Analyzing Amazon Consumer Behavior Survey Data using Power BI

Introduction

Among the many applications of data analytics, I have always been intrigued to crunch consumer market data. Understanding consumer preferences and accordingly, their behavior is something that I am curious about.

While hunting for relevant datasets online to explore this domain, I came across the **Amazon Consumer Behaviour Dataset** on Kaggle which caught my attention. To me, what stood out about this dataset was the level of intricacy and specificity that was recorded by the conducted survey.

The inclusion of metrics like preferred product search methods, cart abandonment factors, dependency on reviews and ratings while purchasing products, etc. were some interesting areas for exploration. Of course, I hand-picked the ones that I deemed to be impactful.
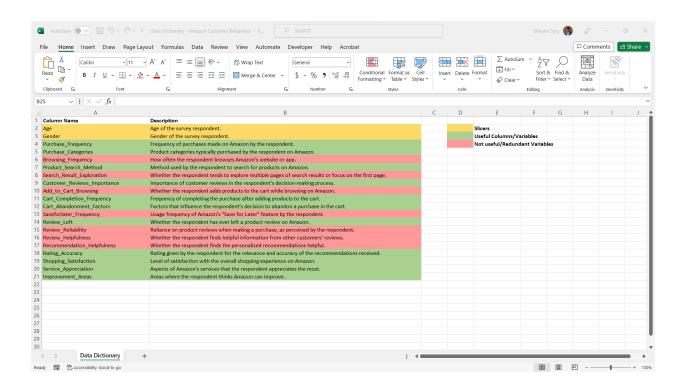
This portfolio project allowed me to break down a survey conducted on a set of Amazon's customers to get insights into their purchasing patterns and their overall satisfaction levels with Amazon by different demographic elements.

Data Cleaning and Preparation

Even though the dataset contained various fascinating fields, I still had to indulge in a good amount of data cleaning, transformation and preparation before I could use it for analysis and visualization.

Some of the important steps I conducted were the following:

- **Creating a data dictionary:** The first step was to clearly and accurately understand what each field/column represented. And also, to get a preliminary assessment of the more important variables for the analysis.
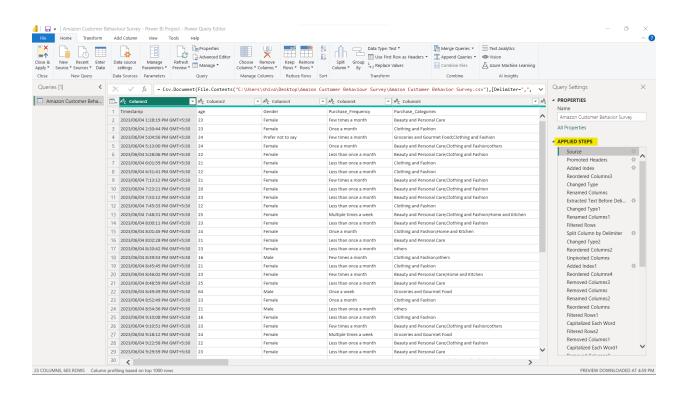

Within the dictionary, I identified and classified the fields into different categories: the ones highlighted in **yellow** were to be used as **slicers** (a tool that helps you pick and show only the data you need in a report by selecting options from a list), the one highlighted in **green** were **useful columns** to be used for analysis, and the ones highlighted in **red** would be **discarded or not considered** for analysis.

This of course was done at a preliminary stage to identify meaningful analytics questions while eliminating the extra fat.
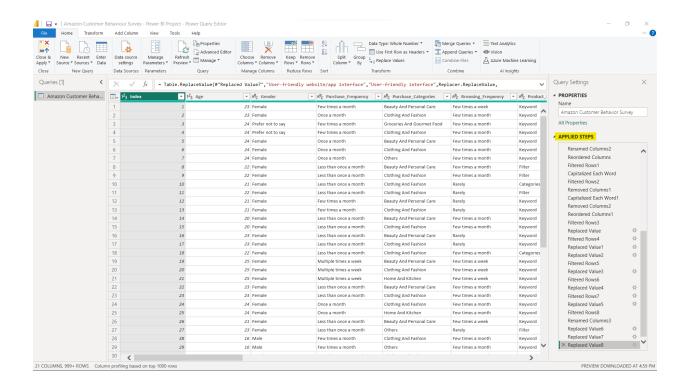
The creation of a data dictionary also contributed to documentation. Documentation is an underrated practice which is immensely useful for bigger projects with multiple stakeholders involved.

- **Using Power Query Editor for data transformation:**

This is how the data looked originally

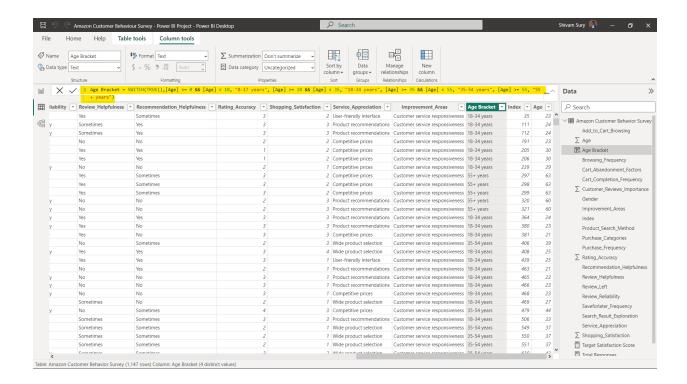And this is how it looked post all the transformations were applied:

Important transformations:

1. promoting the first rows as column headers, changing column data types and removing unneeded/redundant columns.

2. splitting the 'Purchase_Categories' column into different columns since it contained many categories for a single respondent; the column was split using the delimiter and then unpivoted to create unique rows for each category by each respondent.

3. creating an 'Index' column; the data originally didn't include a unique identifier for different survey response submissions; it was important to follow this step after splitting the 'Purchase_Categories' column.

4. renaming some categorical row values to ensure uniformity and clarity

- Using the `SWITCH` function to create a calculated column called '**Age Bracket**'

The dataset comprised different ages for different respondents. For our analysis, I figured that converting individual age units into age brackets would summarize and categorize the age variable- an important demographic criterion.

Glossary:

> 💡 **SWITCH** *function: evaluates conditions and returns a value based on the first true condition.*

Analytics Questions

After completing the data cleaning and processing phase, the stage was to determine the analytics questions which would be answered with elegant and informative data visualizations in the form of a dashboard.

Some of the analytics questions I have tried to answer:

Q. What are the most popular product categories?

Q. What is the preferred method for searching for products?

Q. What is the purchase frequency for consumers?

Q. What is the cart completion frequency (the tendency to complete an order after items have been added to the cart) for consumers?

Q. What are some common factors that contribute to cart abandonment (consumers not completing their orders)?

Q. What do shoppers appreciate most when shopping with Amazon?

Q. What is the overall satisfaction score for consumers with Amazon?

Q. What are some areas of improvement for Amazon?

Analysis and Visualization

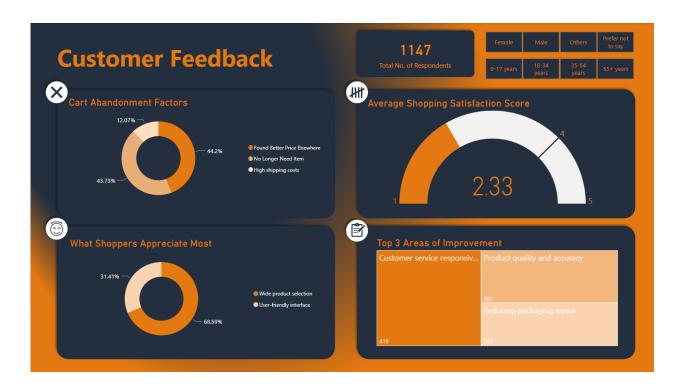Before delving deep into the analysis and interpretation, here's a glance at the dashboards created:

After shortlisting the data fields I wanted to utilize for analysis, I decided to break things down into two components (resulting in two dashboards): evaluating the **Purchasing Behavior** and the **Customer Feedback**.

For the first dashboard showcasing visualizations about Purchasing Behavior, the following were some of the key observations:

- The majority of consumers/users (41.1%) prefer to search for products by typing in the **Keywords** across various age groups and genders, followed by searching for products through the **Categories** dropdown on the website/app.

- **Female** shoppers tend to have a higher **purchase frequency** (38% shop a few times a month) as compared to **male** shoppers (30% of them shop less than once a month).

- The most popular product category for females turned out to be **Beauty and Personal Care (31.5%)** while for **males** it turned out to be **Clothing and Fashion (27.3%)**.

- Age groups **0-17 years** and **18-34 years** have a high shopping frequency in the **Clothing and Fashion** category. For the age group **35-54 years** there was a sharp increase in purchases for the **Home and Kitchen** category. Lastly, for the age group **55+ years**, the most popular product category came out to be **Beauty and Personal Care**.

- Around **77%** of consumers complete their purchases after adding products to their cart.

The second dashboard focuses more on evaluating the customer feedback received through the survey:

- The top three factors for a shopper abandoning their cart and not completing the purchase were the following in the same order: **found a better price elsewhere (44.2%)**, **no longer needed the item (43.7%)** and **high shipping costs (12%)**.

- The average **shopping satisfaction score** was recorded as **2.33** on a scale of 5.

- The most commonly valued aspects of Amazon's service were the presence of a w**ide range of product selections** and a **user-friendly interface** on the website and app.

- Some key areas of improvement for the retail giant would be the following: **customer service responsiveness**, **product quality and accuracy**, and

reducing packing waste.

Recommendations

Based on the findings of the analysis, several recommendations are proposed for Amazon to enhance customer satisfaction and address areas for improvement:

1. Improve the **searchability of products** by investing in advanced keyword search algorithms and refining the search functionality.

2. **Personalize product recommendations** based on gender and age groups to cater to specific consumer preferences and increase customer engagement. It will also be helpful for **retargeting ads**.

3. Enhance **quality assurance** processes to ensure product accuracy, quality, and consistency, thereby improving customer trust and satisfaction.

4. Invest in **customer service training** and resources to **improve responsiveness** and resolve customer queries and concerns more efficiently.

5. Implement **eco-friendly packaging solutions** to reduce waste and align with the growing environmental consciousness of consumers.

Barring these more obvious recommendations that were deduced from the analysis, the most important recommendation would be to **dive deeper** into some of the areas explored in this analysis. The survey data gives us a fair high-level picture of the different elements inspected. But a more low-level and specific investigation would be required for each of these areas, as per importance and level of impact.

For instance, we did observe that customers appreciated a user-friendly interface. But what do they specifically like in the user interface? Is it the placement of different information objects (information architecture)? Is it the presence (or absence) of certain elements?

Or if consumers are not enjoying the customer service responsiveness, what are the identifiable causes for it? And what could be the potential and tailored solution for that?

Conclusion

In conclusion, this project provides a high-level analysis of the Amazon Consumer Behavior Survey dataset, revealing valuable insights into consumer preferences and behaviours.

The findings highlight significant patterns and trends related to product search methods, purchase frequency, cart completion, cart abandonment factors, satisfaction levels, and areas for improvement.

The recommendations put forth aim to assist Amazon in enhancing its offerings, improving customer satisfaction, and refining its strategies to better cater to the needs and preferences of its consumers.

By leveraging the insights gained from this analysis, Amazon can further establish its position as a customer-centric and adaptive e-commerce platform.

References

Menon, S. (2023, July 6). *Amazon consumer Behaviour Dataset*. Retrieved from Kaggle: https://www.kaggle.com/datasets/swathiunnikrishnan/amazon-consumer-behaviour-dataset

Microsoft. (n.d.). *SWITCH function (DAX) - DAX*. Retrieved from Microsoft Learn: https://learn.microsoft.com/en-us/dax/switch-function-dax

Shrewsbury, T. (2023, July 3). *Understanding Consumer Behavior: Mastering Segmentation and Positioning for Effective Marketing*. Retrieved from Medium: https://medium.com/@tylershrewsbury/understanding-consumer-behavior-mastering-segmentation-and-positioning-for-effective-marketing-147e4359a977

Slater, D. (n.d.). *The Imperatives of Customer-Centric Innovation*. Retrieved from Amazon: https://aws.amazon.com/executive-insights/content/the-imperatives-of-

customer-centric-innovation/