

**Data Visualization Assignment**

[support@intellipaat.com](mailto:support@intellipaat.com)

+91-7022374614

US: 1-800-216-8930 (Toll Free)

**Instructions for the Assignment**

1. Use a single file in Jupyter Notebook or Goggle collab for the entire assignment.
2. Download the Assignment file (File> Download > Download as .ipynb), compress it in a Zip folder and submit through LMS only.
3. Use the given dataset (**Stock\_File\_1.csv & Stock\_File\_2.txt**) for this Assignment.
4. Submit the Assignment by 7th February 2022 without fail.
5. This Assignment is an Exam kind and is considered for the evaluation and certification. No support is provided by our technical team.
6. If you have any doubts how to work on this assignment (No technical support), please drop an email to [support@intellipaat.com](mailto:support@intellipaat.com)

**FINAL ASSIGNMENT**

‘**Stock\_File\_1**’, a stock trend forecasting company has just employed you as a Data Scientist. As a first task in your new job, your manager has provided you with a company’s stock data and asked you to check the quality of the data for the next step of analysis. Following are the additional description and information about the data which your manager has shared with you.

* 1. The data set contains six variables namely-
     1. Date
     2. Open
     3. High
     4. Low
     5. Close
     6. Volume
  2. Typically, the stock market opens at 9:15 hours and closes at 15:30 hours. Each stock is defined by an opening price and a closing price which are the prices it opens and closes with. Within the operating hours, the stock price touches a maximum and minimum which are the highest and lowest prices achieved by the stock in the working hours of the stock market. You have access to ten years of monthly stock price data with the Open, High, Low and Close price and the number of stocks traded for each day given by the feature Volume. On some days when there is no trading, the parameters Open, High, Low and Close remain constant and Volume is zero.

Furthermore, your manager also claims that the model prediction is too bad since the data is polluted. Try to impress your new boss by preprocessing the data and by giving a proper rationale behind the steps you would follow. The two datasets should be merged before preprocessing.