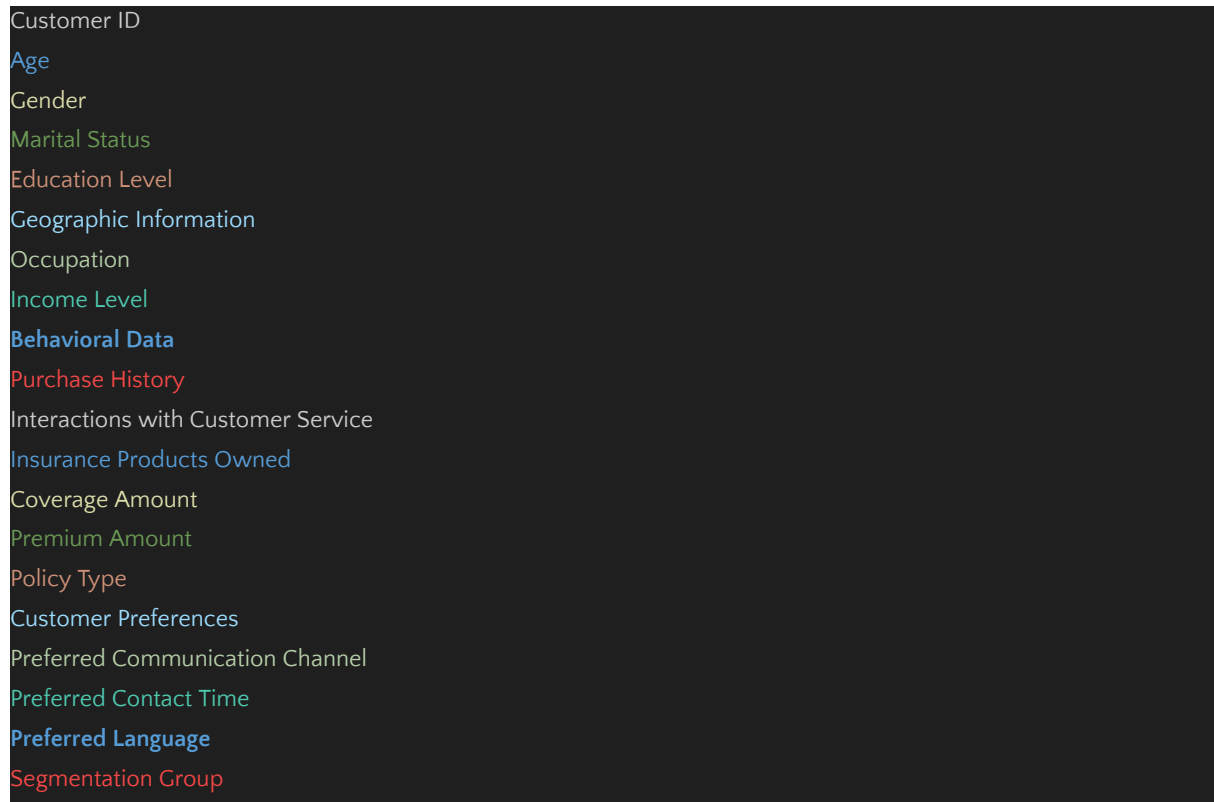


Challenge – Data Scientist

RSK Consulting

El siguiente archivo en formato .csv contiene información referente a un grupo de 53 mil clientes de una empresa de seguros:



Customer ID
Age
Gender
Marital Status
Education Level
Geographic Information
Occupation
Income Level
Behavioral Data
Purchase History
Interactions with Customer Service
Insurance Products Owned
Coverage Amount
Premium Amount
Policy Type
Customer Preferences
Preferred Communication Channel
Preferred Contact Time
Preferred Language
Segmentation Group

El reto consiste en analizar el dataset usando Python, se apreciará sobre todo el uso de tres 3 librerías(pandas, numpy y matplotlib) para la visualización de datos también es posible usar bokeh/plotly/seaborn.

El resultado del análisis debe ser el perfilado del cliente por segmento (Segmentation Group) utilizando técnicas estadísticas descriptivas, calculos de medidas de tendencia central, estadísticos de orden entre otras.

Hints:

- Un perfilado completo debe incluir el estudio de las distintas variables categoricas por ejemplo estado civil(Marital Status) en relación con las variables cuantitativas. Si se desea es posible usar técnicas de aprendizaje no supervisado para explorar la información.

- Se apreciará la capacidad de explotar la información disponible mediante cuadros y gráficos estadísticos de resumen que puedan ser explicados de manera sintética, clara y directa (storytelling)

El archivo en mención puede ser encontrado en el siguiente enlace:

<https://drive.google.com/file/d/1lcDwl4mcm94l2TFkuhblpInGfhMNG-Hc/view?usp=sharing>

Finalmente, el caso será presentado en una reunión de máximo 15 minutos con el Líder de la práctica de Ciencia de Datos de la consultora, de ser satisfactorio el resultado se presentará su candidatura al cliente final. Durante la reunión podrán hacerse preguntas relacionadas a SQL y Power BI.