

Coffee Shop SQL Sales Analysis

Insights & Recommendations

[Next Slide](#)





About Me

**Moamen Islam Talima -
Business Information Systems
NTI trainee
Google Data Analytics learner.**

2

Why I chose this dataset?

- It represents real-life business data (coffee shop sales), making the analysis practical and relatable.

What we expect by the end of this presentation?

- To provide actionable recommendations that can help improve sales, marketing, and customer experience.



3

Python - Coffee Shop Data Preparation

- Exploratory Data Analysis (EDA)
- → Understanding patterns and distributions
- Data Cleaning
- → Handling missing values, nulls, incomplete records, outliers, duplicates, noisy data, inconsistent entries, scaling, and standardization



Key Features:

- File: Coffe_sales.csv
- Records: 3,547 sales transactions

Columns

- hour_of_day → Hour of purchase (0–23)
- cash_type → Mode of payment (Cash / Card)
- money → Transaction amount (in local currency)
- coffee_name → Type of coffee purchased (e.g., Latte, Americano, Hot Chocolate)
- Time_of_Day → Categorized time of purchase (Morning, Afternoon, Night)
- Weekday → Day of the week (Mon–Sun)

- Month_name → Month of purchase (Jan–Dec)
- Weekdaysort → Numeric weekday order (1 = Mon, ..., 7 = Sun)
- Monthsort → Numeric month order (1 = Jan, ..., 12 = Dec)
- Date → Transaction date (YYYY-MM-DD)
- Time → Exact transaction time (HH:MM:SS)



Display basic information about the dataset

```
print("Dataset Shape:", df.shape)
print("\nColumn Names:")
print(df.columns.tolist())
print("\nFirst 5 rows:")
df.head()
```

→ Dataset Shape: (3547, 11)

Check unique values in categorical columns

→ Unique values in categorical columns:

Cash Type: ['card']

Coffee Names: ['Latte' 'Hot Chocolate' 'Americano' 'Americano with Milk' 'Cocoa' 'Cortado' 'Espresso' 'Cappuccino']

Time of Day: ['Morning' 'Afternoon' 'Night']

Weekdays: ['Fri' 'Sat' 'Sun' 'Mon' 'Tue' 'Wed' 'Thu']

Months: ['Mar' 'Apr' 'May' 'Jun' 'Jul' 'Aug' 'Sep' 'Oct' 'Nov' 'Dec' 'Jan' 'Feb']

Basic dataset information

```
print("Dataset Info")
print(df.info())
```

→ Dataset Info
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3547 entries, 0 to 3546
Data columns (total 11 columns):
Column Non-Null Count Dtype

0 hour_of_day 3547 non-null int64
1 cash_type 3547 non-null object
2 money 3547 non-null float64
3 coffee_name 3547 non-null object
4 Time_of_Day 3547 non-null object
5 Weekday 3547 non-null object
6 Month_name 3547 non-null object
7 Weekdaysort 3547 non-null int64
8 Monthsort 3547 non-null int64
9 Date 3547 non-null object
10 Time 3547 non-null object
dtypes: float64(1), int64(3), object(7)



6

Data Cleaning

Find Missing Values

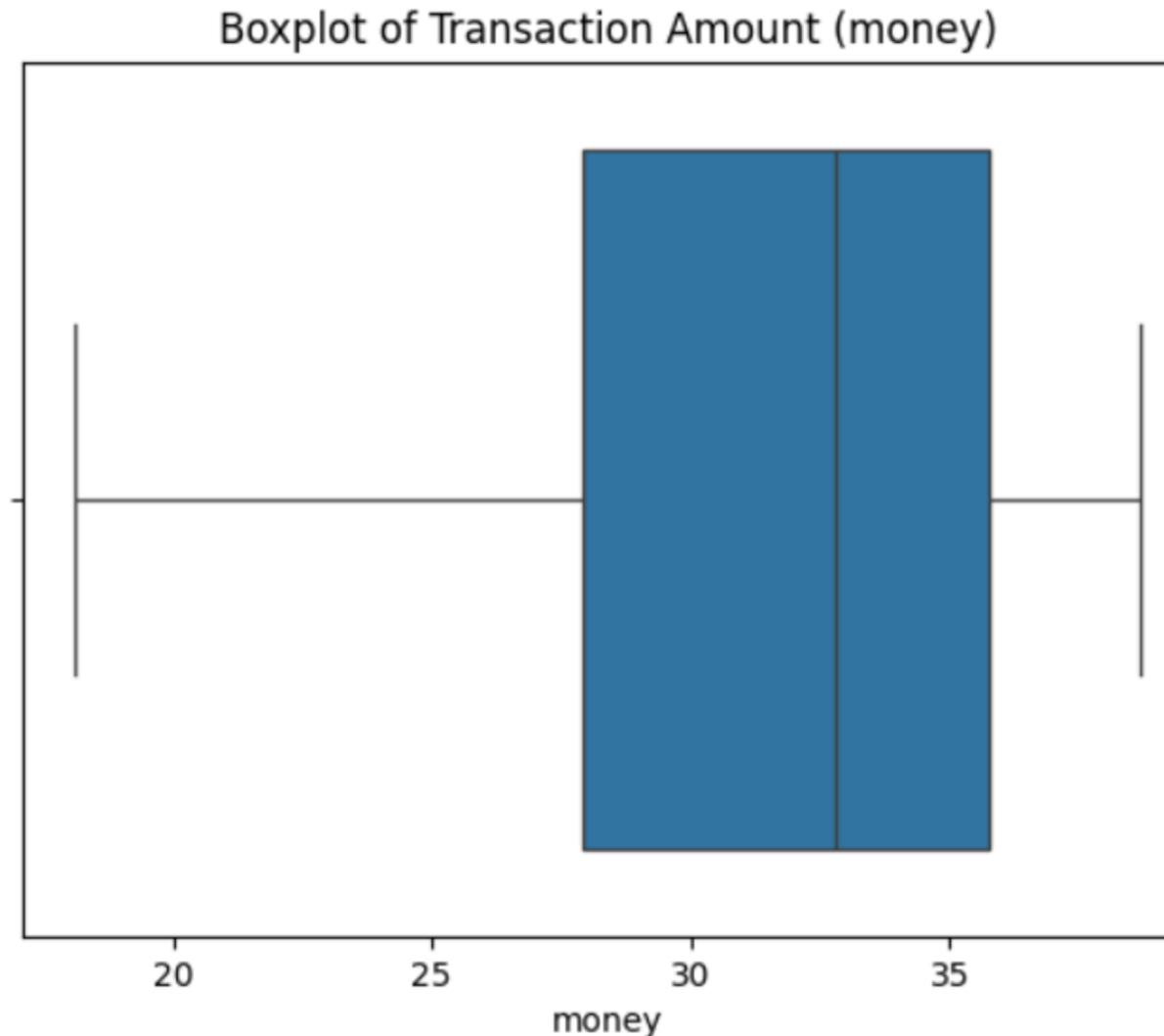
→ Missing Values:

hour_of_day	0
cash_type	0
money	0
coffee_name	0
Time_of_Day	0
Weekday	0
Month_name	0
Weekdaysort	0
Monthsort	0
Date	0
Time	0

dtype: int64

Duplicates & Outliers

Scaling & Outliers in money transactions



→ Number of duplicate rows: 0
Dataset shape after removing duplicates: (3547, 12)

Money column statistics:
Q1: 27.92, Q3: 35.76, IQR: 7.839999999999996
Lower bound: 16.160000000000007, Upper bound: 47.519999999999996
Outliers count: 0

SQL

	hour_of_day	cash_type	money	coffee_name	Time_of_Day	Weekday	Month_name	Weekdaysort	Monthsort
1	10	card	38.7000007629395	Latte	Morning	Fri	Mar	5	3
2	12	card	38.7000007629395	Hot Chocolate	Afternoon	Fri	Mar	5	3
3	12	card	38.7000007629395	Hot Chocolate	Afternoon	Fri	Mar	5	3
4	13	card	28.8999996185303	Americano	Afternoon	Fri	Mar	5	3
5	13	card	38.7000007629395	Latte	Afternoon	Fri	Mar	5	3

Date	Time	Datetime	day_of_week	month	day_of_month	is_weekend	price_category
2024-03-01	10:15:50.520000	2024-03-01 10:15:50.5200000	Friday	3	1	0	High
2024-03-01	12:19:22.539000	2024-03-01 12:19:22.5390000	Friday	3	1	0	High
2024-03-01	12:20:18.089000	2024-03-01 12:20:18.0890000	Friday	3	1	0	High
2024-03-01	13:46:33.006000	2024-03-01 13:46:33.0060000	Friday	3	1	0	Low
2024-03-01	13:48:14.626000	2024-03-01 13:48:14.6260000	Friday	3	1	0	High
2024-03-01	15:39:47.726000	2024-03-01 15:39:47.7260000	Friday	3	1	0	Medium

- Based on the dataset, the analyst (Me) observed that the week starts on Monday, which strongly suggests that these sales belong to a coffee shop located in Europe.
- From this assumption, the seasonal_classification was applied (Spring, Summer, Autumn, Winter), and the local_time context (hour of the day) was also considered in the analysis.



Q1: How much does each drink contribute to total revenue?

Q1.1: What is the revenue per drink?

Q1.2: Which are the Top 3 drinks by revenue?

Q1.3: Are the drinks classified as hot or cold?

```
SELECT
    coffee_name,
    ROUND(SUM(money), 2) AS total_revenue,
    ROUND(100.0 * SUM(money) / (SELECT SUM(money) FROM cleaned_data), 2) AS revenue_percentage
FROM cleaned_data
GROUP BY coffee_name
ORDER BY total_revenue DESC;
```

	coffee_name	total_revenue	revenue_percentage
1	Latte	26875.3	23.94
2	Americano with Milk	24751.12	22.05
3	Cappuccino	17439.14	15.54
4	Americano	14650.26	13.05
5	Hot Chocolate	9933.46	8.85
6	Cocoa	8521.16	7.59
7	Cortado	7384.86	6.58
8	Espresso	2690.28	2.4

Q2: What is the daily average sales per day?

```
SELECT  
    ROUND(AVG(daily_total), 2) AS avg_daily_spending  
FROM (   
    SELECT  
        Date,  
        SUM(money) AS daily_total  
    FROM cleaned_data  
    GROUP BY Date  
) AS daily_sales;
```

	avg_daily_spending
1	294.61



Q3: How many products exist in each price category (Low, Medium, High)?

Q3.1: How many sales happened in each price category?

	price_category	product_count
1	High	4
2	Low	4
3	Medium	5

	price_category	sales_count
1	High	1415
2	Low	1202
3	Medium	930

Q4: Which months (or seasons) drive the most sales?

Meteorological Seasons (Europe)

- Spring → Months 3, 4, 5 (March, April, May)
- Summer → Months 6, 7, 8 (June, July, August)
- Autumn → Months 9, 10, 11 (September, October, November)
- Winter → Months 12, 1, 2 (December, January, February)

	Monthsort	Month_name	month_count
1	3	Mar	494
2	10	Oct	426
3	2	Feb	423
4	9	Sep	344
5	8	Aug	272

```
-- Month frequency
SELECT
    Monthsort,
    Month_name,
    COUNT(*) AS month_count
FROM cleaned_data
GROUP BY Monthsort, Month_name
ORDER BY month_count DESC;
```

Q5: How do sales differ between weekdays and weekends?

Q5.2: What are the peak sales hours during the day (ranked highest to lowest)?

	Day_Type	total_sales	total_revenue
1	Weekday	2658	84175.9987697601
2	Weekend	889	28069.5796718597

	hour_of_day	total_sales	total_revenue
1	10	328	10198.5199298859
2	11	283	8453.09991836548
3	16	278	9031.83978843689
4	9	242	7264.27993583679
5	12	241	7419.61991691589

Recommendations

1. Peak Hours & Staffing

Focus on 9:00 AM –12:00 PM with more staff to reduce waiting time and improve service.

2. Seasonal Inventory Planning

Sales peak in Spring, Autumn, and Winter prepare extra stock of best-selling products.

3. Profitable Products

Invest more in top-selling drinks with stronger marketing and more variety.

4. Weak Products

Remove low-performing items or re-market them to high-end customers.



Recommendations

5. Payment Methods

- Add diverse payment options and promote the ones linked to higher spending.

6. Weekdays vs Weekends

- Weekdays → stronger inventory planning.
- Weekends → use promotions to boost sales

7. Left to be filled by attendees

8. Left to be filled by attendees



End



Thank you