

Lucía Aranda

UTE

EmTech Institute

Proyecto 4

Resultados del Proyecto 04 en Machine Learning

Para la resolución del caso práctico cuatro en el nivel uno, primero seleccione una página web con datos útiles para realizar un web scraper, que fue una página dedicada al tiempo del clima.

Para comenzar se requirió de la instalación de las herramientas beautifulsoup4 y requests.

Posteriormente se utilizaron las librerías BeautifulSoup, requests, re, datetime y pandas para la manipulación de grandes cantidades de datos.

Posteriormente, ya con las herramientas instaladas, procedí a la extracción de datos, donde realice una solicitud con requests.get, para obtener una respuesta, que para comprobar posteriormente la imprimí. Con BeautifulSoup se analizó el contenido para posteriormente colocarla en una tabla, con find_all, y para confirmar imprimí la primera parte de la tabla y obtener la primera serie de datos, que convertir a texto con .text, y filtrar una lista con el contenido del texto, pero quitando al mismo tiempo los espacios vacíos.

Después, con outer_data_list se separó los datos y los guarde en formato de lista, y con un bucle que recorre los datos de cada día, y con una variable nueva "value" asigne la función de los datos filtrados, y lo anexe a la lista con .append, finalmente observar dichos datos.

Con la variable nueva day_data_list se mandaron los datos a un contenedor y crear un nuevo bucle donde se enlistan los valores en una variable para valores numéricos (el contenido), y un bucle para contener los dígitos, después de ser analizados en formato de texto, que se anexó a la primera lista, después asignar el nombre a las columnas que sostendrán la información, y finalmente se crea la variable del nombre con el que se llama la dataset, donde se convergen las listas de contenido y las columnas.

Comienzo el análisis exploratorio con la función .head() confirmó la creación del dataset para ahora manipular los datos, para ello comienzo por conocer la data que

estoy manejando con `.info()`, y ya que para los cálculos estadísticos se requiere un `datatype numeric`, cambio de tipo de dato objeto a numérico, que se asignan automáticamente. Después confirmo imprimiendo la misma información del dataset. Calculo las dimensiones del dataset con `.shape[]`, dentro de los corchetes específico si son las filas o las columnas de acuerdo a su posición en la cadena para por medio de un comentario de pueda imprimir específicamente a que pertenece cada número.

Finalmente calcular la media con `.mean()`, el máximo de cada columna `.max()`, el mínimo de cada columna `.min()`, y la desviación estándar de cada columna `.std()`.

Conclusión

Dentro de este sector económico, se puede llegar a la conclusión de que la mejor temporada para invertir en accesorios y utensilios para el temporal de primavera-verano es en mayo, puesto que en los valores máximos se encuentran en la temperatura y humedad. Con esto se puede llegar a predecir que ventas convendría realizar durante las diferentes estaciones del año para empresas que se dedican a la venta departamental, sin perder ganancias, basándose no solo en años anteriores sino en los datos actuales del clima que se monitorean por medio de dispositivos IoT.