



Data Analyst

Technical Challenge

Welcome to the Data Analyst challenge

This task aims to assess your thinking process when doing data analysis. We want to understand how you approach delivering insights. We know there are multiple ways to solve a problem. Therefore, we are expecting a *variety of answers*.

What matters most is your thought process, justifications and how you present your insights.

What we expect

Here are some guidelines to keep in mind while creating your output:

- **Provide actionable insights.** If there is no apparent action due to your chart, the chart may be unnecessary.
- **Be clear and concise in your report.** We want to understand how you present and communicate your findings to non-technical stakeholders.
- **Send us the insights, all code, and any other supporting files inside a zipped folder** (.zip format).
 - Your code should allow us to understand how you arrived at your insights.
 - **Please do not submit files in any particular BI tool format** (e.g. PBIX, TWB, etc.). Assume we won't be able to open it.
 - Python/R files can be submitted in any format.
- **You only have 10 days to deliver the insights.**
- **We don't expect you to explore all aspects of the data. You can focus on one area.**
- The report should be comprehensive enough to facilitate a conversation during the team round.

Task Introduction

The team have an upcoming update for one of their games, Combat Elite, and are considering adjusting features where necessary.

Combat Elite has 4-player matches which can be against other players and/or bots. There are 3 match types that are available in turns in a cycle to the player.

Combat Elite has products which can be bought with real money, known as in-app purchases (IAPs), as well as products that can be bought with the in-game virtual currency, Gold.

What insights will you give, and what changes can you suggest for this game?

Data

You can find the data sets for this task in a zipped folder **challenge_task_data.zip**. Inside the zip file, you will find three data sets:

1. **data_daily_activity.csv** containing the daily activity of the users. Each row represents a user who opened the game on that specific day. This data set has the following structure:
 - *activity_date* - The calendar date when the user was active
 - *acquisition_date* - The calendar date when the user was active in the game for the first time
 - *platform* - The users' reported platform on each calendar date
 - *user_id* - The unique user ID of each active user
2. **data_matches.csv** contains all the matches played during the analysis period per user. Each row represents a day of matches. This data set has the following structure:
 - *activity_date* - The calendar date when the user was active
 - *user_id* - The unique user ID of each active user
 - *match_type* - The type of match that is played
 - *finish_position* - The finish position of the players (1-4)
 - *bots* - The number of bots in the match (0-3)
 - *n_matches* - The total number of matches per user by calendar date, match type and finish position
3. **data_in_app_purchases.csv** contains all the in-app purchases made during the analysis period per user. Each row represents a transaction. This data set has the following structure:
 - *activity_date* - The calendar date when the user was active
 - *user_id* - The unique user ID of each active user
 - *purchase_number* - The number of the purchase made by the specific user_id
 - *product_group* - The group of the product purchased
 - *dollar_purchase_value* - the revenue that purchased generated in dollars
4. **data_virtual_purchases.csv** contains all the purchases made using gold during the analysis period per user. Each row represents a day of transactions. This data set has the following structure:
 - *activity_date* - The calendar date when the user was active
 - *user_id* - The unique user ID of each active user
 - *product_group* - The group of the product purchased
 - *n_purchases* - The number of the purchase made per user by calendar date, product and product_group
 - *gold_spend* - the total gold spent per user by calendar date, product and product_group

Additional Info

- Due to the high volume of applications, we do not provide feedback on individual submissions.
- You will be asked to present and discuss your insights if you make it to the next round.
- Please be patient about hearing back from us. We will try our best to get back to you soon.