VERTIGO GAMES
DATA ANALYST CASE

Description

Thank you for your interest in Vertigo Games. To better understand your analytical thinking, problem-solving, and technical skills, we have prepared the following case study.

The study consists of two parts: Task 1 involves A/B test modeling and simulation, and Task 2 involves exploratory data analysis. Please attempt to complete both tasks.

Expectations & Deliverables

We expect you to complete this case study as you would approach a project in a professional work environment:

1. **Clean & Reusable Code:** You may use your preferred tool for the analysis (e.g., Python/Jupyter Notebook, R, SQL). It is important that your code is clean, well-commented, and easily reproducible by others. Do not write the code in only one file. Separate it to reusable elements.
2. **GitHub Usage:** We request that you share your solution via a public GitHub repository.
3. **README.md File:** Please include a README.md file in your repository that summarizes the purpose of the analysis, your methodology, assumptions you made, and your key findings. (Task 1 - a: winner variant is A or B)
4. **Screenshots:** Add screenshots of your key findings or visualizations (e.g., graphs) to your README.md file.
5. **Release Version:** Tag the final, completed version of your work as v1.0 on the "Releases" page of your GitHub repository.

Task 1

Assume we are running an A/B test on a difficulty flow, and each variant receives exactly 20,000 installs per day. Write the codes to analyze and create graphics.

Metric	Variant A	Variant B
Daily Purchase Ratio	3.05%	3.15%
eCPM	$9.80	$10.80
Ad Impression/DAU	2.3	1.6
D1 Retention	53%	48%
D3 Retention	27%	25%
D7 Retention	17%	19%
D14 Retention	6%	9%

a) Which variant will have the **most daily active users** after 15 days?
b) Which variant will earn the **most total money** by Day 15?
c) If we look at the **total money earned by Day 30** instead, does our choice change?
d) What if we run a 10-day sale starting on Day 15 (boosting everyone's purchase rate by 1%)? Does this change which variant earns more **total money by Day 30**?
e) On Day 20 we add a new user source. From then on, we get 12,000 users from the original source and 8,000 from this new one. The new users' retention is described by these formulas. With this mix of old and new users, which variant makes more **total money by Day 30**?

- **Variant A (New):** $Retention = 0.58 \cdot e^{-0.12(x-1)}$
- **Variant B (New):** $Retention = 0.52 \cdot e^{-0.10(x-1)}$

f) Which one should you prioritize, and why? If you could only make **one** of these improvements:

1. Run the **temporary 10-day sale** (from d)
2. Add the **new, permanent user source** (from e/f)

Task 2

Using the data set provided to you, we want you to make analyses about the different aspects of the users' characteristics. For example, you can consider segmenting users based on their first-day engagement.
Other options might be: Are there any trends in session duration over time? Do sessions tend to get longer or shorter as time goes on?
We ask you to be more creative and conduct an analysis that uncovers such segmentations or trends. You can use any tool to present and visualize your findings.

Dataset provided shows a summary of the in-game and monetization activities performed daily by the players in our game, including columns and data types below;

Field name	Type
user_id	STRING
event_date	DATE
platform	STRING
install_date	DATE
country	STRING
total_session_count	INTEGER
total_session_duration	FLOAT
match_start_count	INTEGER
match_end_count	INTEGER
victory_count	INTEGER
defeat_count	INTEGER
iap_revenue	FLOAT
ad_revenue	FLOAT