Analysis of Global Electronic Product Sales Data

CC03-Group 5

1. Introduction

Our team is composed of learners and enthusiasts in data analysis and business intelligence, and our goal is to provide insights and decisions through data analysis. Our members have solid data analysis skills and the ability to use data visualization tools. In this task, we are working for a large retail organization selling various categories of products over grocery, home decorations, electronics books, etc. This is a large retailer with sales locations all over the world. They are forward to investigate the market demands and potential consumers. The market research will direct future retail strategies to increase the benefit of shareholders. We will present multiple key descriptive questions and one predictive question in conjunction with the dashboard to address them. During the research, we will find the country has the most sales for electronic product, the best-selling product type in the select country, the most enthusiastic age group purchasing the selected product type, the most enthusiastic region for this persona purchasing the selected product type, the dates have the best sales. The predictive question is the future trend of buying smart phones for American young aged group of consumers.

Our dashboard provides a comprehensive visual display of retail data. It includes multiple images and interactive actions, such as sales distribution maps for different countries, product category sales maps, and purchasing power comparison maps for different age groups. The design of the dashboard is very intuitive and concise, including dynamic interaction functions. Users can explore different scenarios and make predictions for the future according to their needs.

2. Scenario

To come out the report, the work breakdown will target to find the compared high amount sales of products. The analysis of consumers of each transaction will benefit further marketing strategies makings, such as promotions launch, products inventory and advertising. The bestselling product category is electronics. We are conducting important research on data of electronic products. This type of product has the highest market share looking from overview the dashboard, including smartphones, laptops, household appliances, and some commonly used electronic products.

Our dashboard's target audience is the electronic product sales department, including sales managers and sales teams, inventory managers and market analysis team. The analysis can help these audiences make routine decisions and further strategies in future. Sales management team: the sales manager needs to develop sales strategies, set sales targets for the sales team, and analyze sales performance. The dashboard can help them determine the best time to develop promotional activities and sales plans. Inventory team: ensuring reasonable inventory levels and sufficient product supply is crucial. Their decision-making needs are to optimize inventory turnover, ensure lower holding costs, and avoid stockouts. Market analysis team: the decision-making needs of market analysts are to make judgments on market trends and customer preferences. They provide insights to guide product positioning, marketing strategies, and sales forecasting. Our research and visualizations can support these roles in exploring and observing relevant data to help them make data-driven decisions that align with organizational goals.

3. Data description

3.1 Source reliability

Open and transparent: The dataset was sourced from Kaggle: this is a well-trusted community of data analysts, and data scientists. There are huge amount of datasets used for training purposes, competitions as commercial projects, scientific researches. The uploader created the dataset with the help of Python on Google using retailer information with data acquisition regarding online retail. The author described the data set in detail to ensure the reliability and transparency of data in the introduction page on Kaggle. Additional features for setting out this dataset are descriptions of the data fields and detailed descriptions to maintain the consistency and integrity of the data.
High-quality data: The dataset has sufficient data volume up to 300k transactions retail records. The data is high dimension contains 30 attributes.

3.2 Data applicability

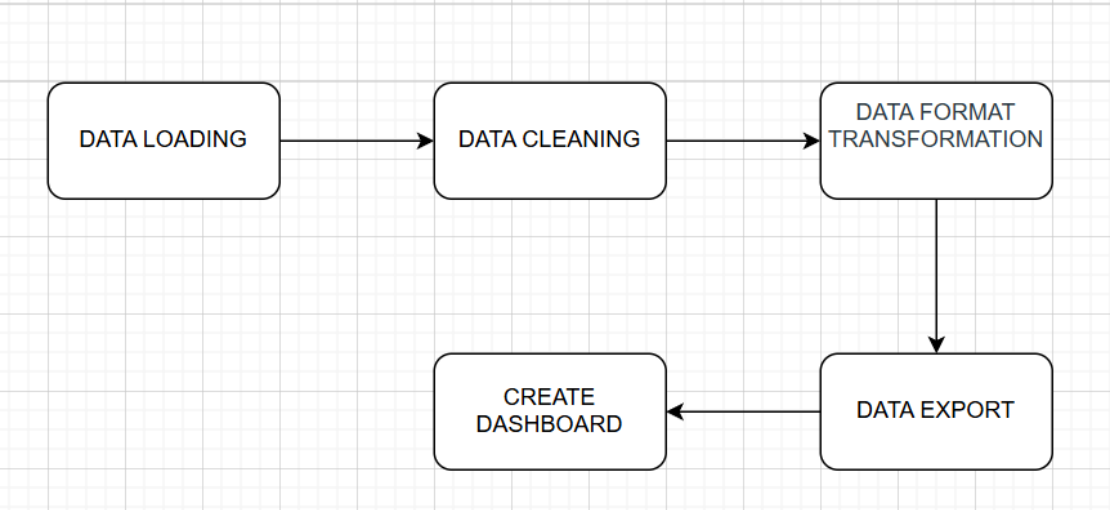
Multidimensional data: The dataset contains data in multiple dimensions, such as the time dimension, the geographical dimension (sales region, etc.), the product dimension (product name, category, etc.), and the customer dimension. Such multidimensional data help us to provide a detailed descriptive and predictive analysis. Time Series data: The dataset contains date-of-sale time series data that allow one to see a trend over time in the data set for predicting with quite a great deal of accuracy. This helps understand the pattern and make accurate predictions based on historical data. All such descriptive questions can be quickly answered by the information of the data set: what is the best-selling product, and how is the regional sales performance? In the meantime, time series data can also support predictive questions such as the impact of discounts on future profits.
The dataset chosen falls within our scenario of data analysis and decision support in the electronics sales department of a big retailer. We find the following reasons satisfactory: Comprehensive coverage of sales and inventory information: The dataset gives detailed sales information and movement, allowing us to describe sales and the features of each persona. By analyzing these data sources in terms of their reliability and applicability, we will be able to ensure that the insights and decision support from the dashboard are built on high-quality and reliable data for a better service to our target audience.

3.3 Data limitations and their potential impact on analysis:

While the dataset has high reliability and applicability to the study, a few limitations can affect the findings: this is an unchanging dataset and is not time-updated, meaning that we cannot analyze the most recent changes or market trends. The results of the conducted analysis may be lagging and, therefore, not correctly reflect the current situation. Data are regionally concentrated and not necessarily representative of the global market conditions, thereby leaving our analysis regionally bound. Generalizability to other regions in sales and marketing strategies cannot be guaranteed. Identifying and discussing data limitations helps one to understand better where the limitation can be within data and take appropriate actions during the analysis process to lower its impacts on results. On the other hand, analysis for a specific region will effectively avoid the limitations.

4. Data Transformation and Integration

As shown in the flowchart, in the process of data source to dashboard, we do the following work:



4.1 Data preprocessing

In the process of data cleaning, we first deal with the missing values and remove the duplicate Transaction ID, and for the data analysis and visualization process, customer id is useless data, so we decide to remove this attribute. Also, due to the ambiguity of total amount and total purchases, we choose to remove these columns and use amount as the amount of each transaction. The specific code implementation is shown in the figure.

```
In 7
df = df.dropna() #dropNA values
# Get unique values for each column
unique_values = {column: df[column].unique() for column in df.columns}

# Display unique values for each column
for column, values in unique_values.items():
    print(f"Unique values in '{column}':")
    print(values)
    print()
```

```
# Calculate the value counts for the 'Transaction_ID' column
transaction_id_counts = df['Customer_ID'].value_counts()

# Filter out the 'Transaction_ID' values that appear more than once
ids_to_remove = transaction_id_counts[transaction_id_counts > 1].index

# Remove rows with 'Transaction_ID' values that need to be removed
df_clean = df[~df['Customer_ID'].isin(ids_to_remove)]
```

4.2 Data Format Transformation

Since the date data in the original data is not conducive to tableau's direct use, we used python to convert the date format.

4.3 Export Data

After the data is fully processed, we use python to export the data to a csv format dataset file, which

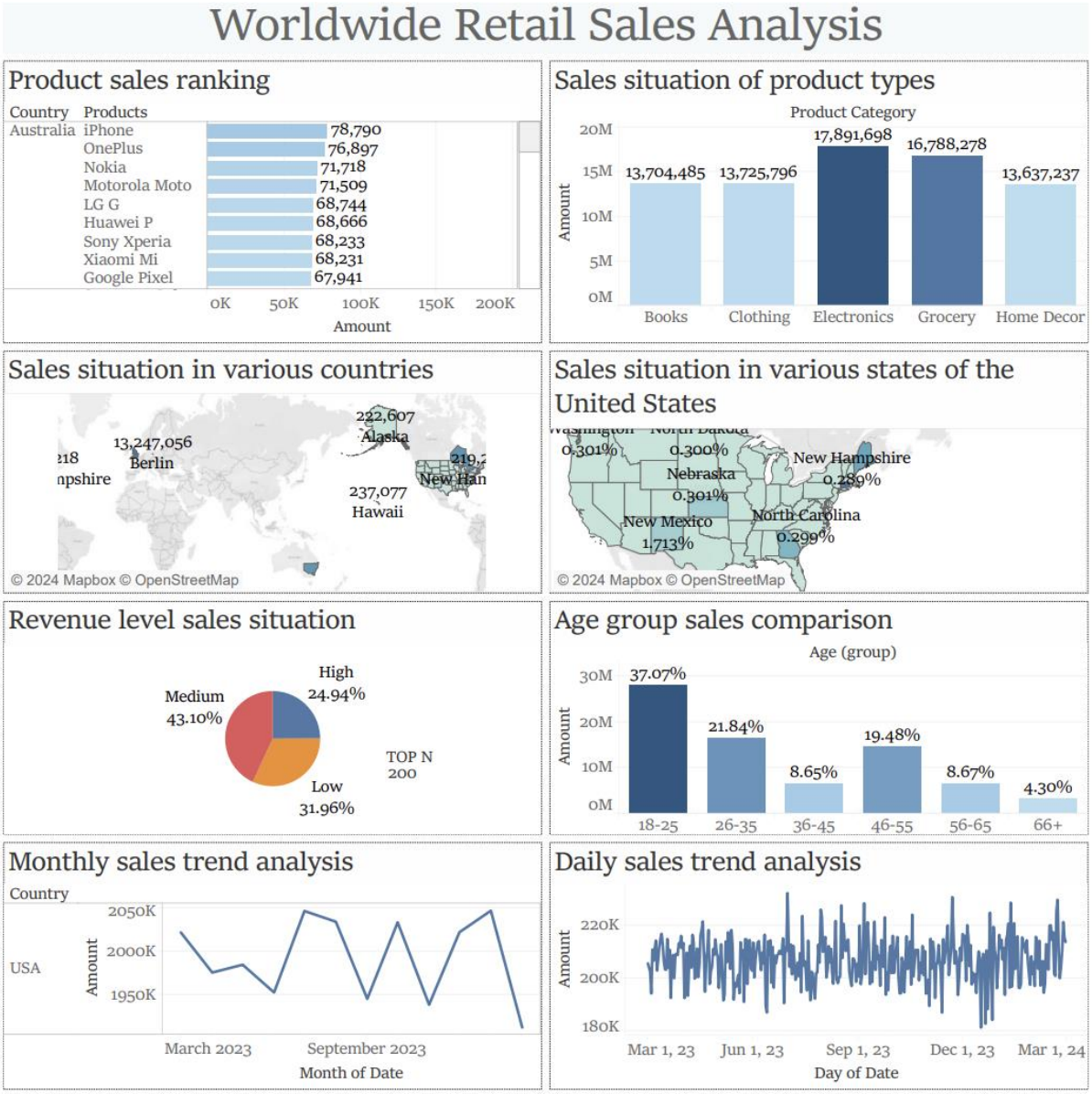
is easy for tableau to use directly. As shown in the figure.

```
df_clean.to_csv('cleaned_data.csv', index=False)
```

4.4 Creating Dashboard

In the process of creating the dashboard, we ranked the sales of each product category through a bar chart and visualized its proportion. Another bar chart is taking products as y-axis which could be modified to visualize combinations of product, product type, countries, etc. Two maps show the sale situation in various countries and in various states of the United States. Taking income, age and as visual targets, the sales percentages of different income groups are shown in pie charts, the sales percentages of different age groups are shown in Bar charts. A line chart is drawn for sales by month and day to show the change of sales over time. This dashboard contains a total of 9 worksheets, among which we add some interactions to connect them with each other. Elements of one worksheet in the dashboard is integrated with other graph's elements by setting the filter actions.

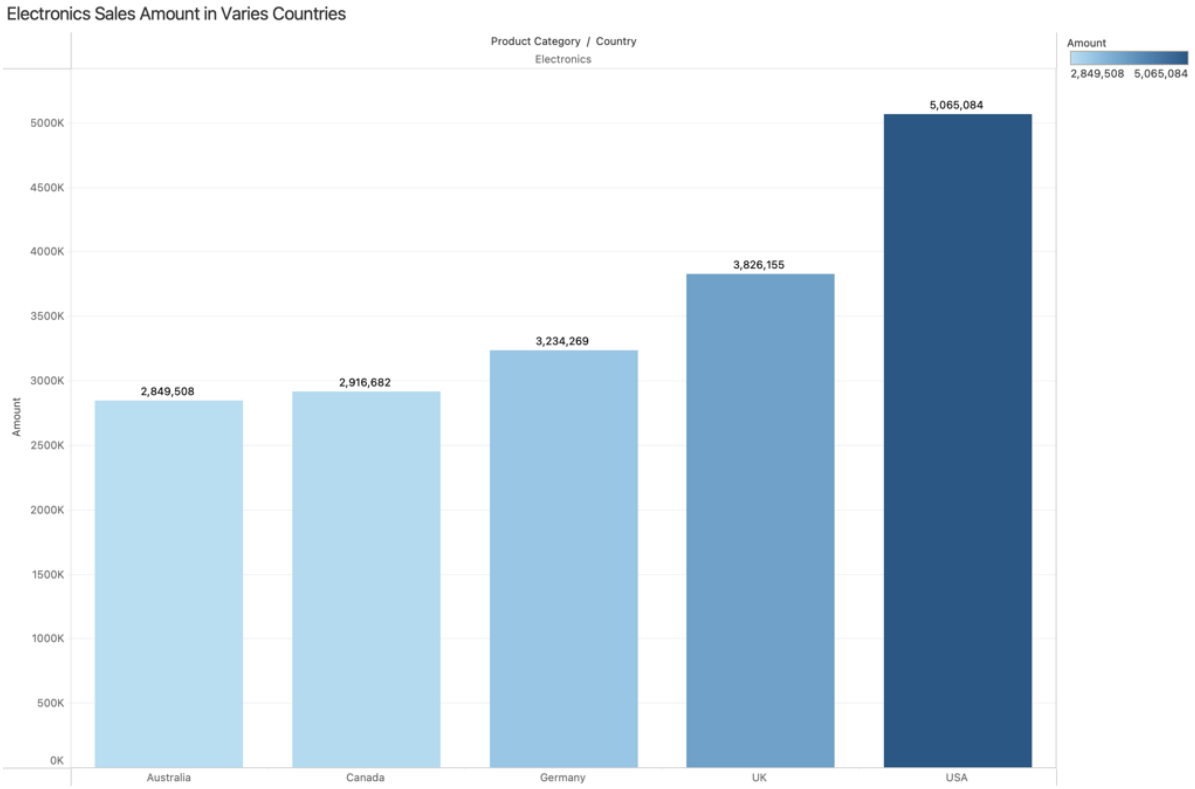
The overview of dashboard is below.



5. Descriptive Questions

5.1 Descriptive Question1

What is the most impressive electronics market in sample countries?



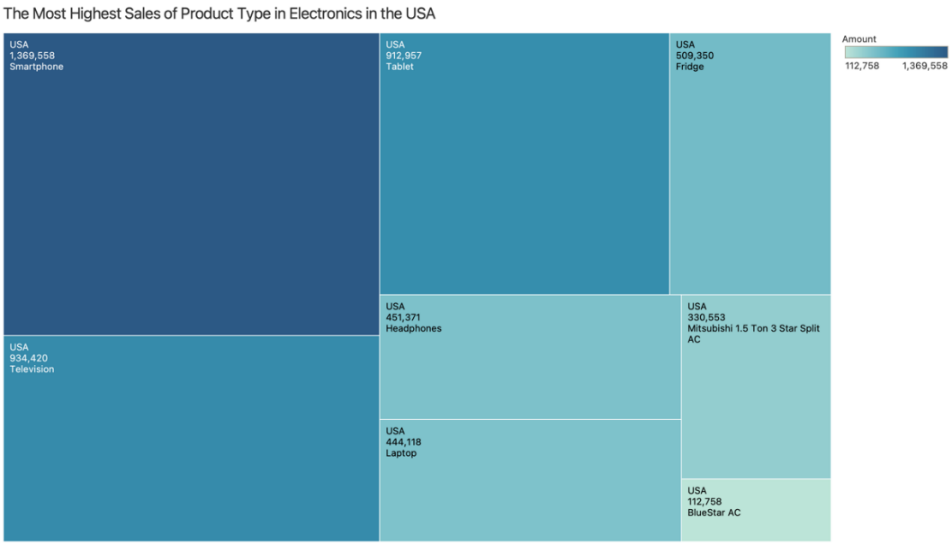
- Description:

  This question was aimed at determining which of the sample countries analyzed had the most significant market value for electronics sales. Using Dashboard's data visualization to see which countries are leading in electronics sales, The retailer's retail team can focus more on those markets, to realize potential market expansion, increase product supply and tailor-made marketing activities for leading countries. The justifications of the best-performing countries would also help to set benchmarks and reasonable performance targets for the rest of the world. This issue is critical for market strategic planning and product resource allocation at the international level. As shown in the figure, in this dataset, the United States has the highest sales of electronic products, at $5,065,084. It is followed by the UK ($3,826,155), Germany ($3,234,269), Canada ($2,916,682), and Australia ($2,849,508).

- Relevant Dataset Parts:
  - Country: Identifies the location of the market.
  - Amount: Measures the sales volume.
  - Product Category: Only sales data for the Electronics category is shown.
- Dashboard Visualization:
  - Bar chart highlighting electronics sales volume by country.

5.2 Descriptive Question 2

In the U.S., what is the highest product type sales in the electronics market?



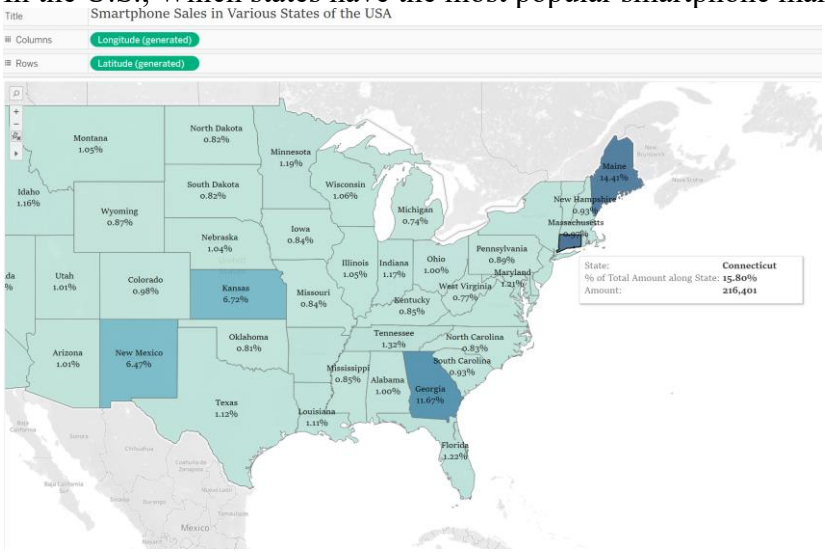
- Description:

This question seeks to determine the best-selling types of electronic products in the U.S. market. Understanding which product categories drive the most sales can help The retailer's product managers and marketing teams prioritize product inventory, design more clearly targeted promotions, and better allocate marketing budgets. By focusing resources on its best-selling product types, The retailer can ensure that the products it sells better meet customer needs, actively and proactively optimize inventory levels, and maximize profits. As can be seen from the figure, smartphones account for the largest share in the US market, with sales of $1,369,558.

- Relevant Dataset Parts:
    - Product_Type: Identifies the type of electronic products.
    - Amount: Measures the sales volume.
    - Country: Ensures the data pertains to the U.S..
    - Product Category: Only sales data for the Electronics category is shown.
- Dashboard Visualization:
    - Treemap showing sales amount by product type of Electronics in the U.S.

5.3 Descriptive Question3

In the U.S., Which states have the most popular smartphone markets?



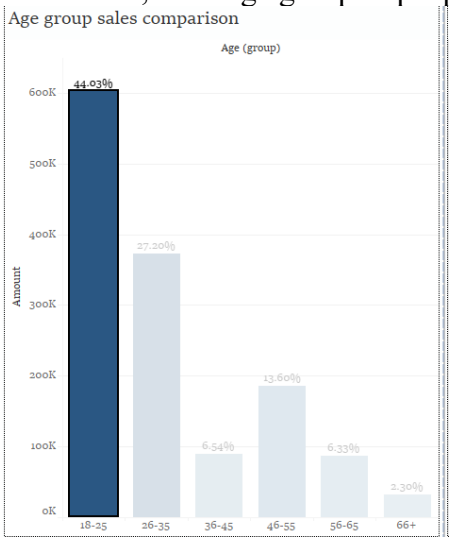
- Description:

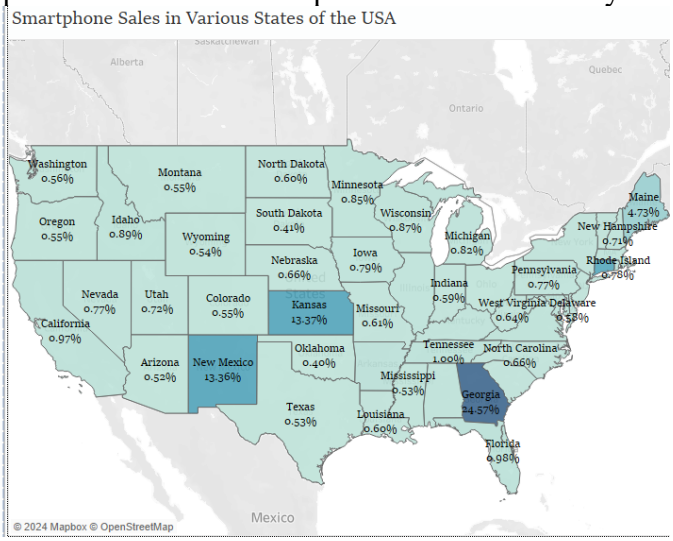
This question can determine which US states have the best-selling markets for smartphones. This information is crucial for marketing teams to use to adjust regional marketing and distribution strategies. Knowing which states have the highest smartphone sales allows the retailer to more effectively adjust its smartphone marketing strategy and optimize logistics and inventory management in the best-selling states. This information also helps to plan promotions for different regions and understand regional differences in consumer preferences. As can be seen from the figure, Connecticut's smartphone sales account for the highest proportion of overall sales in the United States. Connecticut accounts for 15.80% of total sales. The Connecticut has more enthusiastic smartphone market than other states. In this case, it will effectively affect how much inventory to Connecticut in the future coming from marketing department.

- Relevant Dataset Parts:
  - State: Identifies the location of sales within the U.S.
  - Product_Type: Filters data to smartphones.
  - Amount: Measures the sales volume.
  - Country: Ensures the data pertains to the U.S.
  - Product Category: Only sales data for the Electronics category is shown.
- Dashboard Visualization:
  - Map visualization showing smartphone sales by state.

5.4 Descriptive Question4

In the U.S., what age group of people consume most smartphones? Where are they located?





- Description:

This question identifies the age group that buys the most smartphones while determining their location in the United States. This insight is critical for marketing teams to target both consumer demographics and regional marketing. By understanding the age distribution of smartphone consumers and where they are concentrated, The retailer can design more effective marketing campaigns, tailor targeted products, and ensure The retailer has the right products in the right place to meet different needs. As can be seen from the figure, the "18-25" age group is the largest consumer group in the smartphone market, accounting for 44.03% of the total consumption. On this premise, Georgia is the region with the best smartphone sales at 24.57%. As a result, consumers in the "18-25" age group tend to live in Georgia, and marketing teams can develop more promotions and distribution strategies for smartphone sales targeting the Georgian market.

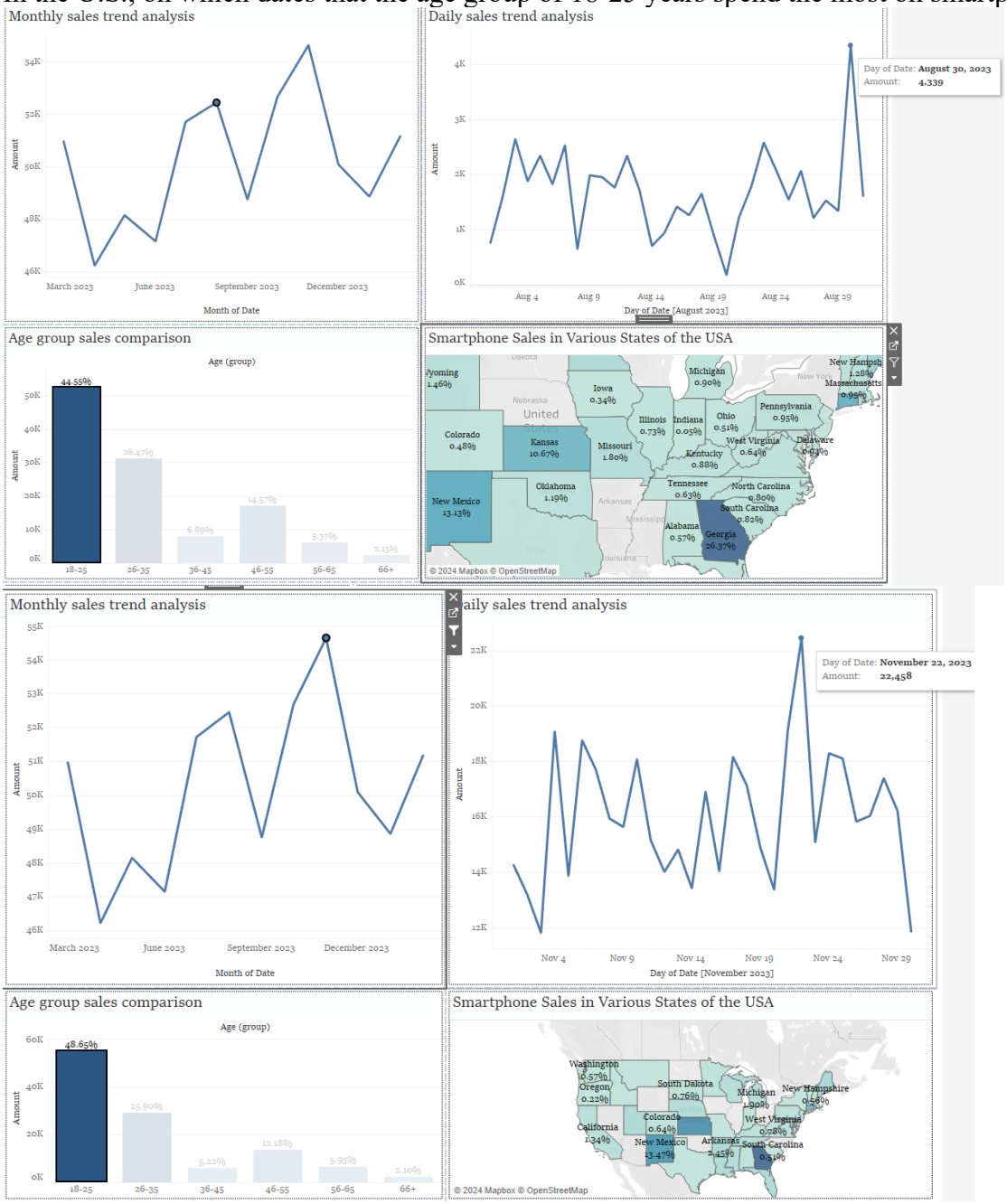
- Relevant Dataset Parts:
  - Age Group: Identifies the age demographics.
  - Product_Category: Filters data to smartphones.

- Amount: Measures the sales volume.
- State: Identifies the location within the U.S..
- Country: Ensures the data pertains to the U.S..

- Dashboard Visualization:
  - Bar chart showing sales by age group and map visualization showing their locations.

5.5 Descriptive Question5

In the U.S., on which dates that the age group of 18-25 years spend the most on smartphones?



- Description:

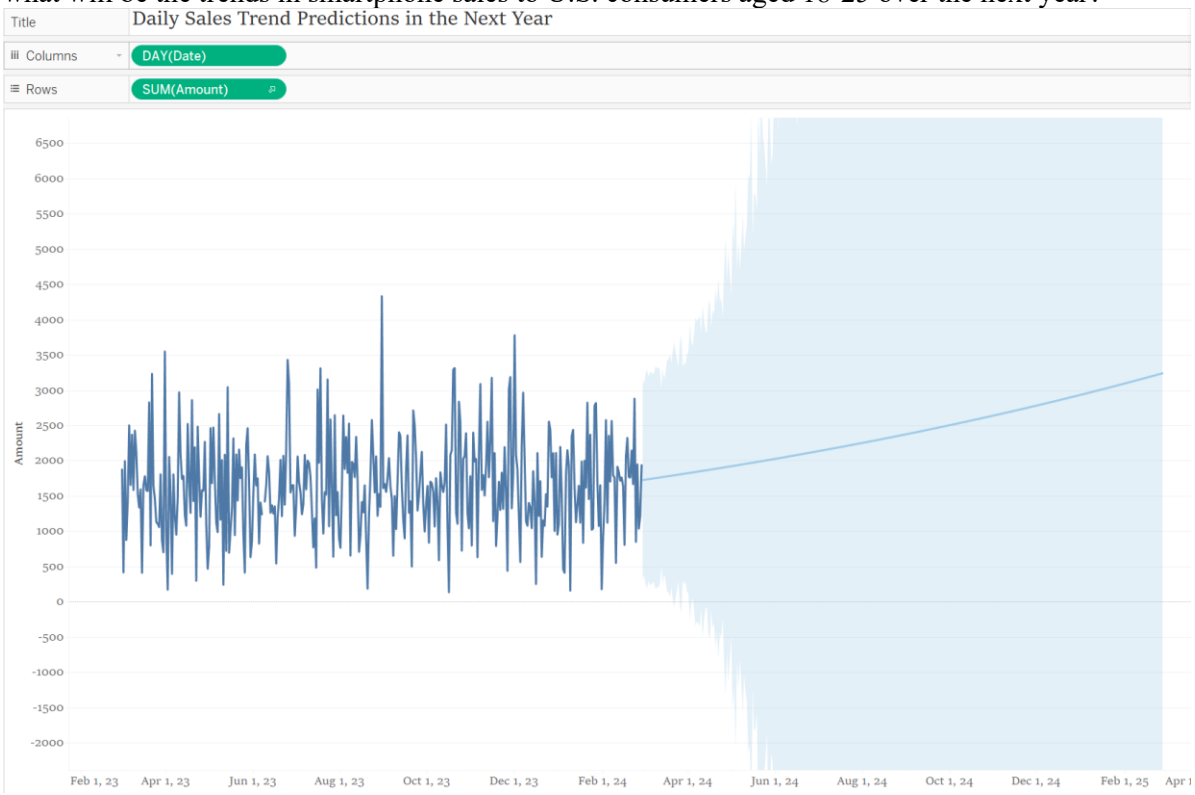
This question identifies specific days when the 18-25 age group spends the most on smartphones in the United States. Two dates reach the peaks: August 30, 2023, and November 22, 2023. The selected consumer segment between the ages of 18 and 25 might be attending school. In the U.S., most schools start in September. This customer persona may be purchasing a new smartphone before the start of the school year. The retail store is most likely to promote value

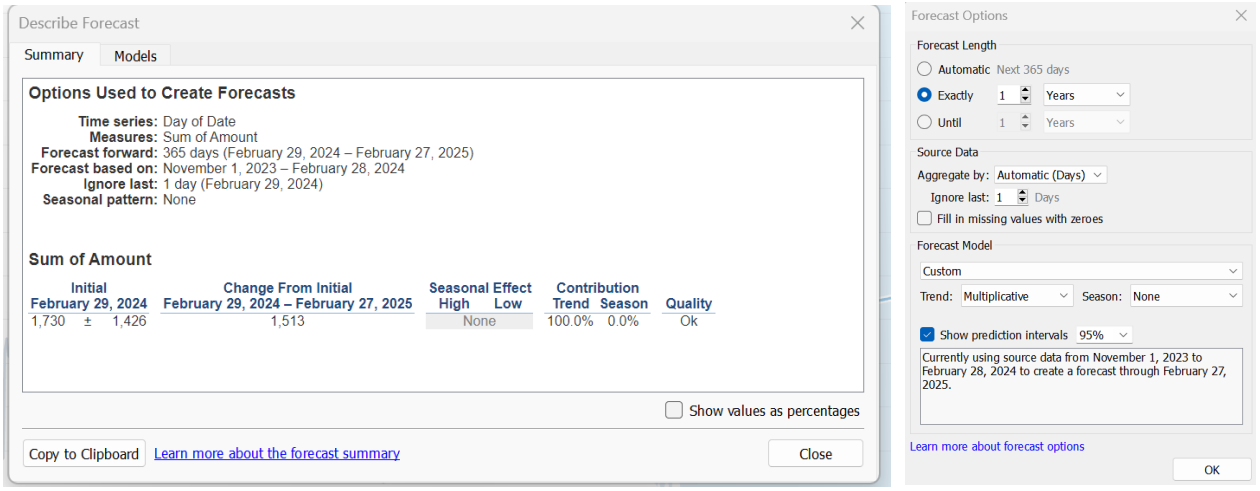
discounts to attract these students. As can be seen from the figure, in August 2023, the consumption amount reached the first peak of the year, and August 30 was the day with the highest sales amount in August. 44.55% of consumers are aged 18 to 25 and purchased smartphones in August.
Viewing the Month sales records, the sales in November is high as well. During the daily sales records in November, the highest sales represent on the November 22 which is identified as 2023's Black Friday.
November 2023 is the second peak consumption period of the year. Due to the "Black Friday" promotions, a large number of products are sold at huge discounts, including smartphones. During these times, consumers tend to pay close attention to market trends and purchase items from their wish lists, taking advantage of the discounts and deals offered. In December, 48.65% of consumers of total data are aged between 18 to 25 and the hottest sale date is November 22 which is the one day ahead the "Black Friday".
These information helps The retailer's marketing team plan promotions and sales events more effectively, as well as manage supply chains and inventory allocations. By targeting these high sales dates with special offers and marketing, The retailer can increase sales and ensure they have enough inventory to meet the increased demand for the persona as students.

- Relevant Dataset Parts:
    - Age Group: Filters data to the 18-25 age group.
    - Product Category: Filters data to smartphones.
    - Amount: Measures the sales volume.
    - Date: Identifies the specific dates of high sales.
    - Country: Ensures the data pertains to the U.S.
- Dashboard Visualization:
    - Line chart showing sales spikes on specific dates.

5.6 Predictive Question

what will be the trends in smartphone sales to U.S. consumers aged 18-25 over the next year?





- Description:

This question aims to predict trends in smartphone sales for U.S. consumers aged 18-25 in the coming year. Understanding these trends is critical for the retailer's management and marketing teams, as understanding them helps plan inventory, design targeted marketing campaigns, and prepare for seasonal sales spikes. Accurate forecasting can lead to better inventory management, reduced excess inventory costs, and optimized sales strategies to meet consumer demand. As can be seen from the figure, the smartphone sales of American consumers aged 18-25 will show an upward trend, based on the data of the previous 12 months and using the time series forecasting method to forecast. Meanwhile, "Quality" is OK, which means that the forecast quality is guaranteed. The forecast results show that sales are expected to increase significantly in the future. Specifically: In the short term (about six months), sales are expected to have a relatively modest rise. It is important to note that actual sales figures show large daily fluctuations, and this volatility is likely to continue in the future. As a result, although the overall trend is upward, there can still be significant fluctuations in daily sales.

- Relevant Dataset Parts:
    - Date: Historical sales data over time, which is essential for trend analysis.
    - Age Group: Filtering the data to focus on the 18-25 age group.
    - Product Category: Specifically looking at smartphone sales.
    - Amount: The sales amount to measure the volume of sales.
    - Country: Ensuring the data is specific to the U.S.
- Predictive Model:

The predictive model used in the Tableau dashboard employs a time series forecasting approach. This model analyzes historical sales data and identifies patterns and trends to project future sales. Specifically, the model use techniques like exponential smoothing to forecast future values. In the context of the Tableau visualization:

1. Time Series Analysis: The model considers daily sales data and projects future sales based on past trends.
2. Trend Lines: The forecast includes upper and lower confidence intervals to account for variability and uncertainty in the predictions.
3. Seasonality Adjustments: If applicable, the model adjusts for seasonal variations, such as increased sales during back-to-school periods or holidays.

- Dashboard Visualization

The dashboard includes a line chart that shows both historical sales data and the forecasted sales trends for the next year.

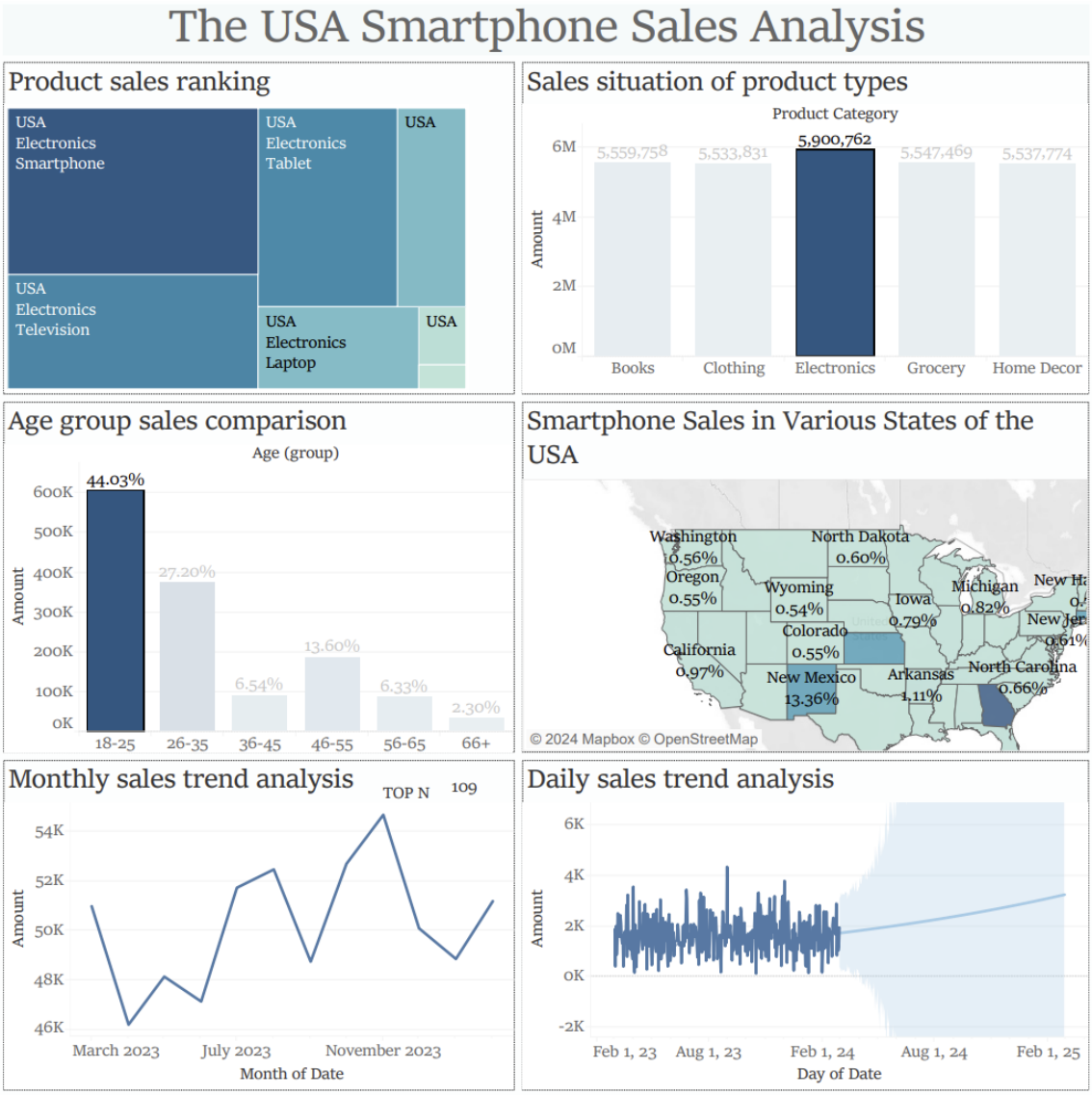
6. Conclusion and Discussion

6.1 Conclusion

The visualizations and outcomes from the dashboard or partial dashboard can clearly answer the questions we are asking. The modifying of graph is helpful to find the solution of the questions. Renaming the graphs can direct the audience to understand the goal of the modified graph. The graphs not related to the questions are removed. The workflow investigates deeper and deeper corresponding

descriptive questions from the most general question to detailed information in region, age group, and time.

The overview of the modified dashboard which can directly answer the questions is below.



The report corresponds to dashboard containing each graph guide the retailer with conclusion: the most electronics sales amount has happened in U.S. market. The active stats are In U.S. market, the most welcomed product is smartphones. The most smartphones are purchased in Connecticut. The most enthusiastic persona are young people who are from 18 to 25 years old and mostly located at in State of Georgia. They would like to purchase new smartphone before “back to school” month in August and “Black Friday. The prediction based on the built-in model in tableau will shows increase in the following one year based on current data.

6.2 Discussion

6.2.1 Predictive methodology

The predictive question needs further approach in further investigations. The current prediction is only based on the built-in methodology in Tableau. In further analysis, it is expected to deploy strategies made by the market analysis team in the future. For example, insert new attributes as discount values and calculated sales amounts to partial transactions. The predictive question could be more specific such as: Conducting 10% to iPhones, 20% off to one plus, 30% off to Google Pixels, how many sales will increase in the future.