Project Title: Best-Selling Video Games Genres on Steam
Analyst: Zachery Riordan

ASK

The first step to any good analysis would be to determine the questions you are trying to answer. For me, those questions are:

- Which Genres have the highest total sales?
- How have sales trends changed over time?
- Can we use these trends to predict future best-selling genres?

Identifying stakeholders is also an important part of this step. For this analysis stakeholders may look like this:

- Game Developers
- Publishers
- Marketing teams

PREPARE

For the second part of my analysis, I needed to prepare for the study, which included finding a dataset. I chose to use the Steams Store Games Dataset available on Kaggle by author Nik Davis. It is a little dated but contains information on a broad variety of games and can be used to discover trends that may still be useful to this day.

PROCESS

For this step of my analysis, I have loaded the dataset into BigQuery to run SQL queries on to obtain the information I needed.

I then took some steps to clean the data a little bit, by splitting genres to avoid games with multiple genres counting for only one, and allowed it to estimate owners. I also wanted to remove any null values. This is the code I used:

```
WITH cleaned AS (
  SELECT *
  FROM `glass-turbine-458018-t8.steam_analytics.steam_data`
  WHERE owners IS NOT NULL
    AND owners != ''
    AND genres IS NOT NULL
    AND genres != ''
    AND release_date IS NOT NULL
),
owners_fixed AS (
  SELECT
```

```
    *,
    (
      CAST(REPLACE(SPLIT(owners, '-')[OFFSET(0)], ',', '') AS INT64) +
      CAST(REPLACE(SPLIT(owners, '-')[OFFSET(1)], ',', '') AS INT64)
    )/2 AS owners_estimate
  FROM cleaned
),
split_genres AS (
  SELECT
    appid,
    name,
    release_date,
    owners_estimate,
    price,
    TRIM(genre) AS genre
  FROM owners_fixed,
  UNNEST(SPLIT(genres, ';')) AS genre
)
SELECT *
FROM split_genres;
```

Analyze

With my cleaned dataset I am now able to analyze the new data to find the answer to my question of which of these genres are the best sellers?

In order to accomplish this I ran this query:

```
SELECT
  genre,
  SUM(owners_estimate) AS total_owners,
  COUNT(DISTINCT appid) AS game_count
FROM `glass-turbine-458018-t8.steam_analytics.steam_cleaned`
GROUP BY genre
ORDER BY total_owners DESC
LIMIT 10;
```

This gave me the top 10 genres/combination of genres that sell the best on Steam.

SHARE (VISUALIZE)

I now needed to prepare the data for visualizations using Google Sheets. To do this I ran a query to give me some cleaned data for the most recent 5 years of the table to allow me to make a line chart:

```
WITH filtered_years AS (
  SELECT
    EXTRACT(YEAR FROM release_date) AS year,
    genre,
    SUM(owners_estimate) AS total_owners
  FROM `glass-turbine-458018-t8.steam_analytics.steam_cleaned`
  WHERE EXTRACT(YEAR FROM release_date) BETWEEN 2013 AND 2018
  GROUP BY year, genre
),
top5_genres AS (
  SELECT genre
  FROM filtered_years
  GROUP BY genre
  ORDER BY SUM(total_owners) DESC
  LIMIT 5
)
SELECT
  f.year,
  f.genre,
  f.total_owners
FROM filtered_years f
JOIN top5_genres t
  ON f.genre = t.genre
ORDER BY f.year ASC, f.total_owners DESC;
```

I made a visualization using this data on Google Sheets linked below:
Genre sales by year

ACT

With the data gathered and analyzed using these steps, I have determined that the best selling genres on steam seem to be Action, Indie, and Adventure. It seems that Action games perform well, however there is a lot more competition in that genre. Indie games have shown significant growth in recent years with less market saturation so this genre may sell better if the popularity continues on this trend. Combining genres also seems relevant, so combining Action elements into an Indie game may perform well.

Reflection

This project gave me hands-on experience applying the full data analysis process—from formulating questions to presenting actionable insights. I used SQL in BigQuery to clean, transform, and analyze a real-world dataset, which improved my skills in writing efficient queries and handling messy data. I also used Google Sheets to create visualizations that helped communicate trends clearly, and these visuals were key to drawing meaningful conclusions. Through this project, I gained a deeper understanding of how data can drive decision-making, especially in dynamic industries like gaming. Overall, this capstone reinforced my confidence in working with data tools and strengthened my ability to extract insights that can inform strategic business decisions.

Tools Used

SQL (BigQuery): Used for Processing and Analyzing

Google Sheets: Used for Visualization