Credit Card Spending habits in India (Using SQL)

CREATED BY- AKHILESH CHATURVEDI

Project Description:

Objective: To present a detailed analysis of a public dataset on spending habits of Indian credit card users. It is basically an Exploratory Data Analysis (EDA) to understand spending habits of credit card users.

Dataset: This data is taken from Kaggle. Contains spending data of a company from Oct 2013 to May 2015

Tech Stack Used: **MySQL Workbench** is used as the database in which csv file is imported to create the database. The queries along with the output are presented in the report. **MS Excel** is used as the visualisation and analysis of the output. **MS Power Point** is used to create the report.

The Dataset: Attributes

1. Index: having unique and distinct 26,053 records.
2. City: having records of 986 Indian cities
3. Date: 2013 to 2015
4. Card Type: Gold, Platinum, Signature, Silver.
5. Expenditure Type: Bills, Entertainment, Food, Fuel and Grocery.
6. Gender: Female and Male
7. Amount

During weekend which city has highest total spend to total number of transactions ratio.

	city	total_amt	total_no_of_transc	transc_ratio
▶	Sonepur	299905	1	299905.0000
	Lanka	298960	1	298960.0000

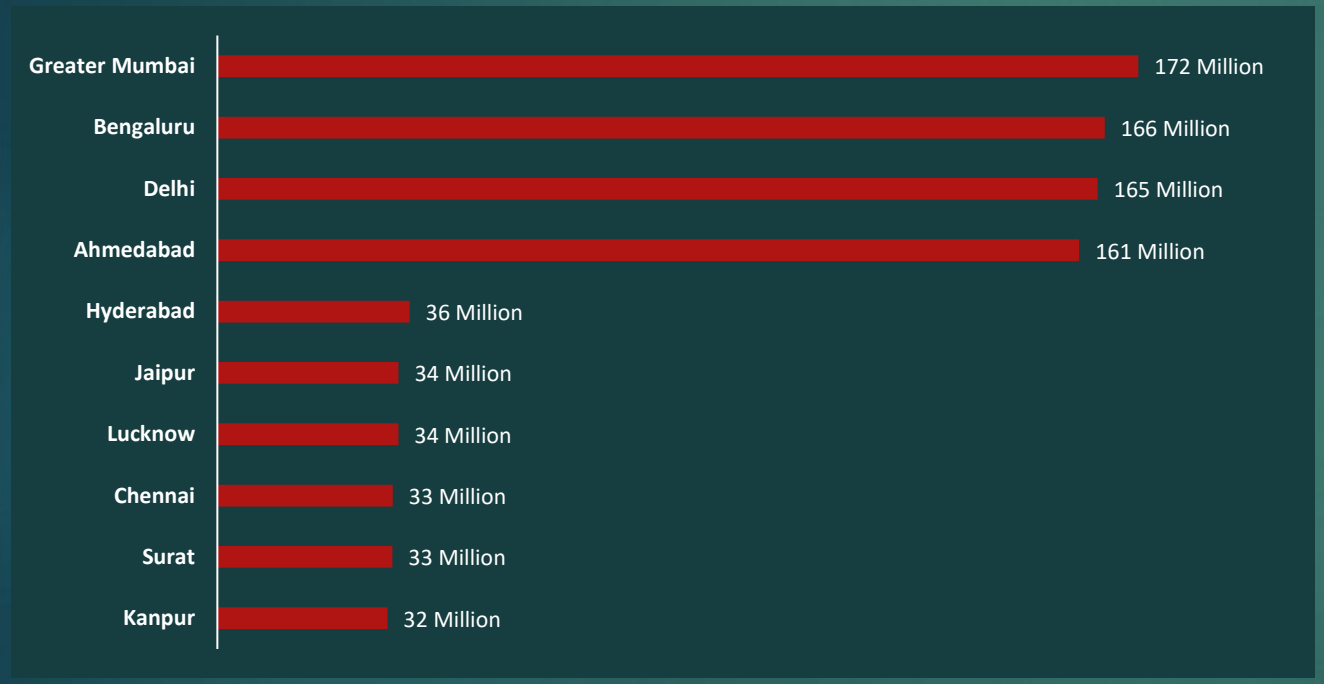
Sonepur, city in Odisha, has highest total spend to total number of transactions ratio, during Weekend.

Because there is only one transactions happened. And in metros there were thousands of transactions.

Here is the **SQL Query** to pull data from SQL Server.

```sql
select city,
    sum(amount) as total_amt,
    count(*) as total_no_of_transc,
    sum(amount) / count(*) as transc_ratio
from credit_in
where dayofweek(date) in (7,1)
group by city
order by transc_ratio desc
limit 2;
```

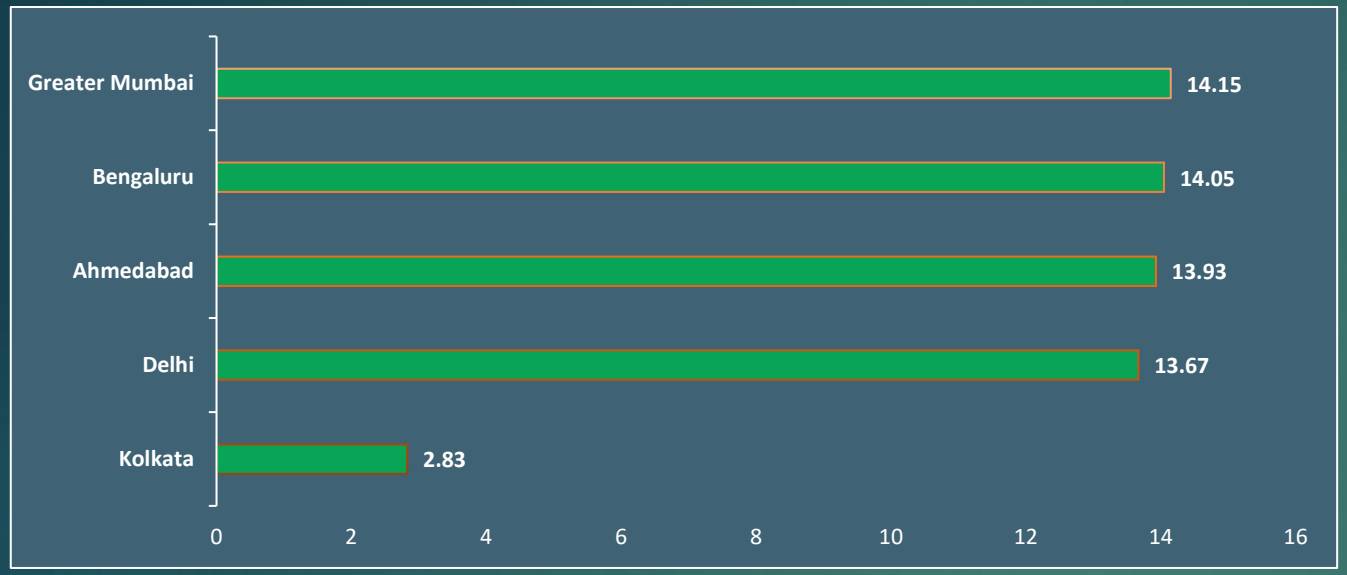
In previous query, if we sort the result according to the total amount, then Greater Mumbai is highest spending city. And the total number of transactions is 1017, during Weekend. (In case of Sonepur, total transaction = 1)



Result Grid | Filter Rows: | Export: | Wrap

city	total_amt	total_no_of_transc	transc_ratio
Greater Mumbai	172418991	1017	169536.8643
Bengaluru	166140238	1050	158228.7981
Delhi	164810553	983	167660.7864
Ahmedabad	161356124	1036	155749.1544
Hyderabad	35996951	242	148747.7314
Jaipur	33909676	232	146162.3966
Lucknow	33877673	217	156118.3088
Chennai	32838233	223	147256.6502
Surat	32766659	211	155292.2227
Kanpur	31834380	215	148066.8837

Top 5 cities with highest spends and their percentage contribution ?



Result Grid | Filter Rows: | Export:

city	SpentCityWise	pct_citywise
Greater Mumbai	576751476	14.15
Bengaluru	572326739	14.05
Ahmedabad	567794310	13.93
Delhi	556929212	13.67
Kolkata	115466943	2.83

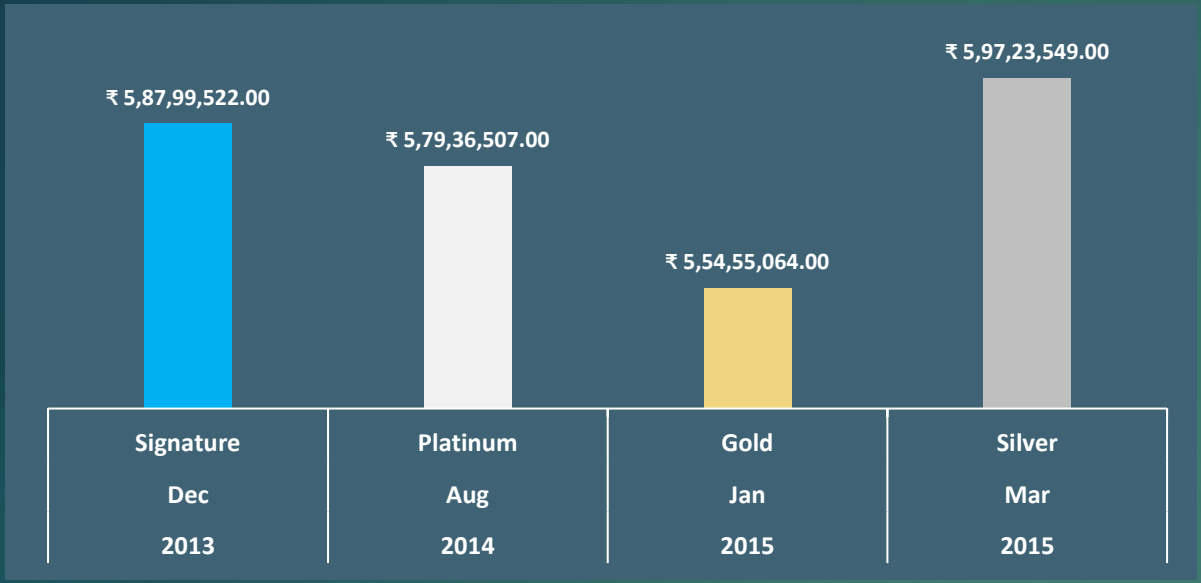
In **Greater Mumbai** the Amount Spent by the credit car owners between **2013 and 2015** is **576 Million** rupees.

Which is **14.15 %** of the total spending between all **986 cities** under the study.

Here is the **SQL Query** to pull data from SQL Server.

```
with cte1 as (
select distinct city,
    sum(amount) over(partition by city order by amount
        desc rows between unbounded preceding and unbounded following) as SpentCityWise,
    sum(amount) over() as grand_total
from credit_in
order by SpentCityWise desc
limit 5)
select city, SpentCityWise,
round( SpentCityWise / grand_total *100,2) as pct_citywise
from cte1;
```

Highest spent month and amount spent in that month for each card types.



In Month of **March 2015**, the total amount spent by the **Silver** card users was highest. (2013-15)

Similarly, highest spending made by **Gold** card users was in **Jan 2015**. (2013-15)

Result Grid | Filter Rows: | Export: | Wrap

trans_year	trans_month	card_type	total_amt	drnk
2013	Dec	Signature	58799522	1
2014	Aug	Platinum	57936507	1
2015	Jan	Gold	55455064	1
2015	Mar	Silver	59723549	1

Here is the **SQL Query** to pull data from SQL Server.

```sql
with cte1 as (
select
    date_format(date, "%Y") as trans_year,
    date_format(date, "%b") as trans_month,
    card_type,
    sum(amount) as total_amt
from credit_in
group by date_format(date, "%Y"), date_format(date, "%b"), card_type),
cte2 as (
    select *,
    dense_rank() over(partition by card_type order by total_amt desc) as drnk
    from cte1)
select * from cte2
where drnk =1
order by trans_year;
```

Fetch transaction details for **gold card** type and expenditure type **entertainment**, when it reaches a cumulative sum of **10 lakhs and above**.

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

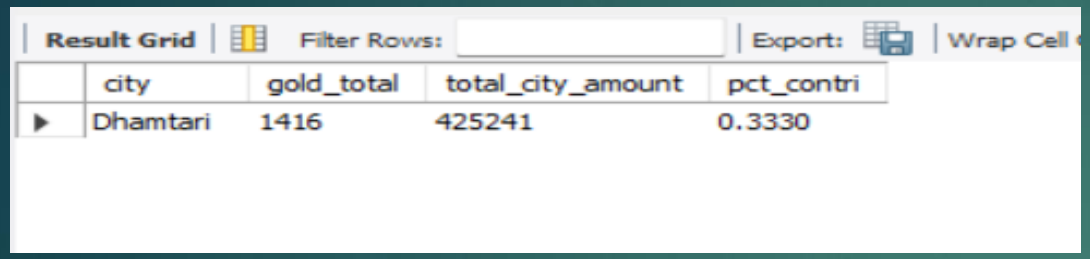
id	city	country	date	card_Type	ExpType	Gender	Amount	cumm_sum	drnk
15118	Phalodi	India	2013-10-04	Gold	Entertainment	M	139361	3414108	1
18278	Sultanganj	India	2013-10-04	Gold	Entertainment	M	131374	3414108	1
19086	Thane	India	2013-10-04	Gold	Entertainment	F	259372	3414108	1
1866	Ahmedabad	India	2013-10-04	Gold	Entertainment	F	78098	3414108	1
7201	Greater Mumbai	India	2013-10-04	Gold	Entertainment	M	201942	3414108	1
8365	Delhi	India	2013-10-04	Gold	Entertainment	M	208425	3414108	1

Cumulative Sum for Gold card type crosses 10 Lakh mark in October 2013 only.

Here is the **SQL Query** to pull data from SQL Server.

```sql
with cte1 as (
select *,
    sum(amount) over(partition by card_type order by date) as cumm_sum
from credit_in),
cte2 as (
select *,
    dense_rank() over(partition by card_type order by cumm_sum) as drnk
from cte1
where cumm_sum >= 1000000 )
select * from cte2
where drnk = 1 and exptype = 'entertainment' and card_type = 'gold';
```

Find city which had Lowest percentage spend for Gold card type.



Result Grid | Filter Rows: | Export: | Wrap Cell

	city	gold_total	total_city_amount	pct_contri
▶	Dhamtari	1416	425241	0.3330

Dhamtari in Chhattisgarh has lowest percentage spend on Gold card type.

Here is the **SQL Query** to pull data from SQL Server.

```sql
with cte1 as (
select city, sum(amount) as gold_total
from credit_in
where card_Type = 'Gold'
group by city ),
cte2 as (
select city,
    sum(amount) as total_city_amount
from credit_in
group by city ),
cte3 as (
select c1.city, c1.gold_total, c2.total_city_amount,
    c1.gold_total /c2.total_city_amount * 100 as pct_contri
from cte1 c1
join cte2 c2 on c1.city = c2.city)
select * from cte3
order by pct_contri asc
limit 1;
```

List City with highest expenditure type and lowest expenditure type.

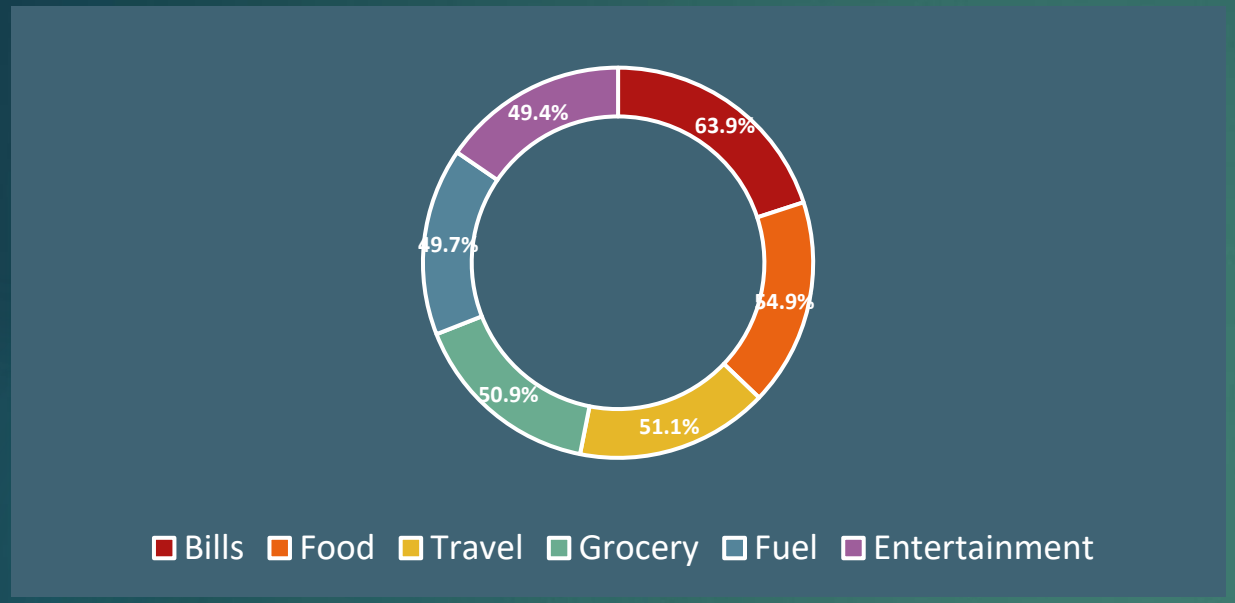
city	highest_exp_type	lowest_exp_type
Achalpur	Grocery	Entertainment
Adilabad	Bills	Food
Adityapur	Food	Grocery
Adoni	Bills	Entertainment
Adoor	Fuel	Bills
Afzalpur	Fuel	Food
Agartala	Grocery	Food
Agra	Bills	Grocery
Ahmedabad	Bills	Grocery
Ahmednagar	Fuel	Grocery
Aizawl	Food	Grocery

For Example: Adilabad city credit card users spend heavily on Bills payment, through Credit Card, whereas least on Food.

Here is the **SQL Query** to pull data from SQL Server.

```sql
with cte1 as (
select city, ExpType, sum(amount) as total_amount
from credit_in
group by city, ExpType),
cte2 as (
select city,
    max(total_amount) as highest_amount_spent,
    min(total_amount) as lowest_amount_spent
from cte1
group by city )
select c1.city,
    max(case when total_amount = highest_amount_spent then exptype end) as highest_exp_type
    ,min(case when total_amount = lowest_amount_spent then exptype end) as lowest_exp_type
from cte1 c1
join cte2 c2 on c1.city = c2.city
group by c1.city
order by c1.city;
```

Find percentage contribution of spends for females by each expense type.



exptype	pct_female
Bills	63.9459
Food	54.9053
Travel	51.1329
Grocery	50.9110
Fuel	49.7104
Entertainment	49.3729

Females spends mostly on Bills in all cities within the specified study period. (2013-15)

Here is the **SQL Query** to pull data from SQL Server.

```sql
with cte1 as (
select ExpType, sum(amount) as total_amount_female
from credit_in
where gender= 'F'
group by ExpType),
cte2 as (
select ExpType, sum(amount) as total_amount
from credit_in
group by ExpType)
select c1.exptype,
    c1.total_amount_female / c2.total_amount * 100 as pct_female
from cte1 c1
join cte2 c2 on c1.exptype = c2.exptype
order by pct_female desc;
```

Find Card type and expense type where there is highest month over month growth in Jan 2014.

	card_type	exptype	trans_year	trans_month	total_amt	prev_month_amt	growth_pct
▶	Gold	Travel	2014	1	2092554	1113534	87.9201

Gold card type in Travel expense category has shown highest 87.92% growth, month over month, in Jan 2014

Here is the **SQL Query** to pull data from SQL Server.

```sql
with cte1 as (
select card_type, exptype,
    date_format(date, '%Y') as trans_year,
    date_format(date, '%c') as trans_month,
    sum(amount) as total_amt
from credit_in
group by  card_type, exptype, trans_year, trans_month
-- order by trans_year, trans_month
),
cte2 as (
select *,
    lag(total_amt, 1) over (partition by card_type, exptype
        order by trans_year, trans_month) as prev_month_amt
from cte1  )
select *,
   (total_amt - prev_month_amt ) / prev_month_amt *100 as growth_pct
from cte2
where trans_year = 2014 and trans_month = 1
order by growth_pct desc
limit 1;
```

Fastest city to reach 500th transaction after first transaction in that city.

	city	tr_start_date	transc_date_500th	no_of_days_to_reach_500th_transc
▶	Bengaluru	2013-10-04	2013-12-24	81

Bengaluru is first city to reach 500th transaction, in just 81 days.

Here is the **SQL Query** to pull data from SQL Server.

```
with cte1 as (
select city,
    min(date) as tr_start_date,
    max(date) as tr_end_date,
    count(*) as total_trans
from credit_in
group by city),
cte2 as (select * from cte1 where total_trans >= 500),
cte3 as (
select city, date, row_number() over(partition by city order by date) as rn
from credit_in
where city in (select city from cte2)),
cte4 as (select c2.city, tr_start_date, tr_end_date, total_trans,
    c3.date as transc_date_500th
from cte2 c2
join cte3 c3 on c2.city = c3.city
where c3.rn = 500)
select city,  tr_start_date, transc_date_500th,
    datediff(transc_date_500th, tr_start_date) as no_of_days_to_reach_500th_transc
from cte4
order by no_of_days_to_reach_500th_transc
limit 1;
```

Thank You