DSC 465 - Data Visualization

(Professor. Eli Brown)

Analysis and Visualization of Global Superstore

Final Report

By

Viz Champs

Luuk Alfons Wassenaar
Nawaaz Sharif
Nandhini Gulasingam
Anisha Patel
Saurin Patel

Table of Contents

Note: Please grade the visualizations listed under III. Visualizations, Analysis and Discussion

I. Introduction

News reports and market research sites like Shopify[1] predict the global ecommerce market is expected to total $5.5 trillion in 2022. Online shopping like the Global Superstore has become a routine part of everyone's lives. Therefore, we decided to focus on a dataset that helped us understand the online retain business.

The Global Superstore dataset we used for our project was obtained from Kaggle.com. Global Superstore specializes in selling furniture, technology, and office supplies all around the world for corporate, consumer and home office segments. This dataset had order information for orders placed by consumers from 2011 to 2014 including returns. The dataset has three different tables; the first table had all the information of products, customers, orders, categories, sub-categories, segments and various other variables. The second table had the returns of the products. Third table had information about the managers for each region. Our dataset included was a mixture of the date/time, geographical, categorical, and numerical variables. Below are variable and record counts or each table:

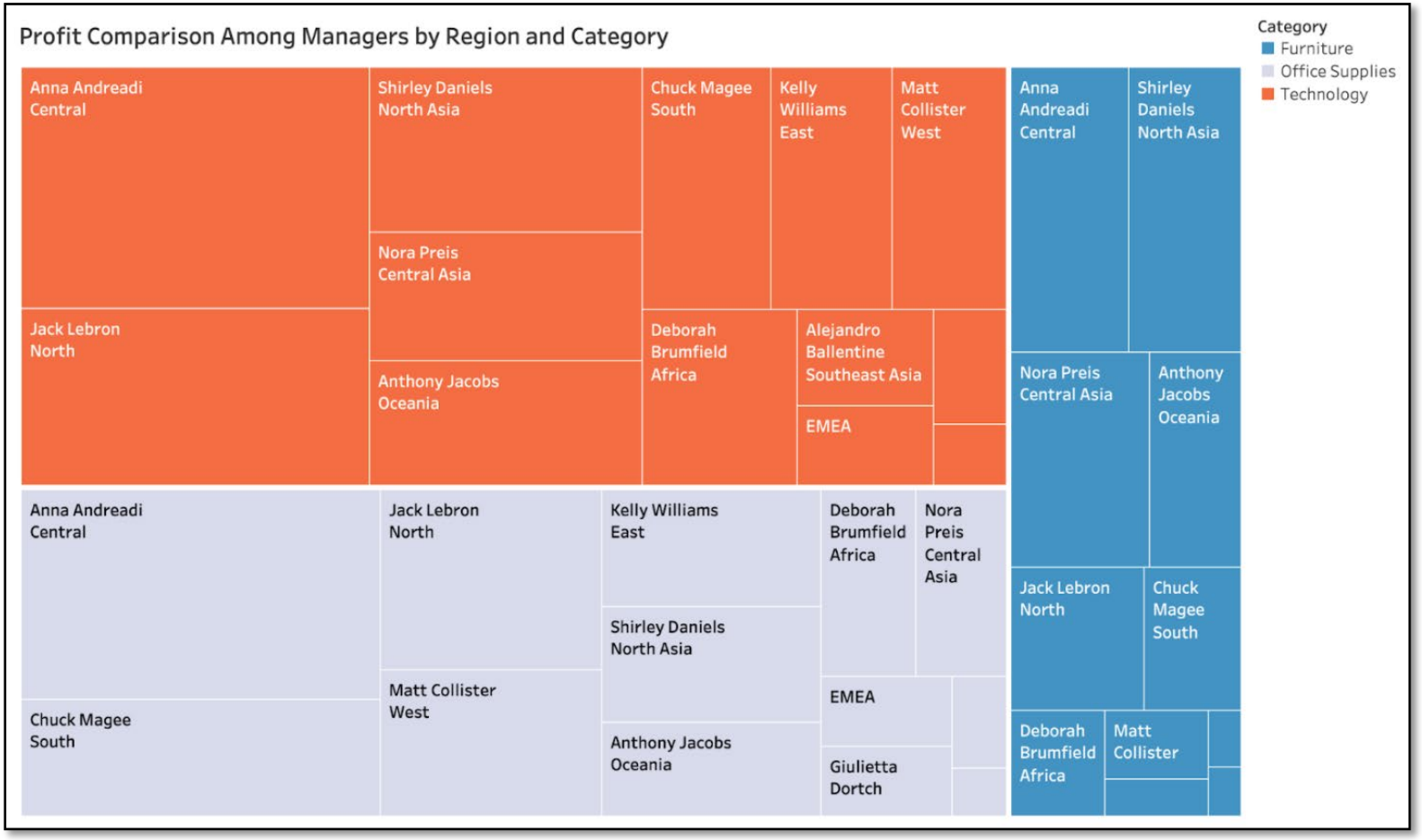
Tables	Variables	Observations	Variables
Orders	24	51,291	24
Returns	3	1175	3
People	2	14	2

Our goal was to come up with different avenues to help the Global Superstore increase their profitability. To this end, identified how sales and orders played a role in profits including identifying the trends by segments and categories for countries. We analyzed returns to understand how to minimize returns to increase profits. We found the best performing managers so we can learn best practices and their strategies that we can be implement in underperforming different regions.

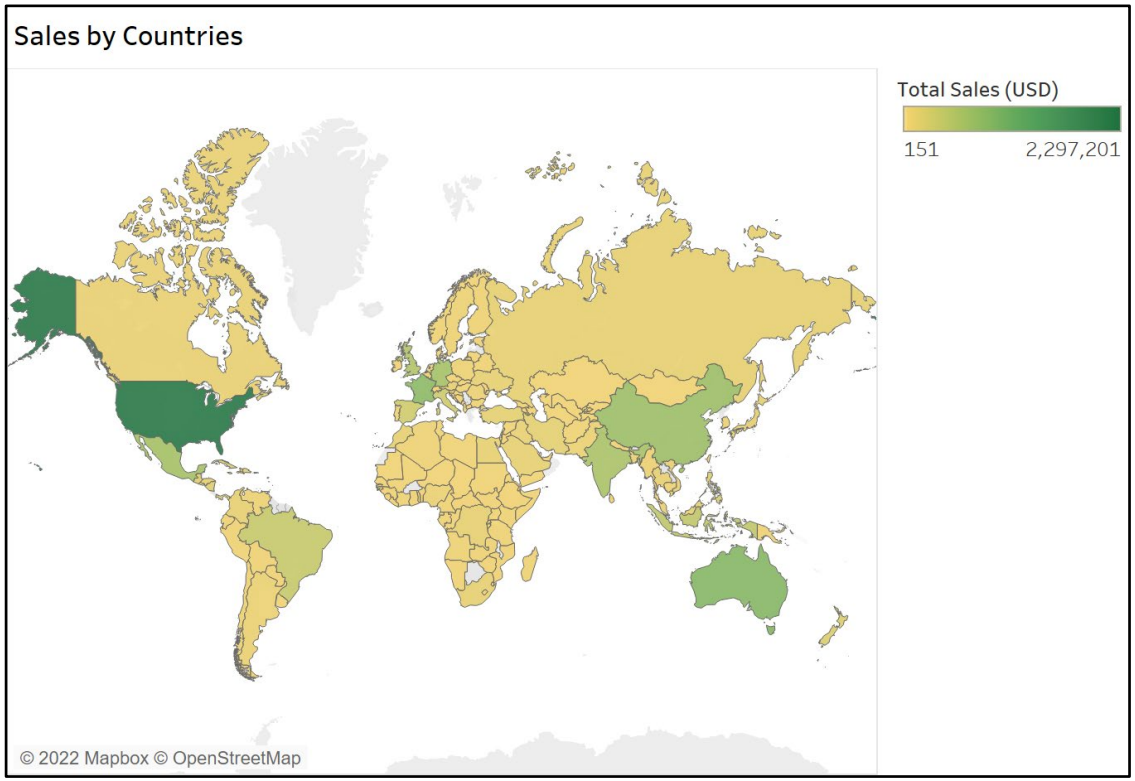
[1] Global ecommerce: Stats and trends to watch to succeed internationally (2022). Shopify Plus. (n.d.). Retrieved March 16, 2022, from https://www.shopify.com/enterprise/global-ecommerce-statistics#2

II. Exploratory Analysis

The exploratory visualizations that were part of milestone-2 and also the following visuals helped us discover hidden patterns.



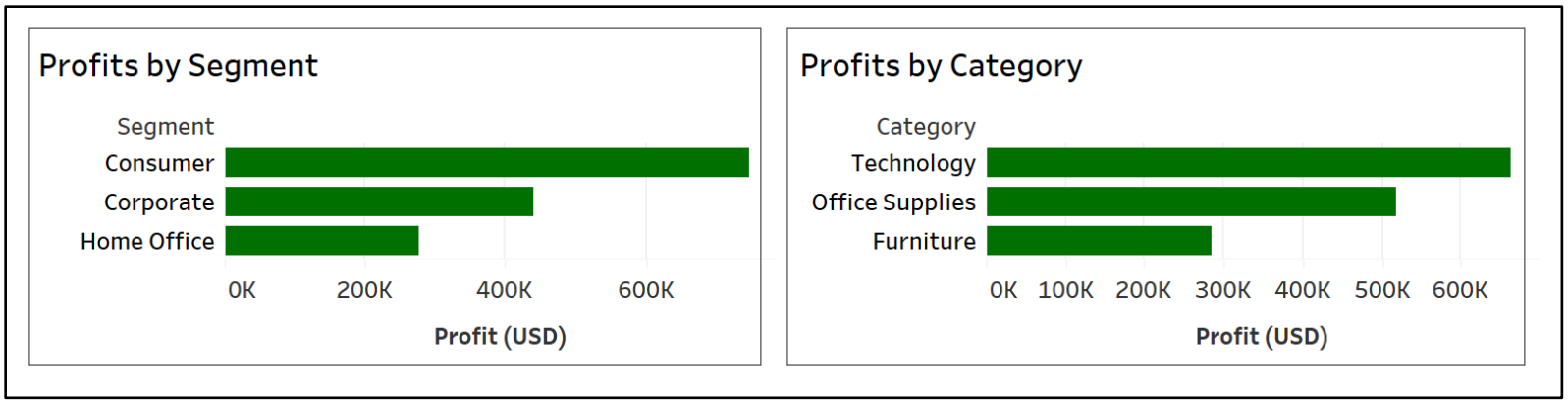
One of the first visualizations can be seen below. This Tree Map created with Tableau focused on the manager section of our research. It visualizes a 4-year profit report for each manager, region, and product category. The size of the boxes represents the amount of profit, where the different product categories are color coded. This initial visualization allowed us to see that there is a major difference in profit among the managers and their region, which we wanted to explore in more detail. Besides this, we could also see profit/market share of product categories (Furniture, Office Supplies, and Technology), and identify which categories brought the most profit and which categories played a small role in the overall profit of the Superstore.



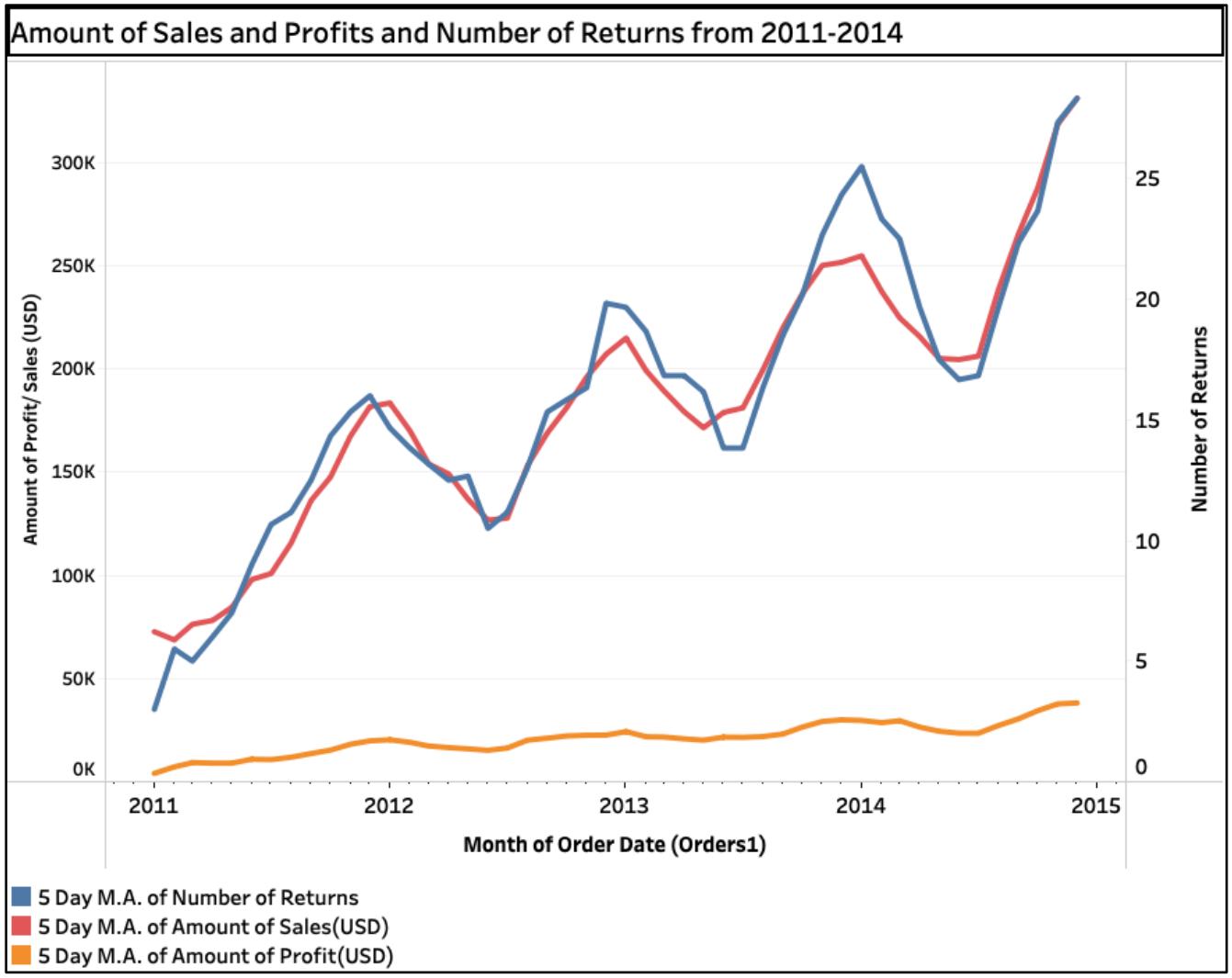
The geospatial map of sales indicated the sales trends by countries with USA leading the pack with 2.2 million USD in sales. India, China, Australia, France, Germany, UK and Mexico also showed high sales of 250K USD or more. The sales revealed only part of the story, as there were many countries such as Turkey, Nigeria, Netherlands, Honduras, Pakistan, etc. that had high sales with low profits or losses.

This prompted us to not only compare the correlation between sales and profits, orders and profits but also what role the three segments (Consumer, Corporate, Home Office) and the three Categories (Furniture, Office Supplies, Technology) played on sales and profits for our explanatory visualizations.

As a result, two explanatory visualization were created – dashboard to show Sales and Profit, and interactive dashboard to show Profit Profile by Countries.



The profits for each segment (Consumer, Corporate, Home Office) and each of the three categories (Technology, Office Supplies, Furniture) revealed the profit profile for each of these segments and categories were different. Therefore, we wanted to look at the segments and categories for sales, profits and returns by various countries and world to see if there were any trends or patterns emerging during our explanatory analysis.

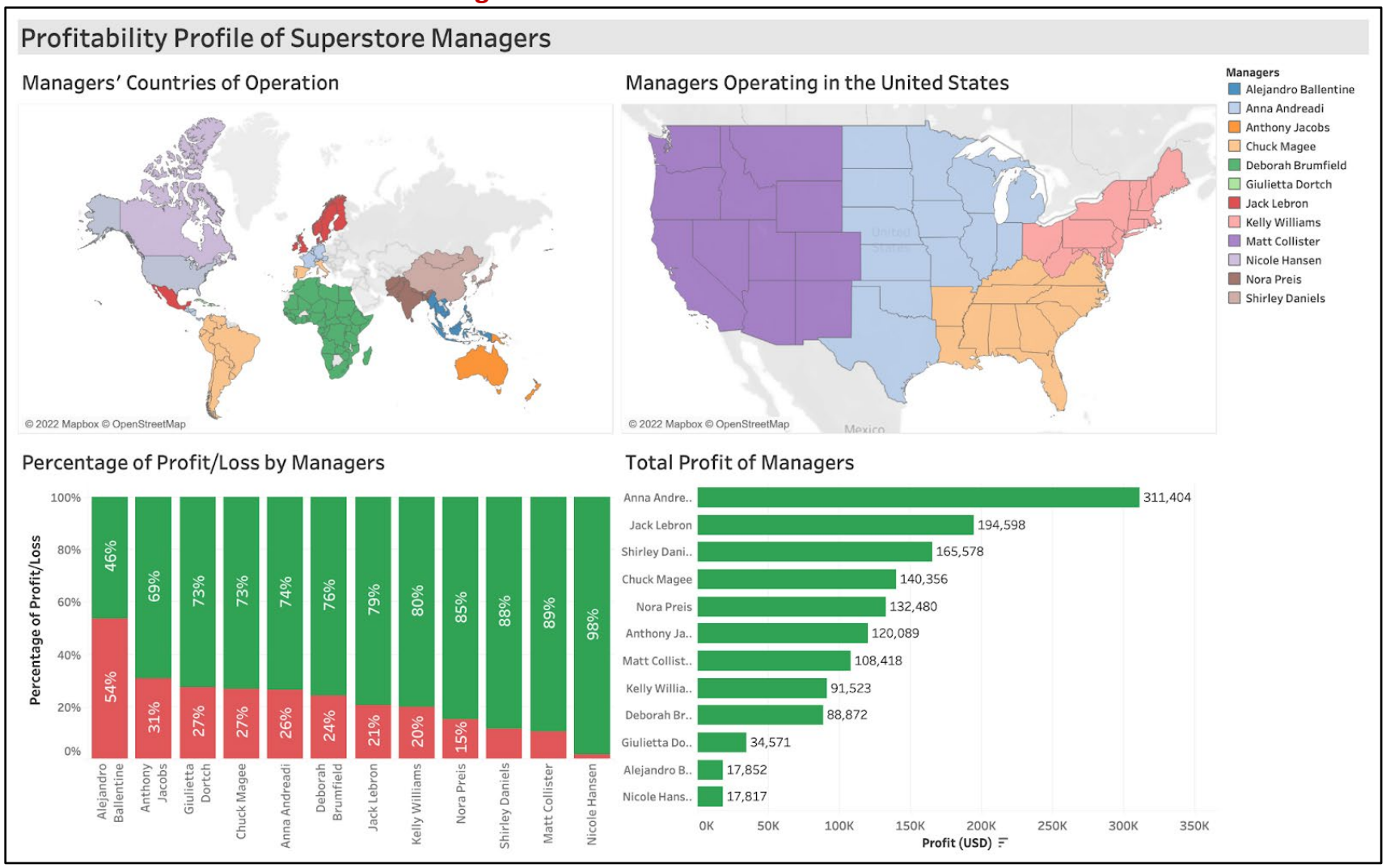


The above explanatory analysis was a time-series analysis of the amount of Sales, Profit and Number of Returns from 2011 to 2014. The amount of Profit and Sales on the left y-axis and Number of Returns on the right y-axis. The profit, sales and number of returns data was smoothed by using the 5-day Moving Average. All three measures increased from 2011 to 2014. At the beginning of 2012, 2013, and 2014 all three measures started decreasing, but increased from mid-year the same year. Sales and Number of Returns following a similar trend whereas Profit had a steady linear growth. This provided a staring point to analyze the returns and its role on increasing profits.

III. Visualizations, Analysis and Discussion

Our group come up with several exploratory visualizations before coming together in our third zoom meeting. This allowed us to see the different components of the dataset and helped brainstorm and discover important but more hidden patterns. Based on this, our goal was to help Global Superstore increase profits. We divided the dataset into five major sections to provide a detailed visual analysis. These sections included: (1) Profit Profile of Managers, (2) Comparison of Sales and Profit, (3) Profit Profile by Categories, Segments, and Profitable Months, (4) Profit and Returns by Countries, and (5) Profits and Returns of Top 4 Countries for Returns

Visualization-1: Profit Profile of Managers

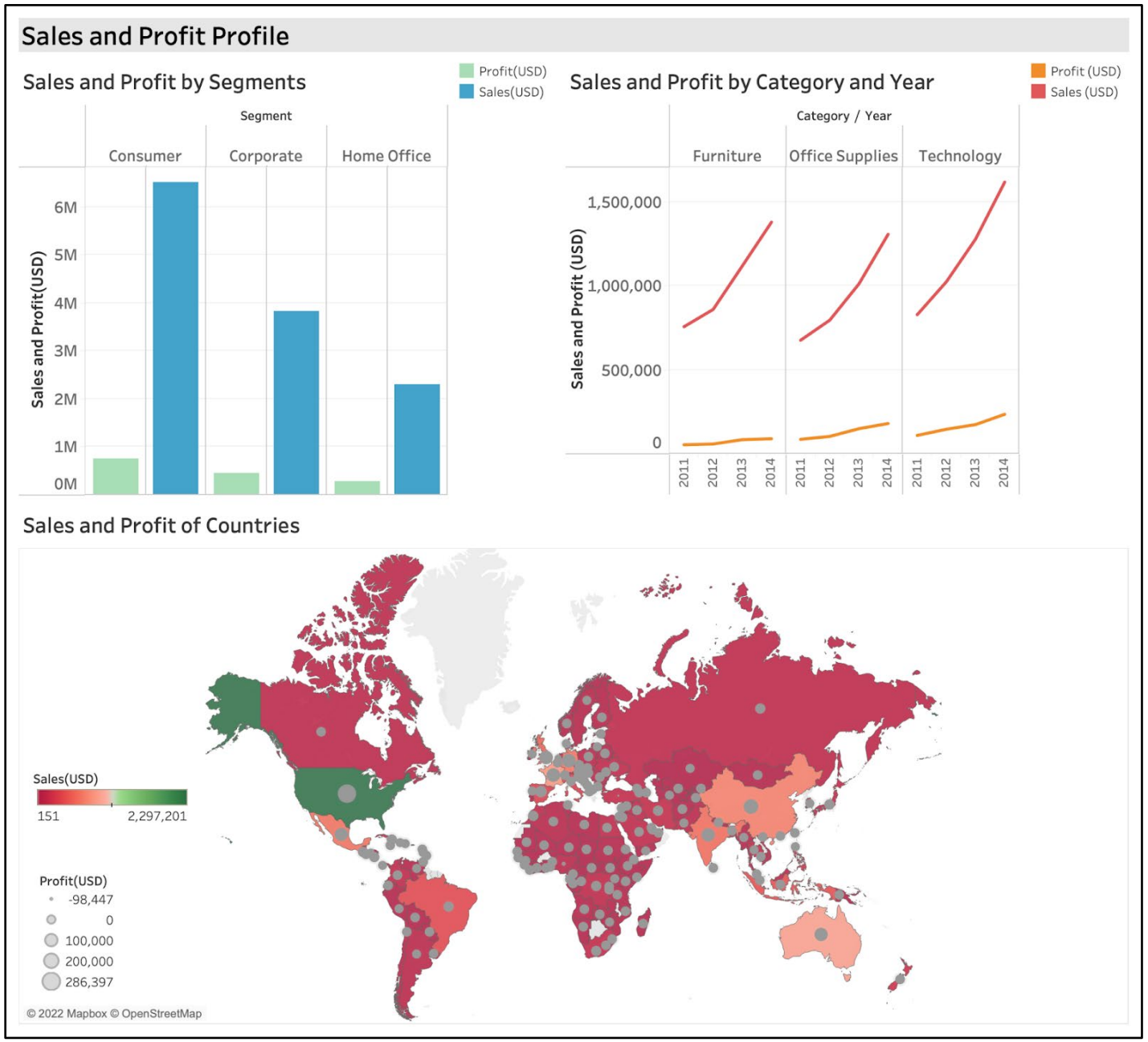


This is a composition of several visualizations which show the profitability of the twelve Superstore managers. The top left and right visuals are geographical map visualizations and show the countries/states that the different managers are overseeing. More specifically, because the maps' areas are colored they are choropleths. The top left visual is a world map. The countries in which the Superstore operates are colored. Each manager has a color assigned to them which can be found in the legend on the right side of the composition. The top right choropleth was created because several managers operate in the United States which is not visible in the world map. The bottom left visual is a stacked bar chart and portrays the profit/loss percentages of each manager. The colors in the bar chart show the percentage loss in red and the percentage profit in green. Each bar represents a manager, which can be found on the x-axis, and the y-axis shows the percentages. To read easily, the total percentage of profit/loss, the actual percentages were added inside the bars. Lastly, the bottom right visual shows the total 4-year profit per manager. The x-axis shows the profit is USD and the y-axis represents the individual managers.

Before the composition was made, the managers' section was split up into individual visuals, which explained different parts of the data. It included a profit point distribution chart, a tree map, and a hierarchical bar chart of discounts given per manager. The topic needed to be narrowed down to fit the profit storyline. Since the point chart was cluttered, the decision was made to make it into a stacked bar chart that showed the overall profit and loss percentage of each manager. Because the percentages can be misleading, a manager with a high profit percentage does not necessarily mean that this manager makes a lot of profit, the decision was made to include the overall profit of each manager. Important insights were found about the performance of managers, which could be addressed in a leadership meeting. The last additions to the composition were the choropleths of the world and the United States, which completed the overall story of profit per managers and their regions. It showed that several managers overlook entire continents and others only a few countries. In combination with the other visuals, conclusions could be made about the overall performance of each manager based on profitability. We can conclude that Alejandro Ballentine, who oversees several Asian countries including Indonesia, Thailand, and Vietnam, has a high loss percentage (bottom left visual) and a low total profit (bottom right visual), seven of the twelve Superstore managers have a loss percentage between 20 and 30%, and Nicole Hansen records the lowest loss percentage out of all managers. However, when we look at her total profit in the bottom right visual. She records the lowest profit over 4 years out of all the managers. Lastly, it is important to note the high profit managers for the superstore. Anna Andreadi tops the list with a profit of $311,404. Other high profit managers are Jack Lebron and Shirley Daniels.

Managers are hugely important in the success of a company. Addressing the overall performance of managers can be confronting, but a key to improving profitability, which is why this composition of visualizations fits into the Superstore profitability storyline.

Visualization-2: Comparison of Sales and Profit



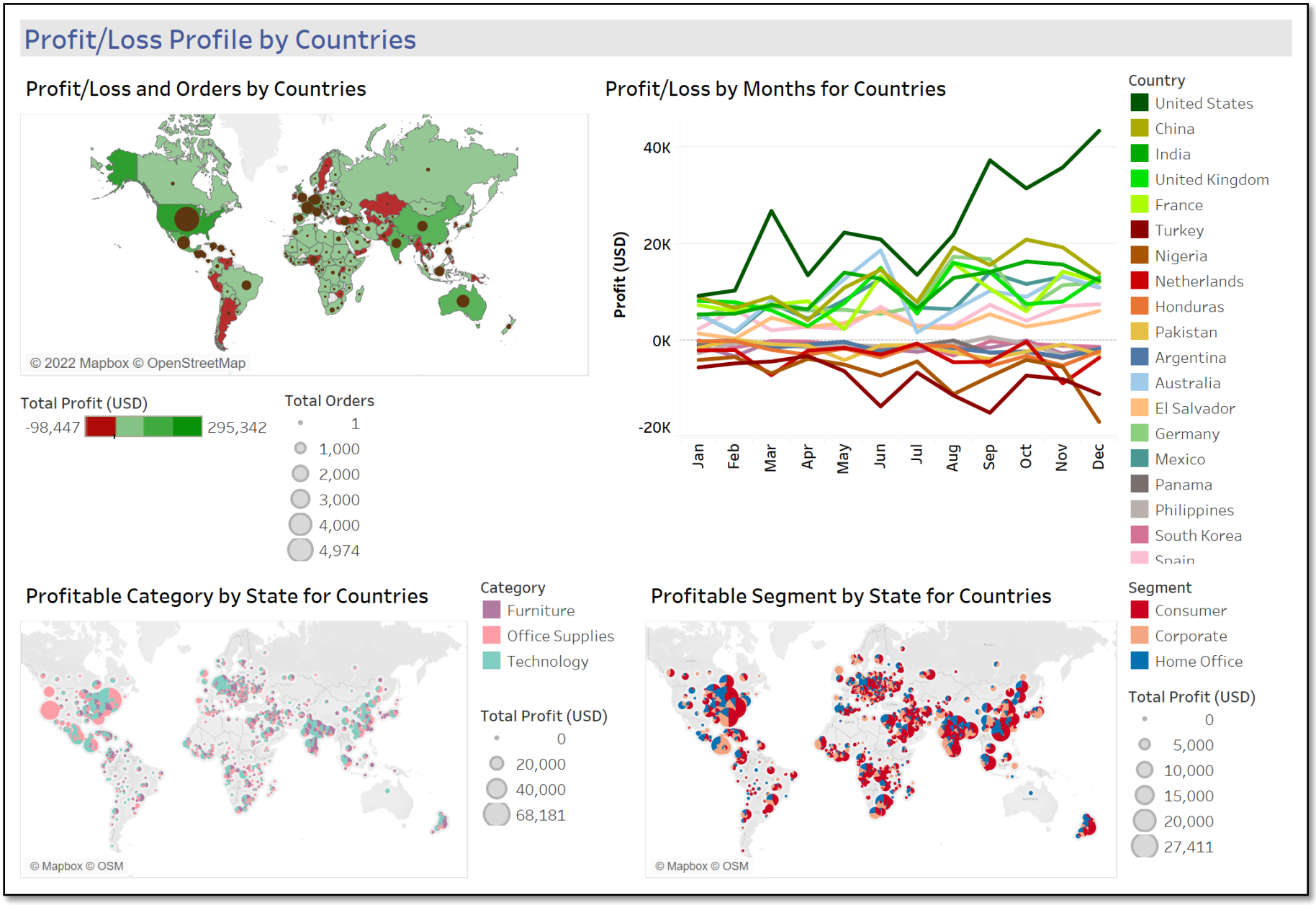
Sales and Profit are important factors for a company to grow and to increase its revenue. We know that Sales and Profits are related, higher the sales, higher the profit. Using the Superstore dataset obtained from Kaggle, my visualizations focus on the relationship between Sales and Profits by segments, categories, and countries.

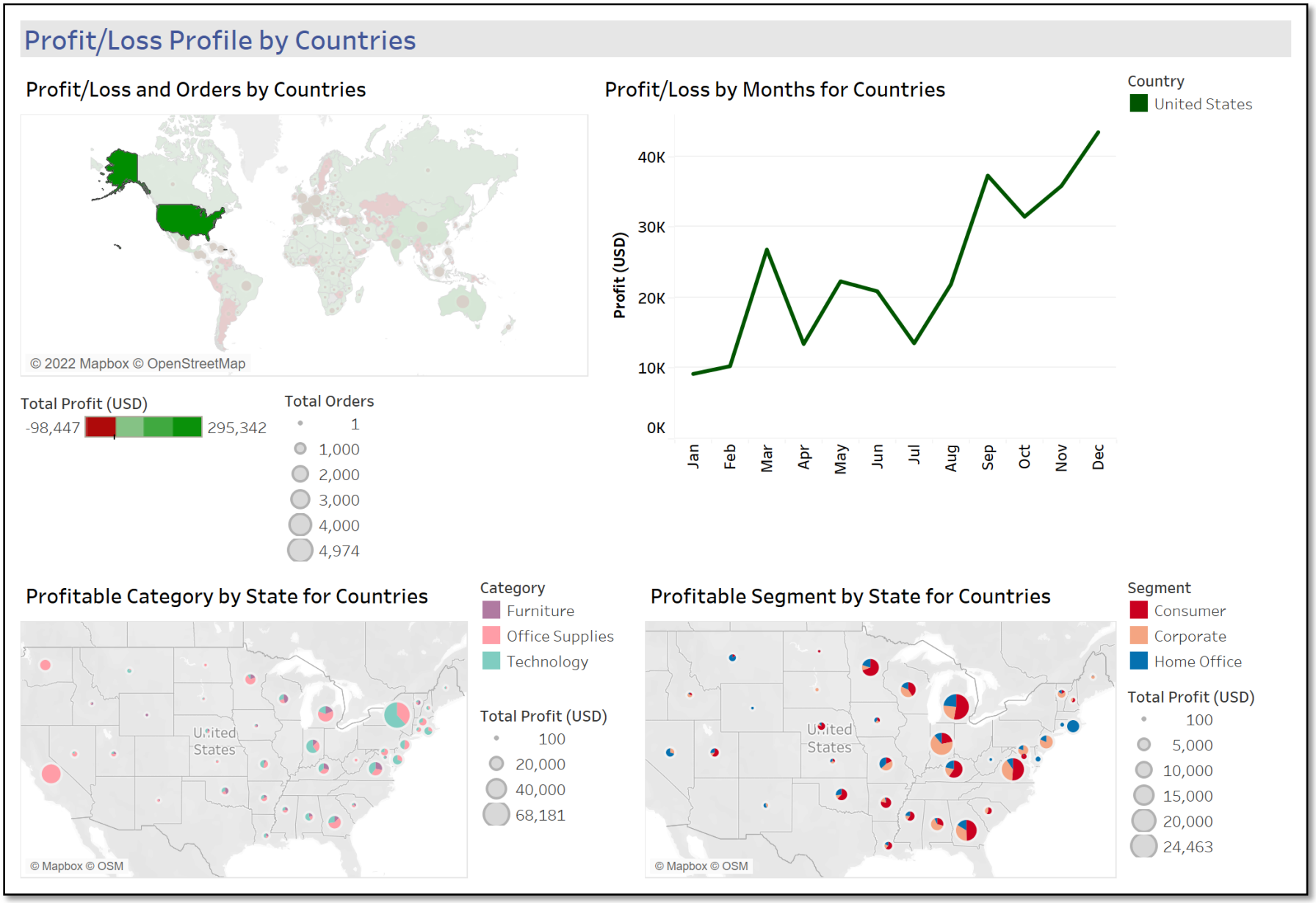
First visualization shows us the Sales and Profits by Segments (USD), in this we have three segments – Consumer, Corporate and Home Office, in this we can see which segment makes the most sales and how much profit they make. The variable segment is an independent variable, and sales and profits are dependent and continuous. In this graph, we can see that the segment- Consumer made the highest sales and so it did make the highest profit. The Superstore can focus on the other two segments - Corporate and Home Office as they have less sales as compared to Consumer, and try to give some promotions, discounts and include more items that may attract customers to buy more of their products. This visualization was made by using the sales and profit variables. The color palette was chosen from color brewer, by choosing the colors from "colorblind safe" option. Bar graphs are quite simple and easy to understand, even by non-technical people and represent data in a very clear way.

In the second visualization of Sales and Profits by Category (USD), categories are divided into three types – Furniture, Office Supplies and Technology. Order Date and category are independent variables, sales and profits are dependent variables and are continuous. In this visualization we can compare the sales and profits made by categories over the years, and how much sales and profit of each category increased over the years. With respect to the sales of furniture, office supplies and technology, we can see that the sales increased and doubled over the years; whereas, the profit of furniture was almost doubled, but for the office supplies and technology the profit made was more than double over the years. This visualization was created by using the sales, profit, and order date variables. The color palette was chosen from color brewer, by choosing the colors from "colorblind safe" option. Line graphs are easy to understand the growth of a variable (sales and profits) over the period in a very easy and effective way.

The last visualization shows us the Sales and Profits of Countries (USD), in this visualization we can see the overall sales made by the countries. Country is an independent variable, sales and profits are dependent variables and are continuous. To get a clear understanding of the sales made by the countries, we can see the color palette and see which country made the highest sales and which one made the lowest sales. The dark green color represents the area with the highest amount of sales recorded and the dark red color represents the area with the lowest amount of sales recorded. The profit made by a country is represented with a circle, the size of the circle represents how much profit that country made; bigger the size of the circle higher the profit and smaller the size of a circle lesser the profit made. The interesting things that I noticed in this visualization are that the USA making the highest amount of sales also made the highest amount of profit; but for countries like Nigeria and Turkey even after making sales more than 50K USD their profit was in negative (no profit made).

Visualization-3: Profit Profile by Countries: Orders, Categories, Segments and Profitable Months





An interactive dashboard was created to show detailed insight for the profit profile for each country using four key aspects — profits by (1) categories, (2) segments, (3) months, and (4) orders. Unlike the static visualizations where it shows only a single snapshot view, the interactivity helped visualize data about any country at a granular level for better understanding of the ground truth. Interactivity also allowed for various levels/layers of information to be

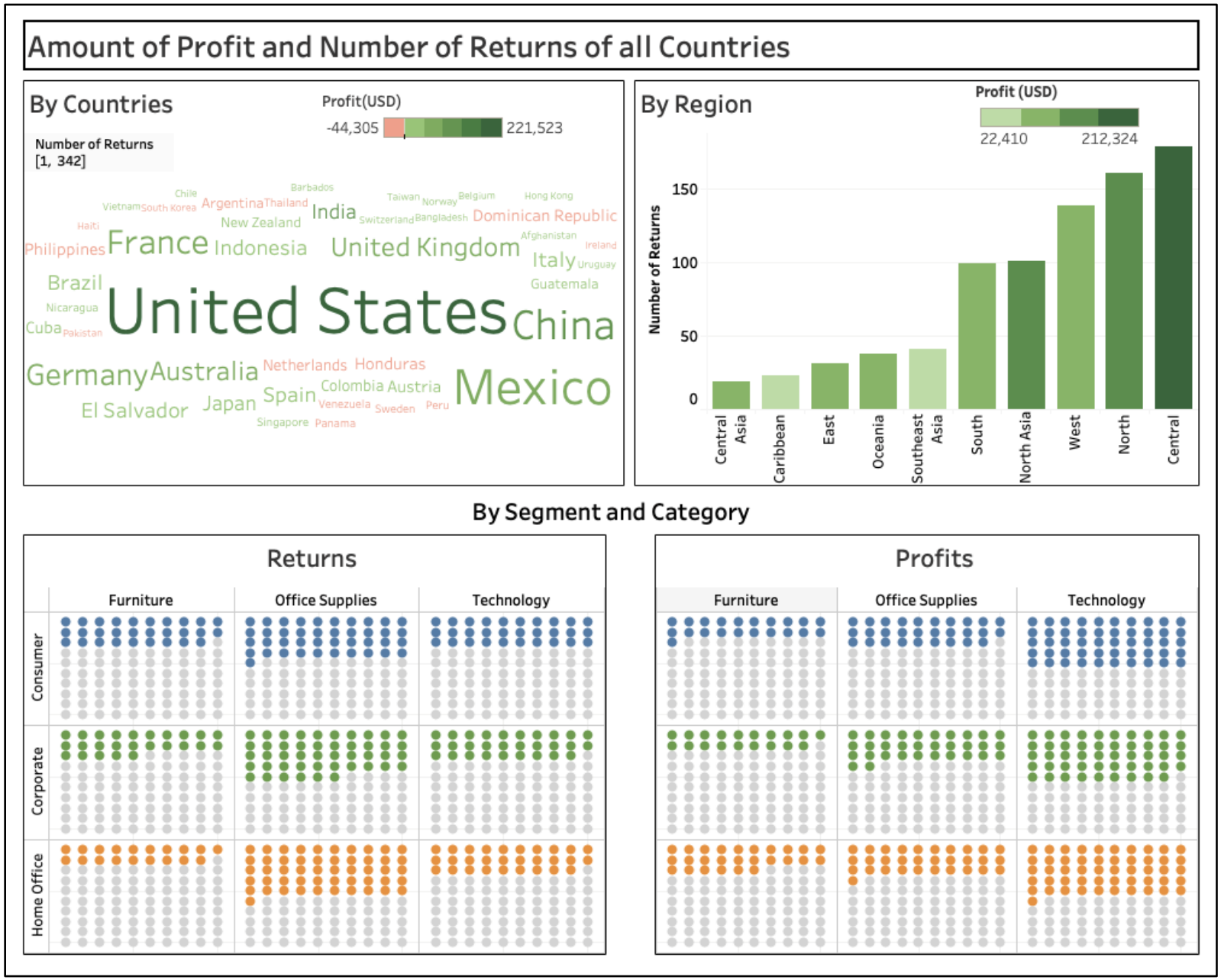
displayed. The tool tip (Appendix:2) provided additional information when the mouse is hovered over in each visualization. The first map on the top-left corner controls the remaining three visualizations. When you click on a country, the remaining visualizations zooms to the relevant granularity for that country. Various filters and actions were used both at the worksheet- and dashboard-level to create the visuals and the interactivity. Colorblind safe colors from colorbrewer was used for all the visuals. An alternate form providing information using tool tip was included for users with visual impairments.

Four visualizations were used to create the dashboard (fist image) shows the full extent of all the visualizations. The top-left visual, "**Profit/Loss and Orders by Countries**" shows the correlation between profit/loss and orders using a geospatial map. Profit/Loss ad orders are aggregated by countries. Profit/loss is shown using a graduated color map where green represents Profit and red represents Loss. The Orders are shown using graduated symbol (size). The user also has the ability to zoom into specific regions/continents or use the tool tip for further analysis. Countries such as USA, China, India and Australia had high orders (750 or more) and high profits (100k or more). Netherlands, Honduras, Pakistan, Kaseskstan and some South American countries had low orders and losses that ranged from 7K to 21K USD. Turkey and Nigeria were exceptions. They had had high orders 632 and 410 respectively, but had high losses of 98K and 8K respectively. The Superstore need to look into why the high number of orders still resulted in losses? Was it due to returns or low priced items?

If USA is selected (clicked) on the top-left map, the dashboard changes to show information about USA (see second image). The top-left map highlights USA indicating the selection. The top-right graph shows "**Profit/Loss by Months for Countries**" for the selected country. The profit/loss are aggregated by months to show the trend using a line graph. For USA, even though there is a slight dip in profits in Oct, overall there is a profit growth from Aug. - Dec. is 20K or higher while Jan., Feb., Apr. and Jul. seem to bring in lower profits that is less than 20K. The Global Superstore managers should focus on either, increasing profits during these months by providing additional promotions, discounts, free shipping etc., or focus on increasing profits even more during the high months to compensate for low months.

When USA is selected (clicked) on the top-left map, you will see both the bottom maps zoom into the country that was selected, and it shows the information at a more granular level – here it shows by state. The bottom-left map shows "**Profitable Categories by State for Countries**". This map displays the 3 categories using a pie chart and the size of the pie chart shows the amount of profits each state brings in. The profits are aggregated by state for each category. The tooltip provides more information of each slice of the pie chart when you have the mouse over it. Based on this map, we can see that most of the profits come from the eastern (northeast and southeast) part of the country with the exception of California and Washington states. Office Supplies is the most dominant profitable category in almost all states followed by Technology. States such as Michigan, Virginia, and Tennessee have almost equal amounts in profits for all three categories. New York (75K) had the highest profits for all 3 categories followed by Michigan (23K), Indiana (19K), Virginia (17K) and Georgia (16K). The only profitable category in California (37K) and Washington (11K) was Office Supplies. The bottom-right map shows "**Profitable Segments by State for Countries**". This map displays the 3 segments using a pie chart, the size of the pie chart shows the amount of profits each state brings in. The profits are aggregated by state for each segment. This map also provides more information of each slice of the pie chart using tool tip when you have the mouse over it. Based on this map, we can see that most of the profits come from eastern part of the Midwest and southeast. Unlike categories, almost all states show profits in all segments, however, Consumer segment dominates profits in most states. The most profitable states for all 3 segments (ranked) include Michigan (24K), Virginia (18.5K), Indiana (18.2K), Georgia (16K), Kentucky (11.2K), Minnesota (10K) and Wisconsin (8.3K). Rohde Island's profit (5.5K) come only from Home Office segment, while New York (6.2K) and Maryland (3.7K) come from both Home Office and Corporate. Global superstore managers should focus on maintaining or increasing their market share in the existing states to increase profiles while devising a strategy to advertise and promote their services in unprofitable states to boost profits. They should also incentivize their managers to find high purchase-volume clients in such states to widen their reach and thereby increase profit margins. (Note: To show the interactivity, the another country – India was included in Appledix:3)

Visualization-4: Profit and Returns by Countries



This Visualization was created to capture the counties, regions, segments and categories with the highest number of returns and highest amount of Profit(USD). This discovery can lead to increase in profit margin for the global superstore by enforcing measures to decrease the number of returns.

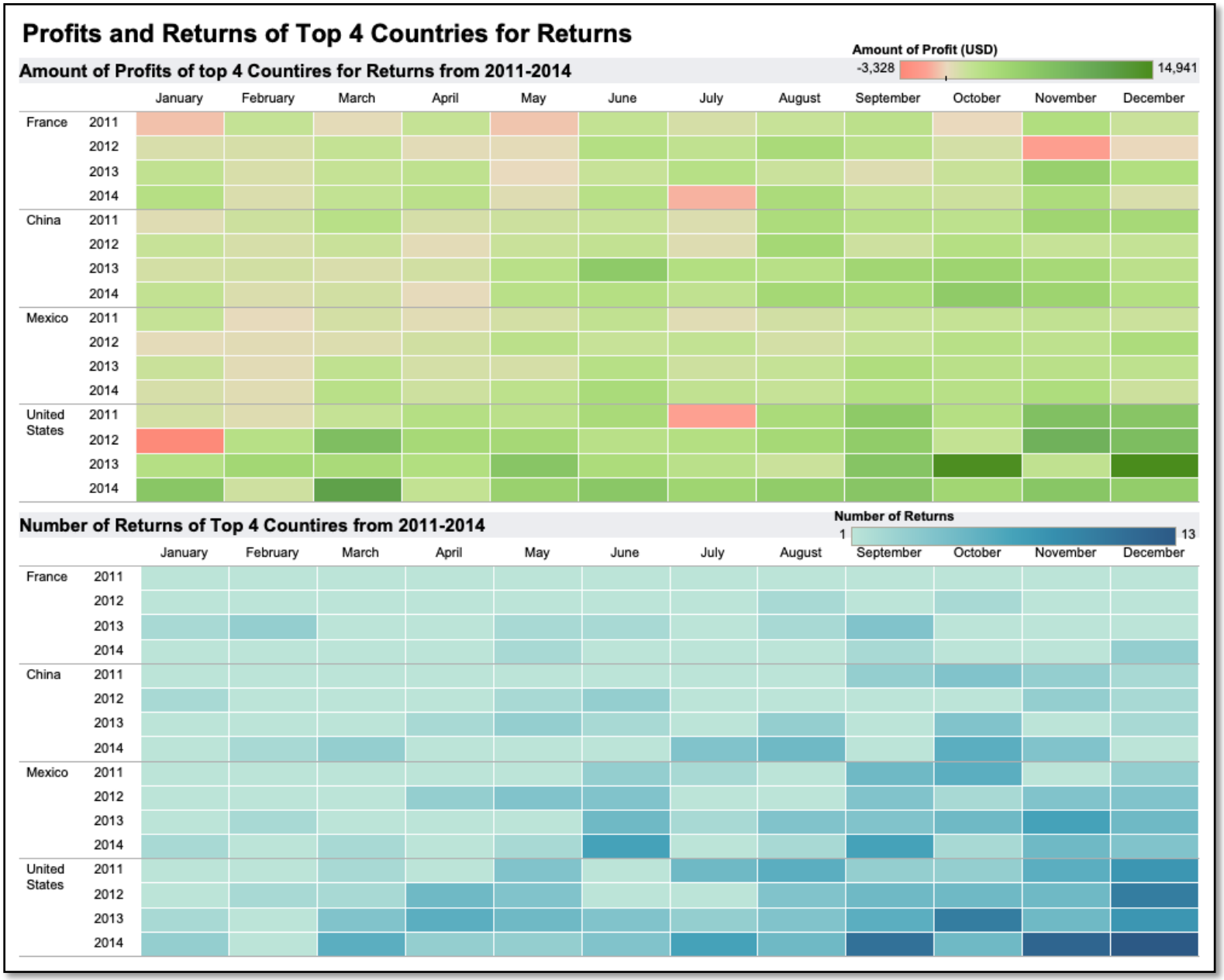
The Orders and the Returns were in two separate tables with a common field of 'OrderID'. The tables were merged with the 'OrderID 'in common. For this visualization, we created a dashboard with a word cloud, bar chart and two waffle charts to show the Profits and Returns of each country, segment, category and region.

In the **Word Cloud graph**, the size of each Word (Country) encodes the sum of the number of returns for each country and the color encodes the sum of profit from 2011 to 2014. The red color represents if that particular nation incurred losses, the green represents the profit. It is clear from the word cloud that the United States, Mexico, and China have the most number of returns. Countries like Pakistan, Ireland , Peru, and Haiti had the least number of returns and incurred losses.

The **Bar Chart** shows the number of Returns and Profit is encoded in color**.** Central, North Asia and the North regions had the most amount of profit. West, North and Central had the most number of returns.

A separate table was created with variables Segment, Category, Number of Returns and Profit. Here, the Profit and returns were in percentages of each Category. In the **Waffle Chart**, each dot represents one percent. Both Waffle charts were created with Category as the columns and Segments as rows. The first Waffle chart shows the number of returns and the second shows the amount of profit. I had different colors for each of the segments so that it is easy to understand. Each color represents a Category (100%). Office Supplies have a high number of returns and lower profit whereas Technology had the lower number of returns and higher profit. In the future, I would add interactivity to this dashboard and maybe instead of a waffle chart, I would try to use Isotope visualization.

Visualization-5: Profits and Returns of Top 4 Countries for Returns



Another type of visualization generated is a heat map suggesting the amount of Profit and Number of Returns of Top 4 countries with the highest number of returns. The Heat Map is generated for the years from 2011-2014 and divided monthly. For this graph, the data comes from the combination of two tables Orders and the Returns with a common field of 'OrderID'. The months and years come from the variable 'OrderDate'. The top 4 cities with the highest number were used. Focusing on preventing returns from countries with higher returns will increase the Profit margin greatly. This is why I chose the top 4 countries. To add to it the legend for amount of Profit suggests the Profit margin in USD, and the legend for the returns represents the number of returns.

For profits the red color shows that specific country had losses in that year. The green color shows the profit. Darker green represents a higher profit margin for that country in that year. The beginning of the year has losses and lower profit margins. From January to December and from 2011 to 2014, the Profit increased for all four countries with a couple of exceptions (eg. June of 2011 for the USA). For the returns, the darker the color, the more the number of returns in that country for that year. From January to December and from 2011 to 2014, the number of returns increases for all countries. Among the top 4 countries, the United States had the highest number of returns and highest profit margin in all the years. France had the lowest number of returns and lower profit margin in all the years. Year 2011 had the lowest profit and number of returns and 2014 had the highest profit and the highest number of returns.

IV. Conclusion

In order for a company to grow, it is important to increase profit and market share. The focus of this research was to support Global Superstore by providing a detailed insight into their profitability, sales, and returns based on their data. The key insights that emerged from each of our visual analysis are :

Visualization-1: Profit Profile of Managers

- Alejandro Ballentine who oversees several Asian countries including Indonesia, Thailand, and Vietnam has the worst overall profitability performance out of all the managers.
- Having a high profitability percentage can be misleading to the overall performance of a manager, as the total profit can be low, which is the case for Nicole Hansen.
- Seven of the twelve Superstore managers have a loss percentage between 20 and 30%. This means that on average every fourth sale a manager in this category does is without profit. By lowering this percentage, the total profit can be increased.

Visualization-2: Comparison of Sales and Profit

- For the categories, sales and profits had linear growth.
- Some countries (Turkey and Nigeria) even after having a decent amount of sales, had no profit. For such countries Superstore should try to increase sales by including/introducing products that are used a lot in those countries.
- The Superstore can give more promotions to Corporate and Home Office customers to increase the amount of sales and profits.
- Even with a limited number of segments and categories, Superstore has made a fair amount of sales and profit for different countries. Customers need to feel that they can get all their shopping done at one place, therefore, by adding more categories to their store like gym equipment, clothing, and things that people use in their day to day life, they can boost up their sales and profits.

Visualization-3: Profit Profile by Countries: Orders, Categories, Segments and Profitable Months

- Higher orders generally yield higher profits for some countries such as USA, China, India and Australia while lower number of orders resulted in low profits for some of the third-world countries like Honduras, Pakistan and Kaseskstan. However, Turkey and Nigeria exceptions to this rule - they both had high orders but also incurred losses. This will need more investigation as to what caused this anomaly.
- Certain months, for e.g USA had lower profits during first half of the year compared to the second half which need to be analyzed as to why the second half yields more profits and if the pattern is similar for all countries or is based on the financial year/calendar for each country.
- Certain states for e.g. in USA showed almost equal amount of profits for all categories (New York, Michigan, Indiana, Virginia and Georgia) while other states had only one of the categories such as Office Supplies (California and Washington). Similar to categories, some U.S. states were profitable in all three segments (Michigan, Virginia, Indiana, Georgia, Kentucky, Minnesota, Wisconsin, etc.), while some states like Rohde Island's profit come solely from Home Office segment, and states like New York and Maryland's profits were from both Home Office and Corporate. There were several states in USA that did not show any profits. It would be prudent to look into the states that do not have any profits to see how to recruit customers and increase market share and there by increase profit in the underserved states. At the same time, try to increase market share for existing states to bring in more profits.

Visualization-4: Profit and Returns by Countries

- The United States, Mexico, China and France had the highest number of returns and the most amount of Profit. These countries need more attention in terms of reducing the number of returns. Global Superstore should find ways to keep customers satisfied with the products to keep the profit margin intact. Countries like Pakistan, Ireland, Peru, and Haiti had the least number of returns and incurred losses. There should be efforts to increase profits in these countries.
- Central region had the highest number of returns and profit. The West region did not have as much profit as the North region, but it had more returns. What can be done to reduce the number of returns from the West region?
- Office Supplies category had the highest number of returns and the lowest profit in all the segments. Global Superstore definitely needs to take measures to keep customers in all three segments satisfied with Office Supplies.

Visualization-5: Profits and Returns of Top 4 Countries for Returns

- The number of returns for all top four countries (USA, Mexico, China, and France) increased from 2011 to 2014. United States being at the top. Profits have also increased from these countries over the years. Global Superstore should emphasize on reducing the number of returns and keeping up the profit margins for those four countries.
- From January to June, for all four years, the number of returns and amount of profit decreased. From July to December, these numbers increased for all four years for all countries. What can be done to increase sales and hence profit in the first half of the year? Also, what can Global Superstore do to decrease the number of returns in the second half of the year?

V. Future Work

Here are various ways we could improve the visualizations in the future:

Visualization-1: Profit Profile of Managers

- As the report mostly focuses on profit from different perspectives, it is hard to go into extreme detail about the manager section. If I would have to do further research, I would want several questions to be answered: Why is Alejandro Ballentine's loss percentage higher than his profit percentage? Why is Giulietta Dortch, Nicole Hansen, and Alejandro Ballentine's total profit so low? How can we make sure that the loss percentage drops under 20% for every manager?

Visualization-2: Comparison of Sales and Profit

- In the future I can improve my visualizations by trying to include more data in these visualizations so that more information and insight is achieved by looking at less visualizations, and add interactivity.

Visualization-3: Profit Profile by Countries: Orders, Categories, Segments and Profitable Months

- The interactive dashboard could be expanded to show more dashboards using tabs, with each showing additional interactive visualizations.
- More detailed analysis need to be performed to identify certain trends and anomalies exist.

Visualization-4: Profit and Returns by Countries

- If I had time, I would add interactivity in the Dashboard. Also, instead of a bar chart, I would use Isotype chart.

Visualization-5: Profits and Returns of Top 4 Countries for Returns

- If I had time, I would make a prediction of the number of returns and amount of profit and present it with interactivity.

Appendix: 1 – Individual Report

Luuk Alfons Wassenaar

In this group assignment, I was responsible for overseeing and researching data surrounding the managers of the Superstore. Before I decided to take the managers section, I did exploratory analysis that included Tree Map visuals of the managers and categories (shown in this report), a hierarchical bar chart of the discounts given (which included creating calculated fields, as the desired discount formats were not given), and a Choropleth of profit and loss per country. After the decision was made to focus on the managers, I made tens of visualizations that were important for the process but were not included into any of the project milestones. These exploratory visualizations included time series, waffle charts, scatterplots, bar charts, and many more. These visualizations led me to create several dashboards of the managers' profitability. One dashboard included the choropleths, and one dashboard included the stacked bar chart of profitability and the bar chart of total profit by manager. Eventually I merged both static dashboards to create the overall story.

Besides contributing to the final report, I was, as every other team member, involved in all the meetings we had throughout the quarter. The purpose of the meetings was to get everyone on the same page and included brainstorming sessions, feedback sessions, and planning sessions. Additionally, I edited our presentation to make it a whole.

Before I entered this class, I had some data visualization experience from an internship. The work I did for this company was purely based on my intuition, because I did not have the background information on colors, distortion, clutter, etc. After this class I am walking away with a baggage full of information that makes me a better data analyst, because I now know what works and what doesn't based on research, not just based on my intuition. Besides that, in my internship I learned how to visualize with AWS QuickSight. Because of this class, I can confidently say I am now a Tableau and R visualizer as well.

FNU Nawaaz Sharif

The variables involved in my visualizations are Sales, Profits, Order Date, Segments, Categories and Countries. I have created a bar chart showing the Sales and Profits by Segments (USD), line graph showing us Sales and Profits by Categories (USD) over the years and a choropleth showing us Sales and Profits by Countries (USD). The visualization techniques used are easy to understand and effective for the audience to provide with the message intended. For the choropleth visualization I chose red and green as it's our common understanding that green represents something good/more, and red represents something not good/less, for the bar chart I chose colors light blue and soft green as they look kind of sober, and for the line graph I chose colors red and orange, didn't use the same colors as people would confuse between them (bar chart and line graph). All graphs complement each other as all of them represent Sales and Profits, seeing where there is less amount of sales we can focus and try to increase the amount of sales in those aspects and increase our profit.

In the future I can improve my visualizations by trying to include more data in these visualizations so that more information and insight is achieved by looking at less visualizations, and also add interactivity.

While working on creating the visualization, I learned how to look at data with different angles and which technique of visualization would show us a clearer view to understand it. Conversely, I also learned that there are also plenty of ways to create misleading or objectively useless visualizations that can easily be avoided, and that there are many ways to cleverly show data and build an appropriate narrative with visualization. Additionally, I learned the importance of creating visualizations for the purpose of generating actionable insights.

Nandhini Gulasingam

During weeks 1-2, I (and Anisha) actively looked and found team members for our team. I setup all the zoom meetings. Before all the meetings, I proactively looked at the project documents and came up with an action plan to discuss at the meeting and to keep all of us on track so that we are able to meet the deadlines on time. After the meetings, I send out detailed instructions on next steps including deadlines that we all agreed during the meetings. During the meetings, everyone provided their ideas and thoughts.

For the MS-0, all members had to bring in a dataset they thought were interesting. I had identified 3 datasets (salaries about data scientists, chocolates dataset, hotel bookings). At the meeting, we reviewed the datasets that Luuk, Anisha, Saurin and I brought, and we decided to go with the dataset Saurin identified – the Global Superstore dataset, as it was multifaceted and would provide us more ways to analyze/visualize the dataset. Luuk, Anisha, Saurin and I (Nawaaz was unable to join the meeting) came up with a group name and submitted the MS-0 (survey) vial D2L.

Once we decided on the dataset, at the second meeting, we decided what areas would be best to analyze/visualize. Luuk was unable to make this meeting. We all decided what areas each of us are going to focus on. After the meeting, I sent out an email with who was responsible for which areas.

By the next meeting, we all had to create exploratory visualizations, and upload our visualizations including a story line for our visualization on the google doc Anisha created. For the MS-2 (exploratory analysis), I had created 20 visualizations to compare USA and World profit/loss by categories, segments, profitable months which included geospatial maps, heat maps and line graphs. At the meeting, each of us went over our visuals to identify a common/overall storyline. For this MS-2 deliverable, for my part, I performed background research on how others are visualizing similar data and provided links and justifications as to why the visualizations I choose were better suited for my part of the analysis. I modified the data description Saurin wrote. Anisha and I redid the datasets/datatypes that Nawaaz provided.

For MS-3, we all decided to focus our analysis/visualizations by world. After getting feedback from professor for MS-2, (he wanted us to also use R to create some type of viz that was discussed after midterm exam), I created a dashboard with 4 visualizations and a mosaic plot (Markets and the three Segments by the three Categories of products sold by the Global superstore for the whole world). Please note that Nawaaz also did a similar mosaic plot using R on his own and professor's feedback was provided for his mosaic plot ☺.

After submitting MS-3, I decided to do an interactive dashboard to do something that was not taught in class. I modified my visuals (you see that in the main document and also in my presentation), for the presentation. We met as a team to review and critique all the visuals so that we can modify it before our presentations. Nawaaz was unable to make the meeting as this was not a meeting that was scheduled during our usual meeting time. Once we critiqued the maps, we decided on the presentation order and Anish put together a list of things that we all should focus on while doing the presentations. I recorded this meeting for Nawaaz and us to review in case we need it and sent it to everyone.

On presentation day, looked at all the visuals. We all helped Nawaaz modify his visuals before the presentation. We recorded our presentation using zoom and I sent the videos to Luuk for editing. Luuk did a great job of editing our multiple takes (especially mine ☺) and uploaded it to d2l. After we recorded our presentation, we discussed what all we needed to do for the final project and allocated tasks for each person. I also caught something in our visuals which were not consistent (adding USD for all monitory fields) and agreed that we changed it on all our visuals consistently.

During week-10, I watched all the presentations and provided detailed feedback on each visualization for 2 groups and a collective feedback for another group. I also check the voice thread for feedback on our project and let the team know about the feedback and what (in my opinion) may need to be modified.

For the final report, first, I fine-tuned my visualization after our presentation, and incorporated the suggestions provided by professor. Then, I wrote the description for my visualization (**Visualization-3: Profit Profile by Countries: Orders, Categories, Segments and Profitable Months)**, and my individual report to go as an appendix in the report. In addition to those, Anisha, Luuk and I worked on the exploratory section of the report. We all worked on the Conclusion and Future work sections that related to our section of the visualization.

I created the templates for all the deliverables, and proofed all the visuals and documents. I also helped various members of our team (except Saurin) to show how to create certain visuals (over lay map to show 2 variables, plugging in colors from colorbrewer into Tableau, some aspects of dashboards, exporting to .png, etc.)

Anisha Patel

In the group, I was responsible for looking at the Returns and discovering any insights from it. I created multiple exploratory visualizations for Returns and Profits including but not limited to Dual Axis Geographical Plots, Cartograms, and Heat Maps . I created a Dashboard of multiple small visuals to incorporate what story each exploratory visualization was trying to tell. I chose the Dashboard of the Returns and Profits for all countries as my final visualization. Along with that, I also contributed to the group by creating a list of the key points that each team member needs to cover while presenting the Visualizations.

The final Dashboard, I created a Word Cloud, Waffle Chart and Bar Chart. Word Cloud showed the countries with the highest number of Returns encoded by size and Profit encoded by the color. The Bar Chart was representing the Region with the highest number of Returns and the color encoded the amount of Profit. The Waffle Charts showed which Segment and Category had the most amount of Profit and highest number of Returns. From the Dashboard it was evident that among the countries, United States and among the Regions, the Central Region had the most Returns and Profit. From the Categories, Office Supplies had the most amount of Returns in all the Segments. Technology Category had the most amount of Profit in all the Segments.

From this project, I learned a lot about data Visualization. I learned how to use Tableau and R from Professor Eli Brown . From Tableau, I learned how to Join two data sets, perform table Calculation, Create dual axis Graphs, and Dashboards. In R, I was able to use ggplot2 extensively to create many different visuals covered in course including Violin and Mosaic Plots. I learned how to effectively tell a story through a visual by incorporating Data, Audience, and Message in my Visual.

Saurin Patel

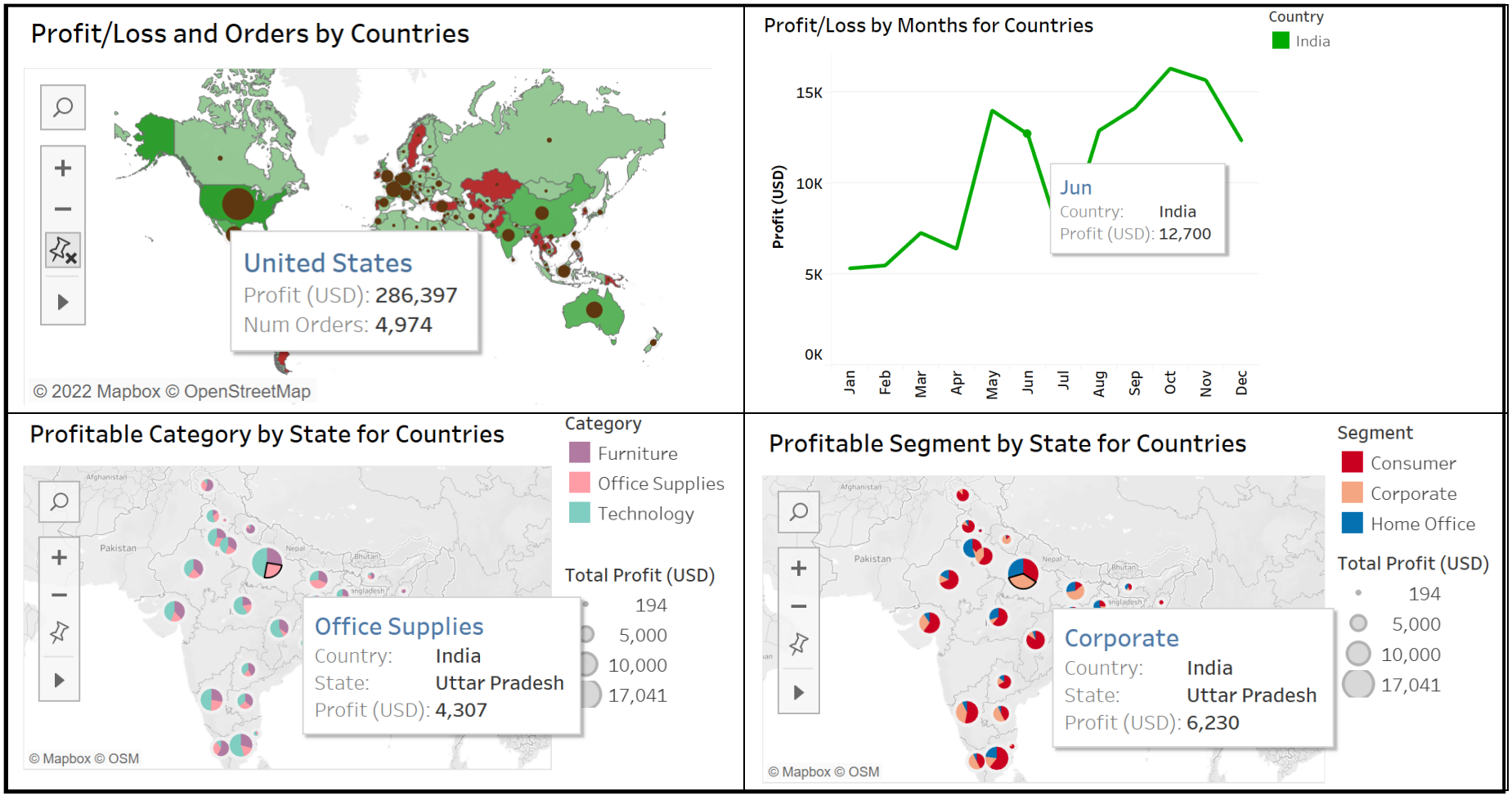
For our project, we took the dataset from Kaggle platform which consists of various tables consisting of different aspects of Global SuperStore data. Our main focus was to explore the Profit Margin and Profit margin patterns for all the countries. My role was to explore the returns and profits for the Global SuperStore. This Data set was time-series and geographical, so I explored different visualizations based on returns and profits.

Talking about my contribution in the Project. I created the Heat Map Visualization. Along with that, I chose this Dataset from Kaggle and the group approved it. I was also responsible for writing the introduction part for all the Milestones, Presentation and Final Report. I also created the PowerPoint presentation for the Presentation.

In the visualization building phase of the project I have created multiple exploratory visualizations to get some interesting patterns out of the data. I created cartograms, bubble charts, choropleth and Dashboards. After finalizing on the Heat map as my final visual, I did multiple iterations to refine the graphs in terms of telling a story. Also, I incorporated the feedback provided by my team members, professor and peers from the class to create a better version of the visual.

In terms of Learning for the course, I have learnt a lot specifically in generating powerful visualizations. The most important part in the learning process was building visualizations in R using the library ggplot2. I have learned tableau as well. Overall taking this course will help me a lot in future in my career as well.

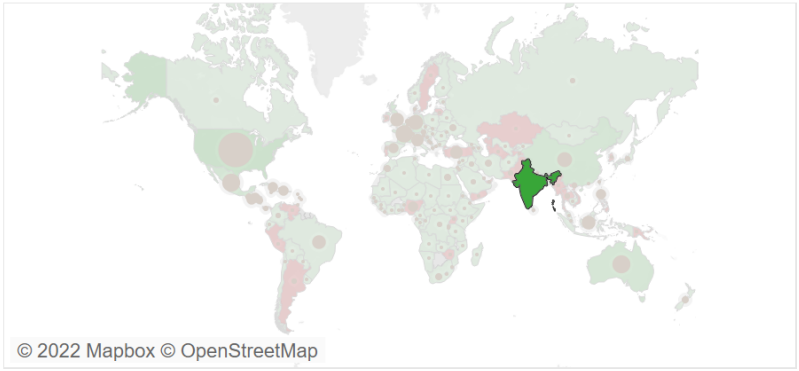
Appendix: 2 – Tooltip for Interactive Visualization



Appendix: 3 – Additional Interactive Visualization showing Different Country (India)

Profit/Loss Profile by Countries

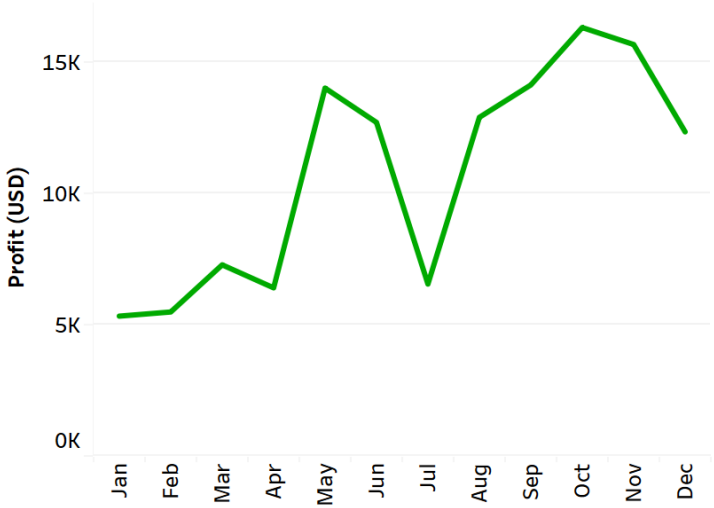
Profit/Loss and Orders by Countries



Total Profit (USD)
-98,447 295,342

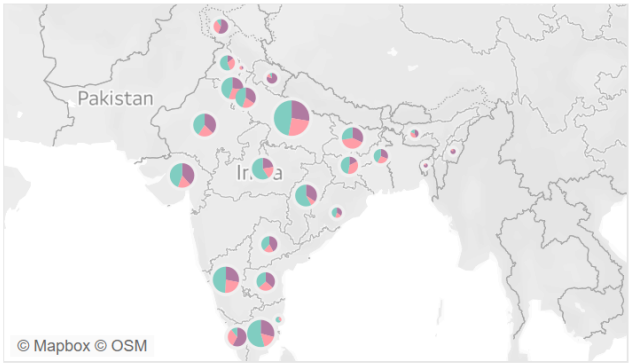
Total Orders
1
1,000
2,000
3,000
4,000
4,974

Profit/Loss by Months for Countries



Country
India

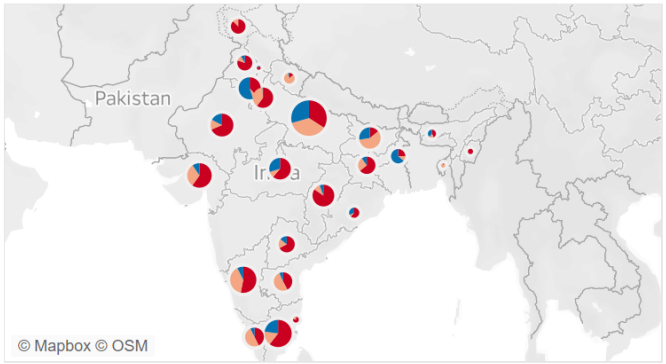
Profitable Category by State for Countries



Category
Furniture
Office Supplies
Technology

Total Profit (USD)
194
5,000
10,000
17,041

Profitable Segment by State for Countries



Segment
Consumer
Corporate
Home Office

Total Profit (USD)
194
5,000
10,000
17,041