Final Project Phase I

Guidelines for Phase I Submission: For open ended questions, 1-3 bullet points should suffice for most answers. You do not need essay-length answers; however, there needs to be enough information, so that we can understand your topic and confirm that you have a cohesive and feasible topic. Make sure your answers are brief, but cohesive and answer all of the questions.

NOTE: Most of the points lost in this phase are due to not reading the instructions. Please make sure to read each question in its entirety.

Q1. Topic - 15 points

Please provide an overview of what your topic is going to be.

Q1.1 - 5 points

What topic have you chosen for your Final Project? If the scope of your project is too narrow, you are subject to lose points.

Answer: I will be researching the pick and ban rates, winrates and viability of recently released characters and items in League of Legends and compare/contrast them with older characters and items in the game, based on the performances of professional players that use them. In doing this, I hope to be able to identify if feature creep (a phenomenon in which newer features in a program are more "overloaded" than older ones) exists in the game, or if the whole issue is overblown.

Because there is a large spectrum of skill levels when it comes to LoL players, I plan to investigate pro players' games, but I may choose not to do this if I cannot find enough data.

Q1.2 - 5 points

Why did you choose this specific topic and what are you looking to learn from the analysis?

Answer: I chose the topic because I spend a lot of time playing League of Legends, or LoL as I will refer to it in this document. LoL is a multiplayer online game that has seen hundreds of changes since its release in 2009, and many of these changes have been seen as questionable by a large portion of the playerbase. One major bone of contention is the concept of "old vs. new"; many players believe that newer items and characters in the game are massively overpowered compared to those released back in the day. However, is this true or is this a case of nostalgia bias as a result of familiarity with older features and a lack of it with the newer ones? In analyzing the data, I hope to be able to draw conclusions on

whether or not the people that have been complaining about new features in the game are actually justified in their discontent.

Q1.3 - 5 points

Explain some of the concrete insights you expect to gather from your data and/or hypothesis you expect to answer.

Answer: Questions I hope to be able to answer with my research include but are not limited to:

- What items are more likely to lead to won games? Because every item in the game is built by some characters and not others, I will balance these numbers by weighting each item winrate according to the number of characters that build that item.
- How does the length of a character's moveset correlate with their winrate? For example, a character that can slow other units, dash, and apply "crowd control" (LoL's special name for things like stunning and knocking units into the air, leaving them unable to act) has a 'longer kit' than a character that can only slow units.
- How does the number of revisions a character/item has received from their release to now correlate with their winrate?
- In LoL, items are purchased using in-game currency called "gold". All items have stats like attack speed or movement speed, and some items have active abilities, which are activated using a button, or passive abilities, which activate on their own according to the conditions of the game. Because an item with no abilities will have better stats than an item with one or more of them, LoL players sometimes refer to items according to their "gold efficiency", which is basically a measure of how good their stats are for the amount you are paying to get them. In general, items with high gold efficiency are considered to be better purchases than those with low efficiency. How has the average gold efficiency of an item changed with time?

Q2. Downloaded Dataset - 15 points

Please provided a brief overview of you downloaded dataset. This should demonstrated that you understand the data contained within the dataset.

Q2.1 - 2 points

Provide the link (url) to your downloaded dataset.

Answer:
https://drive.google.com/drive/u/1/folders/1gLSw0RLjBbtaNy0dgnGQDAZOHIgCe-HH

This Google Drive folder contains data from all proplay games of LoL since 2013. I will be analyzing the .csv file for 2022's games.

Q2.2 - 3 points

What are the dimensions of your downloaded dataset in terms of rows x columns and file size? Ex. 50,000 rows x 20 columns and 5.4mb. If your file is a .json file, state the file size (mb, gb, etc.).

Answer: 149,233 x 123 columns; 75.9 mb

Q2.3 - 5 points

Briefly discuss the structure of your dataset. For .csv or table type datasets list out the column titles you anticipate using and give examples of the data contained within. This is not binding. For json data map out the dictionary and give examples of the data contained within.

Answer: The dataset contains comprehensive data about every major game played in LoL's 2022 pro season. In particular, I think some of the most useful columns are:

- Total gold each player got
- Length of each game (as some characters are better in long games, while others are good in short games)
- The character each player picked
- The character each player banned

Q2.4 - 5 points

Please explain why you chose this specific dataset. How will this data be used in your analysis? Can insights be drawn from this data alone, or will it be combined with other data?

Answer: I chose this dataset because it was the most comprehensive data on pro games I could find. However, for some reason, the file does not include information on what items were purchased during the games despite this information being readily available from both the LoL eSports website and just watching the games. Because of this, I will need to find a way to generate information on what items each player in each match bought and use it with the .csv file.

Q3. Web Requirement #1 (Web-scrape or HTML) - 15 points

Please provided a brief overview of you downloaded dataset. This should demonstrated that you understand the data contained within the dataset.

Q3.1 - 2 points

Provide the link (url) to your downloaded dataset.

Answer: https://leagueoflegends.fandom.com/wiki/List_of_champions

I will be using this webpage and the related pages that can be accessed through the navigation table at the bottom of the page, under the "See also" header.

Q3.2 - 3 points

Briefly explain how you plan to retrieve the data from this source, including the necessary Python libraries/modules.

Answer: Because all the data I need is in a table, I can use BeautifulSoup4 to parse through the table and store the information in it in lists and dictionaries in Python.

Q3.3 - 5 points

Briefly discuss the structure of your dataset. For .csv or table type datasets list out the column titles you anticipate using and give examples of the data contained within. This is not binding. For json data map out the dictionary and give examples of the data contained within.

Answer: This webpage contains some basic info about the characters (referred to as "Champions") in the game. In particular, the useful info on this page includes the characters' names and release dates, as these are metrics I can use to judge how "old" or "new" a champion is.

The related pages include other info, like base stats and growth coefficients. I can use this info to judge how the distribution of stats for each character has changed as the game has progressed.

Q3.4 - 5 points

Please explain why you chose this specific dataset. How will this data be used in your analysis? Can insights be drawn from this data alone, or will it be combined with other data?

Answer: The tables on the wikipages I have linked will help me gather information about the characters themselves. However, I will need to link this data with statistics (e.g. winrates, pickrates, banrates) from other sources.

Q4. Web Requirement #2 (API or JSON) - 15 points

Please provided a brief overview of you downloaded dataset. This should demonstrated that you understand the data contained within the dataset.

Q4.1 - 2 points

Provide the link (url) to your downloaded dataset.

Answer: http://ddragon.leagueoflegends.com/cdn/13.4.1/data/en_US/item.json

Q4.2 - 3 points

Briefly explain how you plan to retrieve the data from this source, including the necessary Python libraries/modules.

Answer: Because I have developer access to LoL's API, I am able to make requests to the API using my private dev key and the requests module on Python. To parse item data, I will use the requests module to generate a copy of this json file in Python, and then use the json module and list comprehensions to get what I need.

Q4.3 - 5 points

Briefly discuss the structure of your dataset. For .csv or table type datasets list out the column titles you anticipate using and give examples of the data contained within. This is not binding. For json data map out the dictionary and give examples of the data contained within.

Answer: The json file contains data on all items in LoL, including ones that have been removed. I will be focusing on the data under the "data" header. Each object under this header represents an item and is named after that item's unique ID. Each object includes:

- the item's name
- the item's description
- what other items a player can build using that item
- the item's gold cost
- the item's stats

Q4.4 - 5 points

Please explain why you chose this specific dataset. How will this data be used in your analysis? Can insights be drawn from this data alone, or will it be combined with other data?

Answer: As I am planning on looking at items' stats and gold efficiency, I will definitely need this dataset during my analysis. On its own it probably won't tell me much, though, so I will combine it with my other findings.

Q5. Additional Datasets - 10 points

If you have found any datasets beyond the three required, please describe them below: (If you do not plan to use any additional datasets please simply write **N/A**)

Q5.1 - 5 points

Provide the links for any additional datasets you might use

Answer: I found some other websites I may or may not consult:

- https://leagueofitems.com/
- https://champsdb.gg/

Q5.2 - 5 points

Briefly explain how you will retrieve data from these sources, and how this data is going to be used for your analysis

Answer: These websites use the same API that I have access to in order to display info on items and characters respectively, but display them in a visually pleasing manner. Because of this, I have no real reason to scrape them (as I could just use the api instead), but I may reference them from time to time if I need something minor.

Q6. Inconsistencies - 15 points

Please list at least 3 inconsistencies you have found in your dataset, and how you plan to address each of them.

Answer:

1. In the .csv file, I noticed that some games only have partial data; most of the important stuff is there, but things like what objectives spawned that game, how many structures they destroyed, etc. are missing for some players in some matches. I don't anticipate this to be a serious issue, but I feel it may very slightly impact the quality of my analysis if I decide to use any of these values.
2. In the .csv file, some rows that are otherwise totally complete are missing player names, which could be as a result of the names containing illegal characters for Excel. Because of this, it could be difficult to research trends in a given player's games. However, there are not a lot of rows like this, so I will manually fill in the blanks by watching the games and finding out who is missing.
3. In the .csv file, some of the values seem to have been inputted incorrectly, as there are negative values in some columns in which it doesn't make sense for them to be there (for example, negative values for gold earned during a game). Again, there are not that many of these errors, so I can probably correct them myself.
4. On the webpage with a list of all champions in the game, all champions are listed according to the date they were originally put into the game. However, there are quite a few champions that received comprehensive updates, meaning that they are different in all but name, with completely different stats and abilities. For example, a character named Udyr was originally added in 2009, but received an update in 2022, and so should not be counted as a 2009 champion. For these champions, I will manually update the dates in my collection of data to reflect when these characters were changed.

Q7. About Your Analysis - 10 points

Provide a BRIEF list of steps of how you plan on performing your analysis and the way you will gather/present your findings. (Non-technical, high-level overview)

Answer: I will:

- Develop some sort of scale to judge how "old" or "new" a character/item is in terms of how many updates it has received or how long it has been in the game for
- Update the csv file to find out what items were bought in the games I choose to analyze
- Use Numpy to find champion winrates over all the professional games played in 2022
- Correlate this data with the items they built and how gold-efficient they were
- Present findings

Q8. About You - 5 points

Q8.1 - 2.5 points

List the names of each of the members of the group working on this project. If you are working alone, there should be one name listed. Failure to list your teammate and group them in Canvas may result in working individually on the project.

Team Member 1: Oluwatobi Adewoye

Team Member 2(If Applicable):

Q8.2 - 2.5 points

Each member of the group should initial below to indicate that you acknowledge this statement:

I affirm that all of the work in this project will be done my me/my team and is not duplicated from any other source. In addition, any references that I use or code that I choose to model after will be appropriately credited and referenced in my project.

Team Member 1 Initials: OA

Team Member 2 Initials (If Applicable):

Total - 100 points