

Data Science Assignment Report

Candidate: ds_Neeraj

❖ Executive Summary: Data Validation and Analysis

The objective of this assignment was to explore the relationship between trader behaviour and market sentiment using two provided datasets. A full data processing and analysis pipeline was built in Google Colab (ds_notebook.ipynb) as requested.

The first and most critical step of any data science workflow is data validation. Upon inspection, this analysis was **blocked by two critical data quality issues** in the provided files.

The accompanying notebook is robust, programmed to handle data sourcing challenges, and includes diagnostic checks that programmatically identify these data issues.

❖ Key Finding 1: Fatal Date Mismatch

The core analysis is impossible because the two datasets do not share any common dates.

- The **Historical Trader Data** (historical_data.csv) contains data from **late 2024 and 2025**.
- The **Bitcoin Market Sentiment Dataset** (fear_greed_index.csv) contains data exclusively from **2018**.

As there are **zero overlapping dates**, it is impossible to merge the datasets to correlate trading behaviour with market sentiment. The diagnostic block in ds_notebook.ipynb confirms this, showing 0 matched rows.

❖ Key Finding 2: Missing 'leverage' Column

The assignment overview explicitly lists 'leverage' as a key metric for analyzing risk.

- The historical_data.csv file **does not contain a 'leverage' column** or any obvious equivalent.

The analysis script was coded to robustly handle this missing data by skipping the leverage-based visualization to prevent errors.

❖ Data Sourcing and Workflow

The historical_data.csv file (80MB+) is too large for a standard GitHub upload. To solve this, the ds_notebook.ipynb script was coded to download the data *directly* from the Google Drive links using the gdown library. This is a robust and repeatable solution that bypasses file-size limits.

❖ Conclusion

The four required charts (PnL, Volume, Side, Leverage) were generated by the script and saved to the outputs/ folder per the instructions. However, as a direct result of the date mismatch, the underlying data is empty, and the charts are therefore blank.

The full analysis pipeline is built and ready. It will produce the required insights as soon as compatible datasets (i.e., from the same time period and including the 'leverage' column) are provided.