

Shopping Behavior and Pattern Visualization

Information Visualization Final Project

Regional Purchase Pattern:

Distribution of Purchases Across US States

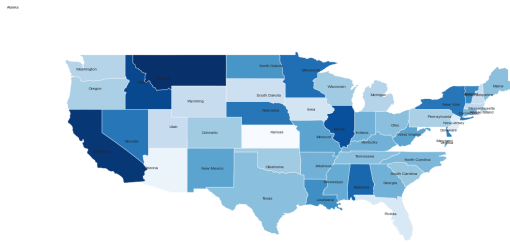
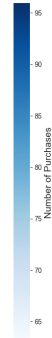


Fig1. Purchase Amount Choropleth Map



Average Purchase Amount by State (USD)

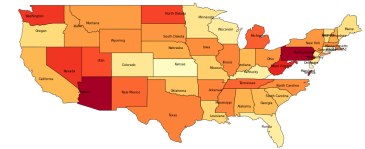


Fig2. Average Purchase Amount Choropleth Map

Figure 1 and 2 shows the choropleth map related to the purchase pattern.

Figure 1 shows the number of purchase choropleth map, while the color represents different purchase amount while darker color means higher number of purchases for each state as shown on the right legend.

Figure 2 shows the average purchase amount choropleth map, the legend on the right indicates how the colors correspond the average purchase amount respect to state.

Findings:

- Revealed the different shopping behaviors about each state
- East region generally have higher purchase amount, especially California and Montana
- Northeast region usually have higher average purchase amount such as Pennsylvania or west Virginia
- Most Southern states show moderate to lower average purchase amounts
- The number of purchases is not highly related to the average purchase amount. For example, California have one of the largest number of purchases but the average purchase amount is relevantly low.

Age related Analysis

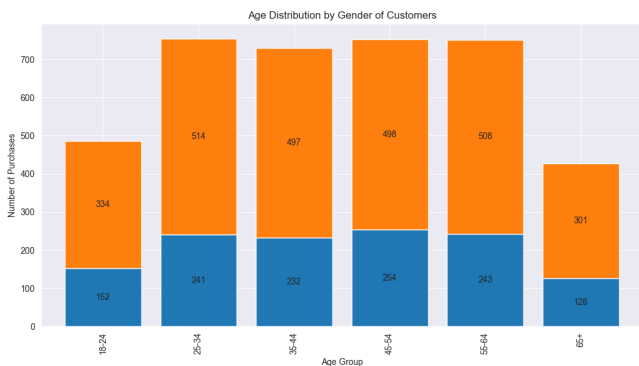


Fig3. Age and Gender Distribution of Customer Histogram

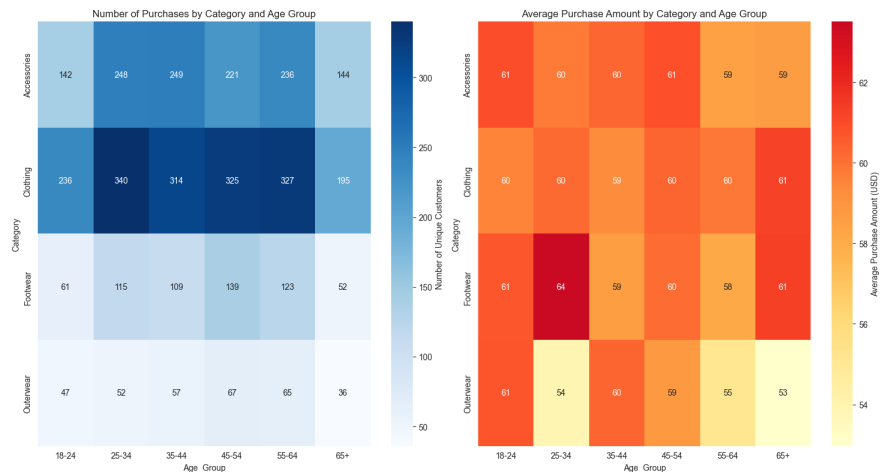


Fig4. Average Purchase amount and Number of Purchases Respect to Category and Age Group Heat Map

The Figure 3 is a histogram shows the number of purchases related to age and gender, while y-axis is number of purchases, and x-axis is the age bin, different colors in the represent different gender, orange for male, and blue for female.

The Figure 4 is a heat map shows the shopping pattern related to different age group, there are multiple fields has been included in this heat map, including category, purchase amount, number of purchases. For both figure, the y-axis represents the category and the x-axis represents the categories, For the left one, the intensity of color represents the total number of purchases for each category, while darker color means higher number of purchases. For the right one, the intensity of color represents the average purchase amount, similarly darker color means higher amounts.

Findings:

- Across all the age group, male always have higher purchase amount.
- In figure 3, age group 25-54 have a similar purchase amount pattern, while age group 18-24 and 65-70 are much lower. The number of purchases heat map in figure 4 also agrees with the result in figure 3.
- Across all the age group, they all show a similar preference for the category, Clothing have significantly higher number of purchases, while outwear have lowest number of purchases.
- Although customer all have similar preference of category to purchase across all age, the average price for each category does not have a clear pattern across the age, the average purchase amount various significantly by both age group and product category.

Product Preference:

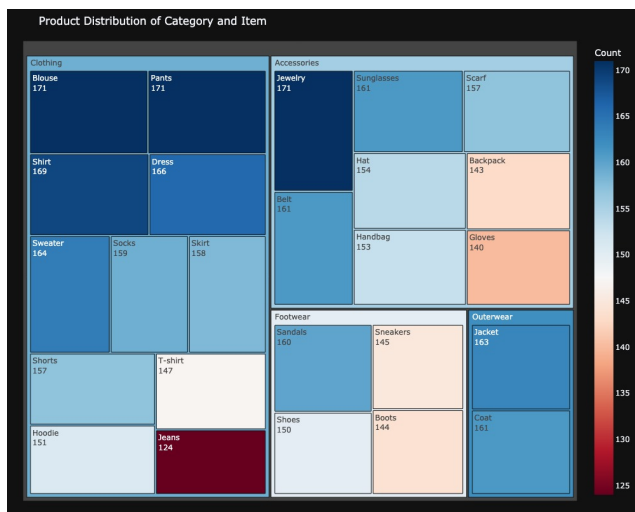


Fig 6. Product Distribution of Category and Item Tree Map

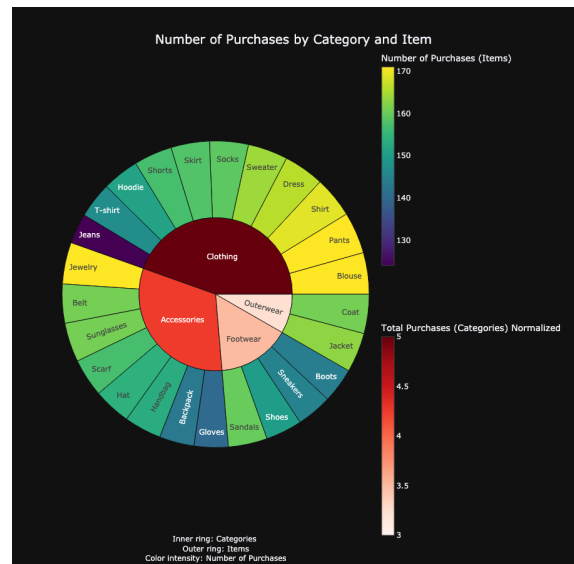


Fig 7. Number of Purchases by Category and Item Sunburst Graph

The figure 6 presents the tree map displays the distribution of products purchased across different categories and items, and the size and color of each rectangle represent the number of purchases.

The figure 7 presents the sunburst graph visualizes the number of purchases by product categories and items. Inner ring shows the product categories, while the outer ring presents the specific item within each category. Color intensity represents the number of purchase, with yellow indicates higher purchase volumes and darker blue indicates lower purchase volumes. The total purchases number for category is normalized to distinct it from other color, also without affect the outer ring color tone.

Findings:

- The tree map and sunburst graph presents same information that agrees with each other. Also, the purchase pattern of category also can match the one in figure 4.
- Clothing have the largest portion, this may due there are various kind of item in the clothing category. Also, Clothing shows the widest range of purchase volumes, from 124(Jeans) to 171(blouse and pants).
- Tree Map and Sunburst graph are both very suitable for visualize hierarchical data structures, like the category-subcategory relationship represents in the figure 6 and 7.

Significance Statement:

I have selected the figures that been presented in this report to show the shopping behavior patterns with different topic that is important to both business analysis and consumer research. Those figure are been selected because they have shown more interesting and advanced visualization technique, and the information it contains helps us to determine the different aspect of the shopping behavior analysis.

The various visualization techniques used in this report, from choropleth maps showing geographic patterns to tree maps and sunburst charts displaying product categories, work together to give us a complete picture of shopping behaviors. The heat maps and histograms help us understand how different age groups shop, while the hierarchical visualizations show what products are popular. Since each type of visualization has its own strengths. Like how choropleth maps are great for seeing regional patterns but tree maps are better for showing category breakdowns, using them together gives us the most comprehensive understanding of shopping patterns. This combination of visualization methods helps businesses better understand their customers and make decisions about their products and marketing.

There are several other graphs been generated, the code for all the figure generation can be found in GitHub.

Data and Method:

The dataset is about various shopping related data. In this dataset, the data including three major types: customer information, product attributes, and shopping behavior metric. The detail dataset is from Kaggle and can be view in:

<https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and-shopping-habits-dataset>

In this project, all graphs are generated with python in Jupyter notebook. A selection of data parameter is used in the visualization process, to better understanding the connection between each fields and explore the advanced visualization techniques.

Github Page:

<https://github.com/yongjin77/ShoppingViz>