**Case Study**

*Data Analyst at JustPlay*

Dear Candidate,

We have prepared a short Case Study to get a better insight into how you would approach the role as Data Analyst at JustPlay. Please answer these questions as best you can, there is no need to be too detailed, as we will be discussing it together in the next interview. You have 1 week to complete the Case Study and please share your answers with us before the next stage of interviews.

If you have any questions, feel free to reach out to us. We really appreciate your time and are looking forward to discussing your answers!

**About JustPlay**

This is the JustPlay app ([iOS](https://apps.apple.com/us/app/justplay-loyalty-program/id6444946155) & [Android](https://play.google.com/store/apps/details?id=com.justplay.app&hl=en&gl=US)) and [here are our published games](https://play.google.com/store/apps/dev?id=6638826862196782417&hl=de&gl=US). Users play games to earn rewards. If the app isn't available in your country, you can find many review videos on youtube that will give you a good overview of how the app works.

Let us describe the app briefly: Users download the app where they find a list of games published by us that will reward them for playing. Once they click on one, they get redirected to the Store and when they start playing, they will start earning coins.

After 3 hours, the player can cash out their first earnings in the JustPlay app. The monetization is 100% ad revenue based (with no IAP revenue), a portion of which is modeled out to go to the player. We follow the user across all the games they will download through the app. There is a timer in the app telling the player when their next cashout will be ready.

**The Task**

For this case study, we are providing you with a dataset that includes daily records of revenue, ad spend, and installs for JustPlay. The dataset has been deliberately left uncleaned and only includes a small subset of data.

Your task is to analyze this data to uncover meaningful insights. You may use any tools for this analysis, but please ensure you share the scripts/notebooks used along with your findings. Present your results either as a presentation or a document, detailing your entire process.

Below are some guiding questions to help you with the analysis:

1. What does an exploratory analysis reveal about this data?
2. How would you address various issues such as data cleanliness, joins, and aggregation levels?
3. What metrics can be derived from this data subset?
4. What insights can you uncover from this data? - **This is a KEY step. Please make sure you analyze the metrics you created.**
5. Which teams would benefit from these findings?

Your task is to explore this data, generate metrics, and derive actionable insights. Ensure your analysis addresses data cleanliness, proper joins, and appropriate aggregation levels. Share your detailed findings and the process you followed for the analysis.

The objective of the task is to allow us to understand how you would approach a data request from its conception all the way to the final results presented to the business stakeholders. There are no right/wrong answers and we want to understand your thought process, how well you understand the underlying data and the assumptions you make. Focus less on the results of the metrics themselves.

**Data Description**

You will find the 3 csv files [here](https://drive.google.com/drive/folders/1f9lxBYKx9wqms_sqX-Og2nzfnLhJ6lIt?usp=sharing). They each contain a single day of data.

* Installs: Each row represents an install event, providing information about user attribution. The column channel in this table corresponds to the column channel in the ad\_spend table, indicating the channel responsible for the install.
* Revenue: Each row represents a single ad impression, with the column Network indicating where the revenue was generated. - *This file contains less than a day of data*
* Ad\_spend: This table includes data on ad spend, installs reported by both the channels (for the metrics marked with the ending \_network) and the attribution provider, and other useful metrics.

**Metadata**

For ad spend data:

* channel: "ad channel name. Also known as ad network",
* campaign: "ad campaign name",
* creative: "name of the ad",
* network\_clicks: "number of clicks on ad",
* network\_impressions: "number of impression on ad",
* network\_installs: "number of installs reported by channel",
* network\_installs\_diff: "difference between installs reported by Adjust and channel. installs - network installs",
* installs: "number of installs attributed by Adjust",
* cost: "The amount of money spent on ads",
* os\_name: “Corresponds to the platform of the user, either android or iOS”
* country\_code: “the country code that Adjust attributes to a particular creative”

For installs:

* campaign: "name of the ad campaign",
* tracker\_name: "tracker name. Adjust link token",
* channel: "ad channel name. Also known as ad network",
* adjust\_id: "unique id generated by Adjust, our app attribution provider",
* limit\_ad\_tracking: "flag to discover if the user limited tracking",
* adgroup: "group name of the ad",
* creative: "name of the ad shown",
* idfa: "unique device identifier assigned by Apple to a user iOS device",
* idfv: "unique identifier of device within vendor. Assigned by Apple to all apps on a single device that are from the same publisher/vendor"
* user\_id: “internal id generated by the JustPlay backend to track a user”
* country\_code: “the country code that Adjust attributes to this particular user/install event”
* os\_name: “Corresponds to the platform of the user, either android or iOS”

For revenue:

* event\_id: "unique event id",
* user\_id: “internal id generated by the JustPlay backend to track a user”
* source: "3rd party tool that is used to convert user's 'attention' to money.",
* amount: "revenue money in usd",
* network: "network name. Only for Applovin revenue",
* ad\_unit\_format: "format of the ad unit. values like banner, rewarded video, interstitials. Only for Applovin revenue",
* package\_name: "package name of the game where the revenue was generated",
* device\_id: “A column that groups device identifiers for users of different platforms. For Android users, the value shown here corresponds to the AppSetId, and for ios users it corresponds to the IDFV”
* advertising\_id: "A column that groups advertising identifiers for users of different platforms. For Android users, the value shown here corresponds to the Google Ad ID (gaid), and for ios users it corresponds to the IDFA",
* ad\_unit\_name: "Name of the applovin MAX Ad Unit from which the revenue was generated",
* Country\_code: “the country assigned to where the revenue was generated. Returned by the network”

Some resources to understand what Adjust does and what attribution means:

* [https://www.adjust.com/glossary/adjust/](http://goog_1537731217/)
* [https://www.adjust.com/glossary/advertisement/](http://goog_1537731217/)
* [https://www.adjust.com/blog/why-do-you-need-mobile-attribution/](http://goog_1537731217/)
* <https://www.adjust.com/glossary/attribution/>