**Outlines**

**Sarah’s:**

1. **Two Options:**
   1. Initially we wanted to utilize both the Naive Bayes Model and the Decision Tree Model. However, after running into some user errors and time restraints, the Naive Bayes Model fell through. Positively though, the Decision Tree Model provided us just what was **needed!**
2. **Decision Tree:**
   1. With a 98% precision and accuracy outcome, machine learning could be used to predict the cheapest prices.
   2. Regarding the relationship between cost of materials and days of the week, there is a 75% precision and accuracy result.
   3. The number of returned queries increases around the weekends due to advertising for materials queried.
   4. Material price by vendor region and or location provides a 73% accuracy and a 69% precision.
   5. In addition, as we collected more data along the way, our accuracy and precision has increased **positively.**
3. **Beginning:**
   1. Our dataset included 27 vendors and 16 materials of interest.
   2. The bar graph details how many materials each vendor carries
   3. Home Depot, Lowes, eBay, and Amazon are the top four vendors that provide the majority of materials of interest.
   4. Both of these are important when considering which vendor is reliable and can help shorten time spent searching for either cheapest price or **materials.**
4. **Q 1 and 2:**
   1. In answering the questions posed at the beginning:
   2. Yes, there are correlations!
   3. In the image, the ordinal pairs (x and y values) illustrate the correlation. Left tells us what the data looks like.
   4. Right tells us if it’s correlated.
   5. The tinier the outcome, the more it’s **correlated!**
5. **Q 3:**
   1. In the visualization, Wednesday has the biggest decrease in material cost, while Sunday has the highest increase.
   2. This could be due to the restocking day being Saturday.
   3. We could definitely predict the day of increase or decrease, however, a regression model would be made instead of the logistic model shown.
   4. But, due to the time restraints, the visual depiction can only be **viewed.**
6. **Q 4:**
   1. Based on the image and matrix, yes there is a correlation between the region of vendor and theirprice.
   2. The southeast has an average price range, while the west has an average price range.
   3. This is good to know because most material comes from the southeast, assuring us we are getting a good **price.**
7. **Copper v. Steel**
   1. The daily fluctuations of copper and steel prices directly affect the materials of interest **price.**

**Alexei’s:**

As you can see we have been very thorough with our data, in doing so we have used:

* Jupyter Notebook, SQL, and pgAdmin were predominately used because these tools work with our analysis of the data, and happen to be common practice industrial wise.
* Instead of Plotly and Leaflet, we used NPM, Angular, Bootstrap, and SciPy, which were used to develop our dashboard, they helped streamline our work and are actually industry utilized.
* Lastly, Tableau is the most user-friendly data visualization program and GitHub is a version control for team project **integration.**

We were very thorough with our project, however due to time constraints, we would have liked to include future improvements, those being:

* The weight of materials and items being queried are something we would have liked to correlate to raw materials and if so, would have liked to query how much of a percentage is the total weight difference.
* Does being a large vendor mean we have a larger percentage of results, and could this be more viable opposed to a smaller company, as well as, how this data will be used depends on the circumstances.
* In addition, if we had a budget, a mega database like AWS could have been used to run our script through a website but it costs.
* A linear model expanding to a regression model would have provided us with data output for the next day instead of the next week
* Finally, using JSON could have potentially given us a more efficient **dataset.**