**Project Report**

**University Of New Brunswick Saint John**

**Dr. Kim**

**Cynthia Haque**

**4-23-2025**

**Explanation of the hypothesis:**

In the dataset I chose each item has a Price as well as a Unit of Issue (UI), which specifies how the item was sold (e.g. PK for ‘packaged’, RL for ‘rolled’, DZ for ‘dozen’). From this we can find the significant correlation between the two variables.

**Hypothesis:** The price of items has a correlation to the Unit of issue.

**Dataset:**

* [**National Stock Number Extract – Catalog**](https://catalog.data.gov/dataset/national-stock-number-extract)

**Explain the dataset:**

**A graph of a number of different colored bars

AI-generated content may be incorrect.**

**Fig.1**

**A screenshot of a phone

AI-generated content may be incorrect.A screenshot of a graph

AI-generated content may be incorrect.**

**Fig.2**

* Bar graph of the Average Prices by Unit of Issue (Fig.1) and a pivot table (Fig.2) to display the data more precisely
* From the pivot table, AY(Assembly) had the highest selling of issue with $4639.45 and the least selling of issue TU(Tube) with $1.06
* The goal: The readers or the audience can easily see the relationship between the two variables

**Descriptive analytics:**

**A graph of value and value

AI-generated content may be incorrect.**

**Fig.3**

* I calculated the p-values for each UI and displayed it with a distribution graph (Fig.3).
* If the p-value is less than 0.05, hence variables are correlated. Inconclusion the variables were less than 0.05, suggesting that there is a genuine difference in average prices between the UI categories being compared.

**Predictive analytics:** A graph with a line and a line

AI-generated content may be incorrect.

**Fig.4**

* I used linear regression to predict the Actual vs Predicted Price of an item. I concluded that the prediction is statistically significant as the r2 = 0.65, which suggest that there is a strong positive correlation.

**Discussion and further research:**

* It was useful to conduct a linear regression model to see that the UI and Price are correlated with one another.
* The Challenges I faced in this project was being able to find a dataset that was of my interest. Another is being able to pick the 2 predictive variables that can be used for my hypothesis. I faced the struggles of creating the ROC curve for my predictive model as I did not have binary values to compare with.
* For future works:
* Explore new features that could improve model accuracy such as additional product attributes (dimensions or weight), customer ratings or reviews
* Experiment with different model types to improve performance such as neural networks
* Consider handling outliers more effectively perhaps through robust regression techniques or by identifying and treating outliers as a separate category
* Consider importing external data sources such as market trends or economic indicators
* To conduct a thorough feature important analysis to better understand which features most influence the models predictions