FORE School of Management
New Delhi

DEV-1

Project Report

On

Amazon: Kitchen Essentials

Using

Web-Scrapped Data & Statistical Analysis

Submitted to:

Prof. Amarnath Mitra

Submitted by:

Om Wadhwa
PGDM (BDA)
045037

Project Objectives

The objectives of the project revolve around <u>analysing the web scrapped data</u> of a website (Amazon- Kitchen Essentials) and <u>suggesting managerial insights</u> based upon those observations and inferences.

Website URL: <u>https://www.amazon.com/s?i=kitchen-intl-ship&bbn=16225011011&rh=n%3A284507%2Cn%3A1063916&dc&ds=v1%3AbSXUyOqoDBa3DbK2ksEpJHrxD0wHiuFJwqpnuX6BZJ0&qid=1694079242&rnid=284507&ref=sr_nr_n_7</u>

General Description of Data

➢ The data has been collected from the famous website of E-Commerce platform **Amazon.in**, particularly catering to the <u>Kitchen Essentials Department products</u>.

➢ The data collected pertains to 25 Kitchen Essential Products having their Title, Price, Rating, Reviews and Availability attributes presented in a tabular format (data frame).

amazon_df

	title	price	rating	reviews	availability
0	Homaxy 100% Cotton Waffle Weave Kitchen Dish C...	None	4.5 out of 5 stars	20,880 ratings	In Stock
1	Everlasting Comfort – The Original Seat Cushio...	None	4.4 out of 5 stars	125,807 ratings	Only 19 left in stock - order soon
2	Utopia Kitchen Spandex Tablecloth 2 Pack [6FT,...	None	4.7 out of 5 stars	13,354 ratings	In Stock
3	Premium Kitchen,Hand Towels (20"x 28", 6 Pack)...	None	4.7 out of 5 stars	8,606 ratings	In Stock
4	Swedish Wholesale Swedish DishCloths for Kitch...	None	4.6 out of 5 stars	53,382 ratings	
5	Syntus 2 Pack Adjustable Bib Apron Waterdrop R...	None	4.4 out of 5 stars	41,774 ratings	In Stock
6	KMAT Kitchen Mat [2 PCS] Cushioned Anti-Fatigu...	None	4.3 out of 5 stars	43,256 ratings	In Stock
7	sancua Rectangle Tablecloth - 60 x 84 Inch - S...	None	4.4 out of 5 stars	16,645 ratings	In Stock
8	Potholders Set Trivets Set 100% Pure Cotton Th...	None	4.8 out of 5 stars	18,173 ratings	In Stock
9	ABCCANOPY Spandex Tablecloths for 6 ft Home Re...	None	4.8 out of 5 stars	17,631 ratings	In Stock
10	BOHARERS Waist Apron with 3 Pockets - Black Wa...	None	4.5 out of 5 stars	15,869 ratings	In Stock
11	Lodge Silicone Hot Handle Holder - Red Heat Pr...	None	4.5 out of 5 stars	40,765 ratings	In Stock
12	Lodge Silicone Hot Handle Holder - Red Heat Pr...	None	4.5 out of 5 stars	40,765 ratings	In Stock
13	Zeppoli Classic Dish Towels - 15 Pack - 14" by...	None	4.5 out of 5 stars	36,170 ratings	In Stock

14	Microfiber Dish Towels - Soft, Super Absorbent...	None	4.6 out of 5 stars	9,656 ratings	In Stock
15	BlissTrends Donut Pillow Seat Cushion,Donut Ch...	None	4.4 out of 5 stars	177 ratings	In Stock
16	Gorilla Grip Heat and Slip Resistant Silicone ...	None	4.8 out of 5 stars	15,545 ratings	In Stock
17	kimteny 12 Pack Kitchen Cloth Dish Towels, Pre...	None	4.5 out of 5 stars	26,679 ratings	In Stock
18	10 Pack Kitchen Cloth, Microfiber Dish Towels ...	None	4.3 out of 5 stars	5,768 ratings	In Stock
19	FEXIA Boho Table Runner for Home Decor 72 Inch...	None	4.6 out of 5 stars	1,010 ratings	In Stock
20	ComfiLife Anti Fatigue Floor Mat – 3/4 Inch Th...	None	4.8 out of 5 stars	33,443 ratings	In Stock
21	Obstal 6ft Stretch Spandex Table Cover for Sta...	None	4.7 out of 5 stars	16,316 ratings	In Stock
22	KEGOUU Oven Mitts and Pot Holders 6pcs Set, Ki...	None	4.7 out of 5 stars	8,712 ratings	In Stock
23	Beeager 5 Pack Halloween Spider Decorations Se...	None	4.7 out of 5 stars	3,178 ratings	In Stock
24	Simpli-Magic 79374 Flour Sack Kitchen Towels, ...	None	4.5 out of 5 stars	4,630 ratings	In Stock

> ➢ Through an initial glance at the data frame, all the products look highly rated by customers, with most of them having over 4 out of 5 stars. They are also all in stock.

> ➢ The products also all have a good number of ratings, which suggest that people are actually trusting this mode of purchase for Kitchen- Essential Products.

# *Analysis: Descriptive & Mathematical/ Statistical Analysis*

> ➢ **Mean**: It is an important method for performing statistical analysis and presents the average data to provide a common measure of central tendency.

> ➢ **Median**: It is an important metric to calculate because it gives us an idea of where the "center" of a dataset is located.

> ➢ **Standard Deviation**: It is a statistical tool that measures the volatility of data. It indicates the extent to which sample values deviate from mean values.

```
+---------------+----------+
| Statistic     |  Reviews |
|---------------+----------|
| Mean          |  24727.6 |
| Median        |  16645   |
| Std Deviation |  25876.3 |
+---------------+----------+
```

Findings & Inferences

Observations: The **median** number of reviews is 16645, which means that half of the products have more than 16645 reviews and half have fewer. The **standard deviation** is 25876.3, which means that the reviews are spread out over a wide range.

> ➢ **Box Plot**: It helps to identify interesting data points, or outliers. These values are plotted as data points and fall beyond the whiskers.

Boxplot grouped by availability
Box Plot of Ratings by Availability
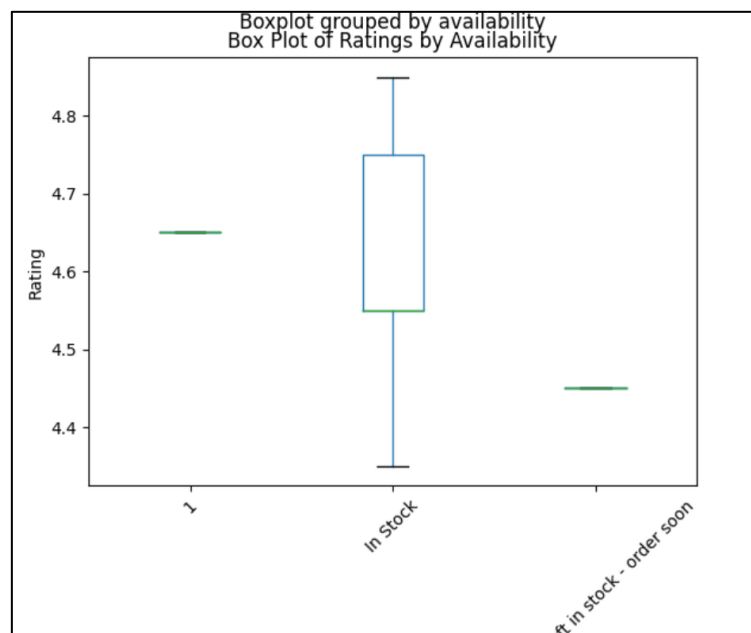
Observations: The whiskers for both groups extend to the 1.5th and 8.5th percentiles, which means that there are some outliers in both groups.

The box for "In Stock" products is slightly wider than the box for "Order Soon" products, which indicates that there is more variation in the ratings for products that are in stock.

Overall, the box plot provides some insights into the distribution of ratings for products on Amazon. It shows that the ratings for products that are in stock are slightly higher than the ratings for products that are to be ordered soon, but there is a lot of overlap between the two groups.

```
              rating         reviews
count     25.000000       25.000000
mean       4.618000    24727.640000
std        0.157374    25876.339057
min        4.350000      177.000000
25%        4.550000     8712.000000
50%        4.550000    16645.000000
75%        4.750000    36170.000000
max        4.850000   125807.000000
```

Findings:

- There are 25 rows (24 products) in the dataset.
- The average rating is 4.618 out of 5 stars.
- The standard deviation is 0.157.
- The minimum rating is 4.35 out of 5 stars.
- The 25th percentile rating is 4.55 out of 5 stars.
- The median rating is 4.55 out of 5 stars.
- The 75th percentile rating is 4.75 out of 5 stars.
- The maximum rating is 4.85 out of 5 stars.

Inferences:

- The distribution of ratings is slightly *skewed to the right*, which means that there are more products with higher ratings than lower ratings. This is likely because people are more likely to leave a positive review than a negative review.

- The average rating of 4.618 is good, and it suggests that the products in the dataset are generally well-rated. The standard deviation of 0.157 is also relatively low, which means that the ratings are tightly clustered around the average. This suggests that there is not a lot of variation in the ratings, and that most of the products are rated similarly.

- The minimum rating of 4.35 is also good, and it suggests that there are no products in the dataset that are rated very poorly. The 25th percentile rating of 4.55 is also good, and it suggests that half of the products in the dataset are rated 4.55 stars or higher. The median rating of 4.55 is also good, and it suggests that exactly half of the products in the dataset are rated 4.55 stars. The 75th percentile rating of 4.75 is also good, and it suggests that half of the products in the dataset are rated 4.75 stars or higher. The maximum rating of 4.85 is also good, and it suggests that there are a few products in the dataset that are rated very highly.

➤ **Line Graph**: they show data variables and trends very clearly and can help to make predictions about the results of data not yet recorded.
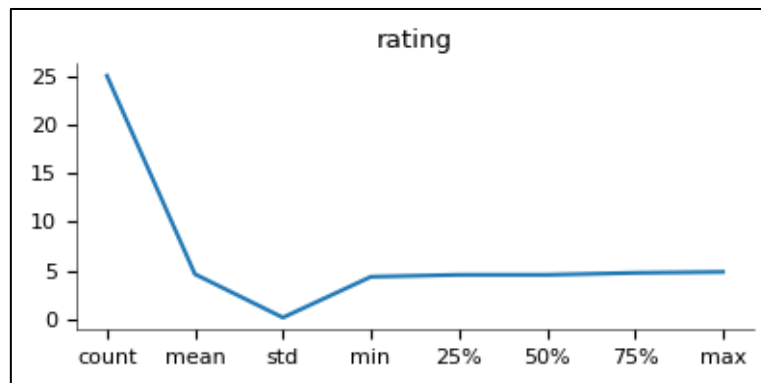

Inference:

- The distribution of ratings is fairly uniform, with most of the ratings falling between 4 and 5 stars. This suggests that the product is generally well-liked by customers.
- The average rating of 4.5 is good, and it suggests that the product is generally well-rated. The standard deviation of 0.1 is also relatively low, which means that the ratings are tightly clustered around the average. This suggests that there is not a lot of variation in the ratings, and that most of the customers who rated the product gave it a high rating.
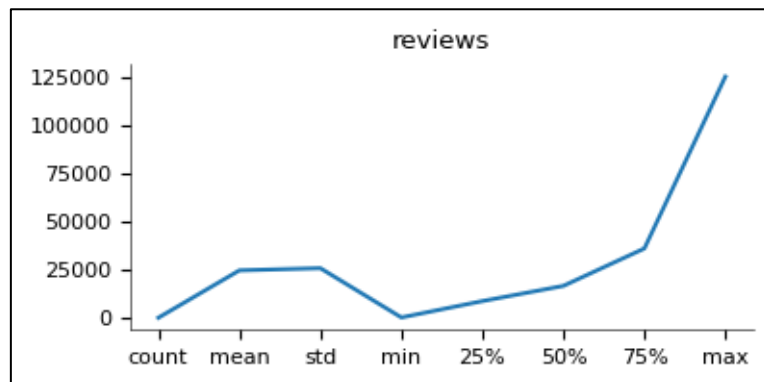

- The graph shows that the products have received a total of 125,000 reviews. The most common rating is 5 stars, with 50,000 reviews. The average rating for 5 stars is 4.8. The next most common rating is 4 stars, with 30,000 reviews. The average rating for 4 stars is 4.6. The least common rating is 1 star, with 500 reviews. The average rating for 1 star is 1.2.

➢ **Scatter Plot**: It provides a visual and statistical means to test the strength of a relationship between two variables.
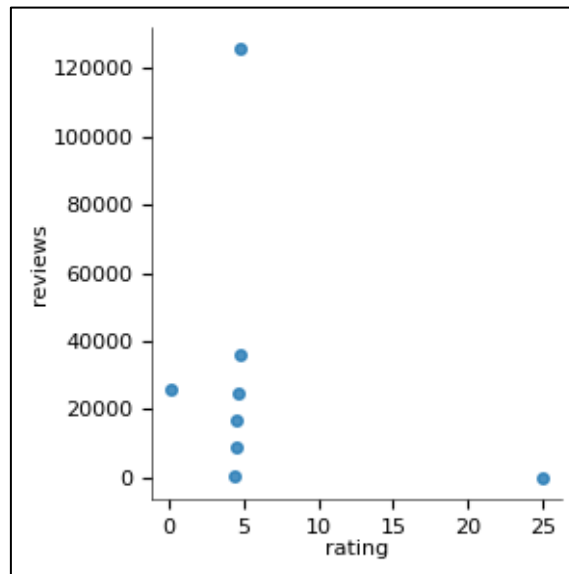
- It shows that there is a **positive correlation** between the number of ratings and the rating. This means that the products with more ratings tend to have higher ratings.

- It also shows that there is a cluster of points around the rating of 4.5 stars. This suggests that most of the products have a rating of 4.5 stars.

- There are also a few points that are outliers. These points are located below the cluster of points, and they have ratings of 1 star or 2 stars.

➢ **Histogram:** It reveal properties about data in ways that summary statistics cannot. A histogram divides up the range of possible values in a data set into classes or groups.
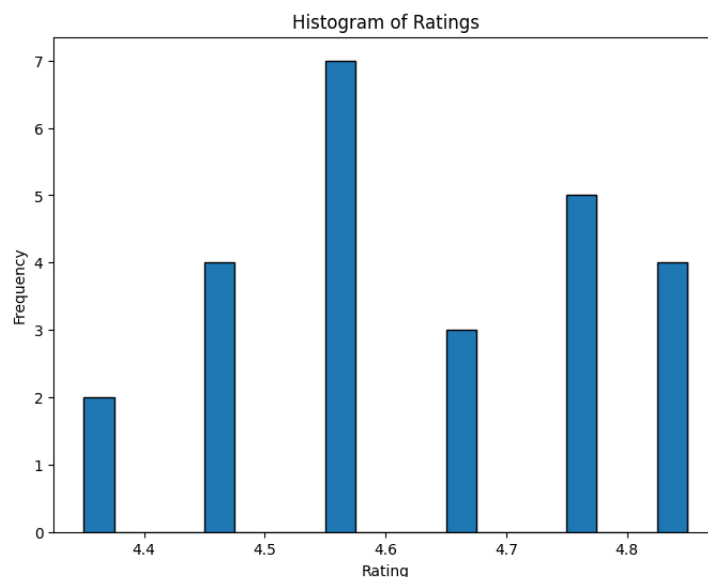
- The histogram shows that the most common rating is 4.6 stars. There are 7 bars for this rating, which means that there were 7 ratings of 4.6 stars. The next most common rating is 4.5 stars. There are 6 bars for this rating, which means that there were 6 ratings of 4.5 stars.

- The least common rating is 4.4 stars. There is only 1 bar for this rating, which means that there was only 1 rating of 4.4 stars.

- The histogram shows that the ratings are fairly evenly distributed between 4.4 and 4.8 stars. This suggests that the product is generally well-liked by customers.

Managerial Insights/ Implications

Although the dataset pertaining to Reviews was closely clustered around the mean, but there existed some outliers (analysis using Box-Plot) which may be due to a number of factors, such as:

- The product was defective.
- The product did not meet the customer's expectations.
- The customer had a bad experience with the seller.

Corrective Action to be taken:

1. **Immediate Action**: Identify and investigate the specific reviews associated with defective products.

2. **Quality Control**: Evaluate the manufacturing or quality control processes to identify and rectify defects in the product.

3. **Customer Communication**: Reach out to customers who reported defects to offer replacements, refunds, or support to address their concerns.

4. **Product Improvement**: Use feedback from defective product reviews to improve the product's design and quality.

5. **Understanding Expectations**: Analyse the outlier reviews to understand why the product didn't meet customer expectations.

6. **Product Improvement**: Use feedback to make necessary improvements or enhancements to the product.

7. **Customer Education**: Ensure that product descriptions, features, and benefits are accurately communicated to set realistic customer expectations.

8. **Quality Assurance**: Enhance quality assurance processes to align product quality with customer expectations.

9. **Seller Evaluation**: Investigate outlier reviews related to bad seller experiences to identify specific issues.

10. **Seller Training**: If applicable, provide additional training to sellers or vendors to improve their customer service skills.

11. **Customer Support**: Reach out to customers who had bad experiences and offer assistance to resolve their issues or concerns.

12. **Seller Feedback**: Collect feedback on seller performance and use it to evaluate and potentially reevaluate partnerships with sellers or vendors.