**Exploración y análisis de series, EDA, y Representación gráfica**

Durante el proceso de revisión, validación y comprensión de los datos, es necesario utilizar diferentes técnicas que permitan identificar discontinuidades, cambios en el comportamiento temporal, y en general revisar los estadísticos característicos de cada serie por parámetro.

La exploración y análisis de datos o EDA, por sus siglas en inglés, se refiere al proceso de investigación, mejoramiento, descubrimiento de patrones de comportamiento, e identificación de anomalías sobre conjuntos de datos; además, permite realizar pruebas de hipótesis, y revisar las suposiciones que frecuentemente realizamos frente a ciertas variables climatológicas, por medio de resúmenes estadísticos, tablas dinámicas y gráficas detalladas. Como buena práctica y antes de crear modelos complejos de análisis o pronóstico, se recomienda entender primero los datos para tener una idea de su comportamiento y relación.

*Tenga en cuenta que, en este video se realiza una explicación general del procedimiento a realizar, y que el procedimiento detallado para completar las actividades requeridas, se encuentra en la guía de clase disponible en el enlace de la descripción.*

**Objetivos**

* Aplicar técnicas de exploración y análisis de datos a series temporales.
* Identificar los tipos de datos contenidos en el conjunto de datos IDEAM.
* Identificar valores nulos en las colecciones de datos para cada objeto o variable.
* Obtener estadísticas generales para las variables numéricas del conjunto de datos.
* Identificar las estaciones contenidas en el conjunto de datos.
* Contar el número de registros por estación para cada parámetro.
* Crear tablas dinámicas por parámetro.
* Crear tablas de correlación por parámetro.
* Realizar análisis segmentados generales de cada parámetro, como visualización conjunta de series temporales y densidades KDE, estadísticos generales, matriz de correlación, estadísticos de correlación. Realizar análisis individuales por estación a partir de sus metadatos, con análisis estadístico para los valores registrados, gráficos de serie de tiempo, caja de bigotes, histograma y densidad.

**Procedimiento general**

El diagrama mostrado en pantalla, contiene el procedimiento general, para la segmentación temporal de las series obtenidas a partir de estaciones terrestres del IDEAM, los análisis de series de tiempo, densidad y correlaciones que se realizan para cada parámetro.

Para iniciar, descargue el script EDA punto py, disponible en la carpeta punto .src de este repositorio, y guárdelo en la carpeta local punto src de su equipo.

**Funcionalidades del script**

* Permite definir la ventana de tiempo para el análisis, a partir de un año inicial y final.
* Generación de reporte detallado, en formato Mark down.
* Análisis masivo de registros por parámetro, generando tablas dinámicas y gráficas generales de análisis.
* Análisis masivo de estaciones por parámetro, generando estadísticos y gráficas detalladas de análisis.

En pantalla se muestran las instrucciones iniciales del script, en la guía de clase encontrará su contenido completo.

En la carpeta local punto datasets de su repositorio local, cree una carpeta con el nombre IDEAM \_ EDA y verifique que en la carpeta IDEAM, del mismo directorio datasets, se encuentre el archivo IDEAM joined punto c s v, que contiene todos los registros IDEAM descargados previamente.

Desde la consola de comandos de Windows, ingrese a la carpeta IDEAM \_ EDA. Luego, ejecute la instrucción mostrada en pantalla, que realizará todo el proceso de análisis.

Durante la ejecución, podrá observar que en la consola se presenta el detalle de los procesos ejecutados y la previsualización de las diferentes tablas.

Durante el proceso de ejecución del script, se genera automáticamente un reporte científico integrado de resultados en formato Mark down, que contiene los resultados de análisis de todos los parámetros contenidos en la base de datos, y diferentes archivos en formato punto c s v, con las tablas dinámicas y análisis de correlación.

Dentro de la carpeta Graph, han sido exportadas las diferentes gráficas generales y detalladas de análisis.

Durante el proceso de ejecución del script, se genera automáticamente un reporte científico integrado de resultados en formato Mark down, que contiene los resultados mostrados en pantalla.

Al inicio del reporte Mark down, encontrará una descripción general del análisis y las librerías utilizadas, también información relacionada con el data frame.

luego, diferentes tablas con estadísticos relacionados con conteo de nulos,

descriptiva de variables numéricas del data frame,

las estaciones disponibles para el análisis,

el conteo de registros discretos por parámetro y estación,

análisis generales por parámetro,

análisis de densidad por parámetro,

estadísticas para los valores registrados en cada estación,

localización geográfica de la estación,

gráficos de análisis para cada estación mostrando la serie de tiempo y caja de bigotes,

histograma de frecuencias y análisis de densidad por estación.

Al final de la descriptiva por estación para cada parámetro, se presentan las tablas dinámicas de agrupamiento,

el análisis de correlación de cada estación contra todas las demás estaciones disponibles,

las estadísticas de las correlaciones,

y un gráfico de análisis de correlación.

Para conocer los resultados de los análisis de temperatura máxima y mínima, evaporación y caudal, consulte el reporte disponible en la guía de clase de esta actividad.

**Actividades complementarias**

*En la guía de clase, se encuentran listadas las actividades adicionales que los estudiantes deben desarrollar y documentar para complementar los conocimientos y alcances definidos en este curso. También encontrarás información y referencias adicionales, te recomiendo realizar estas lecturas, que te ayudarán a fortalecer y complementar tus conocimientos.*

*Para completar la Exploración y análisis de series, EDA, y su Representación gráfica, consulta la guía de clase detallada de esta actividad. Si necesitas ayuda, da clic en el enlace Ayuda o Colabora, que se encuentra en el enlace adjunto de la descripción.*