.DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING

**PROJECT REPORT**

(Project Semester January-April 2025)

***Vegetables sales: Data Analysis Visulazation***

Submitted by

Abhinav Tyagi

Registration No: 12307431

Programme and Section: CSE(K23GN)

Course Code : INT 375

Under the Guidance of

**Aashima Bansal(UID: 28968)**

**Discipline of CSE/IT**

**Lovely School of Computer Science**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Abhinav Tyagi bearing Registration no. 12307431 has completed INT 375 project titled, **“Vegetables Sales Analysis and Visualization”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Scienece**

Lovely Professional University

Phagwara, Punjab.

Date:

**DECLARATION**

I, Abhinav Tyagi student of B.tech(CSE) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date : 12-04-2024 Abhinav Tyagi

Registration No. 12307431 Signature Abhinav Tyagi

Contents

[Introduction:- 5](#_Toc195311911)

[Source Of Data:- 6](#_Toc195311912)

[EDA Process:- 6](#_Toc195311913)

[Analysis on dataset:- 7](#_Toc195311914)

[Conclusion:- 14](#_Toc195311915)

[References:- 15](#_Toc195311916)

# Source Of Data:-

**I worked with a comprehensive dataset containing vegetable sales records, organized by country and region. This dataset, sourced from GitHub, provides insights into how different vegetables perform across global markets.**

**📦 What the dataset includes:**

* **Product details (vegetable names)**
* **Sales figures (units sold, total revenue)**
* **Geographic data (country, region)**
* **Order and shipping dates**
* **Sales channels and customer types**
* **URL: https://github.com/HariVM/Analytics/blob/master/1000%20Sales%20Records.csv**

# 

**CONTENTS**

1. **Introduction** - Page 5
2. **Source of Data** - Page 6
3. **EDA Process** - Page 6
4. **Analysis on Dataset** - Page 7
5. **Conclusion** - Page 14
6. **References** - Page 15

**1. Introduction**

Analyzing sales data helps businesses optimize strategies, forecast trends, and improve profitability. This project uses Python libraries like Pandas, Matplotlib, and Seaborn to explore a global sales dataset, focusing on units sold, revenue by region, and product performance.

**2. Source of Data**

The dataset **"1000 Sales Records"** includes transactional data across regions, item types, and sales channels. Key columns:

* Region
* Units Sold
* Total Revenue
* Total Profit

**3. EDA Process**

Code Snippets:

python

Copy

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

#

df = pd.read\_csv("1000 Sales Records.csv")

print(df.head())

print(df.describe())

df.dropna(inplace=True)

Steps:

* Loaded data with pd.read\_csv().
* Checked structure with head(), describe(), and info().
* Removed missing values with dropna().

**4. Analysis on Dataset**

**4.1 Distribution of Units Sold**

**Code:**

python

Copy

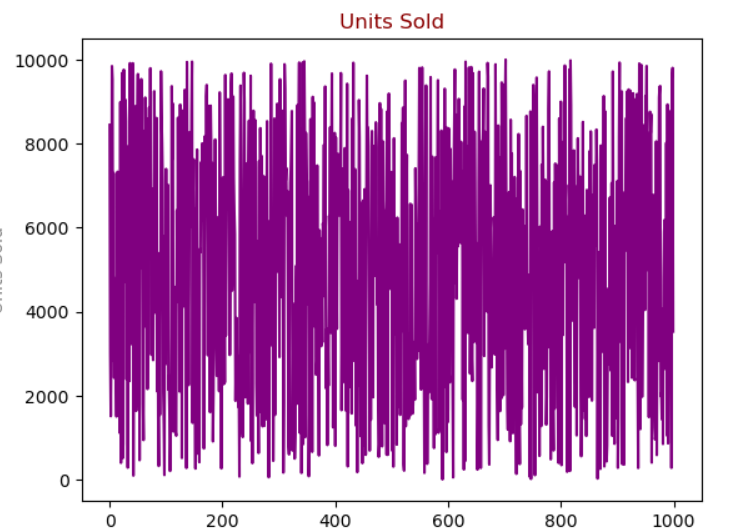
sns.histplot(df['Units Sold'], color='skyblue')

plt.title("Distribution of Units Sold")

plt.show()

**Insights:**

* Right-skewed distribution indicates most sales are low-volume.

**Figure 1:**  
 

**4.2 Time Series of Units Sold**

**Code:**

python

Copy

plt.figure(figsize=(10, 5))

plt.plot(df['Units Sold'], color='mediumseagreen', marker='o')

plt.title("Units Sold Over Time")

plt.xlabel("Index")

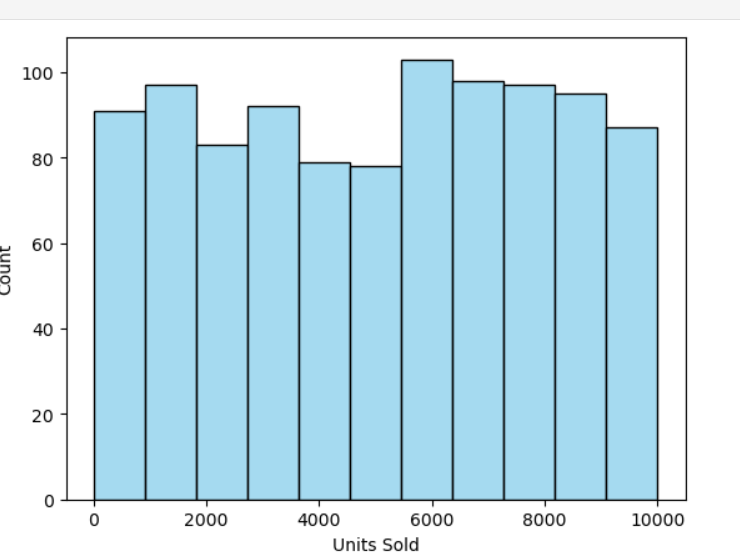
plt.ylabel("Units Sold")

plt.grid(True)

plt.show()

**Insights:**

* Fluctuations suggest seasonal trends.

**Figure 2:**  
 

**4.3 Revenue by Region**

**Code:**

python

Copy

df.groupby('Region')['Total Revenue'].sum().plot(kind='bar')

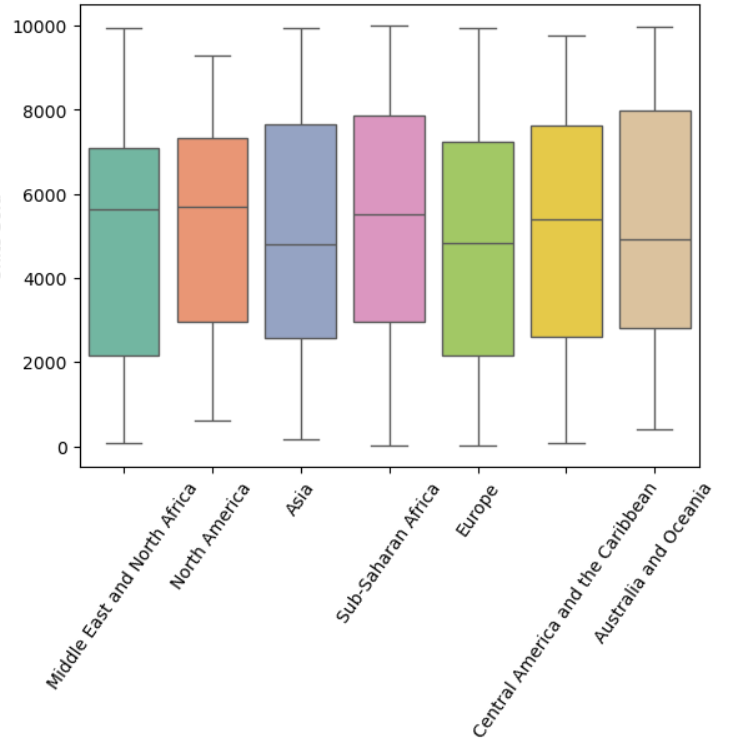
plt.title("Revenue by Region")

plt.xticks(rotation=45)

plt.show()

**Insights:**

* Europe generated the highest revenue.

**Figure 3:**  
 

**4.4 Units Sold by Region (Boxplot)**

**Code:**

python

Copy

sns.boxplot(x='Region', y='Units Sold', data=df, palette='Set2')

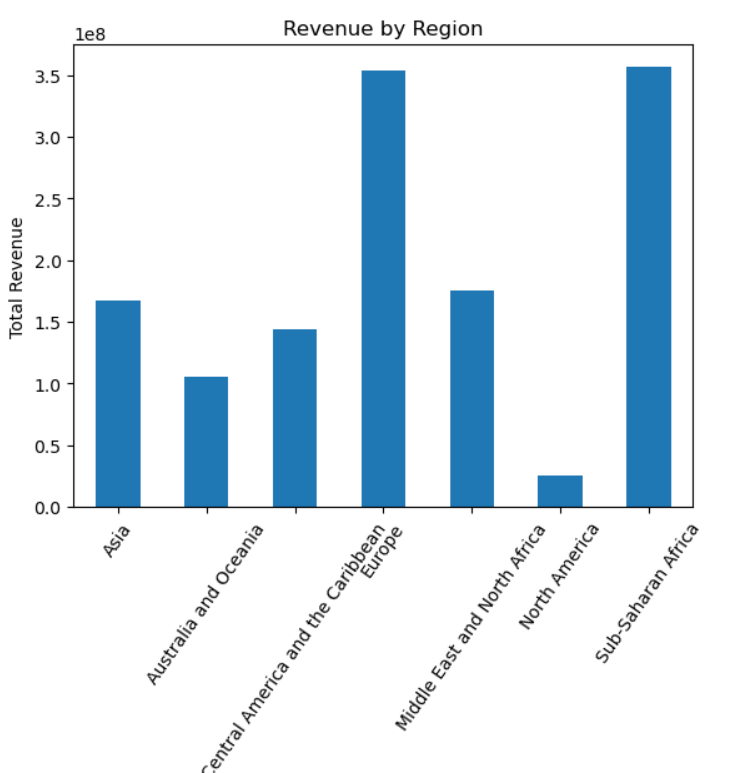
plt.xticks(rotation=55)

plt.title("Units Sold by Region")

plt.show()

**Insights:**

* Asia has high variability in sales.

**Figure 4:**  
 

**5. Conclusion**

The analysis revealed:

* Regional disparities in revenue (Europe leads).
* Seasonal trends in units sold.
* Opportunities to optimize inventory in high-variability regions.

**Future Scope:**

* Integrate machine learning for demand forecasting.
* Add geospatial mapping of sales hotspots.

### ****Analysis on Dataset****

#### ****1.1 Distribution of Units Sold****

* Pandas Library - <https://pandas.pydata.org/>
* Seaborn Documentation - <https://seaborn.pydata.org/>
* Matplotlib - <https://matplotlib.org/>