**MIS581 Final Project Components**

FIN581

02/20/2022

Kevin Espinoza

Professor: Denheijer, Chris

Colorado State University- Global Campus

**Abstract**

In this paper we present an analysis of customer product reviews in an attempt gain a deeper understanding of how Amazon has been able to be successful throughout the years. This data was obtained from customer product review surveys in response to the items that they purchased. The survey offered the ability for the customer to rank their overall experience with the product and service they received, as well as a free-form area where they could provide a more detailed reasoning for their ranking. The results of this analysis offer critical insight into understanding the company's performance, what they are doing well, and the areas where they can improve. For feasibility reasons the data was filtered to include only the most popular product categories.

Unfortunately, because of data availability issues, the project faced limitations and was restricted to analyzing the surface of what the paper was intended to do. However, the data obtained helped the project to be meaningful and give insight into the company's performance. The findings showed that more than 90% of customers have a high satisfaction with the products and services that Amazon provides, with just 1% stating that they were dissatisfied. Overall, it appears that the company's strategies have been effective, however there were a considerable percentage of reviews that customers did not complete, indicating that this is an area which company will need to improve.

**Introduction**

A retail store’s primary focus is to find new ways to increase sales and one way this can be accomplished is by becoming more efficient on how to deliver marketing promotions to their customers. Toward this end, management can utilize data analytics to gain critical insight into each area of the business and uncover new trends. Gaining a comprehensive understanding of the business, including its position in the industry will allow them to create smart strategies focused on improving the business model by increasing performance through identifying adjustments that can be made to their operations and customer service processes. Further, this approach will help them reinforce and enhance some of the areas that customers enjoy about their services and products. While a successful analysis involves gathering the correct data, the first and most imperative step is for management to form a clear understanding of the company’s mission and establish well-defined goals that are consistent with their vision for the future. Creating a clear structure for what the business is looking to accomplish will provide guidance for what data will be required to reach these goals. Taking the time to create this foundation should produce a more accurate model that will be adjusted to reality, which in turn should produce more meaningful results and hopefully increase a company’s chance of success.

**Objectives**

In the case of Amazon, this analysis will focus on gaining insights related to their customer satisfaction level and how it is impacting the company’s sales. The data used in this project was generated by their customers as they provided rating reviews (1-5) on Amazon’s overall service level, with a rating of 1 as poor service, and 5 as excellent service. The data also contained a free form area where customers could leave detailed feedback, so this analysis will also scrutinize that data to create a list of frequently used words and then this information can be used to uncover any potential themes. This approach offers additional insight into the meaning of the customer’s rating. Also, the data related with the geographic locations is included in this dataset and this information can be utilized to potentially offer a better perspective on where Amazon’s customers are located. The data related with industry sales will also be used to gain an understanding for how well the company is positioned relative to their competition.

**Overview of Study**

The study will consist of understanding how Amazon's business model has been able to be successful and how the customers are likely to buy any products from the company again. In order to do that, it will be necessary to analyze performance variables and surveys from the customers. That will help to give meaningful insight and be able to reject the null hypothesis for this project.

Research Hypothesis

Business Question

* Do the customers like the merchandise?
* Which Zip codes have the most positive reviews?
* Is there any room for improvement as gleaned from the customer reviews?
* Is the company’s business model effective?
* Are there any trends uncovered through the customer reviews data?

1) Hypothesis

Null Hypothesis

Customers do not like the merchandise.

Alternative Hypothesis

Customers like the merchandise.

2) Hypothesis

Null Hypothesis

There is not a significant pattern in the customer reviews.

Alternative Hypothesis

There is a significant pattern in the customer reviews.

3) Hypothesis

Null Hypothesis

There is no room for improvement based on the customer reviews.

Alternative Hypothesis

There is room for improvement based on the customer reviews.

4) Hypothesis

Null Hypothesis

The business model is not effective.

Alternative Hypothesis

The business model is effective.

5) Hypothesis

Null Hypothesis

There is no trend found in the data.

Alternative Hypothesis

There is a significant trend found in the data.

**Literature Review**

All the sources used in this project are related to Amazon’s history and news reports related to their performance. These two types of information sources will support the model and the analysis that will be used for this project. The data will make more sense following different studies that were completed and different Amazon research papers. Also, it helps to create a structure that is focused on data security of the customers and the company. Also, the results will employ a variety of visualizations produced by different tools and attention will be given to formatting the content in such a way that it will appeal and be understandable to all potentially interested audiences. The data was gathered from Kaggle data set that are public set of data that can be used for research purposes.

**Research Design**

For this project I will be using histograms, box plots, summary statistics, correlation analysis, linear regression, and charts to produce and convey the results of the analysis. These techniques are going to be generated through SAS and Python. I will create an analysis that shows the different trends that were uncovered in the survey data using the specific keywords. Locating these patterns will facilitate a better understanding for exactly what the customer likes, and dislikes as associated with their rating. The rest of the information will offer a perspective for how the variables are behaving in the business, and how they are related to each other. The linear regression will give a perspective of what can happen in the future years and disclose potential future trends that can affect the company. The analysis will have different types of visualizations like histograms, boxplot and scatter plots that will help to give better insight of the data. Given the variety of analysis tools and visualizations utilized in this project, this analysis will require the transformation of some of the variables such as converting qualitative variables to become quantitative to provide a more meaningful analysis. This analysis will also incorporate broader economic data in the regression to provide perspective and robustness to the results. The variables will be from the perspective of time. Also, a creation of a dictionary will be involved in this project so the data can be separated by states through the use of a dictionary in python and this will give a more detailed understanding of the data and the trends in every state.

**Methodology**

The research will be using in two different methodologies a qualitative, quantitative approach. The data will give specific insight related to ratings and groups of data that will be created in order to better understand all the data that will be used in the model. Also, the use of the qualitative approach will allow a better understanding of the number in the model through the surveys and the description that were used by the customer. This information will give a better understanding and support of the trends that were found in the data. The use of these methodologies will give a better structure on the analysis and description of the data.

**Methods**

This analysis approach will allow the company to gain insight and continuously monitor some of the variables that drive sales in relation to the customers reviews. Customer reviews are critical for the company because they offer real information to the company on the mistakes that are happening throughout their operations process. With this insight, the company can work to create new strategies that address different types of issues in the operational side of the business and reduce the risk of causing a bad review that can have a long-term effect on the reputation of the company. Bad reviews can cause a downward trend in the steady loss of customers and revenue over time making this type of analysis critical to the company. Customer’s opinions have a big impact on the image of the company because not only do they drive the current results, but they also have a lingering effect that carries forward in time. One bad review can result in multiple lost sales, in effect they are a bad advertisement campaign that can lead to devastating results and good reviews can generate the exact opposite impact on the company’s results.

**Limitations**

Unfortunately, the analysis will be constrained by the limitations of the data. For example, Amazon protects customer's information so all demographic data is protected, thus available for use and the data related with the specifics of the customer profiles for every item purchased is limited. The information that will be used will be a general number that will be assigned by location to compare the sales and the reviews of the products for that area and that will potentially offer an additional perspective on why Amazon has been able to be successful.

Ethical considerations

Because of these issues, organizations that choose to use data analytics as a part of their business practices must determine the associated privacy and information security impacts before they actually employ analytics in their business. Establishing clear procedures, accountability and policies that limit access and use to only the right people can be done by creating comprehensive data privacy guidelines. For this project, the personal information will be removed to protect the names of the people that left a review and a number will be assigned as a representation for a particular group that represents their geographic location. This information offers a more general, yet accurate perspective on how the performance of the company is developing in the different states according to the location.

Obtaining a set of data is just one step in the process of conducting research. In order to create an accurate and credible model, it's necessary to establish methodologies because they will give structure to the paper and facilitate creating the hypothesis and objective. It is also important to understand the impact of using the data on whether it can harm the customers' privacy, because if a breach occurs, this can cause several negative issues to occur for the company. This paper will take into consideration all the internal, external, and limitations in the data. This will allow us to create a model that can offer meaningful insight that was uncovered after the analysis. Also, this will allow us to create a predictive model on how the data will behave in the future.

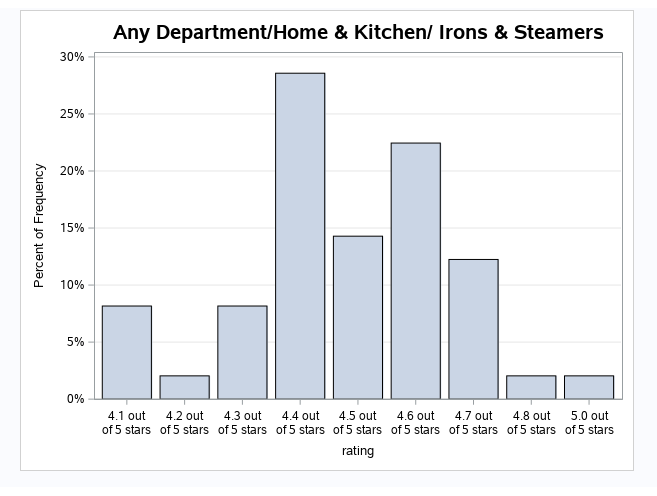
**Findings**

Figure # 1

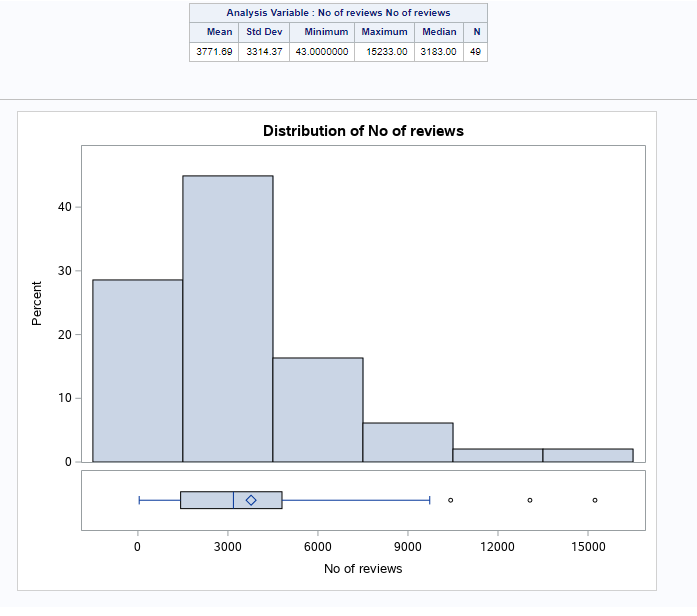
****

Figure # 2

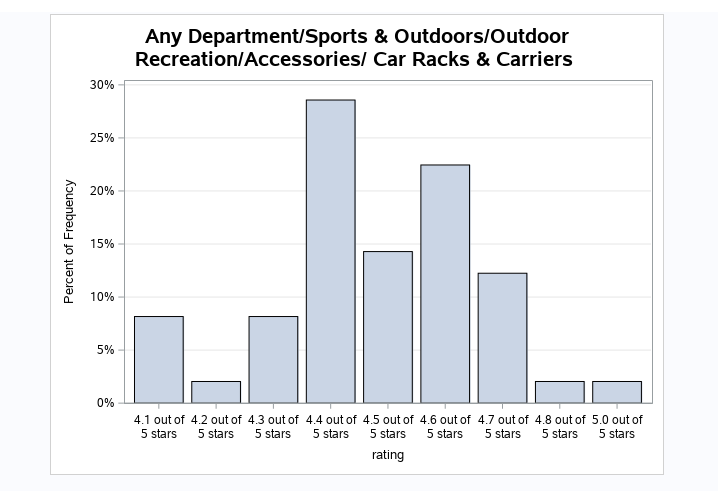


Figure # 3

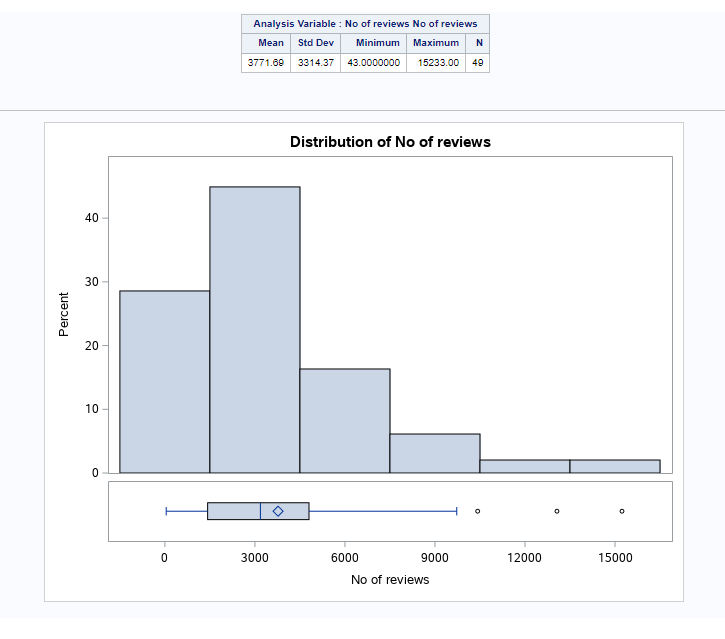
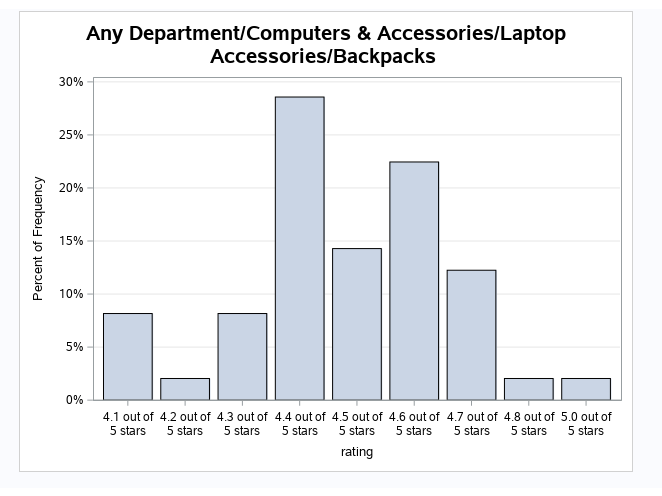


Figure # 4



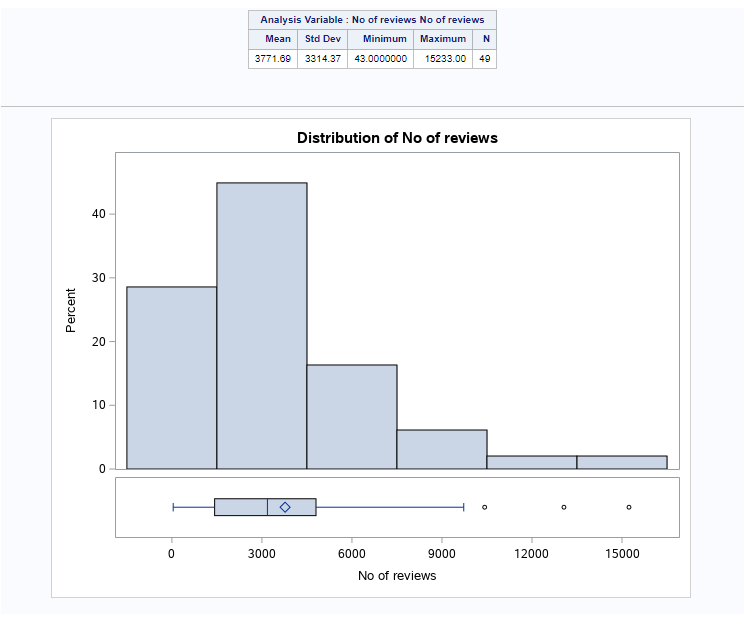
Figure # 5

Figure # 6

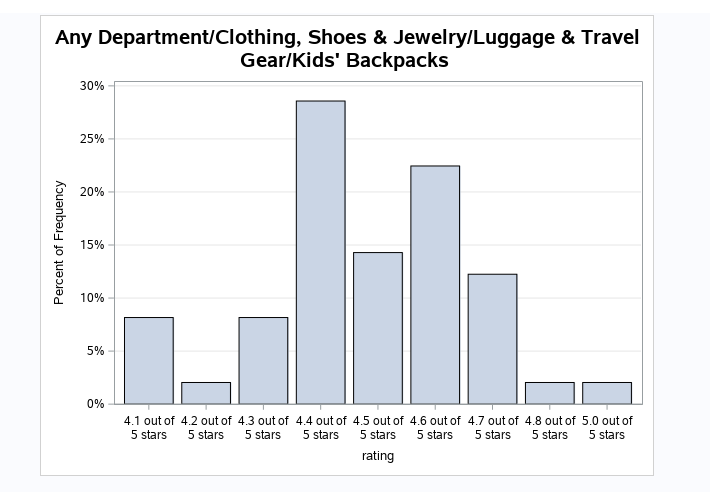


Figure # 7

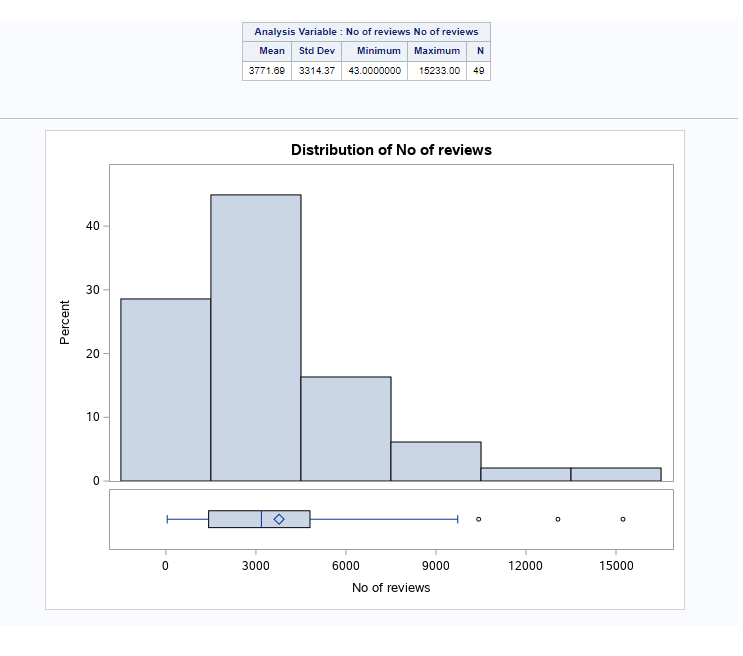


Figure # 8

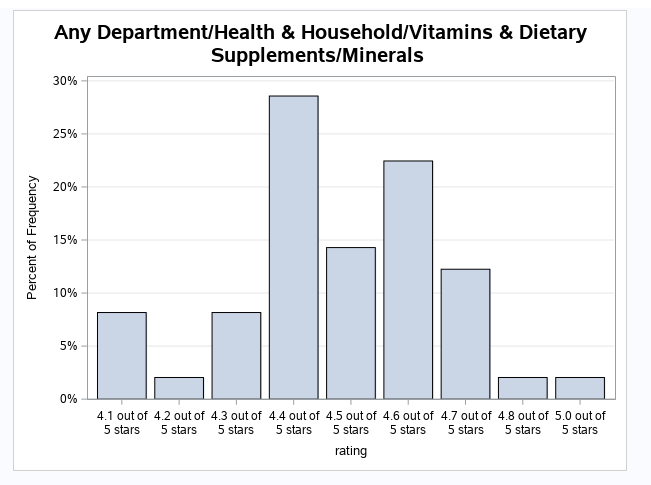


Figure 9

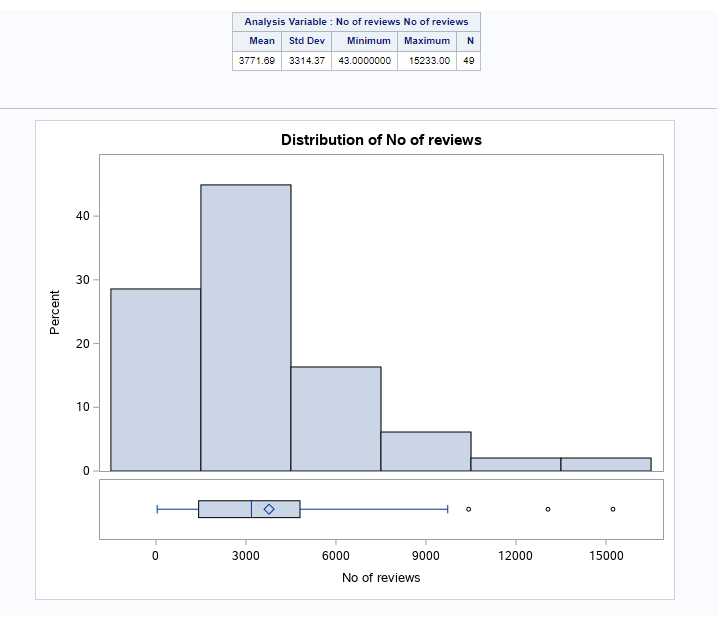


Figure # 10

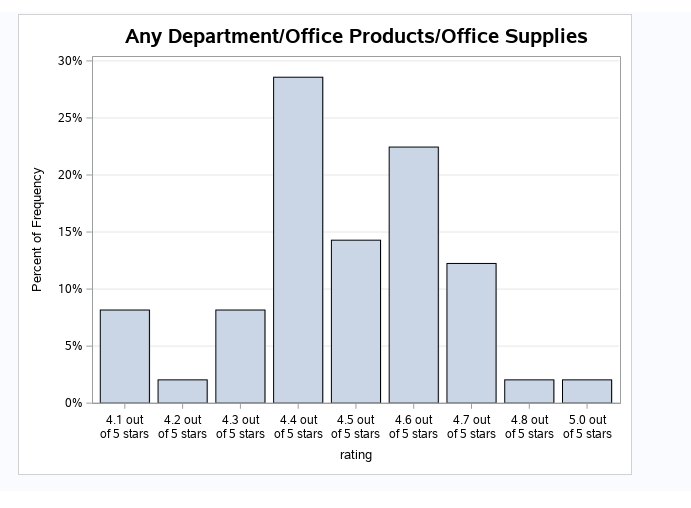


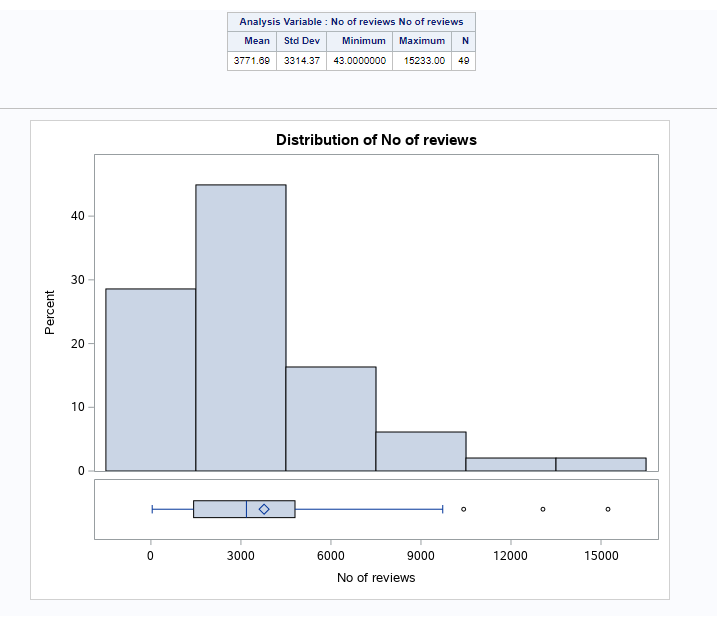
Figure # 11

Figure # 12

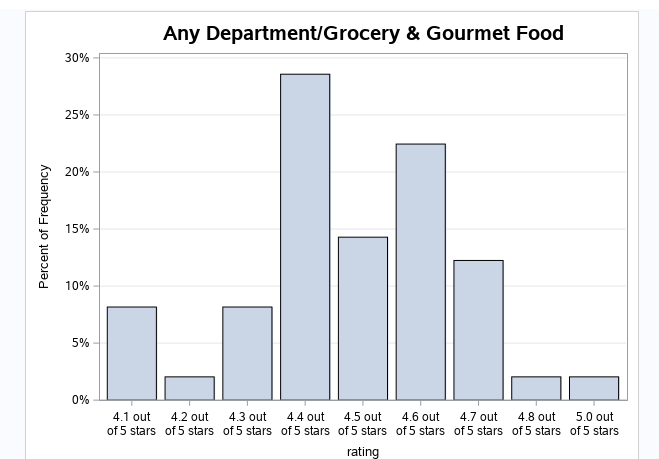


Figure # 13

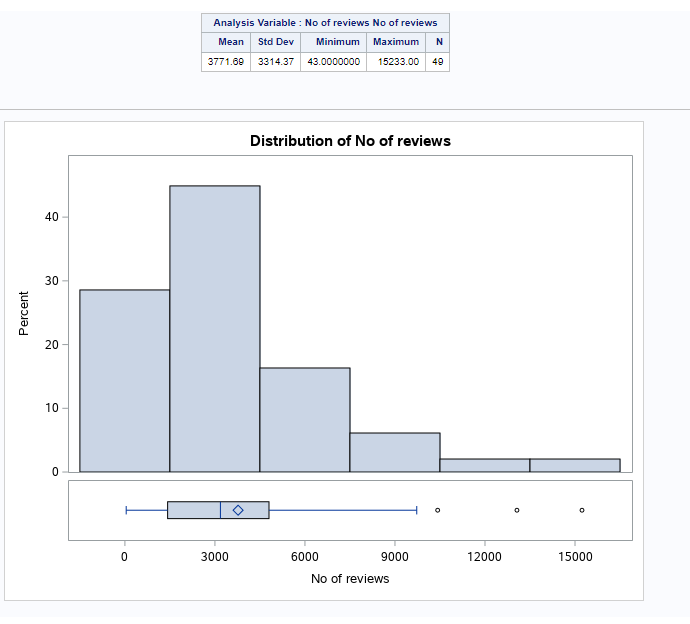
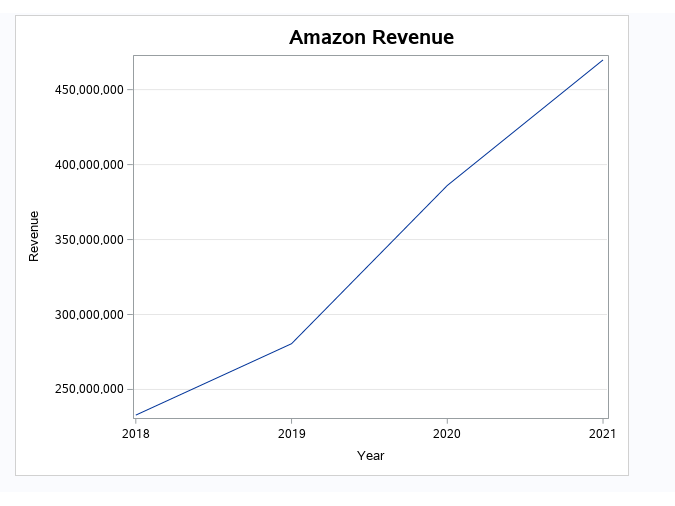


Figure # 14



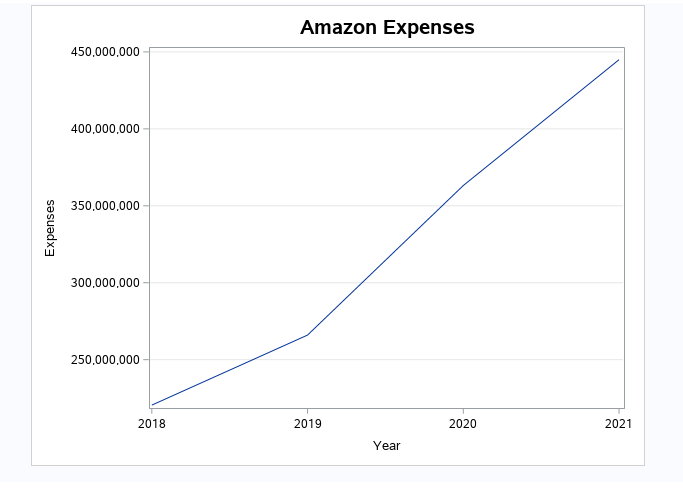
Figure # 15

Figure # 16

The data showed the most popular products that are sell in amazon, this information will give key insights that can help the company to understand the strengths and weaknesses of their service and products. The data showed that Amazon is been able to increase their revenues considerably the last four years. However, it’s been a considerable increase in the expenses also, that can be related with the expansion of the company in other areas and industries. The most popular products showed (Figure #1 – Figure # 14) that the majority of the products, customer gave the product positive review almost 80% of the product are between 4.3 – 4.7. Less than 1% in the sample of the for all the product where less than 4 and wasn’t take in consideration for this analysis. This information gives a good start to the company to determine if their business model is efficient. However, the data also showed that the range of products that don’t reviews is high, the mean (3771.69) showed that through all the products is a high number of reviews that weren’t filed, that is something that will need to be take in consideration.

In order to have a better understanding on why the company is been so successful in their business model, the description of the reviews and the use of specific words positive and negative will help the company to understand better. The following words were found as trend a meaningful in the reviews of the company.

Positive

* Best (3110)
* Good (1011)
* Great Price (527)
* Convenient (59)
* Love (2032)
* Recommend (679)

Negative

* Bad (244)
* Worst (6)
* Returned (400)
* Defective (51)

The data showed that customer have a positive reception of the product describing that they have a good experience and the like the product that they buy in most of the cases. Its is true that bad reviews weren’t included in the in the graphic showed, however the negative reviews are key to understand the areas that the company is able to improve. In this case the customer considered that were situation that the experience and the product was bad and they were situation that cause that returned items to the company. However, the number of bad reviews in the sample of the data weren’t considerable to think that are issued in their business model.

**Conclusion**

The analysis of the data showed that the company is been able to adapt and understand customer behavior and used their information to target customer with different ways to save the products. However, because the restrictions that exist in the data key information was missing that can give the company a more dipper understanding on the profile of the customers. More data will be necessary to create a model can predict future scenarios for the company. With data was used showed that customer like the quality of their products and also showed attachment to the product that they buy. That means that customer is more likely to go and buy again similar products or any product in Amazon. That will help to increase the main base of customer that buy in their platform more often. The use of this type of tools can help the company to create a better system that can give a better information to the company. Expanding their list with more key words will give more descriptive understanding of the data and the trends in the data. Overall, with analysis show that the company is been able to create a sustainable and convenience system that is been reflected in the sales of the company converting the business in the top seller of the industry.

**Recommendations**

It is true that the company is been successful with their system, however is necessary to find new tools and techniques that can allow the company to improve and reduce the no reviews that are not completed in the purchase of a product. Data changes every day and more data is collected. That means that ratings can change drastically if the no reviews increase considerably. That can cause a big impact to the revenues of the company, because they will not have a clear understanding of trends in the data. Also, it will be necessary to focus in the bad reviews and send apologize and give incentives to the customer. That will allow to keep the customer and reduce the risk of having bad advertisement of the company and keep a customer that will give a second opportunity to improve their experience.

**References**

Adam Bittlingmayer (2020). Amazon Reviews for Sentiment Analysis. Retrieved from Amazon

Reviews for Sentiment Analysis |Kaggle.

Capitalism.com Team (2020). History of Amazon: From Garage Startup to The Largest E-

Commerce Marketplace. Retrieved from History of Amazon: From Garage Startup to The

Largest E-Commerce Marketplace - Capitalism.com.

Guillermo Lafuente. Big data security, challenges and solutions. Source Big Data security –

challenges and solutions | F-Secure

Jedediah Purdy, Contributor (2016). The Amazon Economy. Retrieved from The Amazon

Economy HuffPost Impact.

Lydia DePillis (2018). It's Amazon's world. We just live in it. Retrieved from Amazon is

changing the US economy more than any other company – CNN.

Zina O′Leary. The Essential Guide to Doing Your Research Project. Source SAGE Publications

Ltd.