

Muhammad Faizan Tariq

During this project, my contributions focused on cleaning the data mainly handling missing values in the dataset, a critical step to ensure data integrity and analysis reliability. I began by identifying null values in the key columns (Open and Close) and standardized their representation by replacing them with NA. Using the **zoo** package, I applied the `na.approx()` function for linear interpolation, filling missing values based on adjacent entries. This approach preserved the dataset's continuity while minimizing potential biases. To ensure completeness, I re-verified the dataset after imputation, confirming that no null values remained. I also calculated and shared summary statistics for the cleaned data to provide insights into its updated distribution. These steps were essential for preparing the data for subsequent visualization and modelling.

My work complemented the efforts of the team. Muhammad Awais conducted detailed exploratory analysis, identifying key variables and addressing anomalies like zero trading volumes. Muhammad Ahmed ensured the environment setup and validated the dataset's structure, enabling a smooth workflow. Udyanraje Bhosale visualized the data by creating histograms and fitting an exponential decay curve to model the Close price distribution. Muhammad Hasnain explored the relationship between Open and Close prices through scatter plots, linear regression, and Spearman correlation, providing insights into variable dependencies. My contributions were foundational to ensuring data quality, enabling these subsequent analyses to yield meaningful results. Our discussions and collaborative efforts, captured in Git commits, were pivotal to the project's success.