

TRABAJO PRÁCTICO 1

En este primer práctico se trabajará en conocer en detalle el proyecto, realizando una exploración completa de los datos. El concepto de un proyecto de Data Science es "Contar historias con datos", en donde podamos demostrar de manera explícita lo que nos dicen los datos. Para lograr este objetivo organizaremos el trabajo práctico en etapas.

1.- Entender el dominio.

Para lo cual deberá crearse dentro del notebook, los siguientes apartados.

a.- Breve descripción de la problemática.

b.- Contexto comercial, consecuencias que crean las problemáticas, beneficios que aporta este tipo de análisis.

No perder el foco de lo que se busca en este proyecto, suele ser difícil al encontrarnos con muchos datos que no conocemos.

2.- Comenzar a responder algunas preguntas que nos podemos hacer como Data Scientist:

a.- En la práctica, los dataset no son obtenidos de forma sencilla, limpios y listos para realizar análisis de datos. Por lo general se realiza un trabajo de ETL. ETL es un acrónimo que significa "Extract, Transform, Load" (Extraer, Transformar, Cargar). En este trabajo práctico realizaremos una pequeña parte de este dominio, por lo que se deberá transformar los archivos .txt a csv, Importar los dataset, unirlos en función de una variable, cambiar los nombres de las columnas según el documento "Descripción del Dataset".

b.- Una vez realizado el trabajo de ETL es importante analizar el dataset para evaluar con que datos se cuenta, cantidad y calidad.

Tipos de datos (int, float, object, datetime,etc), también evaluar su naturaleza (temporales, económicos, cuantitativos, demográficos, etc).

c.- Evaluar si es posible llegar al objetivo planteado con los datos disponibles.

d.- Cantidad de observaciones y variables en el dataset.

e.- Que periodo de fechas disponemos en el dataset.

f.- ¿Tenemos disponible la variable "Target", que se quiere predecir?, ¿Es necesario realizar operaciones o transformaciones para obtenerla?

g.- Desde el punto de vista histórico o cantidad de muestras ¿es viable realizar un modelo predictivo?

Es importante en esta etapa pensar, planificar y documentar. Una vez que se tiene definido el problema a resolver, se puede avanzar en estudiar en detalle los datos para interpretar sus comportamientos y entender los mas que se pueda del negocio.

3.- Limpieza y revisión de Datos, comenzar con el proceso de Feature Engineering:

En esta etapa es necesario analizar los datos con ojo critico de limpieza. Para ello debemos investigar cada variable:

- a.- Datos nulos. Se eliminan o se imputan.
- b.- Valores iguales a 0. Se eliminan o se imputan.
- c.- Existencia de outliers, analizarlos y evaluar que medidas tomar.
- d.- Existencia de datos innecesarios, eliminarlos.
- e.- Eliminación de datos duplicados.

Creación de variables.

En la mayoría de los casos crear nuevas variables es de gran ayuda para poder analizar y visualizar los datos de manera que facilitan la interpretación y nos permiten determinar insights. Por ello necesario crear las siguientes columnas:

- a. Día de la semana en el cual se realiza la medición.
- b. Mes del año en el cual se realiza la medición.
- c. Consumo mensual por medidor.
- d. Todas las que crean necesarias.

Graficar:

- a. Consumo anual
- b. Datos climáticos anuales
- c. Todos los que crean necesarios (distribuciones, estadísticos, etc)

REQUISITOS A TENER EN CUENTA.

- Los proyectos de Data Science tienen que estar estructurados.

- El trabajo se debe desarrollar en un notebook, el cual debe contar con un índice y sus correspondientes apartados, el código debe ser fácil de leer, estar probado y comentado.
- Es importante tener en cuenta que se pueden realizar diferentes visualizaciones y tipos de análisis, pero en el entregable solo debe estar desarrollado todo lo que sea relevante.

EXITOS!!