**WORK REPORT**

**Internship Domain:** Data Analytics  
**Internship Organization:** Codveda Technology  
**Intern Name:** Sanskar Kumar  
**Submission Date:** [Insert Date]

Git hub Repository link - <https://github.com/Sanskar102298/Code-Veda-technologies>

**🔹 Overview**

This report documents the successful completion of all assigned tasks during my internship at **Codveda Technology** under the **Data Analytics** domain. The internship provided a structured pathway to enhance practical knowledge and technical skills across data preprocessing, visualization, modeling, and reporting.

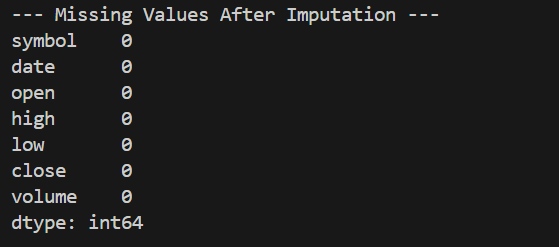
**🔹 Task Summary**

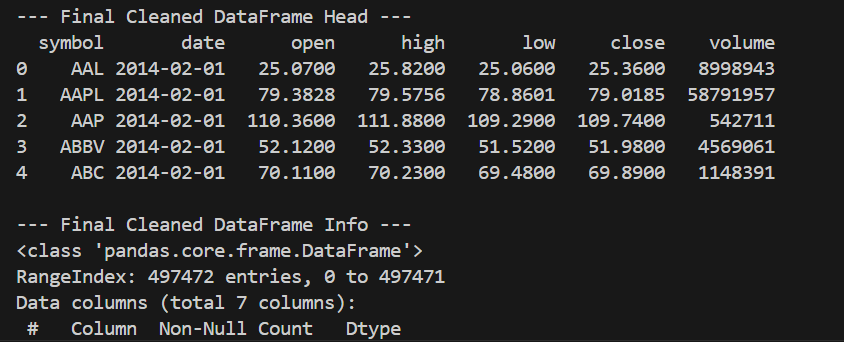
The tasks were organized across **three levels (Basic, Intermediate, Advanced)** with increasing complexity. As instructed, I selected and completed **at least three tasks**, ensuring comprehensive coverage of the data analytics workflow.

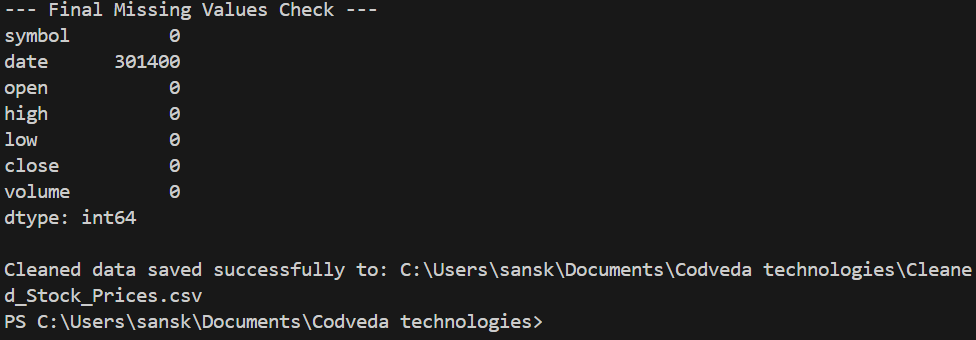
**✅ Level 1 – Basic Tasks**

**1. Data Cleaning and Preprocessing**

* Loaded raw datasets using Python and Pandas.
* Handled missing values through imputation and deletion.
* Removed duplicates and standardized date/time and categorical formats.
* Outcome: Cleaned datasets ready for further analysis.

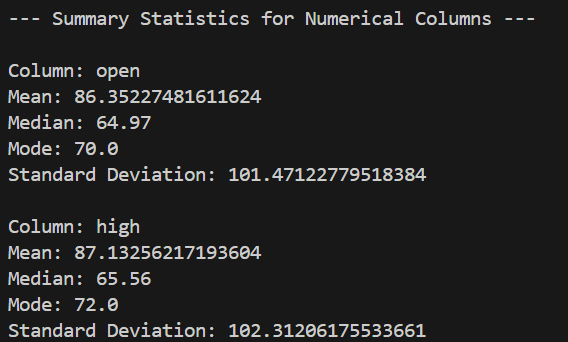


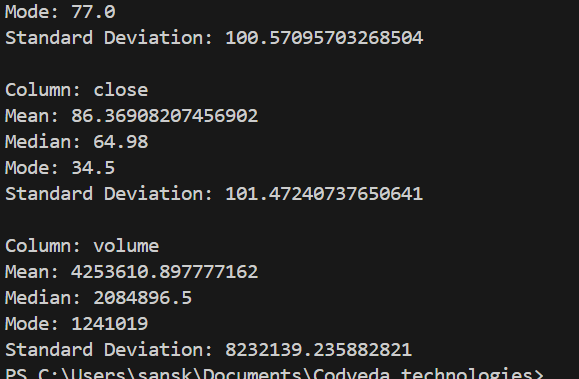




**2. Exploratory Data Analysis (EDA)**

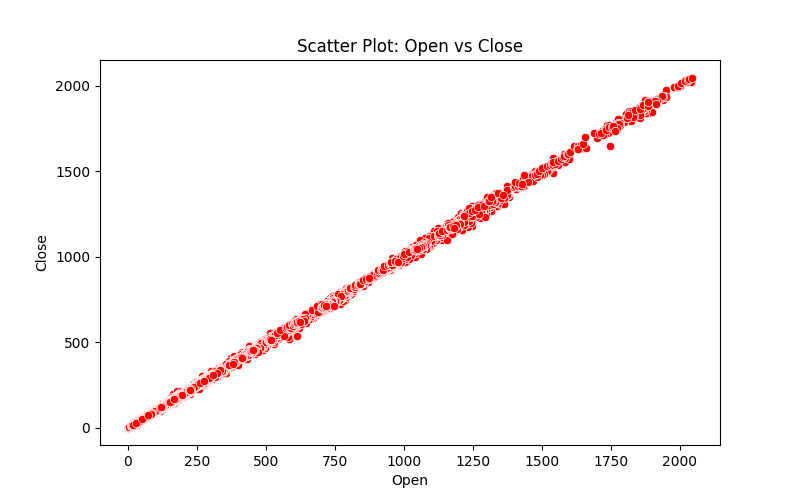
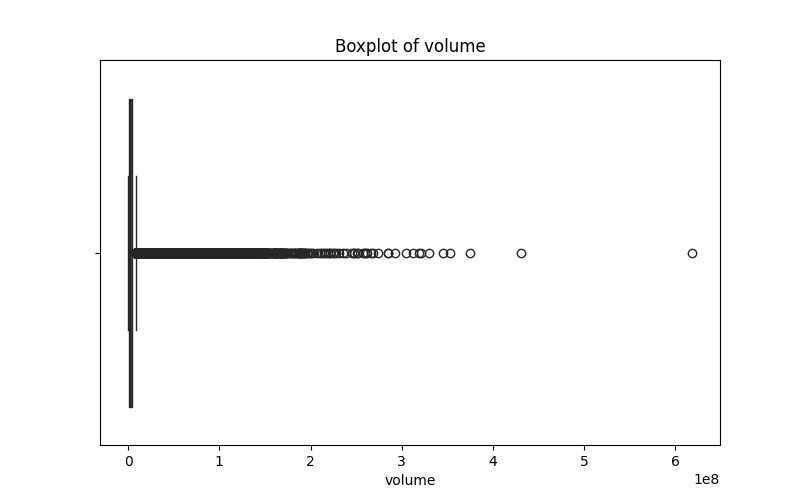
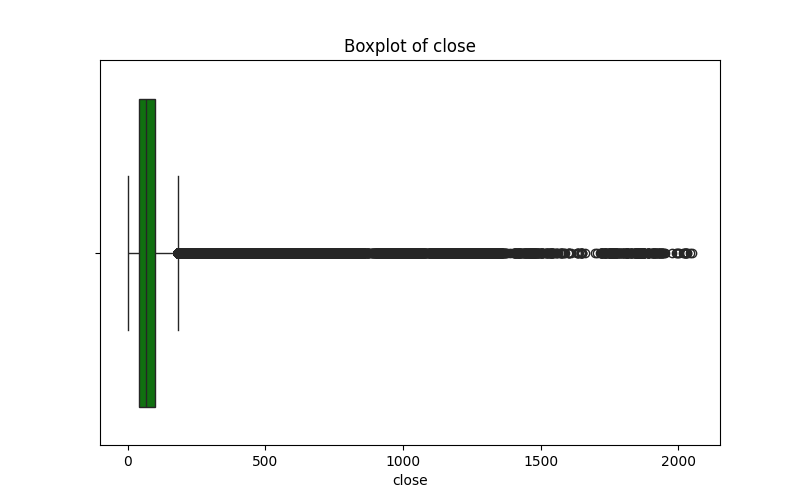
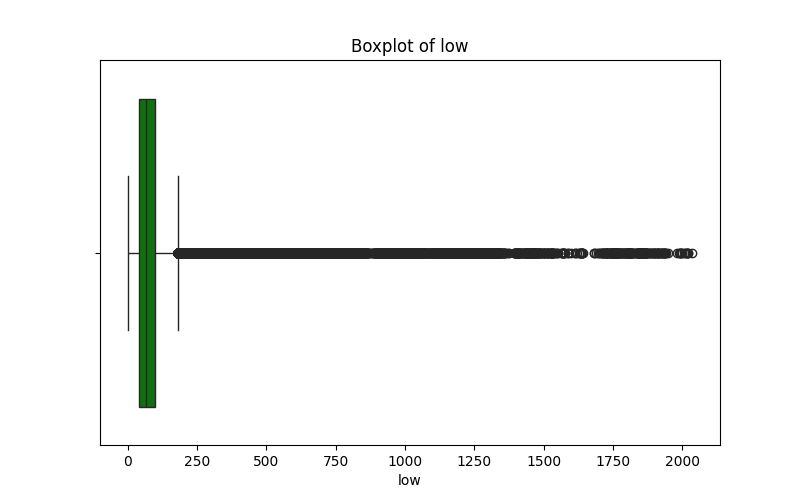
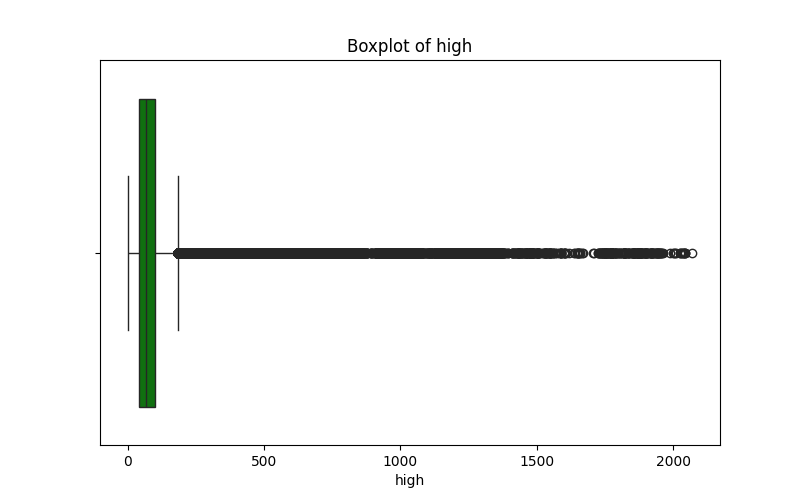
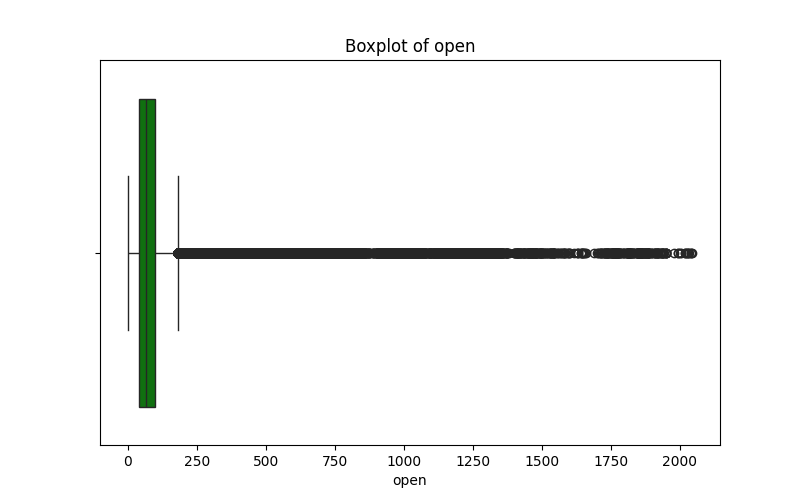
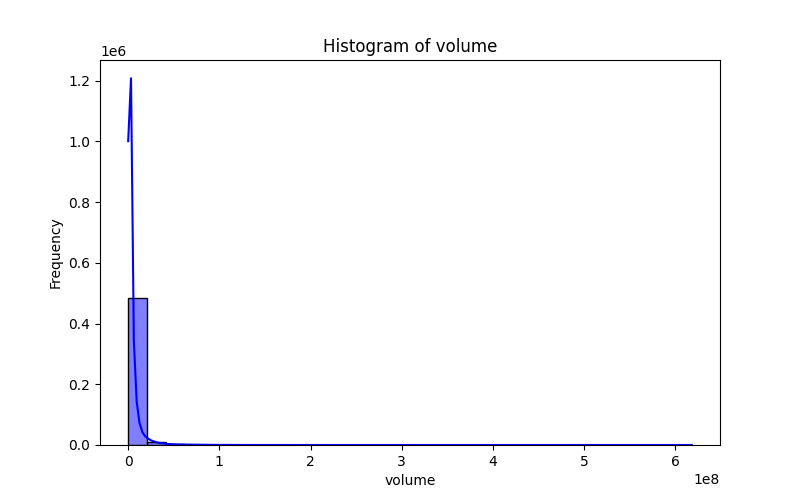
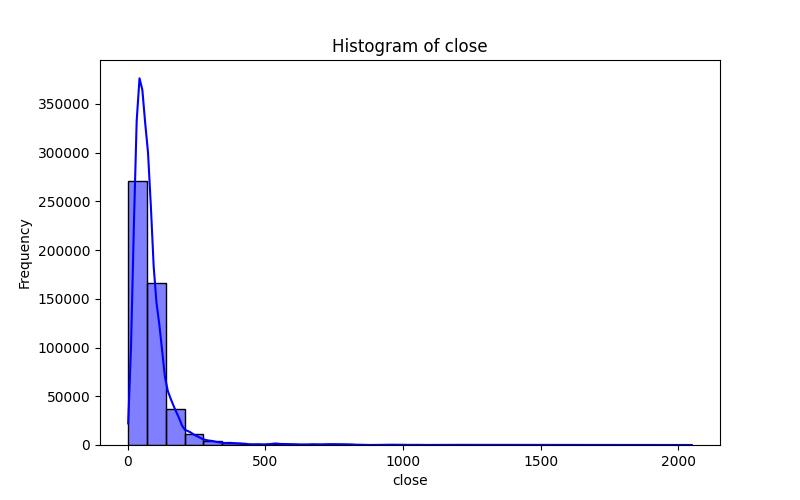
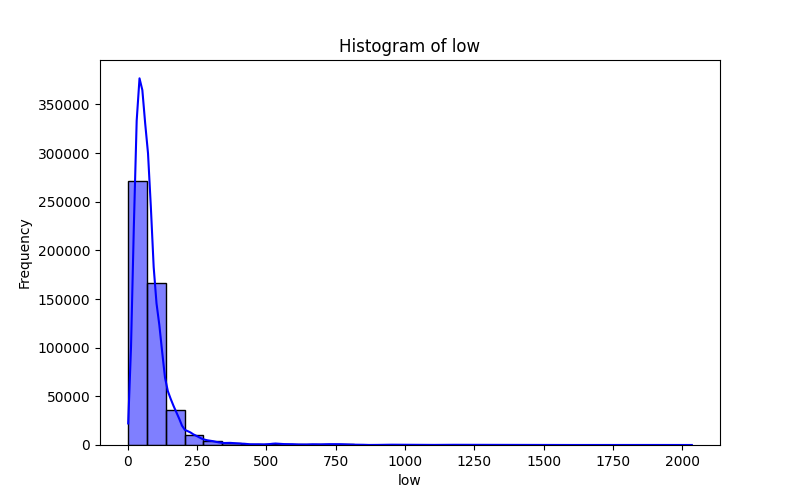
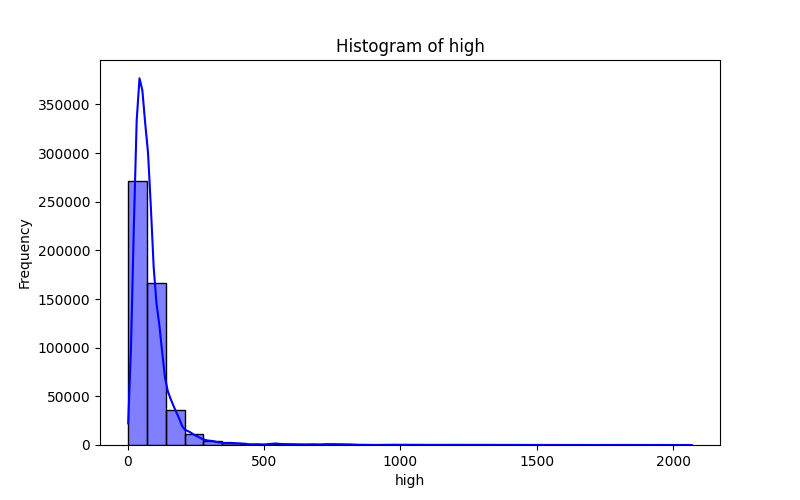
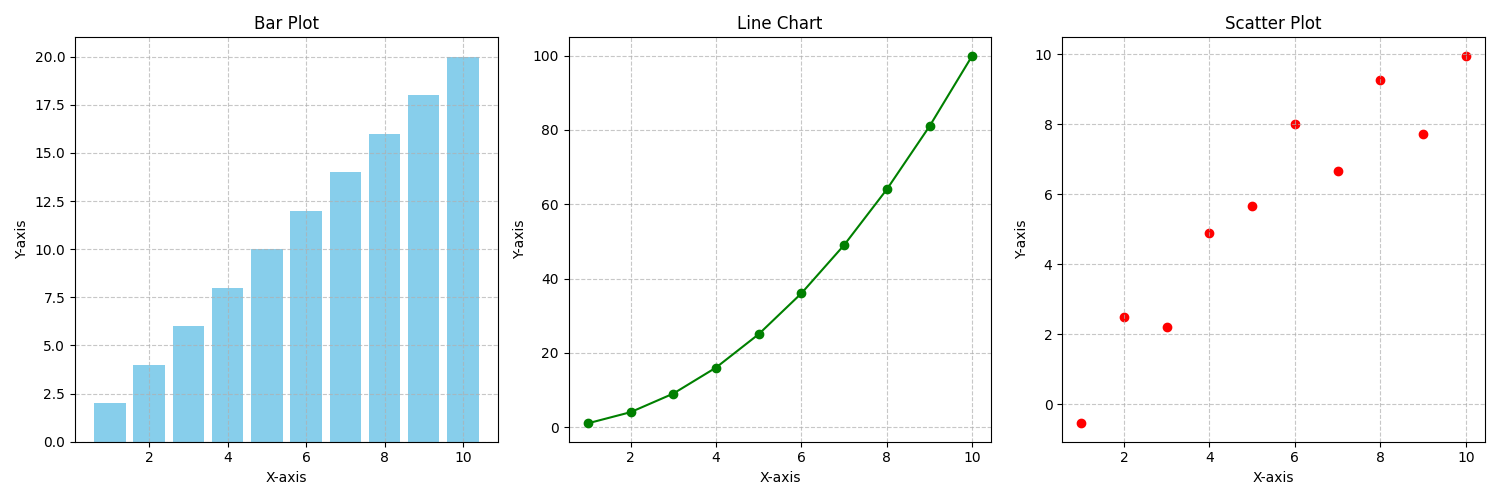
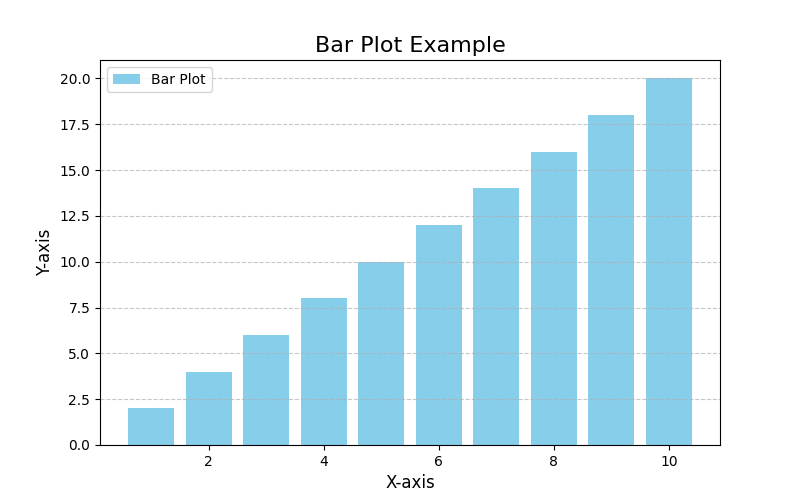
* Computed summary statistics: mean, median, standard deviation, etc.
* Generated histograms, boxplots, and scatter plots using matplotlib and seaborn.
* Identified feature correlations and potential outliers.
* Outcome: Key insights into data distribution and relationships.





**3. Basic Data Visualization**

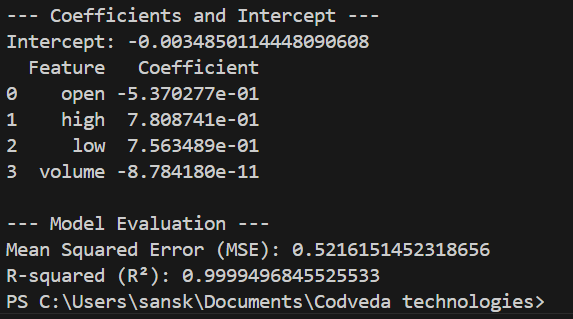
* Developed bar plots, line charts, and customized scatter plots.
* Labeled and styled graphs appropriately for reports.
* Exported plots as high-quality image files for documentation.
* Outcome: Effective visual summaries of data.



**✅ Level 2 – Intermediate Tasks**

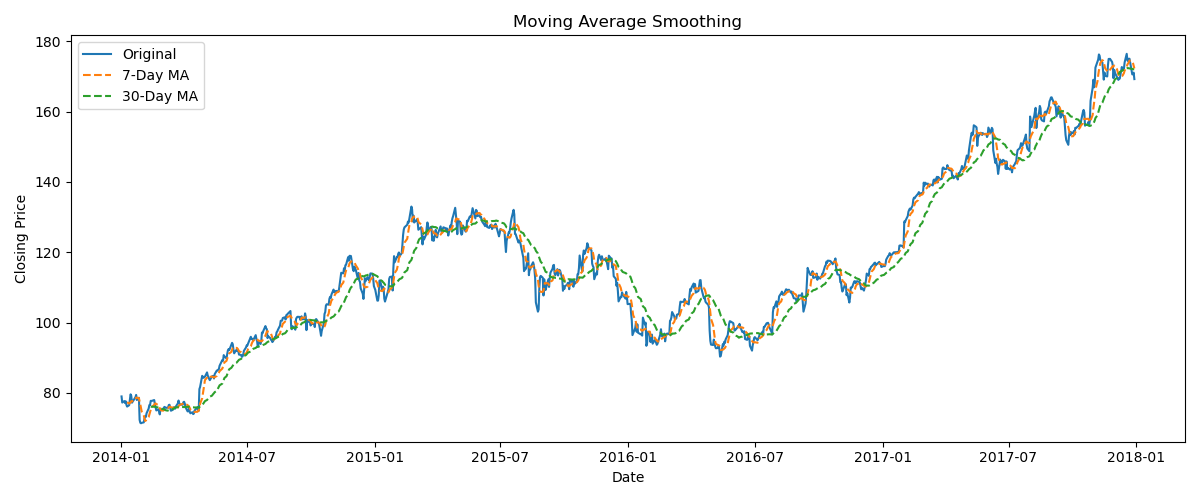
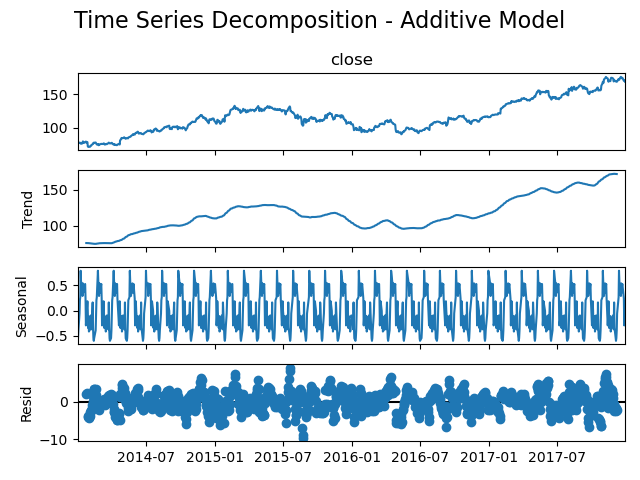
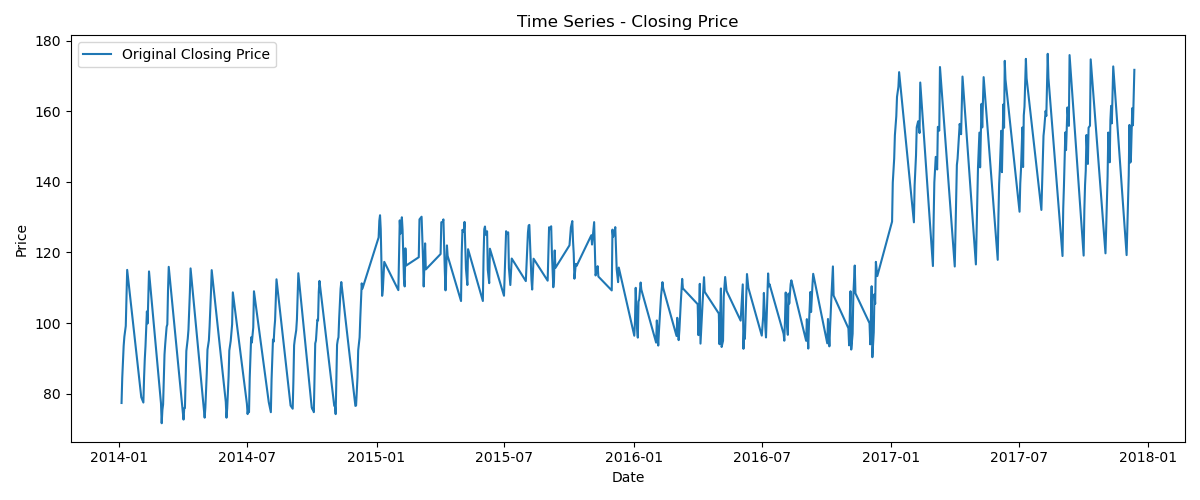
**1. Regression Analysis**

* Split datasets into training and testing sets using train\_test\_split.
* Applied linear regression with scikit-learn.
* Interpreted regression coefficients and validated model with R² and MSE.
* Outcome: A robust linear model with actionable insights.

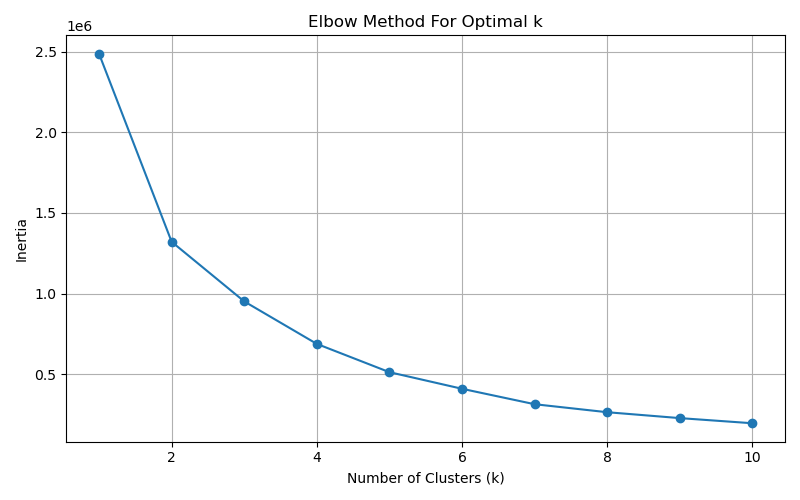
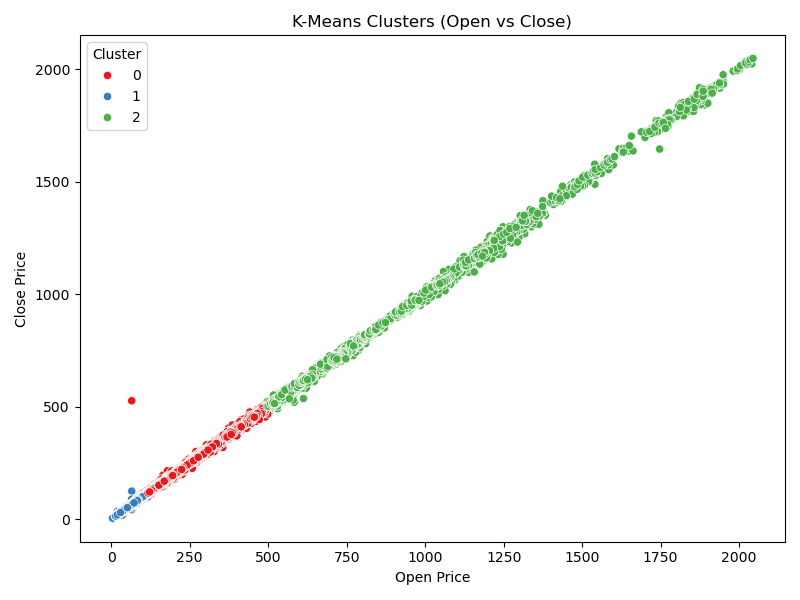


**2. Time Series Analysis**

* Imported time-series datasets (e.g., stock prices).
* Used decomposition to extract trend, seasonality, and noise.
* Applied smoothing techniques (e.g., moving averages) for clearer insights.
* Outcome: Detected meaningful patterns and forecast signals.



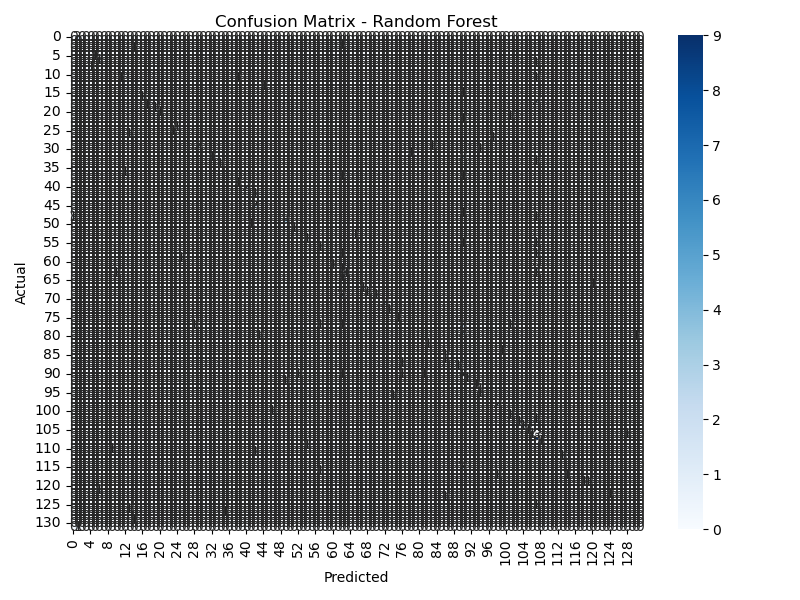
**3. Clustering Analysis (K-Means)**

* Standardized features using StandardScaler.
* Used the elbow method to determine optimal clusters.
* Visualized clusters and interpreted feature-based groupings.
* Outcome: Segmented data effectively for targeted strategies.

**✅ Level 3 – Advanced Tasks**

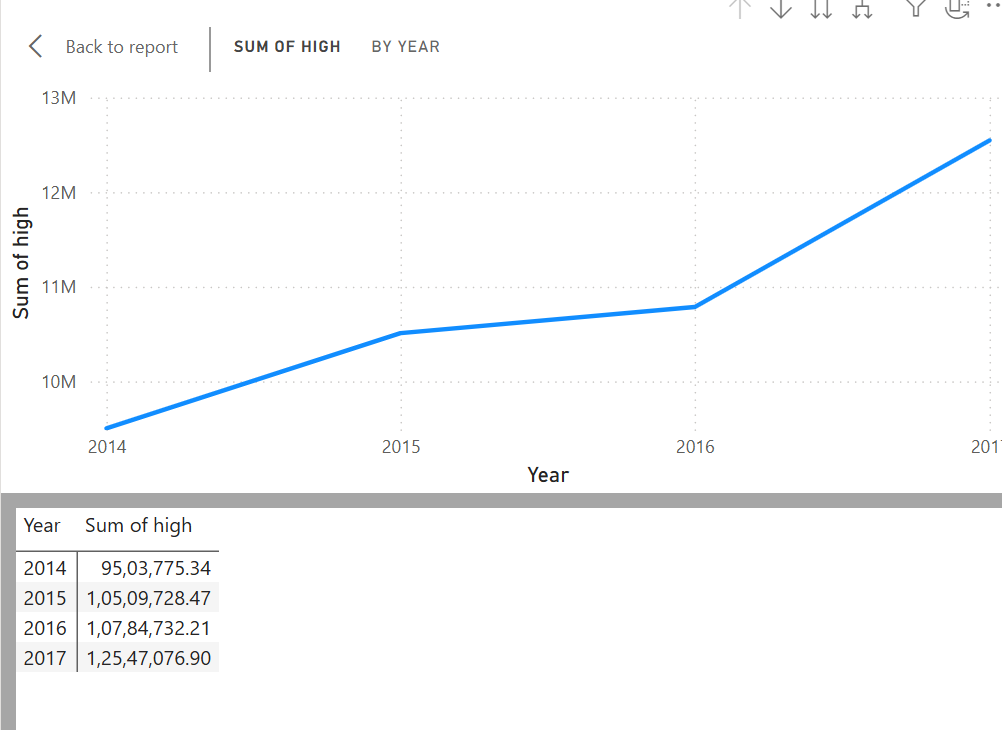
**1. Predictive Modeling (Classification)**

* Preprocessed data with feature encoding and normalization.
* Built multiple models (Decision Tree, Logistic Regression, Random Forest).
* Evaluated performance using accuracy, precision, recall, and F1-score.
* Conducted hyperparameter tuning with GridSearchCV.
* Outcome: High-performing classification model with validated metrics.



**2. Dashboard Building (Power BI/Tableau)**

* Cleaned data within Power BI and Tableau environments.
* Created interactive dashboards with slicers, filters, and drill-throughs.
* Visualized trends and KPIs effectively for business presentation.
* Outcome: Insightful, shareable dashboards for stakeholder engagement.



**3. Natural Language Processing – Sentiment Analysis**

* Processed textual data (tokenization, stopword removal, lemmatization).
* Applied sentiment scoring using TextBlob and NLTK.
* Created visualizations (word clouds, sentiment histograms).
* Outcome: Classified text into sentiment categories with interpretive visuals.

**🔹 Skills Gained**

* **Technical:** Data cleaning, EDA, regression, clustering, classification, NLP, dashboarding.
* **Tools Used:** Python, Pandas, Matplotlib, Seaborn, Scikit-learn, Statsmodels, Tableau/Power BI, NLTK, TextBlob.
* **Soft Skills:** Project documentation, video presentation, LinkedIn professional engagement.

**🔹 Conclusion**

This internship at Codveda Technology significantly enhanced my practical exposure to real-world data analytics tasks. The structured approach, combined with independent problem-solving, has strengthened both my technical proficiency and professional communication skills. I am confident that these experiences will support my future endeavors in the healthcare and analytics domains.