**Estudio de caso: análisis exploratorio de datos**

La empresa A&A Ltda, empieza un proceso de implementación de *Machine Learning*, usted ha sido designado para una de las tareas más importante dentro del proyecto, el cual consisten en realizar el análisis exploratorio de los datos y documentar los resultados encontrados, generando un informe que involucre los procedimientos y los resultados.

El archivo que se analizara corresponde a precios de viviendas y locales para la venta y la colección de datos cuenta con variables que se ven involucradas en ese valor.

La información se encuentra en el siguiente enlace: <https://www.datos.gov.co/Hacienda-y-Cr-dito-P-blico/Inmuebles-Disponibles-Para-La-Venta/72gd-px77/data> la cual usted debe descargar e importar con la herramienta anaconda, el archivo lo puede exportar en formato CSV.

La descarga del archivo se describe en la siguiente imagen:

Interfaz de usuario gráfica, Aplicación, Correo electrónico

Descripción generada automáticamente

También puede encontrar el archivo en los anexos del curso.

El informe debe dar cuenta de:

* Procedimiento para la importación del archivo en formato CSV
* Plante una pregunta objetivo
* Total, de Registros
* Total, de columnas
* Detallado de cada columna
* Identificar cuáles de las columnas son categóricas y numéricas
* Identifique en que columnas existen valores nulos
* Identifique si existen registros duplicados
* Realice un reporte estadístico de los datos numéricos (media, moda, mediana, desviación estándar, cuartiles, entre otros que considere)
* Identifique columnas con valores erróneos
* Utilice gráficos para identificar valores atípicos
* Realice histogramas de frecuencia
* Use la herramienta para gráficos para determinar correlación entre variables
* Realice y explique la eliminación de datos nulos y duplicados
* Agrupe columnas que considere pueden generar información importante
* Cree nuevas columnas a partir de las existentes
* Identifique columnas que no aportan de acuerdo con su pregunta objetivo
* Realice conclusiones sobre las variables que considere tienen mayor relevancia
* Dejar documentado usando gráficos y capturas de pantalla todo el proceso realizado.