COL290: DESIGN PRACTICES

ASSIGNMENT 1 – SUBTASK 1

OBJECTIVE –

Given a stock symbol (say SYM) and number of years (say x) as input from make, need to extract the SYM stock's data (specific columns) for the last x years and save it in different file formats such as csv, txt etc. Then, compare the write time and space used by each file types using a graph to plot the points as space used vs time taken.

LIBRARIES/MODULES USED –

1. jugaad-data – for extracting data for the stock from NSE

2. Pandas – for storing the data extracted into different file formats and filtering the columns.

3. datetime & dateutil.relativedelta – for using x and calculating start date and end date of the data to be written

4. matplotlib – for plotting, storing the graph and labelling axes, points and graph.

5. sys – for taking sys args

4. os – for calculating space used by each file

5. time – for calculating time taken to write in the file

DESIGN CHOICES –

- Maintained 2 lists – sizes and times
  Used these 2 lists to store the size and time for each file type in a ordered way so that it can be retrieved using index
- For each file type, calculate time taken in ms to write in file using start time and end time difference and stored it in times list
- For each file type calculate space required by it in megabytes and stored it in sizes list.
- Used these 2 lists as axes for the graph to be plotted
  Plotted a point in time vs size space for each file type and marked it with a different color stored in color list.

INSIGHT ABOUT FILE TYPES

1. csv and txt – they have almost same write speeds and sizes , they occupy low space and low time relatively

2. html – time taken and space used is very high relative to others

3. json – time taken is least but size used is higher than csv, txt, and parquet

4. parquet – space used is least but time taken is higher than csv, json and txt

LEARNINGS

- Learnt about Makefile for the first time.
- Learnt to architect the assignment and plan about what needs to be done for the first time since col106 assignments were heavily detailed with little to no flexibility for implementations.
- Learnt to use modules such as matplotlib and pandas for the first time.
- Learnt about os, datetime and time modules.
- Understood the importance of version management.