IFT 533: Data Visualization & Reporting for IT
Prof. Asmaa Elbadrawy

Project - Phase 2: Decision Making

Team 13

Shreyas Anil Hingmire
Nishtha Dharmendra Wagh
Shivani Chauhan

Section 1: Choosing Visualization Tools

For the development of our interactive dashboard, we have chosen Tableau as the primary visualization tool over Python's Bokeh library. This decision is based on the following reasons:

1. User Interface and Accessibility:
   Tableau excels with its highly intuitive interface, enabling quick mastery and ease of use. Unlike Python's Bokeh, which requires foundational knowledge of Python programming, Tableau facilitates rapid visualization creation through its drag-and-drop interface, significantly expediting the development process.

2. Visualization Capabilities:
   Tableau offers superior, easy-to-implement visualizations and seamless dashboard integration, making it ideal for complex data interactions and enhancing user engagement.

3. Integration Features:
   Tableau's robust integration with various data sources and advanced data blending simplifies syncing across platforms, ensuring a seamless workflow.

To conclude, Tableau's comprehensive capabilities for handling complex visualizations and its ease of use make it the preferred choice over Python's Bokeh for our dashboard. These features align with our objective to develop responsive and insightful dashboards that can be easily utilized across various departments within the company.

Section 2: Data Preparation and Preprocessing

The specific data preprocessing and preparation procedures used to examine a consumer trends shopping dataset for Tableau visualization are described in this study. These procedures are essential to guaranteeing the precision and usability of our visualizations, which rank states according to purchase frequencies and divide up customer groups.

1. Categorizing Customer Ages:

We will divide the customers' ages into predetermined groups to examine their behavior at various periods of life. To segment the data to represent different consumer preferences and purchase patterns that generally correspond with age, this phase is crucial. The following defines the age groups:

Young Adult: Ages 18-24
Adult: Ages 25-39
Middle-Aged Adult: Ages 40-59
Senior Adult: Ages 60 and above

We will use the ‘Create Group’ function to categorize the age groups.

2. Grouping the states by regions:

To analyze the data from a geographical perspective, we grouped the states into five regional categories based on their locations.

West
East
North
South
Central

Each state, as listed in the Location column of our dataset, will be assigned to one of these regions using a ‘State Region’ calculated field in Tableau. This regional grouping allows for a comparative analysis of consumer trends across different parts of the United States, helping to identify regional preferences and behaviors.

The grouping will be implemented using a calculated field in Tableau.

3. Ranking States by Purchase Frequency:

We will evaluate the states according to the total of prior purchases to see which ones have the highest customer buy frequencies. To do this, a calculated field will be made in Tableau:

Calculated Field Name: `State Rank`
Formula: `RANK(SUM([Previous Purchases]))`

The states with the best customer engagement and sales effectiveness can be found using this ranking.

4. Coloring States Based on Ranking:

We will be using a color-coding technique in our Tableau graphic to visually distinguish the states according to their rankings for purchase frequency. To do this, a second calculated field will be created:

Calculated Field Name: `State Color`
Formula: `IF [State Rank] <= [Number of Loyal States] THEN "Top N" ELSE "Other"`

The top N states, where N is the number of states we have selected to draw attention to as having the highest purchase frequencies, can be visually distinguished owing to this calculated field. To make identification and comparison easier, states in this top tier are colored differently from one another.

5. Top Category by State:

In this data preprocessing step, we will create a calculated field called 'Top Category by State' to identify and highlight the top-selling product category in each state for the selected season(s). The field calculates the total sum of Previous Purchases for each category within a state and season, then compares it to the maximum sum of purchases across all categories within the same state and season. This allows Tableau to dynamically highlight the state based on the highest-selling product category.

Calculated Field Name: `Top Category by State`
Formula:

```
IF {FIXED [Location], [Category], [Season] : SUM([Previous Purchases])} =
   {FIXED [Location], [Season] : MAX({FIXED [Location], [Category], [Season] : SUM([Previous Purchases])})}
THEN [Category]
END
```

Section 3: List of the Final Set of Questions

Dashboard 1: Customer Preferences and Trends

This dashboard will focus on understanding who the customers are, their purchasing habits, and their preferences across different demographics and seasons.

1. **What is the distribution of our customer base across different age categories?**
    - Plot 1: Distribution of Customer Base across Age Categories
2. **How is our customer base distributed across different genders?**
    - Plot 2: Distribution of Customer Base across Gender
3. **How do top-selling product categories change across states during different seasons?**
    - Plot 3: Seasonal Product Category Leaders by State

Dashboard 2: Revenue Drivers and Payment Insights

This dashboard will provide insights into what drives revenue, including high-performing products and preferred payment methods, which are crucial for financial planning and operations.

4. **Which states have the highest overall consumer spending in terms of purchase amount?**
    - Plot 4: Top States by Overall Consumer Spending
5. **Which states have the least overall consumer spending in terms of purchase amount?**
    - Plot 5: Bottom States by Overall Consumer Spending
6. **Which product categories and their corresponding items generate the highest revenue?**
    - Plot 6: Top Revenue Generators by Product Category
    - Plot 7: Top Revenue-Generating Items
7. **What are the most commonly used payment methods among consumers?**
    - Plot 8: Breakdown of Payment Methods Used
8. **How does the number of subscribers compare to non-subscribers?**
    - Plot 9: Comparison of Subscribers vs. Non-Subscribers
9. **How do discount applications vary by product category and shipping method?**
    - Plot 10: Discount Usage Overview
    - Question 9 is being answered by combining the insights from Plot 10 along with those of Plot 13.

Dashboard 3: Customer Satisfaction and Engagement

This dashboard will focus on how customers interact with the company's services and products, highlighting areas of strength and opportunities for improvement in customer engagement.

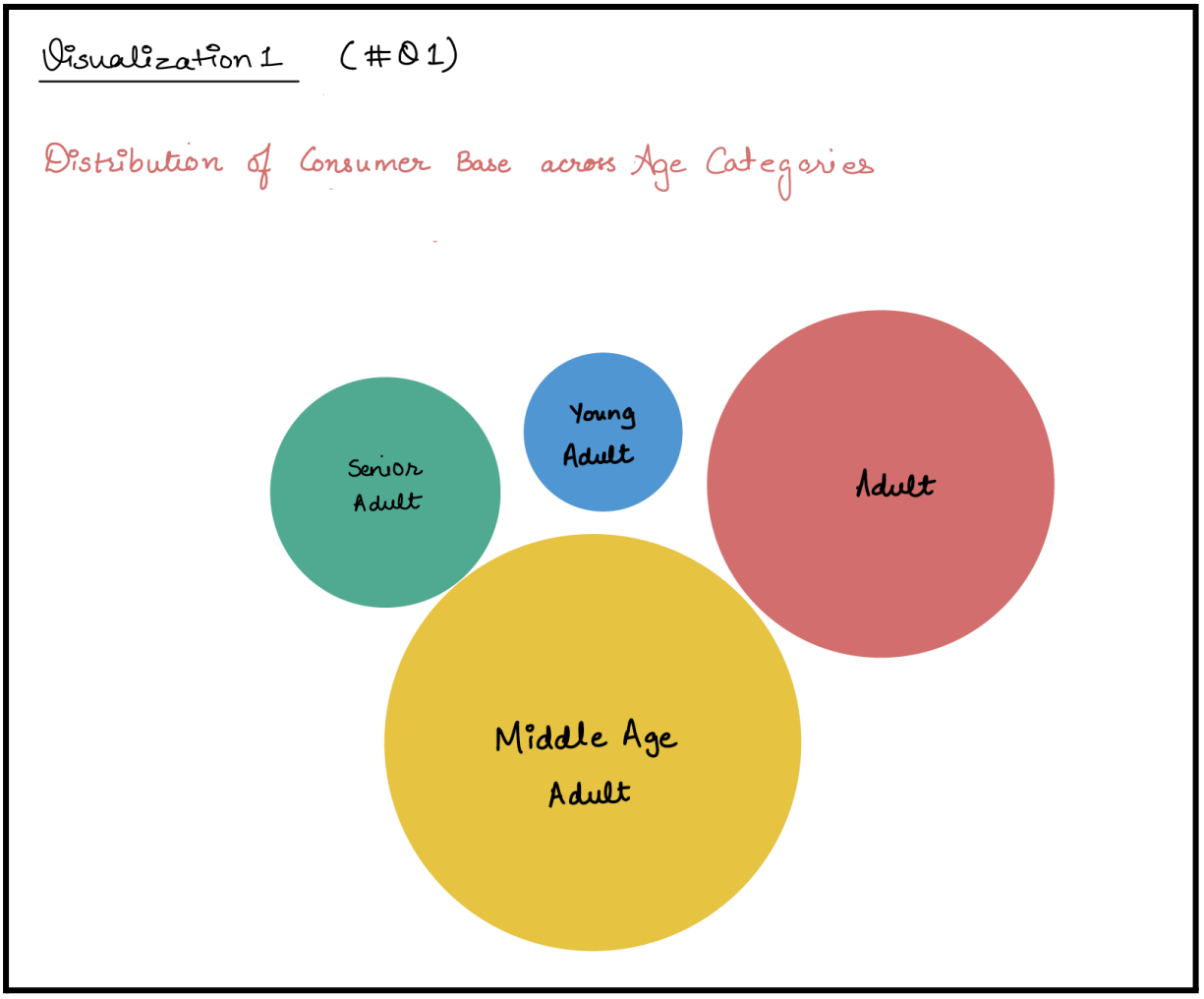
10. **In which states are loyal customers most concentrated based on their previous purchase activity?**
    - Plot 11: Geographic Distribution of Customer Loyalty
11. **How do customer preferences for different shipping types vary across seasons?**
    - Plot 12: Preferred Shipping Methods by Seasons
12. **What are the preferred shipping methods for different product categories?**
    - Plot 13: Preferred Shipping Methods by Product Category

Section 4: Dashboard Plot Drafts

Plot 1 (Question 1) - What is the distribution of our customer base across different age categories?

Plot - Bubble Chart

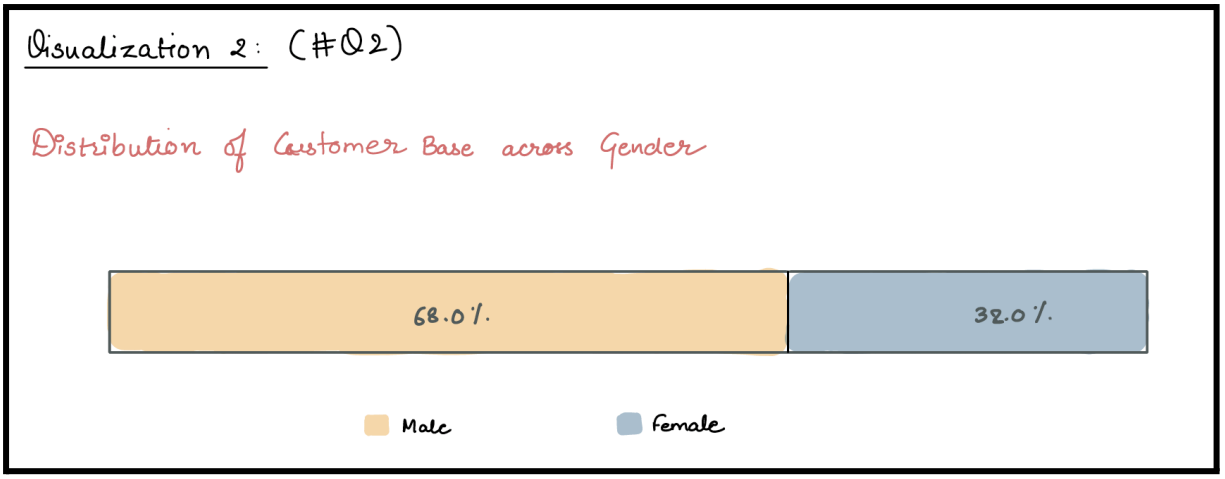
Pre - Attentive Attribute - Size, Color



Plot 2 (Question 2) - How is our customer base distributed across different genders?

Plot - Stacked Bar Chart

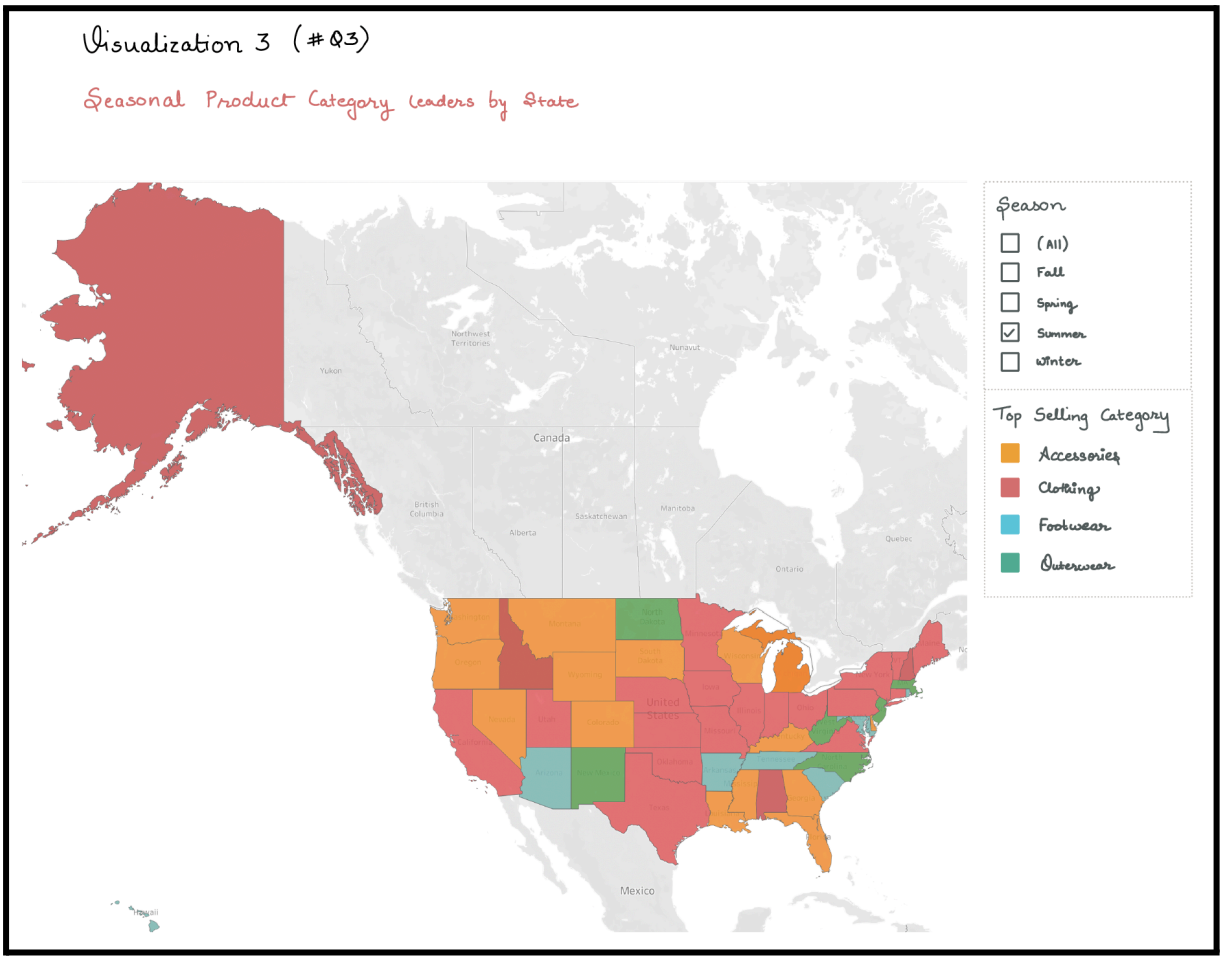
Pre - Attentive Attribute - Color , Length , Position



Plot 3 (Question 3) -How do top-selling product categories change across states during different seasons?

Plot - Choropleth

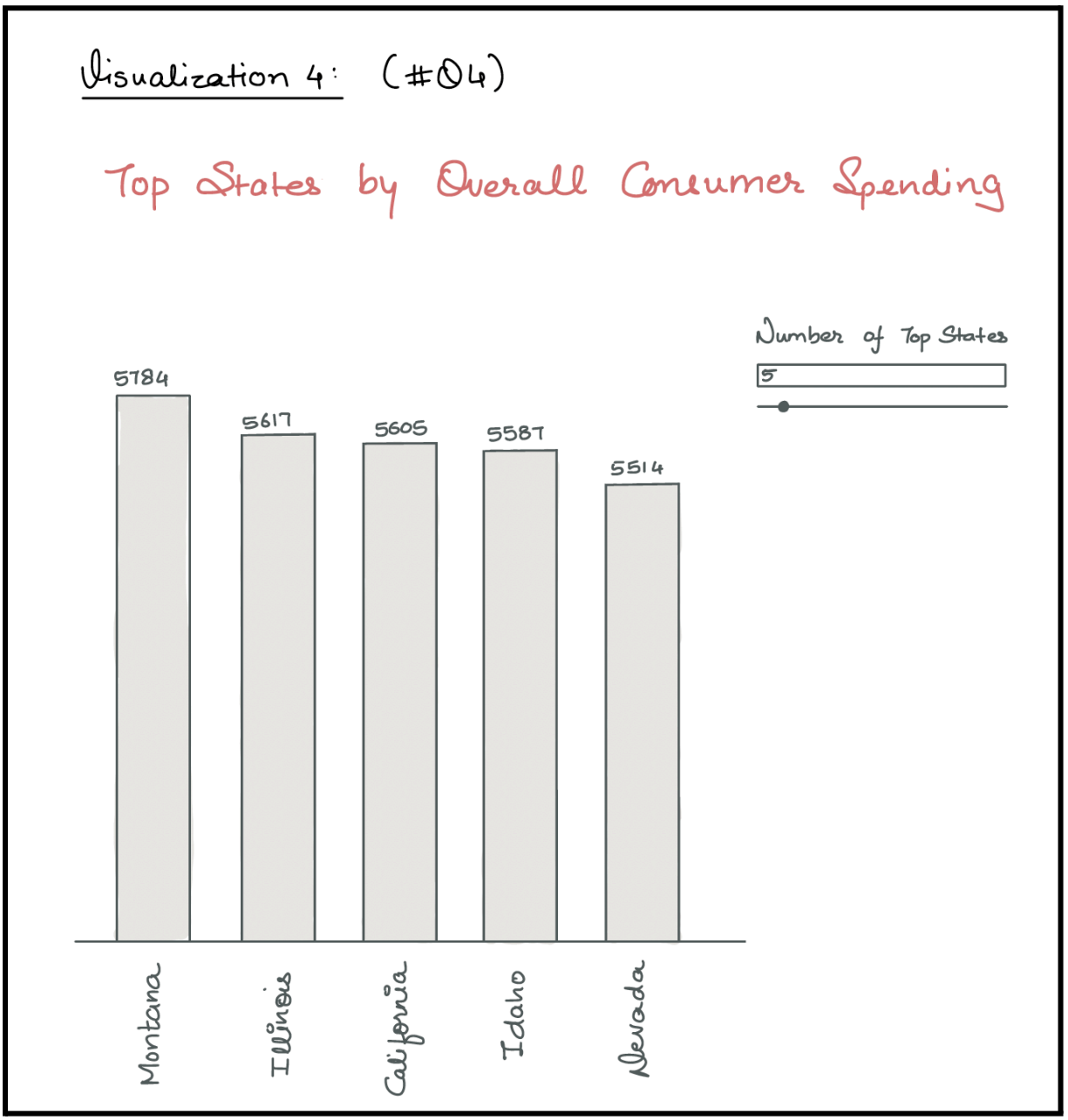
Pre - Attentive Attribute - Categorical Color , Spatial Position



Plot 4 (Question 4) - Which states have the highest overall consumer spending in terms of purchase amount?

Plot - Bar Chart

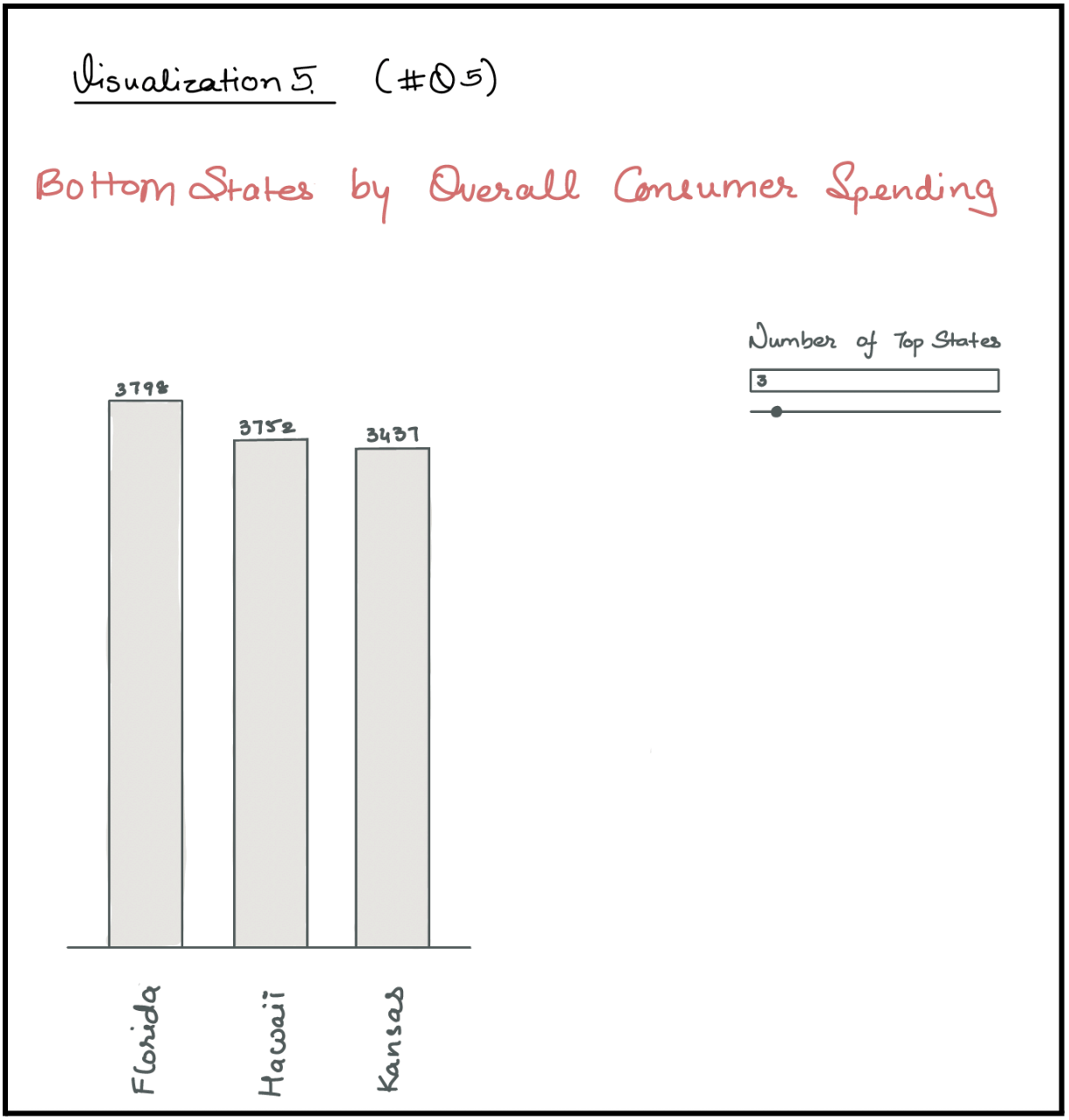
Pre - Attentive Attribute - Length , Position



Plot 5 (Question 5) - Which states have the least overall consumer spending in terms of purchase amount?

Plot - Bar Chart

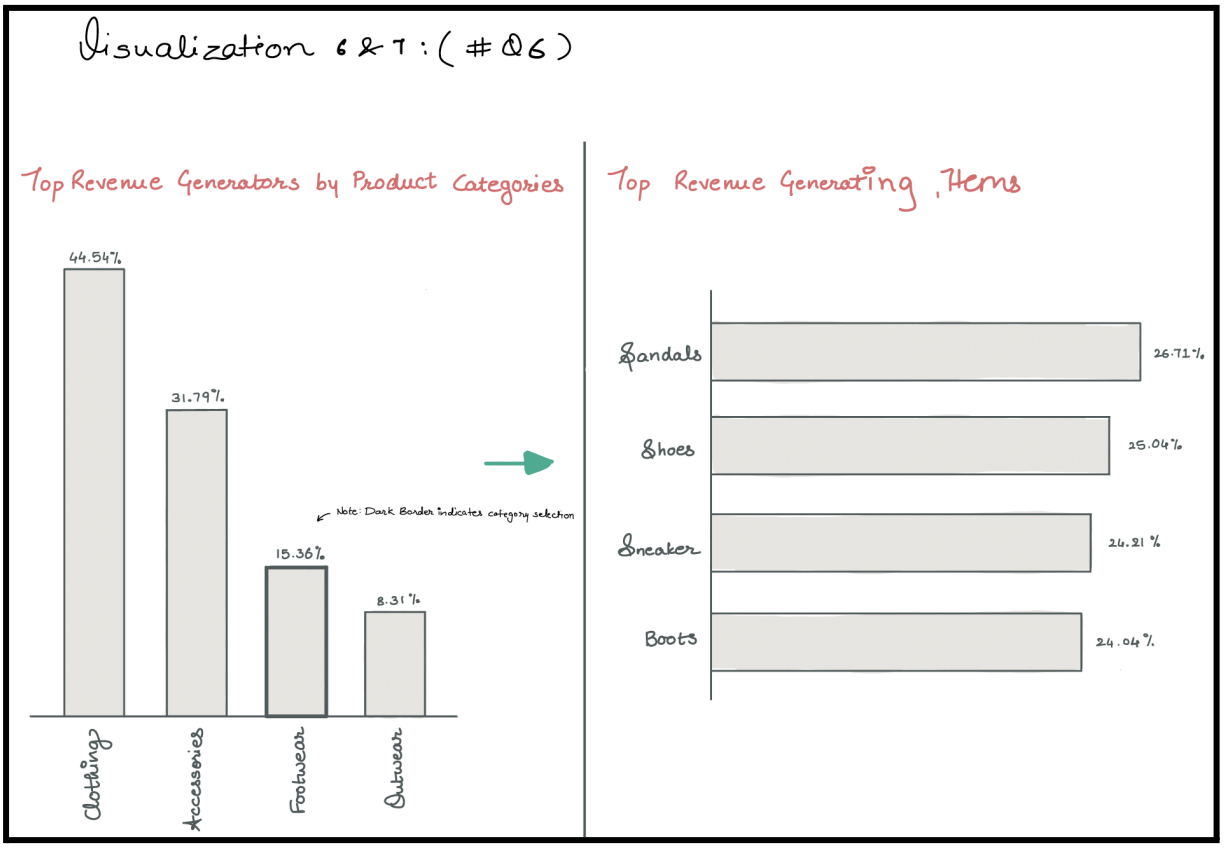
Pre - Attentive Attribute - Length , Position



Plot 6 (Question 6) - Which product categories and their corresponding items generate the highest revenue?

Plot - Bar Chart

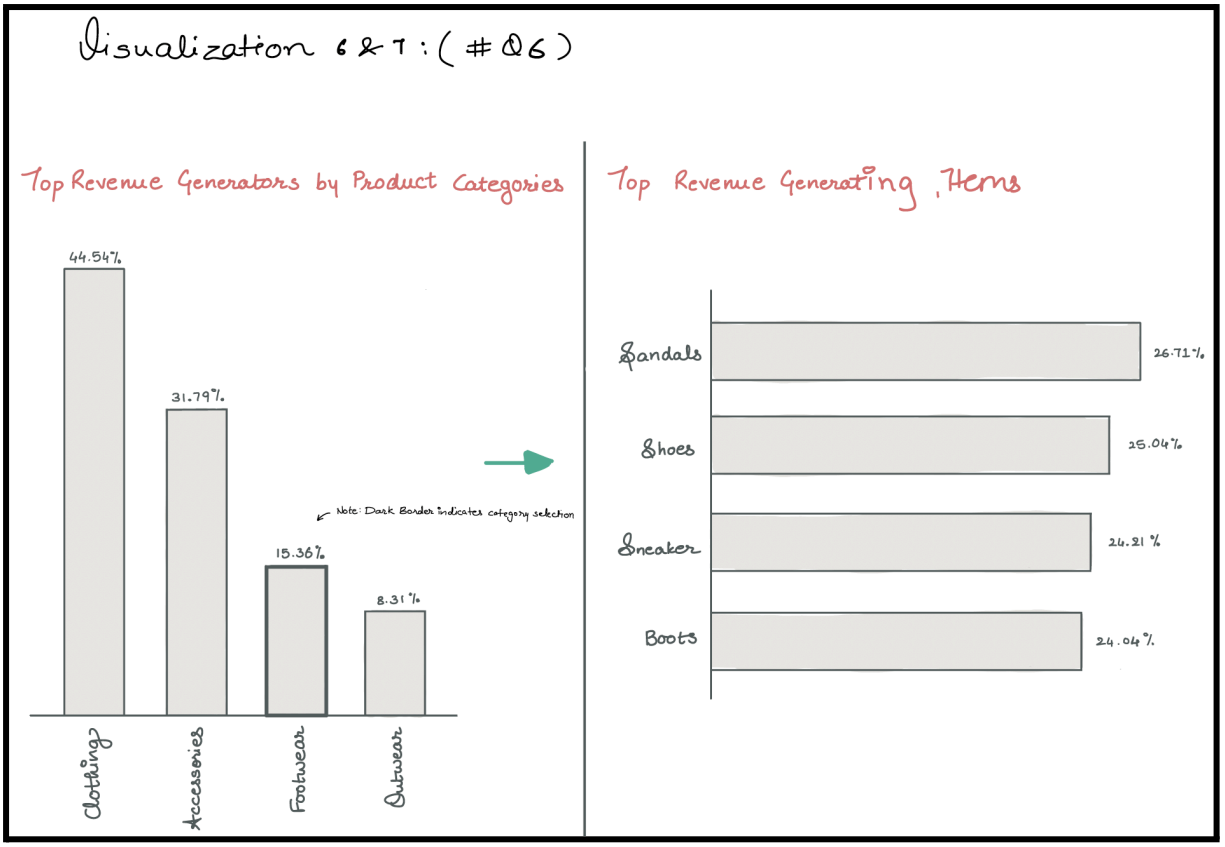
Pre - Attentive Attribute - Length , Position



Plot 7 (Question 6) - Which product categories and their corresponding items generate the highest revenue?

Plot - Horizontal Bar Chart

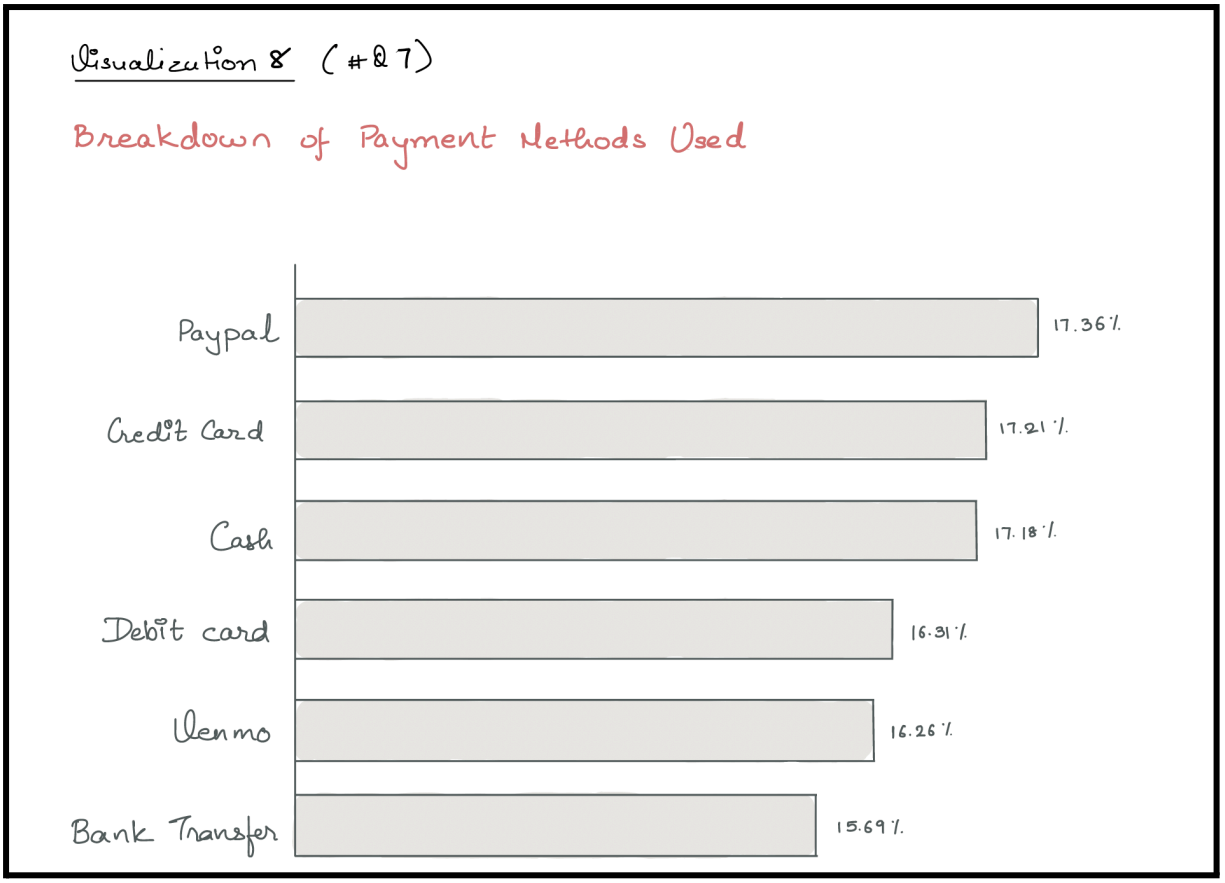
Pre - Attentive Attribute - Length , Position



Plot 8 (Question 7) - What are the most commonly used payment methods among consumers?

Plot - Bar Chart

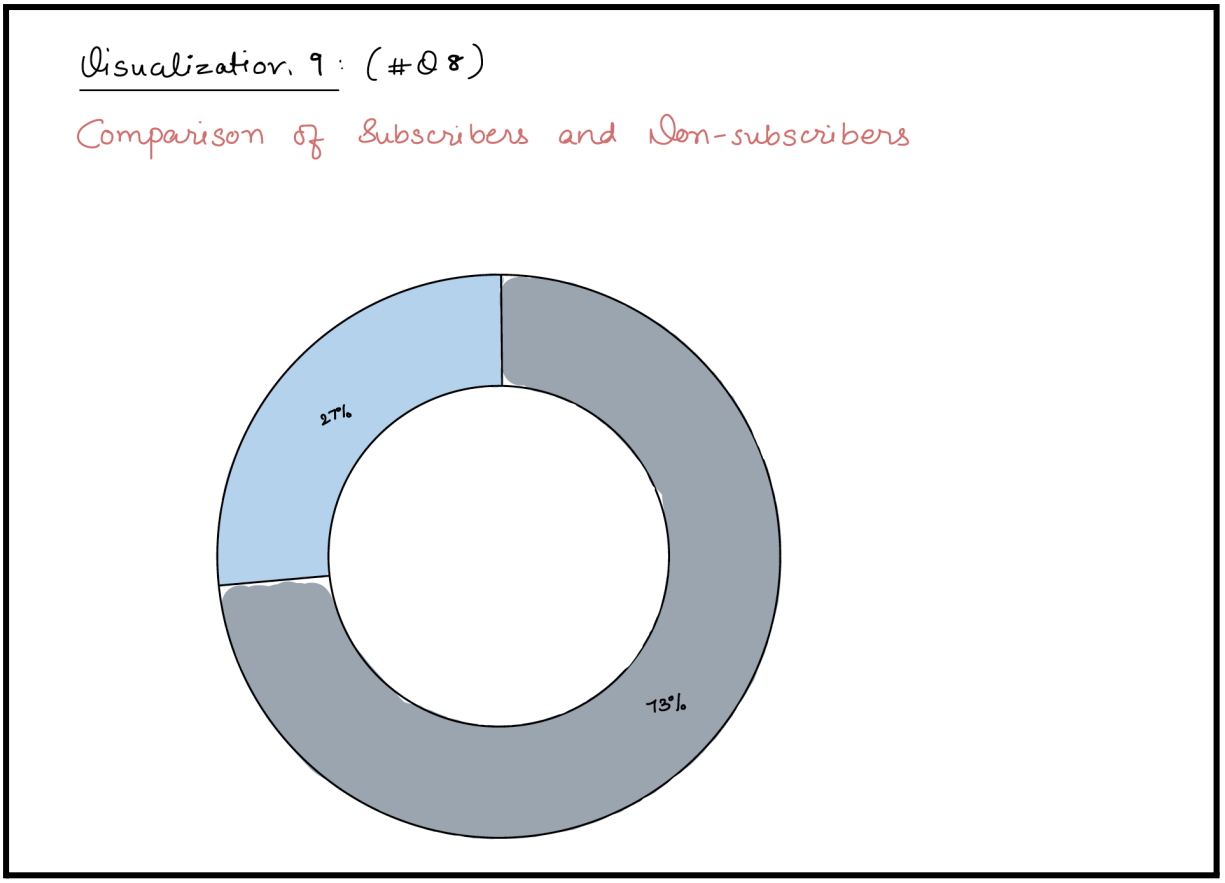
Pre - Attentive Attribute - Length , Position



Plot 9 (Question 8) - How does the number of subscribers compare to non-subscribers?

Plot - Donut Chart

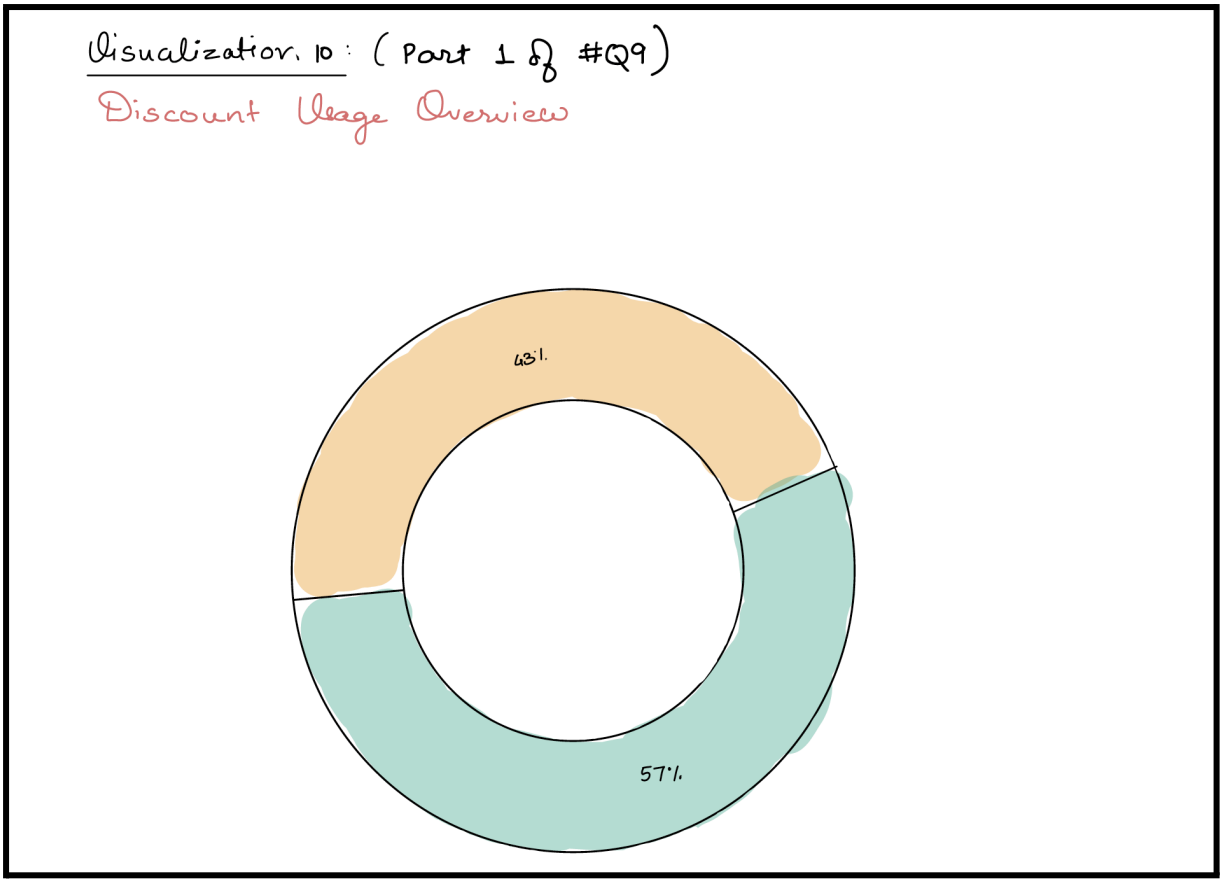
Pre - Attentive Attribute - Color, Size/Area, Orientation



Plot 10 (Question 9) - How do discount applications vary by product category and shipping method?

Plot - Donut Chart

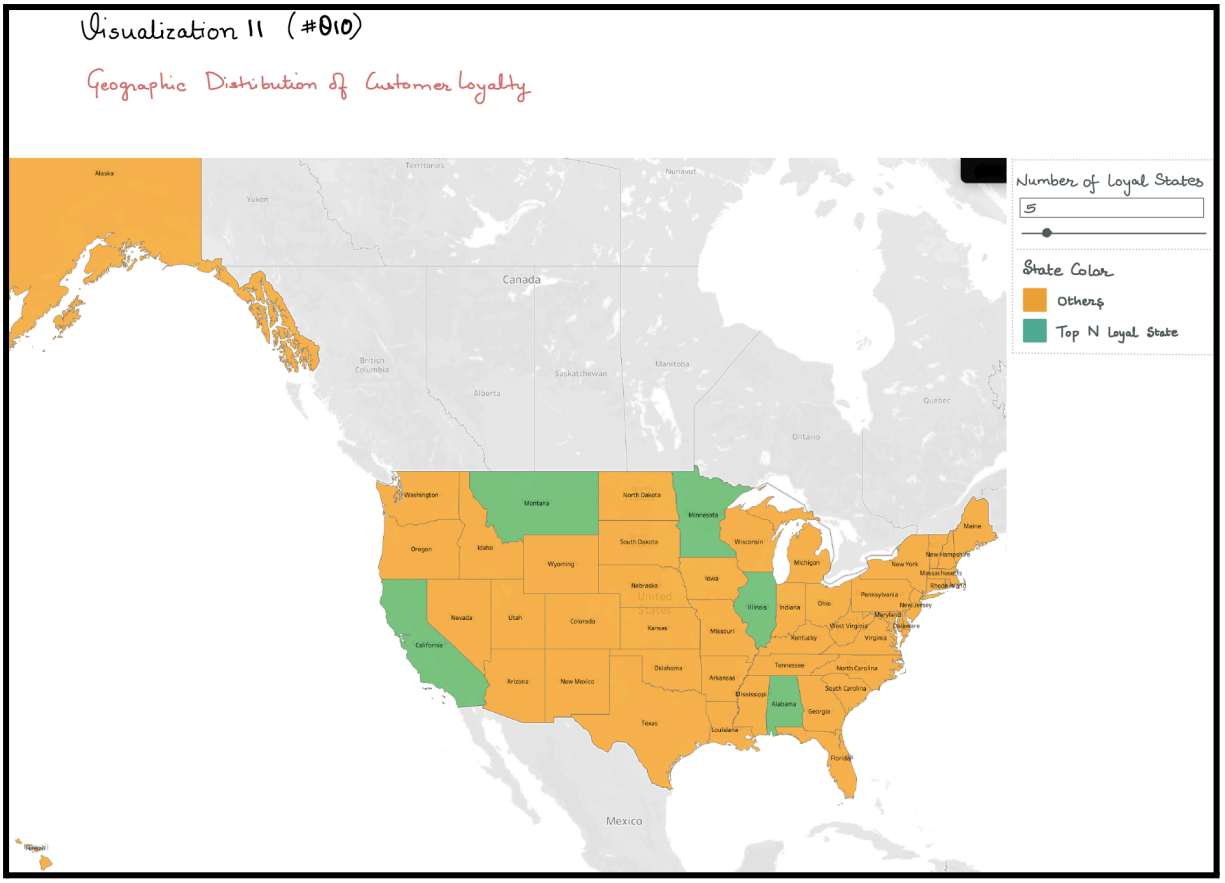
Pre - Attentive Attribute - Color, Size/Area, Orientation



Plot 11 (Question 10) - In which states are loyal customers most concentrated based on their previous purchase activity?

Plot - Choropleth

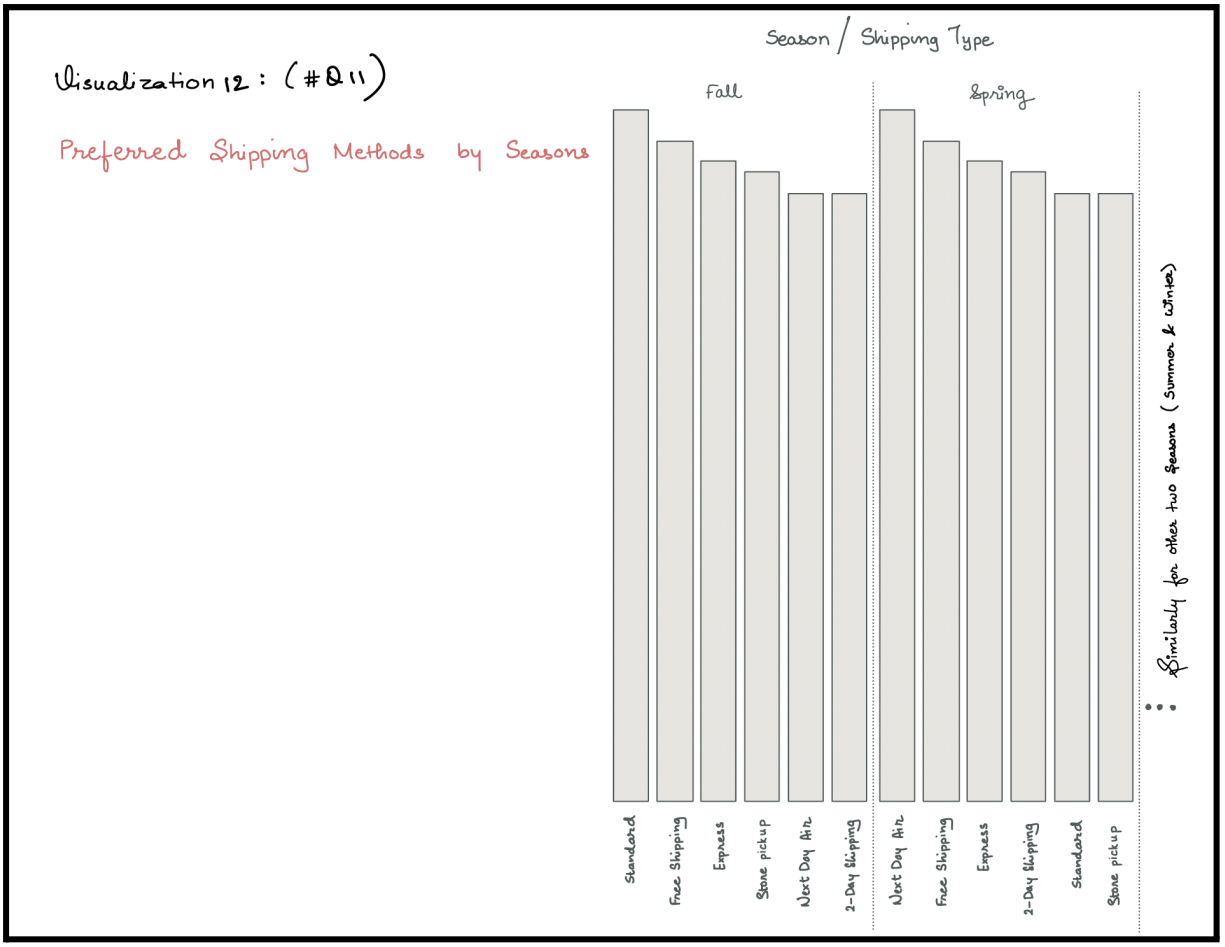
Pre - Attentive Attribute - Categorical Color , Spatial Position



Plot 12 (Question 11) - How do customer preferences for different shipping types vary across seasons?

Plot - Side-by-Side Bar Chart

Pre - Attentive Attribute - Length, Position, Grouping



Plot 13 (Question 12) - What are the preferred shipping methods for different product categories?

Plot - Horizontal Stacked Bar Chart

Pre - Attentive Attribute - Length, Position, Alignment

Visualization 13 : (#Q12) (Part 2 of #Q9)

Preferred Shipping Methods by Product Category

Similarly for other two categories (Footwear & Outwear)

Section 5: Dashboard Interactivity

In order to enhance the user experience and provide dynamic insights into consumer behavior across different states and product categories, several interactivity controls have been integrated into the Tableau dashboard. These controls allow users to tailor the visualization outputs to their specific needs and interests.

1. Interactive Slider for Top States by Total Purchase Amount

A calculated parameter called 'Number of Highest Consumer Spending States' will be created to display the top states by total purchase amounts in the ‘Top States by Overall Consumer Spending’ visualization. This feature is represented by a slider, allowing users to select the number of top-spending states they wish to view. The slider accepts integer values ranging from 1 to 60, with a step size of 1. Adjusting the slider dynamically updates the visual representation in the plot, providing a flexible view of the data based on user selection. Moreover, in order to add interactivity to the slider, we will add a Top N filter on the ‘Location’ attribute. This parameter is coupled with a Filter (By Field) based on the Location (states) attribute, which selects the top N states according to the sum of the purchase amount.

2. Interactive Slider for Bottom States by Total Purchase Amount

A calculated parameter called 'Number of Lowest Consumer Spending States' has been created to display the top states by total purchase amounts in the ‘Bottom States by Overall Consumer Spending’ visualization. This feature is represented by a slider, allowing users to select the number of top-spending states they wish to view. The slider accepts integer values ranging from 1 to 60, with a step size of 1. Adjusting the slider dynamically updates the visual representation in the plot, providing a flexible view of the data based on user selection. Moreover, in order to add interactivity to the slider, we will add a Bottom N filter on the ‘Location’ attribute. This parameter is coupled with a Filter (By Field) based on the Location (states) attribute, which selects the bottom N states according to the sum of the purchase amount.

3. Filter Action for Product Categories and Items

To facilitate an interactive exploration of top-selling categories and items, a Filter Action has been implemented. The source and target fields are based on ‘Category.’ This control links the 'Top Revenue Generators by Product Category’ plot (Source Sheet) with the 'Top Revenue-Generating Items' plot (Target Sheet). By selecting a category in the first visualization, the display in the second visualization updates automatically to show the top-selling items within the chosen category. This filter action allows for seamless navigation between related data points, enabling users to drill down into specific areas of interest with ease.

4. Loyal Customer States Visualization

For visualizing states with the most loyal customers in the ‘Geographic Distribution of Customer Loyalty’ plot, a calculated parameter named 'Number of Loyal States' will be utilized. This parameter is adjustable via a slider, ranging from 1 to 55 with a step size of 1. We will also use a Top N Filter on this calculated field to display the top loyal states based on the sum of ‘Previous Purchases.’ It controls the display on a choropleth map, highlighting the top-ranking loyal states in green and coloring the rest in orange based on the user’s input. For instance, setting the slider to 5 will color the top 5 states green, distinctly marking them as areas with high customer loyalty.

5. Interactive Visualization of Seasonal Product Category Leaders Across States

The visualization, Seasonal Product Category Leaders by State, highlights the top-selling product categories across states based on seasonal trends. Using the calculated field Top Category by State, the states are dynamically colored by the product category with the highest sum of 'Previous Purchases' for the selected seasons. The calculation considers both category and season context to ensure accurate results. An interactive ‘Season’ filter has been incorporated, allowing users to select one or multiple seasons using multi-select checkboxes. This interactivity empowers users to explore how category leaders shift across different states and seasons, providing an intuitive way to analyze seasonal consumer behavior trends.

6. Interactivity for Exploring Discount Usage and Shipping Preferences

The dashboard combining Plot 10 and Plot 13 leverages Tableau's Action Filters to provide interactive insights into the relationship between discount usage and shipping preferences across product categories. Users can explore this connection through two visualizations: a bar chart showing "Preferred Shipping Methods by Product Category" and a donut chart titled "Discount Usage Overview." Selecting a specific product category or shipping method in the bar chart dynamically updates the donut chart to display the percentage of orders within that category where a discount was applied versus not applied. This interactivity is achieved by linking the bar chart as the source sheet and the donut chart as the target sheet, enabling seamless exploration of consumer behavior trends. The source fields and target fields will be ‘Category’ and ‘Shipping Type.’

Section 6: References

Mural Link:
https://app.mural.co/t/ift533projectgroup139235/m/ift533projectgroup139235/1731445252262/1fcf456867882fdef2a0ba00b60c8182abec13af?sender=u77ee4221837d74d15bf70496