1. **Θέλω να ανοίξω ένα αρχείο excel**

Α. Εισάγω την βιβλιοθήκη pandas

**import pandas as pd**

B. Καλώ το αρχείο airline passengers

Dataset=pd.read\_excel(r'D:\OneDrive - HIT\Desktop\MaunaLoaDailyTemps.xlsx')

**Πατάμε run και στα δεξιά πατάμε πάνω στο variable explorer για να δούμε τον πίνακα που φτιάξαμε**

1. **Μετατρέπουμε την ημερομηνία των δεδομένων σε ημέρα για να μπορέσουμε να το χρησιμοποιήσουμε σαν ανεξάρτητη μεταβλητή**

Εισάγουμε την βιβλιοθήκη datetime για να μπορέσει η python να διαβάσει τις ημερομηνίες

**from datetime import datetime**

Στη συνέχεια πρέπει να ορίσουμε ποια στήλη είναι η στήλη με την ημερομηνία χρησιμοποιώντας της ακόλουθη εντολή

**dataset["DATE"]=pd.to\_datetime(dataset["DATE"])**

Τέλος μετατρέπουμε τις ημερομηνίες σε αριθμό εβδομάδων

**dataset['weekday']= dataset["DATE"].dt.weekday**

1. **Για να εφαρμόσουμε πρόβλεψη θα ήταν βολικό να ενώσουμε όλες τις ανεξάρτητες μεταβλητές (x) αριστερά και την εξαρτημένη δεξιά (y).**

Ας δούμε λοιπόν ποιες είναι οι κεφαλίδες των στηλών των δεδομένων μας

**for col in dataset.columns:**

**print(col)**

Κάτω δεξιά θα βγουν τα ονόματα των στηλών:

MinTemp

MaxTemp

AvgTemp

Sunrise

Sunset

Weekday

Άρα ο κώδικας για την τελευταία μορφή του dataset ορίζεται από

**df=dataset[["MinTemp","MaxTemp","Sunrise","Sunset","weekday","AvgTemp"]]**

1. **Διαχωρισμός των δεδομένων σε ανεξάρτητες μεταβλητές x και εξαρτημένη y**

**X = df.iloc[:, 0:5].values**

**y = df.iloc[:, 5:6].values**

1. **Διαχωρισμός των δεδομένων σε train και τεστ**

Ο σκοπός του διαχωρισμού είναι για να μπορέσω να εκπαιδεύσω τον αλγόριθμο μου στα x,y του train set, και στη συνέχεια, να εξετάσω την αποτελεσματικότητα του στα x,y του test set. Δηλαδή, και αφού εκπαιδεύσω τον αλγόριθμο στο train set, θα εισάγω τις τιμές του x του test σετ στον αλγόριθμο και θα συγκρίνω τις προβλέψεις του y του αλγορίθμου με τις πραγματικές τιμές y του τεστ σετ οι οποίες αντιστοιχούν στα x του test set

Για τον διαχωρισμό εισάγω από την βιβλιοθήκη sklearn, την δυνατότητα διαχωρισμού των δεδομένων με τις ακόλουθες εντολές ορίζοντας ότι το train set θα είναι το 80% του συνολικού dataset και το test set το υπόλοιπο 20%.

**from sklearn.model\_selection import train\_test\_split**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.2, random\_state = 1)**

1. **Προσαρμόζω ένα μοντέλο πολλαπλής παλινδρόμησης στα x train και y train**

Εισάγω την βιβλιοθήκη παλινδρόμησης και κάνω προσαρμογή

**from sklearn.linear\_model import LinearRegression**

**regressor = LinearRegression()**

**regressor.fit(X\_train, y\_train)**

1. **Χρησιμοποιώ το προσαρμοσμένο πια μοντέλο για να προβλέψω το y τεστ**

**y\_pred = regressor.predict(X\_test)**

1. **Αξιολογώ το σφάλμα με το MAD**

Εισάγω την βιβλιοθήκη το μέσο απόλυτο σφάλμα και το εφαρμόζω ως η διαφορά αναμεσά σε αυτό που πρόβλεψα και αυτό που υπάρχει στην πραγματικότητα

**from sklearn.metrics import mean\_absolute\_error**

**mae=mean\_absolute\_error(y\_test,y\_pred)**