**INFORMACION PARA LA REALIZACION DEL ANALISIS EXPLORATORIO DE DATOS**

1. **Define el problema**

Indica que problema tienes que resolver: clasificación, regresión y que es lo que tienes que encontrar en función de los datos que tienes. Esto implica identificar la7s variables dependientes y las independientes

Da una lista de hipótesis sobre el problema teniendo en cuenta si quieres hacer una regresión o una clasificación.

1. **Prepara los datos**

* ***Descripción de los datos***
  + Que variables tienes
* ***Procesamiento de los datos***
  + Formatea los datos de forma que puedas trabajar con ellos
  + Limpia los datos de manera que sean consistentes
    - Missing values, eliminación de variables y/o entradas si fuera necesario. Explicar porqué
    - Imputación si es necesario, que método usarías y porqué
    - Identifica outliers y su tipo, razona que medidas tomarías y porqué
* ***Resumen de los datos***
  + - Crea plots univariables de cada atributo
    - Crea plots bivariables de dos a dos
    - Comprueba hipótesis planteadas creando plots en relación a la variable dependiente/salida
* ***Descompón atributos complicados en sus partes***
  + - Descompón fechas en sus componentes
    - Variables categóricas quizás en atributos binarios
    - Crear mejores variables categóricas.
* ***Busca que datos son redundantes***
  + - Existe correlación
    - ¿Eliminarías atributos altamente correlacionadas?
    - ¿Necesitas aplicar una reducción de la dimensionalidad? Porqué, si necesario puedes hacerlo.
* ***Transformación de datos*** 
  + - Siguen tus datos una distribución normal
    - ¿Serían necesarias transformaciones?: raíz cuadrada, cuadrado, estandarización, normalización, discretización.

1. **Conclusiones.**
   * Indica que hipótesis has podido probar, cuales no y que variables utilizarías y porqué para el modelo correspondiente.