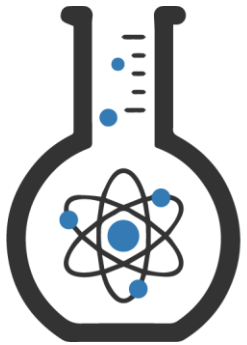


Astradrel E-Commerce

Business Data Analysis Report

Prepared by: Ahmad Khairi Bin Ahmad Khir
Email: ahmadkhairi.khir@gmail.com

Astradrel
E-Commerce

Business Function

Astradrel E-Commerce is a company that specialized in online commerce platform and provide delivery option for its customer. As part of company core businesses, its incorporate delivery arrangement for the product that the customer bought straight to their addresses. The company operated businesses unit that handled any products related to technology, furniture and office supplies and the company main demographics mostly are other company that requires these products. From the sales data the company had from 2008 until 2012, the company wants to find out what metrics and factors that can contribute to boosting their sales and increases profit. A report analysis will be needed to gain insights of the data and make use of the findings to drive business decision for the company.

Astradrel
E-Commerce

Report Objectives

The objective of this data analysis report is to gather information based on sales record data from 2008 until 2012:

1. To identify the business unit that generate the most sales.
2. To identify which product lines that needed to be focused on.
3. To explore the relationship between promotional discount and people spending behavior.
4. To determine state and market segment that most profitable.

Problem Statement

As to answer the point presented in the objectives of this report, several problem statement can be created that help with the insight findings.

1. Which business unit bring the most sales to the company so that the company can prioritize more on the unit in future based on past sales?.
2. In order to increase the company sales for specific product, they need to know which trending product that customer sought after in their catalogues?. The company then can focus more on these product by increasing more supply and generate more revenues in future based on the historical data.
3. Most e-commerce uses promotional events like discount on certain product to boost more traffic to their platform. Does these events affect the spending behaviour of the people to buy more product?. Which discount range group that are bringing most sales to the company?.
4. Demographics and target audience can help company in determining how their marketing should be forming around these factors. Based on the data of past sales, which state in United States and market segment of the customers contributes to the highest order count on the company platform?.

Methodology

Tools that are used in these data analysis process on generating this report are:

1. **Jupyter Notebook** (Python) : For data cleaning, data transformation and export purposes that were done in this analysis
2. **Tableau**: Business Intelligence tools used to create the visualization and dashboard for this analysis.

Astradrel
E-Commerce

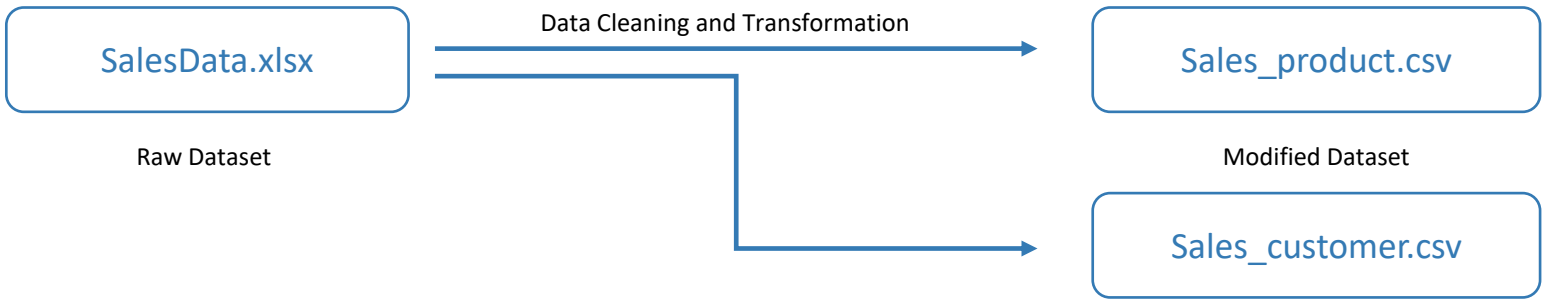
Methodology

Data Cleaning and Transforming:

The dataset received needed some cleaning to carried out due to duplicated column and improper formats. The columns that contains missing values were dropped, new columns were added to help understand the data better and dataset were then processed into two new dataset named “Sales_product.csv” and “Sales_customer.csv” for easier analysis.

Data Exporting:

The processed modified dataset were then exported into Tableau to create visualization. The “Sales_product.csv” contains the details of order transaction records and “Sales_customer” contains the customer details with “Order id” connecting these two dataset.



Astradrel
E-Commerce

Methodology

Using Tableau to create visualization and dashboard:

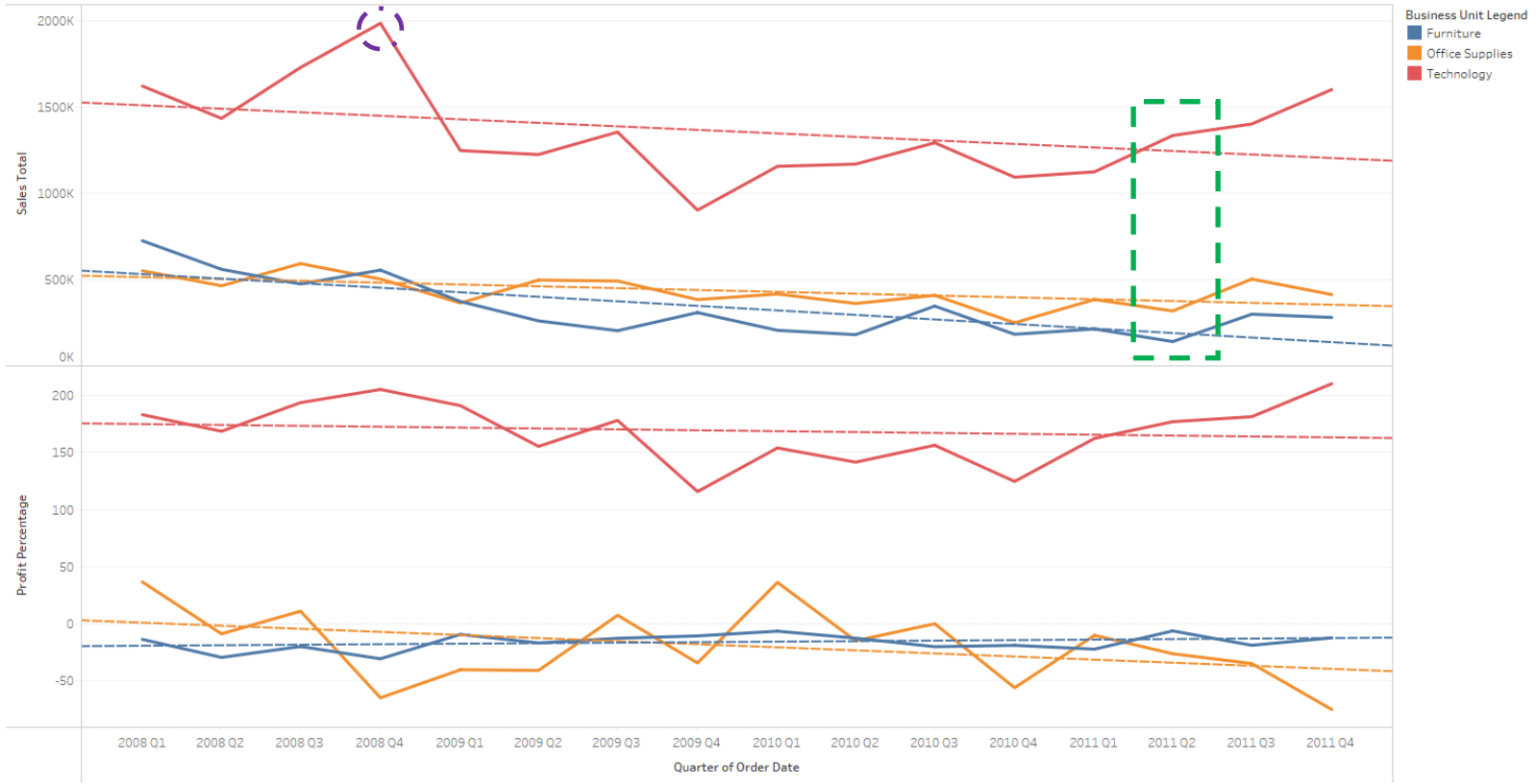
With the dataset readied for analyzation, both data sets were loaded into tableau and several worksheet containing the visualization which answer the problem statement as above. The visualization were created based on suitable type of graph to explain the metrics and legend were put in place to helps the viewer to understand the visual and interpret from it. From the worksheet created, an interactive dashboard are formed using all the worksheet visual created. The visual and dashboard created were as followed:

Index	Visualization Title
1	Quarterly Total Sales and Profit Percentage Trend
2	Top 10 Product Line Total Sales
3	Discount Utilization Rate over Discount Group
4	Quarterly Discount Utilization Trend
5	Order Traffic for each States
6	Order Traffic over Market Segment
7.	Astradrel E-Commerce Business Dashboard

Astradrel
E-Commerce

Quarterly Total Sales and Profit Percentage Trend

The **decreasing trendline** for all business unit for total sales and profit percentage shows a sign of small declining over period of years. The company should take **affirmative action** in addressing the issues.



The trends of sum of Sales Total and sum of Profit Percentage for Order Date Quarter. Color shows details about Business Unit. The data is filtered on Order Date Year, which has multiple members selected. The view is filtered on Business Unit, which keeps Furniture, Office Supplies and Technology.

Total Sales insight

The Technology business unit records an **all-time total sales high** in **2008 Q4**. All other business unit seems to maintain relatively flat total sales throughout the year

Even though all business unit **gain some traction** in total sales at the start of **2011 Q2**, the overall trendline still shows that the company had total sales plummet on previous year. The total sales plummet below the trendline are present greatly in Technology products

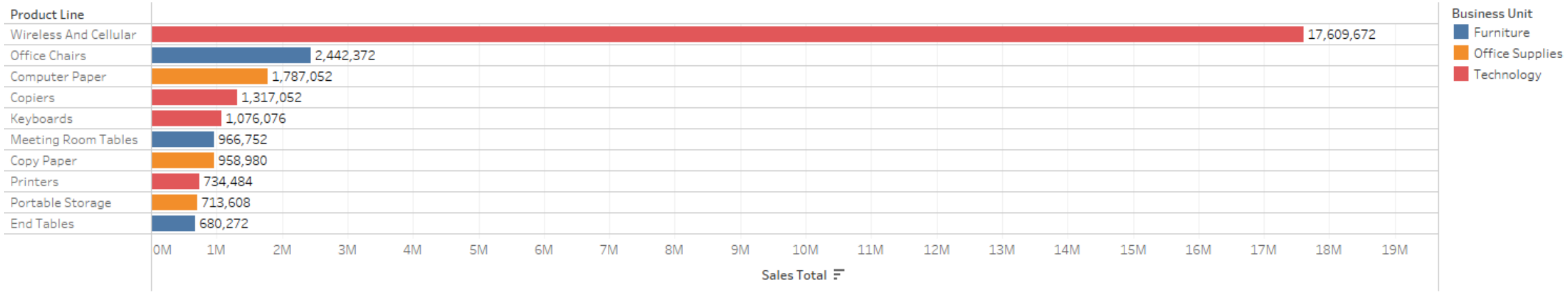
Profit Percentage insight

At the start of **2011 Q2**, as the total sales for all business unit increases, only the profit percentage to total sales of **Technology** shows sign of increasing whereas the **Furniture** remain stagnant and office supplies losses. This shows that **Office Supplies** while the total sales increases it does not reflect in profit gain due to the fact of cost of goods are greater than the revenue generated from the sales.

Astradrel
E-Commerce

Top 10 Product Line Total Sales

In the bar chart below, we can see top 10 product line that generate the most total sales. **Wireless And Cellular** had the highest total sales with staggering **17,609,672 total sales** throughout the year which made up almost the entire total sales generated by the company. The company should prioritize on this product line to maximize the sales output.

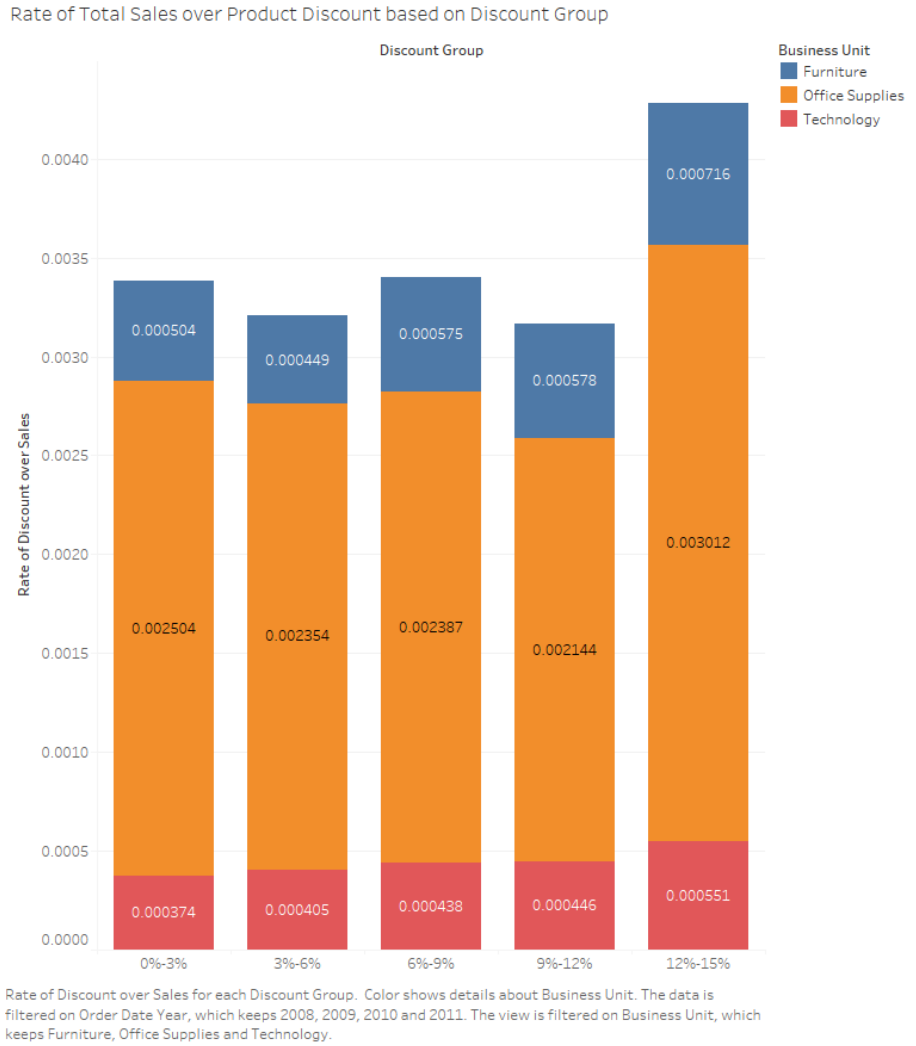


Sum of Sales Total for each Business Unit broken down by Product Line and Product Group. Color shows details about Business Unit. The marks are labeled by sum of Sales Total. The data is filtered on Order Date Year and TOP 10. The Order Date Year filter keeps 2008, 2009, 2010 and 2011. The TOP 10 filter keeps True. The view is filtered on Business Unit, which keeps Furniture, Office Supplies and Technology.

From the bar chart, we can see that trending product line that the customer sought after are the product line that belongs to the **Technology** business unit with 4 product line were in the top 10 product sold by the company. By increasing the inventory for these product line from this business unit, the company may increase the total sales in future.

Astradrel
E-Commerce

Discount Utilization Rate over Discount Group



Rate of Discount over Sales for each Discount Group. Color shows details about Business Unit. The data is filtered on Order Date Year, which keeps 2008, 2009, 2010 and 2011. The view is filtered on Business Unit, which keeps Furniture, Office Supplies and Technology.

Promotional event such as discount can helps in boosting sales. In viewing the relationship between the rate of discount value over total sales **(Discount Utilization)** categorizes for each discount group shown on the stacked bar chart on the left.

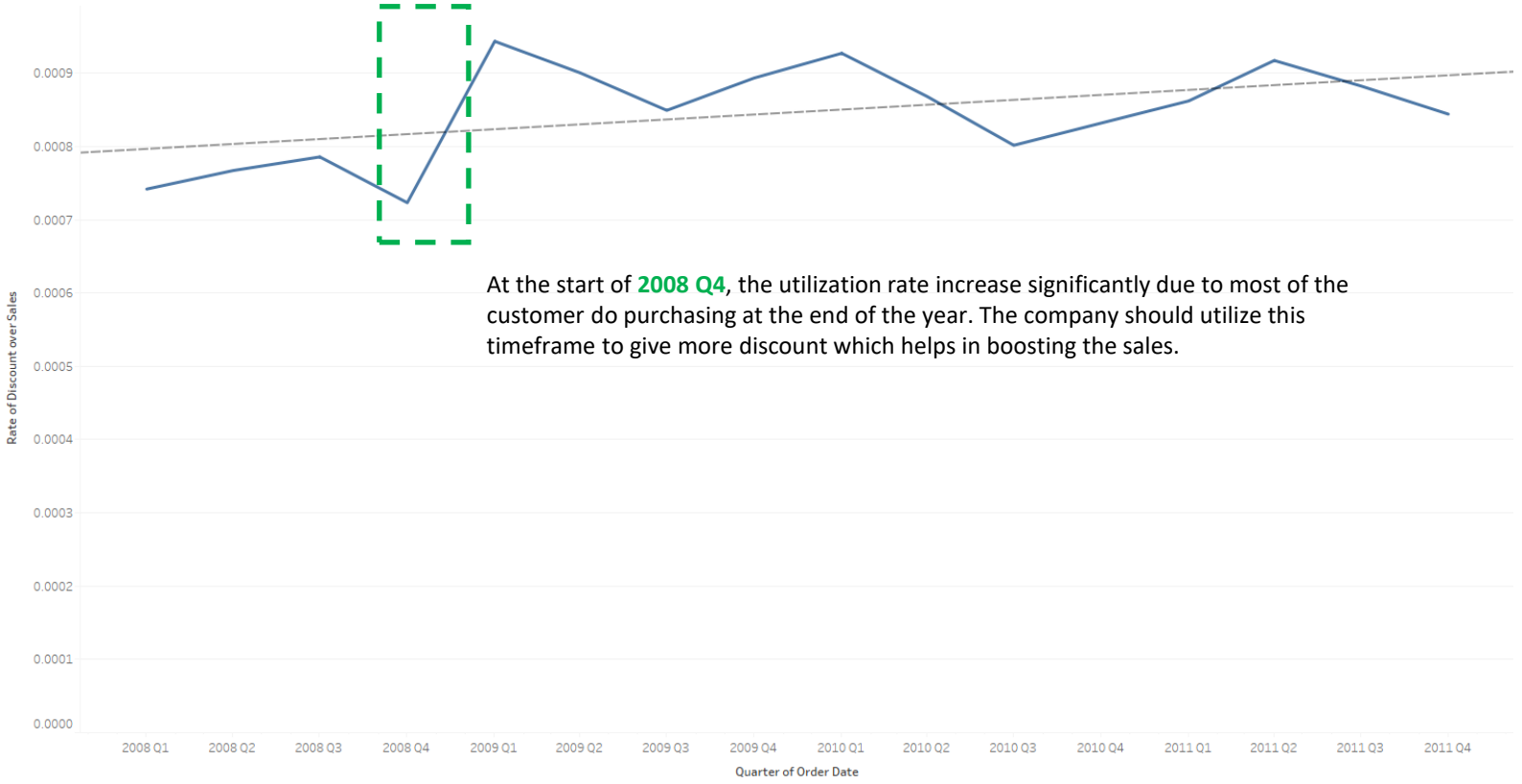
The total sales for each business unit increases greatly when the discount offered were 12% until 15%. With **0.000716**, **0.003012** and **0.000551** for respective business unit product, we can see that most of the customer responded to the higher promotional discount which translated into how many total sales generated over that discount value.

From the chart observation, most of the customer responded to the higher discount value for **Office Supplies** product which is why there were sales increased for that discount value. Overall, the **Furniture** and **Technology** product retain the same number of discount utilization regardless of the discount value.

Astradrel
E-Commerce

Quarterly Discount Utilization Trend

In order to support the finding on previous bar chart, we take a look in the discount utilization over the year. The trendline shows in increases in customer's respond towards promotional discounts which is translated into increasing **discount utilization rate**. The company should focus more on incorporating more promotional discount in future year as to increases customer traffic and total sales.



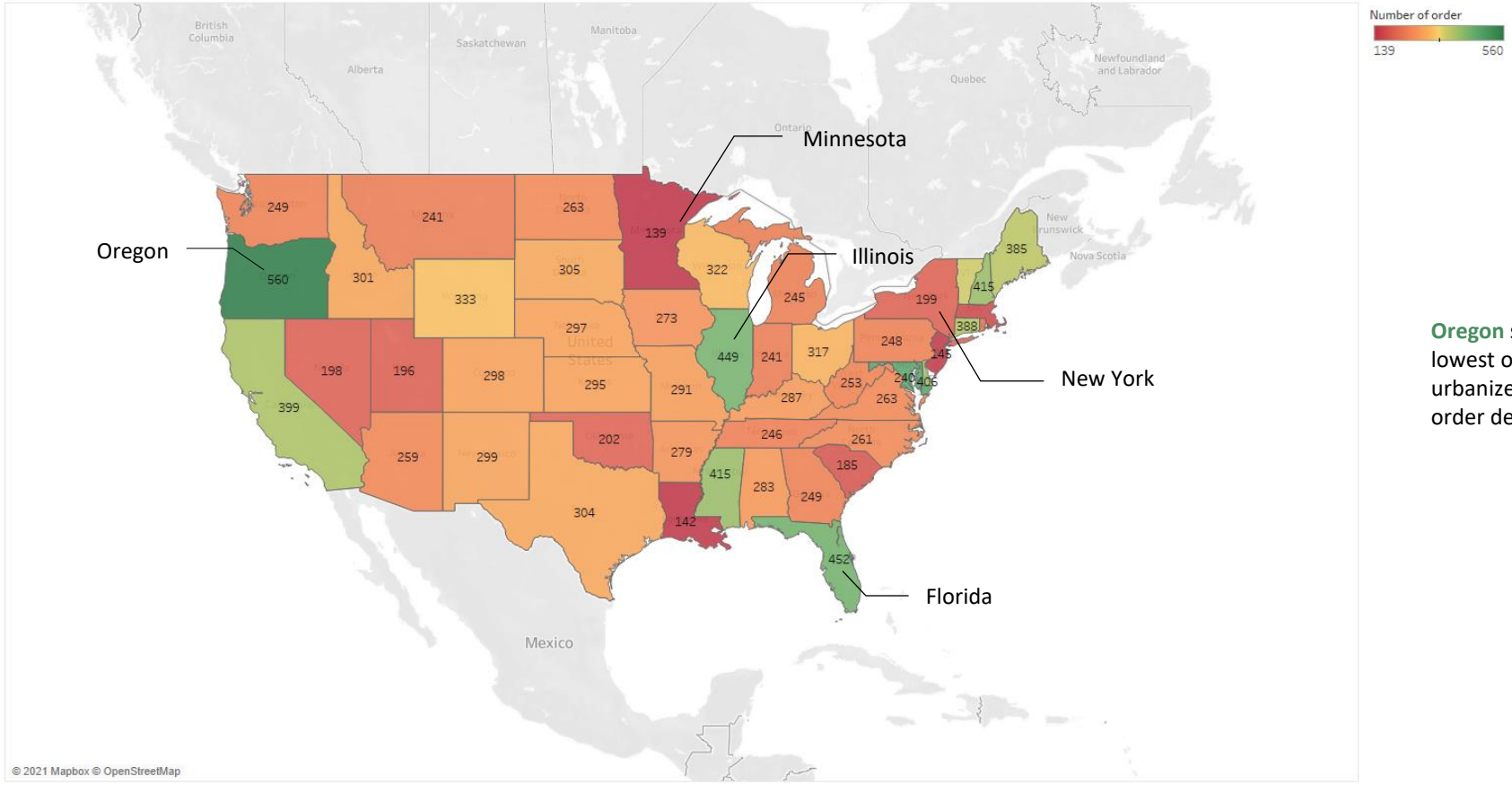
At the start of **2008 Q4**, the utilization rate increase significantly due to most of the customer do purchasing at the end of the year. The company should utilize this timeframe to give more discount which helps in boosting the sales.

The trend of Rate of Discount over Sales for Order Date Quarter. The data is filtered on Business Unit and Order Date Year. The Business Unit filter keeps Furniture, Office Supplies and Technology. The Order Date Year filter keeps 2008, 2009, 2010 and 2011.

Astradrel
E-Commerce

Order Traffic for each States

The heatmap below shows the distribution of the order traffic based on each state of United States of America during the period of 4 years. From the observation made, state like **Oregon**, **Illinois** and **Florida** had the **highest order traffic** compared to other states which beneficial for the company in terms of total sales by prioritizing these state. The states with the **lowest order traffic** also needed some marketing measure to increase the customer awareness on the company product.

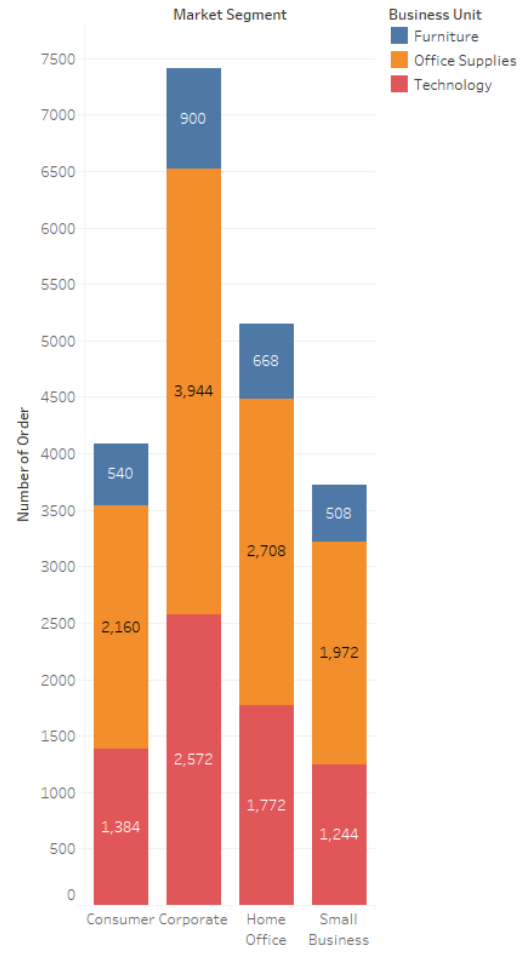


Map based on Longitude (generated) and Latitude (generated). Color shows count of Name. The marks are labeled by count of Name. Details are shown for State. The data is filtered on Business Unit and Order Date Year. The Business Unit filter has multiple members selected. The Order Date Year filter has multiple members selected.

Oregon state had the highest order traffic while **Minnesota** had the lowest order traffic. The higher order traffic usually comes from more urbanized state except for **New York** state which had lower traffic order despite being the most urbanized state of all.

Astradrel
E-Commerce

Order Traffic over Market Segment



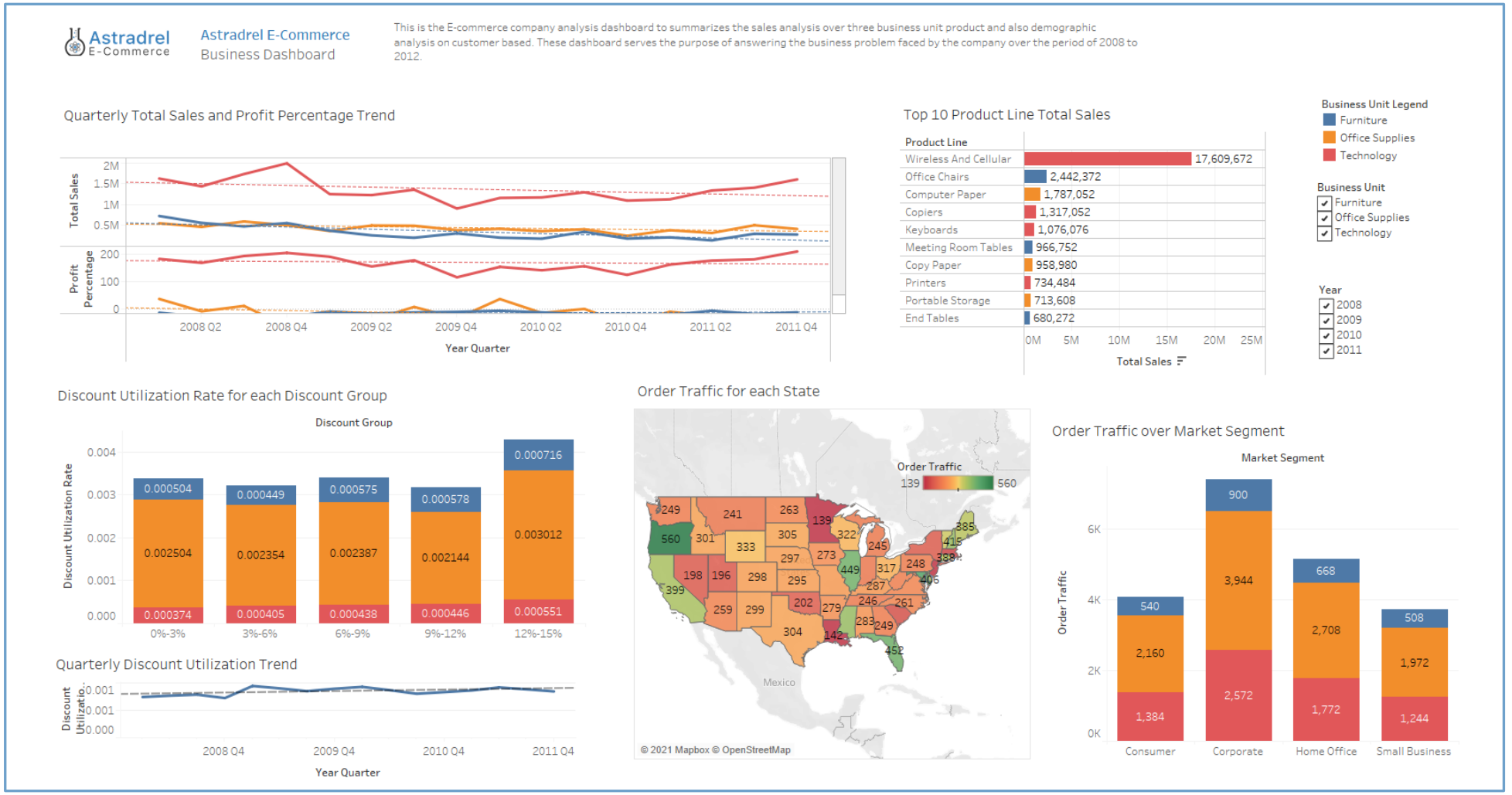
Count of Order Id (Sales customer.csv) for each Market Segment. Color shows details about Business Unit. The data is filtered on Order Date Year, which has multiple members selected. The view is filtered on Business Unit, which has multiple members selected.

From the observation made into what market segment made up of the total order traffic throughout the 4 years, the highest market segment would be the **Corporate**. It is worth noting that the **Corporate** had substantial amount of order on **Technology** product compared to other market segment.

Although Corporate had the largest amount of order compared to other market segment, these large amount may come from the size of the Corporate market segment itself and it does not mean that other market segment perform poorly from lower order traffic. All the market segment should be given equal amount of prioritization and marketing campaign measure such as discount promotion or product awareness should be done to increases all order traffic for all market segment.

Astradrel
E-Commerce

Astradrel E-Commerce Business Dashboard



The dashboard created were used to track important business metrics that gave insight on what is happening on the business performance. The real time tracking with the ability to feed with new data makes the dashboard ideal for driving business decision every set of time period.

The dashboard can be viewed on:

https://public.tableau.com/views/AstradrelE-Commerce/AstradrelE-CommerceDashboard?:language=en-US&:display_count=n&:origin=viz_share_link

Astradrel
E-Commerce

Appendix – Data Cleaning Python Code

```
# Import the necessary libraries

import pandas as pd
import numpy as np
import os

## Data Cleaning

os.getcwd()

df_sales = pd.read_excel("SalesData.xlsx")

column_list = list(df_sales.columns)

column_list

df_sales.head()

df_sales.info()

# drop columns that contain missing values
df_sales.isna().sum()

df_sales_modified = df_sales.dropna(axis=1,how='all')

df_sales_modified.info()

# Columns that needed to be dropped due to redundancy and duplication

dropped_columns = ['Transaction (2)', 'Gross Profit Percentage (2)', 'Sales Total (Numeric) (2)',
   'Gross Profit (Numeric) (2)', 'Profit as a Percent of Sales (2)', 'Filtered out at 2:36:28 PM (2)',
   'Sales Total Percent Format (2)', 'Gross Profit Percent Format (2)',
   'Comparison Period (2)', 'Volume (2)', 'Transaction (3)', 'Time Period (2)', 'Product hierarchy (2)',
   'Gross Profit Percentage (3)', 'Sales Total (Numeric) (3)', 'Gross Profit (Numeric) (3)',
   'Profit as a Percent of Sales (3)', 'Sales Total Percent Format (3)', 'Gross Profit Percent Format (3)', 'Comparison Period (3)', 'Volume (3)',
   'Products (2)', 'Transaction', 'Gross Profit Percentage', 'Sales Total (Numeric)', 'Gross Profit (Numeric)',
   'Sales Total Percent Format', 'Gross Profit Percent Format', 'Volume', 'Inventory Status', 'Comparison Period',
   'Delivery Time', 'Forecast Month', 'Forecast Quarter']

df_sales_modified = df_sales_modified.drop(columns=dropped_columns,axis=1)
```

```
# With the removal of the unnecessary columns from the dataset, the dataset only retain the relevant informations

print("The total number of columns before the removal is {}".format(df_sales.shape[1]))
print("The total number of columns after the removal is {}".format(df_sales_modified.shape[1]))

# Lets change the value for the order priority to be standardized with the rest of datasets

value_change = list(df_sales_modified['Order Priority'].unique())
value_change

value_new = ['Low', 'Medium', 'Unspecified', 'High', 'Urgent']
df_sales_modified['Order Priority'] = df_sales_modified['Order Priority'].replace(value_change,value_new)

df_sales_modified.info()

# Change decimal places for Profit Percent of Sales column to 2 decimal places

df_sales_modified['Profit as a Percent of Sales'] = df_sales_modified['Profit as a Percent of Sales'].round(2)


# Change the Discount column into percentage

df_sales_modified['Discount'] = (df_sales_modified['Discount']*100)

## Data Transformation

# Creating Discount Group Column

df_sales_modified['Discount Group'] = df_sales_modified['Discount'].apply(
                        lambda x: "0%-3%" if 0<=x<3 else "3%-6%" if 3<=x<6 else "6%-9%" if 6<=x<9 else
                            "9%-12%" if 9<=x<12 else "12%-15%")

df_sales_modified['Discount Group'].unique()
```

Astradrel
E-Commerce

Appendix – Data Cleaning Python Code

```
# Creating Delivery Period

df_sales_modified['Delivery Period'] = (df_sales_modified['Delivery Date'] - df_sales_modified['Order Date']).dt.days
df_sales_modified['Delivery Period']

df_sales_modified['Delivery Period'].unique()

# Creating Delivery Remarks based on defined set rule

df_sales_modified['Delivery Remarks'] = df_sales_modified['Delivery Period'].apply(
                        lambda x: "Fast" if 0<=x<30 else "Slow" if 30<=x<120 else "Very Slow")
df_sales_modified['Delivery Remarks']

# Rename column Profit as a Percent of Sales to Profit Percentage

df_sales_modified = df_sales_modified.rename(columns={'Profit as a Percent of Sales': 'Profit Percentage'})

df_sales_modified.info()

# Create new dataset from the original dataset

# Let split the dataset into two dataset for easier analyzing

new_dataset = ['Name', 'State', 'Region', 'Market Segment']

df_sales_product = df_sales_modified.drop(columns=new_dataset, axis=1)
df_sales_customer = df_sales_modified[['Order Id', 'Name', 'State', 'Region', 'Market Segment']]

df_sales_product.head(2)

df_sales_customer = df_sales_customer.drop_duplicates().reset_index().drop(columns='index')

df_sales_customer.head(2)
```

```
# Rearranging both dataframe columns so it can be exported as tables for BI tool analyzation

rearrange_product = ['Record No','Order Id', 'Business Unit', 'Product Group', 'Product Line', 'Product', 'Discount',
            'Discount Group', 'Sales Total', 'Gross Profit', 'Profit Percentage', 'Order Date', 'Delivery Date',
            "Delivery Period", 'Delivery Remarks', 'Order Priority', 'Ship Container', 'Shipping Method']

df_sales_product = df_sales_product[rearrange_product]
df_sales_product.head(2)

df_sales_product.tail(2)

df_sales_customer = df_sales_customer[['Order Id', 'Name', 'Market Segment', 'State', 'Region']]
df_sales_customer.head(2)

# Shape of these two dataset

print('Product Table:',df_sales_product.shape)
print('Customer Table:',df_sales_customer.shape)

# Exporting the both tables to csv

df_sales_product.to_csv(r'Sales_product.csv', index = False)
df_sales_customer.to_csv(r'Sales_customer.csv', index = False)
```