



Almedia's Data Analytics Case

Almedia is looking for data scientists, analysts, and engineers to start building its data team.

This case aims to provide some context as to what the company's day to day challenges are (within data), whilst allowing the candidate to demonstrate their knowledge, skills & creativity.

You will be presented with a database simulating real data, as well as insights into the company's business model and operations.

Your Task

Understand the data, identify business opportunities, and to derive actionable insights. Using data driven solutions to optimize the workflow and generate value to the company.

While you are free to present your results in any format, ensure that your presentation includes context around problem framing, reasoning, and any relevant inputs. This will help a neutral yet technical evaluator assess your solution more effectively.

Once completed, please email your TA partner with it attached. We will do our best to respond within 5 working days.

Be creative, and good luck!



The Company

Almedia's main product is [Freecash](#)

Freecash allows users to earn real money by playing games, participating in surveys and promotions.

On signup, the user can explore an offerwall, pick and choose an offer of an app they might like, start playing & earning.

Each offer has customized tasks, which when achieved, reward the user with a certain amount of coins, which in turn, can be used to cash out \$\$ (1000c → 1\$).

Companies advertise with us, because we are able to provide quality customers at a competitive rate compared to traditional user acquisition channels.

In short, our business model is:

1. We spend money to acquire new users.
2. Some of these users engage in our partner's offers, installing their apps.
We get paid for those users.
3. Users may achieve certain goals, in which case they get rewarded by us with coins.
Later on, they may exchange those coins for money.

While some goals are easily achievable (such as installing an app), others might require the user to dedicate a considerable amount of time, or even to make in-app purchases. Both of which are metrics our partners try to optimize for when considering which acquisition channel to invest in.



Departments & Processes

Currently, the company is experiencing rapid growth, with most of its 50 employees having recently joined. Departments, teams, and processes are undergoing frequent revisions to enhance and streamline our operations.

Data initiatives are not limited to specific areas, but there is a significant focus on operations and product, primarily optimizing offers and related tasks.

In essence, we want to focus on determining which offers to present to which users, the tasks included in those offers, and how to best structure incentives for these tasks to boost user engagement and meet our and our advertisers' performance goals.

Most of those processes so far have been done manually, to great success of the company - but we understand that there is still significant margin to be gained by leveraging data.



The Data

The data shared is a subset of a replica of our production database.

While we are not sharing all of its content, and anonymizing part of it, the provided data should suffice for a well versed candidate to explore different hypotheses, build, and propose solutions.

Entities shared are:

- **actions.csv:**

The actions taken by users within an app during a given time period. Data for one game is shared for a period of 3 months. Completion rates can change over time (f.e. The game could have become more difficult or easier, or changes in rewards, offer display and other factors might lead to different user behavior).

Please download the CSV from the following sheet: [Actions.csv](#)

The following columns are available:

- Action_id - id of an action
- Install_id - id of an install
- User_id - user id
- action_type - can be one of the following (progress, install, purchase). Progress means that the user reached the next level in the game, purchase - user did an in-app purchase.
- Event_name - name of the action (e.g. new level is stored in event_name)
- Action_value - monetary value of an action (relevant for in-app purchases)
- Action_dt - date of action
- Mobile_os - user device's OS, takes values A or B

Offer Tasks:

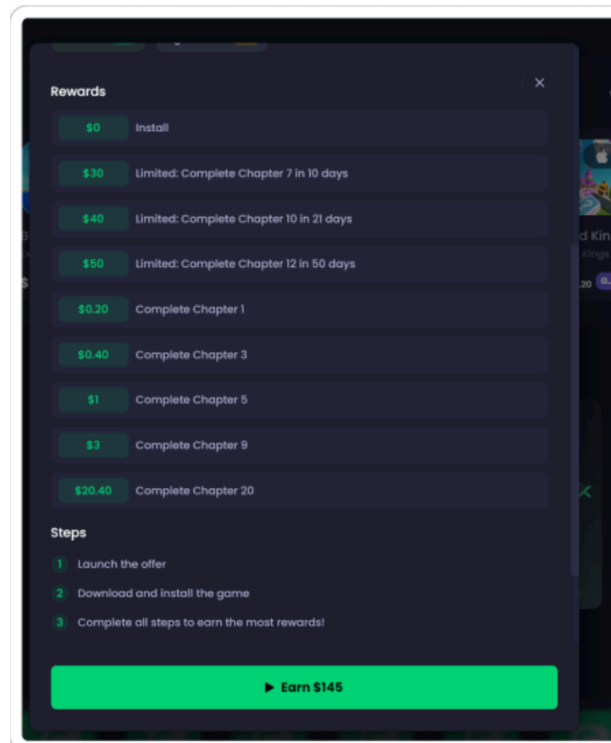
Assume that user rewards follow the following structure: Almedia's revenue is 10\$ for each install and users get \$10/20/40 for accomplishing the following tasks:

Task Name	Revenue	User Reward
Install	\$10.00	0
Reach Level 35 in 10 days	-	\$10
Reach Level 100 in 20 days	-	\$20
Reach Level 175 in 50 days	-	\$40



Users need to achieve each task by the date limit (from offer start) in order to be rewarded. If a user doesn't complete the first task in time, they can still complete the other ones.

For the user it would look something similar to:



Sample Offers' Task Structure



Objective

Given our business model - *The Company* section - and the **data shared**, please answer the following questions:

- Based on the revenue and reward structure defined above and the purchase data in `actions.csv`, calculate the total revenue, total rewards earned and the total amount spent (on in-app purchases) for each unique user.
- If you were an analyst at Almedia, which 4-5 specific KPIs would you prioritize tracking to monitor overall performance? Justify why each chosen KPI is relevant.
- Examine the provided data to identify significant trends or insights related to user activity and business performance. Provide 2-3 key observations concerning user behavior patterns, level completion rates, unit economics and in-app purchase behavior.
- Using the available data, propose a user segmentation and explain how segmenting users in this way could help inform targeted marketing campaigns, optimize in-app offer structures, or refine the reward payout system for better efficiency or engagement.
- Compare the user behaviour between users on operating system A versus operating system B with regards to KPI metrics defined earlier. Specify the statistical tests and/or visualizations you would use to support your analysis and validate whether any observed differences are statistically significant.

We are looking for candidates to perform a suitable analysis of the data provided. Bear in mind, demonstration of **business understanding** and **narration** will be preferred over the implementation of submissions with inaccurate complex techniques which don't deliver the aforementioned.

Consideration of both business or technical stakeholders is important, however please primarily focus on technical stakeholders for now. Present your results accordingly.

While we predominantly use Python and SQL for ideation, you are free to consider any language / technology for the proposed solutions.

We expect the solution to be in the form of Jupyter notebook and slides (for panel interview with main results of the analysis). If your solution satisfies our minimum criteria, you will be invited to a follow up interview where you will present and discuss your solution. Solutions that do not satisfy our criteria, but that show potential, will receive an in-depth feedback.

Have fun & good luck