

Lecciones aprendidas

Data Science with Python
Prepare and Explore the
Data

Edwin Spencer

Acciones ejecutadas

- Entendimiento del dataset a través de los comandos `Head()`, `describe()`, `info()`, `columns()` y `dtypes`
- Validación de los datos nulos a través del comando `isnull().any()` detectando que no existían valores nulos.
- Además se realizó la verificación de si existen valores duplicados con el comando `duplicated().any()` sin encontrar ningún valor duplicado.
- Por otra parte procedí a modificar el tipo de de las siguientes variables a `category`:
 - SEX
 - EDUCATION
 - MARRIAGE
 - Las demás variables quedaron de igual forma.
- Además, se realizó análisis básico de los datos a través de la librería `matplotlib` dando como resultado los siguientes datos:
 - Más del 50% de los clientes tienen créditos por menos de 200 mil dólares.
 - Del total de los clientes, más del 50% de los clientes son Mujeres.
 - Más del 50% son solteros.
 - Un poco más del 40% son graduados de universidad.
 - Una mayor cantidad de clientes se encuentran entre los 20 y 30 años.

Incidencias presentadas y experiencias generadas

- Se me dificultó poder realizar el renombre de una de las columnas (`default payment next month`), por más que ejecuté diferentes comandos no lo logré reflejar en el dataset final.
- A la hora de realizar la modificación de las variables a `category` se me complicó un poco, luego entendí que debo de realizar la asignación de la variable nuevamente porque si no lo que hace es eliminar la columna del dataset.
- En términos generales, la experiencia con Python a sido enriquecedora y todo lo aprendido en R definitivamente a apoyado a la rápida adopción de este nuevo lenguaje,