



**PROYECTO N° 1:**  
**PLANTEAMIENTO DE**  
**PROYECTO Y DE**  
**DATASET-**  
**DATAFRAME**

**Alumno: Abel Montiel Aguilar.**

# 1 º DATA-SET/FRAME

## 1. Bank:

a. **Asunto:** Util para predecir si alguien merece un préstamo.

b. **Contenido:**

- i. Contiene 41.188 filas
- ii. Contiene 21 columnas

c. **Columnas:**

i.	0	age	41188 non-null int64
ii.	1	job	41188 non-null object
iii.	2	marital	41188 non-null object
iv.	3	education	41188 non-null object
v.	4	default	41188 non-null object
vi.	5	housing	41188 non-null object
vii.	6	loan	41188 non-null object
viii.	7	contact	41188 non-null object
ix.	8	month	41188 non-null object
x.	9	day_of_week	41188 non-null object
xi.	10	duration	41188 non-null int64
xii.	11	campaign	41188 non-null int64
xiii.	12	pdays	41188 non-null int64
xiv.	13	previous	41188 non-null int64
xv.	14	poutcome	41188 non-null object
xvi.	15	emp.var.rate	41188 non-null float64
xvii.	16	cons.price.idx	41188 non-null float64
xviii.	17	cons.conf.idx	41188 non-null float64
xix.	18	euribor3m	41188 non-null float64
xx.	9	nr.employed	41188 non-null float64
xxi.	20	y	41188 non-null object

d. **Potencial:** Este conjunto de datos es ideal para construir y entrenar modelos predictivos, como clasificadores, que pueden prever la probabilidad de que un cliente acepte un préstamo. Podrías explorar algoritmos de aprendizaje automático, como regresión logística o árboles de decisión, para este propósito.

Es fundamental realizar un análisis exploratorio de datos para comprender las relaciones entre las variables y decidir qué características son más relevantes para la predicción.

e. **PowerBI: Algunas sugerencias:**

i. **Análisis Descriptivo:**

- 1. Utiliza gráficos de barras para visualizar la distribución de categorías en variables categóricas como "job", "marital", "education", etc.
- 2. Crea histogramas para entender la distribución de variables numéricas como "age", "duration", "campaign", etc.

ii. **Análisis de Correlación:**

- 1. Emplea gráficos de dispersión para explorar relaciones entre variables numéricas.
- 2. Calcula y visualiza la matriz de correlación para entender la fuerza y dirección de las relaciones.

iii. **Análisis Temporal:**

1. Utiliza gráficos de líneas para representar la evolución temporal de variables como "month" y "day\_of\_week".
2. Examina la relación entre el tiempo de contacto ("duration") y la respuesta del cliente.

**iv. Segmentación de Datos:**

1. Aplica filtros dinámicos para segmentar datos por características específicas (por ejemplo, trabajo, estado civil) y observar cómo estas segmentaciones afectan la aceptación del préstamo.

**v. Machine Learning:**

1. Incorpora visualizaciones que resuman el rendimiento del modelo, como la precisión, la sensibilidad y la especificidad.
2. Utiliza gráficos de importancia de características para destacar qué variables son más influyentes en las predicciones.

**vi. Panel de Control Interactivo:**

1. Crea un panel interactivo que permita a los usuarios ajustar parámetros del modelo, como umbrales de decisión, y ver cómo afecta esto a las predicciones.

**f. METODO : SMART:**

**i. Específico (Specific):**

1. Objetivo: Predecir la probabilidad de aceptación de préstamos por parte de clientes.
2. Enfoque: Utilizar variables demográficas, de empleo y de historial de contactos para desarrollar un modelo predictivo.

**ii. Medible (Measurable):**

1. Métrica de éxito: La precisión del modelo en predecir si un cliente aceptará o no un préstamo.
2. Indicadores clave de rendimiento (KPI):
  - a. Precisión del modelo.
  - b. Sensibilidad y especificidad.
  - c. Análisis de la curva ROC.

**iii. Alcanzable (Achievable):**

1. Datos Disponibles: Se cuenta con un conjunto de datos con 41,188 registros y 21 variables relevantes.
2. Capacidad técnica: Se utilizarán algoritmos de aprendizaje automático estándar (regresión logística, árboles de decisión) implementados en herramientas como Power BI.

**iv. Relevante (Relevant):**

1. Importancia del Problema: La capacidad de predecir la aceptación de préstamos permite a las instituciones financieras optimizar sus estrategias de marketing y gestión de clientes.
2. Impacto Social: Ayuda a garantizar que los préstamos se otorguen de manera más precisa, evitando decisiones financieras desfavorables para los clientes.

**v. Temporal (Time-bound):**

1. Plazo: Desarrollar, entrenar y evaluar el modelo dentro de los próximos meses.
2. Iteraciones: Realizar al menos dos iteraciones del modelo, ajustando parámetros según sea necesario.

## 2 º DATA-SET/FRAME

### 2. Madrid:

- a. **Asunto:** Util para valorar precio de vivienda
- b. **Contenido:**
  - i. Contiene 38.754 filas
  - ii. Contiene 22 Columnas
- c. **Columnas:**

i. 0	Título	38751 non-null object
ii. 1	Tipo de inmueble	38751 non-null object
iii. 2	Dirección	38742 non-null object
iv. 3	Precio	38754 non-null int32
v. 4	Habitaciones	38754 non-null int32
vi. 5	Subtítulo	38742 non-null object
vii. 6	Barrio	38742 non-null object
viii. 7	Municipio	38742 non-null object
ix. 8	Distrito	38742 non-null object
x. 9	Euros/m2	38754 non-null int32
xi. 10	Metros cuadrados contruidos	38754 non-null int32
xii. 11	Baños	38754 non-null int32
xiii. 12	Planta	38754 non-null int32
xiv. 13	Ascensor (Sí/No)	38742 non-null object
xv. 14	Obra nueva (Sí/No)	38742 non-null object
xvi. 15	Piscina (Sí/No)	38742 non-null object
xvii. 16	Terraza (Sí/No)	38742 non-null object
xviii. 17	Parking (Sí/No)	38742 non-null object
xix. 18	Parking incluido en el precio (Sí/No)	38742 non-null object
xx. 19	Aire acondicionado (Sí/No)	38742 non-null object
xxi. 20	Trastero (Sí/No)	38742 non-null object
xxii. 21	Jardín (Sí/No)	38742 non-null object
- d. **Potencial:** Este conjunto de datos es valioso para evaluar el mercado inmobiliario en Madrid. Puedes realizar análisis descriptivos para entender la distribución de precios y características de las viviendas. Además, podrías explorar modelos predictivos para predecir el precio de las viviendas basándote en las variables disponibles.
- e. **PowerBI:** Aquí hay algunas sugerencias:
  - i. **Análisis Descriptivo:**
    - 1. Utiliza gráficos para visualizar la distribución de precios.
    - 2. Crea gráficos de barras para mostrar la frecuencia de diferentes características, como el tipo de inmueble, el número de habitaciones, etc.
  - ii. **Análisis de Correlación:**
    - 1. Explora la relación entre el precio y otras variables como metros cuadrados, número de baños, etc.
  - iii. **Análisis Espacial:**
    - 1. Utiliza mapas para visualizar la distribución geográfica de las viviendas y sus precios.
  - iv. **Filtros Dinámicos:**

1. Aplica filtros para segmentar datos por características específicas y observar cómo afectan al precio.

**v. Machine Learning:**

1. Experimenta con modelos de regresión para predecir el precio de las viviendas.

**vi. Panel de Control Interactivo:**

1. Crea un panel interactivo que permita a los usuarios explorar diferentes aspectos del mercado inmobiliario en Madrid.

**f. Método: SMART:**

**i. Específico (Specific):**

1. Objetivo: Valorar el precio de las viviendas en Madrid.
2. Enfoque: Utilizar variables como tipo de inmueble, número de habitaciones, ubicación, etc., para desarrollar un modelo de valoración.

**ii. Medible (Measurable):**

1. Métrica de éxito: La precisión del modelo en predecir el precio de las viviendas.
2. Indicadores clave de rendimiento (KPI):
  - a. - Precisión del modelo.
  - b. - Análisis de error.

**iii. Alcanzable (Achievable):**

1. Datos Disponibles: Se cuenta con un conjunto de datos con 38,754 registros y 22 variables relevantes.
2. Capacidad técnica: Se utilizarán técnicas de análisis exploratorio y modelos predictivos implementados en herramientas como Power BI.

**iv. Relevante (Relevant):**

1. Importancia del Problema: Evaluar el precio de las viviendas es crucial para compradores, vendedores y agentes inmobiliarios.
2. Impacto Social: Facilita decisiones informadas sobre inversiones inmobiliarias.

**v. Temporal (Time-bound):**

1. Plazo: Desarrollar y evaluar el modelo en los próximos meses.
2. Iteraciones: Realizar al menos dos iteraciones del modelo, ajustando según sea necesario.

## 3 º DATA-SET/FRAME

### 3. Crimes

- a. **Asunto:** Util para conocer datos sobre la criminalidad en Chicago desde 2001 hasta ahora

- b. **Contenido:**
  - i. Contiene 7.846.809filas
  - ii. Contiene 22 Columnas

- c. **Columnas**
  - i. 0 ID int64
  - ii. 1 Case Number object
  - iii. 2 Date object
  - iv. 3 Block object
  - v. 4 IUCR object
  - vi. 5 Primary Type object
  - vii. 6 Description object
  - viii. 7 Location Description object
  - ix. 8 Arrest bool
  - x. 9 Domestic bool
  - xi. 10 Beat int64
  - xii. 11 District float64
  - xiii. 12 Ward float64
  - xiv. 13 Community Area float64
  - xv. 14 FBI Code object
  - xvi. 15 X Coordinate float64
  - xvii. 16 Y Coordinate float64
  - xviii. 17 Year int64
  - xix. 18 Updated On object
  - xx. 19 Latitude float6
  - xxi. 20 Longitude float64
  - xxii. 21 Location object

- d. **Potencial:** Este conjunto de datos proporciona una visión detallada de la actividad criminal en Chicago, lo que permite análisis temporales, geospaciales y de tipos de delitos.

- e. **PowerBI: Aquí hay algunas sugerencias:**

- i. **Análisis Temporal:**

- 1. Utiliza gráficos de líneas para representar la evolución temporal de la criminalidad en Chicago.
    - 2. Examina la variación de tipos de delitos a lo largo de los años.

- ii. **Análisis Geoespacial:**

- 1. Utiliza mapas para visualizar la distribución geográfica de los crímenes en diferentes áreas de Chicago.

- iii. **Análisis de Tipos de Delitos:**

- 1. Crea gráficos de barras para mostrar la frecuencia de diferentes tipos de delitos.

- iv. **Filtros Dinámicos:**

- 1. Permite a los usuarios filtrar datos por año, tipo de delito, ubicación, etc.

- v. Machine Learning:**
    - 1. Explora modelos predictivos para predecir la frecuencia de crímenes en el futuro.
  - vi. Panel de Control Interactivo:**
    - 1. Crea un panel interactivo que permita a los usuarios explorar diferentes aspectos de la actividad criminal en Chicago.
- f. Método: SMART:**
- i. Específico (Specific):**
    - 1. Objetivo: Analizar la criminalidad en Chicago desde 2001 hasta ahora.
    - 2. Enfoque: Explorar patrones temporales y geoespaciales, así como la distribución de tipos de delitos.
  - ii. Medible (Measurable):**
    - 1. Métrica de éxito: Comprender y visualizar patrones claros en los datos de crímenes.
    - 2. Indicadores clave de rendimiento (KPI):
      - a. - Frecuencia de delitos por año.
      - b. - Distribución geográfica de delitos.
  - iii. Alcanzable (Achievable):**
    - 1. Datos Disponibles: Se cuenta con un conjunto de datos extenso con información detallada sobre crímenes.
    - 2. Capacidad técnica: Se utilizarán técnicas de análisis exploratorio y visualización de datos.
  - iv. Relevante (Relevant):**
    - 1. Importancia del Problema: Analizar la criminalidad es crucial para la seguridad pública y la toma de decisiones informada.
    - 2. Impacto Social: Permite la implementación de estrategias efectivas para abordar y prevenir la delincuencia.
  - v. Temporal (Time-bound):**
    - 1. Plazo: Realizar análisis y presentar resultados en los próximos meses.
    - 2. Iteraciones: Realizar iteraciones según sea necesario para profundizar en el análisis.