

1 Definición del proyecto

En este proyecto se busca que el/la estudiante realice un análisis exploratorio de datos (en ingles EDA) sobre datos tomados de un sensor. Se debe realizar inicialmente una limpieza de los datos, determinar si hay datos faltantes y evaluar si es pertinente generar un conjunto de datos sin datos faltantes. Explorar por meses, días y años si existe algún patrón que se repita en el tiempo. Buscar correlaciones entre los datos y analizar de manera estadística los datos.

1. Cargar la base de datos y determinar cuales son las variables, fecha de inicio y fecha final, mínimos, máximos, desviaciones estándar, y medias de cada variable medida. En algunos casos se debe transformar las columnas de la fecha ya que vienen divididas por año, mes, día y hora, en diferentes columnas.
2. Graficar líneas de tendencia (plot) de cada variable en el tiempo. Que se puede decir a primera vista sobre los datos en el tiempo? (son cíclicos, existe alguna tendencia, hay años con comportamientos anormales comparado con los otros años.)
3. Generar un conjunto nuevo de datos filtrando los datos faltantes (dropna). Evaluar si es pertinente realizar esta operación.
4. Determinar si existen datos atípicos empleado BoxPlots sobre cada variable (columna) para todos los registros (filas).
5. Hacer un análisis de correlación entre las variables, y adicionalmente analizar los histogramas de cada variable.
6. Realizar BoxPlots de cada variable analizando el comportamiento para cada mes a lo largo de los años. Por ejemplo, un gráfico para el mes de enero, con un Boxplot por cada año, si son 3 años, entonces en la misma figura de enero, deben aparecer 3 BoxPlots de cada año.
7. Agrupar los datos por año y por mes, calculando la media despues del agrupamiento. Hacer un grafico de barras de cada mes a lo largo de los años. Por ejemplo si son 3 años, graficar la media de cada mes (3 barras) de una de las variables de la base de datos.

Si la base de datos solo tiene un año de análisis, cambiar año por mes, y mes por “día de la semana” en los puntos anteriores.

Se debe enviar el notebook con los análisis y desarrollos de cada punto. Incluir una sección de conclusiones.

2 Paginas de ayuda

Complete Guide on Time Series Analysis in Python, <https://www.kaggle.com/code/prashant111/complete-guide-on-time-series-analysis-in-python>.