Assignment – Student Data

Shaun Pritchard

Rasmussen College

QMB3200

Kimberly Powell

July,12 2020

**Assignment – Student Data**

**Describe & purpose of each data file:**

Upon examining the specific files, I came to some determinations about their validity and importance in predicting store sales based on historical data collected from these files. There were four files the train, test, stores, and features CSV files. Each file had its specific qualities that I will soon discuss.

I would like to discuss what important factors facilitate good datasets for predicting store sales and qualifying historical data. Categorizing data especially in historical context. Analyzing the number of units of each product being sold. Overtime can help determine if one product is being profitable While others are not. Sales by margins are data that can help determine which products are most profitable by profit margins. Cells by Price Point are similar to products at different prices and it's two similar products have different volumes and their distribution this could be something that can be capitalized on through optimization. To facilitate historical data that will prove to predict anything from store sales for profit. It is important not to overlook one of the most important data aspects any business should be collecting. And that is the data based on the customers purchasing specific products. Whether it's behaviors shopping patterns buying patterns all the correlating with sells patterns for feedback based on specific products he's a very important aspect of churning historical data into data that will yield profitable results.

**Missing data sets**

With that being said the features.csv Data file adds more data than any other file well there could be correlations between the temperature the dates any markdowns and prices of products that could have lied to the behavioral influence of consumers. Display all I had 12 data points that it collected. Some of the data points were unclear such as unemployment and CPI. So it seemed like they had felt for marked down products and fuel prices what kind of tribute to distribution. The *isHoliday* field with the historic dates could tell the business if specific product units where marginal. except for no specific product, data was gathered.

The test.csv data file could be attributed to the features.csv data file only to correlate the different stores and departments well it is riveting dates the stores and departments had holidays or no holidays. This kid is attributed to the other data sets with the features understanding the specific stores and departments from historical. I do not feel enough data was collected to be specific. Also, by the name of the file test, it would seem as though this data is being used for specific testing to use with training data in modeling algorithmic solutions.

The store.csv data file contribute a little bit more. Revealing the amount in different stores being 45 categorizing them by types A, B, and C while showing the capacity per each of the 45 stores. this data could be used in correlation with the features.csv file to determine the scale and volume of profit-based by margins I just got, locations, and specific holidays. though there is no specific product data historical data still could have been used for this purpose.

The train.csv data file had five dates what's a tribute to the date the holiday that apartment in the store, but it also showed the weekly cells and correlation with the store. And the features that CS we could use historical data to determine more viability of margin profit and sales that could be used for future predictions.

**Justification for how the additional data sets will help solve the problem:**

I believe that there was more specific data that could have been provided such as products themselves, the volumes distribution labels, margins based on profit, loss, manufacturing, and distribution. We would have had more details specifically based on the future forecast in predictive data that could yield valuable results. But most importantly if we had more consumer data feedback even negative feedback like products return cricket associates more analytical decisions for future predictions. We could have predicted future buying patterns and behaviors of customers with this type of data as well.