

# Data Analytics and Insights

## Overview

As a part of our evaluation exercise, we are sharing some questions along with some dataset for you to evaluate. Some of the questions are intentionally kept vague and may require you to make some presumptions. In such cases, when presenting your solution, please state those assumptions very clearly. We will be looking at your approach as well as the final output of the problem, so think about how you will create solutions and scale them as well.

There is **no timeline for submission of your solutions** but the sooner you submit the better will be your chances

The top link of the data dictionary tables points to the dataset that would be relevant for a particular problem. You should be able to download it from there.

## Submission

Once you are satisfied with your solution, please submit a detailed **Python script notebook** along with all insights generated from this task in a **Word document, PDF, or PPT**. and share the link in the same email where you got the assignment (**Question number 1,3,4,5,6 to be done using python and question 2 to be done using SQL**)

If you need any help with the problem or otherwise, please respond back to the same thread where you got the problem from.

All the best!

---

## 1. Transaction data analysis

In many applications, monitoring user wallet activity is critical for identifying potential fraud or misuse. By analyzing transaction data, we can spot patterns where a user's behavior significantly deviates from the usual trends. Such deviations may indicate suspicious activities, such as fraudulent transactions, misuse of the system, or irregular behavior, which require further investigation.

The problem aims to analyze user transaction behavior following their first transaction. Specifically, we need to answer three key questions regarding the timing and amounts of subsequent transactions as below :

1. How long did it take for the user to make their second transaction on average?
2. How many percent of users who did the first transaction also did the second transaction?
3. What was the average transaction amount after the first transaction?  
For example, if a user's first transaction was 100 INR and they subsequently made transactions of 50, 100, and 200 INR, the average would be  $(50 + 100 + 200) / 3$ .

1 <a href="#">txn_data.csv</a>	
Column Name	Comments
adv_id	<a href="#">Advertising ID</a> of the user, consider it as an unique user id of the user
value_in_paise	Value of the transaction in paisa
payment_gateway	Name of the payment gateway
payment_method	Name of the payment method
payment_status	Status of the payment
created_at	Transaction creation date
app_id	Unique id of the app
from_currency_amount	Amount of the currency to be converted from
from_currency	Payment from the currency type
to_currency_amount	Amount of the currency converted to
to_currency	Converted to the type of currency
transaction_fee_amount	Amount of transaction fee we are charging users to pay gateway
transaction_fee_currency	Currency of transaction fee

## 2. Identify High-Value Transactions

This problem focuses on analyzing transaction data to understand user spending patterns and detect any anomalies or spikes in transaction amounts. By identifying transactions that are significantly higher than the user's recent transaction history, we can flag potential issues, such as unusual spending behavior, fraud, or system errors.

So based on this overview give the solution for below problem statement :

1. Write an SQL query to identify transactions where the value\_in\_paise is at least twice the average of the previous three transactions for the same user.

## 2 high\_val\_txn.csv

Column Name	Comments
adv_id	Advertising ID of the user, consider it as an unique user id of the user
value_in_paise	Value of the transaction in paisa
payment_gateway	Name of the payment gateway
payment_method	Name of the payment method
payment_status	Status of the payment
created_at	Transaction creation date
app_id	Unique id of the app
from_currency_amount	Amount of the currency to be converted from
from_currency	Payment from the currency type
to_currency_amount	Amount of the currency converted to
to_currency	Converted to the type of currency
transaction_fee_amount	Amount of transaction fee we are charging users to pay gateway
transaction_fee_currency	Currency of transaction fee

## 3. Identify Unusual Wallet Activity for Potential Fraud Detection

In many applications, detecting unusual wallet activity is crucial for identifying potential fraudulent behavior. By analyzing users' transaction patterns, we can pinpoint those whose wallet activity deviates significantly from the norm. Such anomalies may indicate suspicious actions, including fraudulent transactions or misuse of the system.

So based on this overview give the solution for below problem statement :

1. Identify users whose wallet activity is significantly different from others, and explain how this could relate to fraud detection.

## 3 wallet\_data.csv

Column Name	Comments
adv_id	Advertising ID of the user

transaction_type	Can be DEBIT OR CREDIT, - DEBIT means the amount is deducted from the Greedygame account and transferred to the user account. - CREDIT means the amount is deducted from the user account and transferred to the Greedygame account.
value	Amount in currency
currency	Currency name of the transaction
source_entity_type	Id of the source entity like different ways using which amount can be Debit or Credit from users account.
created_at	Date on which the transaction was created

## 4. Analyze Upvote and Downvote Data to Extract Key Insights

In many applications, understanding user engagement, offer quality, and overall sentiment is essential. Analyzing upvote and downvote data can provide valuable insights into how users interact with content and how they perceive it. By examining patterns in upvotes and downvotes, we can uncover trends that help in assessing user satisfaction and content relevance. Users can access upvote and downvote options only after completing a reward. For example, if a user engages with an offer from Mobikwik, they will receive the option to vote on the offer only after starting and interacting with it. This allows users to provide feedback based on their actual experience with each offer.

So based on this overview give the solution for below problem statement :

1. We have upvote and downvote data. Based on this, provide as many insights as possible from an analytical perspective

<a href="#">4_voting_data.csv</a>	
Column Name	Comments
adv_id	<a href="#">Advertising ID</a> of the user
offer_id	The Unique ID of the offer is a unique identifier for a specific product or offer. Here, an "offer" refers to various types of tasks associated with it, such as installing an app, completing KYC, and more. To explore this further, you can try out our BharatCash app by installing it from the app store link: <a href="#">BharatCash app</a>
app_id	Unique ID of the app
state	Types of the comment as UP and DOWN , where UP means Positive comment and DOWN means

	Negative comment
comment	Comment by users after offer completion
ip	IP address from where the comment was created
country_code	Country from where comment was created
created_at	Date on which the comment was created

## 5. Comparing Revenue Per User Between Referral and Non-Referral Groups

In many applications, understanding the impact of different user acquisition strategies on revenue generation is crucial. One common strategy is referral programs, where existing users encourage others to join. The success of such strategies can be measured by comparing the Revenue Per User (RPU) between users who joined through referrals and those who joined organically.

By analyzing the differences in revenue between these two groups, we can gain insights into how effective the referral program is in driving higher revenue. This analysis can help inform future marketing and acquisition strategies to maximize revenue growth and improve user engagement.

So based on this overview give the solution for below problem statement :

1. In this problem, we aim to compare the Revenue Per User (RPU) between users who joined the Bharat Cash app via referrals and those who joined without referrals. We have data that captures the total users, revenue from referred users, and overall revenue, and our goal is to derive insights about how the referral strategy impacts user revenue generation.

### [5\\_1\\_bharat\\_cash\\_overall\\_daily\\_login.csv](#)

The total number of users who logged into the Bharat Cash app

Column Name	Comments
adv_id	Advertising ID of the user
app_id	Unique id of the app
day	Date when an user came on app so its like DAU

### [5\\_2\\_from\\_referral\\_bharatcash.csv](#)

The number of users who joined the app via referral

Column Name	Comments
referre_adv_id	Advertising ID of the user who is doing referrals.
created_at	Timestamp when referral was created

[5 3 rev\\_from\\_referral\\_bharatcash.csv](#)

The total revenue generated by users who joined via referral

[5 4 rev\\_overall\\_bharatcash.csv](#)

The total revenue generated by all users (both referred and non-referred)

Column Name	Comments
adv_id	Advertising ID of the user
revenue_in_usd	Revenue generated from an user in \$

## 6. Detecting Discrepancies in User Signup and Click Locations

In many applications, understanding user behavior is critical to ensuring a smooth and secure user experience. One key aspect of this behavior is the consistency of user information across different interactions with the system. Discrepancies in user data, such as differences between a user's registration details and their activity patterns, can indicate potential issues, including fraud, misuse of system resources, or unintended behaviors.

By identifying and analyzing such inconsistencies, we can take corrective actions, improve security, and maintain the integrity of the system. This analysis can be particularly useful in detecting suspicious activity, ensuring that the platform functions fairly for all users.

1. We have access to users' click IDs and their signup data. Using this information, the goal is to identify any users who show discrepancies between their signup location and the location from which they make clicks. For example, a user may have signed up from one country, but their interactions with the app (clicks) may originate from a different country.
2. The task is to detect such discrepancies, as they could indicate suspicious activity, such as fraud, account takeover, or use of VPNs. By identifying these inconsistencies, we can take steps to address potential fraud or other irregular behaviors on the platform.

<https://dashboard.pubscale.com/json/country.json> this link has info about country code and its name so use it to get country name from country code

[6 1 clicks\\_country.csv](#)

This defined the country code from which the click was generated

Column Name	Comments
click_id	Unique click ID generated by an user
country_code	Country code from which this click was generated

#### [6 2 clicks.csv](#)

Clicks data gives info about users clicks and ip on a particular reward

Column Name	Comments
adv_id	Advertising ID of the user
click_id	Unique click ID generated by an user
reward_id	Unique ID of the reward
ip	IP address from where click was generated

#### [6 3 login log.csv](#)

login\_log means the users who came on our apps any day so consider it as DAU.

Column Name	Comments
adv_id	Adv id of the user
app_id	Unique id of the app
day	Date when an user came on app so its like DAU

#### [6 4 user signup location.csv](#)

User\_signup\_location means that which user is signup from which country code

Column Name	Comments
adv_id	Adv id of the user
country_code	Signup country code of a user