**UNIVERSIDAD CATÓLICA BOLIVIANA”SAN PABLO”  
MAESTRÍA EN CIENCIA DE DATOS Versión 2  
MATERIA MINERIA DE DATOS 1**

**DATASET DE PRÁCTICA**

**DATASET CREDITO**

* ~~Quitar filas duplicadas~~
* ~~Quitar columnas duplicadas~~
* ~~Identificar columnas que tienen valores constantes~~
* ~~Identifique el porcentaje de valores faltantes en las columnas~~
* Para el dataset hacer un análisis gráfico documentado de las variables:
* Histograms
* Quartiles
* Scatter plots
* Distributions
* Boxplots
* Bar plot
* Heatmaps
* Aplicar expresión matemática para el cálculo (CreditAmount / Duración)\*residence\_since
* Discretizar duración, credit\_amount, age, number\_of\_dependents.
* Obtenga la matriz de correlación de las variables numéricas.
* Obtenga la tabla de relación de frecuencias entre los atributos job y purpose.
* Mapear a valores diferentes para atributos own\_telephone, foreign\_worker.
* Normalizar el campo instalment\_commitment
* Estandarizar el campo instalment\_commitment
* De los gráficos Boxplot obtenidos quitar outliers por medio de Inter-quartil range
* Utilice One-hot encoding en campo purpose.
* Obtenga las instancias que cumplan la condición de un monto de crédito mayor a 3000
* Obtenga las instancias que cumplen con la condición de que job sea igual a “skilled” y que el monto de crédito sea mayor a 2000.
* Obtenga suma total del camp credit\_amount agrupado por el campo purpose “new car”.
* Obtenga age y class para los registros que tienen campo purpose “education” y housing “for\_free”.
* Obtenga el total y el promedio de credit amount agrupado por housing y class.