

# **Analysis on Amazon Product Reviews using the SQL Dataset located on Azure**

Nishigandha Wankhade

## Introduction

In this project, we analyze customer reviews from Amazon to gain deeper insights into customer sentiment, review patterns, and overall product ratings. The analysis is performed using a SQL dataset located on Azure, leveraging various SQL queries and analytical techniques to uncover meaningful trends and insights.

### Key Objectives of the Project:

- Understanding the distribution of customer ratings and sentiments.
- Identifying the length and excitement levels of customer reviews.
- Analyzing the relationship between numerical ratings and sentiment scores.
- Detecting confusing or contradictory reviews.
- Applying advanced window functions for deeper analytical insights.

### Data Sources:

The data used for this project is sourced from Amazon's product review datasets, which include the following key columns:

- Review Text: Detailed customer feedback.
- Overall Rating: Numerical ratings (1-5) given by customers.
- Sentiment: Preprocessed sentiment labels (Positive, Neutral, Negative).
- Sentiment Score: Numeric representation of the sentiment.
- Word Count, Review Length, Exclamation Count: Text-based metrics for each review.

This project aims to provide a comprehensive understanding of customer feedback, helping businesses improve their products and customer experiences through data-driven insights.

### **\*\* Previewing the Dataset\*\***

```
SELECT TOP 5 * FROM [dbo].[amazon_reviews];
```

reviewText	overall	category	summary	sentiment	sentiment_score	review_length	word_count	exclamation_count
As advertised. Reasonably priced	5	All_Beauty_5	Five Stars	positive	0.2	32	4	0
Like the odor and the feel whe...	5	All_Beauty_5	Good for the face	negative	-0.028124999999999999	207	44	0
I bought this to smell nice after...	1	All_Beauty_5	Smells awful	negative	-0.06666666666666667	142	30	0
HEY!! I am an Aqua Velva Man ...	5	All_Beauty_5	Truth is There IS Nothing Like a...	positive	0.21740646258503404	1417	266	2
If you ever want to feel pamper...	5	All_Beauty_5	Bulgari Shampoo	positive	0.25	237	45	1

- **Purpose:** Retrieve the first 5 rows from the **amazon\_reviews** table to get a quick overview of the dataset structure and sample data.
- **Insight:** Useful for understanding the schema and columns available for analysis.

### **1. To find Overall Rating Distribution**

```
SELECT overall, COUNT(*) AS review_count
FROM [dbo].[amazon_reviews]
GROUP BY overall
ORDER BY overall DESC;
```

overall	review_count
5	3675
4	272
3	90
2	52
1	92

- **Purpose:** Count the number of reviews for each overall rating (typically from 1 to 5).

- **Insight:** Helps visualize the distribution of customer satisfaction and identify potential biases in the ratings.

## 2. To find how many Positive / Negative reviews are overall? (Sentiment Distribution)

```
SELECT sentiment, COUNT(*) AS review_count  
FROM [dbo].[amazon_reviews]  
GROUP BY sentiment  
ORDER BY review_count DESC;
```

sentiment	review_count
positive	3767
neutral	243
negative	171

- **Purpose:** Count the number of positive, neutral, and negative reviews based on sentiment analysis.
- **Insight:** Useful for understanding the overall customer sentiment towards the product.

## 3. To Find Top 10 Longest Reviews

```
SELECT TOP 10 reviewText, word_count  
FROM [dbo].[amazon_reviews]  
ORDER BY word_count DESC;
```

- **Purpose:** Identify the 10 reviews with the highest word count.
- **Insight:** These reviews might provide more detailed feedback, potentially capturing a broader range of customer experiences.

reviewText	word_count
I **LOVE** my Essies. They are by far my favorite nail polish brand (and one of only two brands I will buy--only others are impossible on my nails). Even t...	962
Overall, this is an amazing value and an awesome kit. The toothbrush is just wonderful and comes with plenty of heads for different cleaning/issues.. The...	692
I got this in a beauty subscription box. I've had it for so long I forget which box. I needed a new exfoliator so I grabbed this from my stash. Here are the i...	640
This is an average-sized bar of scented French-milled, French-made soap designed to over a more luxurious experience than consumer-grade soaps. I wi...	551
I've always prided myself on my good teeth. I thought that flossing and brushing was enough...that was until I went to the dentist, and he said I had the ...	429
I've always prided myself on my good teeth. I thought that flossing and brushing was enough...that was until I went to the dentist, and he said I had the ...	429
I love Pre de Provence bar soaps, and I love sandalwood. By reading lots of reviews, I notice that many people do not like potent scents, and prefer light,...	405
This is probably the "smoothest" tasting mouthwash I have ever used. It is a subtle mint flavor, yet still leaves your mouth feeling fresh. "Smooth" fits the...	396
This is probably the "smoothest" tasting mouthwash I have ever used. It is a subtle mint flavor, yet still leaves your mouth feeling fresh. "Smooth" fits the...	396
I have this in "Berry Seductive" but I've had a great experience with it and would certainly purchase it in other colors. As you can see from the product im...	393

#### 4. Find the Top 10 Most Excited Reviews by Exclamation Marks

```
SELECT TOP 10 reviewText, exclamation_count
FROM [dbo].[amazon_reviews]
ORDER BY exclamation_count DESC;
```

reviewText	exclamation_count
Williams Lectirc Shave is a great product that has served men's electric shaving needs for several decades! BUT DON'T buy it here! WHY? I didn't pay clo...	20
Everything about Kerastase is fantastic!! The Chrome Riche line promotes healthy shiny hair. Love it!! I also buy Kerastase Fusio-Dose Systeme Profession...	9
Everything about Kerastase is fantastic!! The Chrome Riche line promotes healthy shiny hair. Love it!! I also buy Kerastase Fusio-Dose Systeme Profession...	9
Everything about Kerastase is fantastic!! The Chrome Riche line promotes healthy shiny hair. Love it!! I also buy Kerastase Fusio-Dose Systeme Profession...	9
All time favorite!!!! Wish they still carried this!!!	7
Absolutely LOVE, LOVE, LOVE this product!!!! It smells AMAZING and feels GREAT on the body! Will be repurchasing this item MANY, MANY more times!	7
Absolutely LOVE, LOVE, LOVE this product!!!! It smells AMAZING and feels GREAT on the body! Will be repurchasing this item MANY, MANY more times!	7
Absolutely LOVE, LOVE, LOVE this product!!!! It smells AMAZING and feels GREAT on the body! Will be repurchasing this item MANY, MANY more times!	7
All time favorite!!!! Wish they still carried this!!!	7
Absolutely LOVE, LOVE, LOVE this product!!!! It smells AMAZING and feels GREAT on the body! Will be repurchasing this item MANY, MANY more times!	7

- **Purpose:** Identify the top 10 reviews with the most exclamation marks.
- **Insight:** Reviews with more exclamation marks might indicate stronger emotions, either positive or negative.

#### 5. Sentiment vs Rating

```
SELECT ROUND(overall, 0) AS rating, AVG(sentiment_score) AS avg_sentiment_score
FROM [dbo].[amazon_reviews]
GROUP BY ROUND(overall, 0)
```

ORDER BY rating;

rating	avg_sentiment_score
1	-0.04212334707433698
2	0.13250344818217616
3	0.11670258371888062
4	0.2594572792952947
5	0.41492256500906416

- **Purpose:** Calculate the average sentiment score for each overall rating.
- **Insight:** Useful for identifying any mismatches between numerical ratings and the actual tone of the text.

## 6. Average Review Length by Sentiment

SELECT sentiment, AVG(review\_length) AS avg\_review\_length

FROM [dbo].[amazon\_reviews]

GROUP BY sentiment

ORDER BY avg\_review\_length DESC;

sentiment	avg_review_length
positive	219
negative	213
neutral	37

- **Purpose:** Calculate the average length of reviews for each sentiment type.
- **Insight:** Longer reviews may indicate more invested feedback or a stronger emotional response.

## 7. To find Confusing Reviews (low rating but positive text)

SELECT reviewText, overall, sentiment\_score

```
FROM [dbo].[amazon_reviews]
WHERE overall <=2 AND sentiment_score >= 0.5
ORDER BY sentiment_score DESC;
```

Query succeeded: Affected rows: 0

- **Purpose:** Identify reviews with low ratings but high sentiment scores.
- **Insight:** These reviews can be confusing, as they express positive sentiment but assign a low numeric rating.

#### ===== ADVANCED ANALYSIS USING WINDOWS FUNCTIONS =====

##### 8. To rank each Review based on word count

```
SELECT
reviewtext, word_count,
RANK() OVER(ORDER BY word_count DESC) AS word_count_rank
FROM [dbo].[amazon_reviews];
```

reviewtext	word_count	word_count_rank
I **LOVE** my Essies. They are by f...	962	1
Overall, this is an amazing value a...	692	2
I got this in a beauty subscription ...	640	3
This is an average-sized bar of sce...	551	4
I've always prided myself on my g...	429	5
I've always prided myself on my g...	429	5
I love Pre de Provence bar soaps, a...	405	7
This is probably the "smoothest" t...	396	8

- **Purpose:** Assign a rank to each review based on its word count.
- **Insight:** Useful for quickly identifying the most detailed reviews.

### 9. To find how sentiment averages out as we move up the sentiment scale.

```
SELECT
    reviewText, sentiment_score,
    AVG(sentiment_score) OVER(ORDER BY sentiment_score ROWS BETWEEN UNBOUNDED
    PRECEDING AND CURRENT ROW) AS running_avg_sentiment
FROM [dbo].[amazon_reviews]
ORDER BY sentiment_score;
```

reviewText	sentiment_score	running_avg_sentiment
too bad this is discontinued	-0.6999999999999998	-0.6999999999999998
too bad this is discontinued	-0.6999999999999998	-0.6999999999999998
too bad this is discontinued	-0.6999999999999998	-0.6999999999999998
too bad this is discontinued	-0.6999999999999998	-0.6999999999999998
Electric razors don't get a closer sh...	-0.6999999999999998	-0.6999999999999998
Electric razors don't get a closer sh...	-0.6999999999999998	-0.6999999999999998
my wife loves this bath soap, too b...	-0.4958333333333324	-0.6708333333333333
my wife loves this bath soap, too b...	-0.4958333333333324	-0.6489583333333333
my wife loves this bath soap, too b...	-0.4958333333333324	-0.6319444444444444
my wife loves this bath soap, too b...	-0.4958333333333324	-0.6183333333333334

- **Purpose:** Calculate a running average of sentiment scores.
- **Insight:** Shows how the average sentiment score changes as you move from the lowest to the highest sentiment.

### 10. To find the percentile position of each review's length

```
SELECT
    reviewText,
    review_length,
    PERCENT_RANK() OVER(ORDER BY review_length ASC) AS length_percentile
```



FROM [dbo].[amazon\_reviews];

- **Purpose:** Calculate the percentile rank of each review based on its length.
- **Insight:** Helps identify where a review's length stands relative to the entire dataset.

reviewText	review_length	length_percentile
:)	2	0
Ok	2	0
:)	2	0
Ok	2	0
:)	2	0
Ok	2	0
:)	2	0
Ok	2	0
:)	2	0
yum	3	0.00215311004784689
yum	3	0.00215311004784689
yum	3	0.00215311004784689
yum	3	0.00215311004784689
Fine	4	0.003110047846889952
AAA+	4	0.003110047846889952
Fine	4	0.003110047846889952