# CS 432/532: Final Project Report

**Amazon Products:** Analyzing Data Related To Reviews

**Team member(s):** Nicole Chan

Ⅰ. N+1 NOSQL QUERIES

**Query 2 (Nicole Chan):** Find products with the highest average rating within a two year period that mentions a specific color in their reviews.

Attributes: name, reviews.rating, reviews.text

**Query 3 (Nicole Chan):** Find the correlation of reviews written by specific users with more people finding it helpful by their usage of technological words in the review.

Attributes: reviews.username, reviews.numHelpful, name, reviews.text

Ⅱ. NOSQL DATABASE AND DATASET

This database stores all of the data relating to Amazon product reviews, with 27 total attributes used to organize the available data. These attributes include: product id, asins, brand, category, colors, dateAdded, dateUpdated, dimension, ean, keys, manufacturer, manufacturerNumber, name, prices, reviews.date, reviews.doRecommend, reviews.numHelpful, reviews.rating, reviews.sourceURLs, reviews.text, reviews.title, reviews.userCity, reviews.userProvince, reviews.username, sizes, upc, and weight. All of these attributes describe a certain product purchased by a customer and the customer's opinion on the product. This dataset provides an insight into customer sentiment on Amazon through a collection of roughly 1,600 randomly chosen reviews for various products. The data comes from actual Amazon ratings and includes details like the review text itself, usernames, and the product's overall rating. By analyzing the review text, Amazon can identify patterns for specific product information or customer experiences on Amazon more easily and efficiently.

Ⅲ. PROJECT OUTCOME

Query 2 analyzes product reviews focusing on reviews between January 1, 2016 and January 1, 2018. It groups products by name, calculates the average rating, and extracts the review text. It then matches reviews mentioning specific colors chosen including black, blue, red, green, tan, gray, or grey, and assigns a "colors" field to each product. Products with reviews mentioning no colors are excluded from the final output. The last part of the query projects only relevant information for the new table created including product name, average rating, reviews, and colors, and sorts the results by average rating in descending order so it shows products with the highest average rating first.

| | product | avgRating | reviews | color |
|---|---|---|---|---|
| 2 | Kindle Paperwhite | 5 | I have to say upfront - I | black |
| 3 | Kindle E-reader - Black | 5 | I don't know why there | grey |
| 4 | All-New Fire 7 Tablet with Alexa | 5 | Much improved screer | green |
| 5 | Amazon Echo Dot Case (fits Echo Dot 2nd Generation only) - Indigo Fabric | 4.666666667 | I'm a huge fan of the Ec | Blue |
| 6 | Amazon Echo Dot Case (fits Echo Dot 2nd Generation only) - Indigo Fabric | 4.666666667 | Changes the whole loc | tan |
| 42 | All-New Amazon Fire HD 8 Tablet Case (7th Generation | 3.166666667 | 1. While a great idea tc | black |
| 43 | All-New Amazon Fire HD 8 Tablet Case (7th Generation | 3.166666667 | I've had the new Fire H | blue |
| 44 | All-New Amazon Fire HD 8 Tablet Case (7th Generation | 3.166666667 | In this day and age of r | black |
| 45 | All-New Amazon Fire HD 8 Tablet Case (7th Generation | 3.166666667 | This is an improvemen | blue |
| 46 | Fire Kids Edition Tablet | 2.5 | I purchased 3 of these | blue |

The results of this show that the highest average rating of 5 is when those products had black, gray, and green in the review text. On the bottom of the table, it seems that reviews that had the colors blue and black had lower average ratings. The purpose of this query is to understand color preferences for these products and identify any relationship between color and customer perception of the products they rated.

Query 3 analyzes product reviews to identify users who write reviews using technical terms. It groups reviews by username and creates "reviews" containing helpfulness rating, product name, and review text for each review by the user. It iterates through each review in the "reviews" and identifies how many technical words (screen, feature, quality, light, display, sound, device, play, app, camera, microphone, tablet, echo, kindle) are mentioned in the review text. It then removes reviews with a helpfulness rating of 0 or lower. Finally, the final output keeps only relevant information like username, total helpfulness rating across all reviews, product name, review text, and the number of technical words found in each review, and sorts the results by total helpfulness in descending order, showing the users with the most helpful reviews first.

| | username | techWords | numHelpful | productName | review |
|---|---|---|---|---|---|
| 2 | NF | 15 | 997 | Kindle Fire HD 7" | An Amazon.com official commented on th |
| 3 | Michael S | 11 | 988 | Amazon Echo - Black | LOVE OUR NEW ECHO! I have been watch |
| 4 | Earthling1984 | 17 | 975 | Fire HD 6 Tablet | For the low price, this tablet really does m |
| 5 | JJCEO | 24 | 971 | Kindle Paperwhite | I have loved and used my Kindle Keyboard |
| 6 | Bryant R. | 10 | 966 | Echo Dot (2nd Generation) - Black | So I don't normally write reviews but I just |
| 356 | Alex Johnson | 0 | 3 | Certified Refurbished Kindle E-reader - Bl | Very useful for those who enjoy reading a |
| 357 | Kindle Customer | 1 | 3 | All-New Amazon Kid-Proof Case for Amaz | I am happy in having this good things. |
| 358 | Grandmaof4 | 2 | 3 | Amazon 5W USB Official OEM Charger an | According to the info, the Paperwhite cha |
| 359 | Eilhard Molina | 0 | 2 | All-New Amazon Kid-Proof Case for Amaz | Super cool! |
| 360 | bkgrigsby725 | 3 | 2 | Amazon Tap - Alexa-Enabled Portable Blu | Love that it's 'interactive' as long as I'm c |
| 361 | Len Burroughs | 0 | 2 | Amazon 5W USB Official OEM Charger an | very overpriced! |

The results of this query reveal that the number of people who find the reviews more helpful using a lot more technical terms is higher than the reviews with less technical terms. These results prove that technical details enhance the clarity and value of reviews for other users and potential buyers of the product.

## IV. REFERENCES

https://www.kaggle.com/datasets/yasserh/amazon-product-reviews-dataset