

PROYECTO Nº 1: **PLANTEAMIENTO DE PROYECTO Y DE DATASET-DATAFRAME**

**Alumno: Abel Montiel Aguilar.**

**1 º DATA-SET/FRAME**

1. **Bank:** 
   1. **Asunto**: Util para predecir si alguien merece un préstamo.
   2. **Contenido**:
      1. Contiene 41.188 filas
      2. Contiene 21 columnas
   3. **Columas**:
      1. 0 age 41188 non-null int64
      2. 1 job 41188 non-null object
      3. 2 marital 41188 non-null object
      4. 3 education 41188 non-null object
      5. 4 default 41188 non-null object
      6. 5 housing 41188 non-null object
      7. 6 loan 41188 non-null object
      8. 7 contact 41188 non-null object
      9. 8 month 41188 non-null object
      10. 9 day\_of\_week 41188 non-null object
      11. 10 duration 41188 non-null int64
      12. 11 campaign 41188 non-null int64
      13. 12 pdays 41188 non-null int64
      14. 13 previous 41188 non-null int64
      15. 14 poutcome 41188 non-null object
      16. 15 emp.var.rate 41188 non-null float64
      17. 16 cons.price.idx 41188 non-null float64
      18. 17 cons.conf.idx 41188 non-null float64
      19. 18 euribor3m 41188 non-null float64
      20. 9 nr.employed 41188 non-null float64
      21. 20 y 41188 non-null object
   4. **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.

* 1. **PowerBI: Algunas sugerencias:**
     1. **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.
     2. **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.
     3. **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.
     4. **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.
     5. **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.
     6. **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.
  2. **METODO : SMART:** 
     1. **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.
     2. **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):
           1. Precisión del modelo.
           2. Sensibilidad y especificidad.
           3. Análisis de la curva ROC.
     3. **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.
     4. **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.
     5. **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**

1. **Madrid:**
   1. **Asunto**: Util para valorar precio de vivienda
   2. **Contenido**:
      1. Contiene 38.754 filas
      2. Contiene 22 Columnas
   3. **Columas**:
      1. 0 Título 38751 non-null object
      2. 1 Tipo de inmueble 38751 non-null object
      3. 2 Dirección 38742 non-null object
      4. 3 Precio 38754 non-null int32
      5. 4 Habitaciones 38754 non-null int32
      6. 5 Subtítulo 38742 non-null object
      7. 6 Barrio 38742 non-null object
      8. 7 Municipio 38742 non-null object
      9. 8 Distrito 38742 non-null object
      10. 9 Euros/m2 38754 non-null int32
      11. 10 Metros cuadrados construidos 38754 non-null int32
      12. 11 Baños 38754 non-null int32
      13. 12 Planta 38754 non-null int32
      14. 13 Ascensor (Sí/No) 38742 non-null object
      15. 14 Obra nueva (Sí/No) 38742 non-null object
      16. 15 Piscina (Sí/No) 38742 non-null object
      17. 16 Terraza (Sí/No) 38742 non-null object
      18. 17 Parking (Sí/No) 38742 non-null object
      19. 18 Parking incluído en el precio (Sí/No) 38742 non-null object
      20. 19 Aire acondicionado (Sí/No) 38742 non-null object
      21. 20 Trastero (Sí/No) 38742 non-null object
      22. 21 Jardín (Sí/No) 38742 non-null object
   4. **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.
   5. **PowerBI**: Aquí hay algunas sugerencias:
      1. 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.
      2. **Análisis de Correlación:**
         1. Explora la relación entre el precio y otras variables como metros cuadrados, número de baños, etc.
      3. **Análisis Espacial:**
         1. Utiliza mapas para visualizar la distribución geográfica de las viviendas y sus precios.
      4. **Filtros Dinámicos:**
         1. Aplica filtros para segmentar datos por características específicas y observar cómo afectan al precio.
      5. **Machine Learning:**
         1. Experimenta con modelos de regresión para predecir el precio de las viviendas.
      6. **Panel de Control Interactivo:**
         1. Crea un panel interactivo que permita a los usuarios explorar diferentes aspectos del mercado inmobiliario en Madrid.
   6. **Método: SMART:**
      1. **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.
      2. **Medible (Measurable):**
         1. Métrica de éxito: La precisión del modelo en predecir el precio de las viviendas.
         2. Indicadores clave de rendimiento (KPI):
            1. - Precisión del modelo.
            2. - Análisis de error.
      3. **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.
      4. **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.
      5. **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**

1. **Crimes**
   1. **Asunto**: Util para conocer datos sobre la criminalidad en Chicago desde 2001 hasta ahora
   2. **Contenido**:
      1. Contiene 7.846.809filas
      2. Contiene 22 Columnas
   3. **Columas**
      1. 0 ID int64
      2. 1 Case Number object
      3. 2 Date object
      4. 3 Block object
      5. 4 IUCR object
      6. 5 Primary Type object
      7. 6 Description object
      8. 7 Location Description object
      9. 8 Arrest bool
      10. 9 Domestic bool
      11. 10 Beat int64
      12. 11 District float64
      13. 12 Ward float64
      14. 13 Community Area float64
      15. 14 FBI Code object
      16. 15 X Coordinate float64
      17. 16 Y Coordinate float64
      18. 17 Year int64
      19. 18 Updated On object
      20. 19 Latitude float6
      21. 20 Longitude float64
      22. 21 Location object
   4. **Potencial**: Este conjunto de datos proporciona una visión detallada de la actividad criminal en Chicago, lo que permite análisis temporales, geoespaciales y de tipos de delitos.
   5. **PowerBI: Aquí hay algunas sugerencias:**
      1. **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.
      2. **Análisis Geoespacial:**
         1. Utiliza mapas para visualizar la distribución geográfica de los crímenes en diferentes áreas de Chicago.
      3. **Análisis de Tipos de Delitos:**
         1. Crea gráficos de barras para mostrar la frecuencia de diferentes tipos de delitos.
      4. **Filtros Dinámicos:**
         1. Permite a los usuarios filtrar datos por año, tipo de delito, ubicación, etc.
      5. **Machine Learning:**
         1. Explora modelos predictivos para predecir la frecuencia de crímenes en el futuro.
      6. **Panel de Control Interactivo:**
         1. Crea un panel interactivo que permita a los usuarios explorar diferentes aspectos de la actividad criminal en Chicago.
   6. **Método: SMART:**
      1. **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.
      2. **Medible (Measurable):**
         1. Métrica de éxito: Comprender y visualizar patrones claros en los datos de crímenes.
         2. Indicadores clave de rendimiento (KPI):
            1. - Frecuencia de delitos por año.
            2. - Distribución geográfica de delitos.
      3. **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.
      4. **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.
      5. **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.