Hector Manuel Gonzalez Villarreal

178679

Ciencia de Datos

Apuntes Modulo 4

Abrir y leer archivos

En Python es sencillo abrir archivos con la función open, tenemos 2 parametros el path del archivo y el modo de apertura, para leer o escribir

Recordemos que para crear cambios de líneas e indentaciones usamos como en todos los lenguajes caraceteres especiales

La función open también se puede usar para escribir, solo que en vez de usar su función read o usamos funciones write. Podemos leer solo parte de los archivos usando readln y si no queremos todo el renglón podemos especificar el numero de caracteres a retraer.

En los ejercicios veo ya comandos mas avanzados como el await al traer el archivos, y eso ya me hace mas sentido de la fuerza del lenguaje, mi transición de C# la estaba viendo muy sencilla ya que solo estaba ejecutando comanditos separados, No veo de momento un writeln como lo hay en c# muy practico para no escribir los caracteres especiales.

Veo que el copy es un barrido manual, me extraña que no haya función de copy como tal o no es del alcance del curso

Es importante cerrar los archivos después de terminar de trabajar con ellos. Podemos usar la función close o usar bloques de código con el with

Empiezan los laboratorios a ser mas interesants con bloques de código completo donde tenemos que recordar ciclos y asignaciones. Pero no me agrada que venga la solución escrita en una liga de solución.

Pandas

Python usa alias en los importes para acortar el uso de las librerías

Con pandas podemos usar archivos con formatos comunes como csv o Excel

El concepto de dataframe me parece interesante ya que nos ayuda a darle o a identificar el formato de los datos que contienen los archivos

Me agrada el uso de los Brackets para obtener columnas o “celdas” especificas

La selección de las columnas puede ser por “nombres” de columnas o si relación numérica

Para hacer un distinct como en sql esta la función unique()

Podemos usar operadores directo a los dataframes como filtros usando > < etc

Podemos los resultados de los dataframes con las funciones to\_tipodearchivo

Interesante la manera directa y sencilla de manipular datos desde Python empiezo a ver una de las diferencias entre este lenguaje y los tradicionales que he usado siempre, es una combinacon de C# con SQL integrado