ANALYSIS ON CREDIT CARD DATA USING SQL

Data Set

About the Dataset

This dataset provides a detailed overview of credit card transactions across India, highlighting spending patterns by gender, card type, cities with the highest expenditures, and expense categories. It enables analysis of customer preferences, spending habits, and trends, offering valuable insights for business intelligence and deeper exploration of consumer behavior.

Columns

Column Name	Description
City	The city where the transaction took place. (String)
Date	The date when the transaction took place. (Date)
Card Type	The type of credit card used for the transaction. (String)
Exp Type	The type of expense associated with the transaction. (String)
Gender	The gender of the cardholder. (String)
Amount	Amount spent on transaction. (Number)

Goal 1: Analyse Consumer Trends and Interests

Query 1: Top Spending Cities

```
SELECT Top 10
    city,
    SUM(convert(int, amount)) AS total_spent,
    COUNT(*) AS transaction_count
FROM
    transactions
GROUP BY
    city
ORDER BY
    total_spent DESC
```

	city	total_spent	transaction_count
1	Greater Mumbai	576751476	3493
2	Bengaluru	572326739	3552
3	Ahmedabad	567794310	3491
4	Delhi	556929212	3482
5	Kolkata	115466943	773
6	Lucknow	115334476	759
7	Chennai	114730600	774
8	Hyderabad	114493477	784
9	Surat	114486151	749
10	Kanpur	114370532	764

Code Breakdown:

1. **SELECT Clause**:
    - Fetches the city, calculates the total amount spent (SUM(convert(int, amount))), and counts the number of transactions (COUNT(*)). The convert(int, amount) ensures the amount is treated as an integer for summation.
2. **FROM and GROUP BY Clauses**:
    - Queries data from the transactions table and groups it by city, so the aggregate functions (SUM and COUNT) are calculated per city.
3. **ORDER BY and LIMIT**:
    - Orders the results by total_spent in descending order to prioritize cities with the highest spending.
    - Limits the output to the **top 10 cities** using SELECT Top 10.

Query 2: Most Popular Expense Type by Gender

```
SELECT
    gender,
    [Exp Type],
    COUNT(*) AS transaction_count,
    SUM(convert(int, amount)) AS
FROM
    transactions
GROUP BY
    gender, [Exp Type]
ORDER BY
    gender, transaction_count DESC;
```

	gender	Exp Type	transaction_count	total_spent
1	F	Food	3012	452817279
2	F	Bills	2860	580035469
3	F	Fuel	2617	392282421
4	F	Grocery	2428	365646998
5	F	Entertainment	2384	358663333
6	F	Travel	379	55865530
7	M	Fuel	2640	396853400
8	M	Food	2451	371906730
9	M	Entertainment	2378	367774203
10	M	Grocery	2326	352560925
11	M	Bills	2218	327037004
12	M	Travel	359	53390081

Code Breakdown:

1. **SELECT Clause**:
    - Extracts gender and [Exp Type] (expense type), calculates the number of transactions (COUNT(*)), and computes the total amount spent (SUM(convert(int, amount))).
2. **FROM and GROUP BY Clauses**:
    - Queries data from the transactions table and groups it by gender and [Exp Type], so the aggregations (COUNT and SUM) are calculated for each combination of gender and expense type.
3. **ORDER BY Clause**:
    - Sorts the results first by gender alphabetically and then by transaction_count in descending order to prioritize the highest transaction counts within each gender group.

Query 3: Total Usage Count of Each Card Type

```sql
SELECT
    [Card Type],
    COUNT(*) AS Usage_Count
FROM
    Transactions
GROUP BY
    [Card Type]
ORDER BY
    Usage_Count DESC;
```

	Card Type	Usage_Count
1	Silver	6840
2	Signature	6447
3	Platinum	6398
4	Gold	6367

Code Breakdown:

1. **SELECT Clause**:
    - Retrieves the [Card Type] and calculates the number of transactions (COUNT(*)) for each card type, naming this count as Usage_Count.
2. **FROM and GROUP BY Clauses**:
    - Queries data from the Transactions table and groups the results by [Card Type] so that the transaction count is calculated per card type.
3. **ORDER BY Clause**:
    - Sorts the grouped results by Usage_Count in descending order, showing the most used card types first.

Insights:

Greater Mumbai leads with ₹576.75 crore spent across 3,493 transactions, followed closely by Bengaluru and Ahmedabad. These cities dominate consumer spending, reflecting high purchasing activity in metropolitan areas.

Women spend the most on "Bills" (₹580.03 crore) and "Food" (₹452.81 crore), while men primarily spend on "Fuel" (₹396.85 crore) and "Food" (₹371.91 crore). Food consistently shows high transaction counts across both genders, indicating its importance as a key expenditure category.

Silver" cards lead with 6,840 transactions, followed by "Signature" and "Platinum" cards with 6,447 and 6,398 transactions, respectively. Despite being premium, "Gold" cards are the least used among the listed types, highlighting a preference for mid-tier cards like "Silver."

Goal 2: Detect Potential Credit Card Fraud

Query 1: Unusually High Spending Transactions

```
SELECT
    city,
    gender,
    amount,
    [Card Type],
    [Exp Type],
    date
FROM
    transactions t
WHERE
    amount > (SELECT AVG(convert(int,amount)) + 6 * STDEV(cast(amount as money))
              FROM transactions
              WHERE city = t.city AND gender = t.gender)
order by city;
```

	city	gender	amount	Card Type	Exp Type	date
1	Ahmedabad	F	995634	Platinum	Bills	27-Aug-14
2	Ahmedabad	F	980146	Gold	Bills	25-Mar-15
3	Ahmedabad	F	996291	Gold	Bills	17-Apr-15
4	Bengaluru	F	977951	Signature	Bills	19-May-15
5	Bengaluru	F	965721	Platinum	Bills	05-Aug-14
6	Bengaluru	F	987935	Platinum	Bills	18-Jan-15
7	Bengaluru	F	983839	Platinum	Bills	19-Jan-15
8	Bengaluru	F	973026	Signature	Bills	13-Oct-13
9	Bengaluru	F	984466	Signature	Bills	06-Mar-14
10	Delhi	F	996754	Gold	Bills	27-Oct-13
11	Delhi	F	974384	Gold	Bills	18-Apr-14
12	Delhi	F	981404	Signature	Bills	15-Feb-15
13	Delhi	F	973519	Gold	Bills	18-Oct-14
14	Delhi	F	994537	Signature	Bills	07-Jun-14

Code Breakdown:

1. **Main Query**:
    - Selects city, gender, amount, [Card Type], [Exp Type], and date from the transactions table (t). Filters the data to only include rows where the amount is significantly higher than a threshold defined in the subquery.
2. **Subquery**:
    - Calculates the threshold as the average (AVG) amount plus six times the standard deviation (STDEV) for transactions within the same city and gender as the row in the main query. Uses correlated subquery logic (WHERE city = t.city AND gender = t.gender) to dynamically compare against the city and gender of the current row.
3. **Filtering and Sorting**:
    - The WHERE clause ensures only transactions with unusually high amounts (outliers) are included. Results are ordered by city alphabetically for structured output.

Query 2: Transactions with unusual expenditure type

```
SELECT city, Gender, [Exp Type], amount
FROM transactions t
WHERE [exp type] IN ('Travel') and amount > 2*(SELECT AVG(convert(int,amount))
                FROM transactions
                WHERE city = t.city AND gender = t.gender)
order by city;
```

	city	Gender	Exp Type	amount
1	Hyderabad	F	Travel	299377
2	Hyderabad	M	Travel	297354
3	Jaipur	F	Travel	294370
4	Kanpur	M	Travel	295684
5	Lucknow	F	Travel	299618
6	Pune	M	Travel	295713
7	Pune	M	Travel	290173

Code Breakdown:

1. **SELECT Clause**:
    - Retrieves city, Gender, [Exp Type], and amount from the transactions table (t).
2. **WHERE Clause**:
    - Filters for rows where [Exp Type] is 'Travel' and amount is greater than **twice** the average amount for transactions matching the same city and gender (calculated using a correlated subquery).
3. **Subquery**:
    - Computes the average amount (AVG(convert(int,amount))) for transactions within the same city and gender group as the current row (city = t.city AND gender = t.gender).
4. **ORDER BY Clause**:
    - Orders the results by city in alphabetical order for better readability.

Insights: There are some unusually high spending transactions that are deviating more than 5 times the standard deviation from the mean. Also, there are some transactions spending more amount in the less expenditure type area like “travel” which might be a case of potential credit card fraud.

Goal 3: Predict Spending Patterns for Promotional Campaigns

Query 1: Highest Monthly Spend by Card Type

```
SELECT
    MONTH(CONVERT(DATE, date, 106)) AS month,
    [card type],
    SUM(convert(int,amount)) AS total_spent,
    COUNT(*) AS transaction_count
FROM
    transactions
GROUP BY
    MONTH(CONVERT(DATE, date, 106)),
ORDER BY
    month, total_spent DESC;
```

month	card type	total_spent	transaction_count
1	Platinum	112784373	683
1	Gold	110146204	709
1	Silver	109359598	690
1	Signature	98919381	640
2	Silver	107231486	658
2	Signature	94637714	587
2	Gold	91703346	602
2	Platinum	89940078	586
3	Silver	113547912	710
3	Platinum	104760684	660
3	Gold	99773193	644
3	Signature	91972657	608

month	card type	total_spent	transaction_count
4	Platinum	106793699	658
4	Silver	106532762	672
4	Signature	102329658	624
4	Gold	100738615	633
5	Silver	100700833	656
5	Signature	96602648	613
5	Gold	90290513	584
5	Platinum	87861615	583
6	Silver	51437691	334
6	Gold	48104062	296
6	Platinum	46761272	300
6	Signature	45018014	309

month	card type	total_spent	transaction_count
7	Signature	52510876	342
7	Silver	51377700	334
7	Platinum	48026689	328
7	Gold	46066151	306
8	Silver	58831694	374
8	Platinum	57936507	361
8	Signature	51965786	337
8	Gold	49719139	321
9	Silver	53451091	325
9	Signature	53339323	339
9	Gold	53155064	348
9	Platinum	49615955	309

month	card type	total_spent	transaction_count
10	Signature	115694418	717
10	Silver	108427432	698
10	Gold	98265077	623
10	Platinum	97459080	627
11	Platinum	104189636	649
11	Signature	102285132	651
11	Gold	100119324	661
11	Silver	97513876	661
12	Silver	111201638	728
12	Signature	107765498	680
12	Platinum	101509431	654
12	Gold	96458848	640

Code Breakdown:

1. **SELECT Clause**:
    - Extracts the month (MONTH(CONVERT(DATE, date, 106))) from the date column, card type ([card type]), total amount spent (SUM(convert(int,amount))), and the number of transactions (COUNT(*)).
2. **FROM and GROUP BY Clauses**:
    - Queries data from the transactions table and groups it by month and card type (MONTH(CONVERT(DATE, date, 106)), [card type]) to calculate aggregated metrics (SUM and COUNT) for each month-card type combination.
3. **ORDER BY Clause**:
    - Sorts results first by month (ascending) and then by total_spent in descending order, prioritizing card types with the highest spending within each month.

Query 2: Top spending during holidays grouped by city and card type

```
WITH RankedTransactions AS (
    SELECT
        city,
        [card type],
        MONTH(CONVERT(DATE, date, 106)) AS month,
        SUM(CONVERT(INT, amount)) AS total_spent,
        COUNT(*) AS transaction_count,
        ROW_NUMBER() OVER (PARTITION BY MONTH(CONVERT(DATE, date, 106))
                           ORDER BY SUM(CONVERT(INT, amount)) DESC) AS rank
    FROM
        transactions
    WHERE
        MONTH(CONVERT(DATE, date, 106)) IN (10, 11, 12) -- /Diwali/Year-End
Holidays
    GROUP BY
        city, [card type], MONTH(CONVERT(DATE, date, 106))
)
SELECT
    month,
    city,
    [card type],
    total_spent,
    transaction_count
FROM
    RankedTransactions
WHERE
    rank <= 3
ORDER BY  month, rank;
```

month	city	card type	total_spent	transaction_count
10	Ahmedabad	Signature	19200929	114
10	Greater Mumbai	Silver	18418457	98
10	Bengaluru	Signature	16772604	107
11	Bengaluru	Silver	18270269	112
11	Greater Mumbai	Platinum	16367504	95
11	Ahmedabad	Platinum	16184318	94
12	Greater Mumbai	Silver	18353161	113
12	Ahmedabad	Gold	16977430	97
12	Bengaluru	Silver	16783151	110

Code Breakdown:

1. **Common Table Expression (CTE)**:
    - **RankedTransactions CTE**: Computes aggregates for each city and [card type] within specific months (10, 11, 12) and calculates their rank using ROW_NUMBER() for each month, ordered by total_spent (descending).
    - Includes:
        - SUM(CONVERT(INT, amount)) for total spending.

        - COUNT(*) for the number of transactions.
        - ROW_NUMBER() to rank entries within each month.

2. **WHERE Clause in the CTE**:
    - Filters transactions to only include those in October, November, and December (Diwali/Year-End Holidays).
3. **Main Query**:
    - Selects the top-ranked (rank <= 3) city and [card type] combinations from the RankedTransactions CTE for each month.
4. **ORDER BY Clause**:
    - Orders the results by month and then by rank, ensuring a structured view of top-performing cities and card types for each holiday month.

Insights:

The month-wise analysis reveals that total spending is highest for Silver cards, followed by Platinum cards, indicating a preference for mid-tier cards over premium options like Gold. Additionally, during holiday periods, the top spending cities are Ahmedabad, Greater Mumbai, and Bengaluru.

Recommendations:

1. Promote Mid-Tier Cards (Silver and Platinum):

- Introduce targeted rewards or cashback programs for Silver and Platinum cardholders to further capitalize on their popularity.

- Offer exclusive perks for frequent users to enhance loyalty and encourage higher usage.

2. Boost Gold Card Adoption:

- Enhance Gold card benefits with competitive offers like higher reward rates or complimentary services to attract more users.

- Implement campaigns highlighting the unique advantages of premium cards to convert mid-tier cardholders.

3. Focus Marketing Efforts on Top Spending Cities:

- Launch region-specific campaigns during holidays in Ahmedabad, Greater Mumbai, and Bengaluru to maximize customer engagement and spending.

- Collaborate with local businesses and merchants in these cities to offer tailored discounts for cardholders.