```
# show duplicate rows
df.duplicated()
```

	0
0	False
1	False
2	False
3	False

```
[ ]  # Convert relevant columns to appropriate data types
     df['Order Date'] = pd.to_datetime(df['Order Date'], format='%d/%m/%Y')
     df['Ship Date'] = pd.to_datetime(df['Ship Date'], format='%d/%m/%Y')
     df['Postal Code'] = df['Postal Code'].astype(object)


     ## Standaralize Column Names:
     # replace space in column headers with _
     df.columns = df.columns.str.replace(' ', '_')
     # replace - in column headers with _
     df.columns = df.columns.str.replace('-', '_')
     # Check Changes
     df.info()
```

```
# inspecting row with the NaN value
df[df.isna().any(axis=1)]
```

	Row_ID	Order_ID	Order_Date	Ship_Date	Ship_Mode	Cus
2234	2235	CA-2018-104066	2018-12-05	2018-12-10	Standard Class	

Findings Regarding Outliers

-Outliers is not Focused in a Certain Time or Region

-Furniture Category has largest number of Outliers While Office Supplies has the least

-Consumer Segment has largest number of Outliers While Home Office has the least

```
[ ]  #Export the Cleaned Data
     df.to_csv("Superstore Sales Dataset- Cleaned.csv", index=False)
```

```
[ ]  # Finding Outliers Using IQR (Interquartile Range) Method
     Q1 = df['Sales'].quantile(0.25)
     Q3 = df['Sales'].quantile(0.75)
     IQR = Q3 - Q1

     # Define outlier Bounds
     lower_bound = Q1 - 1.5 * IQR
     upper_bound = Q3 + 1.5 * IQR

     # Identify outliers
     iqr_outliers = df[(df['Sales'] < lower_bound) | (df['Sales'] > upper_bound)]
     print(f"Total transactions: {len(df)}")
     print(f"Outliers detected: {len(iqr_outliers)} ")
     print(f"Outliers Percentage: {len(iqr_outliers)/len(df):.1%}")
     print(f"Outlier Bounds: > ${upper_bound:.2f} or < ${lower_bound:.2f}")
```

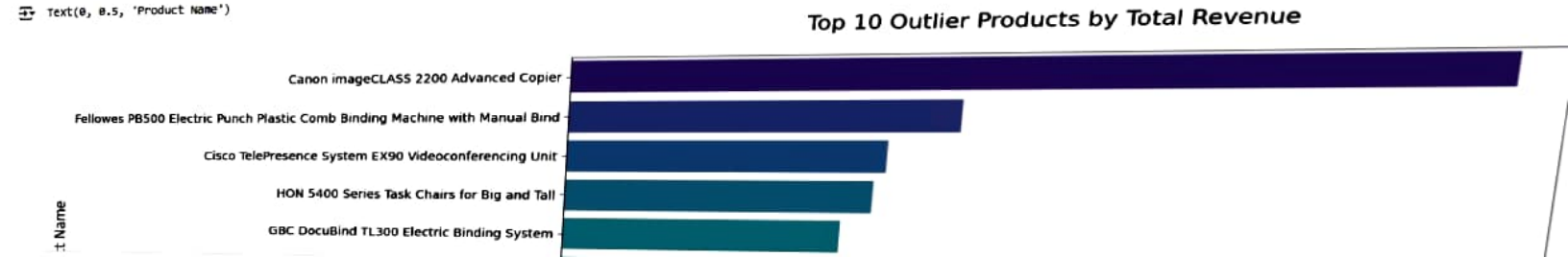
```
[ ] # Get top 10 outlier products by count

    # Visualize 10 outlier products by count
    top_outlier_products = iqr_outliers.groupby('Product_Name')['Sales'].sum()\
        .sort_values(ascending=False)\
        .head(10)

    # Create visualization
    plt.figure(figsize=(12, 6))
    sns.barplot(x=top_outlier_products.values, y=top_outlier_products.index, hue=top_outlier_products.index, palette="viridis", legend=False)

    # Add annotations and formatting
    plt.title('Top 10 Outlier Products by Total Revenue', fontsize=16, pad=20)
    plt.xlabel('Total Revenue ($)', fontsize=12)
    plt.ylabel('Product Name', fontsize=12)

    Text(0, 0.5, 'Product Name')
```



```
# Standardize 'Ship Mode' categories
df['Ship_Mode'] = df['Ship_Mode'].str.strip().str.title()

# Standardize 'Country' categories
df['Country'] = df['Country'].str.strip().str.title()
```

```
[ ]
# Count Each category of ship mode
df['Ship_Mode'].value_counts()
```

```
                count
Ship_Mode
Standard Class   5859
Second Class     1902
First Class      1501
Same Day          538

dtype: int64
```

Superstore Sales Dashboard

City

Segment

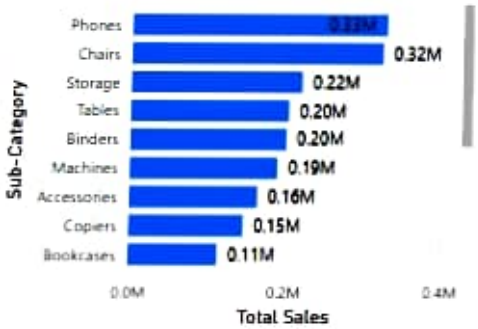
Homepage | Sales Analysis | Customers | Summary

Filters:- All All

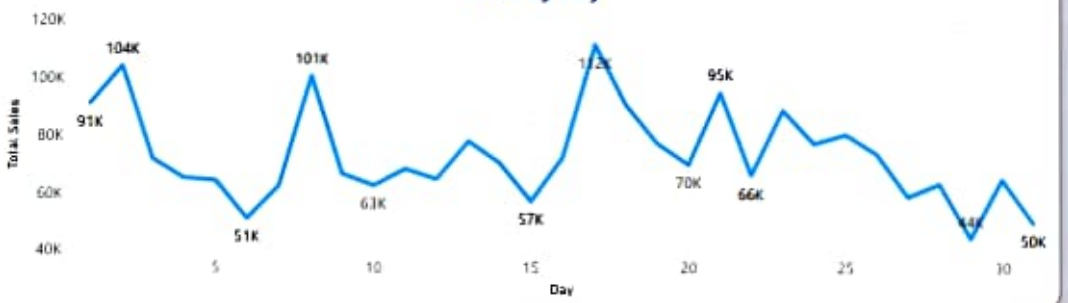
2015 | 2016 | 2017 | 2018

Jan | Feb | Mar | Apr | May | Jun

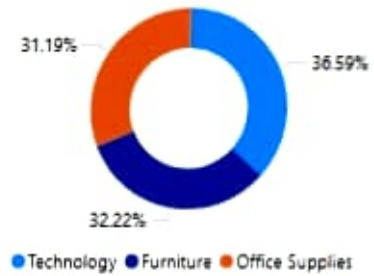
Sales by Sub-Category



Sales by Day

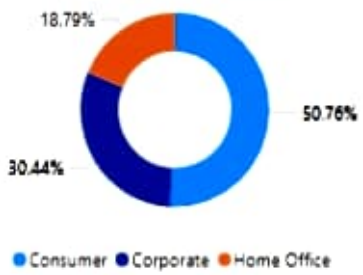


Sales by Category



Sales by State

Sales by Segment



Superstore Sales Dashboard

City

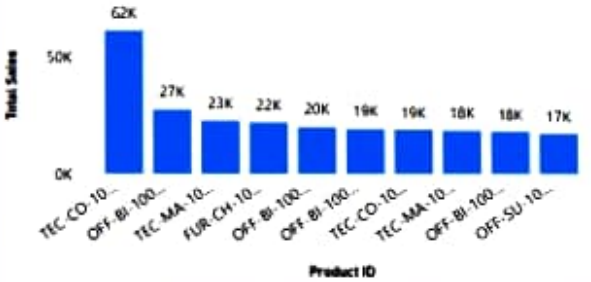
Segment

Filters:- All All

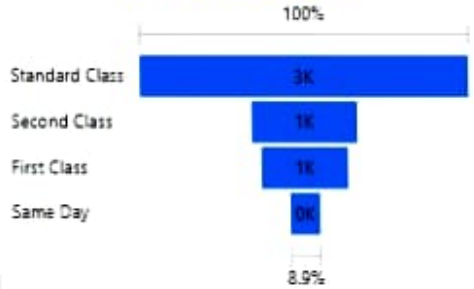
Homepage | Sales Analysis | Customers | Summary

2015 | 2016 | 2017 | 2018

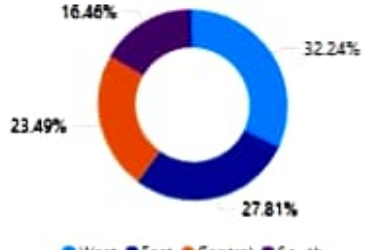
Top 10 Selling Products



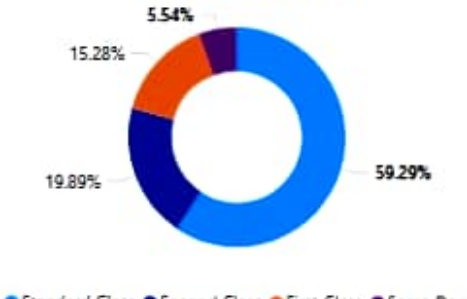
Orders by Ship Mode



Orders by Region



Sales by Ship Mode



Statistics Summary

Order ID	Customer Name	Region	State	Segment	Total Sales
CA-2015-100293	Neil Französisch	South	Florida	Home Office	91.06
CA-2015-100867	Eugene Hildebrand	West	California	Home Office	321.55
CA-2015-100881	Daniel Raglin	West	New Mexico	Home Office	302.38
CA-2015-100972	Dennis Bolton	West	Utah	Home Office	166.44
CA-2015-101364	Tamara Willingham	East	New York	Home Office	296.71
CA-2015-101392	Ann Steele	West	Washington	Home Office	269.36
CA-2015-101476	Shirley Daniels	East	New York	Home Office	69.99
CA-2015-101833	Frank Gastineau	East	New York	Home Office	34.44
CA-2015-103219	Benjamin Farhat	East	New York	Home Office	56.52
CA-2015-103744	Michael Grace	Central	Texas	Home Office	11.36
Total					2,261,536.78

Superstore Sales Dashboard

Homepage | Sales Analysis | Customers | Summary

Filters:-

City: Akron

Segment: All

2015 | 2016 | 2017 | 2018

Aberdeen
Abilene
Akron
Albuquerque
Alexandria
Allen
Allentown
Altoona

Total Sales	Total Customers	Avg. Sales Per Order	Avg. Sales per Customer
2.72K	9	303	303

Welcome to the Superstore Sales Dashboard, your key resource for analyzing sales performance and customer trends. The dashboard features four pages:

Homepage: Displays essential KPIs such as Total Sales, Total Orders, and Average Sales metrics, along with a line chart of Total Sales by Month.

Sales Analysis Page: Provides insights into sales by Sub-Category, Category, State, and Segment.

Customers Page: Highlights customer metrics, including Total Customers by Ship Mode and the Top 10 Customers by Orders.

Summary Page: Offers a comprehensive overview with Top Selling Products, Orders by Ship Mode, and detailed statistics.

Homepage | Sales Analysis | Customers

Superstore Sales Dashboard

Homepage | Sales Analysis | Customers | Summary

Filters:-

City: All

Segment: All

2015 | 2016 | 2017 | 2018

Total Sales	Total Orders	Total Customers	Avg. Sales Per Order	Avg. Sales per Customer
459.44K	1019	567	451	810

Welcome to the Superstore Sales Dashboard, your key resource for analyzing sales performance and customer trends. The dashboard features four pages:

Homepage: Displays essential KPIs such as Total Sales, Total Orders, and Average Sales metrics, along with a line chart of Total Sales by Month.

Sales Analysis Page: Provides insights into sales by Sub-Category, Category, State, and Segment.

Customers Page: Highlights customer metrics, including Total Customers by Ship Mode and the Top 10 Customers by Orders.

Summary Page: Offers a comprehensive overview with Top Selling Products, Orders by Ship Mode, and detailed statistics.

Superstore Sales Dashboard

Homepage | Sales Analysis | Customers | Summary

Filters:-

City: All

Segment: All

2015 | 2016 | 2017 | 2018

Total Sales	Total Orders	Total Customers	Avg. Sales Per Order	Avg. Sales per Customer
2.26M	4922	793	459	2852

Welcome to the Superstore Sales Dashboard, your key resource for analyzing sales performance and customer trends. The dashboard features four pages:

Homepage: Displays essential KPIs such as Total Sales, Total Orders, and Average Sales metrics, along with a line chart of Total Sales by Month.

Sales Analysis Page: Provides insights into sales by Sub-Category, Category, State, and Segment.

Customers Page: Highlights customer metrics, including Total Customers by Ship Mode and the Top 10 Customers by Orders.

Summary Page: Offers a comprehensive overview with Top Selling Products, Orders by Ship Mode, and detailed statistics.

Superstore Sales Dashboard
City
Filters:-
All
Homepage
Segment
Home Office
Consumer
Corporate
Home Office
Sales Analysis
Customers
Summary
2015
2016
2017
2018
Total Sales
424.98K
Total Orders
894
Avg. Sales Per Order
475
Avg. Sales per Customer
2872
by Month
Total Sales
50K
0K
17K
11K
57K
33K
21K
29K
19K
19K
48K
55K
62K
55K
January
February
March
April
May
June
July
August
September
October
November
December
Month
Welcome to the Superstore Sales Dashboard, your key resource for analyzing sales performance and customer trends. The dashboard features four pages:
Homepage: Displays essential KPIs such as Total Sales, Total Orders, and Average Sales metrics, along with a line chart of Total Sales by Month.
Sales Analysis Page: Provides insights into sales by Sub-Category, Category, State, and Segment.
Customers Page: Highlights customer metrics, including Total Customers by Ship Mode and the Top 10 Customers by Orders.
Summary Page: Offers a comprehensive overview with Top Selling Products, Orders by Ship Mode, and detailed statistics.

Customers by Ship Mode

Top 10 Customers by Orders

Customers by Segment

Customers by State

Customer Name	Customer ID	Total Sales	Total Orders
Alex Avila	AA-10315	5,563.56	5
Allen Armold	AA-10375	1,056.39	9
Andrew Allen	AA-10480	1,790.51	4
Anna Andreadi	AA-10645	5,086.94	6
Aaron Bergman	AB-10015	886.16	3
Adam Bellavance	AB-10060	7,755.62	8
Adnan Barton	AB-10105	14,473.57	10
Aimee Bixby	AB-10150	966.71	5
Alan Barnes	AB-10165	1,113.84	8
Alejandro Ballentine	AB-10255	914.53	9
Ann Blume	AB-10600	1,515.86	4
Alyssa Crouse	AC-10420	925.80	3
Total		**2,261,536.78**	**4922**