# Data Analyst Case Study

Welcome to the Jollify Games Data Analyst Case Study phase. This segment provides an illuminating view into the data with which we engage and the diverse challenges we confront. The initial two challenges encompass a broad spectrum, assessing your analytical acumen and finesse in communication. As we venture into the third challenge, we aim to gauge your capacity for discerning insights within the realm of gaming and your proficiency in the realm of data analysis. This is particularly relevant as you assume the role of a Data Analyst at Jollify Games, collaborating closely with Product Managers.

**Challenge 1: Unveiling Insights from Data Tables**

Below, you will find the contents of five distinct .csv tables, conveniently located at the end of the CSV file. Employing these five tables, we anticipate your execution of the following calculations utilizing SQL or Python: Number of sessions, count of users with sessions, count of purchases, total purchase revenue, revenue per purchase, revenue per purchasing user, count of sessions with purchases, and the total count of new users.

- **User Table:** Columns - user_id, user_country_id, user_join_datetime
- **Purchase Table:** Columns - purchase_id, user_id, purchase_datetime, purchase_revenue, purchase_session_id
- **Country Table:** Columns - Country_id, country_name
- **Media Source Table:** Columns - user_id, media_source_name
- **Session Table:** Columns - session_id, user_id, session_datetime

Additionally, incorporating visualization techniques can significantly elevate your analysis. Feel free to employ tools like Excel, Power BI, or Python libraries to create visual representations of your findings. Such visualizations can provide a more intuitive understanding of the data patterns and trends you uncover, enhancing the clarity and impact of your insights.

**Answer:** At attachment

**Challenge 2: Deciphering User Experience Anomaly**

**Question:** Imagine that the design team at a prominent gaming studio has recently introduced an innovative feature aimed at enriching the user experience on their platform. The primary

objective behind this feature is to amplify the average duration users spend on the site, thereby augmenting the likelihood of sales. However, in the initial weeks post-launch, instead of the projected surge, there is an evident decline in the average session durations on the platform.

This unexpected outcome has perplexed both the Design team and the Data Analytics team. They have turned to you for assistance in unraveling the potential causes behind this anomaly. Kindly share and elucidate your hypotheses.

**Answer:** To formulate a hypothesis, it's essential to identify various KPIs that may be linked to the KPI in question. Let's delve into the processes to achieve this. The most straightforward approach would be to compare site usage data before and after the update. For this situation, the immediate metrics that come to mind are: page view count, average session duration, bounce rate, and click event data. For potentially related metrics, I'd seek input from the design team or senior management. Even if a direct comparison of before and after data provides clear insights, it's essential to analyze it from multiple perspectives:

- Setting Control Limits
- Evaluating Descriptive Statistics
- Establishing Data Confidence Intervals
- Conducting Correlation Analysis
- Assessing Data Distributions

Furthermore, to pinpoint the root cause, I'd employ certain methodologies. If data is sourced from different interfaces, it's vital to ensure their reliability and consistency, possibly using techniques like Gage R&R.

Once we've undergone these procedures, our areas of investigation will become more apparent. Without an intimate understanding of the company's industry and customer demographics, generating a hypothesis becomes challenging. Thus, I'd pinpoint one or more KPIs to aid in brainstorming potential hypotheses. Given that 'average session duration' is continuous data, I'd leverage variation analysis methods like Regression, ANOVA, and t-test. I'd integrate a scatterplot into the regression analysis for continuous-continuous data and employ boxplot charts for continuous-categorical data. Naturally, I'd rely on the data distribution assessments I've conducted to guide these choices.

If the goal isn't to derive a deeply meaningful hypothesis, my primary suggestion would be to consider A/B testing, assuming it hasn't been implemented yet. Beyond the evident advantages of A/B testing, my rationale for this recommendation is simple: as an analysis enthusiast, having an additional dataset for comparison, even if post-update results align with the A/B test, is invaluable.

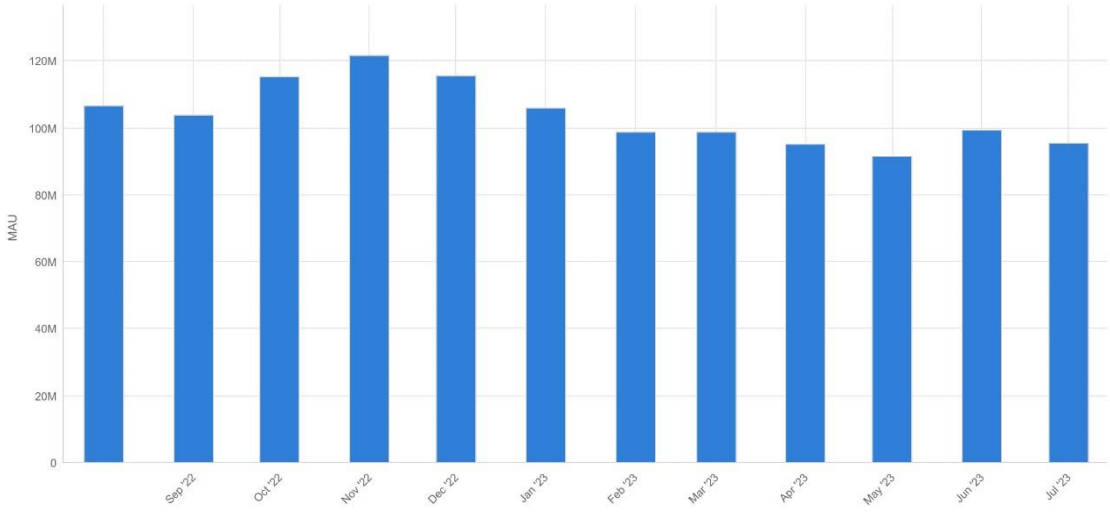
Lastly, a thorough review of user feedback is pivotal before making any decisive conclusions based on the data. It enhances our probability of deriving meaningful insights before diving deep into analysis. While data offers essential insights, the users' perspectives are paramount in decision-making. To extract feedback from users, incentives like small rewards in mobile games could prove beneficial. Favoring commenters slightly more could encourage feedback. If feedback volume is substantial, implementing a script or app to conduct sentiment analysis on the received data might be apt. An aspect extraction technique can help differentiate between significant and trivial comments.

I wanted to show how far I could go with these analyses in terms of statistics/data. However, in the business world, desires of the party requesting the analyses should shape and determine the analysis process. If it's not specified, the analysis should be formed based on prior communications in a way that might be ideal for them.

**Challenge 3: Delving into Graphical Insights**

You are now presented with a pair of graphs that vividly illustrate various facets of performance related to a mobile game. Your task is to conduct a comprehensive analysis of each graph and extract insightful conclusions from these visual representations. It is paramount to recognize that these graphs are interrelated and exert mutual influence on each other. Your analysis should encompass the intricate connections between these visual depictions and the consequential impact they have.

**3.1) Monthly Active Users**




**Questions:**

- What can you infer from the graph regarding the user engagement and retention?
- Identify any notable peaks or valleys in the graph and discuss possible factors that may have influenced these fluctuations.

## Answer:

Starting in August, there's an upward trend that culminates in a peak by November. Following this peak, a steady decline is observed up until June. The heightened user engagement in October and November might be influenced by factors such as year-end gifting or holiday periods.

A decline around the new year is not uncommon and can be anticipated in many industries. The sustained decrease from January through May might suggest difficulties in retaining customer interest or could be a result of increased market competition.

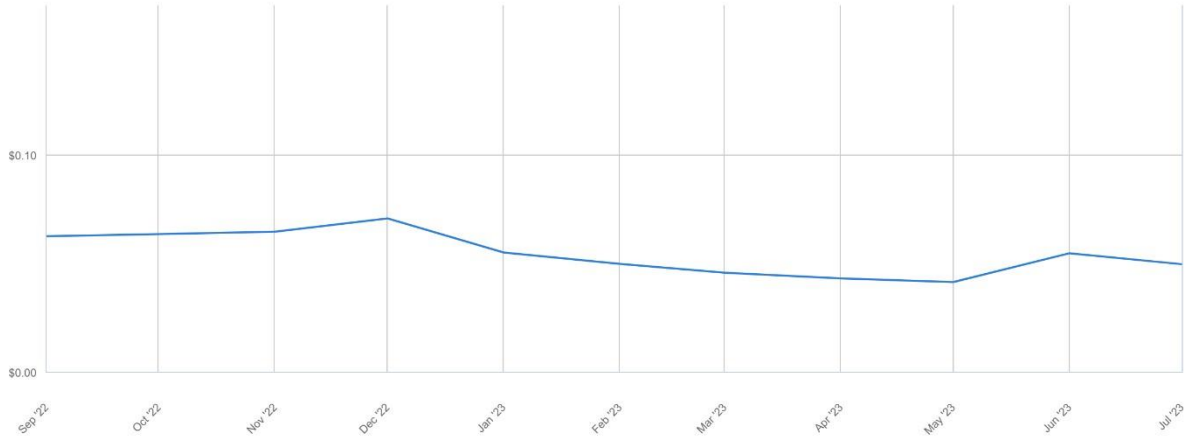
The spike observed in June might suggest a successful initiative or change, but its impact appears to be transient. By contrast, the drop in July warrants investigation to ascertain if it stems from internal

company challenges. To bolster these observations, analyzing data over a longer timeframe, such as 2 or 3 years, can provide a clearer picture.

In sum, several factors could be at play, including seasonal trends, changes to the platform, sales promotions, industry competition, and even global events. For example, if this data were from the Turkish market in real-time, a post-new year decline might be investigated for potential impacts from significant events, such as an earthquake. Additionally, political events like election cycles can have economic impacts, potentially leading to a quieter year like 2023. Technical issues are another consideration.

Lastly, before drawing definitive conclusions, integrating customer feedback can provide invaluable insights.

**3.2) ARPDAU**




**Questions:**

- Does the ARPDAU show any consistent trends or fluctuations over time?
- Identify any noteworthy changes in the ARPDAU and provide possible explanations for these variations.

## Answer:

The insights provided for Monthly Active Users (MAU) are also valid here. The additions this table can give us might be:

- There is a likelihood of a strong correlation between MAU and ARPDAU.
- By cross-referencing, we can obtain more accurate insights.
- It allows us to provide references.

For instance, if MAU remains steady while there is a decline in ARPDAU, it can show that we haven't lost users, but our sales and marketing methods might be reconsidered. Likewise, it should be interpreted according to internal/external user comments

**Thank you for embarking on this intriguing journey through the Jollify Games Data Analyst Case Study. We eagerly await your insights and solutions. Your solutions must include all files in their entirety.**