Analyzing Sales and Customer Data for Adidas Products

MIS 634 C

Business Integration and Data Integration

Created by:
Ancy Jacob
Neha Sinha
Priyanka Barde
Ritesh Patil

Table of Contents

Introduction

Adidas is a multinational corporation that designs and manufactures sportswear, footwear, and accessories. The company has a global presence and is known for its innovative designs, high-quality products, and strong brand reputation. As with any business, analyzing sales and customer data is critical for Adidas to understand their customer base and to make informed decisions that can drive growth.

The goal of this project is to analyze sales and customer data for Adidas products to gain insights into their sales performance, customer behavior, and trends. This analysis will provide valuable insights that can be used to develop effective marketing strategies, improve customer engagement, and optimize product offerings.

To accomplish this goal, we will collect and analyze various types of data such as sales data, customer data, and product data. Sales data will include information on the quantity and revenue generated from the sales of Adidas products, while customer data will include demographics, purchase history, and customer feedback. Product data will include information on product features, pricing, and popularity.

Using statistical analysis and data visualization techniques, we will examine the relationships between different variables and identify patterns and trends. This analysis will help us answer key questions such as: Which products are selling the most? Who are the most valuable customers? What marketing channels are most effective in driving sales?

Overall, this project will provide valuable insights into the performance of Adidas products and the behavior of their customers. By understanding these insights, Adidas can make data-driven decisions that will improve their sales performance, enhance customer engagement, and strengthen their brand reputation.

Objective

The objective of this project is to analyze the Adidas Clean dataset and uncover valuable insights that can inform strategic and tactical business decisions. Through a thorough examination of the sales data, we aim to identify the most popular product categories and individual products that are driving revenue growth for the company.

One key objective of this project is to determine the sales method that has generated the highest revenue in each city. This information can be used to optimize sales channels and develop targeted marketing strategies that are tailored to the unique characteristics of each market.

Another objective is to determine the total revenue generated by each retailer for the top 3 selling products. This analysis can help Adidas identify the most successful retailers and develop partnerships that can drive further revenue growth. Additionally, this information can be used to identify potential areas of improvement for retailers that may be struggling to sell Adidas products.

In order to achieve these objectives, we will use statistical analysis and data visualization techniques to examine the relationships between different variables and identify patterns and trends. This analysis will provide insights into the performance of Adidas products across different markets and retailers, and will help guide strategic and tactical business decisions that can improve revenue growth and enhance the company's competitive position.

Overall, this project represents an important opportunity for Adidas to gain valuable insights into their sales performance and customer behavior. By leveraging these insights, the company can make data-driven decisions that will enhance their brand reputation and drive long-term growth.

Dataset

The Adidas US Sales Dataset is a comprehensive collection of sales data that includes information on product categories, regional sales, sales methods, retailer total sales, sales by trend, and regional units sold. The dataset was sourced from Kaggle, a platform for data science and machine learning enthusiasts.

The Product Categories variable contains information on the different types of Adidas products sold in the United States. This includes sportswear, footwear, and accessories.

The Regional Sales based on Product variable provides insights into the performance of different product categories in different regions of the United States. This information can be used to identify regions that are driving revenue growth and develop targeted marketing strategies that are tailored to the unique characteristics of each market.

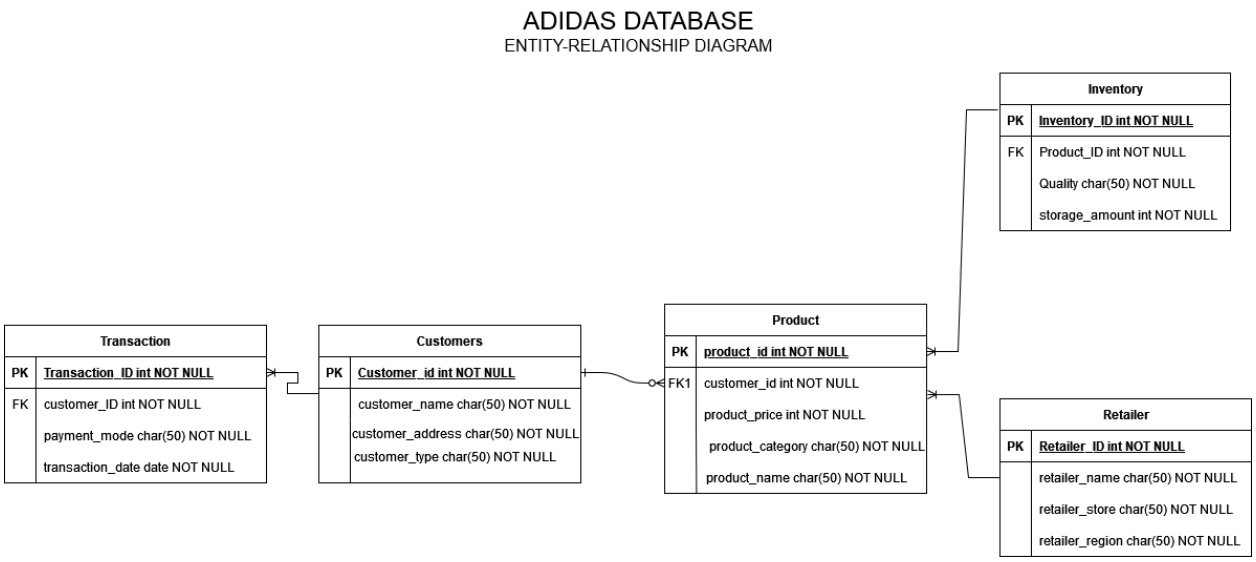
The Sales Method variable indicates the various channels through which Adidas products are sold in the United States. This includes online sales, direct sales, and sales through retail partners.

The Retailer Total Sales variable provides insights into the performance of different retailers in the United States. This information can be used to identify successful retail partners and develop partnerships that can drive further revenue growth.

The Sales by Trend variable provides insights into the performance of Adidas products in different fashion trends. This information can be used to identify popular fashion trends and develop products that are tailored to the preferences of different consumer segments.

Finally, the Regional Units Sold variable provides information on the number of Adidas products sold in different regions of the United States. This information can be used to identify regions that have high demand for Adidas products and develop supply chain strategies that can ensure adequate product availability.

Entity Relationship Diagram



The above diagram shows us the Entity relationship diagram for the Adidas dataset.

Entities: In this dataset, there are five entities listed as follows -

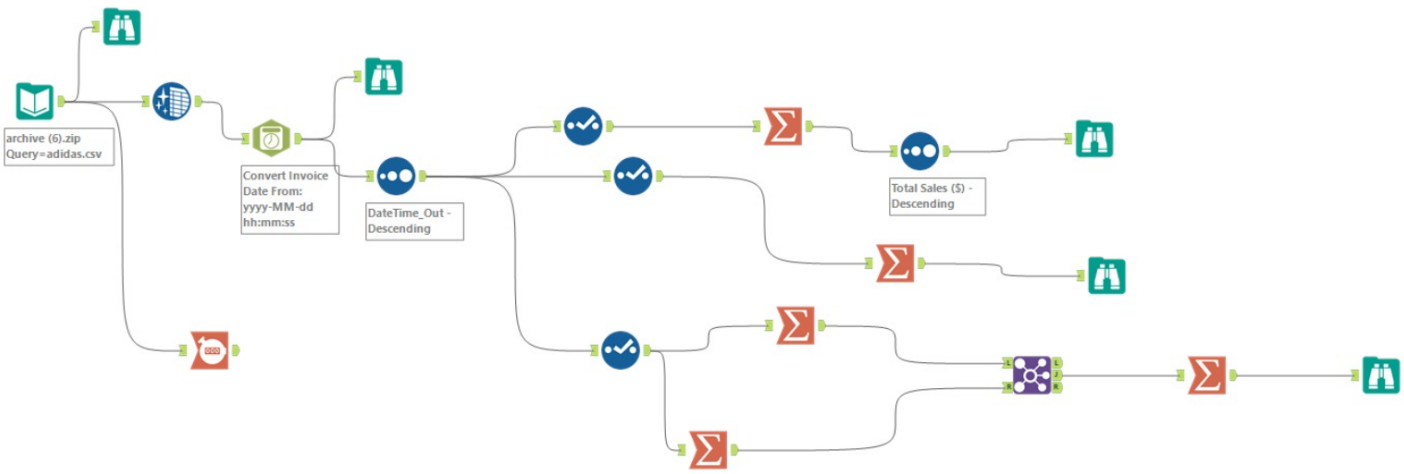
- Transaction (Transaction_ID (PK), payment_mode, transaction_date)
- Customers (Customer_ID (PK), customer_name, customer_address, customer_type)
- Product (product_ID (PK), product_price, product_category, product_name)
- Inventory (Inventory_ID (PK), Quality, storage_amount)
- Retailer (Retailer_ID (PK), name, store, region)

Relationships:

- Transaction:Customers - Many to One
- Customers:Product - One to Many
- Product:Inventory - Many to One
- Product:Retailer - Many to One

Alteryx:

Alteryx is a data analytics and workflow automation tool that allows users to blend, clean, and analyze data from multiple sources.
Alteryx flow steps include inputting data, selecting, filtering,sorting, summarizing, joining, creating formulas, and outputting data. There are also tools for data preparation, spatial data, and integration with other software platforms.



Tools Used in Alteryx:

- **Input data**
- **Browse**
- **Count**
- **Date and Time**
- **Sort**
- **Select**
- **Summarize**
- **Join**

3 of 3 Fields | Cell Viewer | 30 records displayed, 869 bytes

Record	Region	Total Sales ($)	Product
1	West	55014604	Men's Street Footwear
2	West	52191037	Women's Apparel
3	Northeast	51024984	Men's Street Footwear
4	West	50006334	Men's Athletic Footwear
5	West	42520104	Women's Street Footwear
6	Midwest	38322793	Men's Street Footwear
7	Northeast	37543055	Women's Apparel
8	Southeast	36019231	Men's Street Footwear
9	West	35694000	Men's Apparel
10	West	34517062	Women's Athletic Footwear
11	Southeast	31491156	Women's Apparel
12	South	29607182	Women's Apparel
13	Northeast	28874237	Men's Athletic Footwear
14	South	28444555	Men's Street Footwear
15	Midwest	28206342	Women's Apparel
16	Southeast	27777019	Men's Athletic Footwear
17	Northeast	25744371	Men's Apparel
18	South	25710544	Men's Athletic Footwear
19	Southeast	24461480	Men's Apparel
20	Northeast	23341173	Women's Street Footwear

Record	Product	Sum_Units Sold
1	Women's Apparel	433827
2	Men's Apparel	306683
3	Men's Athletic Footwear	435526
4	Women's Street Footwear	392269
5	Women's Athletic Footwear	317236
6	Men's Street Footwear	593320

Record	Retailer ID	Retailer	Sales Method	Sum_Right_Sum_Units Sold
1	1185732	Kohl's	In-store	510615
2	1128299	Foot Locker	In-store	179375
3	1185732	Amazon	In-store	510615
4	1185732	Walmart	Outlet	261193
5	1185732	Kohl's	Outlet	261193
6	1128299	Walmart	In-store	179375
7	1185732	Sports Direct	Online	479123
8	1128299	Walmart	Outlet	214818
9	1185732	Foot Locker	Outlet	261193
10	1128299	Kohl's	Online	231709
11	1197831	Walmart	Online	205647
12	1185732	Amazon	Online	479123
13	1185732	Sports Direct	Outlet	261193
14	1197831	Sports Direct	Outlet	275723
15	1185732	West Gear	Online	479123
16	1128299	Kohl's	Outlet	214818
17	1128299	Foot Locker	Online	231709
18	1185732	Kohl's	Online	479123
19	1128299	Walmart	Online	231709
20	1185732	West Gear	In-store	510615

MongoDB

Data is stored in a JSON-like format using the NoSQL database management system MongoDB. Large and complicated data sets can be accommodated thanks to its excellent scalability and flexibility. MongoDB supports dynamic schema design and can handle unstructured data. Sharding and replication features are included for high availability and performance in large-scale deployments, and it interfaces nicely with well-known web development technologies. Overall, MongoDB is a strong and adaptable database system with benefits over conventional relational databases.

To import the Adidas dataset with 4035 records to MongoDB, we can follow these steps:

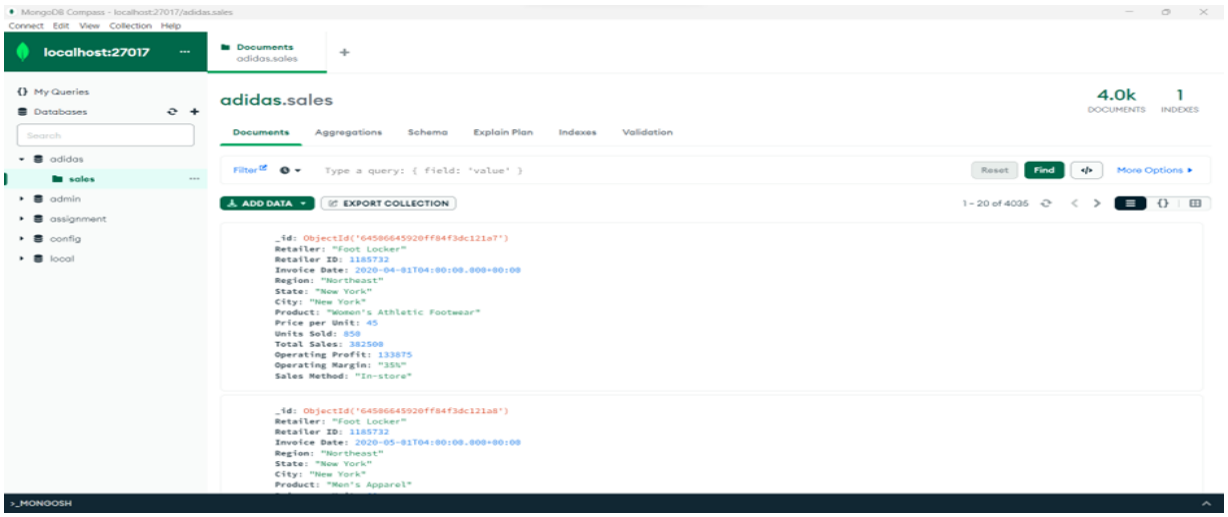
1. Open the MongoDB shell or MongoDB Compass.

2. Create a new database named "Adidas" by running the following command: use **Adidas**

3. Create a new collection named "sales" by running the following command: db.createCollection("**sales**")

4. Import the dataset to the "sales" collection using the following command:mongoimport --db Adidas --collection sales --type csv --headerline --file /path/to/adidas_dataset.csv. Replace "/path/to/adidas_dataset.csv" with the actual path to the dataset file on your system.

After executing the above commands, the Adidas dataset should be successfully imported to the "sales" collection in the "Adidas" database.



After doing the analysis:

Q1.Sales method has generated the highest revenue in each city?

```
db.sales.aggregate([
 {
        $group: {
        _id: { city: "$city", sales_method: "$sales_method" },
        total_revenue: { $sum: "$revenue" }
        }
 },
 {
        $sort: {
        "_id.city": 1,
        total_revenue: -1
        }
 },
 {
        $group: {
        _id: "$_id.city",
        top_sales_method: { $first: "$_id.sales_method" },
        top_revenue: { $first: "$total_revenue" }
        }
```

```
 }
])
```

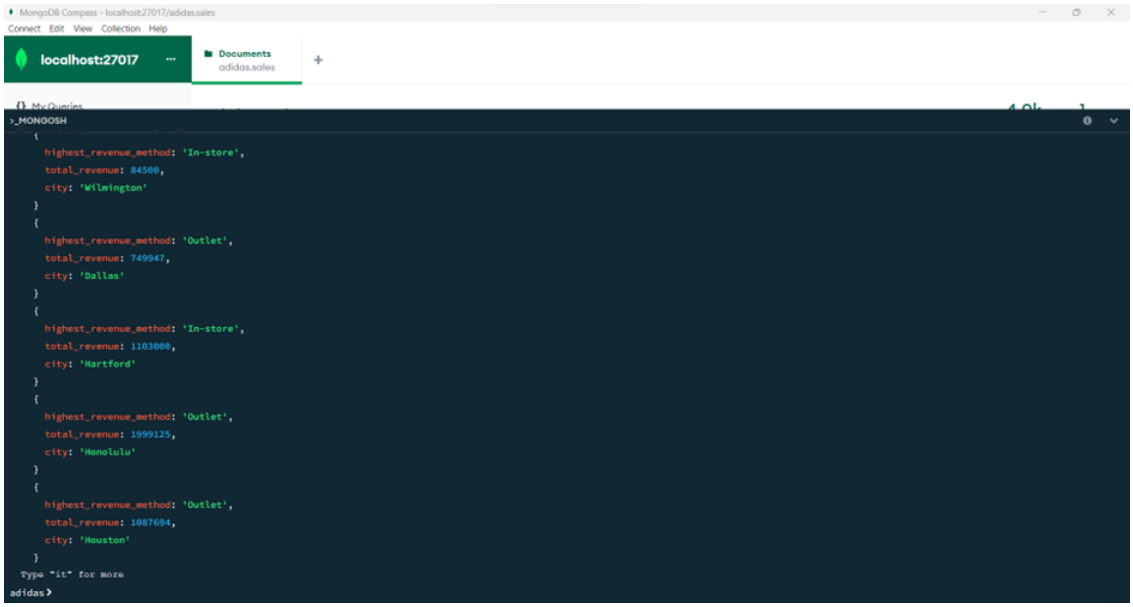
Q.2) Total revenue generated by each retailer for the top 3 selling products?

```
db.sales.aggregate([
 {
        $match: {
        product: { $in: [ "product1", "product2", "product3" ] }
        }
 },
 {
        $group: {
        _id: { retailer: "$retailer", product: "$product" },
        total_revenue: { $sum: "$revenue" }
        }
 },
 {
        $sort: {
        "_id.product": 1,
        total_revenue: -1
        }
 },
 {
```

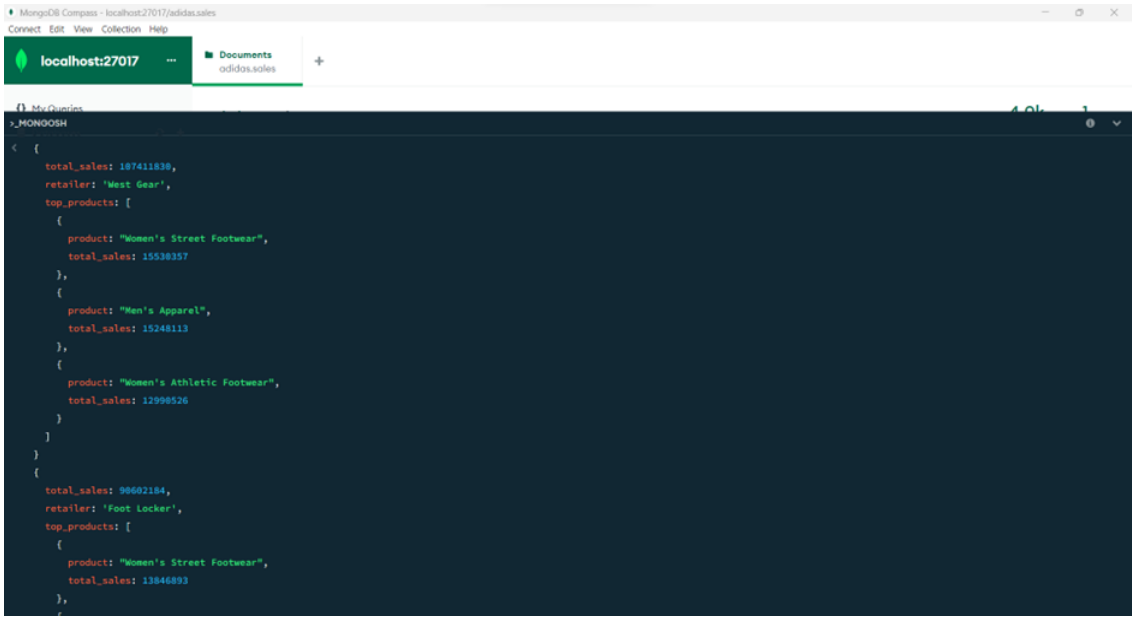
```
        $group: {
        _id: "$_id.retailer",
        top_products: {
        $push: {
        product: "$_id.product",
        revenue: "$total_revenue"
        }
        },
        total_revenue: { $sum: "$total_revenue" }
        }
 },
 {
        $project: {
        _id: 0,
        retailer: "$_id",
        top_products: { $slice: [ "$top_products", 3 ] },
        total_revenue: 1
        }
 }
])
```

OUTPUT:

QUESTION 1



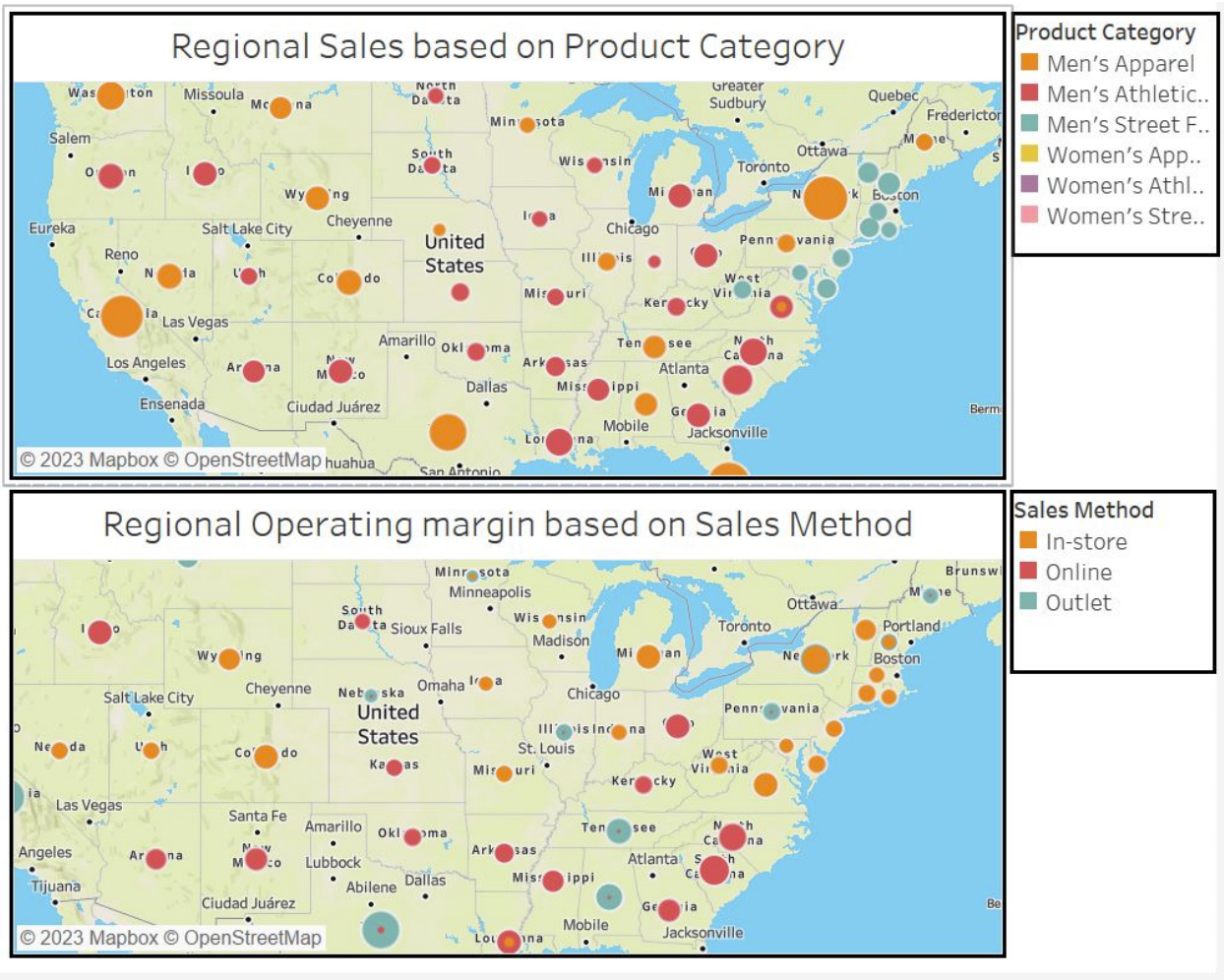
QUESTION 2



Visualizations

We have created the visualizations using Tableau visualization tool.

Visualization 1



Purpose:

This visualization represents the regional sales and operating margin according to the product category and the sales method.

In the first graph, we can see the regional sales based on the product category and in the second, we can see the regional operating margin based on the sales method.

Information:

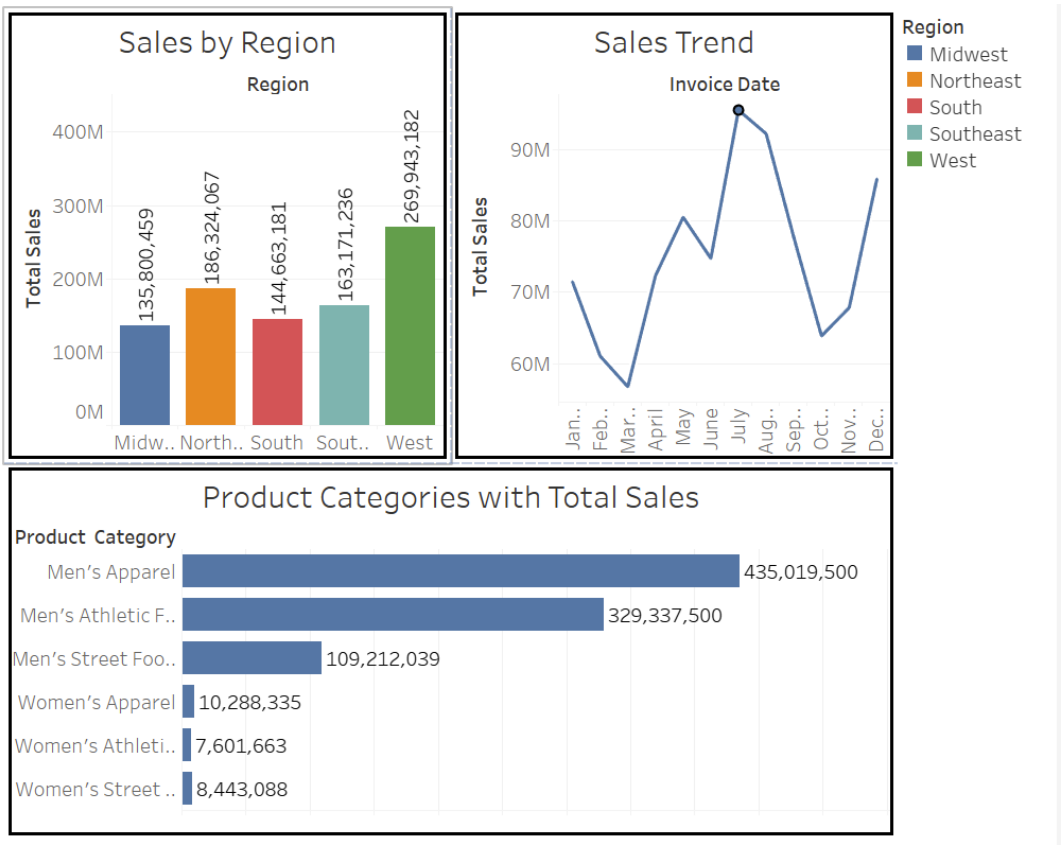
This is a geographic visualization. In the first graph, different colors represent different product categories and in the second one, different colors represent sales methods used.

Insights:

Some insights we can get from these graphs are -

- Men's street footwear products are more famous in the east coast of the United States.
- Sales method of outlets is not very famous uniformly all over the united states.
- Highest sales are done in New York, California, and Texas.

Visualization 2



Purpose:

This dashboard comprises 3 visualizations. The first graph gives us information about sales done by region. Then Monthly Sales Trend is given. The last graph gives the total sales based on the various Product Categories.

Information:

This dashboard consists of two bar graphs, horizontal and vertical both and a line chart. Line chart best represents any type of sales trend.

Insights:

Some of the insights we can take from this dashboard are:

- Highest sales of over 90 Million USD is done in July.
- The most famous category amongst customers is Men's Apparel as we can see in the second visualization.
- Also, highest sales are gained in the west region of the USA

Conclusion

In conclusion, we have analyzed and manipulated Adidas dataset using Alteryx, MongoDB, and Tableau. We have cleaned the data, analyzed it using MongoDB and created visualizations to take insights from them to be able to make more helpful decisions for the improvement of revenue. We learned how to completely do the data analytics process.