

## STATISTICAL ESSENTIALS

### Laboratorio de Aplicación N°01:

- Se entrega el siguiente set de datos, loan\_prediction-II.csv, el cual es un dataset de una entidad bancaria dedicada al otorgamiento de préstamos personales. Se entrega el siguiente set de datos de una empresa del sector financiero que se ocupa de todo tipo de préstamos hipotecarios. Tienen presencia en todas las áreas urbanas, semiurbanas y rurales. El cliente primero solicita un préstamo hipotecario y luego la compañía valida la elegibilidad del cliente para el préstamo.

**La empresa desea automatizar el proceso de elegibilidad de préstamos (en tiempo real)** en función de los detalles del cliente proporcionados al completar el formulario de solicitud en línea. El dataset se encuentra en loan\_prediction-II.csv, el cual es un dataset que cuenta con las principales características de los clientes en cuanto a características sociodemográficas y de mercado.

La empresa necesita saber a quién sería conveniente aceptarle el préstamo hipotecario, para así poder rentabilizar su capital y no tener problemas de morosidad.

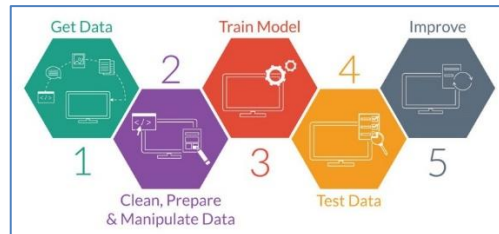
- Las variables que se disponibilidad son:

Variable	Descripción de la variable
Loan_ID	Identificador del cliente
Gender	Género de la persona.
Married	Estado Civil de la persona.
Dependents	Número de hijos (Dependientes).
Education	Nivel educativo del cliente.
Self_Employed	Tiene empleo actual?.
ApplicantIncome	Ingreso actual del cliente de acuerdo a su trabajo dependiente.
CoapplicantIncome	Ingreso actual del cliente de acuerdo a trabajos independientes.
LoanAmount	Monto del préstamo.
Loan_Amount_Term	Término del préstamo.
Credit_History	Tiene historial crediticio?
Property_Area	Área de residencia.
Nationality	Nacionalidad o foráneo?.
Loan_Status	Estado del préstamo.

## **Entregables del Laboratorio:**

1.- Definir el problema de la naturaleza que se tiene a continuación, además de los objetivos de negocio bien definidos.

2.- ¿Qué tipo de variables se utilizan en el problema de negocio?



3.- Realizar un informe sobre las principales medidas de tendencia central y de dispersión. Cuál de los estados de préstamo presenta mayor homogeneidad respecto a su ingreso como dependiente. (3 pts.)

4.- Evaluar la completitud de los datos, sacando indicadores para los datos antes de la imputación y después. Comente sus resultados. (Imputación datos). (3 pts.)

5.- Evaluar los outliers o casos atípicos univariados para las variables cuantitativas. Muestre los principales gráficos para identificarlos y muéstrelos. (3 pts.)

6.- Evaluar la colinealidad o multicolinealidad en las variables, identificando aquellas en las cuáles su coeficiente de asociación supere el punto de corte de  $\pm 0.3$ . (3 pts.)

7.- Desarrolle las mejores transformaciones de variables para las variables cuantitativas, fundamente el porqué de las transformaciones y muestre los hallazgos más relevantes. (4 pts.)

8.- Realice un análisis de asociación de la variable dependiente respecto a alguna variable cualitativa y otra cuantitativa. Comente sus resultados. (4 pts.)

