## **Project Title: Supermarket Sales Analysis**

### **Goal**:

Analyze Point-of-Sale (POS) data to uncover business insights such as:

* Peak sales hours
* Best performing branches
* Payment method popularity
* Product line profitability

## **Dataset**

**Name**: Supermarket Sales Dataset  
**Source**: Kaggle  
**Link**: [Supermarket Sales Dataset on Kaggle](https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales)

**File**: supermarket\_sales.csv

## **Tools Used**

* Python
* Pandas
* Matplotlib
* Seaborn

## **Project Folder Structure**

supermarket-sales-analysis/  
├── data/  
│ └── supermarket\_sales.csv  
├── supermarket\_sales\_analysis.py  
├── requirements.txt  
├── README.md  
└── .gitignore

## **Step-by-Step Python Code**

### 1. Import Libraries

import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
from datetime import datetime  
  
sns.set(style="whitegrid")

### 2. Load and Explore Dataset

df = pd.read\_csv("data/supermarket\_sales.csv")  
  
print("First 5 rows:")  
print(df.head())  
print("\nDataset Info:")  
print(df.info())

### **3. Preprocessing & Feature Engineering**

# Convert Date and Time  
df['Date'] = pd.to\_datetime(df['Date'])  
df['Time'] = pd.to\_datetime(df['Time']).dt.time  
df['Hour'] = pd.to\_datetime(df['Time'], format='%H:%M:%S').dt.hour

### **4. Peak Sales Hours**

plt.figure(figsize=(10,6))  
sns.histplot(df['Hour'], bins=24, kde=False)  
plt.title("Sales Volume by Hour of Day")  
plt.xlabel("Hour")  
plt.ylabel("Number of Sales")  
plt.xticks(range(0, 24))  
plt.tight\_layout()  
plt.show()

### **5. Best Performing Branches**

branch\_sales = df.groupby('Branch')['Total'].sum().sort\_values(ascending=False)  
  
plt.figure(figsize=(8, 5))  
sns.barplot(x=branch\_sales.index, y=branch\_sales.values)  
plt.title("Total Sales by Branch")  
plt.xlabel("Branch")  
plt.ylabel("Total Sales")  
plt.tight\_layout()  
plt.show()

### 6. Payment Method Popularity

plt.figure(figsize=(7, 5))  
sns.countplot(data=df, x='Payment', order=df['Payment'].value\_counts().index)  
plt.title("Payment Method Distribution")  
plt.xlabel("Payment Method")  
plt.ylabel("Number of Transactions")  
plt.tight\_layout()  
plt.show()

### 7. Product Line Profitability

product\_sales = df.groupby('Product line')['Total'].sum().sort\_values()  
  
plt.figure(figsize=(10, 6))  
sns.barplot(x=product\_sales.values, y=product\_sales.index)  
plt.title("Total Sales by Product Line")  
plt.xlabel("Total Sales")  
plt.ylabel("Product Line")  
plt.tight\_layout()  
plt.show()

## requirements.txt

pandas  
matplotlib  
seaborn

## README.md

# Supermarket Sales Analysis  
  
This project analyzes POS data from a supermarket chain to uncover:  
- Peak sales hours  
- Best performing branches  
- Payment method popularity  
- Product line profitability  
  
## Dataset  
- [Kaggle - Supermarket Sales](https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales)  
  
## How to Run  
1. Clone this repo  
2. Add `supermarket\_sales.csv` inside `data/` folder  
3. Install dependencies: `pip install -r requirements.txt`  
4. Run: `python supermarket\_sales\_analysis.py`  
  
## Tools  
- Python, Pandas, Seaborn, Matplotlib

## **Steps to Upload to GitHub**

1. Go to [GitHub](https://github.com) → Click **New Repository**
2. Name it: supermarket-sales-analysis
3. Locally:

cd supermarket-sales-analysis  
git init  
git add .  
git commit -m "Initial commit - supermarket sales analysis"  
git remote add origin https://github.com/YOUR\_USERNAME/supermarket-sales-analysis.git  
git branch -M main  
git push -u origin main