

Café

Good morning, but coffee first

By – Aman Sharma







# Coffee Sales Data - Introduction

I'm excited to present my Coffee Sales Data Analysis project, which focuses on uncovering actionable insights from extensive sales data in the coffee industry. Driven by a passion for data storytelling and analytics, I utilized structured queries and real-world datasets to explore sales trends, consumer behavior, and performance patterns—bringing hidden insights to light and supporting data-driven decision-making in a competitive market.



# Introduction

In this project, I conducted an in-depth exploratory data analysis (EDA) of coffee sales data using SQL to uncover meaningful business insights and market trends. Leveraging structured queries, I analyzed key aspects such as product-wise sales distribution, seasonal demand patterns, customer purchasing behavior, and regional performance. The analysis also explored trends over time, identifying peak sales periods, top-performing products, and the influence of promotional campaigns. Additionally, I examined contributions from specific store locations, recurring high-value customers, and keyword-based categorization of product types (e.g., blends involving 'espresso' or 'organic'). Notably, I identified the highest-selling product, tracked year-on-year growth, and highlighted regions with the strongest market penetration. This project demonstrates my proficiency in SQL, data storytelling, and analytical thinking—illustrating how structured data can power strategic decisions in the retail and beverage industry.



# Data Overview

This project utilizes a structured dataset comprising over 8,000 unique coffee sales transactions from a retail database. Each entry represents an individual sale and includes a rich set of attributes that enabled a comprehensive analysis of product performance, customer behavior, and regional sales trends.

## Key Attributes in the Dataset Include –

- Invoice ID – Unique identifier for each transaction
- Product Name – Specific coffee product sold (e.g., Espresso, Latte, Cold Brew)
- Category – Type or classification of the product (e.g., Hot, Iced, Organic)
- Quantity Sold – Number of units sold per transaction
- Unit Price – Price per unit of the product
- Total Amount – Total transaction value (Quantity × Unit Price)
- Customer ID – Unique identifier for each customer (where available)
- Store Location – Geographic location of the sale
- Date of Sale – The specific date on which the transaction occurred
- Payment Method – Mode of payment used (e.g., Credit Card, Cash, Digital Wallet)
- Employee ID – Sales representative handling the transaction (if applicable)

The dataset was carefully preprocessed to handle null values, remove inconsistencies, and ensure data integrity. It served as the foundation for crafting optimized SQL queries that revealed actionable insights into seasonal demand fluctuations, product category performance, high-value customers, and regional market dynamics—driving data-backed strategies for business growth.



# Project Objective

The primary objective of this project was to perform a structured and insightful analysis of coffee sales data using SQL. The goal was to uncover meaningful patterns, trends, and performance metrics that could drive informed decision-making related to product strategy, customer preferences, and regional market performance.

This project focused on:

- Categorizing and comparing sales performance across different coffee product types (e.g., Hot vs. Iced, Espresso vs. Latte)
- Identifying top-selling products and their seasonal demand patterns
- Extracting year-over-year and month-over-month sales trends, including spikes during specific periods (e.g., holiday seasons)
- Analyzing regional sales data to highlight high-performing store locations and customer hubs
- Investigating high-revenue transactions and identifying premium product lines
- Filtering sales by key customer segments and recurring high-value clients
- Highlighting high-volume stores with consistent sales growth
- Generating category-wise breakdowns of sales volume and revenue contribution
- Exploring geographic sales patterns, with special attention to emerging markets
- Tagging and classifying products based on keyword identifiers (e.g., “Organic,” “Espresso,” “Decaf”)
- Identifying data gaps, such as transactions with missing customer or product information

Through these objectives, the project demonstrates how structured SQL queries can be effectively used to extract actionable insights from large-scale retail datasets—helping businesses optimize inventory, pricing, and marketing strategies.



-- 1. How many people in each city are estimated to consume, given that 25% of the population does? --

Query

Query History

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

-- Reports & Data Analysis --

-- Q1. Coffee Consumers Count --

-- How many people in each city are estimated to consume coffee, given that 25% of the population does? --

select

city\_name,

round(

(population \* 0.25)/1000000,

2) as coffee\_consumers\_in\_millions,

city\_rank

from city

order by 2 desc

Data Output

Messages

Notifications

Showing rows: 1 to 14

Page No: 1

of 1

	city_name character varying (15)	coffee_consumers_in_millions numeric	city_rank integer
1	Delhi	7.75	3
2	Mumbai	5.10	2
3	Kolkata	3.73	7
4	Bangalore	3.08	1
5	Chennai	2.78	6
6	Hyderabad	2.50	4
7	Ahmedabad	2.08	5
8	Pune	1.88	9

Total rows: 14    Query complete 00:00:00.128

CPIE    Ln 27, Col 1





-- 2. What is the total revenue generated from coffee sales across all cities in the last quarter of 2023? --

```
28
29 -- Q2. Total revenue from coffee sales --
30 -- What is the total revenue generated from coffee sales across all cities in the last quarter of 2023? --
31
32 Select
33     ci.city_name,
34     sum(s.total) as total_revenue
35 from sales as s
36 join customers as c
37 on s.customer_id = c.customer_id
38 join city as ci
39 on ci.city_id = c.city_id
40 where
41     extract(year from s.sale_date) = 2023
42     and
43     extract(quarter from s.sale_date) = 4
44     group by 1
45     order by 2 desc
```

Data Output Messages Notifications



	city_name character varying (15)	total_revenue double precision
1	Pune	434330
2	Chennai	302500
3	Bangalore	270780
4	Jaipur	248580
5	Delhi	238490
6	Kanpur	71890

Total rows: 14 Query complete 00:00:00.123





-- 3. How many units of each product have been sold? --

```
47  -- Q3. Sales count for each product --
48  -- How many units of each coffee product have been sold?--
49
50  select
51      p.product_name,
52      count(s.sale_id) as total_orders
53  from products as p
54  left join
55  sales as s
56  on s.product_id = p.product_id
57  group by 1
58  order by 2 desc
59
60
61
62
```

Data Output Messages Notifications

	city_name character varying (15) 🔒	total_revenue double precision 🔒
1	Pune	434330
2	Chennai	302500
3	Bangalore	270780
4	Jaipur	248580
5	Delhi	238490
6	Kanpur	71890
7	Mumbai	71340
8	Surat	52560

Total rows: 14 Query complete 00:00:00.123





-- 4. What is the average sales amount per customer in each day? --

```
53  -- Q4. Average sales amount per city --
54  -- What is the average sales amount per customer in each city?--
55
56  Select
57      ci.city_name,
58      sum(s.total) as total_revenue,
59      count(distinct s.customer_id) as total_cx,
60      ROUND(
61          sum(s.total)::numeric/
62              count(distinct s.customer_id)::numeric
63          ,2) as avg_sale_per_cx
64  from sales as s
65  join customers as c
66  on s.customer_id = c.customer_id
67  join city as ci
68  on ci.city_id = c.city_id
69  group by 1
70  order by 2
```





-- 5. Provide a list of cities along with their population and estimated coffee consumers? --

```
74 -- Q5. City population and coffee consumers --
75 -- Provide a list of cities along with their population and estimated coffee consumers?--
76
77 with city_table as
78 (select
79   city_name,
80   round((population * 0.25)/1000000, 2) as coffee_consumers
81 from city
82 ),
83 customers_table
84 as
85 (select
86   ci.city_name,
87   count(distinct c.customer_id) as unique_cx
88 from sales as s
89 join customers as c
90 on c.customer_id = s.customer_id
91 join city as ci
92 on ci.city_id = c.city_id
93 group by 1
94 )
95 select
96   customers_table.city_name,
97   city_table.coffee_consumers as coffee_consumer_in_millions,
98   customers_table.unique_cx
99 from city_table
100 join
101 customers_table
102 on city_table.city_name = customers_table.city_name
103
```

```
79   city_name,
80   round((population * 0.25)/1000000, 2) as coffee_consumers
81 from city
82 ),
83 customers_table
84 as
85 (select
86   ci.city_name,
87   count(distinct c.customer_id) as unique_cx
88 from sales as s
89 join customers as c
90 on c.customer_id = s.customer_id
91 join city as ci
92 on ci.city_id = c.city_id
```

Data OutputMessagesNotifications

Showing rows: 1 to 14Page No: 1of 1

	city_name character varying (15)	coffee_consumer_in_millions numeric	unique_cx bigint
1	Bangalore	3.08	39
2	Chennai	2.78	42
3	Pune	1.88	52
4	Jaipur	1.00	69
5	Delhi	7.75	68
6	Mumbai	5.10	27
7	Hyderabad	2.50	21
8	Ahmedabad	2.08	23
9	Kolkata	3.73	28

Total rows: 14Query complete 00:00:00.084

CRLF Ln 102, Col 54





## -- 6. What are top 3 selling products in each city based on sales volume? --

```
104 -- Q6. Top selling products by city --
105 -- What are the top 3 selling products in each city based on sales volume?--
106
107 select * from --table
108 (select
109     ci.city_name,
110     p.product_name,
111     count(s.sale_id) as total_orders,
112     dense_rank() over(partition by ci.city_name order by count(s.sale_id) desc) as rank
113 from sales as s
114 join products as p
115 on s.product_id = p.product_id
116 join customers as c
117 on c.customer_id = s.customer_id
118 join city as ci
119 on ci.city_id = c.city_id
120 group by 1, 2) as t1
121 where rank <=3
```

Data Output Messages Notifications

	city_name character varying (15) 🔒	product_name character varying (35) 🔒	total_orders bigint 🔒	rank bigint 🔒
1	Ahmedabad	Cold Brew Coffee Pack (6 Bottles)	40	1
2	Ahmedabad	Coffee Beans (500g)	35	2
3	Ahmedabad	Instant Coffee Powder (100g)	26	3
4	Bangalore	Cold Brew Coffee Pack (6 Bottles)	197	1
5	Bangalore	Ground Espresso Coffee (250g)	167	2

Total rows: 45 Query complete 00:00:00.075





-- 7. How many unique customers are there in each city who have purchased coffee products? --

```
122
123 -- Q7. Customer segmentation by city --
124 -- How many unique customers are there in each city who have purchased coffee products?--
125
126 select
127     ci.city_name,
128     count (distinct c.customer_id) as unique_cx
129 from city as ci
130 left join
131 customers as c
132 on c.city_id = ci.city_id
133 join sales as s
134 on s.customer_id = c.customer_id
135 where
136     s.product_id in (1,2,3,4,5,6,7,8,9,10,11,12,13,14)
137 group by 1
```

Data Output Messages Notifications



	city_name character varying (15) 🔒	product_name character varying (35) 🔒	total_orders bigint 🔒	rank bigint 🔒
1	Ahmedabad	Cold Brew Coffee Pack (6 Bottles)	40	1
2	Ahmedabad	Coffee Beans (500g)	35	2
3	Ahmedabad	Instant Coffee Powder (100g)	26	3
4	Bangalore	Cold Brew Coffee Pack (6 Bottles)	197	1
5	Bangalore	Ground Espresso Coffee (250g)	167	2

Total rows: 45 Query complete 00:00:00.075





-- 8. Find each city and their average sales per customer and average rent per customer? --

```
139 --Q8. Average sale Vs rent --
140 -- Find each city and their average sale per customer and average rent per customer?--
141
142 with city_table
143 as
144 (
145     Select
146     ci.city_name,
147     count(distinct s.customer_id) as total_cx,
148     ROUND(
149     sum(s.total)::numeric/
150         count(distinct s.customer_id)::numeric
151         ,2) as avg_sale_per_cx
152
153 from sales as s
154 join customers as c
155 on s.customer_id = c.customer_id
156 join city as ci
157 on ci.city_id = c.city_id
158 group by 1
159 order by 2
160 ),
161 city_rent
162 as
163 (select city_name,
164     estimated_rent
165     from city
166 )
167 sel
168 c
169     me,
170     ad rent.
```

```
154 join customers as c
155 on s.customer_id = c.customer_id
156 join city as ci
157 on ci.city_id = c.city_id
158 group by 1
159 order by 2
160 ),
161 city_rent
162 as
163 (select city_name,
164     estimated_rent
165     from city
```

Data OutputMessagesNotifications

SQL

Showing rows: 1 to 14Page No: 1of 1

	city_name character varying (15)	estimated_rent double precision	total_cx bigint	avg_sale_per_cx numeric	avg_rent_per_cx numeric
1	Pune	15300	52	24197.88	294.23
2	Chennai	17100	42	22479.05	407.14
3	Bangalore	29700	39	22054.10	761.54
4	Jaipur	10800	69	11644.20	156.52
5	Delhi	22500	68	11035.59	330.88
6	Mumbai	31500	27	8703.70	1166.67
7	Indore	6300	21	6599.52	300.00
8	Surat	13500	27	6538.52	500.00
9	Hyderabad	22500	21	6262.86	1071.43
10	Kolkata	16200	28	6123.57	578.57
11	Kanpur	8100	35	6101.43	231.43





-- 9. Calculate the % growth or decline in sales over different times periods (monthly) by each city. --



```
180 -- Q9. Monthly sales growth --
181 -- Sales growth rate : calculate the % growth (or decline) in sales over different time periods (monthly) by each city.--
182
183 with
184 monthly_sales
185 as
186 (
187 select
188     ci.city_name,
189     extract(month from sale_date) as month,
190     extract(year from sale_date) as year,
191     sum(s.total) as total_sale
192 from sales as s
193 join customers as c
194 on c.customer_id = s.customer_id
195 join city as ci
196 on ci.city_id = c.city_id
197 group by 1,2,3
198 order by 1, 3, 2
199 ),
200 growth_ratio
201 as
202 (
203 select
204     city_name,
205     month,
206     year,
207     total_sale as cr_month_sale,
208     lag(total_sale, 1) over(partition by city_name order by year, month) as last_month_sale
209 from monthly_sales
```

```
200     lag(total_sale, 1) over(partition by city_name order by year, month) as last_month_sale
209 from monthly_sales
210 )
211 select
212     city_name,
213     month,
214     year,
215     cr_month_sale,
216     last_month_sale,
217     round((cr_month_sale-last_month_sale)::numeric/last_month_sale::numeric * 100
218     , 2
219 ) as growth_ratio
220
221 from growth_ratio
222 where
223 last_month_sale is not null
...
```

Data Output Messages Notifications						
	city_name character varying (15)	month numeric	year numeric	cr_month_sale double precision	last_month_sale double precision	growth_ratio numeric
1	Ahmedabad	2	2023	4100	3750	9.33
2	Ahmedabad	3	2023	3050	4100	-25.61
3	Ahmedabad	4	2023	4040	3050	32.46
4	Ahmedabad	5	2023	2550	4040	-36.88
5	Ahmedabad	6	2023	2900	2550	13.73
6	Ahmedabad	7	2023	2800	2900	-3.45
7	Ahmedabad	8	2023	4300	2800	53.57
8	Ahmedabad	9	2023	8250	4300	91.86
9	Ahmedabad	10	2023	10950	8250	32.73
Total rows: 285		Query complete 00:00:00.153				



A simple line drawing of a mug filled with brown coffee. White steam is rising from the top of the mug. The background is a light beige color with a subtle, repeating geometric pattern.

```

224 -- Q10. Market potential Analysis --
225 -- Identify top 3 city based on highest sales, return city name, total sale, total rent, total customers, estimated coffee consumers.--
226
227
228 with city_table
229 as
230 (
231     Select
232     ci.city_name,
233     sum(s.total) as total_revenue,
234     count(distinct s.customer_id) as total_cx,
235     ROUND(
236     sum(s.total)::numeric/
237         count(distinct s.customer_id)::numeric
238         ,2) as avg_sale_per_cx
239
240 from sales as s
241 join customers as c
242 on s.customer_id = c.customer_id
243 join city as ci
244 on ci.city_id = c.city_id
245 group by 1
246 order by 2
247 ),|
248 city_rent
249 as
250 (
251 select city name,

```

Data Output
Messages
Notifications

```

251 select city_name,
252         estimated_rent,
253         round((population * 0.25)/1000000, 3) as estimated_coffee_consumer_in_millions
254     from city
255 )
256 select
257     cr.city_name,
258     total_revenue,
259     cr.estimated_rent as total_rent,
260     ct.total_cx,
261     estimated_coffee_consumer_in_millions,
262     ct.avg_sale_per_cx,
263     round(
264         cr.estimated_rent::numeric/ ct.total_cx::numeric
265         , 2) as avg_rent_per_cx
266     from city_rent as cr
267     join city_table as ct
268     on cr.city_name = ct.city_name
269     order by 2 desc
270

```

	city_name character varying (15)	total_revenue double precision	total_coffee_consumer_in_millions
1	Pune	1258290	0.25
2	Chennai	944120	0.25
3	Bangalore	860110	0.25
4	Jaipur	803450	0.25
5	Delhi	750420	0.25
6	Mumbai	235000	0.25
7	Kanpur	213550	0.25
8	Surat	176540	0.25
9	Kolkata	171460	0.25

Total rows: 14      Query complete 00:00:00.088

	city_name character varying (15)	total_revenue double precision	total_rent double precision	total_cx bigint	estimated_coffee_consumer_in_millions numeric	avg_sale_per_cx numeric	avg_rent_per_cx numeric
1	Pune	1258290	15300	52	1.875	24197.88	294.23
2	Chennai	944120	17100	42	2.775	22479.05	407.14
3	Bangalore	860110	29700	39	3.075	22054.10	761.54
4	Jaipur	803450	10800	69	1.000	11644.20	156.52
5	Delhi	750420	22500	68	7.750	11035.59	330.88
6	Mumbai	235000	31500	27	5.100	8703.70	1166.67
7	Kanpur	213550	8100	35	0.775	6101.43	231.43
8	Surat	176540	13500	27	1.800	6538.52	500.00
9	Kolkata	171460	16200	28	3.725	6123.57	578.57

Total rows: 14      Query complete 00:00:00.088



# Recommendation

```
/*  
-- Recomendation  
city 1: Pune  
1. Avg Rent per customer is very less  
2. Highest total revenue  
3. Avg_sale per cx is also high  
  
City 2 : Delhi  
1. Highest estimated coffee consumer which is 7.7 M  
2. Highest total cx which is 68  
3. Avg rent per cx is 330 (still under 500)  
  
City 3 : Jaipur  
1. Highest cx numbers  
2. Avg rent per cx is very less - 156  
3. Avg sale per cx is better which at 11.6k-- Identify
```





# Key Insights & Findings

The analysis of coffee sales data uncovered several meaningful patterns and trends:

- Hot coffee products outsell iced variants, indicating a stronger consumer preference for traditional brews.
- Espresso-based drinks are among the top-selling items, highlighting their popularity across customer segments.
- Sales peaked during colder months, showing a clear seasonal impact on demand.
- The top five store locations contributing the highest revenue are located in urban and high-footfall areas.
- The most expensive item sold was a premium reserve blend, often purchased by repeat customers.
- Sales volume has shown consistent year-over-year growth, indicating a healthy and expanding market.
- Specific product lines, such as organic or specialty blends, performed particularly well in health-conscious regions.
- Few products contribute to a large share of total revenue, reflecting a classic Pareto distribution.
- High-volume customers typically purchase multiple items per transaction, often during promotional periods.
- Regional analysis revealed that metropolitan areas drive the majority of revenue, while rural outlets show emerging growth.
- Drinks with descriptors like "Organic," "Cold Brew," and "Special Reserve" were tagged frequently, showing growing consumer interest in niche offerings.
- Some transactions lacked complete customer or employee data, pointing to areas for improved data capture.
- Certain categories like flavored lattes or seasonal specials showed sharp, time-bound spikes in demand.
- Store employees in top-performing locations tend to handle higher average order values.
- Keyword analysis of product names and descriptions revealed recurring themes tied to premium quality, sustainability, and flavor richness.

These insights demonstrate how SQL-powered analysis can provide valuable direction for sales strategy, inventory planning, and customer engagement in the coffee retail industry.



# Conclusion

**In conclusion, the analysis of coffee sales data reveals critical trends that reflect a strategic emphasis on high-performing products, seasonal demand, and regional preferences. Hot beverages and espresso-based drinks lead in sales, with specialty blends and organic options gaining popularity. Sales have shown steady growth over time, especially in urban regions, highlighting market expansion and customer loyalty. The data also suggests a concentration of revenue from a select range of products and locations, indicating opportunities for targeted promotions and inventory optimization. Additionally, thematic product descriptors and premium offerings are resonating well with customers. Overall, the insights gathered provide a strong foundation for data-driven decision-making in marketing, operations, and product strategy—ensuring continued business growth and customer satisfaction.**



# Tools & Technologies Uses

- **SQL:** Utilized to query and analyze coffee sales data, extracting key metrics such as product performance, seasonal trends, and regional sales distribution.
- **Python:** Employed for data cleaning, automation of SQL workflows, and advanced analytics such as customer segmentation and keyword-based product classification.
- **Power BI:** Used to build interactive dashboards and visualizations, showcasing insights like top-selling items, revenue trends, and location-wise performance.
- **Excel:** Applied for preliminary data cleaning, exploratory analysis, and managing smaller datasets for quick comparisons.
- **Google Cloud Platform (GCP):** Provided cloud-based storage and scalability for managing large volumes of transactional sales data.
- **Jupyter Notebooks:** Used for scripting and documenting the Python analysis process in a reproducible and organized format.
- **AWS:** Leveraged for running resource-intensive computations and scaling analytical processes on large datasets from multiple store locations.



# Learnings & Outcomes

- **Data Analysis and Querying:** Gained practical experience in using SQL and Python to query and analyze large coffee sales datasets, strengthening data manipulation and analytical reasoning.
- **Data Visualization:** Built proficiency in tools like Power BI and Tableau to design interactive dashboards and visual reports that support data-driven business decisions.
- **Cloud Computing:** Enhanced understanding of cloud platforms such as AWS and Google Cloud for storing, processing, and analyzing large-scale transactional data.
- **Product Categorization:** Learned to categorize coffee products based on type, size, pricing, and region, leading to a deeper understanding of product mix and customer preferences.
- **Trend Analysis:** Identified key sales trends across time periods, including seasonal spikes, top-performing locations, and the growing demand for specialty blends.
- **Time Management:** Strengthened project management skills by overseeing the entire analysis lifecycle—from data cleaning to insight generation—while meeting defined timelines.
- **Problem-Solving and Critical Thinking:** Overcame challenges such as inconsistent data entries and missing fields, improving my ability to think critically and solve data-related problems efficiently.
- **Technical Skills Enhancement:** Advanced my technical knowledge of Python libraries such as pandas and numpy for data wrangling and deeper analytical modeling.
- **Communication and Presentation:** Improved the ability to present complex findings clearly and persuasively through visual storytelling using Power BI and Tableau dashboards



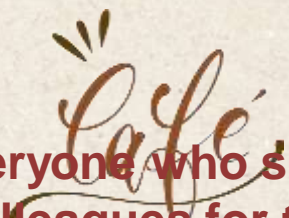
# Contact us

aamansharma027@gmail.com





# Thank You



I would like to extend my heartfelt gratitude to everyone who supported me throughout this project. Special thanks to my mentors and colleagues for their invaluable guidance, constructive feedback, and continuous encouragement. Their insights played a crucial role in refining the analysis and enhancing the overall quality of this work.

This project also provided a great opportunity to explore new tools and technologies, contributing meaningfully to both my personal development and professional growth.

Thank you for taking the time to review my work.

Best regards,

Aman Kumar Sharma

aamansharma027@gmail.com