


Data Collection and Preprocessing Phase

Date	03 October 2024
Team ID	LTVIP2024TMID24963
Project Title	Time Series Analysis For Bitcoin Price Prediction
Maximum Marks	2 Marks

Data Quality Report:

The Data Quality Report will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
Historical Bitcoin Price Dataset	Missing Values: Check for any missing or NaN values in the columns, especially in important fields like "Close" and "Adj Close."	Medium	Identify and fill missing values using interpolation or forward-fill techniques to maintain continuity in the time series.
Historical Bitcoin Price Dataset	Outliers: Verify for any extreme outliers in columns like "Open," "Close," or "Volume" that could skew predictions.	High	Use statistical techniques (e.g., IQR or Z-score) to detect and, if necessary, remove or cap outliers.
Historical Bitcoin Price Dataset	Inconsistent Date Range: Ensure all dates are present, without gaps, to avoid any break in continuity.	High	Interpolate missing dates with estimates or add placeholder rows with NaN values that can be filled through interpolation.
Historical Bitcoin Price Dataset	Data Formatting Issues: Currency symbols in columns like "Open," "High," "Low," etc., which may affect calculations.	Medium	Remove currency symbols and ensure all columns are in numeric format for accurate computations.
Historical Bitcoin Price Dataset	Volume Unit Standardization: Extremely large numbers in "Volume" may need verification for unit consistency. 	Medium	Confirm volume units (e.g., millions or raw values) and adjust as needed to ensure consistency across time periods.