ANALYZING AMAZON SALES DATA

Project Report By: Aadithya Ram

Project Details

Project Title	Analyzing Amazon Sales data
Technologies	Data Science
Domain	E-commerce
Project Difficulties level	Advanced

Objectives and Problem Statement

Objectives

01

ETL

Extract – Transform - Load

02

Sales Trend

Month-Wise, Year-Wise, Yearly-Month Wise

03

Analysis

Key Metrics, Factors & Attributes

Problem Statement

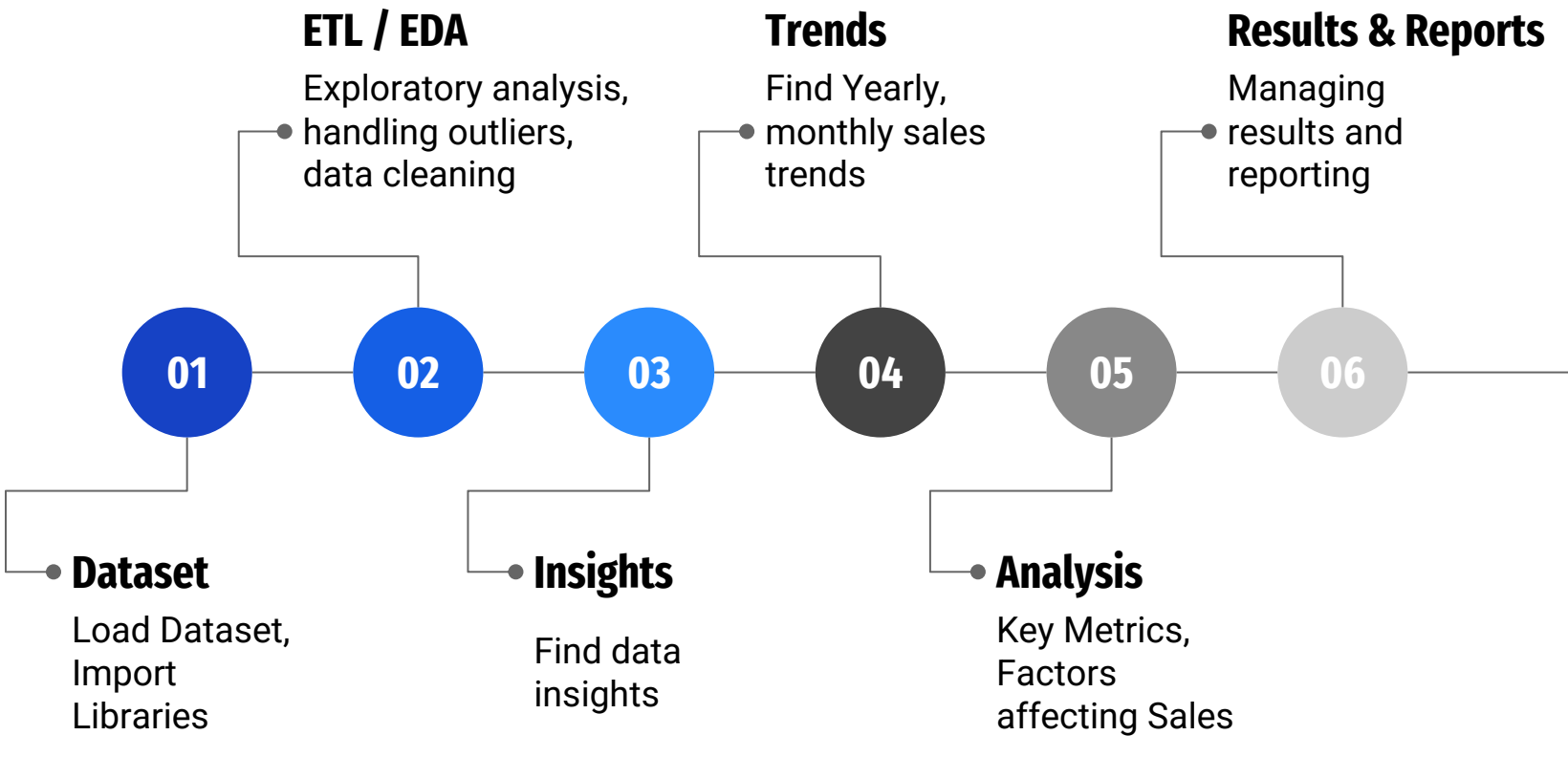
Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profits. Sales management today is the most important function in a commercial and business enterprise.

Do ETL: Extract-Transform-Load some Amazon dataset and find for me
Sales-trend -> month-wise, year-wise, yearly_month-wise

Find key metrics and factors and show the meaningful relationships between

attributes. Do your own research and come up with your findings.

Project Architecture



Dataset

[166]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

#Amazon Sales Dataset
dataset_path = '/Users/aadithyaram/Desktop/amznsalesdata.csv'
sales_data = pd.read_csv(dataset_path)
sales_data.head()
```

[166]:

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010	669165933	6/27/2010	9925	255.28	159.42	2533654.00	1582243.50	951410.50
1	Central America and the Caribbean	Grenada	Cereal	Online	C	8/22/2012	963881480	9/15/2012	2804	205.70	117.11	576782.80	328376.44	248406.36
2	Europe	Russia	Office Supplies	Offline	L	5/2/2014	341417157	5/8/2014	1779	651.21	524.96	1158502.59	933903.84	224598.75
3	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	6/20/2014	514321792	7/5/2014	8102	9.33	6.92	75591.66	56065.84	19525.82
4	Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	2/1/2013	115456712	2/6/2013	5062	651.21	524.96	3296425.02	2657347.52	639077.50

ETL

```
[167]: sales_data.shape
```

```
[167]: (100, 14)
```

```
[168]: sales_data.columns
```

```
[168]: Index(['Region', 'Country', 'Item Type', 'Sales Channel', 'Order Priority',
       'Order Date', 'Order ID', 'Ship Date', 'Units Sold', 'Unit Price',
       'Unit Cost', 'Total Revenue', 'Total Cost', 'Total Profit'],
      dtype='object')
```

```
[169]: sales_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Region          100 non-null    object 
 1   Country         100 non-null    object 
 2   Item Type       100 non-null    object 
 3   Sales Channel   100 non-null    object 
 4   Order Priority  100 non-null    object 
 5   Order Date      100 non-null    object 
 6   Order ID        100 non-null    int64  
 7   Ship Date       100 non-null    object 
 8   Units Sold      100 non-null    int64  
 9   Unit Price      100 non-null    float64
 10  Unit Cost       100 non-null    float64
 11  Total Revenue   100 non-null    float64
 12  Total Cost      100 non-null    float64
 13  Total Profit    100 non-null    float64
dtypes: float64(5), int64(2), object(7)
memory usage: 11.1+ KB
```

```
[170]: sales_data.isnull().sum()
```

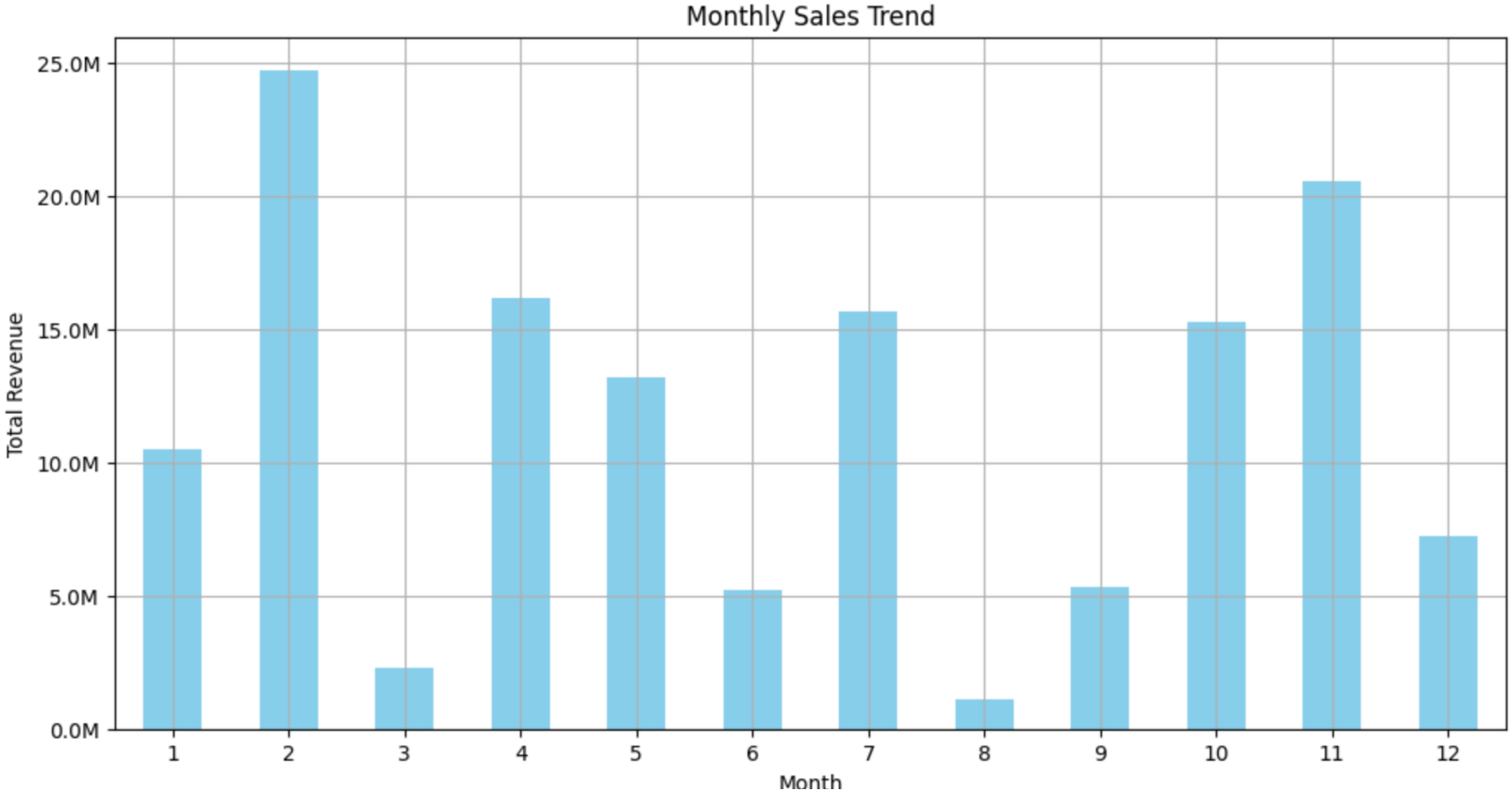
```
[170]: Region            0
       Country           0
       Item Type         0
       Sales Channel     0
       Order Priority    0
       Order Date        0
       Order ID          0
       Ship Date         0
       Units Sold        0
       Unit Price        0
       Unit Cost         0
       Total Revenue     0
       Total Cost        0
       Total Profit      0
       dtype: int64
```

```
[171]: sales_data.describe()
```

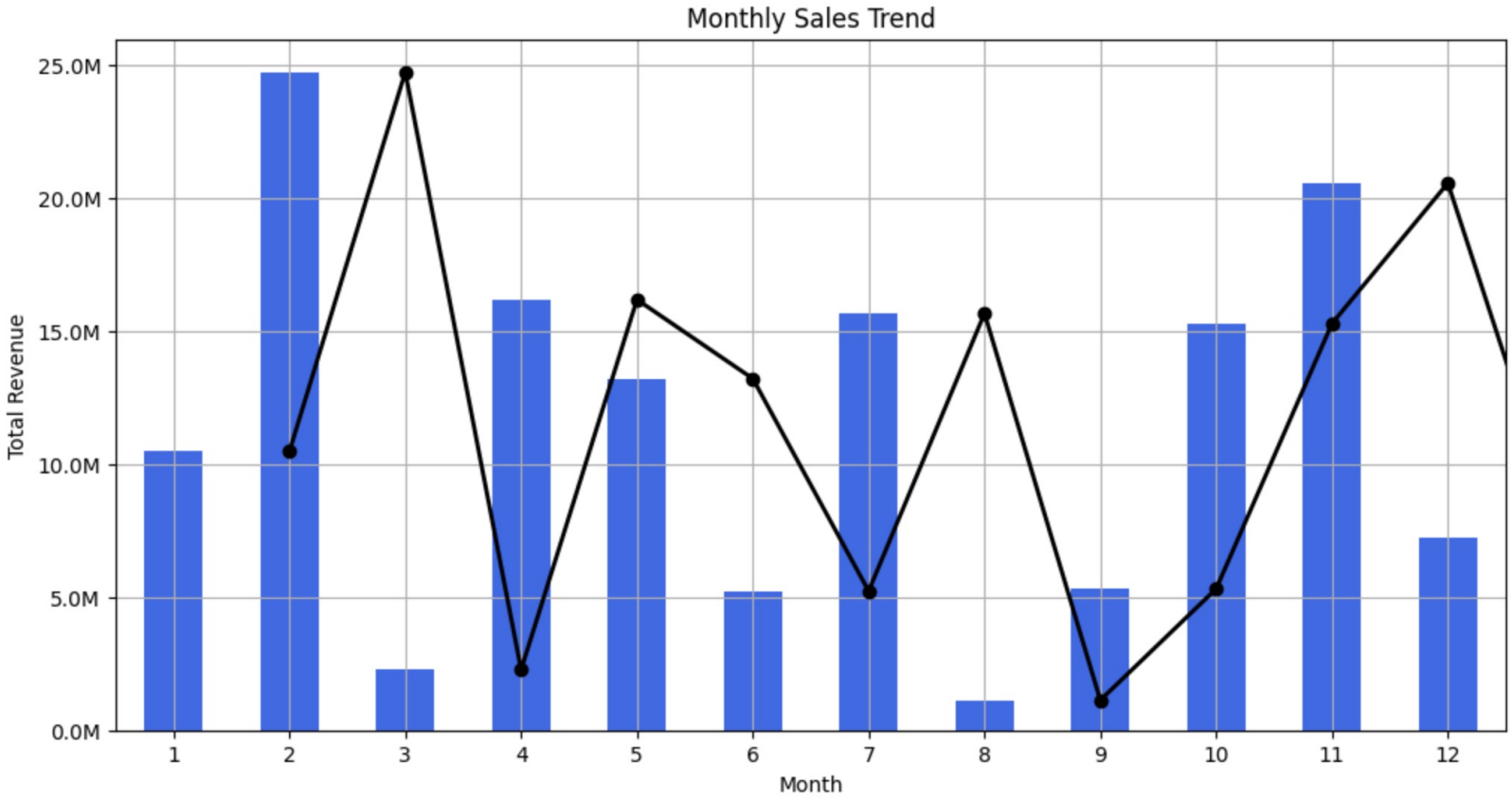
[171]:

	Order ID	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
count	1.000000e+02	100.000000	100.000000	100.000000	1.000000e+02	1.000000e+02	1.000000e+02
mean	5.550204e+08	5128.710000	276.761300	191.048000	1.373488e+06	9.318057e+05	4.416820e+05
std	2.606153e+08	2794.484562	235.592241	188.208181	1.460029e+06	1.083938e+06	4.385379e+05
min	1.146066e+08	124.000000	9.330000	6.920000	4.870260e+03	3.612240e+03	1.258020e+03
25%	3.389225e+08	2836.250000	81.730000	35.840000	2.687212e+05	1.688680e+05	1.214436e+05
50%	5.577086e+08	5382.500000	179.880000	107.275000	7.523144e+05	3.635664e+05	2.907680e+05
75%	7.907551e+08	7369.000000	437.200000	263.330000	2.212045e+06	1.613870e+06	6.358288e+05
max	9.940222e+08	9925.000000	668.270000	524.960000	5.997055e+06	4.509794e+06	1.719922e+06

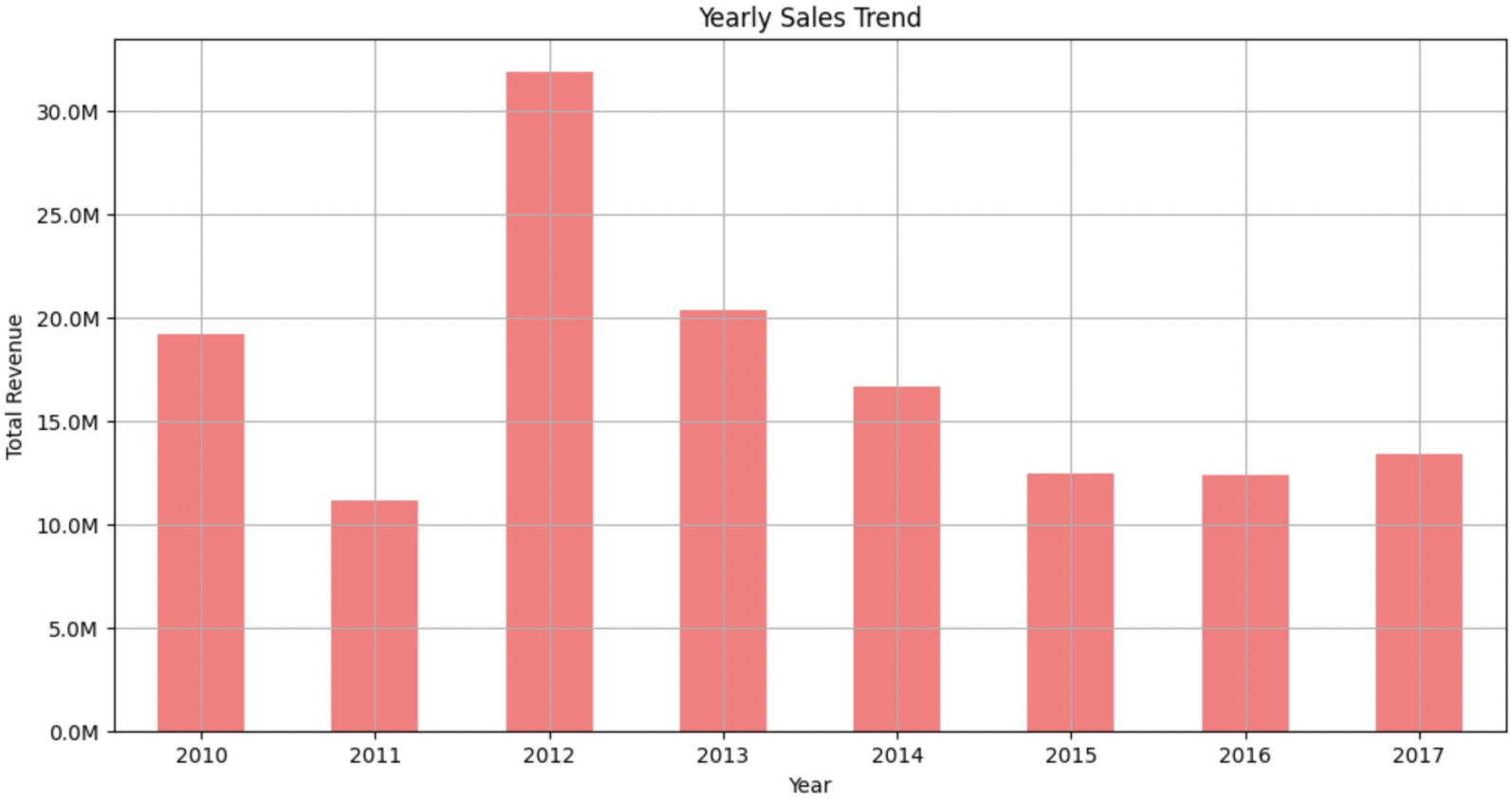
Monthly Sales Trend



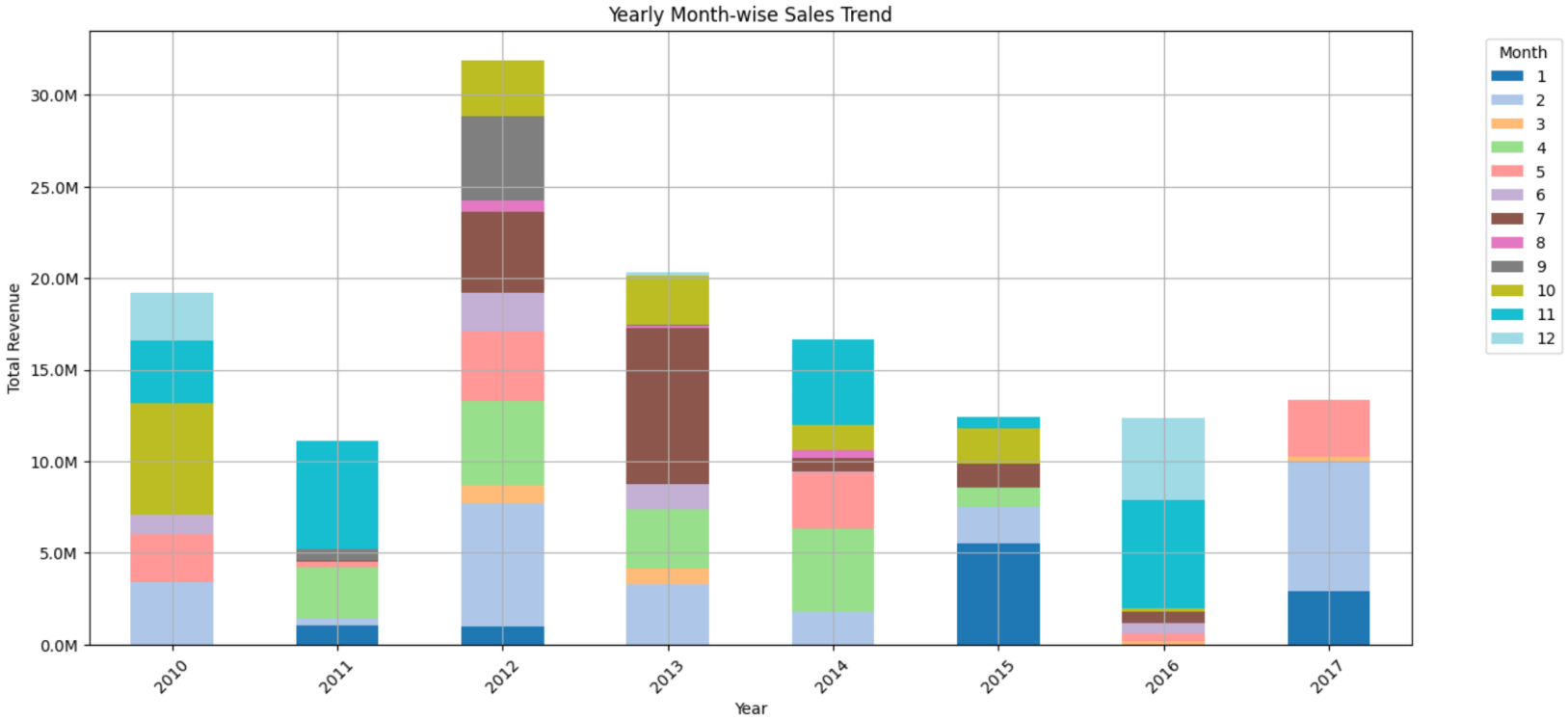
Monthly Trend Line



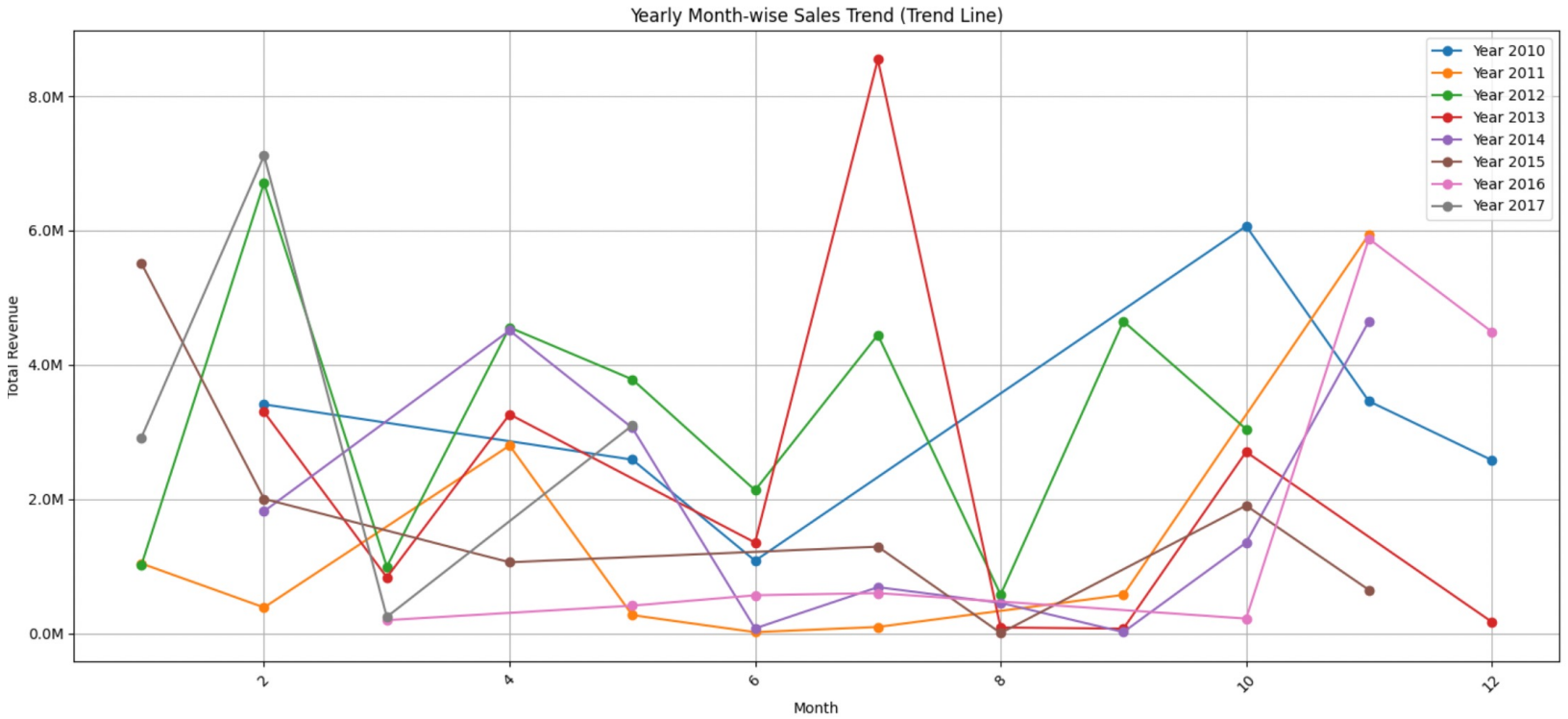
Yearly Sales Trend



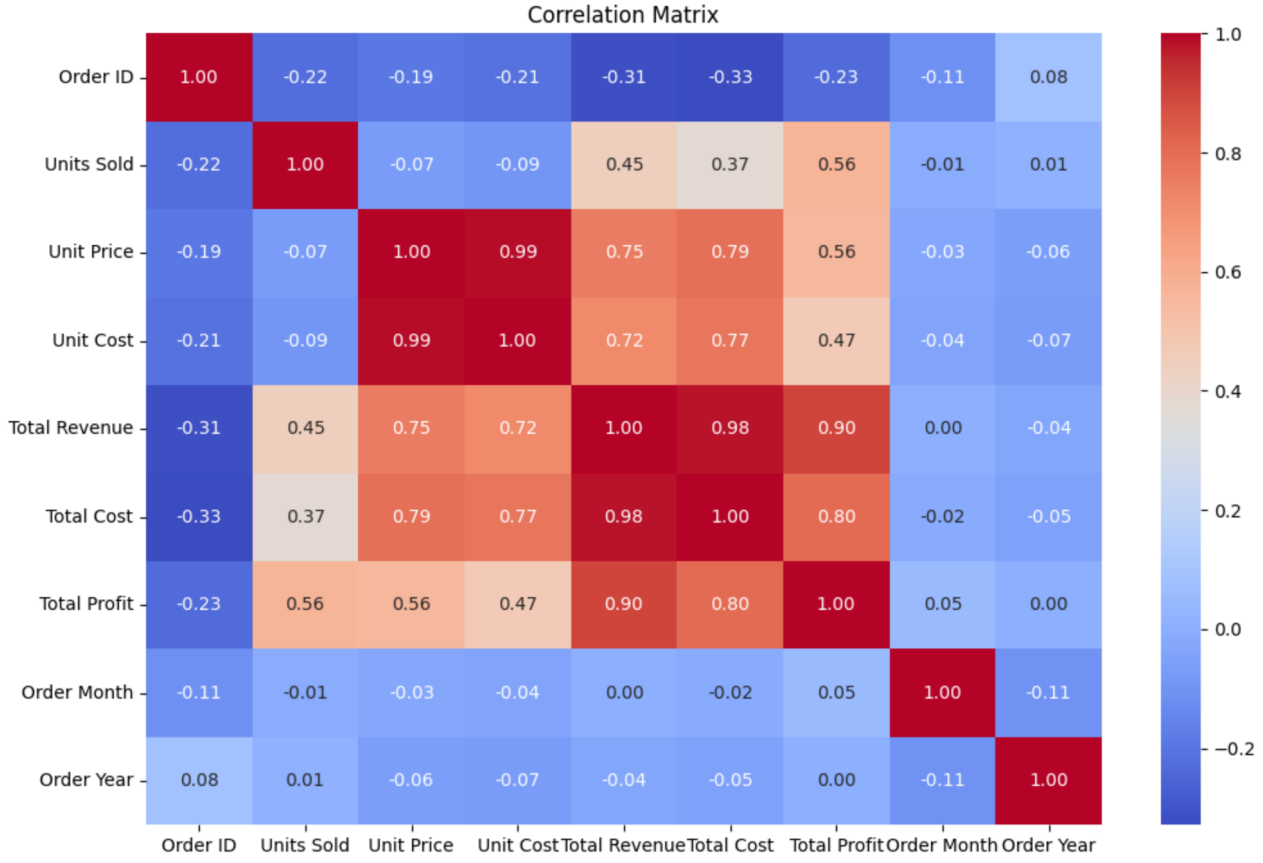
Yearly-Month Wise Trend



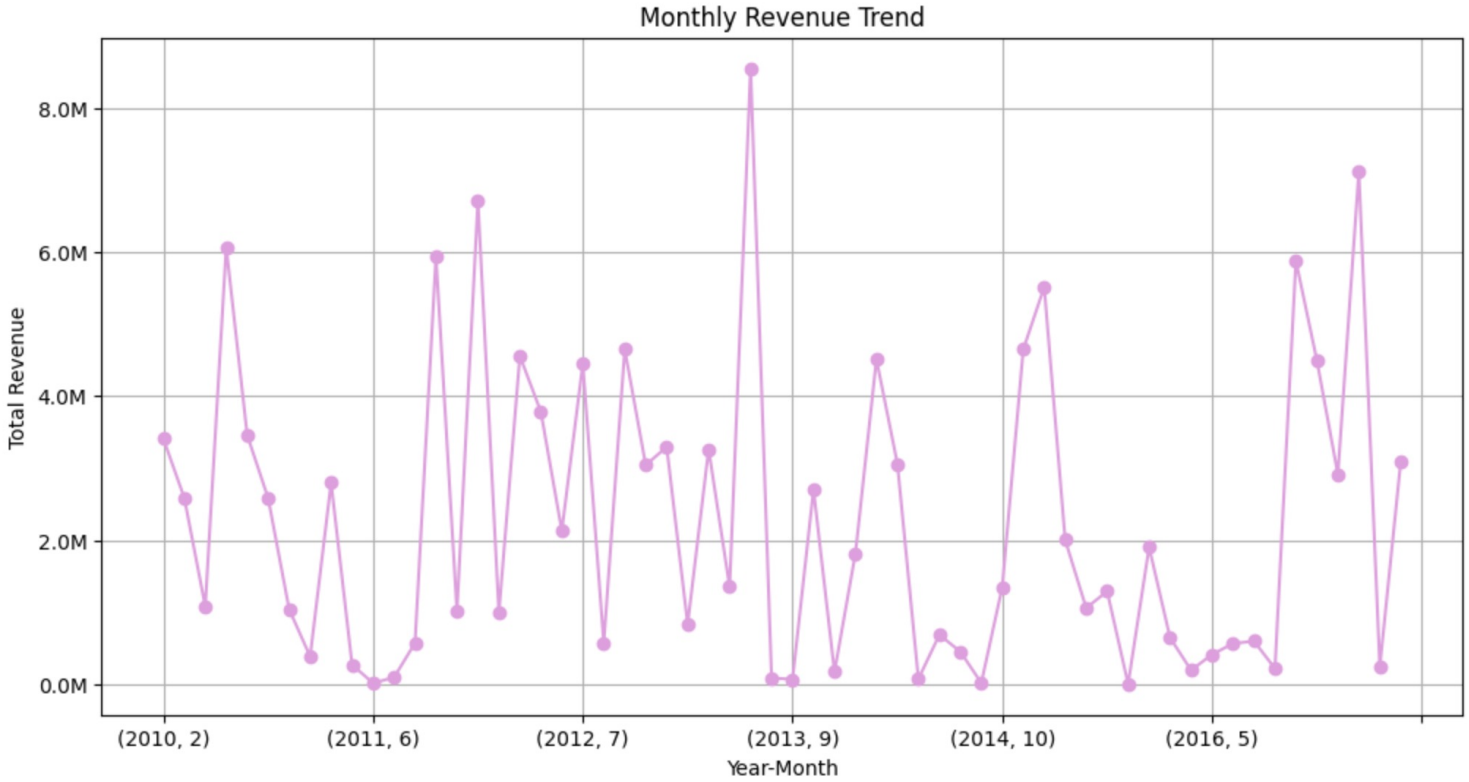
Yearly-Month Wise Trend



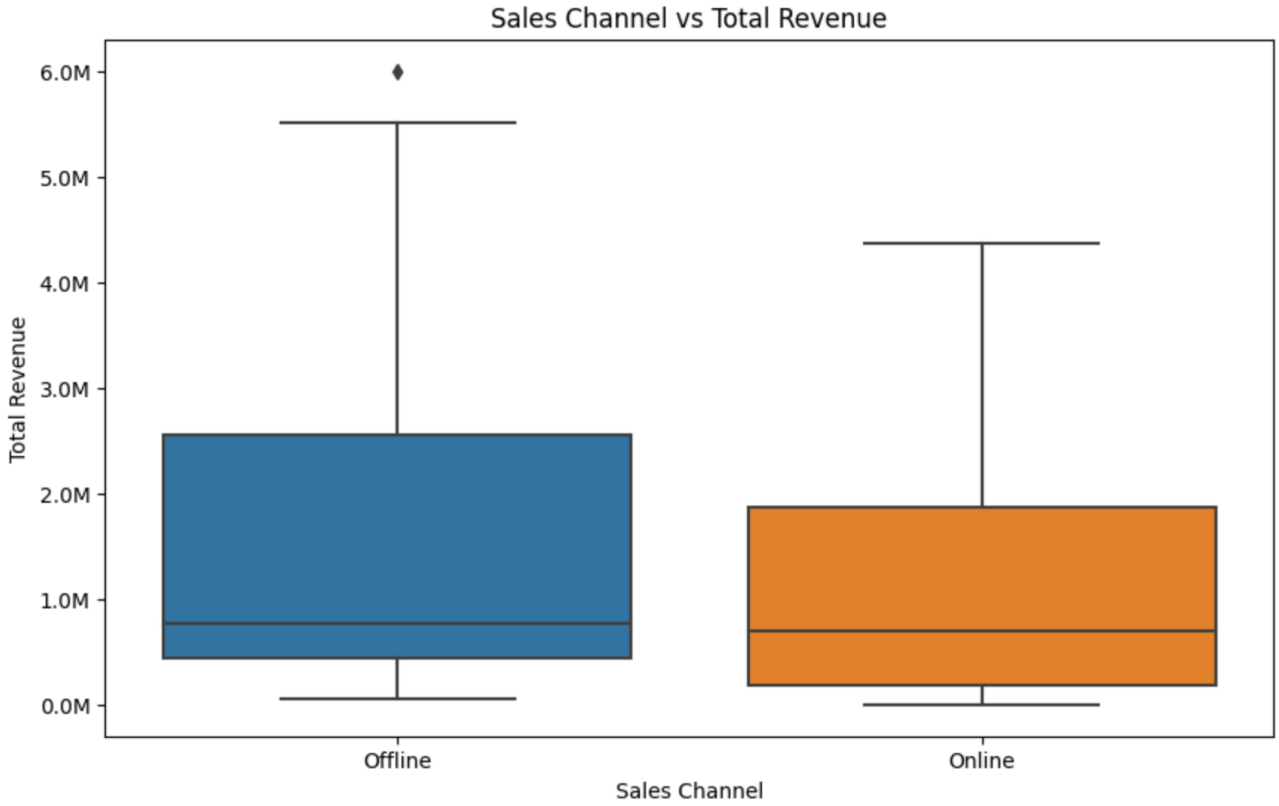
Correlations



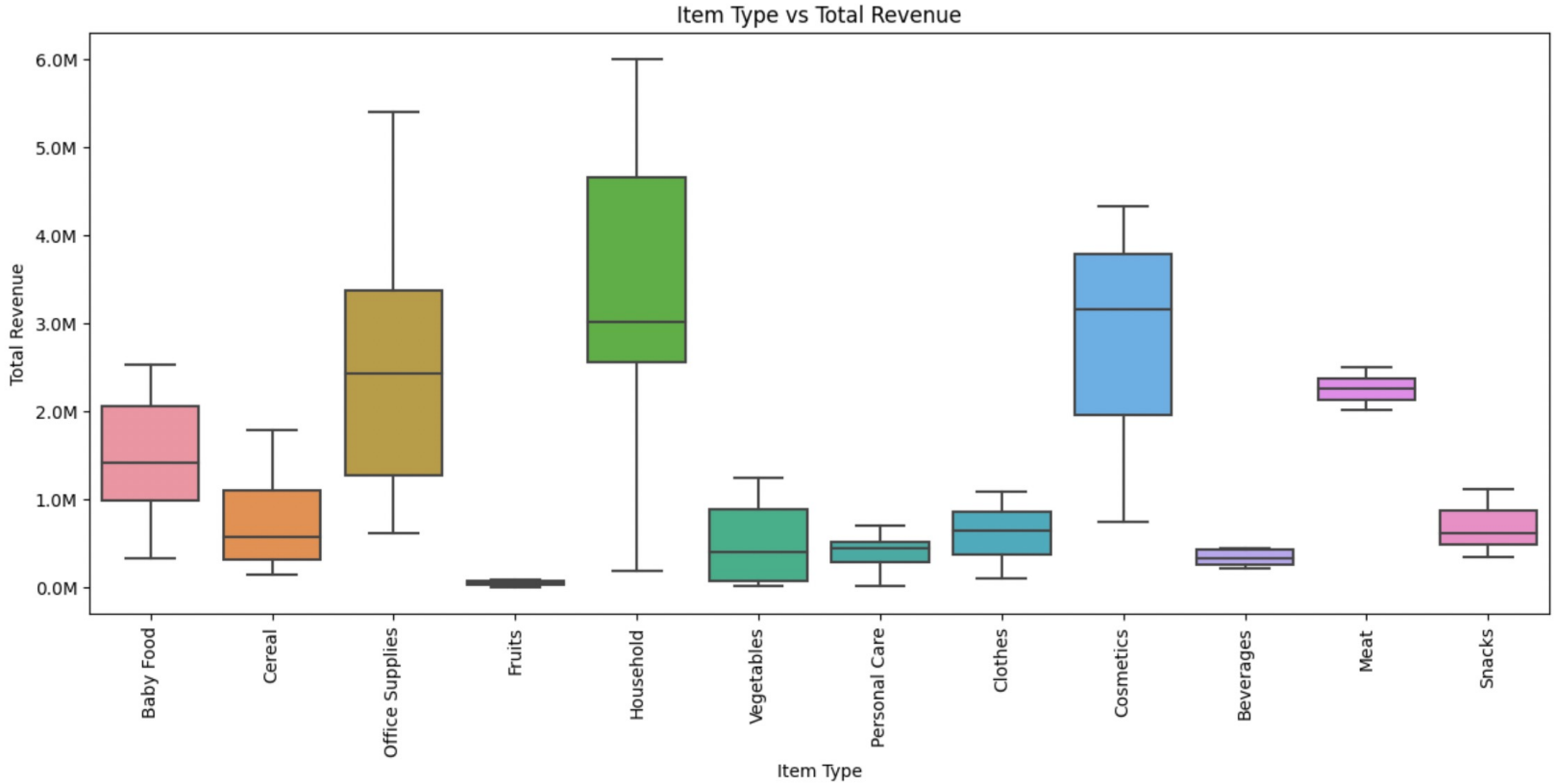
Revenue Trends



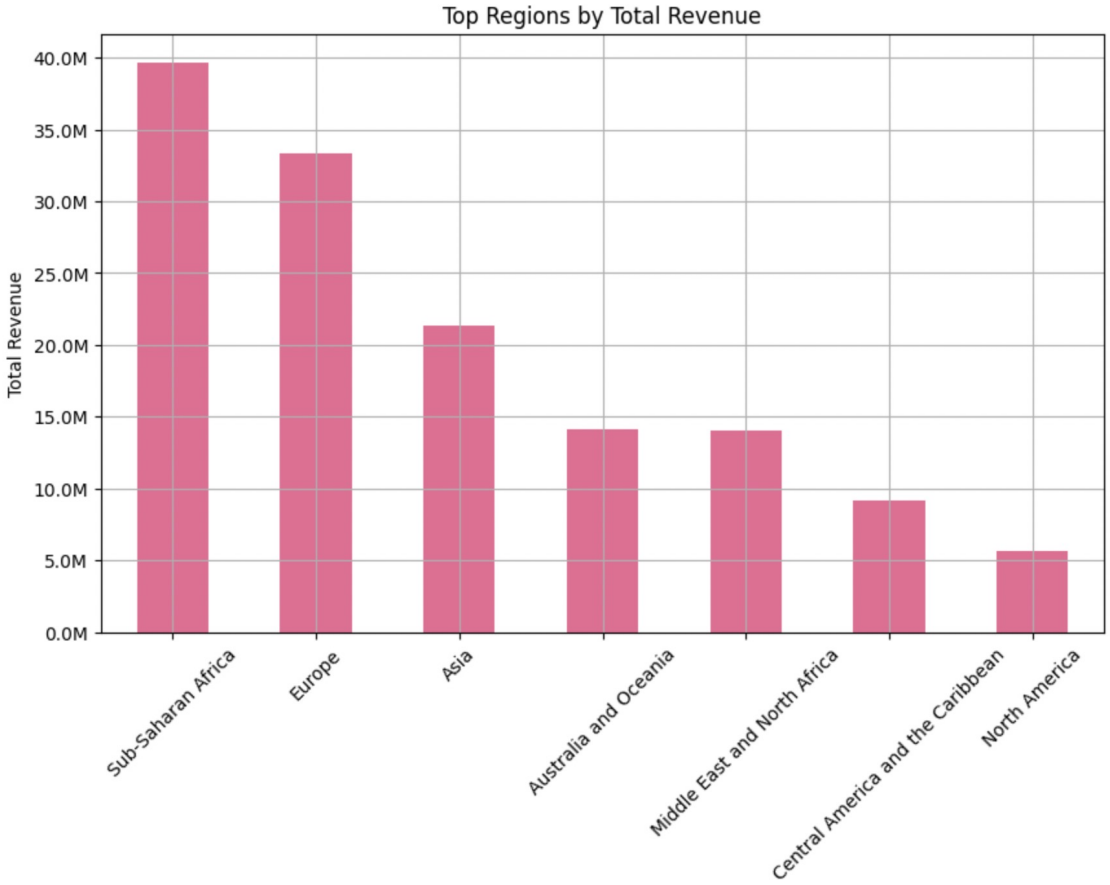
Revenue Trends



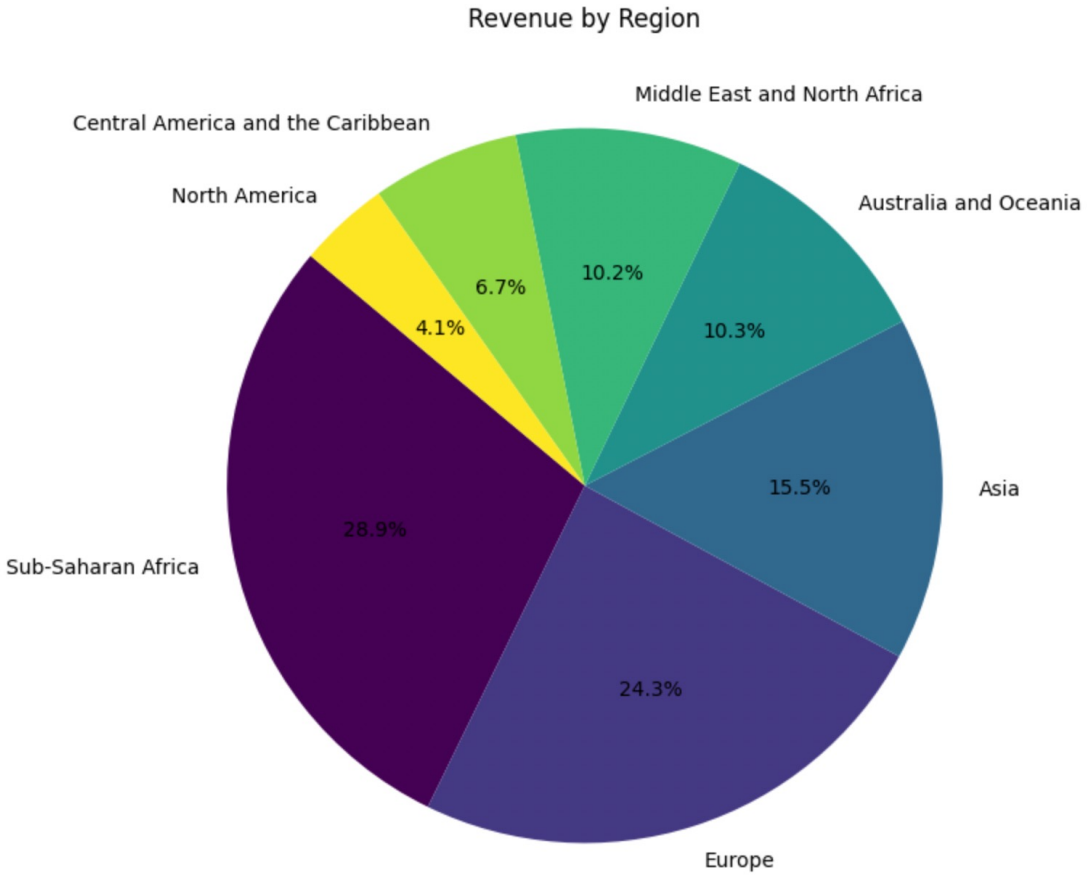
Revenue Trends



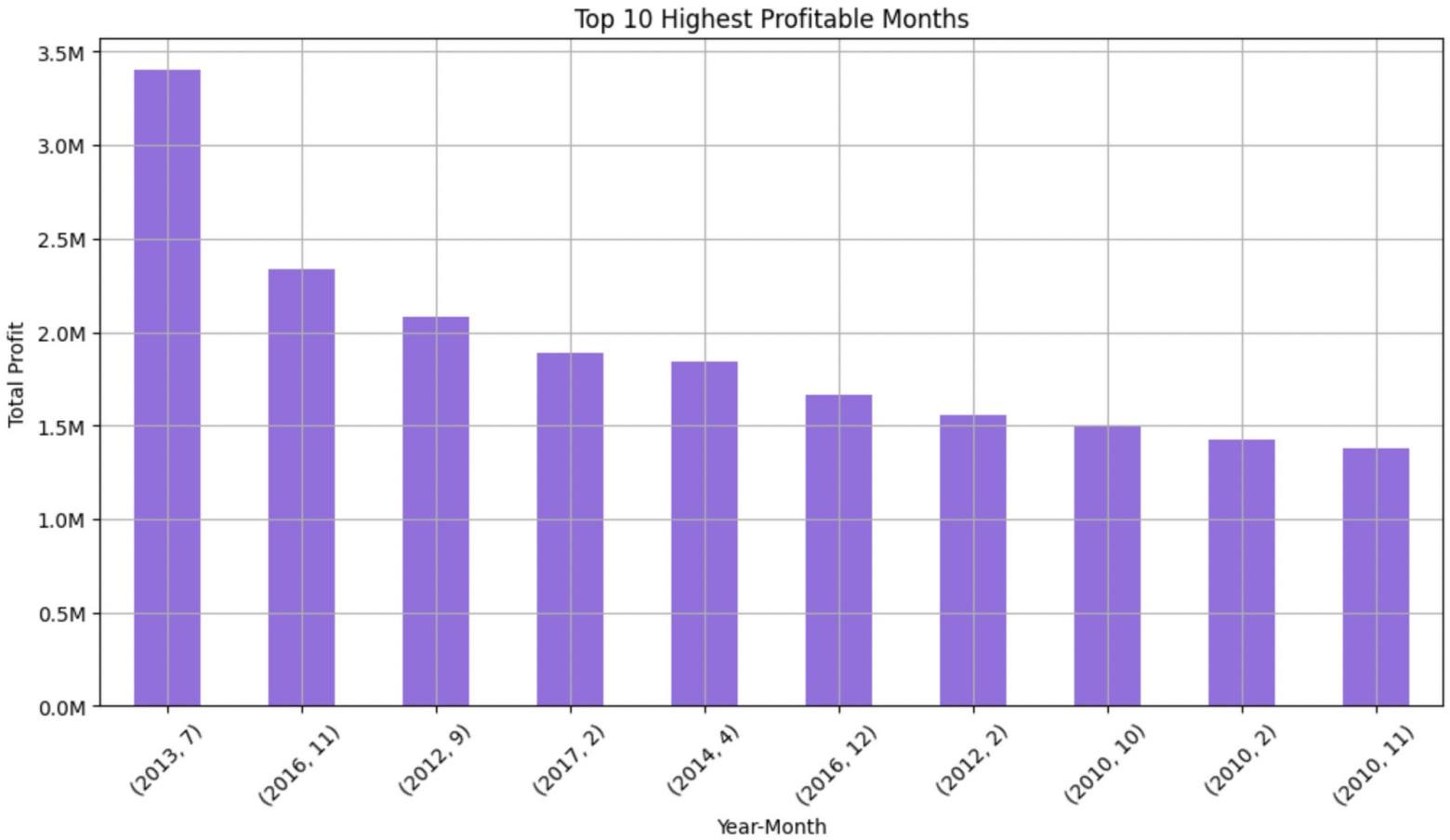
Revenue Trends



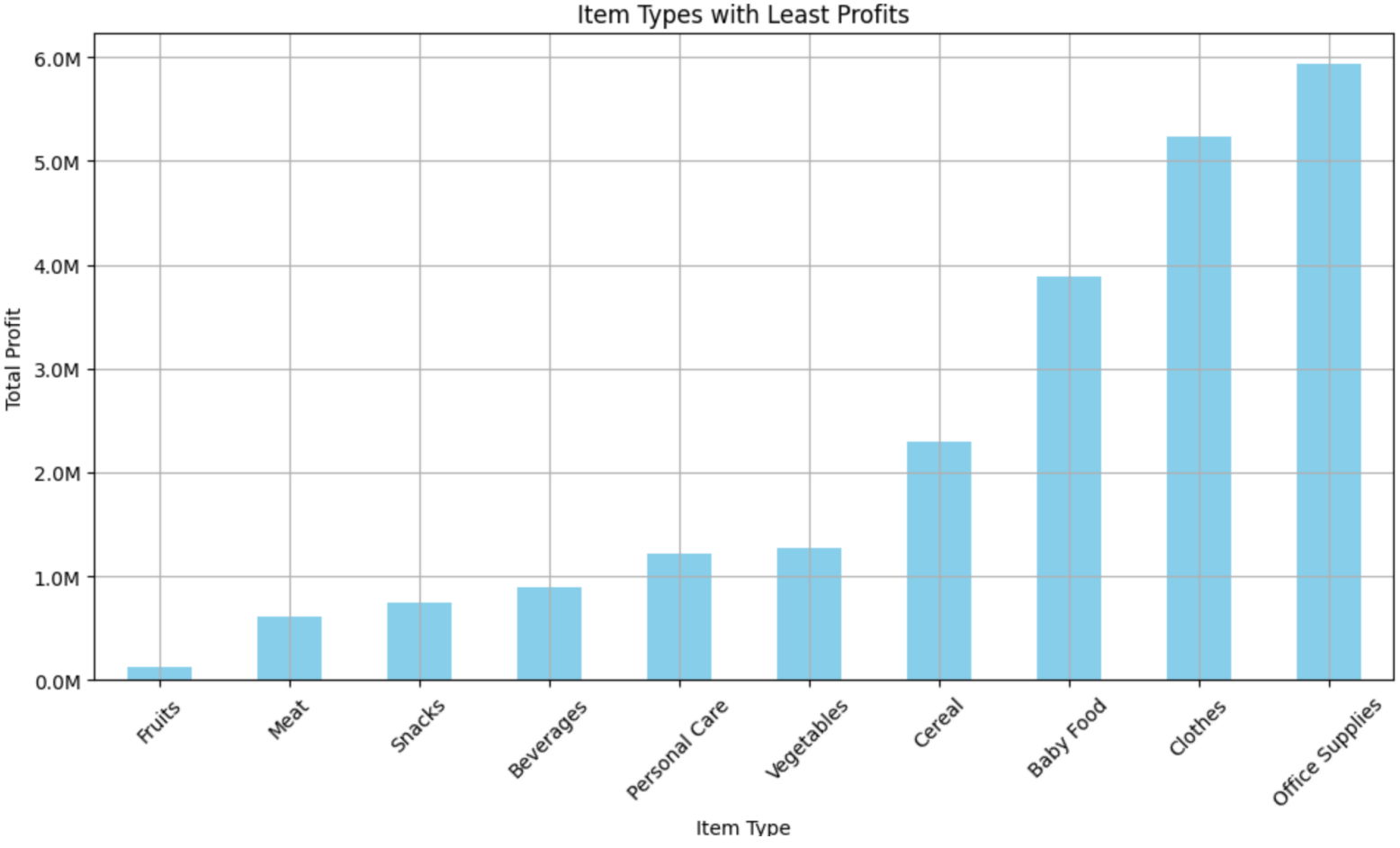
Revenue Trends



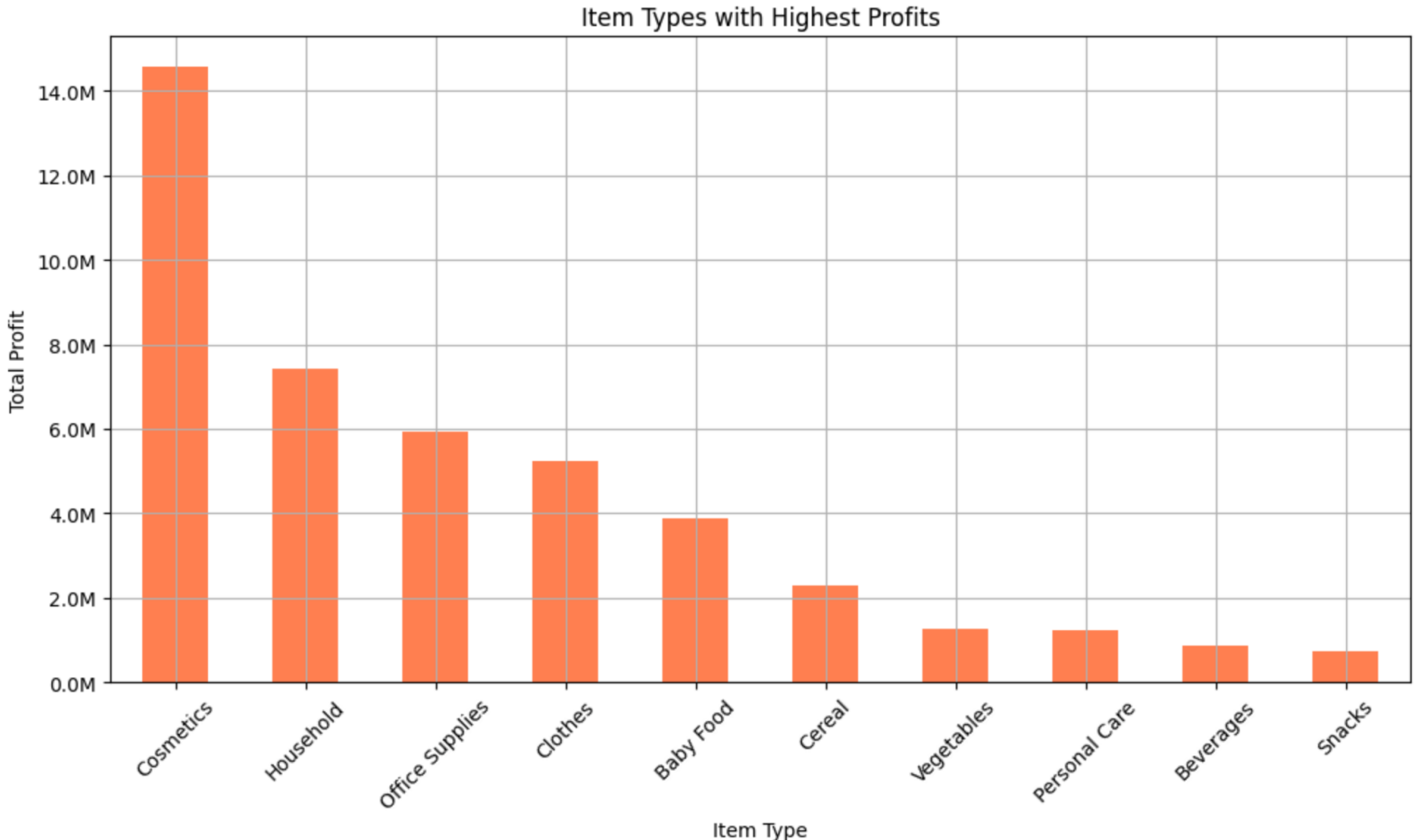
Key Factors


Top 10 Highest Profitable Months
Total Profit
3.5M
3.0M
2.5M
2.0M
1.5M
1.0M
0.5M
0.0M
(2013, 7)
(2016, 11)
(2012, 9)
(2017, 2)
(2014, 4)
(2016, 12)
(2012, 2)
(2010, 10)
(2010, 2)
(2010, 11)
Year-Month

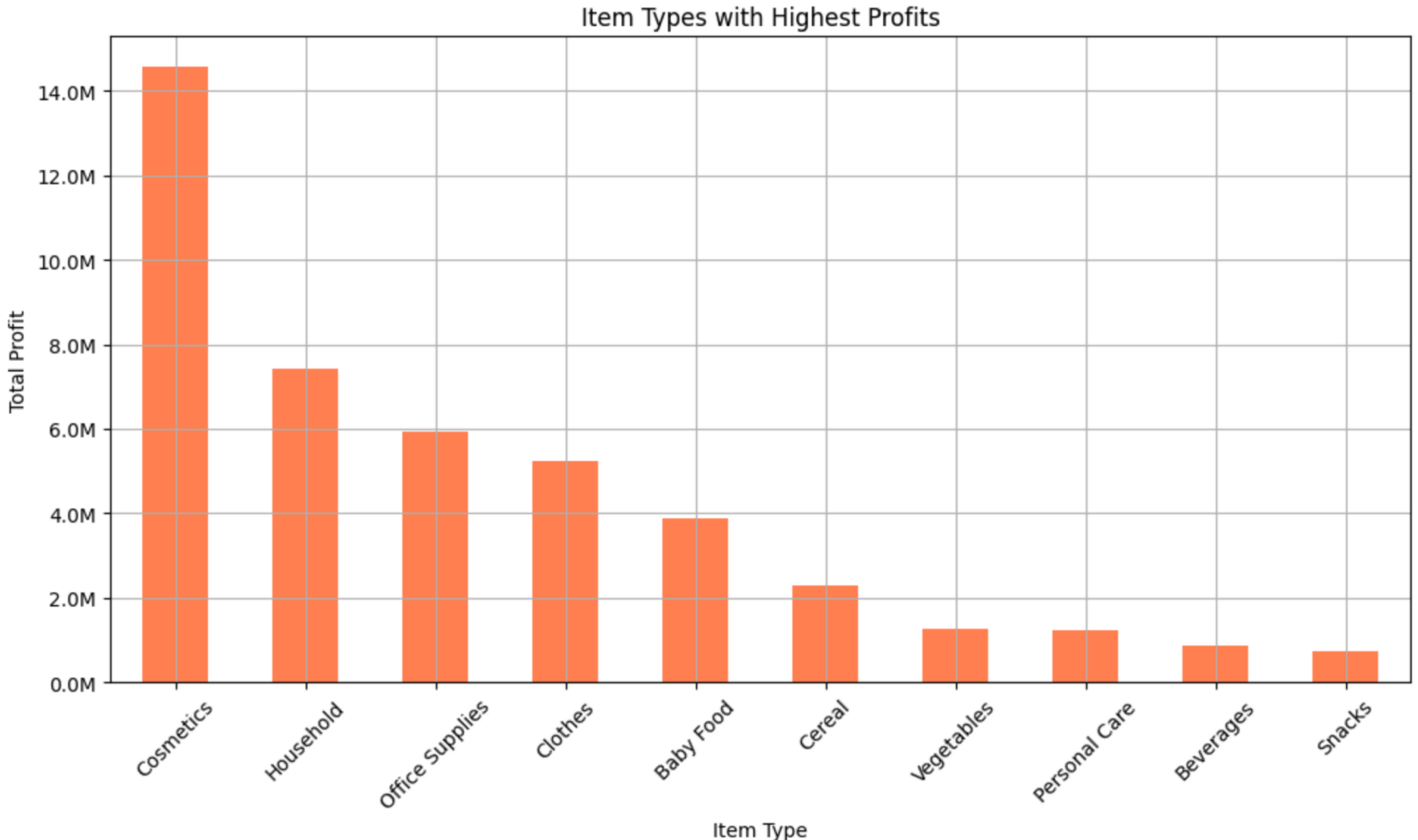
Key Factors


Item Types with Least Profits
Total Profit
6.0M
5.0M
4.0M
3.0M
2.0M
1.0M
0.0M
Fruits
Meat
Snacks
Beverages
Personal Care
Vegetables
Cereal
Baby Food
Clothes
Office Supplies
Item Type

Key Factors


Item Types with Highest Profits
Total Profit
14.0M
12.0M
10.0M
8.0M
6.0M
4.0M
2.0M
0.0M
Cosmetics
Household
Office Supplies
Clothes
Baby Food
Cereal
Vegetables
Personal Care
Beverages
Snacks
Item Type

Key Factors


Item Types with Highest Profits
Total Profit
14.0M
12.0M
10.0M
8.0M
6.0M
4.0M
2.0M
0.0M
Cosmetics
Household
Office Supplies
Clothes
Baby Food
Cereal
Vegetables
Personal Care
Beverages
Snacks
Item Type

Key Factors


Revenue vs. Profit by Region
Total Profit
Total Revenue
0.0M
0.2M
0.5M
0.8M
1.0M
1.2M
1.5M
1.8M
0.0M
1.0M
2.0M
3.0M
4.0M
5.0M
6.0M
Australia and Oceania
Central America and the Caribbean
Europe
Sub-Saharan Africa
Asia
Middle East and North Africa
North America

Key Factors


Revenue vs. Profit by Region
Total Profit
Total Revenue
0.0M
0.2M
0.5M
0.8M
1.0M
1.2M
1.5M
1.8M
0.0M
1.0M
2.0M
3.0M
4.0M
5.0M
6.0M
Australia and Oceania
Central America and the Caribbean
Europe
Sub-Saharan Africa
Asia
Middle East and North Africa
North America

THANK YOU

Report by - Aadithya Ram

Full Code - https://github.com/Aadithya-4010002/Amazon-Sales-Data-Analytics
LinkedIn – linkedin.com/in/aadiithyyya