Our Scenario:

We were hired as data scientists by a small startup game production company that has a budget of $150k to spend on their first Android mobile game. To maximize return on their investment, they want to know what type of game to develop, and how much they should charge to maximize profit (via paid downloads).

Our Hypothesis:

There is an optimal game category and price point, which will maximize return on investment.

Our process:

Google doesn’t have an API to get the information from the Play Store directly, which made things a bit difficult. However, we got lucky and found a Python library on PyPi called, “Play Scraper”. It functions as an API, and combs through the Google Play Store by using variables in Python. Luckily, it was able to get the detailed game information that we needed to deliver a conclusion to our ‘employers’.

Steps we took:

1. Installed the Play Scraper library.
2. Created a loop that would find the top 25 paid games in all gaming categories on the Google Play Store.
3. We took the app/game ID’s from the top 25 games in each gaming category, and pulled the following information for our research:  
   1. The game’s Title
   2. The game’s rating
   3. The amount of reviews for each game
   4. The price of each game
   5. The amount of installs each game has (Google groups them in increments of 5k)
   6. The game’s developer
   7. The link to the game on the play store for reference
4. We then took all the information and put it into a data frame so we could access it and answer the following questions:  
   1. What is the optimal game category (Arcade, Action, Racing, Simulation, etc)?
   2. What is the optimal price point?
   3. Does the game’s rating score have an impact on the game’s success?

Our limitations:

1. >90% of all games on the Google Play Store are free. They make revenue via ads and ‘microtransactions’ aka in-game purchases. Because we don’t have access to the data on revenue generated from free games, we were limited to coming to conclusions based on available information.
2. Lack of an API made it impossible to get an exact download amount for each game.

Our findings:

1. We were surprised to find that there was a significant outlier among the games, which, by itself, contributed to more than double of any other CATEGORY. This outlier was Minecraft, which had a total revenue of over $69,000,000. We then decided to make Minecraft its own category to compare it with the rest of the data instead of excluding it completely because there is a possibility that any game could experience a similar phenomenon.
2. The highest grossing category was Action games.
3. The highest grossing price point was $4.99

Our conclusion:

1. The data showed that games in the Action category and $4.99 price point performed the best overall, and grossed the highest.