**Credit Card Transaction of India**

I downloaded the dataset of credit\_card\_transaction from Kaggle (https://www.kaggle.com/datasets/thedevastator/analyzing-credit-card-spending-habits-in-india)

I did some modifications while importing the data in SQL Server Management Studio 19.

* Change the data type of transaction\_date from nvarchar to datetime.
* Change the data type of transaction\_id from integer to nvarchar.
* There are 7 columns in the Credit Card Transaction (transaction\_id, city, transaction\_date, card\_type, exp\_type, gender, amount)

I write some basic queries to know about the given dataset and these are my findings-

* Date Ranges between - 2013-10-04 (October 2013) and 2015-05-26. (May 2015)
* There were 4 types of Card (Silver, Signature, Gold, Platinum)
* There were 6 types of Expense Type (Entertainment, Food, Bills, Fuel, Travel, Grocery)
* We have data of 986 different cities from India

I solved listed problem related to the given dataset

**1- Write a query to print top 5 cities with highest spends and their percentage contribution of total credit card spends?**

WITH cte1 AS

(

SELECT city, sum(amount) AS total\_spent

from credit\_card\_transactions

GROUP BY city)

,total\_amount AS

( SELECT sum(CAST(amount AS BIGINT)) AS total\_amount

FROM credit\_card\_transactions)

SELECT TOP 5 \*, ROUND((total\_spent/total\_amount) \* 100 ,2) AS percentage\_contribution

FROM cte1 INNER JOIN total\_amount

ON 1=1

ORDER BY total\_spent DESC

**2- Write a query to print highest spend month and amount spent in that month for each card type?**

WITH cte AS

(

SELECT card\_type, DATEPART(Year, transaction\_date) as yr, DATEPART(Month, transaction\_date) as mn, sum(amount) AS total\_spent

FROM credit\_card\_transactions

GROUP BY card\_type, DATEPART(Year, transaction\_date), DATEPART(Month, transaction\_date) )

SELECT \* FROM (SELECT \*, DENSE\_RANK() OVER(PARTITION BY card\_type ORDER BY total\_spent DESC) AS rnk

FROM cte) a WHERE rnk =1

**3- Write a query to print the transaction details (all columns from the table) for each card type when it reaches a cumulative of 100000 total spent? (we have 4 rows in the output for each card type)**

WITH cte as

(

SELECT \*, sum(amount) OVER(PARTITION BY card\_type ORDER BY transaction\_date, transaction\_id ASC) AS total\_spent

FROM credit\_card\_transactions )

, cte2 AS

(

SELECT \*, RANK() OVER(PARTITION BY card\_type ORDER BY total\_spent ASC) AS rnk

FROM cte

WHERE total\_spent > 1000000 )

SELECT transaction\_id, card\_type, total\_spent

FROM cte2

WHERE rnk=1

**4- Write a query to find city which had lowest percentage spend for gold card type?**

WITH cte AS

( SELECT city, card\_type, sum(amount) as total\_spent

FROM credit\_card\_transactions

WHERE card\_type = 'Gold'

GROUP BY city, card\_type )

, cte2 AS

( SELECT sum(amount) as total\_amount

FROM credit\_card\_transactions

WHERE card\_type = 'Gold' )

, cte3 AS(

SELECT c1.city, c1.card\_type, c1.total\_spent, c2.total\_amount, (c1.total\_spent/c2.total\_amount)\*100 as contribution

FROM cte as c1 INNER JOIN cte2 as c2

ON 1=1 )

, cte4 AS (

SELECT \*, RANK() OVER(ORDER BY contribution ) as rnk

FROM cte3 )

SELECT city, total\_spent, contribution

FROM cte4

WHERE rnk =1

**5- Write a query to print 3 columns: city, highest\_expense\_type , lowest\_expense\_type? (example format : Delhi , bills, Fuel)**

WITH cte as

(

SELECT city, exp\_type, SUM(amount) as spent

FROM credit\_card\_transactions

GROUP BY city, exp\_type )

, cte2 AS

( SELECT city, exp\_type, spent, RANK() OVER(PARTITION BY city ORDER BY spent DESC) max\_spent,

RANK() OVER(PARTITION BY city ORDER BY spent) min\_spent

FROM cte )

SELECT city, MAX(CASE WHEN max\_spent = 1 THEN exp\_type END) AS highest\_expense\_type,

MAX(CASE WHEN min\_spent = 1 THEN exp\_type END) AS lowest\_expense\_type

FROM cte2

GROUP BY city

**6- Write a query to find percentage contribution of spends by females for each expense?**

WITH cte AS

(

SELECT exp\_type,gender, SUM(amount) as spent

FROM credit\_card\_transactions

GROUP BY exp\_type, gender )

, cte2 AS

(

SELECT exp\_type, gender, spent, SUM(spent) OVER(PARTITION BY exp\_type) as total\_spent

FROM cte )

SELECT exp\_type, gender, ROUND((spent/total\_spent) \* 100 ,2) AS female\_contribution

FROM cte2

WHERE gender = 'F'

**7- Which card and expense type combination saw highest month over month growth in Jan-2014?**

WITH cte AS

( SELECT card\_type, exp\_type, DATEPART(Year, transaction\_date) as yr, DATEPART(Month, transaction\_date) as mn, sum(amount) AS curr\_month\_spent

FROM credit\_card\_transactions

GROUP BY card\_type, exp\_type, DATEPART(Year, transaction\_date), DATEPART(Month, transaction\_date) )

, cte2 AS

( SELECT \*, LAG(curr\_month\_spent,1,0) OVER(PARTITION BY card\_type, exp\_type ORDER BY yr, mn) prev\_month\_spent

FROM cte )

SELECT TOP 1 \*, (curr\_month\_spent-prev\_month\_spent ) / prev\_month\_spent AS mon\_growth

FROM cte2

WHERE prev\_month\_spent IS NOT NULL AND yr = 2014 AND mn =1

ORDER BY mon\_growth DESC

**8- During weekends which city has highest total spend to total no of transactions ratio?**

SELECT TOP 1 city, (SUM(amount)/COUNT(transaction\_id)) as ratio

FROM credit\_card\_transactions

WHERE DATEPART(weekday, transaction\_date) IN (1,7)

-- WHERE DATENAME(weekday, transaction\_date) IN ('Saturday','Sunday')

--(Slower that the first where condition as filteration works faster with integer value)

GROUP BY city

ORDER BY ratio DESC

**10- Which city took least number of days to reach its 500th transaction after the first transaction in that city?**

WITH cte AS

( SELECT city, transaction\_date, ROW\_NUMBER() OVER (PARTITION BY city ORDER BY transaction\_date, transaction\_id) as rnk

FROM credit\_card\_transactions )

, cte2 AS

( SELECT city, MIN(transaction\_date) as first\_transaction, MAX(transaction\_date) as last\_transaction

FROM cte

WHERE rnk =1 OR rnk=500

GROUP BY city

HAVING count(rnk) =2 )

, cte3 AS

( SELECT \*, DATEDIFF(day, first\_transaction, last\_transaction) as number\_of\_days

FROM cte2 )

, cte4 AS

( SELECT \*, RANK() OVER(ORDER BY number\_of\_days) as rnk

FROM cte3 )

SELECT \* FROM cte4

WHERE rnk =1

once you are done with this create a github repo to put that link in your resume. Some example github links:

https://github.com/ptyadana/SQL-Data-Analysis-and-Visualization-Projects/tree/master/Advanced%20SQL%20for%20Application%20Development

https://github.com/AlexTheAnalyst/PortfolioProjects/blob/main/COVID%20Portfolio%20Project%20-%20Data%20Exploration.sql