



PRESENTED BY JAYESH

Python + MySQL

SQL Project

Project Overview

The goal of this project is to analyze a dataset using Python and MySQL, focusing on cleaning raw data, querying relevant insights from a database, and generating actionable business insights.

Table Structure

- invoice_id: Unique identifier for each transaction.
- branch, City: Details about the store location.
- category: Product category.
- unit_price, quantity: Product pricing and sales quantities.
- date, time: Transaction timestamps.
- payment_method: Method of payment (e.g., Ewallet, Cash, Credit card).
- rating: Customer ratings.
- profit_margin: Profit margin percentage.

Q1. Find the top 5 total number of transactions for each branch.

Query :

```
select branch,  
count(*) as no_transaction  
from walmart_tbl  
group by branch  
order by no_transaction desc  
limit 5
```

Output :

branch	no_transaction
WALM058	2151
WALM009	2115
WALM030	2061
WALM069	1998
WALM074	1890

Q2. Calculate the total revenue for each city, include only cities where total revenue upto 500.

Query :

```
select city,sum(total) as total_revenue
from walmart_tbl
group by city
having SUM(total) > 500;
```

Output :

city	total_revenue
San Antonio	224555.039999999992
Harlingen	73177.200000000001
Haltom City	92091.690000000006
Bedford	75415.410000000002
Irving	56133.9900000000005
Denton	82128.51
Cleburne	106022.430000000002
Canyon	74289.06
Grapevine	109804.139999999997

Q3. Find the earliest and latest transaction dates for each payment method.

Query :

```
SELECT payment_method, MIN(date) AS earliest_date,  
MAX(date) AS latest_date  
FROM walmart_tbl  
GROUP BY payment_method;
```

Output :

payment_method	earliest_date	latest_date
Ewallet	01/01/19	31/12/23
Cash	01/01/19	31/12/23
Credit card	01/01/19	31/12/23

Q4. Find the total profit margin for all transactions.

Query :

```
SELECT SUM(profit_margin) AS total_profit_margin  
FROM walmart_tbl;
```

Output :

total_profit_margin
35327.07000003194

Q.5 Retrieve all records where the payment method is not "Credit card".

Query :

```
SELECT *  
FROM walmart_tbl  
where payment_method != "Credit card";
```

Output :

invoice_id	branch	city	category	unit_price	quantity	date	time	payment_method	rating
1	WALM003	San Antonio	Health and beauty	74.69	7	05/01/19	13:08:00	Ewallet	9.1
2	WALM048	Harlingen	Electronic accessories	15.28	5	08/03/19	10:29:00	Cash	9.6
4	WALM064	Bedford	Health and beauty	58.22	8	27/01/19	20:33:00	Ewallet	8.4
5	WALM013	Irving	Sports and travel	86.31	7	08/02/19	10:37:00	Ewallet	5.3
6	WALM026	Denton	Electronic accessories	85.39	7	25/03/19	18:30:00	Ewallet	4.1
7	WALM088	Cleburne	Electronic accessories	68.84	6	25/02/19	14:36:00	Ewallet	5.8
8	WALM100	Canyon	Home and lifestyle	73.56	10	24/02/19	11:38:00	Ewallet	8
11	WALM013	Irving	Fashion accessories	14.48	4	06/02/19	18:07:00	Ewallet	4.5
12	WALM035	San Angelo	Electronic accessories	25.51	4	09/03/19	17:03:00	Cash	6.8

Q6. Calculate the total revenue based on city 'San Antonio'.

Query :

```
SELECT city, SUM(total) AS total_revenue  
FROM walmart_tbl  
GROUP BY city  
having city = 'San Antonio';
```

Output :

city	total_revenue
San Antonio	224555.039999999992

Q.7 Find the city with the highest total revenue.

Query :

```
SELECT city, max(total) AS max_revenue  
FROM walmart_tbl  
GROUP BY city  
order by max_revenue desc  
limit 1;
```

Output :

city	max_revenue
McKinney	993

Q.8 For each branch,
find the top category by
total revenue.

Query :

```
select *
from (
    SELECT branch,category,total,
    row_number() over(partition by branch order by total desc) as ranked
    from walmart_tbl
    ) as row_data
where ranked = 1;
```

Output :

branch	category	total	ranked
WALM001	Sports and travel	766	1
WALM002	Electronic accessories	524.7	1
WALM003	Home and lifestyle	906.5	1
WALM004	Electronic accessories	609.56	1
WALM005	Home and lifestyle	973.8	1
WALM006	Sports and travel	641.9	1
WALM007	Food and beverages	894.80000000000001	1
WALM008	Home and lifestyle	973.7	1
WALM009	Electronic accessories	680.31000000000001	1
WALM010	Food and beverages	647.68	1

Q.9 Find the ranks of all city with the highest total revenue.

Query :

```
SELECT city, sum(total),  
rank() over(order by sum(total) desc) as ranked_city  
from walmart_tbl  
group by city;
```

Output :

city	sum(total)	ranked_city
Weslaco	417166.109999999975	1
Waxahachie	366329.96999999999	2
Plano	231195.05999999997	3
San Antonio	224555.03999999992	4
Port Arthur	220719.32999999993	5
Richardson	220145.39999999999	6
Rockwall	216699.29999999998	7
Round Rock	209946.06000000006	8
Schertz	207858.87000000014	9
San Marcos	199120.58999999994	10

Q.10 Find the top 3 branches with the highest average rating.

Query :

```
SELECT branch, avg(rating) as avg_rating
from walmart_tbl
group by branch
order by avg_rating desc
limit 3;
```

Output :

branch	avg_rating
WALM004	7.001666666666666
WALM076	6.812162162162164
WALM059	6.732911392405057

Thank You!

Gmail : jayesh281103@gmail.com

Linkdin : www.linkedin.com/in/jayesh2803/

Github : github.com/jayeshpawar28