ShopScope USA

E-Shopping Explorer: Interactive visualizations of US consumer trends

———————

June 2024

BIRLING Romain

CHRISTOPHE Hugues

ALLOCIO Jeanne

INTRODUCTION

In this process book, we outline the development of our project for the Data Visualization course at EPFL. The initial phase of our project involved selecting the dataset we wanted to visualize. After exploring various ideas, we decided to work with a detailed database on Consumer Behavior and Shopping Habits in the USA. We then pondered what interesting information we could extract from this dataset and what kinds of visualizations we could create. Our preliminary research revealed that most existing analyses and visualizations were limited to simple graphs. Therefore, we aimed to take our analysis a step further by conducting a more thorough, engaging, and insightful exploration. We planned to make our analysis more interactive, focusing on highlighting correlations and connections between different features. Once our project topic was clearly defined, we delved into the dataset to fully understand the data available to us and to verify that there were no issues with the dataset. This allowed us to start considering various visualizations we wanted to include in our project.

DATASET AND DATA ANALYSIS

We found our dataset on Kaggle. After detailed preprocessing and numerous analyses, we realized that the dataset was very clean and required minimal modifications or adjustments. This discovery allowed us to quickly focus on the ideas we wanted to highlight and the visualizations we wanted to implement.

CHOOSE OF DATA VISUALIZATIONS

After choosing the datasets and doing the preprocessing of the data, we brainstormed about what kind of visualizations we want to make. Considering the story we wanted to tell, we decided to create our website and our visualizations this way:

- The central feature of our project will be an interactive map that displays in a color-coded way insightful information that will be sectioned using filters (ex: display in a color-coded way the average number of orders per state)

Users will also be able to interact with the map in two ways:

- o Hover Interaction: As users hover their mouse over any state on the map, a tooltip will appear. This tooltip will provide succinct insights into the state's specific consumption patterns: the average number of orders, the average amount spent or the most ordered category of item.

- o Click Interaction: Clicking on a state will select this state for the next visualizations. The next part will host a deeper analytical dive into the specific characteristics of consumption in that state. It will include detailed breakdowns of consumer demographics, preferred product types, and purchasing trends, alongside comparisons with national averages.

- Deep dive page : Once you have clicked on a map state, you will the go to the next part for further analysis. This part will take the form of a dashboard with several graphs. You can filter to display 2 different dashboards:

  - o Customer type analysis: this dashboard will include:
    - A pie chart displaying the male and female distribution.
    - A pie chart displaying the distribution of payment methods used.
    - A diagram displaying the age distribution.
    - A bar chart displaying the subscription status.
    - A bar chart displaying the shipping type distribution.

  - o Product preference analysis: this dashboard will include:
    - A chart showing the distribution of spending power across categories.
    - A bar chart displaying the distribution of product categories purchased.
    - A line chart showing seasonal buying trends.
    - A scatter plot showing the relationship between age and average spending per order.
    - A pie chart displaying the item category distribution.
    - A pie chart showing the ordered item size distribution

We also thought that to enhance user experience, we will include a filter bar on the left of the dashboard and above the map to allow users to customize the data displayed on the graphs.

- In addition, at the bottom of the page, we thought about creating a section for comparing states. It's interesting to be able to study a report in detail, but it's also very interesting to be able to compare customer types and product preferences between reports.

IMPLEMENTATION OF DATA VISUALIZATIONS

After carefully considering everything we wanted to visualize, we began the implementation phase. The first step was creating the website's skeleton using React. We designed the layout that we envisioned, starting with a homepage and introduction. Scrolling down, users first encounter our map, which serves as our initial interactive visualization, followed by the dashboard below.

INTERACTIVE U.S MAP

Our first visualization is a comprehensive map of the USA, displaying each state. Initially, we planned to implement a simple map showing the average number of orders per state. However, we thought it would be more useful for users to have the flexibility to choose what data they wish to view on the map. Consequently, we added a filter that allows users to select between viewing the average number of sales, the total purchase amount, or the average purchase amount. Once a metric is selected, the map uses a color legend to display these values, facilitating an easy comparison between states.

Hovering over a state triggers a tooltip that shows the state's name and the currently visualized metric value. Furthermore, this map also serves as a tool for further visualizations. Users can select states they wish to analyze more deeply. This selection process was designed with the idea of encouraging users to investigate why certain states show higher sales volumes and to explore the characteristics of these states in more detail in the dashboard.

States selected by the user are highlighted in blue, while unselected states are shown in red. There is also a "Deselect All" button at the bottom right of the map, allowing users to clear their selections in one click.
Once the state(s) have been selected, the user can scroll down to begin a more detailed analysis of customer and product characteristics within those states.

INTERACTIVE DASHBOARD

The dashboard is our second principal visualization. We decided to split the visualization into two parts. On the left, we created a sidebar that houses all the filters users can adjust. Adjacent to this, we constructed the dashboard composed of the graphs we detailed earlier. We initially planned to include two distinct dashboards: one focusing on customer types and another on product preferences. Initially, we considered adding a button in the filter bar that would allow users to select which dashboard they wished to view. However, we ultimately preferred the idea of using tabs to switch between the dashboards, leaving the left sidebar solely for filters applied to the graphs.

Regarding the graphs, we implemented exactly the graphs we had planned. Each graph is interactive; hovering over a graph displays a tooltip with a description of the visualized value to simplify the reading experience. Furthermore, all our graphs are interactive and thus linked to the filters on the left side of the dashboard.

As for the filters, we initially thought to include the following: gender, age, item category, season, subscription status, shipping method, and payment method. However, after several reviews, we realized that filtering by shipping method and payment method did not add any interesting insights. Instead, we overlooked an interesting filter—clothing size—which we then decided to include. Ultimately, we implemented the following filters: gender, age, item category, season, subscription status, and size.

Additionally, at the top of the filter bar, we decided to display the U.S. states that had been previously selected on the map as a reminder of what the user is currently analyzing.
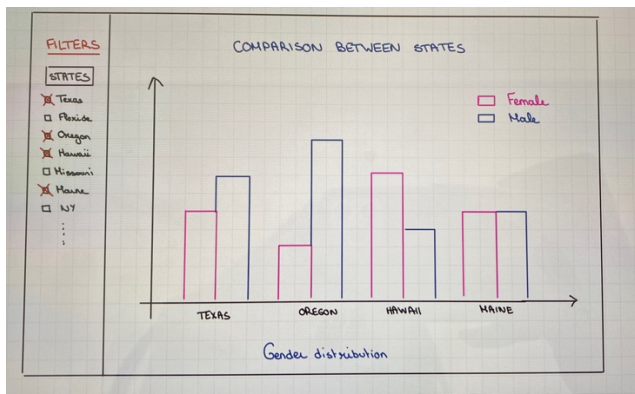
Sketch in Milestone 2
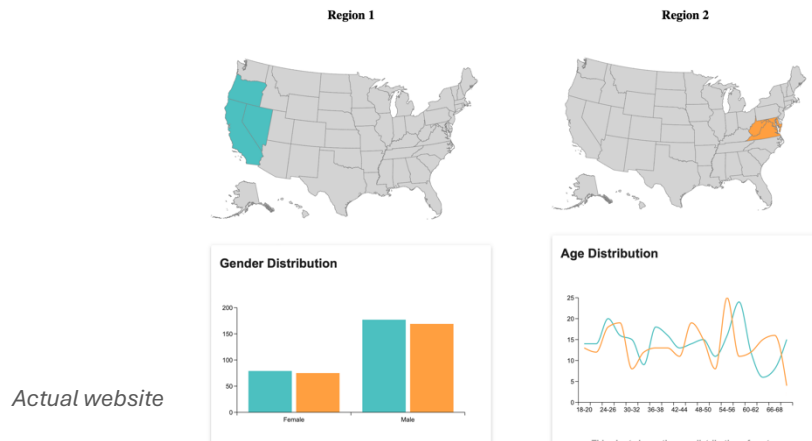

Actual website
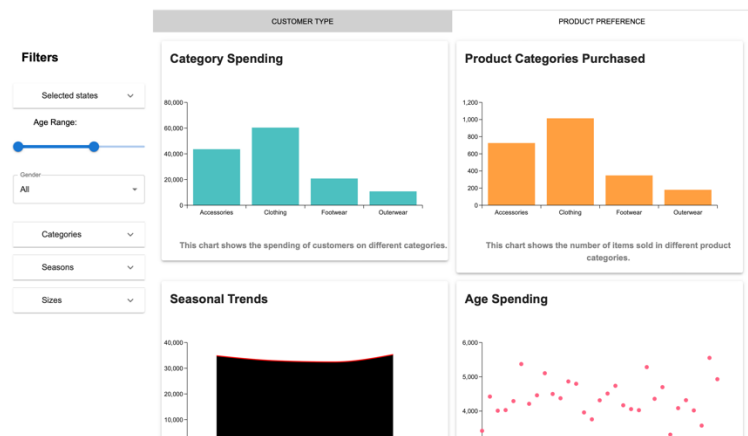
STATE-TO-STATE COMPARISON

Our final major visualization allows users to select two regions—each comprising one or several states—to compare them directly. We recognized that it would be insightful for users not only to analyze the performance of individual states but also to compare them against each other.

This visualization consists of two maps that enable the selection of states for each region to be analyzed. Below these maps, you'll find three critical metrics displayed in graphical form: a comparison of gender distribution across the two regions, a comparison of age distribution, and finally, a comparison of spending by category in each region.

In conclusion, we successfully implemented the visualizations we wanted and had described in Milestone 2.

Sketch in Milestone 2

Actual website

WEBSITE IMPROVEMENT

Once all visualizations were integrated into our website, our final step was to enhance the site's visual appeal. To achieve this, we designed an attractive homepage by incorporating some color schemes. We added titles and detailed descriptions for each section to guide users through their exploration of the site seamlessly. Finally, we enhanced the aesthetics and functionality of our dashboard graphs using the MUI library, making them not only more visually appealing but also easier to analyze.

TOOLS USED

Here are the tools we used, and we will use for our data and for the visualizations.

Data Analysis:
- Python

Visualizations:
- React
- React simple map

- D3.js
- Nextjs
- Papa parse library
- MUI x-charts (for charts)
- Material Ui Kit

CHALLENGES THAT WE FACES

One of the main challenges we faced was effectively linking the data from our CSV file to our visualizations. Initially, we tried several different techniques, starting with creating pre-sorted lists from our CSV and reading them for each graph. However, we soon realized that this approach was not optimal. We then explored other methods and eventually succeeded in integrating the data by using an auxiliary library, Papa Parse.

Another challenge we encountered during the visualization process arose when creating the USA map. We struggled to simultaneously display the metric selected by the user and indicate which states had been selected by the user. It was necessary to find a clear way to differentiate between selected and non-selected states, especially since the map was already color-coded to represent different metrics. We had to carefully balance the visual elements to ensure that both pieces of information were easily discernible and did not interfere with each other.

PEER ASSESMENT

Student's name	SCIPER	Tasks
BIRLING Romain	310054	readme, dashboard, visual improvement of charts, map, screencast
CHRISTOPHE Hugues	311252	site skeleton, link between data and dashboard, dashboard, map
ALLOCIO Jeanne	311844	data processing, readme, map, site visual improvement, process book