Lexus Carton - Business Engineering Management Science

Nov 2, 2023

Intro. Data Science – Fall 2023

Project Update #1

This project has three parts: Data acquisition, Preprocessing, and ethical consideration. Preprocessing has several parts.

Acquisition:

The datafile is a comma separated file called "Amazon Customer Behavior Survey," located from Kaggle.com. The Survey was conducted through Google Forms, and the initial data file has 608 rows and 23 columns. Because it is an anonymous survey, participants did not include their geographical residencies. Lack of inhabitant knowledge limits the availability to target selling to countries or cities. Additionally, the Survey covered a period of 2023/06/04 to 2023/06/16. The investigation will know not much about consumer preferences in contrast to economic weather.

Preprocessing:

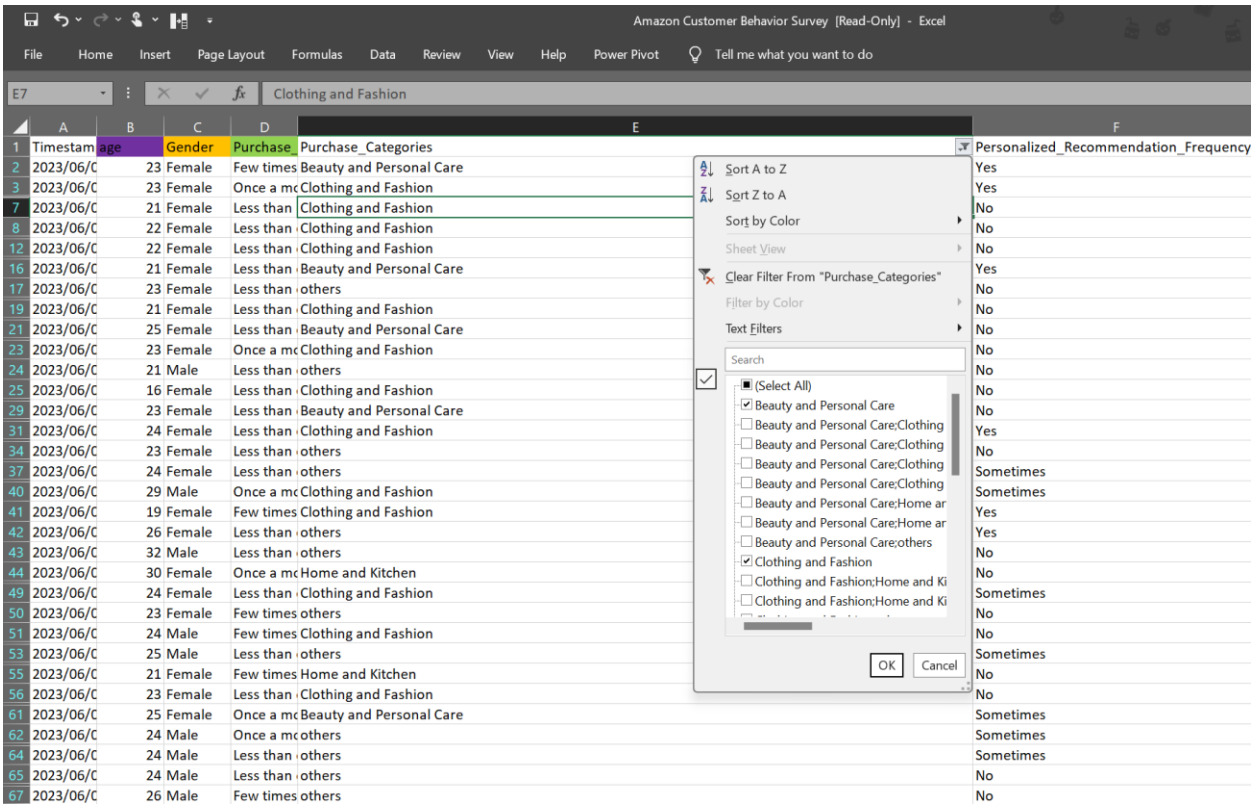
The first of the process was to look at the data set in Excel. I categorized all the columns into Binary, Numerical, and Categorical variables to make visualization more manageable. There were no missing variables, so they did not have to be imputed with median or mode.

Binary	**Numerical (1-5 stars)**	**Categorical**
Search_results_Exploration, Review_Left	Customer_Review_Importance, Rating accuracy,Shopping_Satisfaction, age, Rating Accuracy	Cart_Abandoment_Factors, Product_search_methods, Purchase_Categories,, Personalized_recommenation_frequency,Gender, Review_reliability, Review_helpfulness ,Purchase_Frequnecy, Service_Apprreciaton ,Browsing_Frequency Add_to_Cart_Browsing Cart_Completion_Frequency Saveforlater_Frequency

		Recommendation_Helpfulns Improvement_Areas

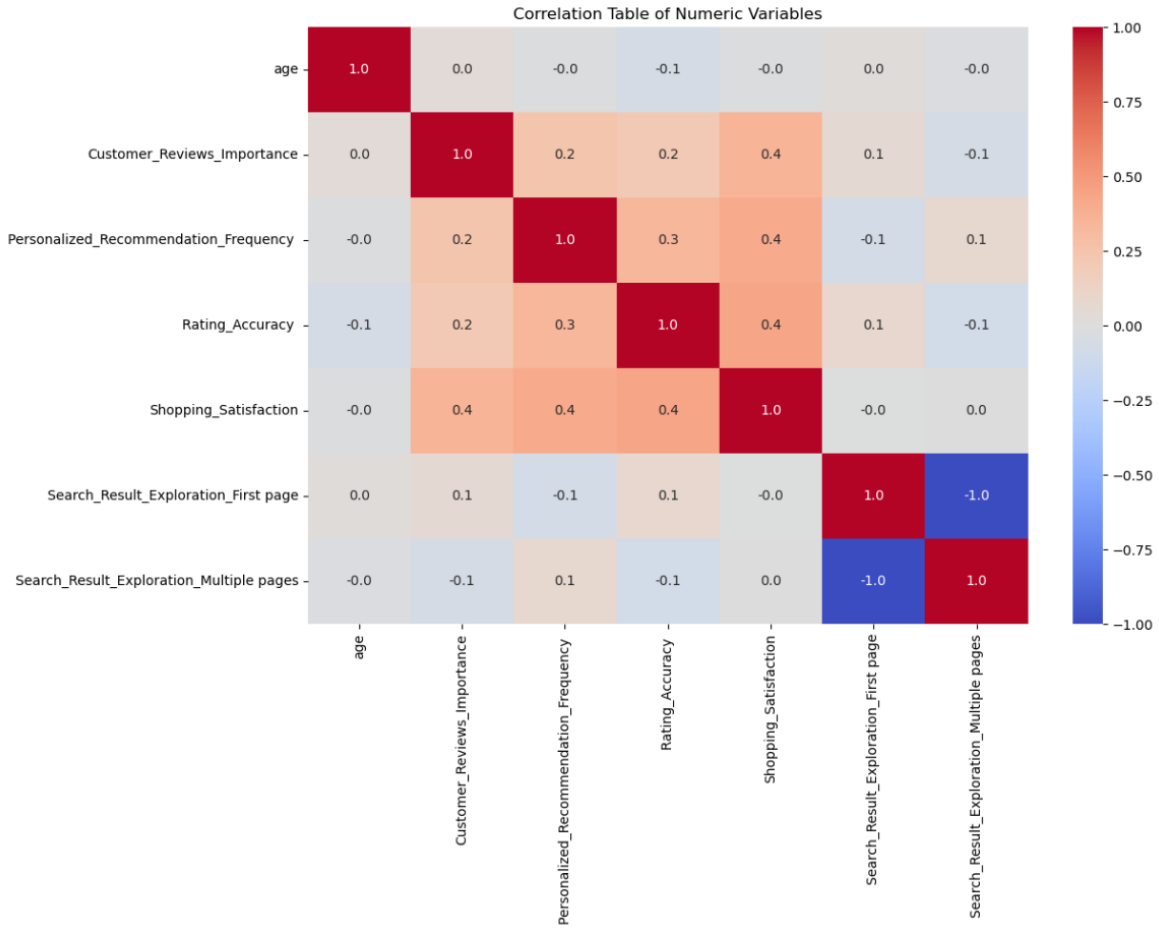
I filtered the Purchase Categories because there are several, for example, Beauty and Personal Care Beauty Personal; Clothing or Beauty Personal, Home, and Kitchen. The filter was making the Purchase Category very exact for some survey responses. The Filter categories became Beauty and Personal care, Clothing and Fashion, Groceries and Gourmet Food, Home and Kitchen, and Other. I thought this was acceptable because the goal of the project is to look for the general target market.

The cleaned file became "C:\Users\Lexus\Desktop\Amazon Customer Behavior Survey FIltered.csv" . The number of rows only changed from 608 to 603, and it made the visualizations look better.

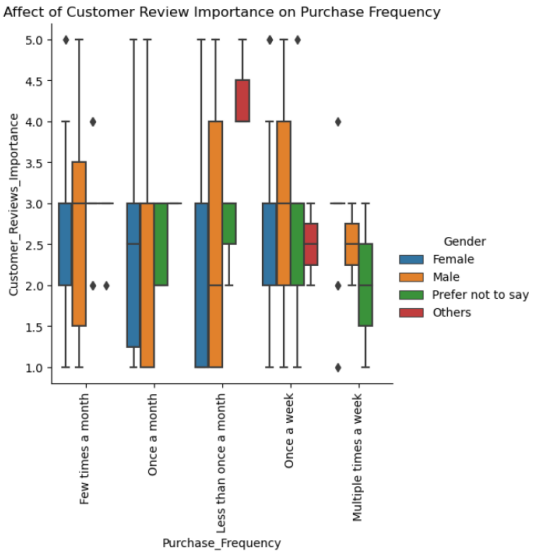


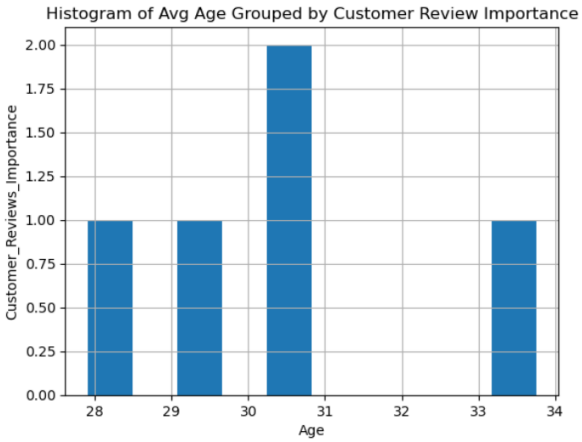
	A	B	C	D	E	F
1	Timestam	age	Gender	Purchase_	Purchase_Categories	Personalized_Recommendation_Frequency
2	2023/06/0	23	Female	Few times	Beauty and Personal Care	Yes
3	2023/06/0	23	Female	Once a mo	Clothing and Fashion	Yes
7	2023/06/0	21	Female	Less than	Clothing and Fashion	No
8	2023/06/0	22	Female	Less than	Clothing and Fashion	No
12	2023/06/0	22	Female	Less than	Clothing and Fashion	No
16	2023/06/0	21	Female	Less than	Beauty and Personal Care	Yes
17	2023/06/0	23	Female	Less than	others	No
19	2023/06/0	21	Female	Less than	Clothing and Fashion	No
21	2023/06/0	25	Female	Less than	Beauty and Personal Care	No
23	2023/06/0	23	Female	Once a mo	Clothing and Fashion	No
24	2023/06/0	21	Male	Less than	others	No
25	2023/06/0	16	Female	Less than	Clothing and Fashion	No
29	2023/06/0	23	Female	Less than	Beauty and Personal Care	No
31	2023/06/0	24	Female	Less than	Clothing and Fashion	Yes
34	2023/06/0	23	Female	Less than	others	No
37	2023/06/0	24	Female	Less than	others	Sometimes
40	2023/06/0	29	Male	Once a mo	Clothing and Fashion	Sometimes
41	2023/06/0	19	Female	Few times	Clothing and Fashion	Yes
42	2023/06/0	26	Female	Less than	others	Yes
43	2023/06/0	32	Male	Less than	others	No
44	2023/06/0	30	Female	Once a mo	Home and Kitchen	No
49	2023/06/0	24	Female	Less than	Clothing and Fashion	Sometimes
50	2023/06/0	23	Female	Few times	others	No
51	2023/06/0	24	Male	Few times	Clothing and Fashion	No
53	2023/06/0	25	Male	Less than	others	Sometimes
55	2023/06/0	21	Female	Few times	Home and Kitchen	No
56	2023/06/0	23	Female	Less than	Clothing and Fashion	No
61	2023/06/0	25	Female	Once a mo	Beauty and Personal Care	Sometimes
62	2023/06/0	24	Male	Once a mo	others	Sometimes
64	2023/06/0	24	Male	Less than	others	Sometimes
65	2023/06/0	24	Male	Less than	others	No
67	2023/06/0	26	Male	Few times	others	No

A correlation table of numeric variables will help determine multicollinearity and display the variables with the highest predictive power .None of the variables are at 0.5 but several are at 0.4.The variables have to do with customer experience and review legitimacy.

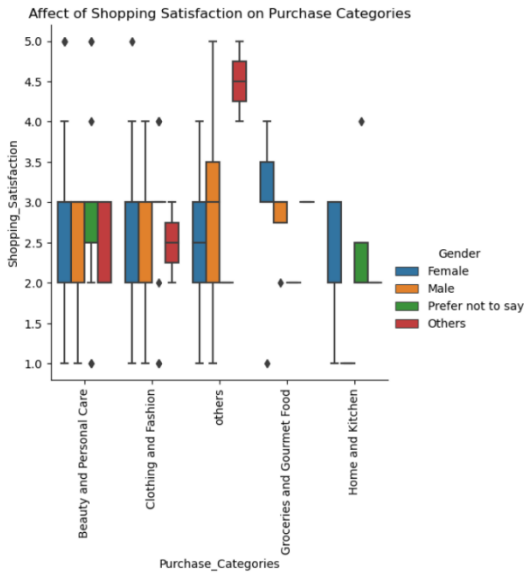


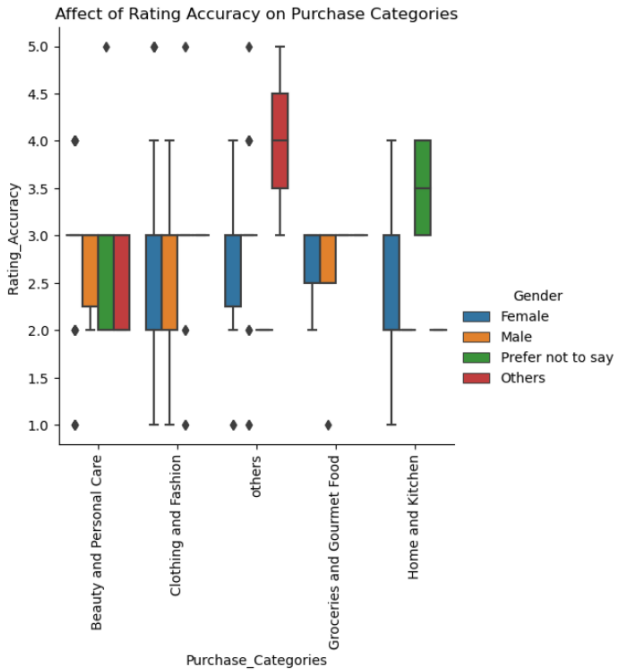
The next two graphics represent what are demographics for and how often they utilize reviews. 1 is the lowest and 5 is the highest. Individuals in their early 30s utilize Amazon's services most often. In all frequencies, Males purchase than any Gener preference.





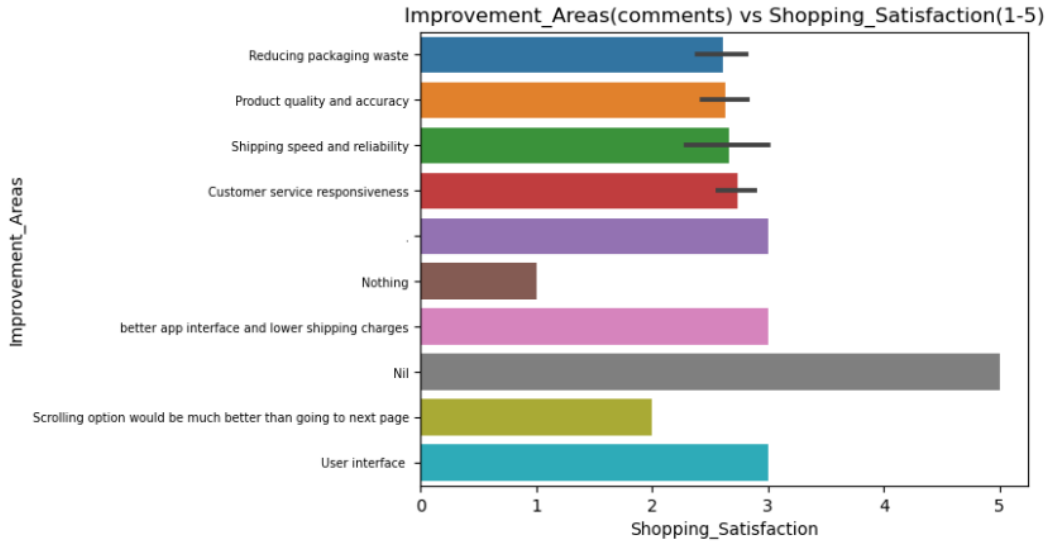
The two boxplots visualize Purchase Categories vs. Rating Accuracy and Shopping satisfaction. The boxes are color-coded based on gender: Female, Male, Prefer not to say, and Other. The Shopping satisfaction for females is present, but the rating accuracy vs purchase category is not. It could be possible there is a high volume of inaccurate ratings on beauty care products, thus inflating the product's value and consumer confidence when consumers shouldn't trust the reviews. The same case is for Others, or the more niche category. In either of the graphs, I prefer not to say it is not represented. Many males are represented in the shopping satisfaction vs. purchase category but are absent in the rating accuracy vs. purchase category. A similar situation may be present as products from shopping satisfaction are

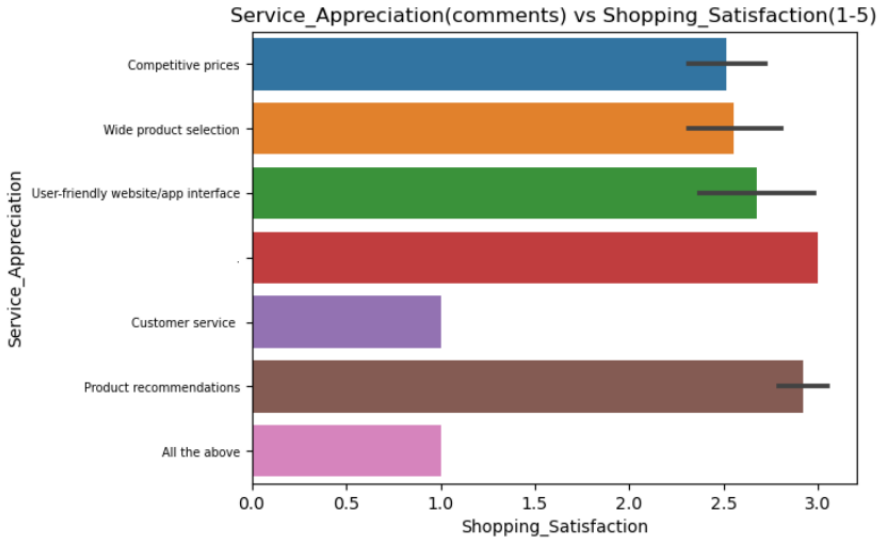




The following two visualization focus on purchase categories. In the first bar chart, Improvement Areas vs Shopping satisfaction, The highest is Nil(a catch-all category) at five stars, the second highest categories are User interface, better app interface, and (.) at a score of 3 three stars. Customers like to have little to no human interaction. This is apparent in Service appreciation. The symbol (.) denotes no comment; so assumtions must be that there is generaly no improvement needed.

In the second bar chart: Service Appreciation vs Shopping Satisfation, the highest column is (.) with product recommendations falling slightly under three. The lowest is Customer service and all of the above at 1. In the first two graphs, this assumes consumers prefer revies for products with seamless digital e-commerce. One can only guess what these catch all comments would be, or why one customer had a notible expierance for an aprreication over the other.





E

Ethical considerations

How many children (3-18) are utilizing Amazon services?

While evaluating the CSV and filtering the age column by ascending, I noticed that the age range was 3-67. Looking at the survey data produced from the three-year-old participants who's gender was prefer not to say. They purchased a Home and kitchen product, and the Rating_Accuracy was a three, and the Shopping_satifaction was a 2. Perhaps this was a parent shopping for their child. Another instance of this is survey participants ages sixteen-seventeen.

The sixteen-year-old user had 1s for both Rating_Accuracy and Shopping_Satifaction, while the seventeen-year-old users had 4s and 5s for both Rating_Accuracy and Shopping_Satisfaction. Individuals could be negligent when providing feedback for services, as in it was too good or too bad. This mind set could trickle up into age groups if they are given a product or service that doesn't fit one of their needs.

How many jobs could be replaced by automation?

In the two bar graphs comparing different Shopping_Satisfaction, it is apparent that customers want little to no interaction with customer service providers. However, society is seeing automation replace more jobs every day, and it will cause more job displacement in human resource based / Shopping satisfaction was the highest for Improvement categories like better app interface and lower shipping charges as well as irrelevant product suggestions. This is apparent in-Service appreciation. The symbol (.) denotes no comment; but customers. One can only guess how they actually felt.

References:

- Menon.S. (date).Amazon Consumer Behavior Dataset. Kaggle Amazon consumer Behaviour Dataset (kaggle.com) o Data set used for investigation.
    - Dataset used for investigation.