**Tesla Case Study Documentation**

**Project Overview: Tesla Stock Dataset Analysis (Sprint 1)**

**Group no:** 03

**Overview**

The code notebook analyses Tesla’s data to uncover key trends, insights, and metrics using Python-based data analysis libraries. This project involves analysing Tesla stock data to gain insights into historical price trends, trading volumes, and relationships between various stock attributes. The analysis includes data preprocessing, visualization, feature engineering, and exploratory data analysis (EDA). The goal is to prepare the data for predictive modelling or deeper financial analysis.

**Dataset Description**:

The dataset used for this project is Tesla stock data, which includes attributes such as:

* **Date:** The trading date
* **Open:** Opening price
* **High:** Highest price of the day
* **Low:** Lowest price of the day
* **Close:** Closing price
* **Adj Close:** Adjusted closing price
* **Volume:** The number of shares traded

**Libraries Used:**

The following Python libraries are utilized in this project:

1. **pandas:** For data manipulation and analysis
2. **matplotlib:** For plotting graphs and visualizations
3. **seaborn:** For advanced visualization and heatmaps
4. **numpy:** For numerical operations
5. **scipy.stats:** For statistical analysis (skewness, kurtosis, z-scores)
6. **warnings:** To manage warnings during execution
7. **datetime:** For handling and formatting date data
8. **sklearn.preprocessing:** For scaling and normalizing data
9. **sklearn.decomposition:** For Principal Component Analysis (PCA)
10. **sklearn.discriminant\_analysis:** For Linear Discriminant Analysis (LDA)

**Key Code Sections:**

**Code Explanation**

**1. Importing Libraries**

The first code block imports all necessary libraries to ensure smooth data processing and visualization.

**CODE:**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

from scipy.stats import skew, kurtosis, zscore

import warnings

import datetime as dt

**Purpose**: Sets up the environment by loading the required libraries.

**2. Data Loading and Description**

* The dataset is loaded using pandas.read\_csv().
* Data attributes are inspected using:
  + df.head() to preview the first few rows.
  + df.shape for dimensions.
  + df.columns for column names.
  + df.dtypes for data types.

**3. Data Cleaning and Transformation**

* **Missing Values**: Checked using df.isnull().sum().
* **Unique Values**: Extracted for the Date column to verify data integrity.
* **Sorting**: Data is sorted by Date for time-series analysis.
* **Correlation**: A correlation matrix is computed (excluding Date) to identify relationships among numerical features.

**4. Data Visualization**

* **Histograms**: Plotted for Close and Volume to observe distributions.
* **Line Plots**: Created for:
  + Closing price trends over time.
  + Open, High, Low, and Close prices to observe overall trends.
* **Boxplot**: Visualized the distribution of prices (Open, High, Low, Close).
* **Correlation Matrix Heatmap**: Displayed interdependencies between numerical attributes using a heatmap.

**5. Data Preprocessing**

* **Scaling**: Applied both Min-Max Scaling and Standardization to normalize numerical attributes.
* **Standardization**: Scaled features to have a mean of 0 and standard deviation of 1.
* **Normalization**: Min-max scaling to fit values between 0 and 1.

**6. Exploratory Data Analysis (EDA)**

* **Bivariate Analysis**:
  + Heatmap of correlations.
  + Scatter plot of Volume vs Close.
  + Pairplot for key numerical features (Open, High, Low, Close, Volume).

**7. Feature Engineering**

* **Dimensionality Reduction**:
  + Applied PCA to reduce high-dimensional data into 2 principal components (PCA1, PCA2).
* **Trend Analysis**:
  + Generated a Trend column to indicate daily stock movement (Uptrend, Downtrend, Neutral).
  + Encoded Trend as numerical labels for further analysis.

**8. PCA Scatter Plot**

* Scatter plot created for PCA1 vs PCA2, with points coloured by Trend\_Label to visualize clustering and trends in stock data.

**Key Highlights**

* The project emphasizes preprocessing for handling missing values, scaling, and creating derived features like trends and PCA components.
* Various visualizations provide insights into the dataset, such as price distributions, trends over time, and relationships between attributes.
* Feature engineering methods like PCA add depth for dimensionality reduction and advanced analysis.

**Conclusion**

This project provides a comprehensive analysis of Tesla’s dataset. By visualizing trends and performing statistical analysis, it offers actionable insights into stock behaviour and patterns. Future work could involve predictive modelling or deeper analysis of factors influencing stock prices. This documentation ensures clarity and understanding of the data processing pipeline and prepares the groundwork for subsequent sprints in the project.