* Overall, write a coherent narrative that tells a story with the data as you complete this section.

I would say the growth of supermarkets in bigger cities and smaller are increasing and the competition in field is very stiff. I picked this data set because it met the requirements and I found it interesting. This dataset is show the complete sales of three different markets where the company recorded it for 3 months. So, essentially this data is for 3 branches for 3 months of data.

* Summarize the problem statement you addressed.

Being as this is a time series I wanted to find out what times / days/ months(along with other factors, tied in or solo) sales go up and what time sales go down. This will all be done using these variables in our data set: "Unit.price", "Total", "Rating", 'gross.income' , cogs(cost of goodsold.

* Summarize how you addressed this problem statement (the data used and the methodology employed).

How did I address this problem? First I imported the data set. Once it was imported then I changed the date time format to something more readable. After that I did some data cleaning making a subset of the entire columns and came out with these 5: "Unit.price", "Total", "Rating", 'gross.income'. Once that was completed I started out with a correlation chart that gave me some insight as to what different variables had some relationship with each other. See chart below.Chart, shape, bubble chart

Description automatically generated

After gathering insight from the correlation chart(keep in mind correlations aren’t super great in summarizing but it is a good starting point.) My first graph is for a variable called SalesByDay that I created from the Date column in the data set. I transformed it into a readable date for R. This showed the relationship between Date and Units. After messing around with correlation I wrote in a boxplot that shoes Branch X Rating and it looks like this:.Chart, diagram, box and whisker chart

Description automatically generated

Moving a long after grabbing more insight to me dataset I chose to do a graph of customers by gender and to see if they are embers or not members, and it looks like this:Chart, bar chart

Description automatically generated

Out of curiosity I also graphed the different types of payment which is unnecessary but still cool.Chart, bar chart

Description automatically generated

Next I wanted to see the total sales by branch and here is what I came up with:A screenshot of a cell phone

Description automatically generated

Next up we will look at the relationship between density and rating.

dA picture containing chart

Description automatically generated

Moving along to sales based on category :

Chart

Description automatically generated

* Summarize the interesting insights that your analysis provided.

Insights:

Brach B has to worst rating when comparing A, B and C. in regard to sales by the hour – most of this happened around 2pm. If you go to my boxplot you can see that Home and Lifestyle, Health and Beauty and electronics are outperforming food and beverages and fashion accessories. In the last graph I have shone you can see that fashion accessories is the top contributor and health and wellness are at the bottom. We can also see that Food and Beverage have the highest average rating where sports and travel are the lowest.

* Summarize the implications to the consumer (target audience) of your analysis.

This data can be used to predict when the slowest time at the super market is if the user prefers it empty or just uncomfortable around other people. It could also predict when it will be busy so that the company can bring on extra staff, like during holidays. . It could also predict sales by day/month for the companies to do what they want with that data as it might mean stocking up on particular items that are popular.

* Discuss the limitations of your analysis and how you, or someone else, could improve or build on it.

I think overall this is a pretty straight forward data set and it didn’t require much cleaning but to get back to the topic I would say people who are comfortable using Machine Learning techniques could run some really cool models on this data set which I did not do. I did some linear modeling and tons of graphs but people could just further in and come up with some cool conclusions.

* In addition, submit your completed Project using R Markdown or provide a link to where it can also be downloaded from and/or viewed.