede usar diferentes métodos de transformación, como OrdinalEncoder, OneHotEncoder, StandardEncoder, NormalizerEncoder, etc.
 - La selección de las variables y de estos métodos depende del modelo de su solución y será evaluada su pertenencia según los datos.
6. Selecciona el número de grupos adecuados para agrupar usando K-means:
 - Ejecuta el algoritmo K-means con distintos números de grupos y almacena los resultados.
 - Utiliza una estrategia para determinar el número adecuado de grupos como K Elbow o Silhouette-Plot.
7. Segmenta el DataFrame original creando nuevos DataFrames con los empleados separados por grupo:
 - Crea una nueva tabla resumen con los estadísticos adecuados que describa el comportamiento de las variables por grupo.
8. Interpreta los resultados obtenidos:
 - Genera diferentes visualizaciones que ayuden a mostrar las características que tienen en común los empleados dentro de cada grupo.
 - Redacta una historia con las conclusiones más importantes del análisis de manera que ayude a la toma de decisiones.
9. Descarga tu libreta de Google Colab con **extensión .ipynb**
10. Nombra tu archivo siguiendo la nomenclatura que se te indica en **Formato de entrega de la actividad**.
1. Copia y publica únicamente la liga a tu libreta de Google Colab presionando el ícono  ubicado en la parte inferior de la pantalla de Plataforma.

Evaluación:

Tu reporte será evaluado con base al cumplimiento de los requerimientos y a su contenido. Aunque es importante la redacción de las conclusiones, el mayor peso lo tendrán los distintos fragmentos de código que incluyas en tu libreta Google Colab.

Formato de entrega de la actividad:

Al finalizar los puntos anteriores realiza lo siguiente:

- Descarga tu libreta de Google Colab (extensión .ipynb) y guárdalo en un archivo con la nomenclatura DS_C6_SC2_Nombre completo (sin espacios).

Instrucciones Reto	Rúbrica de evaluación	Base de datos
Descargar PDF	Descargar PDF	Kaggle.com