ANALYZING AMAZON SALES DATA

UNIFIED MENTOR INTERNSHIP - PROJECT 1

BY
M POOJA

CONTENTS

INTRODUCTION

The Amazon sales data encompasses a comprehensive set of information capturing various facets of sales transactions. It includes details such as region, country, item type, sales channel, order priority, order date, order ID, ship date, units sold, unit price, unit cost, total revenue, total cost, and total profit. This rich dataset enables a thorough examination of sales trends, regional variations, and the impact of different factors on the overall profitability of the products. The data serves as a valuable resource for uncovering insights into customer preferences, market dynamics, and operational efficiency within the context of the Amazon retail platform. Analyzing this dataset can provide crucial business intelligence for optimizing strategies, reducing costs, and enhancing overall sales performance.

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 the Sales-trend -> month-wise, year-wise, yearly_month-wise

Find key metrics and factors and show the meaningful relationships between attributes.

METHODOLOGY USED

The Amazon sales data analysis was conducted using **Python programming language** within the **Google Colab** environment. The methodology involved several key steps.

- Firstly, the dataset was **imported and cleaned** using existing library to ensure data integrity. The temporal aspects of the data, such as order dates, were processed to extract month, year, and yearly-month information for trend analysis.
- **Visualization techniques**, primarily leveraging the matplotlib library, were employed to create insightful charts such as **bar charts, pie charts, and trendline graphs** to represent sales trends and key metrics.
- **Numpy** was used for numerical operations, enhancing the efficiency of data manipulation.
- Throughout the analysis, a combination of descriptive statistics and visualizations was employed to **identify patterns, correlations, and meaningful relationships** within the dataset.

The goal was to derive actionable insights for optimizing business strategies and ***improving overall sales performance in the Amazon marketplace***.

INFERENCES

SALES FOR EVERY MONTH

Calculating Sales for every month

```
month_sales=data.groupby('month')['Total Revenue'].sum()
month_sales
```

```
month
1     10482467.12
```

ps://colab.research.google.com/drive/1g60MbAwjC0LD0Fx3KMbVIL

/24, 7:26 PM

```
2     24740517.77
3      2274823.87
4     16187186.33
5     13215739.99
6      5230325.77
7     15669518.50
8      1128164.91
9      5314762.56
10    15287576.61
11    20568222.76
12     7249462.12
Name: Total Revenue, dtype: float64
```

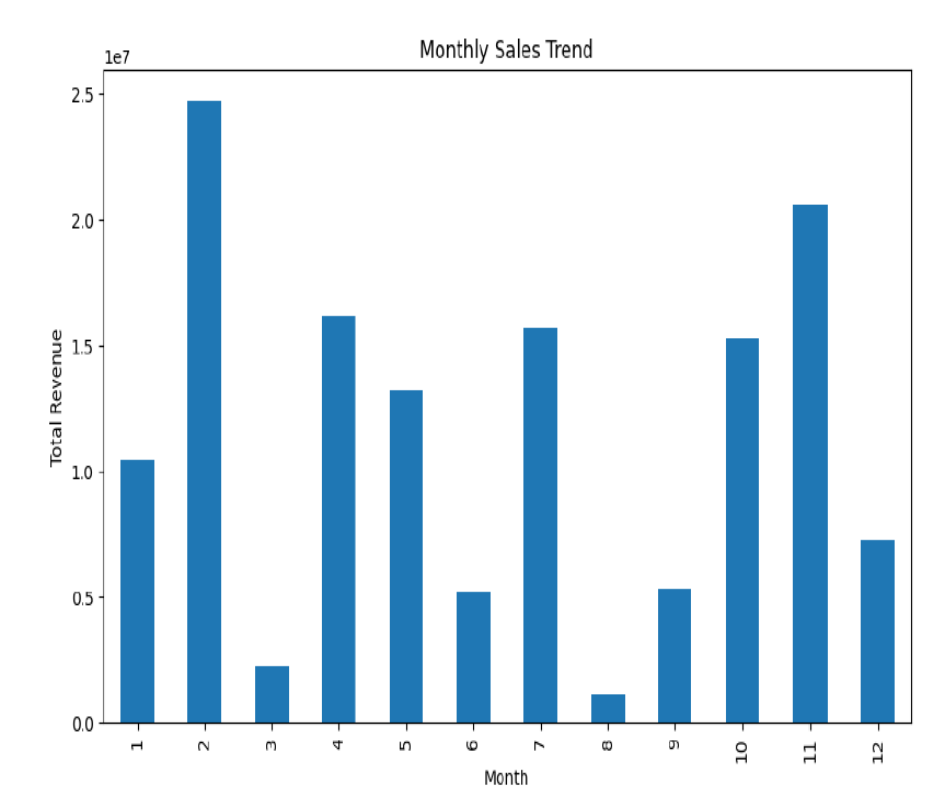
SALES FOR EVERY YEAR

Calculating sales for every year

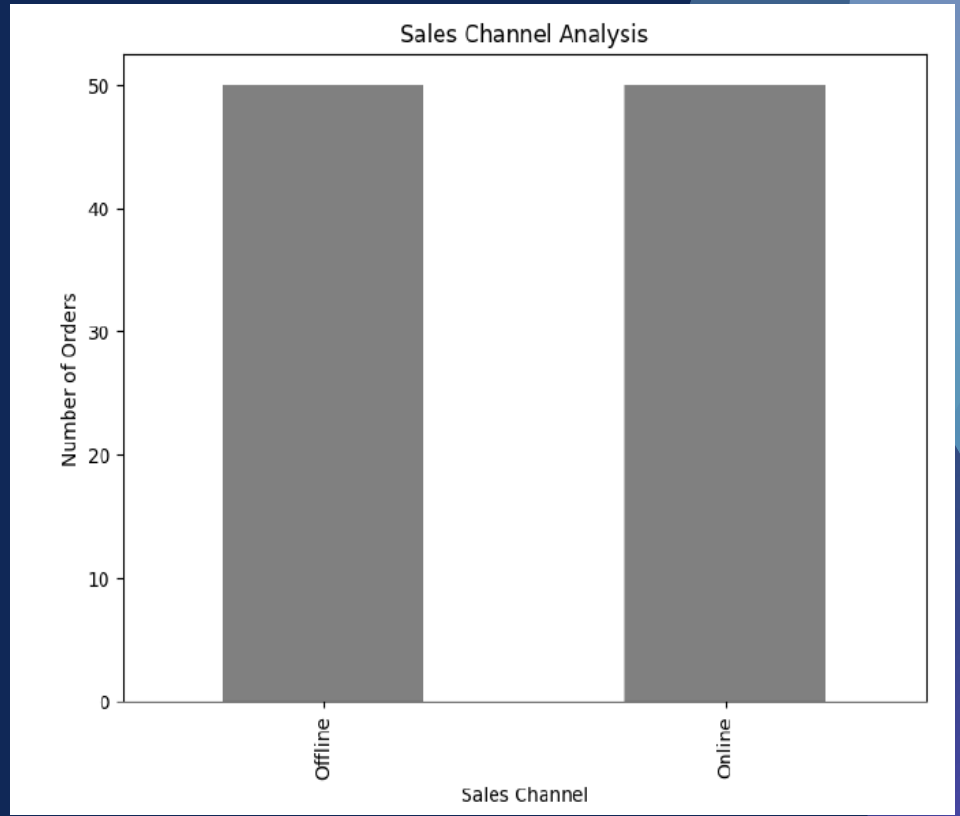
```
year_sales=data.groupby('year')['Total Revenue'].sum()
year_sales
```

```
year
2010    19186024.92
2011    11129166.07
2012    31898644.52
2013    20330448.66
2014    16630214.43
2015    12427982.86
2016    12372867.22
2017    13373419.63
Name: Total Revenue, dtype: float64
```

MONTHLY SALES TREND



SALES CHANNEL ANALYSIS



AVERAGE SALES REVENUE

1373487.6831

TOP 5 DEMANDED ITEMS

```
Item Type
Cosmetics          36601509.60
Office Supplies    30585380.07
Household          29889712.29
Baby Food          10350327.60
Clothes             7787292.80
Name: Total Revenue, dtype: float64
```

AVERAGE PROFIT ATTAINED

36.212

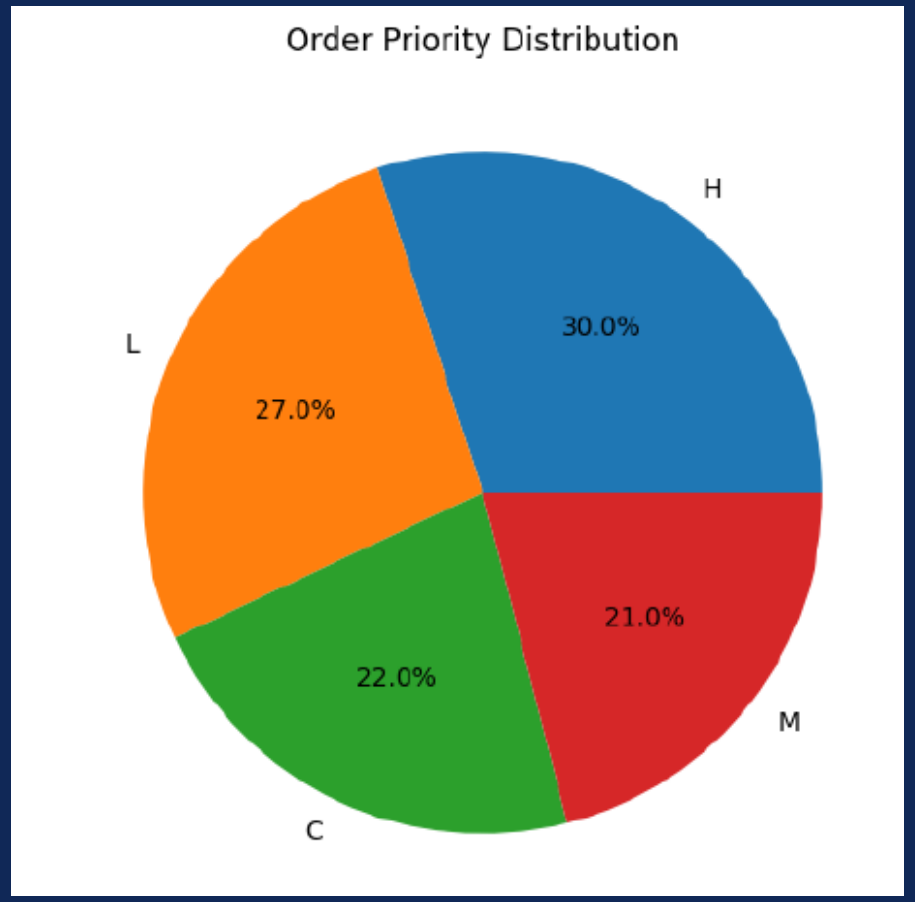
MAXIMUM PROFIT ATTAINED

67.204

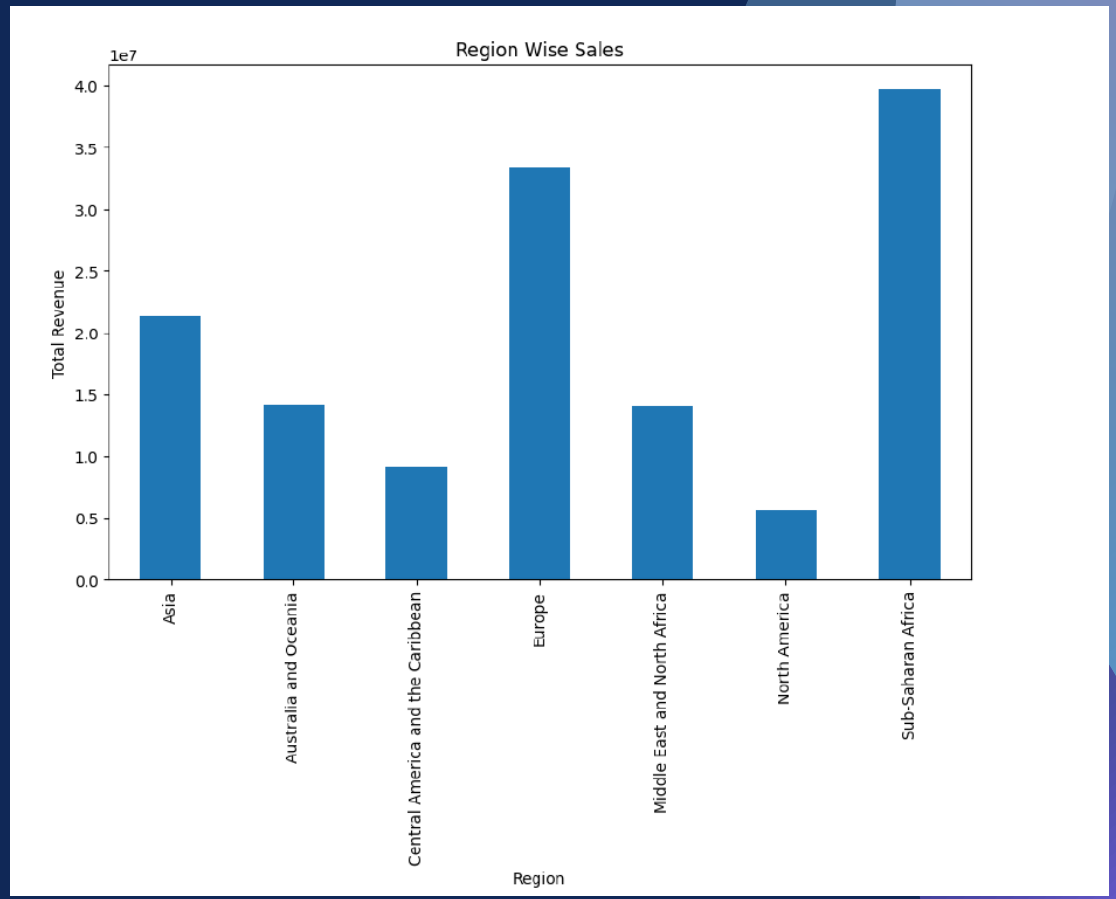
MINIMUM PROFIT ATTAINED

13.558

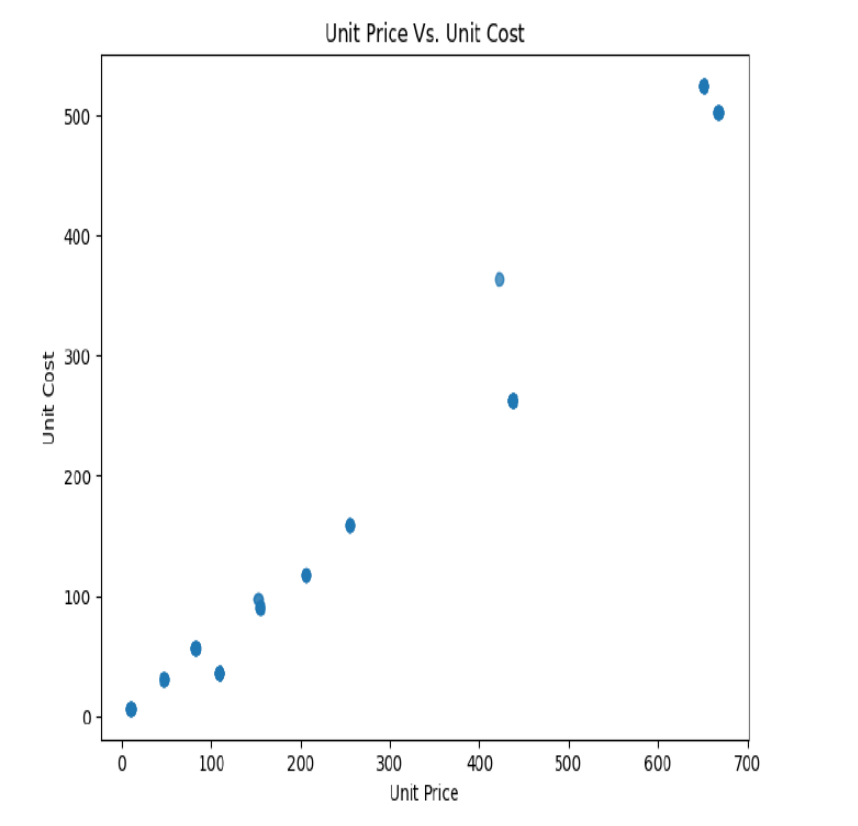
PRIORITY DISTRIBUTION



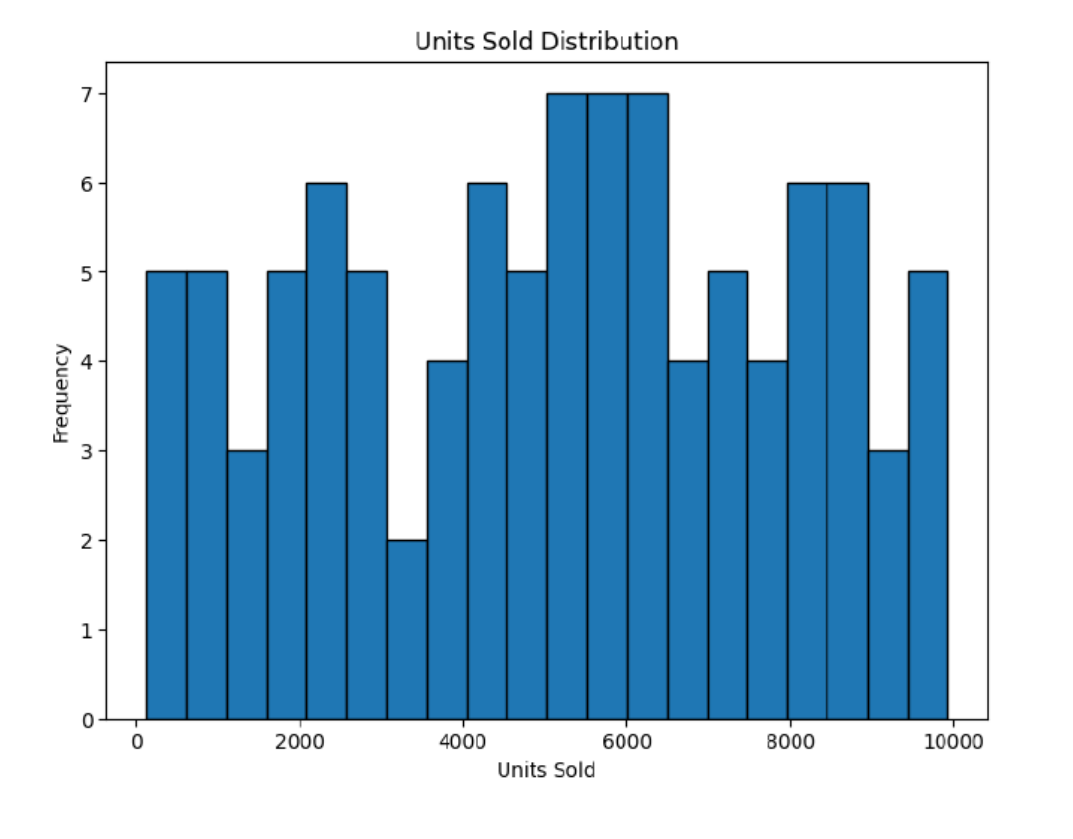
REGION WISE SALES ANALYSIS



UNIT PRICE vs COST PRICE



DISTRIBUTION OF UNITS SOLD



REALATION BETWEEN THE ATTRIBUTES

	Order ID	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit	month	year
Order ID	1.000000	-0.222907	-0.190941	-0.213201	-0.314688	-0.328944	-0.234638	-0.111219	0.081752
Units Sold	-0.222907	1.000000	-0.070486	-0.092232	0.447784	0.374746	0.564550	-0.007995	0.012455
Unit Price	-0.190941	-0.070486	1.000000	0.987270	0.752360	0.787905	0.557365	-0.031917	-0.061791
Unit Cost	-0.213201	-0.092232	0.987270	1.000000	0.715623	0.774895	0.467214	-0.042016	-0.071567
Total Revenue	-0.314688	0.447784	0.752360	0.715623	1.000000	0.983928	0.897327	0.003835	-0.037128
Total Cost	-0.328944	0.374746	0.787905	0.774895	0.983928	1.000000	0.804091	-0.015617	-0.050899
Total Profit	-0.234638	0.564550	0.557365	0.467214	0.897327	0.804091	1.000000	0.051366	0.002196
month	-0.111219	-0.007995	-0.031917	-0.042016	0.003835	-0.015617	0.051366	1.000000	-0.106715
year	0.081752	0.012455	-0.061791	-0.071567	-0.037128	-0.050899	0.002196	-0.106715	1.000000

CONCLUSION

In conclusion, the analysis of **Amazon sales data** using Python in **Google Colab** has yielded valuable insights into sales trends, regional variations, and key metrics influencing profitability.

The project successfully employed data cleaning techniques, temporal processing, and visualization methods to uncover patterns and relationships within the dataset. The identified insights, ranging from month-wise and year-wise trends to the impact of different factors on sales, can inform strategic decision-making and contribute to business optimization.

Moving forward, there is immense potential for **further exploration**, such as advanced **predictive modeling to forecast future sales trends, incorporating machine learning algorithms** for more nuanced insights, and exploring additional dimensions of the dataset for a comprehensive understanding.

The robust methodology employed in this project lays the foundation for ongoing analyses and enhancements, positioning it as a valuable resource for continuous improvement and adaptation to the dynamic landscape of Amazon sales.

THANK YOU

M POOJA
poojaofficial968@gmail.com
www.linkedin.com/in/pooja-m-083556284