**Data Analysis Portion:**

You will have two files to analyze with this final, “April\_Channels\_2021.xlsx” and “Twitch\_2018.xlsx”. Each file has their own set of questions to find. For full credit, you will need to submit the answers the questions in an ipynb file with the pandas module.

**Point Deductions:**

* 5% will be marked off if pandas module is not utilized
* 5% will be marked off if ipynb was not utilized
* 20% will be marked off if Python was not utilized

**Questions:**

April\_Channels\_2021.xlsx

1. Who had the highest value for:
   1. Time Streamed
   2. All-Time Peak Viewers
   3. Hours Watched
   4. Total Followers
   5. Total Views
2. Who had the most Followers Gained and the most Followers Lost during the month of April? I
3. See if there is a correlation between Time Streamed and Followers Gained. It would be interesting to see if the more time streamed means the more followers gained. Hypothetically speaking, these channels are at the top of Twitch, so I would expect there to be very little correlation.  You will need some type of chart to support your answer.
4. Do average viewers multiplied by time streamed equate (roughly) to hours watched? I would believe that this calculation is true, but it would be interesting to see if there are any outliers. For instance, if someone has a really high average viewer value, but a low time streamed, their hours watched may not be equal. Please make sure to support your answer
5. Supposedly TheGrefg holds the highest record for all-time peak viewers at 2.47 million concurrent viewers? This is over 3x the next highest peak viewership… Let’s prove it by building a bar chart of the top 10 channels and their All-Time Peak Viewers
6. What was the total time people streamed on Twitch?
7. What was the total hours watched on Twitch?
8. Who had the highest amount of followers during the month of March?

Twitch\_2018.xlsx

1. Which games made the top 500 every month of 2018? (Which games had values for each month in the Peak Viewers Per Month tab?)
2. Which game had the Max Viewers for each Month?
   1. Output Example:  
      January: Game  
      Februrary: Game

March: Game  
etc etc etc

1. According to this data, which game series was more popular during the year of 2018:
   1. Call of Duty
   2. Battlefield
   3. FIFA
   4. GTA
   5. Mario
   6. Pokemon
   7. Final Fantasy

\*\*\* Use a Pie Chart to support your answer \*\*\*

1. Which Lego game was the least popular throughout 2018?
2. Which month had the most viewers?
3. We are looking for which month has the highest number of viewers. Is there a reason why that month would be higher than the rest? If there isn’t please prove it. Suggestion: use a graph to show which month has the highest volume of viewers.
4. What is Pokémon: Let’s Go, Pikachu!/Eevee! ?
5. Is your favorite game on this list?
6. If you graph out the Peak Viewers Per Month you will have a graph like the one on the main tab. Recreate the graph in Python so it only shows the top 10 games instead of the first 255 games.

**Notes & Hints:**

* If you do not use Python to answer the questions, you will need a nice write up using a Word Doc (Google Docs) and Excel (Google Sheets) file.
* Doesn’t matter if you used Python or Excel, your write up needs to be clear on what answer goes with what question. A great format would be to write out the question, show your work, state the answer.
* Easier it is for me to see the proof of your work, the higher your grade will be.
* When you submit the files, make sure to include your data files as well. I am going to pull the files down and then run whichever file you name mainLastName.py
  + For example:
    - mainLastName.py
    - clean2018.csv
    - cleanApril.csv
    - LastNameWord.docx (This will have all of your answers and graphs)
    - mainLastName.ipynb
* All graphs need: Title and axis title
* If a legend will help understanding the graph, please add one
* Go back and look at your old notes
* Use Excel and Sheets to help you find the answer then prove it in Python.
* April\_Channels\_2021.xlsx: This data represents the top 500 channels based on their rank for the month of April 2021.
  + A channel's rank is an aggregation of its average concurrent viewers, followers, views, and stream time over the month
* Twitch\_2018.xlsx: This data is slightly older, but a lot better for doing analysis on a month-by-month basis.
  + Within the csv file, there are multiple sheets (one for each month) that represent the top 500 games/categories (based on watch time) for each month. It is worth noting though that because of trends and how popular games are, there are games that are in some months but not others because they didn't make the top 500 cut.
  + Since this dataset is a little more verbose/complex, I went ahead and threw together the "Peak Viewers Per Month" sheet, as an example. This takes every game from all 12 months and gets the peak viewers for each month for that game. Some games ("60 Parsecs!" for example) only have values for a few months, because they only made the top 500 list for those months.