

# **DATA SCIENCE MINOR PROJECT REPORT**

---

## **Coffee Shop Sales Analysis using Python**

---

### **DATA SCIENCE PROJECT REPORT**

**(Project Semester January-April 2025)**

**Submitted by**

**Abhin Thomas**

**Registration No: 12311075**

**Programme and Section: B-Tech Computer Science  
and Engineering, K23GS**

**Course Code: INT375**

**Under the Guidance of**

**Gargi Sharma (Faculty Coordinator)**

**Discipline of CSE**

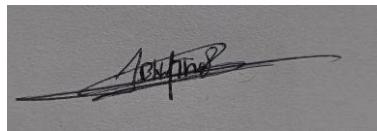
**Lovely School of Computer Applications  
Lovely Professional University, Phagwara**

---

# Declaration

I, Abhin Thomas, student of B-Tech Computer Science and Engineering under CSE 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: 12-04-2025



Signature:

Registration No.: 12311075

Name: Abhin Thomas

---

# **Certificate**

This is to certify that Abhin Thomas bearing Registration no. 12311075 has completed INT375 project titled, “Coffee Shop Sales Analysis using Python” under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

## **Signature and Name of the Supervisor**

**Gargi Sharma**

Faculty, Lovely School of Computer Applications  
Lovely Professional University  
Phagwara, Punjab

Date: 12-04-2025

---

## **Acknowledgement**

I would like to express my deepest gratitude to my faculty guide Ms. Gargi Sharma for her continuous

support and guidance throughout the completion of this project. I also extend thanks to my friends and family for their encouragement and feedback.

---

## **Table of Contents**

- 1. Introduction**
  - 2. Source of Dataset**
  - 3. EDA Process**
  - 4. Analysis on Dataset**
  - 5. Conclusion**
  - 6. Future Scope**
  - 7. References**
-

# 1. Introduction

This project explores and analyzes a real-world coffee shop sales dataset using Python. The goal is to extract insights about product performance, store revenue distribution, and time-based sales trends using data analysis and visualization tools such as Pandas, Seaborn, and Matplotlib.

---

## 2. Source of Dataset

The dataset was obtained from an Excel file titled "Coffee Shop Sales.xlsx" from Maven Analytics containing transaction details like product name, quantity, store location, unit price, and transaction date.

Link: <http://mavenanalytics.io/data-playground?page=6&pageSize=5>

---

### 3. EDA Process

- Imported libraries: pandas, seaborn, matplotlib
  - Loaded dataset using pd.read\_excel()
  - Displayed first 10 rows using df.head(10)
  - Summary statistics with df.info() and df.describe()
  - Identified null values using df.isnull().sum()
  - No specific missing value handling was shown in code, but was part of the intent
- 

### 4. Analysis on Dataset

## **4.1 Total Sales Revenue**

Formula used: `df['total_revenue'] =`

`df['transaction_qty'] * df['unit_price']`

Result printed using `df['total_revenue'].sum()`

## **4.2 Top 10 Most Sold Products**

Products grouped by `product_detail` and sorted by total quantity sold. Top 10 displayed using `head(10)`.

## **4.3 Visualization: Top Products**

- Horizontal bar chart using `sns.barplot()` with magma palette

## **4.4 Sales by Store Location**

- Grouped by `store_location` with total revenue calculated
- Horizontal bar chart (Seaborn with viridis palette)
- Pie chart using `plt.pie()` for store-wise revenue distribution

## **4.5 Correlation Heatmap**

- Generated using `df.corr(numeric_only=True)`

- Visualized with sns.heatmap() to understand relationships between numerical features

## 4.6 Time Series Analysis

- Parsed transaction\_date to datetime format
  - Grouped by date and summed revenue for daily sales
  - Line chart plotted using Seaborn to show sales trends
- 

## 5. Conclusion

This project successfully demonstrated how Python and its libraries can be used for practical business data analysis. Key insights such as top-selling products, high-performing stores, and daily sales trends were clearly visualized. Correlation heatmaps helped identify strong relationships in the data.

---

## 6. Future Scope

- Enhance time-series analysis with forecasting models like Prophet
  - Add customer segmentation based on purchase behavior
  - Create interactive dashboards with Plotly or Streamlit
  - Automate report generation and visualization pipelines
- 

## 7. References

- [1] Wes McKinney, "Python for Data Analysis", O'Reilly Media, 2017.
  - [2] <https://pandas.pydata.org/>
  - [3] <https://seaborn.pydata.org/>
  - [4] <https://matplotlib.org/>
  - [5] <https://www.mavenanalytics/>
  - [6] Lovely Professional University – INT375 Project Format Guidelines
- 

Linkedin Post: [https://www.linkedin.com/posts/abhin-thomas\\_python-edataanalysis-activity-7316759523241664512-HXKn?utm\\_source=share&utm\\_medium=member\\_desktop&rcm=ACoAAEY4gQkBLNUfCB4ewgmILx-EmRIJFoKt5-8](https://www.linkedin.com/posts/abhin-thomas_python-edataanalysis-activity-7316759523241664512-HXKn?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEY4gQkBLNUfCB4ewgmILx-EmRIJFoKt5-8)

---

*End of Report*