

Q 1(a)

[12 Marks]

Using the topic from your CA682 visualisation assignment, apply the Generic Data Analytics Pipeline to describe how the data may have been Gathered, Processed, Analysed, Presented and Preserved. Give a brief description of the activities at each stage (1-2 sentences) and identify any specific tools that you did or would use. If you didn't specifically perform any stage then you can make assumptions or predictions about the actions and tools.

If you didn't complete a visualisation assignment then write about a scenario based on analysing *student feedback to DCU in relation to online teaching in 2020*.

Gathering

Processing

Analysis

Presenting

Preserving

Gathering:

Activities: Information on stock prices, volume of trading and other related financial variables were obtained from [Alpha Vantage] through its (API). The data variables include historical and current stock data from Tesla's IPO and up to 2024.

Tools: API requests using Python `requests` library and data in JSON format received from the websites.

Processing:

Activities: The raw JSON data was migrated through an ETL process into a Relational Database Management System (RDBMS). Activities such as identifying and handling missing values, converting the data type to float, scaling the target variable closing prices for better comparison also came under data cleaning process. New columns created were `month_id`, `year_id`, and % change in closing price for better analysis blueprint.

Tools:

- For ETL: `SQLAlchemy`, MySQL.
- For processing: `pandas`, `ast`, and `json`.

Analysis:

Activities: Hence, normalized closing prices and trade volumes were used as variables, to match Tesla's ambitious growth progression to that of Amazon, Apple, IBM, and Microsoft. One of the reasons for normalizing the results was to compare the organizations with equal starting points with respect to price levels.

Tools: main data manipulation tool in this analysis was done with the help of `pandas`.

Presenting

Activities: Visualizations were created to highlight Tesla's unique growth, using a line chart to compare normalized stock performance over time. Design considerations included color-coding, subtle gridlines, and selective x-axis points for clarity.

Tools: `matplotlib`, `seaborn` libraries.

Preserving

Activities: The processed and visualized data, along with the analysis, was stored in a database for easy updates and reference. Incremental loading ensured that new data could be added without redundancy or system strain.

Tools: There was MySQL for data storage; Jupyter Notebook for documentation as well as reproducibility purposes.