INTRODUCTION TO DATA MANAGEMENT

PROJECT REPORT

(Project Semester January-April 2025)

COFFEE SHOP SALES ANALYSIS

Submitted by

Ritik Kumar

Registration No- 12308707

Programme and Section BTECH -CSE

K23GW

Course Code – INT-217

Under the Guidance of

Baljinder Kaur

(30453)

Discipline of CSE/IT

Lovely School of COMPUTER SCIENCE ENGINEERING

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Ritik Kumar bearing Registration no. 12308707 has completed INT-217 project titled, "COFFEE SHOP SALES ANALYSIS" 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.

Name of the Supervisor: Baljinder Kaur
Designation of the Supervisor: Assistant Professor
School of Computer Science Engineering
Lovely Professional University
Phagwara, Punjab.

Date: 10th April, 2025

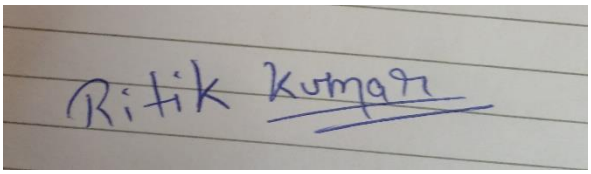
DECLARATION

I, Ritik Kumar, student of B.Tech 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: 10th April, 2025

Registration No.: 12308707

Signature:



Acknowledgement

I would like to express my sincere gratitude to all those who supported and guided me throughout the completion of the Coffee Shop Sales Analysis Project.

First and foremost, I would like to thank my mentor/guide Baljinder Ma'am for their valuable insights, constant encouragement, and expert guidance throughout the course of this project. Their constructive feedback and suggestions greatly improved the quality of my work. I would also like to thank LOVELY PROFESSIONAL UNIVERSITY for providing me with the opportunity and resources to carry out this project successfully. A special thanks to my peers and teammates for their cooperation, support, and helpful discussions that contributed to the overall development of this project.

Lastly, I am thankful to my family and friends for their motivation, patience, and unwavering support during the project duration. This project has been a great learning experience and has helped me gain a deeper understanding of sales analytics and business intelligence tools, which I believe will be beneficial in my future endeavors.

Coffee Shop Sales Analysis

1. Introduction:

The Coffee Shop Sales Dataset captures fine-grained electronic data for a chain of coffee shops across various locations. It encompasses 149,116 single transactions occurring over one complete calendar year of 2023, thus providing an excellent opportunity to conduct exploratory data analysis and gain business insight.

- Each transaction contains important attributes such as:
    - Date and time of transaction
    - Quantity and unit price
    - Total amount of sale
    - Product category, classification, and description
    - Store location and hour of purchase

- The data are organized in such a way to facilitate a multitude of analyses:

    - Time Series: A variable set including month, day of the week, and hour allows assessing trends on a daily, weekly, or monthly basis to establish when the business is busiest and to compare seasonality effects.

    - Product Level: Categorization of data into groups like Coffee, Tea, and Drinking Chocolate can be further disintegrated into product types and details of specific items.

    - Location-Wise Performance: Multiple locations allow for comparative studies by region, performance, and customer preferences.

    - Revenue & Sales Tracking: Price total per transaction enables easy aggregation of revenue and insights into profitability.

This dataset presents a unique opportunity for exploratory data analysis and application in advanced business intelligence concepts, including demand forecasting, inventory planning, and targeting marketing strategies.

2. SOURCE OF DATASET:

This dataset was obtained from **Maven Analytics**, a platform that provides high-quality, real-world datasets for data analysis, business intelligence, and data science practice.

- You can access the dataset via the Maven Analytics Data Playground using the following link: https://mavenanalytics.io/data-playground?order=date_added%2Cdesc&page=6&pageSize=5

3. Exploratory Data Analysis (EDA) Process:

The Exploratory Data Analysis (EDA) process provides a systematic approach to understand the patterns, trends, and relationships within the Coffee Shop Sales dataset. Through this process, we aim to uncover key insights related to product performance, customer behavior, sales timing, and store operations.

- Initial Data Exploration:

The dataset contains over 149,000 transaction records, spanning six months (January to June 2023). Each record captures detailed information, including:

- Transaction date and time
- Product details (category, type, name)
- Quantity sold and price per unit
- Store location
- Total transaction value

We began by loading the dataset and exploring its structure using basic summary functions. This step involved checking the shape of the dataset, reviewing sample rows, understanding column names, and evaluating the types of data contained.

- Data Quality Checks:

Before diving deeper into analysis, we ensured the dataset was clean and reliable.

- Missing Values: No missing values were found in any column.
- Duplicate Records: There were no duplicate entries.
- Consistent Data Types: All fields had appropriate data types:
    - Transaction Date: date and time
    - Price, Total Price, Quantity: numeric
    - Categorical fields (e.g., product names, store locations): string (object)

This meant that no additional cleaning or type conversion was necessary, and the dataset was analysis-ready from the start.

➢ Feature Engineering:

To enhance analysis, we created new time-related features derived from the Transaction Date and Transaction Time:

- Month: Extracted to observe monthly sales trends.
- Weekday: Helps identify which days generate the most traffic.
- Hour: Indicates the time of day for each transaction, supporting hourly sales analysis.

These new features enabled more flexible time-based groupings and deeper insights into customer behavior.

➢ Univariate Analysis:

In this phase, we analyzed each feature individually to understand the distribution and frequency of values:

- Transaction Frequency: Count of transactions per month, weekday, and hour.
- Sales Volume: Number of items sold across categories and product types.
- Revenue: Total revenue from different products and store locations.
- Price Distribution: Range and variation in unit prices of products.

We found that:

- Some products and categories were significantly more popular than others.
- Certain store locations consistently performed better.
- Specific hours and days had noticeably higher sales.

➢ Bivariate and Multivariate Analysis:

This step explored relationships between multiple variables to uncover more complex patterns:

- Sales by Product Category and Store: Identified top-performing combinations.
- Hourly Trends by Store: Showed when each store had peak customer visits.
- Monthly Revenue Growth: Helped spot seasonal trends or growth.
- Weekday vs. Quantity Sold: Compared product demand across weekdays.

Grouping and aggregation methods (like group by) were used to summarize sales volume and revenue across different dimensions.

➢ Visual Analysis:

While some insights can be gathered numerically, visualizations were essential for communicating trends clearly. Planned visualizations include:

- Bar charts for best-selling products and categories
- Line charts for monthly revenue trends
- Heatmaps for hourly sales across stores
- Pie charts for category-wise revenue share

These visuals help reveal patterns not immediately obvious from raw data.

➢ Key Insights Generation:

The final EDA step involved interpreting results and documenting actionable insights. Examples might include:

- Which product category contributes the most to total revenue?
- What times of day have the highest customer traffic?
- Which store location performs best overall?
- Are there any slow-moving products or time slots?

These insights guide strategic business decisions such as inventory management, staffing, marketing focus, and pricing strategies.

4. Analysis on Dataset:

A. Track Total Performance:

i. Introduction:

Tracking overall sales performance helps evaluate how well the business is doing in terms of revenue and quantity sold over time. This high-level view gives insight into general trends and business health.

ii. General Description:

This analysis involves aggregating total sales revenue and quantities over the dataset's entire time period. It includes monthly breakdowns to show trends and highlights potential peak periods or slumps.

iii. Specific Requirements, Functions, and Formulas:

- groupby('Month') to track performance over time
- sum () of Total Price and Quantity
- Sorting by Month order (Jan–Jun)
- Columns used: Month, Quantity, Total Price

iv. Analysis Results:

- The business generated consistent revenue across all months, with a slight peak in March and May.
- Quantity sold and revenue are positively correlated.
- June showed a small dip, possibly indicating seasonality or shifting customer behavior.

v. Visualization:

- Line Chart: Monthly Revenue and Quantity Sold
- Bar Chart: Total Sales and Quantity per Month

B. Analyze Best-Selling Products:

i. Introduction:

Identifying the top-selling products helps in making decisions about inventory, promotions, and featured items. This analysis reveals what items customers love most.

ii. General Description:

This analysis sums up the quantity sold and revenue generated for each product. It focuses on top 10 products in terms of volume and income.

iii. Specific Requirements, Functions, and Formulas:

- groupby('Product_detail')
- Aggregate: sum(Quantity), sum(Total_Price)
- Sort descending by Quantity and Total_Price
- Columns used: Product_detail, Quantity, Total_Price

iv. Analysis Results:

- Ethiopia Rg Lg and Americano Lg were the most sold products.
- Flat White Lg and Cappuccino Md also performed well in revenue.
- Most top products fall under the Coffee category.

v. Visualization:

- Horizontal Bar Chart: Top 10 Products by Quantity
- Bar Chart: Top 10 Products by Revenue

C. Monitor Sales by Time and Day:

i. Introduction:

Analyzing when sales happen the most helps identify peak hours and busiest days. This insight supports staffing, promotions, and store management.

ii. General Description:

This analysis focuses on sales by hour and weekday. It helps determine high-traffic times and days of the week with highest revenue.

iii. Specific Requirements, Functions, and Formulas:

- groupby('Hours'), groupby('Weekday')
- Aggregate: sum(Total_Price)
- Sorting by time order and weekday (Mon to Sun)
- Columns used: Hours, Weekday, Total_Price

iv. Analysis Results:

- Peak hours: 11:00 AM to 1:00 PM — lunchtime rush
- Peak days: Saturday and Friday
- Sales dip slightly during early morning and late evening

v. Visualization:

- Line Chart: Revenue by Hour
- Bar Chart: Revenue by Weekday

D. Evaluate Sales by Location:

i. Introduction:

Store-wise sales analysis helps evaluate the performance of each branch and identify location-based preferences or trends.

ii. General Description:

The dataset includes three distinct store locations. This analysis compares them by total revenue and quantity sold.

iii. Specific Requirements, Functions, and Formulas:

- groupby('Store_location')
- Aggregate: sum(Quantity), sum(Total_Price)
- Sort by Total_Price
- Columns used: Store_location, Quantity, Total_Price

iv. Analysis Results:

- New York store leads in both revenue and sales volume.
- San Francisco follows closely.
- Los Angeles has the lowest overall sales but might have unique product popularity.

v. Visualization:

- Bar Chart: Revenue by Store
- Bar Chart: Quantity Sold by Store

E. Customer Purchase Behavior:

i. Introduction:

Understanding customer behavior is crucial for personalizing marketing, optimizing stock, and improving customer experience.

ii. General Description:

This analysis explores:

- Product categories customers prefer
- Most frequent purchase times
- Basket size (quantity per transaction)

iii. Specific Requirements, Functions, and Formulas:

- groupby('Product_category') and groupby('Hours')
- Histogram of Quantity per transaction
- Analyze co-purchased categories if available
- Columns used: Product_category, Hours, Quantity, Transaction_Id

iv. Analysis Results:

- Customers strongly prefer Coffee and Tea
- Most purchases occur late morning to early afternoon
- Basket size typically ranges from 1 to 2 items per transaction

v. Visualization:

- Pie Chart: Product Category Preference
- Histogram: Quantity per Transaction
- Line Chart: Purchase Frequency by Hour

5. Conclusion:

The Coffee Shop Sales analysis provided valuable insights into the business performance across products, time periods, and store locations. Through comprehensive exploratory data analysis and visual storytelling, we identified critical patterns and trends that can drive strategic decision-making.

• Consistent Monthly Performance:
Sales remained steady throughout the first half of 2023, with noticeable peaks in March and May, suggesting potential marketing or seasonal effects.

• Best-Selling Products:
Coffee-based beverages, particularly large-sized options like *Ethiopia Rg Lg* and *Americano Lg*, dominated both sales volume and revenue, highlighting strong customer preference for premium coffee items.

• Peak Sales Periods:
The highest sales occurred during late mornings and early afternoons, especially around 11 AM to 1 PM, indicating optimal times for promotional campaigns or increased staffing.

• Top-Performing Location:
The New York store led in both quantity and revenue, showing strong customer engagement, while Los Angeles showed potential for growth.

• Customer Behavior:
Coffee and Tea were the most preferred categories, and the average basket size was modest, pointing to quick individual purchases rather than bulk buying.

These findings provide a roadmap for optimizing inventory, staffing schedules, marketing initiatives, and product offerings. Moving forward, leveraging such data-driven insights can significantly enhance customer satisfaction and operational efficiency.

6. Future Scope:

While the current dashboard and analysis provide meaningful insights into sales performance and customer behavior, there is significant potential to enhance this project further through additional features and deeper analytics. Future enhancements could include:

• Expansion to Full-Year Data:
Extending the dataset to include all 12 months of 2023 (or multiple years) would allow for more accurate trend analysis, seasonal forecasting, and annual performance comparisons.

• Customer Segmentation:
Incorporating customer-level data (e.g., loyalty programs, demographics, or feedback) could enable personalized marketing strategies and customer lifetime value analysis.

• Inventory and Supply Chain Analysis:
Integrating inventory data could help monitor stock levels, reduce waste, and optimize ordering based on product demand patterns.

• Profitability Analysis:
Including cost data for each product would allow for gross margin calculations, helping identify the most profitable items, not just the best-sellers.

• Forecasting Models:
Implementing time series forecasting techniques in Excel or other tools (like Power BI or Python) could predict future sales, support demand planning, and prepare for peak periods.

• Interactive Dashboard Features:
Enhancing interactivity with dynamic filters, slicers, and dropdowns can improve user experience, allowing business users to explore data in real-time based on selected criteria (e.g., store, category, time range).

• Cross-Platform Integration:
Migrating or complementing the dashboard using advanced BI tools like Power BI or Tableau could allow for more scalable, cloud-based sharing, and richer visuals.

7. Dashboard Image:

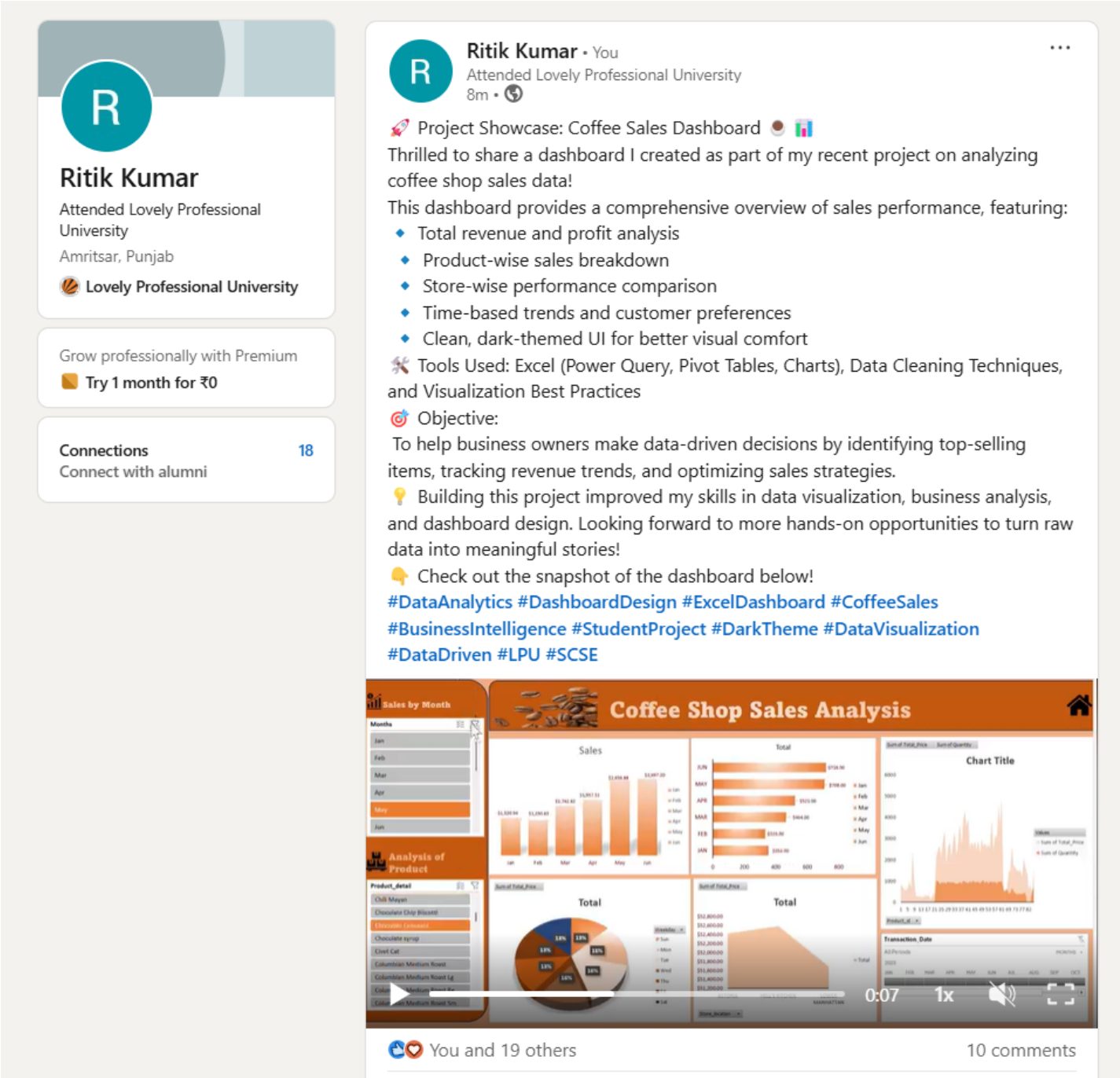
8. Dataset Image:

Transaction_Id	Transaction_Date	Transaction_Time	Quantity	Store_id	Store_location	Product_id	Price	Product_category	Product_type	Product_detail	Total_Price	Month	Weekday	Hours
1	01-01-2023	07:06:11	2	5	Lower Manhattan	32	$3.00	Coffee	Gourmet brewed coffee	Ethiopia Rg	$6.00	Jan	Sun	7
2	01-01-2023	07:08:56	2	5	Lower Manhattan	57	$3.10	Tea	Brewed Chai tea	Spicy Eye Opener Chai Lg	$6.20	Jan	Sun	7
3	01-01-2023	07:14:04	2	5	Lower Manhattan	59	$4.50	Drinking Chocolate	Hot chocolate	Dark chocolate Lg	$9.00	Jan	Sun	7
4	01-01-2023	07:20:24	1	5	Lower Manhattan	22	$2.00	Coffee	Drip coffee	Our Old Time Diner Blend Sm	$2.00	Jan	Sun	7
5	01-01-2023	07:22:41	2	5	Lower Manhattan	57	$3.10	Tea	Brewed Chai tea	Spicy Eye Opener Chai Lg	$6.20	Jan	Sun	7
6	01-01-2023	07:22:41	1	5	Lower Manhattan	77	$3.00	Bakery	Scone	Oatmeal Scone	$3.00	Jan	Sun	7
7	01-01-2023	07:25:49	1	5	Lower Manhattan	22	$2.00	Coffee	Drip coffee	Our Old Time Diner Blend Sm	$2.00	Jan	Sun	7
8	01-01-2023	07:33:34	2	5	Lower Manhattan	28	$2.00	Coffee	Gourmet brewed coffee	Columbian Medium Roast Sm	$4.00	Jan	Sun	7
9	01-01-2023	07:39:13	1	5	Lower Manhattan	39	$4.25	Coffee	Barista Espresso	Latte Rg	$4.25	Jan	Sun	7
10	01-01-2023	07:39:34	2	5	Lower Manhattan	58	$3.50	Drinking Chocolate	Hot chocolate	Dark chocolate Rg	$7.00	Jan	Sun	7
11	01-01-2023	07:43:05	1	5	Lower Manhattan	56	$2.55	Tea	Brewed Chai tea	Spicy Eye Opener Chai Rg	$2.55	Jan	Sun	7
12	01-01-2023	07:44:35	2	5	Lower Manhattan	33	$3.50	Coffee	Gourmet brewed coffee	Ethiopia Lg	$7.00	Jan	Sun	7
13	01-01-2023	07:45:51	1	5	Lower Manhattan	51	$3.00	Tea	Brewed Black tea	Earl Grey Lg	$3.00	Jan	Sun	7
14	01-01-2023	07:48:19	1	5	Lower Manhattan	57	$3.10	Tea	Brewed Chai tea	Spicy Eye Opener Chai Lg	$3.10	Jan	Sun	7
15	01-01-2023	07:52:36	2	5	Lower Manhattan	87	$3.00	Coffee	Barista Espresso	Ouro Brasileiro shot	$6.00	Jan	Sun	7
16	01-01-2023	07:59:58	2	5	Lower Manhattan	47	$3.00	Tea	Brewed Green tea	Serenity Green Tea Lg	$6.00	Jan	Sun	7
17	01-01-2023	07:59:58	1	5	Lower Manhattan	79	$3.75	Bakery	Scone	Jumbo Savory Scone	$3.75	Jan	Sun	7
18	01-01-2023	08:00:18	1	8	Hell's Kitchen	42	$2.50	Tea	Brewed herbal tea	Lemon Grass Rg	$2.50	Jan	Sun	8
19	01-01-2023	08:00:39	2	8	Hell's Kitchen	59	$4.50	Drinking Chocolate	Hot chocolate	Dark chocolate Lg	$9.00	Jan	Sun	8
20	01-01-2023	08:11:45	1	8	Hell's Kitchen	61	$4.75	Drinking Chocolate	Hot chocolate	Sustainably Grown Organic Lg	$4.75	Jan	Sun	8
21	01-01-2023	08:17:27	2	8	Hell's Kitchen	33	$3.50	Coffee	Gourmet brewed coffee	Ethiopia Lg	$7.00	Jan	Sun	8
22	01-01-2023	08:24:26	2	5	Lower Manhattan	56	$2.55	Tea	Brewed Chai tea	Spicy Eye Opener Chai Rg	$5.10	Jan	Sun	8
23	01-01-2023	08:24:26	1	5	Lower Manhattan	69	$3.25	Bakery	Biscotti	Hazelnut Biscotti	$3.25	Jan	Sun	8
24	01-01-2023	08:29:38	1	8	Hell's Kitchen	56	$2.55	Tea	Brewed Chai tea	Spicy Eye Opener Chai Rg	$2.55	Jan	Sun	8
25	01-01-2023	08:31:23	1	8	Hell's Kitchen	40	$3.75	Coffee	Barista Espresso	Cappuccino	$3.75	Jan	Sun	8
26	01-01-2023	08:33:08	1	5	Lower Manhattan	43	$3.00	Tea	Brewed herbal tea	Lemon Grass Lg	$3.00	Jan	Sun	8
27	01-01-2023	08:33:08	1	5	Lower Manhattan	76	$3.50	Bakery	Biscotti	Chocolate Chip Biscotti	$3.50	Jan	Sun	8
28	01-01-2023	08:35:03	2	5	Lower Manhattan	45	$3.00	Tea	Brewed herbal tea	Peppermint Lg	$6.00	Jan	Sun	8
29	01-01-2023	08:35:03	1	5	Lower Manhattan	71	$3.75	Bakery	Pastry	Chocolate Croissant	$3.75	Jan	Sun	8
30	01-01-2023	08:41:57	2	8	Hell's Kitchen	40	$3.75	Coffee	Barista Espresso	Cappuccino	$7.50	Jan	Sun	8
31	01-01-2023	08:52:03	1	8	Hell's Kitchen	38	$3.75	Coffee	Barista Espresso	Latte	$3.75	Jan	Sun	8
32	01-01-2023	08:52:32	1	5	Lower Manhattan	26	$3.00	Coffee	Organic brewed coffee	Brazilian Rg	$3.00	Jan	Sun	8
33	01-01-2023	08:54:33	2	5	Lower Manhattan	27	$3.50	Coffee	Organic brewed coffee	Brazilian Lg	$7.00	Jan	Sun	8
34	01-01-2023	08:56:27	2	5	Lower Manhattan	29	$2.50	Coffee	Gourmet brewed coffee	Columbian Medium Roast Rg	$5.00	Jan	Sun	8
35	01-01-2023	08:57:46	1	8	Hell's Kitchen	54	$2.50	Tea	Brewed Chai tea	Morning Sunrise Chai Rg	$2.50	Jan	Sun	8
36	01-01-2023	08:58:55	1	5	Lower Manhattan	26	$3.00	Coffee	Organic brewed coffee	Brazilian Rg	$3.00	Jan	Sun	8
37	01-01-2023	09:00:12	1	5	Lower Manhattan	55	$4.00	Tea	Brewed Chai tea	Morning Sunrise Chai Lg	$4.00	Jan	Sun	9
38	01-01-2023	09:00:24	2	8	Hell's Kitchen	40	$3.75	Coffee	Barista Espresso	Cappuccino	$7.50	Jan	Sun	9
39	01-01-2023	09:07:13	1	8	Hell's Kitchen	45	$3.00	Tea	Brewed herbal tea	Peppermint Lg	$3.00	Jan	Sun	9
40	01-01-2023	09:08:09	2	8	Hell's Kitchen	45	$3.00	Tea	Brewed herbal tea	Peppermint Lg	$6.00	Jan	Sun	9
41	01-01-2023	09:08:13	1	8	Hell's Kitchen	24	$3.00	Coffee	Drip coffee	Our Old Time Diner Blend Lg	$3.00	Jan	Sun	9

9. LinkedIn:

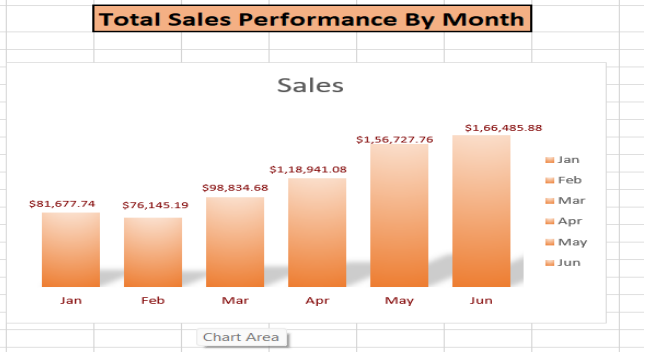
https://www.linkedin.com/feed/update/urn:li:activity:7317216158665842689/

10. LinkedIn Screenshot:

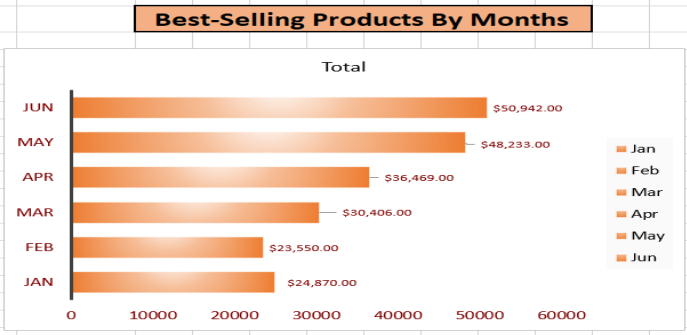


11. Pivot Charts:

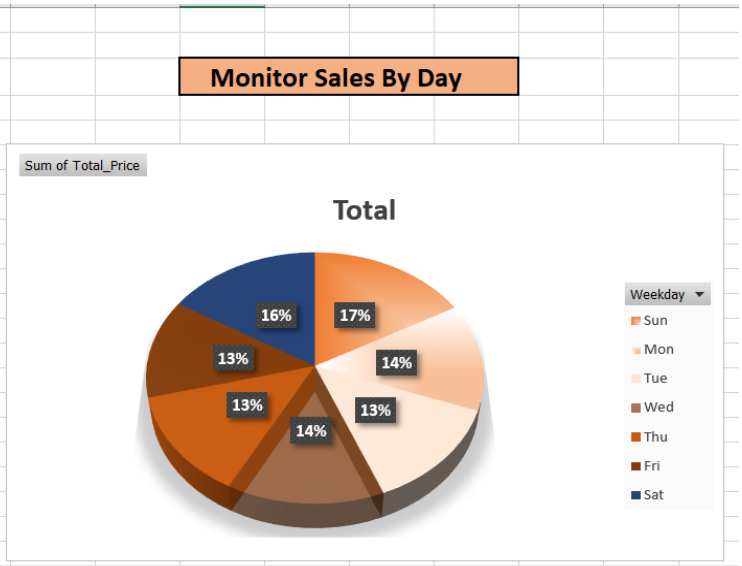
Row Labels	Sum of Total_Price
Jan	81677.74
Feb	76145.19
Mar	98834.68
Apr	118941.08
May	156727.76
Jun	166485.88
Grand Total	698812.33



Product_detail	(All)
Row Labels	Sum of Quantity
Jan	24870
Feb	23550
Mar	30406
Apr	36469
May	48233
Jun	50942
Grand Total	214470

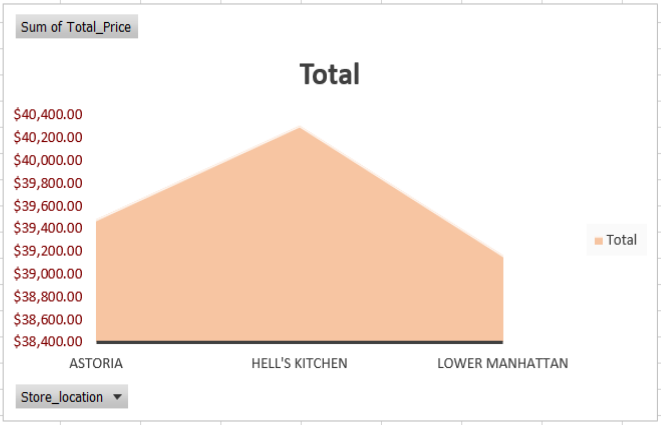


Row Labels	Sum of Total_Price
Sun	20038.74
Mon	16422.8
Tue	15789.23
Wed	16470.64
Thu	15716.61
Fri	15193.23
Sat	19309.83
Grand Total	118941.08



Row Labels	Sum of Total_Price
Astoria	39477.61
Hell's Kitchen	40304.14
Lower Manhattan	39159.33
Grand Total	**118941.08**

Evaluate Sales By Location



Row Labels	Sum of Total_Price	Sum of Quantity
1	630	35
2	594	33
3	354	24
4	429.45	21
5	405	27
6	714	34
7	651.75	33
8	1980	44
9	856.5	38
10	210	21
11	196.9	22
12	241.65	27
13	232.7	26
14	187.95	21
15	231.25	25
16	223.75	25
17	285	30
18	229.95	21
19	140.8	22
20	319.2	42
21	413.23	31
22	1478	739
23	1882.5	753
24	2055	685
25	1628	740
26	2295	765
27	2688	768
28	1434	717
29	1900	760
30	2034	678
31	1705	775
32	2301	767
33	2446.5	699
34	1673.35	683
35	2176.2	702
36	2718.75	725
37	2196	732

Customer Purchase Behavior