

Store Sales Dataset Analysis Project

A comprehensive project analyzing store sales data to identify trends, forecast future sales, and provide actionable insights to decision-makers.

Submitted to DEPI



Prepared By

Asmaa Mahgoub

Hamdy Awad

Saad El-Mohamady

Sarah El-Alfy

Seif El-Nakieb

Project Framework



Data Preprocessing

Build data model and clean dataset (Week 1)



Analysis Questions

Define key analytical questions for business decisions (Week 2)



Forecasting Questions

Generate predictive insights with visualization (Week 3)



Visualization & Reporting

Build Power BI dashboard and prepare final presentation (Week 4)

PROJECT PROJECT TIMELINE



DATA PREPROCESSING

Focus on data cleaning and preparing the dataset for analysis.



DATA VISUALIZATION

Develop visualizations to present findings.



PHASE 03:

Focus on generating predictive insights using machine learning models.

Finalize the dashboard and prepare the presentation.



Data Preprocessing Strategy

Data Model Design

Organize into fact and dimension tables. Define relationships between sales, products, and regions.

Cleaning Strategy

Standardize formats. Treat outliers using statistical methods. Validate data integrity.

Tool Implementation

Use Python (pandas) for manipulation. Apply Matplotlib & Seaborn for visualization.

Analysis Questions

Product Analysis

- Highest sales categories
- Sub-category performance
- Top 10 products by revenue

Geographic Distribution

- Top regions and states
- City-level performance

Customer Insights

- Segment profitability
- Highest value customers
- Average order value



Forecasting Questions



Sales Trends

Forecast 6-month sales by product category

Seasonal Patterns

Identify predictable peak sales periods

Regional Growth

Project highest growth regions

Segment Growth

Predict customer segment distribution changes

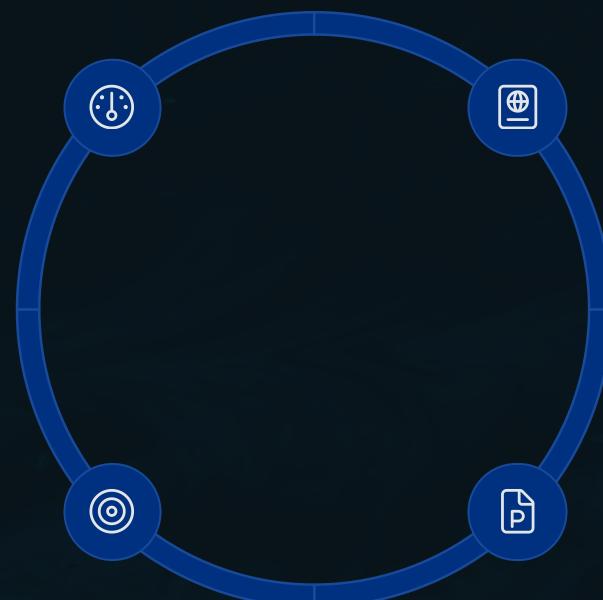
Visualization & Reporting

Interactive Dashboard

User-friendly Power BI interface
with drill-down capabilities

Action Strategies

Business implications and
recommended actions



Final Report

Comprehensive documentation of
methodology and findings

Stakeholder Presentation

Concise, visual slideshow
highlighting key insights

Data Cleaning Process

```
ing row with the NaN value  
a().any(axis=1)]
```

ID	Order_ID	Order_Date	Ship_Date
235	CA-2018-104066	2018-12-05	2018-12-10
275	CA-2016-162887	2016-11-07	2016-11-09

```
ant columns to appropriate data type  
] = pd.to_datetime(df['Order Date'],  
= pd.to_datetime(df['Ship Date'], f  
] = df['Postal Code'].astype(object)
```

```
Column Names:  
in column headers with _  
.columns.str.replace(' ', '_')  
column headers with _  
.columns.str.replace('-', '_')
```

ndas as pd
tplotlib.p
aborn as s

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

0	False
1	False
2	False
3	False

```
# 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
```

stigating Outliers

```
ding Outliers Using IQR (Interquartile Range) Method  
df['Sales'].quantile(0.25)  
df['Sales'].quantile(0.75)
```

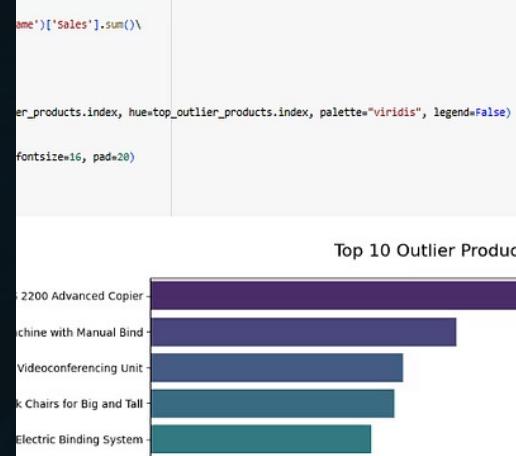
- Q3 - Q1

```
ine outlier Bounds  
_bound = Q1 - 1.5 * IQR  
_bound = Q3 + 1.5 * IQR
```

entify outliers

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

```
transactions: 9800  
ers detected: 1145  
ers Percentage: 11.7%  
er Bounds: > $500.64 or < -$272.79
```



arding Outliers

n a Certain Time or Region

argest number of Outliers While Offic

largest number of Outliers While Hor

ed Data

core Sales Dataset- Cleaned.csv", i

The team implemented rigorous data cleaning protocols to handle missing values, standardize formats, and remove duplicates, ensuring high-quality analysis.

Dashboard Deliverables

Superstore Sales Dashboard

Filters: City: All, Segment: All, Year: 2015, 2016, 2017, 2018

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

Total Sales by 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.

Homepage | Sales Analysis | Customers | Summary | +

Superstore Sales Dashboard

Filters: City: Akron, Segment: All, Year: 2015, 2016, 2017, 2018

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

Total Sales by 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.

Homepage | Sales Analysis | Customers | Summary | +

Main Form

Main Form

Applying Filters

Dashboard Navigation Tabs



Summary



Customer Analysis



Sales Analysis

Thank You