1. CONJUNTO DE DATOS

   Para la resolución de la practica se proporciona una tabla que contiene características de personas que aplicaron para solicitar un crédito, adicional a esa información contamos con una variable "tgt" que representa a nuestra variable objetivo que toma el valor de 1 si se otorgó el crédito y 0 si se rechazó la solicitud. :

	ID_CLIENT	ID_SHOP	SEX	MARITAL_STATUS	AGE	QUANT_DEPENDANTS	EDUCATION	FLAG_RESIDENCIAL_PHONE	AREA_CODE_RESIDENCIAL_PHONE
0	2	15	F	S	18.0	0	NaN	Y	31
1	4	12	F	C	NaN	0	NaN	N	31
2	5	16	F	S	28.0	0	NaN	Y	31
3	6	24	M	S	26.0	0	NaN	N	31
4	7	55	F	S	22.0	0	NaN	Y	31
...	...	...	...	...	...	...	...	...	...
50995	29197	16	F	S	72.0	0	NaN	Y	31
50996	43802	23	F	S	31.0	0	NaN	Y	31
50997	1881	4	F	C	51.0	0	NaN	Y	31
50998	42857	25	F	C	71.0	0	NaN	Y	50

   El diccionario de datos se anexa en un excel donde podrá encontrar la descripción de cada variable.

2. Calidad de datos

   - Realice el etiquetado de las variables de acuerdo a su tipo
   - Revisar y eliminar duplicados por ID_CLIENT y defina cuál es la mejor estrategía para eliminar los demás registros que tengan el mismo ID_CLIENT.
   - Completitud
   - Revisión de valores fuera de la naturaleza de la variables (no válidos) y conversion a NaN
   - Completitud
   - Realice la limpieza de variables y realice transformaciones a tipo de dato int o float en continuas (haga normalización de categorías si es necesario)
   - Eliminación de variables que posean un completitud inferior al 80%

   Los resultados o hallazgos encontrados en cada punto de la sección de calidad de datos , debe estar comentada en el PDF de entrega, por ejemplo para el primer punto : "Contamos con 4 variables de tipo continuo que se enlistan a continuación : ... "

3. Análisis Exploratorio de Datos

   Realice análisis interesantes sobre los datos proporcionados , generé gráficas representativas.
   Se deben realizar al menos el análisis de cuatro variables , los análisis deben estar acompañados de una gráfica por variables analizada y en el PDF de entrega deben estar las gráficas acompañadas con la descripción de qué representa ese análisis.

4. Datos anómalos

   - Realice la identification de datos anómalos y elimine aquellos que sean catalogados como outliers por al menos dos métodos , realice una tabla resumen como la que se generó en clase para conocer el porcentaje por método y el porcentaje final, en el PDF se debe agregar el número de outliers eliminado por variable además de una descripción pequeña. Por ejemplo : "La variable AGE presentó un total de 100 outliers , por otro lado la variable PERSONAL_NET_INCOME contó con menos de 50 outliers ..."

   - Además se debe añadir los gráficos del histograma antes de la remoción y después de la remoción de outliers de todas las variables continuas con datos anómalos.

5. Datos faltantes

   - Genere su conjunto de entrenamiento y prueba estratificado , donde el conjunto de prueba tenga el 30% de la información
   - Realice la imputación de valores ausentes , sobre las variables que lo requieran (para las continuas seleccione uno de los métodos posibles y realice KS para conocer el mejor valor a usar)
   - En el PDF agregue las variables imputadas y el valor que se le asignó a los valores faltantes, además de mencionar que método utilizó.

6. Ingeniería de datos

   - Genere nuevas variables a partir de la información que posee
   - En el PDF indique qué variables se crearon

ENTREGABLES

- PDF con resultados
- Código en python (Notebook) , limpio, ordenado , comentado y bien estructurado, sin errores en el código. En el notebook deben mostrarse todas las gráficas que contiene el PDF además del conjunto de entrenamiento y prueba.