IT ELECTIVE 2 (DATA SCIENCE)

Group 10

Group Members:

Kenzo C. Ragundiaz
Debbie Kaye Alagao
Trisha Mae Lasanas
Angeli Rivera

Acitivity 1:

Data Mining Scenario Challenge

Data Type Assigned

The data type that has been assigned to us is **“World Wide Web Data.”**

Scenario Creation

Fictitious Organization:

Move2Go Logistics Inc.

Move2Go Logistics Inc. (fictitious company as suggested by the instructions to “invent a realistic scenario”) is a blooming supply chain company that uses World Wide Web Data to further enhance its services to its customers worldwide.

Data Set

- The data set gathered for this group activity is from Move2Go Logistics Inc.'s data, which was gathered from the web.

- Move2Go Logistics' data set includes the supply chains data used by itself, showcasing the **types of payment**, **sales per customer**, **countries**, **product categories**, **products bought**, **shipping mode**, and **country origin of the product** from the 53-column-sized data set.

- The data set covers the operations within Move2Go, which is beneficial for solving its business problems and predicting future user interactions and trends.

Order Country (Where the product was ordered)

Africa	Asia	Oceania	Europe	North America	South America
Ivory Coast	Bahrain	Australia	Germany	Canada	Bolivia
Egypt	Cambodia	New Zealand	Belarus	Cuba	Brazil
Ghana	China		Bulgaria	El Salvador	Chile
Guinea	Egypt		Croatia	United States	Colombia
Kenya	India		Denmark	Guatemala	Ecuador
Lesotho	Indonesia		Finland	Honduras	Trinidad and Tobago
Liberia	Iraq		France	Jamaica	Uruguay
Madagascar	Israel		Hungary	Nicaragua	Venezuela

Order Country (Where the product was ordered)

Africa		Asia		Europe	
Morocco	Togo	Laos	Singapore	Ireland	Romania
Mozambique	Uganda	Malaysia	Thailand	Italy	Russia
Nigeria	Zambia	Mongolia	Vietnam	Lithuania	Sweden
Portugal		Myanmar (Burma)	Yemen	Norway	Ukraine
Rwanda		Nepal		Poland	
Senegal		Philippines		Portugal	
Somalia		Russia		Slovakia	
South Africa		South Korea		United Kingdom	

Customer State (Where the customer resides)

Code	Country/State	Code	Country/State	Code	Country/State
AR	Argentina	HI	Hawaii	MD	Moldova
AZ	Azerbaijan	ID	Indonesia	MI	Malawi
CA	Canada	IL	Israel	MN	Monaco
CO	Colombia	IN	India	MO	Morocco
CT	Central African Republic	KS	South Korea	MT	Malta
DE	Germany	KY	Cayman Islands	NC	New Caledonia
FL	Liechtenstein	LA	Laos	ND	North Dakota
GA	Georgia	MA	Madagascar	NJ	New Jersey

Customer State (Where the customer resides)

Code	Country/State	Code	Country/State
NM	New Mexico	SC	Seychelles
NV	Nevada	TN	Tunisia
NY	New York	TX	Texas
OH	Ohio	UT	Utah
OR	Oregon	VA	Vatican City
PA	Paraguay	WA	Washington
PR	Puerto Rico	WI	Wisconsin
RI	Rhode Island	WV	West Virginia

Problems

- Move2Go Logistics Inc. is facing uncertainty where other departments experience an abundance of sales while others are struggling to gain sales equal to the other departments.

- Move2Go Logistics Inc. is also wondering what type of products are being bought the most, what products are mostly left untouched, which countries are their most loyal customers, and which countries don’t subscribe to their shipping modes, choosing only the standard class.

- Move2Go Logistics Inc. is unsure if its customers are just consumers, belong to a corporation, or are just people working from home — unsure on how they can create innovations to reality to refurbish their products.

Goal

- With the problems stated, Move2Go Logistics Inc. wishes to solve these dilemmas to improve its services, gain trust, gain customers, and earn more profit to further assist people worldwide.

- Move2Go Logistics Inc. has an idea where it plans on using Power BI to **preprocess the data**, **visualize**, and create a **dashboard** showing the patterns and trends of the worldwide web data set gathered by creating charts and graphs for the data visualizations.

- With the assistance of Power BI, Move2Go can clean its data to check specific areas where the business is booming and to investigate the fields where it needs to improve its services. Based on the visualizations in the dashboard, Move2Go can reflect and take immediate actions on the spotted weaknesses in its business to maintain a well-balanced and successful business.

Data Description

The World Wide Web (WWW) data gathered is **structured data** that is used by supply chain companies. The data set has both **numerical data** (quantitative data) and **categorical data** (qualitative data). The quantitative data revolves around the sales, the customer ID, the days for shipping, the customer coordinates, etc. The qualitative data includes the product bought, customer category, type of shipping mode, department from which the product was bought, country, etc.

The structured data will be beneficial to the company to find its weaknesses, and the strengths to further harness.

The data set was gathered from the web as it is a World Wide Web (WWW) data. We found out that this data set has the potential to expand, and strengthen the market focusing on promoting products to the right people.

Why Was the Data Collected?

The data was collected to analyze customer behavior, product demand, shipping preferences, and overall business performance. It aims to identify trends, address disparities in sales among departments, and improve decision-making based on data-driven insights.

DATA MINING TECHNIQUES

- **Clustering (K-Means, DBSCAN)**
  - Segment departments based on sales performance (high-performing vs. low-performing).
  - Group products based on purchase frequency to identify popular and unpopular items.
  - Categorize customers into different profiles (consumers, corporations, work-from-home individuals).
  - Identify countries with high and low engagement based on purchase behavior.

DATA MINING TECHNIQUES

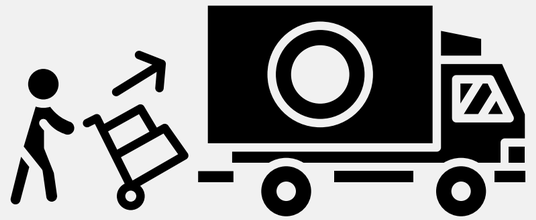
- **Regression Analysis (Multiple Linear Regression)**
  - Analyze factors influencing uneven sales across departments (e.g., product category, shipping mode, customer type).
  - Evaluate customer preference for standard vs. other shipping modes based on available attributes.

DATA MINING TECHNIQUES

- **Association Rule Mining (Apriori, FP-Growth)**
  - Identify frequent product purchase patterns to determine which products are often bought together and which are left unsold.
  - Discover customer behavior trends related to shipping mode preferences.
  - Uncover relationships between customer profiles and product categories to enhance targeted marketing strategies.

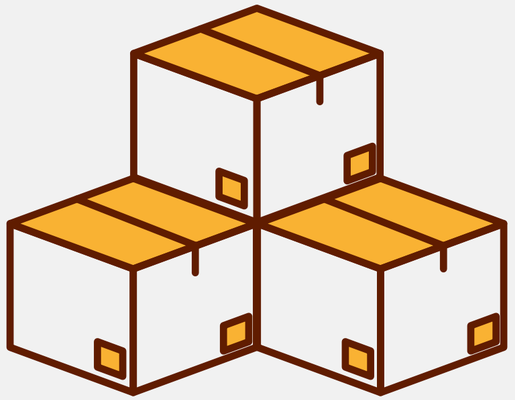
Expected Outcome

- **Identifies performance gaps among departments**, enabling strategic resource allocation and operational improvements.
- **Predicts customer demand and purchasing trends,** supporting better inventory management and business planning.
- **Reveals purchasing patterns and customer preferences**, allowing for optimized product bundling, marketing strategies, and improved customer engagement.



Benefits

- Uses data-driven insights to support better decision-making.
- Strengthens product promotion and sales strategies.
- Allocates resources efficiently to struggling departments.
- Improves inventory management by preventing shortages and overstocking.
- Boosts customer satisfaction with personalized product recommendations.



By leveraging these insights, Move2Go can refine its marketing strategies, improve operational efficiency, and increase revenue. The ability to predict trends and adjust business strategies accordingly will enhance customer retention, streamline logistics operations, and create a sustainable growth path for the company.

Potential Challenges

- **Data Quality Issues**

  Move2Go might encounter datasets that may contain missing values, inconsistent data, and errors that might affect the result of what the organization is trying to analyze, deeming it an inaccurate and unreliable insight.

- **Scalability Problem**

  Since Move2Go is available to service customers around the world, the data would likely grow rapidly as the organization expands. Millions of data transactions may cause an issue by straining computational resources used and slowing down the processing of transactions.

Potential Challenges

- **Privacy Concerns**

The organization collects sensitive personal information from the customers to analyze their preferences and what they might like within the web. However, this may raise privacy concerns with regulations like GDPR.

- **Ethical Considerations**

Personalization and market targeting using customer data might also be an issue, as it raises ethical questions about manipulation in buying more products on the platform.

Potential Challenges

- **Results Interpretability**

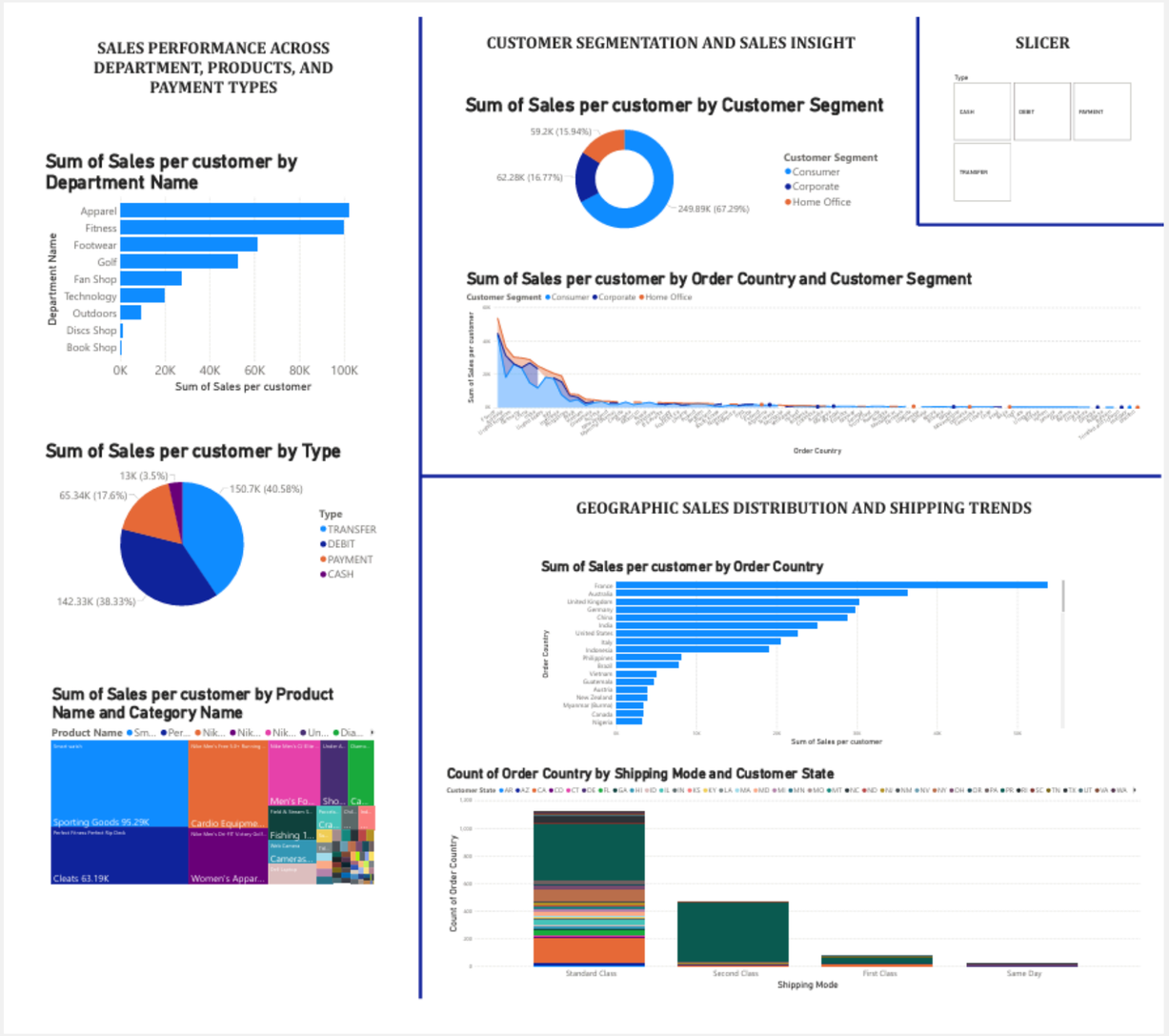
  Using certain data mining techniques like clustering and regression to make up insights may produce results that are not easy to interpret. The results should be understood by the stakeholders so the can act on the produced results without reluctance.

Is It Costly?

Yes, having to implement a data mining solution can be costly for Move2Go Logistics Inc. It would need to set up infrastructure (e.g., licenses, cloud platform), hire professionals like data scientists and analysts to produce improved insights and decision-making, and maintain the cleaning of missing values, errors, etc. All of that is mentioned would be costly, but it will for sure be for the improvement of the operations that would surely benefit the company in the long run.

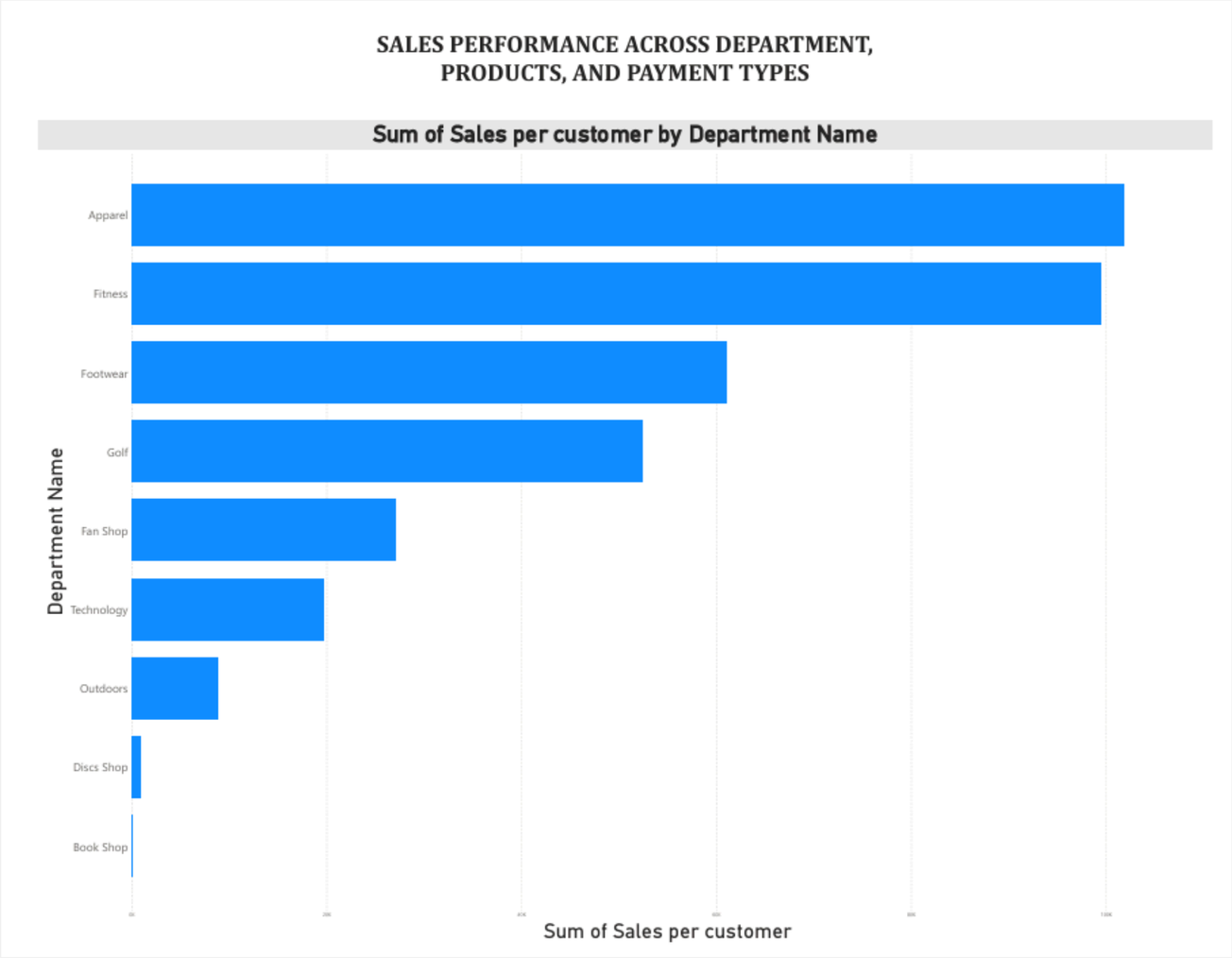
VISUALIZATION: Dashboard

- Summary of the visualizations.
- Made with Power BI.
- Cleaned data set's graphs, and charts.
- Has a slicer for easier data navigation.



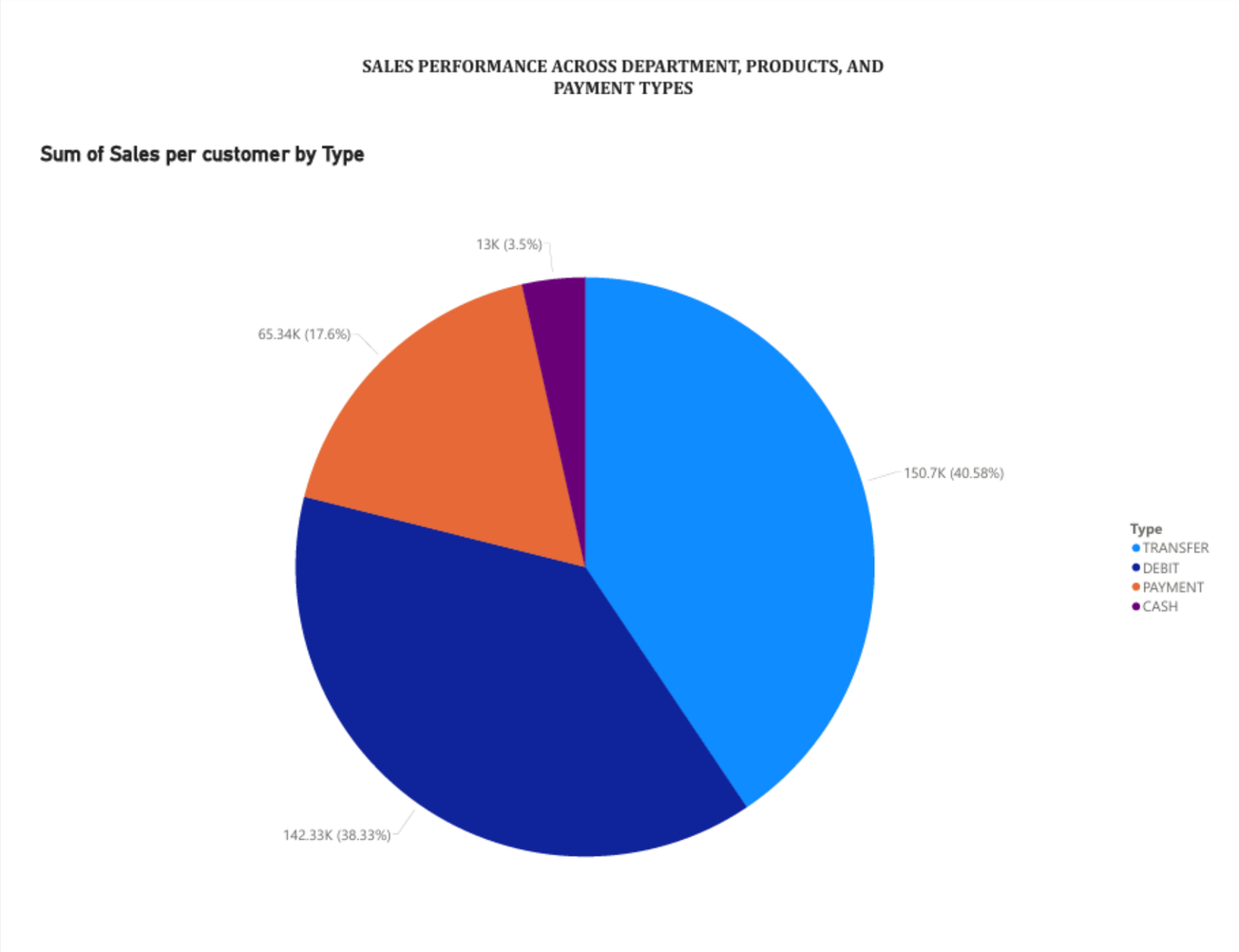
VISUALIZATION

- To determine which department has the most sales among customers worldwide.
- Department Name
- Sales per Customer



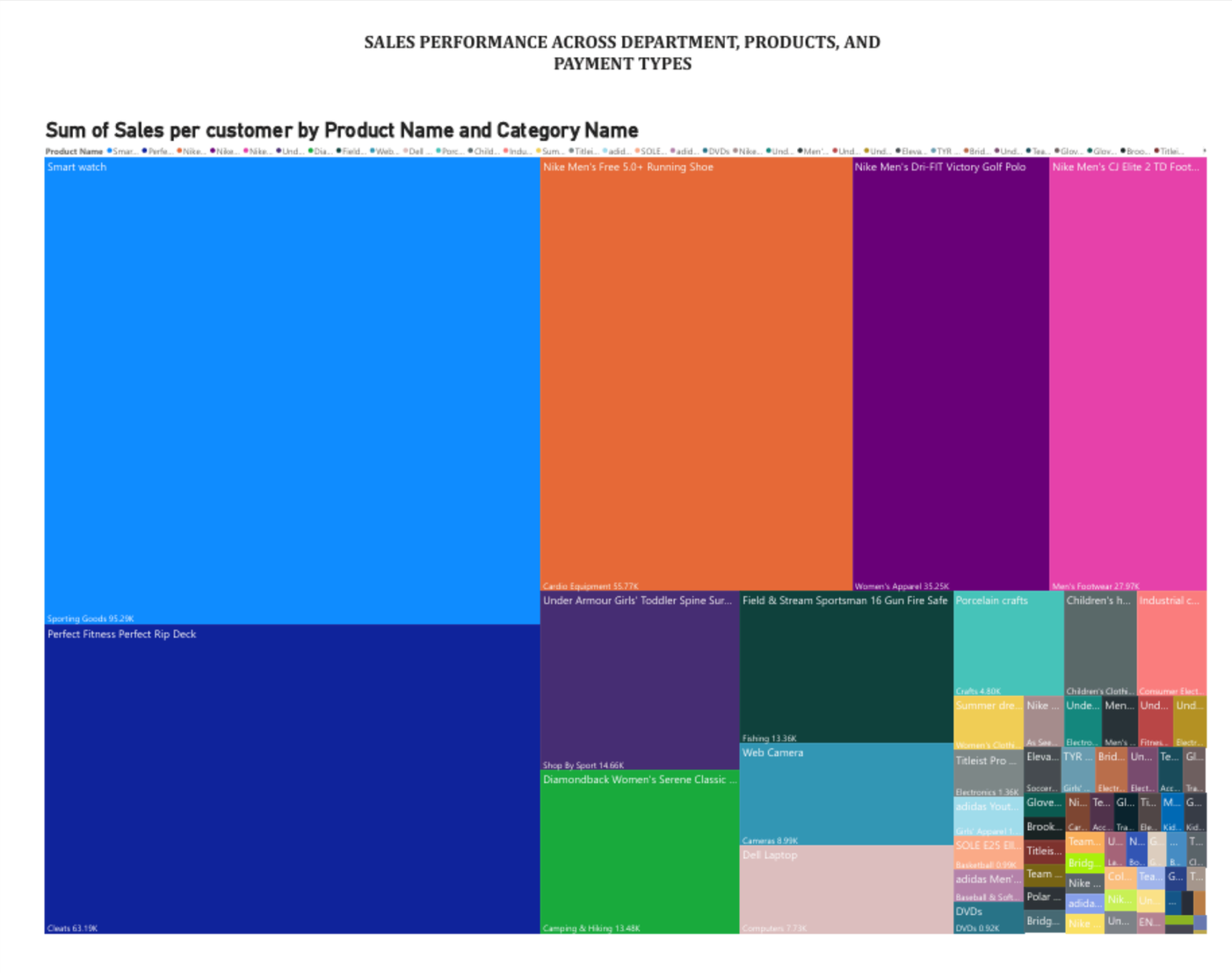
VISUALIZATION

- To determine which type of payment is used frequently
- Sales per customer
- Type/mode of payment



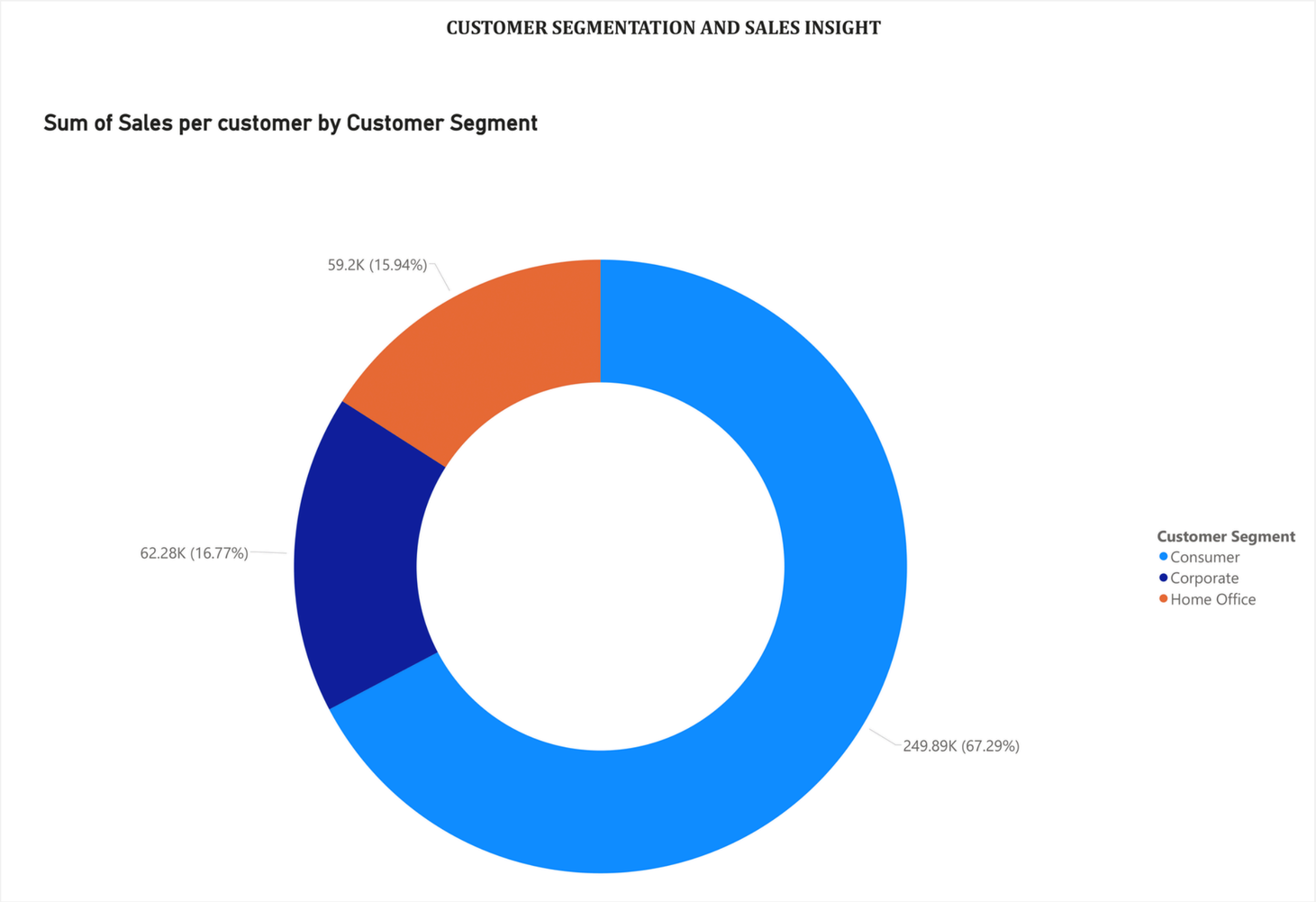
VISUALIZATION

- To determine what product has the most sales among products and Department name
- Product name
- Category name
- Sales per customer



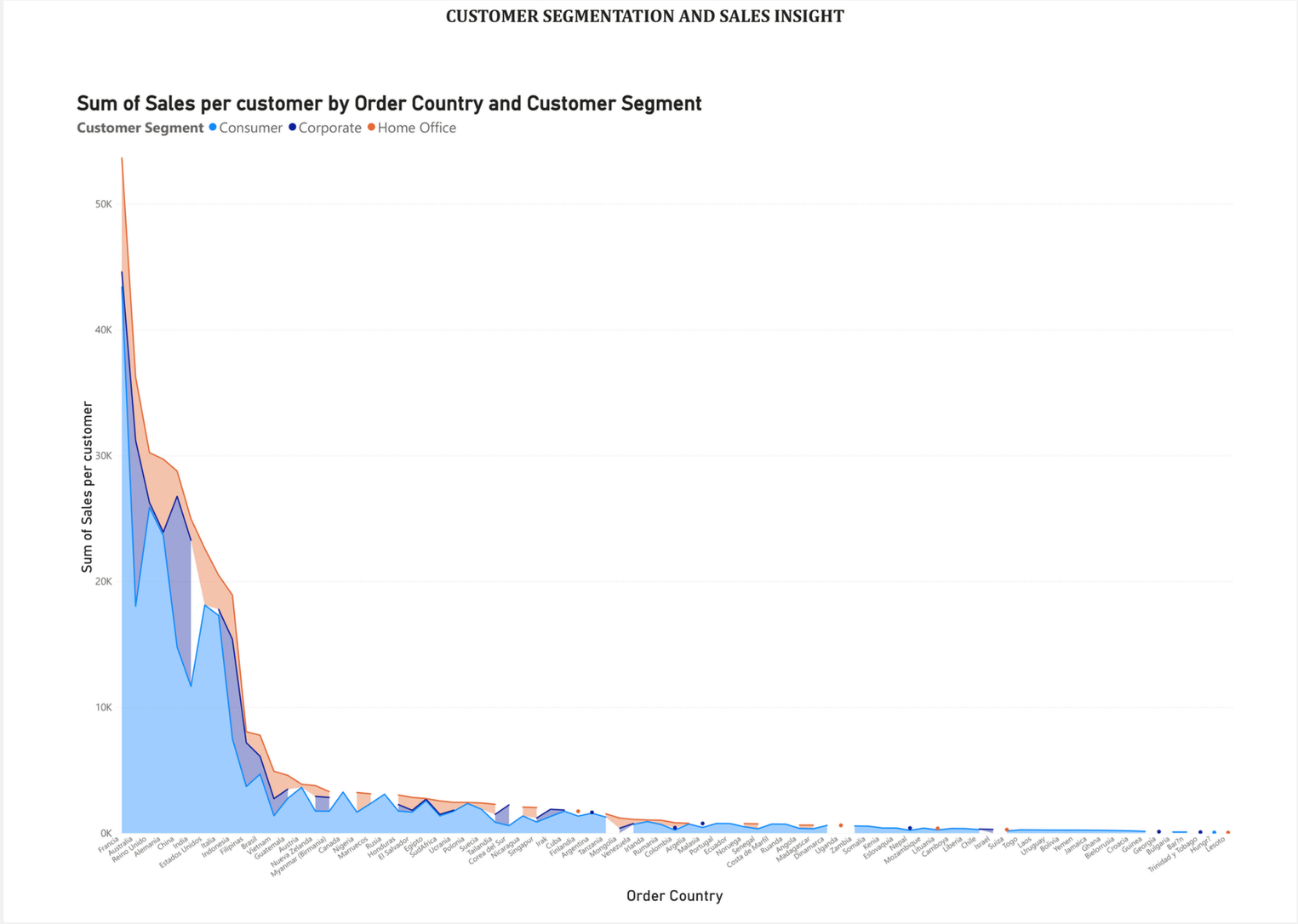
VISUALIZATION

- To determine where the consumer belongs based on customer segment
- Customer Segment
- Sales per Customer



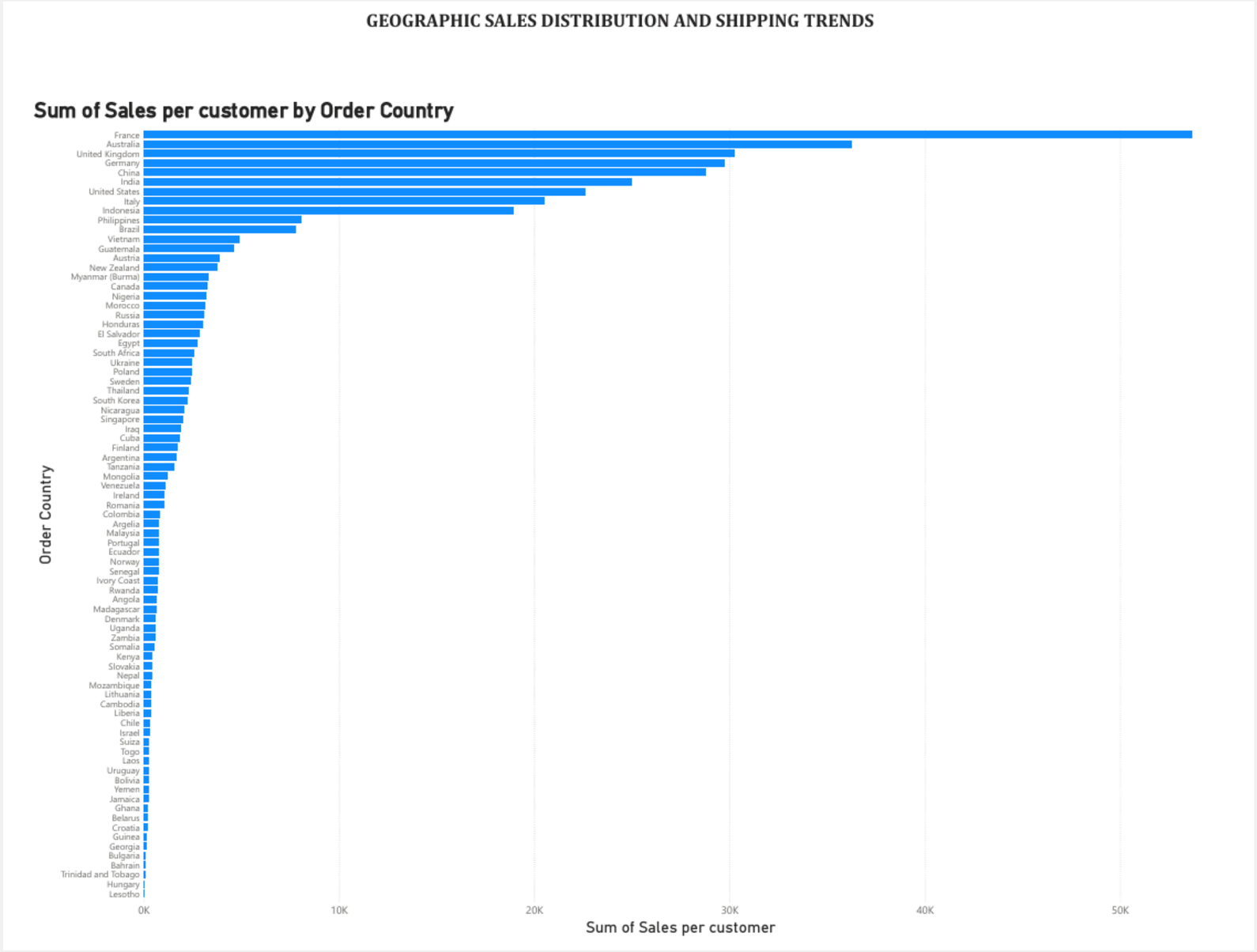
VISUALIZATION

- To determine the sum of sales by country and specified by customer segment
- Customer segment
- Order country
- Sales per customer



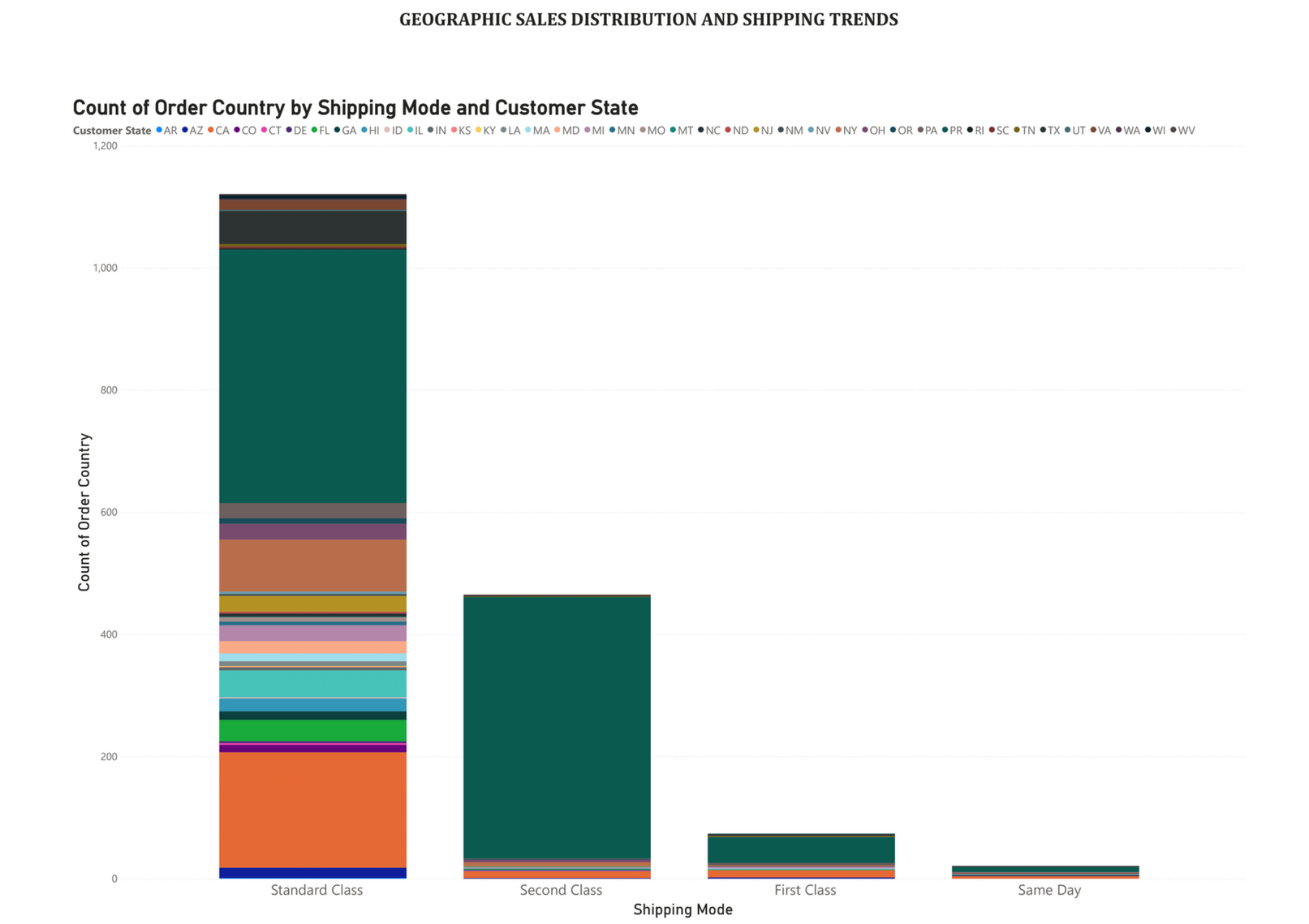
VISUALIZATION

- To determine which country has the most sales among the countries.
- Order country
- sales per customer



VISUALIZATION

- To determine which country avails the most used shipping mode among standard class, second, first, and same day.
- Customer State
- Order country
- Shipping mode



Thank you very much!

BSIT 3-5: Group 10