PSET 9: Analyzing Headline News Impact on the Stock Market

Mandatory Tasks (Total 100 Points)

1. **Data Loading and Preprocessing (10 Points)**:

   - Load **stock_n_hl_news.csv** into DataFrame **df**.

   - Convert 'Date' column to datetime format.

2. **Market Trend Column (15 Points)**:

   - In **df**, create a column 'Trend'.

   - Label 'Trend' as 'Bullish' if NASDAQ index is higher than the previous day, 'Bearish' otherwise.

   - Assume 'Bullish' for the first date in the dataset.

3. **Sentiment Analysis (25 Points)**:

   - Perform NLP analysis on 'Headline' to obtain compound sentiment scores.

   - Store scores in 'StmScore'. Categorize as 'positive' (>0.7), 'neutral' (-0.7 to 0.7), 'negative' (<-0.7) in 'Sentiment'.

4. **Linear Regression Analysis (25 Points)**:

   - Analyze the relationship between AAPL stock prices and 'StmScore'.

   - Apply linear regression; assess if sentiment scores predict AAPL prices.

   - Interpret results and summarize findings in a Markdown cell.

5. **Logistic Regression for Trend Prediction (25 Points)**:

   - Predict 'Trend' based on 'StmScore'.

   - Apply logistic regression; split data into training and testing sets.

   - Discuss model performance and key takeaways in a Markdown cell.

Optional Extra Credit Tasks (Up to 30 Points, No Cap at 100)

1. **Advanced EDA (Up to 20 Points)**:

   - **Descriptive Statistics**:

     - Compute summary statistics (mean, median, standard deviation) for key columns (AAPL, NASDAQ, NYA, SP500, DJI, sentiment scores).

     - Identify outliers or anomalies.

     - **Markdown Summary**: Explain observations of outliers or anomalies.

   - **Distribution Analysis**:

- Create histograms for the distributions of stock indices and sentiment scores.

- Check and discuss skewness or patterns.

- **Markdown Summary**: Describe distribution observations.

- **Correlation Analysis**:

  - Calculate and visualize correlations among stock indices and sentiment scores.

  - Use correlation matrices for visualization.

  - **Markdown Summary**: Discuss correlation findings.

- **Trend Analysis**:

  - Plot time series graphs for stock trends and sentiment scores.

  - Highlight patterns or shifts using preattentive attributes.

  - **Markdown Summary**: Explain trend analysis observations.

2. **Advanced Analysis (Up to 10 Points)**:

   - **Impact on AAPL Prices**:

     - Use linear regression to assess the impact of indices (NASDAQ, NYA, SP500, DJI) on AAPL stock prices.

     - Check for multicollinearity using VIF.

   - **Predictive Modeling**:

     - Predict AAPL stock prices with regression models; assess accuracy.

   - **Markdown Summary**:

     - Detail analysis and findings, highlighting multicollinearity issues if present.