# Dream Games
# Data Scientist Case Study

---


**Summary**

Row Match!  is a game in which players are tasked with helping the King restore his royal castle by rebuilding it through a series of match-3 levels and obstacles, with new rooms, royal chambers, and gardens making up the different levels of the game. This task will be about the game that was defined in the previous sentence. You can find the details about the dataset in the Dataset Description part of this document.


**Task Information**

As a data scientist in the team, we'd like you to perform a deep dive into different datasets using any tools of your choice.

1. We want you to evaluate the current status of Row Match! via an exploratory data analysis, highlight your findings in the dataset and suggest actions and pinpoint positive or negative findings based on your analysis. (We are expecting you to calculate cohorted and daily metrics such as ARPU, Retention, DAU, ARPDAU, ROAS etc. to understand how the main KPIs of the game are evolving).

2. There was an A/B test with 2 groups. In the dataset, you can find the ab_test_enter, revenue and session tables for the users who participated in this A/B test. Based on the data, you are expected to decide which variant is better in terms of monetization and engagement metrics for Row Match and support the decision with your findings.

3. We want you to create a model about user behavior based on the activity of the user in the first 7 days. The objective is to predict whether a user is going to make a purchase in 30 days after installation (We are expecting you to provide a detailed plan on how to approach this task, including data preparation, feature engineering, model selection, and evaluation).

You are expected to deliver your analysis, which includes the following elements:

- The code script (SQL, Python, R, etc)
- Your report, findings, summaries, charts
- Explanation of your approach, methods, and comments

**Evaluation**

We mostly consider these factors to evaluate your study:

- Your capabilities to explore and make sense of given datasets while generating meaningful and clearly explained insights
- The amount and conclusiveness of evidence you provide to support your analysis
- Your ability to use different data analysis & science tools or techniques
- Model selection, validation, and evaluation techniques

The expectation is usage of SQL along with a programming language based on your preferences (Python, R, etc) during the analysis.

**Dataset Description**

**q1_table_install -** install event triggered when the user installs the game

---

event_time: timestamp of the event occurrence
user_id: unique identifier for the user
platform: platform of the game (ios/android)
network: a network which shows where the user is acquired
country: country of the user

---

**q1_table_level_end -** level_end event triggered every time user finishes a level

---

event_time: timestamp of the event occurrence
user_id: unique identifier for the user
platform: platform of the game (ios/android)
level: current level of the user
status: level status of the user (win, fail, quit)
time_spent: time spend of the user in the current level
moves_made: total number of moves made by the user
moves_left: total number of moves that left at the end of the level

**q1_table_session** - session event sent for every session of the user

---

event_time: timestamp of the event occurrence
user_id: unique identifier for the user
platform: platform of the game (ios/android)
coin_status: coin amount of the user
time_spent: time spend of the user
level: current level of the user

---

**q1_table_revenue** - triggered when the user made a purchase

---

event_time: timestamp of the event occurrence
user_id: unique identifier for the user
platform: platform of the game (ios/android)
package_type: type of the package
revenue: dollar amount of the purchase

---

**q1_table_cost** - cost of the user acquisition activity

---

date: date of the cost of user acquisition
network: network which shows where user is acquired
platform: platform of the game (ios/android)
country: country of the user
cost: dollar amount of the marketing spend

---

**q2_table_ab_test_enter** - triggered when user enters an AB Test

---

test_entry_timestamp: timestamp of the users AB test entrance
install_timestamp: timestamp of the user install
user_id: unique identifier for the user
platform: platform of the game (ios/android)
group_id: ID of the AB Test group

---

**q2_table_ab_test_revenue** - triggered when a user in the AB test made a purchase

---

event_timestamp: timestamp of the event occurrence
user_id: unique identifier for the user
platform: platform of the game (ios/android)
package_type: type of the package
level: current level of the user
dollar_amount: dollar amount of the purchase

---

**q2_table_ab_test_session** - sent for every session for a user in the AB test

---

event_timestamp: timestamp of the event occurrence
user_id: unique identifier for the user
platform: platform of the game (ios/android)
time_spent: time spend of the user
level: current level of the user

---

**q3_table_user_metrics** - Day 7 metrics and Day 30 revenue value of the users

---

user_id: unique identifier for the user
platform: platform of the game (ios/android)
network: network which shows where user is acquired
country: country of the user
age: age of the user
time_spend: total time spend of the user until Day 7
coin_spend: total coin spend of the user until Day 7
coin_earn: total coin earn of the user until Day 7
level_success: total level count of the user that has been successfully completed until Day 7
level_fail: total level try count of the user that has been not successfully completed until Day 7
level_start: total level try count of the user until Day 7
booster_spend: total amount of booster spend of the user until Day 7
booster_earn: total amount of booster earn of the user until Day 7
coin_amount: latest(Day 7) coin amount of the user
event_participance: event participation status of the user(0: no participation, 1: participation)
shop_open: number of shop open of the user until Day 7
d30_revenue: total revenue of the user until Day 30

---

**Database Connection**

The data is provided in Google Big Query and you have granted access to your gmail account. “Analytics” dataset has been created for this case. You can find all the tables which are described above in the given dataset.

https://console.cloud.google.com/bigquery?project=casedreamgames

You can access BigQuery with the link above.  Also, in the email, you are going to receive a JSON file for connecting the database via Python, R, etc.  For simplicity, you can find how to connect via Python script in the following section.

**Example - Python Connection**

```
#import necessary libraries
from google.cloud import bigquery as bq
import pandas as pd

#client connection
c = bq.Client.from_service_account_json("/path/to/json/document/file.json")
query = """ select * from `casedreamgames.case_db.q1_table_install` limit 100 """
#query results to dataframe
df = c.query(query).to_dataframe()
```