✓ **LevelUp Milestone** | GameJet Transactions

INTRODUCTION: In order to make the most of a given dataset, you may need to perform engineering of additional calculated fields. Performing calculations to transform the data lets an analyst extract new information that could not be obtained from the data as originally stored.

HOW IT WORKS: Follow the prompts in the questions below to investigate your data. Post your answers in the provided boxes: the **yellow boxes** for the queries you write, and **blue boxes** for text-based answers. When you're done, export your document as a pdf file and submit it on the Milestone page – see instructions for creating a PDF at the end of the Milestone.

RESOURCES: If you need hints on the Milestone or are feeling stuck, there are multiple ways of getting help. Attend Drop-In Hours to work on these problems with your peers, or reach out to the HelpHub if you have questions. Good luck!

PROMPT: In this Milestone, you'll help a mobile game company summarize and describe some general patterns in the app usage and spending for one of their games. Revenue from this game is supported by a traditional microtransaction model, where users can spend real money on in-app purchases for items that allow players to access more game features.

Monetization is a difficult problem for companies in the mobile games space, and they need to balance converting as many users from free users to paid users, without being so aggressive that they actually end up driving away more users and actually make less revenue. But before they can start developing strategies for improving the game from a business perspective, the team needs your help to summarize what has been done so far!

— Data Set **Description**

The data for this Milestone (`game_jet.*`) describes activity for users of a mobile-based game with microtransactions. In the game, two currencies called 'passes' and 'gems' are used to unlock more game content and to purchase cosmetic items. These currencies can be earned slowly over regular play, but gems can also be purchased as in-app purchases. You will be working with three tables in this Milestone: `users`, `sessions`, and `iaps`.

The `users` table lists users who have downloaded the game app, one per row. The table contains six columns, of which include:

- **udid** – unique id for the user / device
- **install_date** – date of user installing the app
- **lang** – two-letter code for native language for device
- **country** – two-letter code for user's country

The `sessions` table records every session taken by a user where they opened the app. There are four columns in this table:

- **udid** – user / device id
- **ts** – timestamp for start of session
- **date** – pre-truncated date for session
- **session_num** – cumulative session number for the user

The `iaps` (In-app purchases) table records all purchases made by users, exchanging real money for in-game currency. There are six columns in this table:

- **udid** – user / device id
- **ts** – timestamp for purchase
- **date** – pre-truncated date for session
- **prod_type** – product type, can be gems, passes, or a value pack.
- **prod_qty** – quantity code that is proportional to each product (e.g. a quantity of 1 for gems might actually represent 10 gems)
- **rev** – revenue / price of the purchase, in cents

— Task 1: App usage behaviors

How long does a user spend with the app in terms of the number of sessions or the days that they stick with the app? This can have implications for how much time we have to 'hook' a player and give them a reason to spend money on the app.

A. Query the `users` data table to discover the total number of users represented in the data.

(paste your query below 👇)

```
SELECT COUNT(udid)
FROM game_jet.users

If you're unsure about the user ids being unique, you could SELECT
COUNT(DISTINCT udid) instead.
```

(write your **answer** below 👇)

There are 22 576 users in the dataset.

B. Next, write a query to count the number of distinct users in the sessions data table. Recalling your answer from Task 1A, how many users downloaded the app but didn't actually open it? HINT: A user who downloaded the app will always add a tally in the users table, but won't show up in sessions if they never use the app.

(paste your query below 👇)

```
SELECT COUNT(DISTINCT udid)
FROM game_jet.sessions

The DISTINCT keyword is required in this case.
```

(write your **answer** below 👇)

> There are only 22 544 unique users in the sessions table, which
> means there were 32 users that did not open the app at all.

C. Write a query that returns the number of sessions made by each user. Sort the
output by the number of sessions made, from largest to smallest. How many
sessions did the user with the most sessions have with the app?

(paste your query below 👇)

```
SELECT
    udid,
    COUNT(*) AS n_sessions
FROM game_jet.sessions
GROUP BY udid
ORDER BY n_sessions DESC
```

(write your **answer** below 👇)

> The user with the most sessions used the app 1939 times.

D. We should also be interested in how many sessions were taken by a 'typical'
app user. The average number of sessions with the app is 32.1. Modify your
query from part C to only return users who have made more than that many
sessions. How many users are there? (Read this from the SQL app interface.)
Based on that number, how well does the average represent a 'typical'
player? HINT: Your query should include the HAVING keyword here!

(paste your query below 👇)

```
SELECT
```

```
    udid,
    COUNT(*) AS n_sessions
FROM game_jet.sessions
GROUP BY udid
HAVING COUNT(*) ≥ 33
ORDER BY n_sessions DESC
```

(write your **answer** below 👇)

There are 5262 users who have logged at least 33 sessions, or about 23% of all users who opened the app. This means about 77% of users logged fewer than the average. This is an indication that 33 sessions is actually quite a bit more engaged than what we might really want to consider as a 'typical' player.

If you look into the data further, you'll find that the median number of sessions with the app is 8. The median is the point that divides the data in half; about half the users spend 7 or fewer sessions with the app, and the other half 8 or more. That's pretty far from the average, which gets pulled up by those few users with hundreds of sessions!

— Task 2: In-App Purchases

How large is our user base in terms of who is spending money? Is it a lot of users spending a little, or is it a few users who are spending a lot? Different types of users will likely be responsive to different types of tactics to get them to spend more money.

A. Using the `iaps` table, write a query to determine the number of users who have made at least one purchase. Remember that each time a user makes a purchase, a new row is added to the `iaps` table.

```
SELECT COUNT(DISTINCT udid)
FROM game_jet.iaps
```

B. Both tables share values in the `udid` column. Write a query that left joins the `iaps` table to the `users` table and returns the total (sum) amount of money spent on in-app purchases, grouped by user id. Your query should return every user in the users table, regardless of whether or not the user made a purchase; the total spent for users with no record in the `iaps` table should be a *null* value.

```
SELECT
   u.udid,
   sum(i.rev) as total_spent
FROM
   game_jet.users AS u
   LEFT JOIN game_jet.iaps AS i
     ON u.udid = i.udid
GROUP BY u.udid
```

C. Modify your query in part B to create a new feature called **persona** that segments users into four purchasing personas:

- a "free player" who does not spend any money on in-app purchases,
- a "minnow" who spends less than $20 on in-app purchases,
- a "dolphin" who spends between $20 and $100 on in-app purchases,
- and a "whale" who spends at least $100 on in-app purchases.

Remember: revenue is recorded in terms of cents, so $20 is equal to 2000 cents and $100 is equal to 10 000 cents.

(paste your query below 👇)

```
SELECT
  u.udid,
  CASE
    WHEN sum(i.rev) is NULL THEN 'free player'
    WHEN sum(i.rev) < 2000  THEN 'minnow'
    WHEN sum(i.rev) < 10000 THEN 'dolphin'
  ELSE 'whale' END AS persona,
  sum(i.rev) as total_spent
FROM
  game_jet.users AS u
  LEFT JOIN game_jet.iaps AS i
    ON u.udid = i.udid
GROUP BY u.udid
```
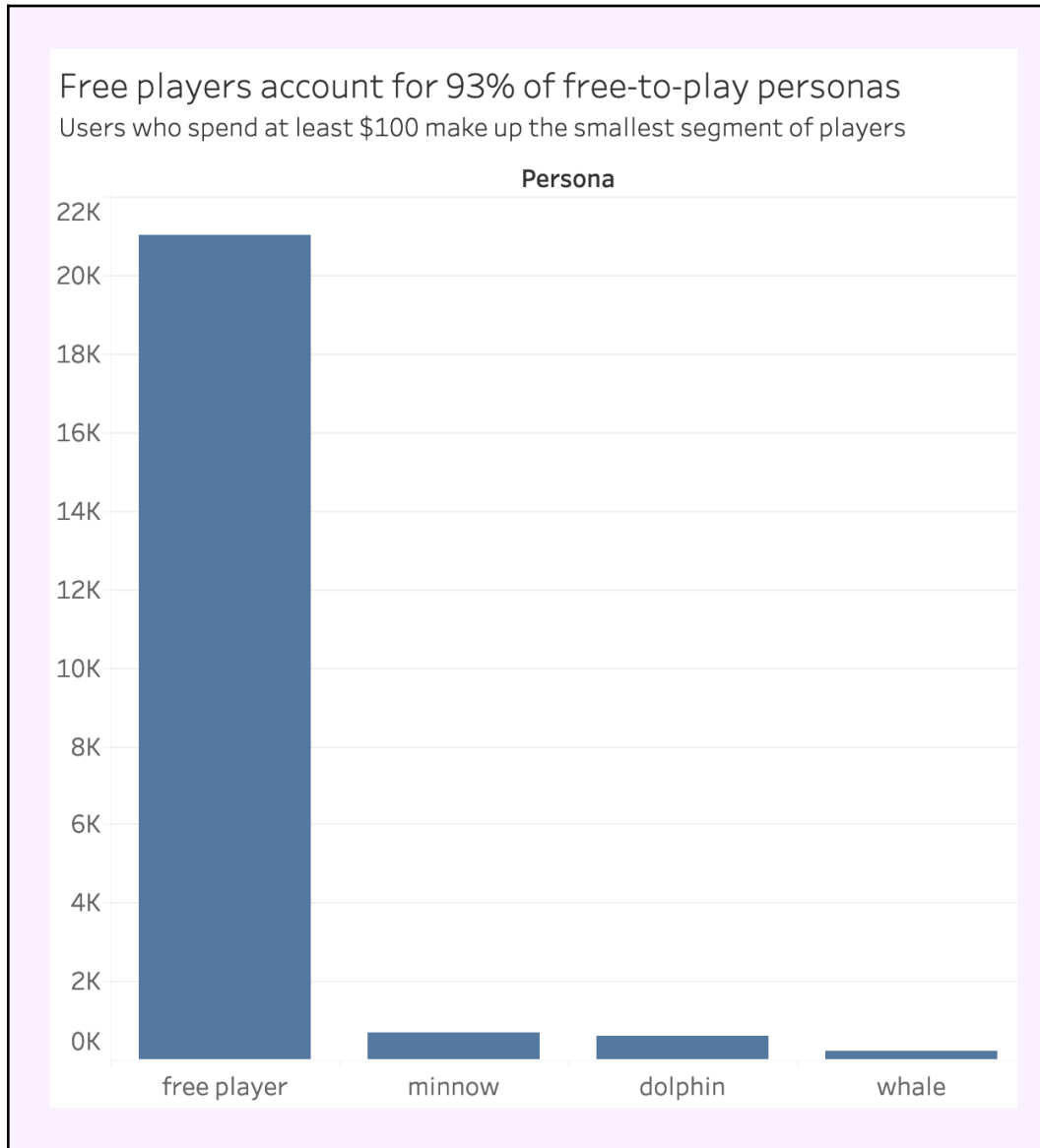
— Task 3: Analyzing Free-to-Play Personas

The remainder of this Milestone will be completed in a Tableau Workbook where we can use visualizations to aid in our analysis. **Click this link to navigate to the workbook you'll use to complete the remainder of this Milestone.** Once you've published your Tableau Workbook in the folder named Upload Workbooks Here, paste the Share Link in the box below.

SHARE LINK

Continue to post your answers in the provided boxes: purple boxes for your visualizations, and blue boxes for text-based answers.

A. On Sheet 1, create a bar chart to illustrate the total number of each purchasing persona. Which persona accounts for the majority of users?

HINT: There are a few ways to do this! One suggestion is to drag the `Persona` pill to the Rows AND Columns! But convert the Columns pill to be a *measure*. You'll want the count here.

Free players account for 93% of free-to-play personas
Users who spend at least $100 make up the smallest segment of players

Persona

(write your **answer** below 👇)

Over 21 000, or about 93% of users in our dataset are 'free-players', making it by far the largest segment of free-to-play personas. The 'whales' – users who spend the most

money individually – account for the smallest share of free-to-play personas, less than 1%.

B. Create a second bar chart to illustrate the total amount of money spent, grouped by each purchasing persona. Since free players don't spend any money on in-app purchases, filter your visualization to show only the total revenue generated by minnows, dolphins, and whales. Which persona accounts for the majority of in-app purchase revenue?
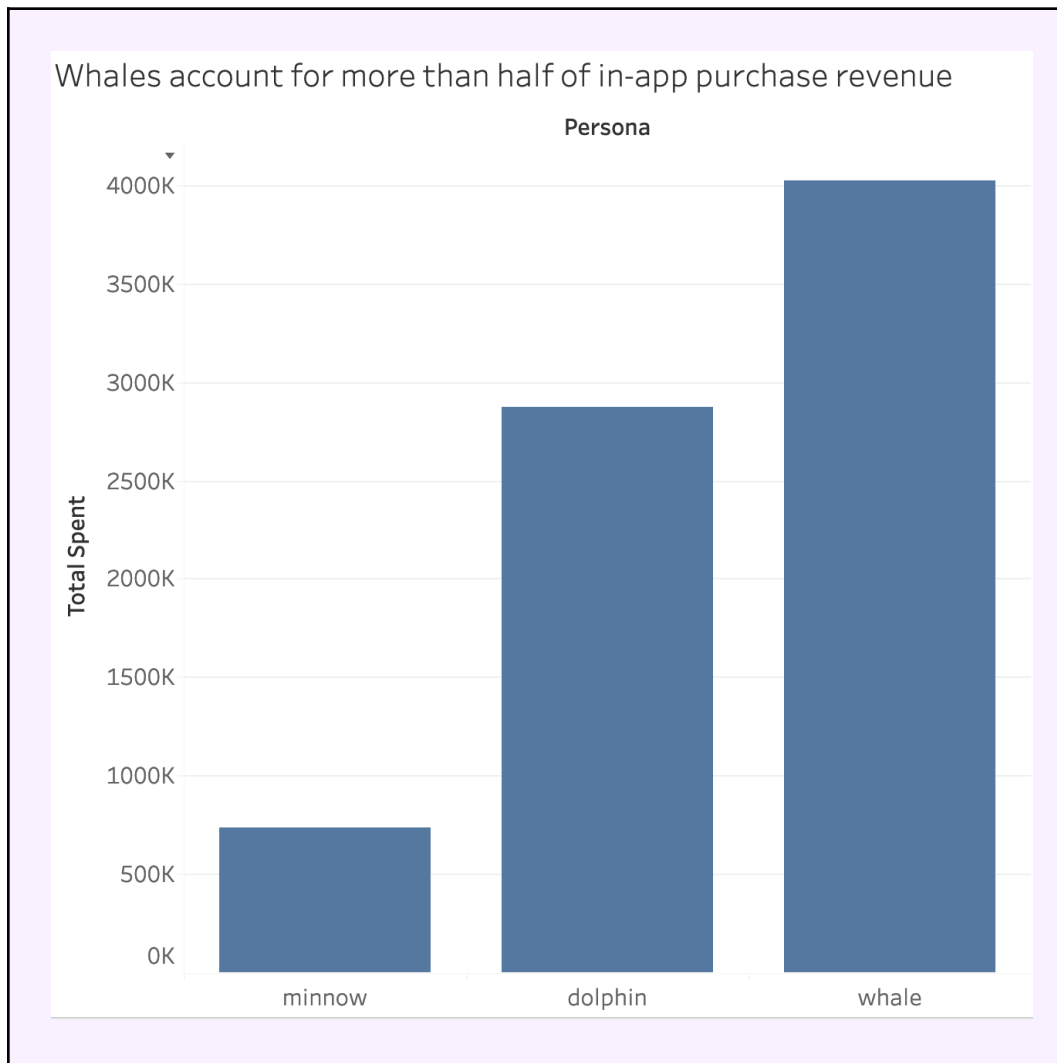
Whales account for more than half of in-app purchase revenue

(write your **answer** below 👇)

C. You might notice in this last graph that your y-axis is pushing 4 000 000! But the units are cents, not dollars! Can you create a calculated field in your Tableau workbook that converts the total amount spent from cents to dollars and use this field to adjust your visualization? HINT: $1 = 100 cents.

(write your **answer** below 👇)

Your calculated field total_spent_dollars should be
[Total Spent]/100

— Task 4

The answers to the questions in the Milestone tasks above should paint a picture of the game app being mostly supported off of a relatively small, dedicated group of players. Most users only engage with the app for a short amount of time and end up not buying anything. Even a fair amount of users who do buy something don't make too many purchases.

Another angle of attack you might be interested in looking at is how long it takes for a user to make their first purchase, if they make any. You can do this in two steps. First, take the difference between the user's install date and the date of each purchase they made to get the amount of days between when they started playing, and when they made a purchase. Their first purchase will then be the purchase with the smallest difference in dates. How many users make a purchase on the first day they have the app? Are there any other periods where we could possibly entice game players to make purchases?

(paste your query below 👇)

```sql
SELECT
  u.udid,
  MIN(i.date - u.install_date) AS
days_to_first_purchase
FROM game_jet.users AS u
INNER JOIN game_jet.iaps AS i
  ON u.udid = i.udid
GROUP BY u.udid
HAVING MIN(i.date - u.install_date) = 0
```

(write your **answer** below 👇)

724 users, or about **47%** of users who made at least one
purchase, did so on their first day of using the app.

— Submission

Great work completing this Milestone! To submit your completed Milestone, you will
need to download / export this document as a PDF and then upload it to the
Milestone submission page. You can find the option to download as a PDF from the
File menu in the upper-left corner of the Google Doc interface.