

Final Project

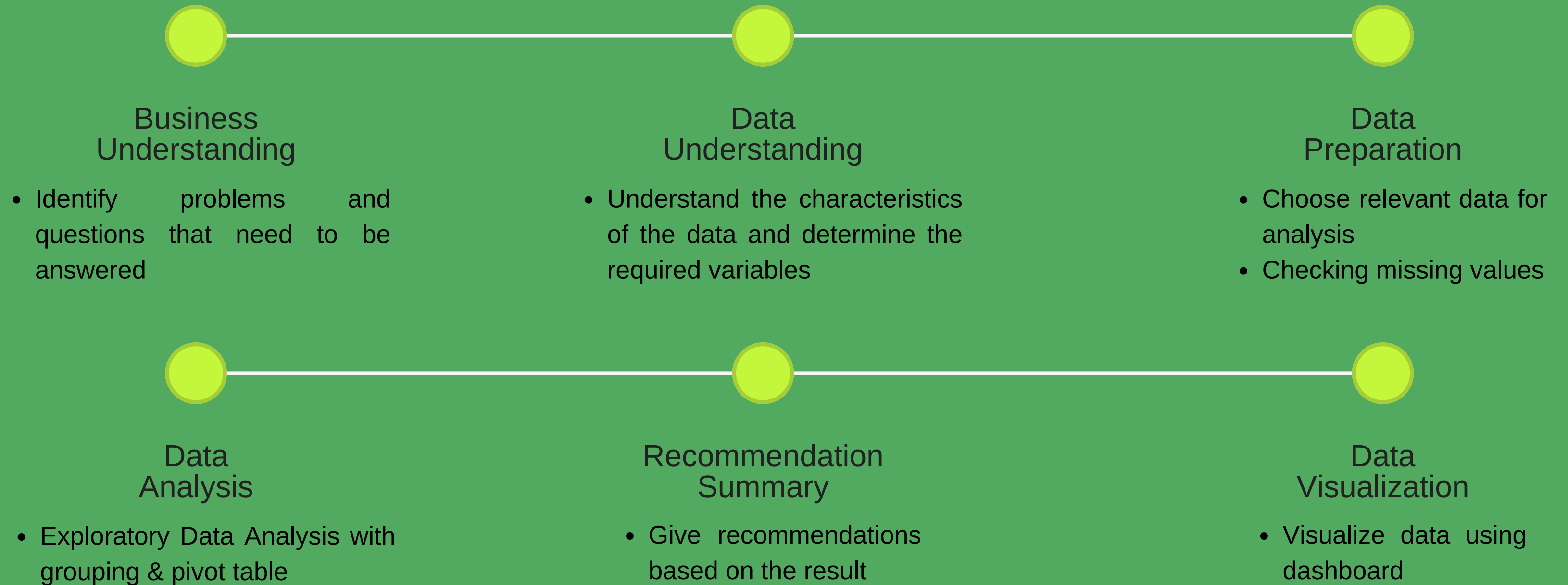
# Exploring the Sales Landscape: Optimizing Sales and Profitability

Anita Mila Oktafani

DQLab Live Class Data Analyst with SQL & Python in Google Platform



# Methodology



## Dataset Source

Superstore Sales:

[https://docs.google.com/spreadsheets/d/1DBpK4Pg3xjLNB\\_I\\_AaEsGLV0groj0XXfLzxJrCz2Imic/edit#gid=410161316](https://docs.google.com/spreadsheets/d/1DBpK4Pg3xjLNB_I_AaEsGLV0groj0XXfLzxJrCz2Imic/edit#gid=410161316)

## Tools Analysis



Google Colaboratory



Looker Studio

# Business Understanding



1. What is the overall sales and profit trend?
2. Which are the top 5 states with the most total sales?
3. Based on region, what product category has the highest profit and losses?
4. What segments are the main revenue contributors in sales?
5. Which ones do customers use the most of shipping mode?

# Data Understanding



Superstore dataset which includes product, customer, sales and discount information. The main variables used are:

- Sales: Product sale
- Quantity: The number of sales for a particular product in that order.
- Discount: The amount of discount given on the product.
- Profit: The profit generated from the transaction.
- Category : Product category (e.g., electronics, furniture, office).
- Order\_Date: The date when the order was created or the transaction was carried out.
- Ship\_Date: The date when the order was delivered
- Ship\_Mode: Type of goods delivery service
- Segment: Customer segmentation
- State: The customer's state of residence
- Region: Territory of the state located

# Data Preparation

- Checking Missing Values

```
data.isna().sum()
```

```
Order_ID      0
Customer_ID   0
Postal_Code    0
Product_ID    0
Sales          0
Quantity       0
Discount       0
Profit         0
Category       0
Sub-Category  0
Product_Name   0
Order_Date    0
Ship_Date     0
Ship_Mode     0
Customer_Name  0
Segment       0
Country/Region 0
City          0
State         0
Region        0
dtype: int64
```

There are no missing value detected

- Choose relevant data for analysis

```
data = data.drop(data.columns[:4], axis=1)
data = data.drop(['Product_Name', 'Customer_Name', 'Country/Region', 'City'], axis=1)
data.head()
```

	Sales	Quantity	Discount	Profit	Category	Sub-Category	Order_Date	Ship_Date	Ship_Mode	Segment	State	Region
0	261.9600	2	0.00	41.9136	Furniture	Bookcases	11/8/2019	11/11/2019	Second Class	Consumer	Kentucky	South
1	731.9400	3	0.00	219.5820	Furniture	Chairs	11/8/2019	11/11/2019	Second Class	Consumer	Kentucky	South
2	14.6200	2	0.00	6.8714	Office Supplies	Labels	6/12/2019	6/16/2019	Second Class	Corporate	California	West
3	957.5775	5	0.45	-383.0310	Furniture	Tables	10/11/2018	10/18/2018	Standard Class	Consumer	Florida	South
4	22.3680	2	0.20	2.5164	Office Supplies	Storage	10/11/2018	10/18/2018	Standard Class	Consumer	Florida	South

- Convert the 'Order\_Date' and 'Ship\_Date' column to datetime data type

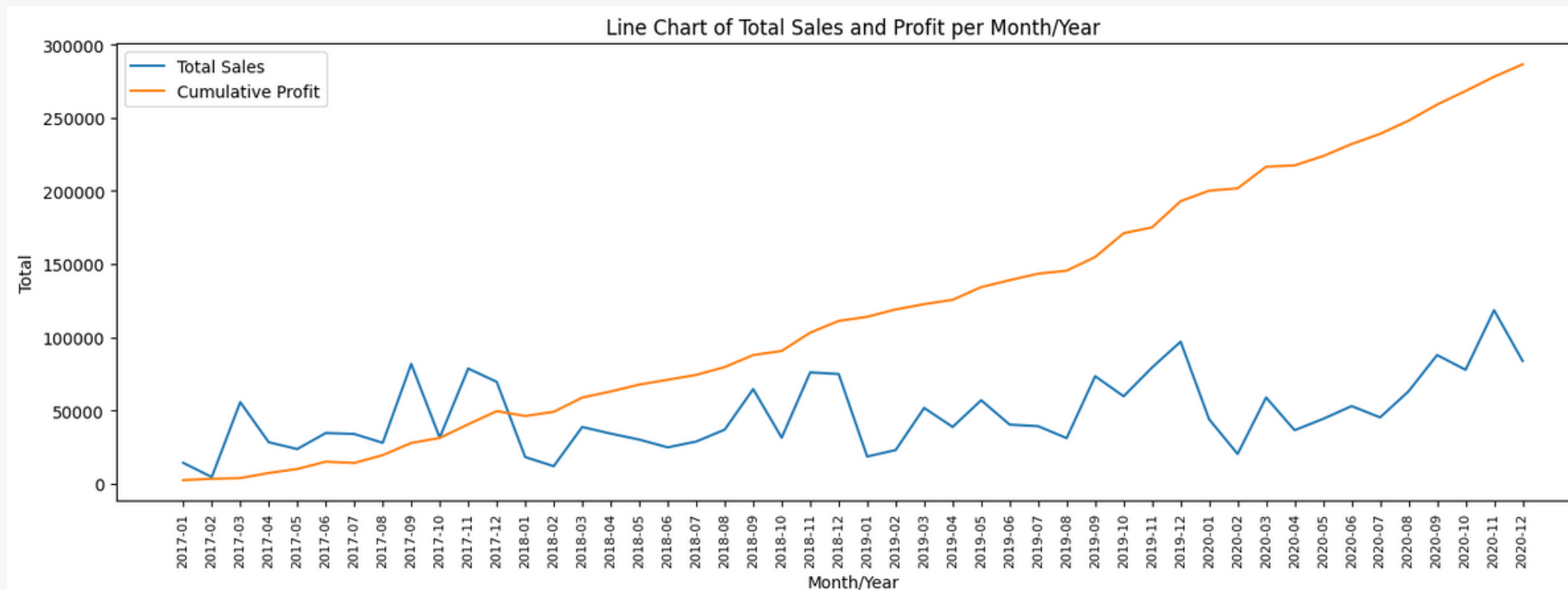
```
data["Order_Date"] = pd.to_datetime(data["Order_Date"])
data["Ship_Date"] = pd.to_datetime(data["Ship_Date"])
```



# 1. What is the overall sales and profit trend?

```
# Getting month year from order_date
data['month_year'] = data['Order_Date'].apply(lambda x: x.strftime('%Y-%m'))

# Grouping month_year by sales
data_temp = data.groupby('month_year').agg({'Sales': 'sum', 'Profit': 'sum'}).reset_index()
data_temp['Cumulative_Profit'] = data_temp['Profit'].cumsum()
```



- This **sales** level fluctuates almost the same every year which is known as a **seasonal trend**. These increases could occur due to several factors. For example, the increase in March could possibly occur due to spring promotions, the increase in September could possibly occur due to promotions for the new school year, and the increase in November/December could possibly occur due to promotions to welcome Christmas.
- Meanwhile, the **profits** continue to **increase every year**. The growth in profits annually can reflect consistent business growth, which may result from increased sales, improved operational efficiency, or a successful marketing strategy.

## 2. Which are the top 5 states with the most total sales?

```
# Grouping states by sales
state_sales = data.groupby('State')['Sales'].sum()

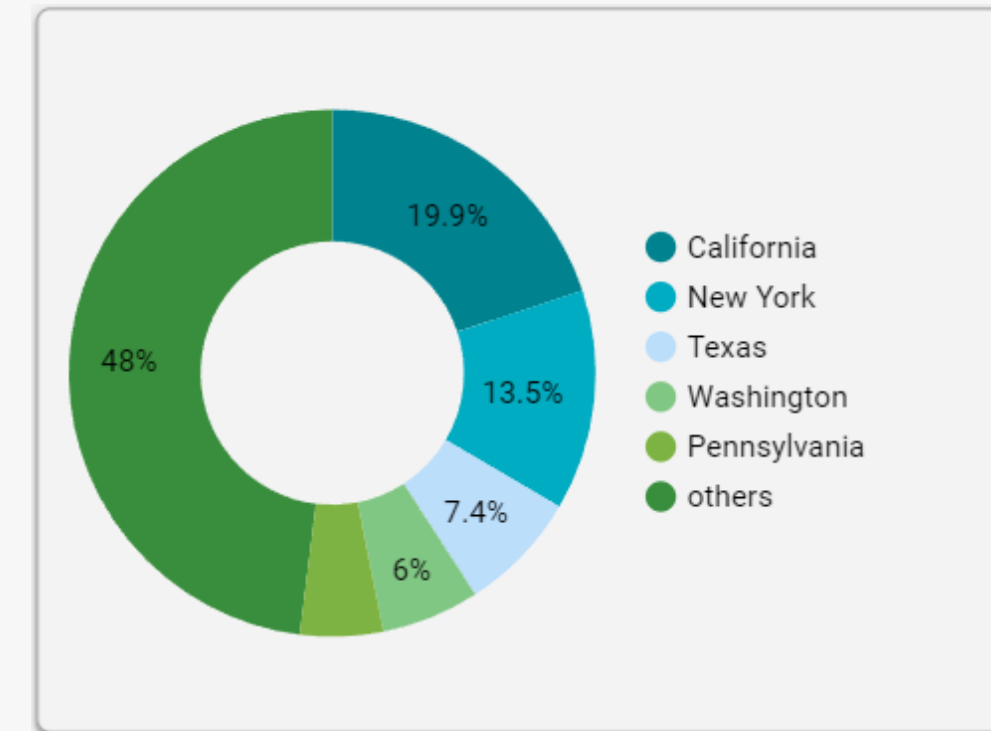
# Count states by sales in percentage
state_sales_percentage = ((state_sales / state_sales.sum()) * 100).apply(lambda x: f'{x:.1f}')

state_sales = pd.merge(state_sales, state_sales_percentage, on='State')
state_sales = state_sales.rename(columns={
    'Sales_x': 'Total_Sales',
    'Sales_y': 'Percentage_Sales'
})

# Top 5 state by sales
top_5_state = state_sales.nlargest(5, 'Total_Sales')

print("Top 5 Highest State by Sales:")
print(top_5_state)
```

```
Top 5 Highest State by Sales:
      State  Total_Sales  Percentage_Sales
California  457687.6315         19.9
New York    310876.2710         13.5
Texas        170188.0458          7.4
Washington  138641.2700          6.0
Pennsylvania 116511.9140          5.1
```



19.9% the highest total sales came from California, followed by New York, Texas, Washington, and Pennsylvania.



### 3. Based on region, what product category has the highest profit and losses?

```
# Grouping profit by region
region_profit = pd.pivot_table(
    data=data,
    index= 'Region',
    values= 'Profit',
    aggfunc= 'sum'
).reset_index()

print("Total Profit by Region:")
print(region_profit)
```

```
Total Profit by Region:
   Region  Profit
0  Central 39706.3625
1    East 91522.7800
2   South 46749.4303
3    West 108418.4489
```

- From the four regions, the central region has the lowest total profit. Followed by the southern region, then the eastern region and the **highest total profits** are in the **western region**.

```
# Grouping profit by category
category_profit = pd.pivot_table(
    data=data,
    index= 'Category',
    values= 'Profit',
    aggfunc= 'sum'
).reset_index()

print("Total Profit by Category:")
print(category_profit)
```

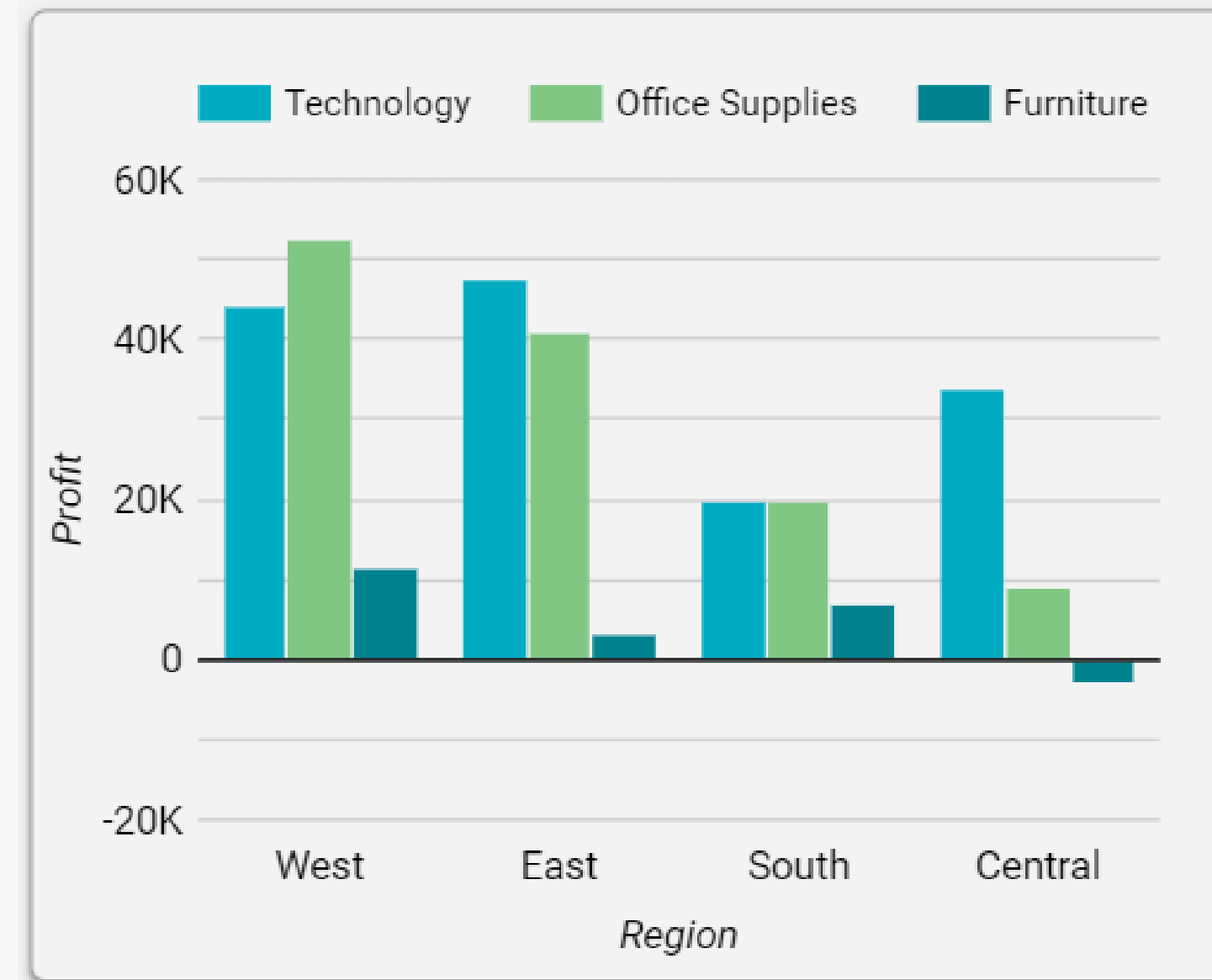
```
Total Profit by Category:
   Category  Profit
0   Furniture 18451.2728
1 Office Supplies 122490.8008
2   Technology 145454.9481
```

- The **highest profits** are dominated by **Technology**, and furniture is the category with the lowest profit.

```
# Grouping profit by region and category
regcat_profit = pd.pivot_table(
    data=data,
    index= ['Region', 'Category'],
    values= 'Profit',
    aggfunc= 'sum'
).reset_index()

print("Total Profit by Region and Category:")
print(regcat_profit)
```

```
Total Profit by Region and Category:
   Region Category   Profit
0  Central  Furniture -2871.0494
1  Central Office Supplies  8879.9799
2  Central  Technology  33697.4320
3    East  Furniture   3046.1658
4    East Office Supplies  41014.5791
5    East  Technology  47462.0351
6  South  Furniture   6771.2061
7  South Office Supplies  19986.3928
8  South  Technology  19991.8314
9    West  Furniture  11504.9503
10   West Office Supplies  52609.8490
11   West  Technology  44303.6496
```



- The furniture category provides the lowest profits across all regions, even in the **central region** the furniture category actually experiences **losses**.

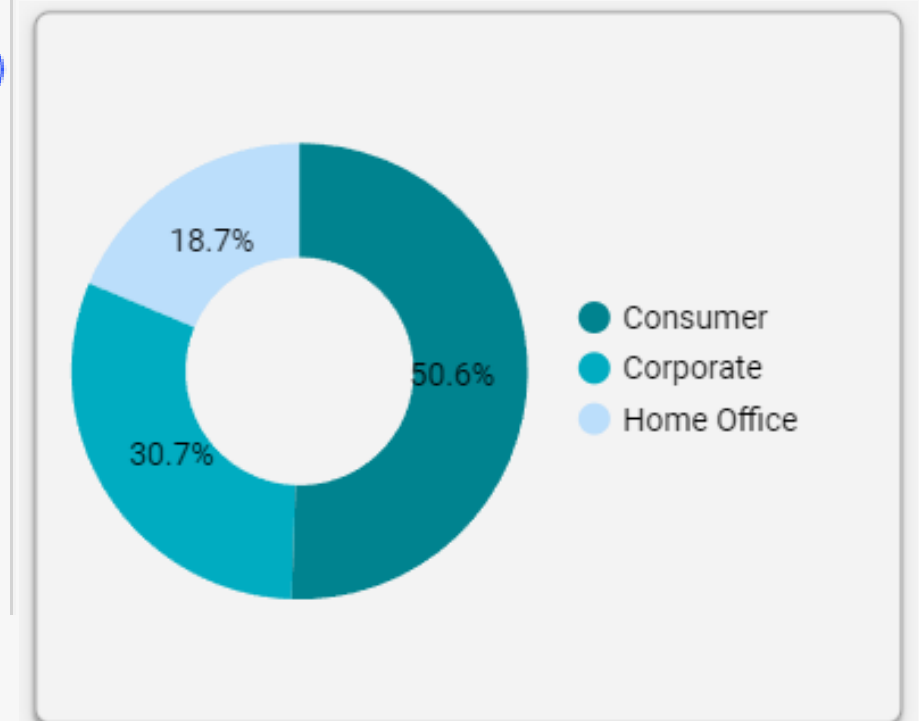
## 4. What segments are the main revenue contributors in sales?

```
# Grouping states by sales
segment_sales = data.groupby('Segment')['Sales'].sum().astype(int)

# Count states by sales in percentage
segment_sales_percentage = ((segment_sales / segment_sales.sum()) * 100).apply(lambda x: f'{x:.1f}')

segment = pd.merge(segment_sales, segment_sales_percentage, on='Segment')
segment = segment.rename(columns={
    'Sales_x': 'Total_Sales',
    'Sales_y': 'Percentage_Sales'
})
print("Percentage of Sales by Segment:")
print(segment)
```

```
Percentage of Sales by Segment:
      Total_Sales  Percentage_Sales
Segment
Consumer      1161401             50.6
Corporate      706146             30.7
Home Office    429653             18.7
```



- The **primary revenue** contributor comes from the **Consumer** segment, accounting for 50.6% of total sales, followed by the Corporate segment at 30.7% and then the Home Office segment at 18.7%.

```
# Grouping sales by region and category
segmen_sub_sales = pd.pivot_table(
    data=data,
    index= ['Segment', 'Sub-Category', 'Category'],
    values= 'Quantity',
    aggfunc= 'sum'
).reset_index()

segmen_sub_sales = segmen_sub_sales[segmen_sub_sales['Segment'] == 'Consumer']
segmen_sub_sales.sort_values(by=['Quantity'], inplace=True, ascending=False)

print("Sub-Categories Purchased by Consumers:")
print(segmen_sub_sales)
```

```
Sub-Categories Purchased by Consumers:
   Segment Sub-Category      Category  Quantity
3  Consumer    Binders  Office Supplies    3015
12 Consumer     Paper  Office Supplies    2602
9   Consumer  Furnishings    Furniture    1834
13 Consumer     Phones    Technology    1685
2   Consumer      Art  Office Supplies    1625
14 Consumer     Storage  Office Supplies    1619
```

- Purchases by **Consumers** are dominated by the **Office Supply** category, with the largest purchase being **3015 units of Binders** products.

## 5. Which shipping mode do customers use the most?"

```
# Grouping date_diff and order_id by ship_mode
avg = data.groupby('Ship_Mode')['Date_Diff'].mean().apply(lambda x: f'{x:.0f}')
order = data.groupby('Ship_Mode')['Order_ID'].nunique()

ship = pd.merge(avg, order, on='Ship_Mode')
ship = ship.rename(columns={
    'Date_Diff': 'Average Days',
    'Order_ID': 'Total Order'
})

ship.sort_values(by=['Average Days', 'Total Order'], inplace=True, ascending=[False, False])

print(ship)
```

Ship_Mode	Average Days	Total Order
Standard Class	5	2994
Second Class	3	964
First Class	2	787
Same Day	0	264

- The longest delivery time is Standard Class, while the fastest delivery is Same Day (the same day as the purchase).
- Even though **Standard Class** has the longest shipping time, it has **the largest number of users**. This may be influenced by cheaper shipping prices. Then followed by Second Class and First Class.
- Meanwhile, **Same Day** has the **fewest users**, even though the delivery time is the fastest. This may be influenced by higher shipping prices.

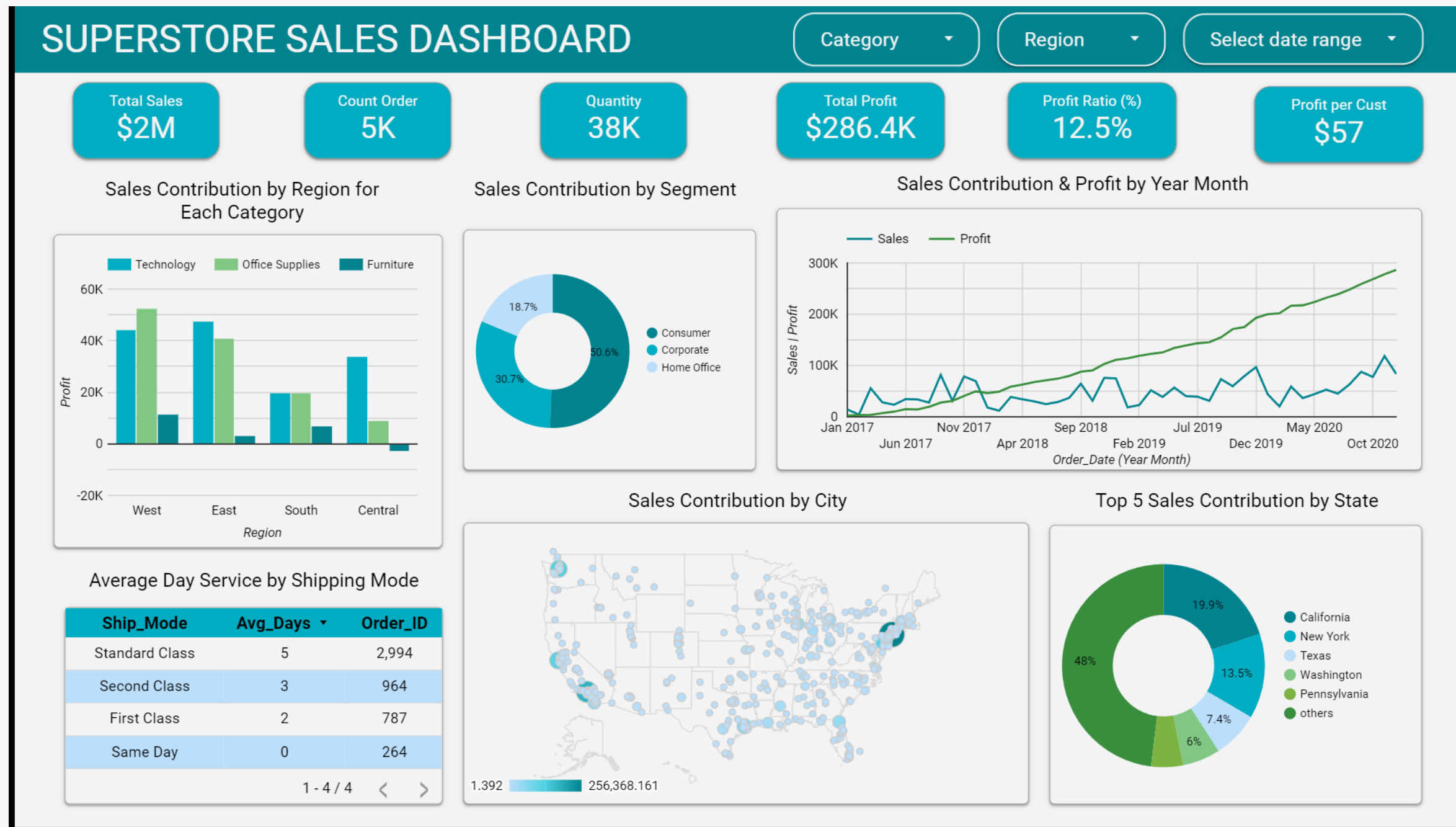
# Recommendation Summary



1. Based on **seasonal patterns**, prepare for sufficient stock availability during seasonal peaks. Consider price adjustments or discounts during seasonal periods to stimulate demand.
2. Evaluate marketing strategies and identify factors that may limit sales growth. Then adjust marketing and sales strategies based on **market characteristics and needs in each state**.
3. There is a need to evaluate and possibly **restructure the furniture category to increase profitability**. Some steps that can be taken include:
  - **Review product prices** to see whether they are in line with production costs and competitive in the market.
  - **Update furniture products** to better suit market needs and trends.
  - **Changing marketing strategies** to be more attractive to increase customer interest in furniture products



# Data Visualization



You can access the dashboard via this following link:

<https://lookerstudio.google.com/s/rcuEqIZvVjQ>

Linkedin:

<https://www.linkedin.com/in/anitamilaoktafani/>

Github:

<https://github.com/anitamila/Final-Project-Live-Class-DQLab>

*Thanks!*

