1. How did you manipulate the data, and why? Illustrate your answer with plots.

* The data was rolled up into Monthly basis for Performing Exploratory Analysis.
* The Plot folder contains the Plots and Python Code for Data Manipulation

1. How did you perform NLP, if any?

NLP was Performed on Keyword Column in Dataset.

* Term Frequency (TF) was Calculated appearing in the Document
* Word Count for All the Individual Words were calculated appearing in the Document

NLP Folder in the attachment contains 2 Excel file along with Python Code. Details Mentioned below

* Term Frequency Excel file: Contains the Word and Term Frequency for all words appearing in the Document
* Word Count Excel file: Contains the Word and Count of all Words appearing in the Document in descending order
* Python Code

1. How did you model the problem, and why?

* Decision Tree Regressor Technique was used to model the problem.
* Decision Tree Regressor yielded a high R-Square value.
* higher value of R Square explains more variability in the dataset.
* Machine Learning Folder contains Python Code

1. How did you evaluate your model? What were the results of the evaluation?

To Evaluate our Model, the following Metrics were used

* R-Square
* Mean Absolute Error
* Root Mean Square Error

Results for Model Evaluation are described below

|  |  |
| --- | --- |
| Performance Metric | Value |
| R-Square | 0.82 |
| Mean Absolute Error | 1300.0 |
| Root Mean Square | 13827.0 |

1. If you had extra time, what would you do next?

I would use Neural Networks to improve R-Square value and minimize the errors