

**Análisis de Big Data:
EAA361A**

Laboratorio 6

Profesor: Cristián Vásquez

Ayudante: Pablo González

Tema 1: Describiendo los datos

El set de datos ‘Credit.csv’ tiene información de una muestra de cliente de una entidad financiera. Utilizando la información contenida en dicho set, responda las siguientes preguntas:

1. Identifique cuáles variables son cuantitativas y cuáles son cualitativas.
2. Genere una función que le permita describir las variables en función del tipo al que ésta corresponde. Su función debe ser capaz de reportar medidas de localización, dispersión y forma para todas las variables cuantitativas.
3. Utilizando la función antes generada, presente una estadística descriptiva de las variables en estudio (excluyendo el identificador).

Tema 2: Matrices de diseño

El set de datos ‘presupuesto.csv’ tiene estadísticas sobre las ventas de un producto específico en 200 tiendas de retail, además se dispone del presupuesto publicitario de cada una de las tiendas para los diferentes canales: televisión, radio y periódicos. Utilizando dicha información, responda las siguientes preguntas:

1. A través de la función `lm()`, realice un modelo de regresión lineal considerando como variable respuesta Balance y como variables independientes Income, Age, Student y Rating.
2. Repita el proceso, pero esta vez realice la estimación a través de operaciones matriciales.

Tema 3: Métricas de data mining

1. Construya la variable 'IndSales' a partir de la variable 'sales', con rangos < 10.38 , $10.38-12.90$ y $17.40-\text{Más}$. Describa esta variable desde la perspectiva de una variable de agrupación.
2. Reporte la matriz de correlación de todas de las variables. Nótese que no nos interesa la correlación con la variable 'Obs'.
3. Reporte la dependencia de cada variable con el resto. Comente los resultados.
4. Reporte matriz dependencia directa de a pares de variables. Comente los resultados
5. Finalmente, reporte el coeficiente de dependencia efectivo