What is Arknights?

Arknights, initially made in China but now available globally, is a tower defense mobile game. The spin with this is that instead of using “towers,” it uses unlockable characters to defend against incoming enemies. Players can unlock these characters in the shop with an in-game currency known as orundum or real-world money. Many of these characters are defined by their given rarity, such as being a 6-star character which is very rare, 5 stars being rare, 4 stars being uncommon, and 3 stars being standard. Different stats and skills make them distinct; however, 6 stars are what you should be looking for as they are powerful. Arknights also is very story-driven, with a lot of levels that present worldbuilding, character arcs, and new threats. Arknights also pump out events for their player base, so they can unlock possible cosmetics for their characters and obtain more in-game currency to unlock new or previous characters. However, Arknights maintains balance within their game by including an energy system where players can not complete its content within one day, so players can keep coming back! Although this might seem negative, it isn’t because Arknights wants players not to spend countless hours on their game; instead, they want players to complete some missions, then log off and continue the rest of their day. Arknights also provides an environment known as the “base,” where players can modify the environment, train their characters to become stronger, and even collect resources to evolve and level up their characters. Overall, players keep playing Arknights for their fantastic character designs and their advantageous system of obtaining characters. So far, Arknights has been a popular game with a player count of around 40 million since April 2020, and it keeps growing today!

Name: Andrew Angulo
Data Systems Final Semester Proposal

Central Question

Given that Arknights is a very time-consuming game, we're being brought up with the problem that some players don't have the time to play this game every day and possibly obtain the characters they want. Thus we can ask: Are some characters better than others based on their stats, and how much would the player have to invest in the game to collect that character? Answering this question will save people time when looking for a character they would like to unlock and seeing if unlocking that character is worth getting.

Data Sets

Data Set: Interactive Operator List - GamePress | Arknights Wiki
Description: This website includes tabular data on the character's rarity, their name, their role, and their given stats. | **[**Columns: 2, Rows: 269, Data Format: Long**]**
Date Accessed: Oct 29, 2022
Credit: GamePress

Data Set: Banner List (Gacha) - GamePress | Arknights Wiki
Description: This website includes tabular data of the upcoming banners, which contain characters that players can look forward to unlocking. | **[**Columns: 5, Rows: 9, Data Format: Wide**]**
Date Accessed: Oct 29, 2022
Credit: GamePress

Summary:
All of these websites provide and offer tabular data with all the given information sourced from the game Arknights. Which we can then translate into SQL and pandas dataframe.

Which data set do you intend to use?
I intend to use all of the data sets as they allow me to answer my central question with the best accuracy.

Name: Andrew Angulo
Data Systems Final Semester Proposal

Outline

Given the data sets above and the proposed question, the main focus would be to collect and store all of the data within Arknights by scraping it from the websites. Afterward, we can use SQL and pandas to accurately represent our data. Once we achieve all of this, we can step further into getting our answer to our question. Further below are some functions we could create to help us further answer our question.

Possible Functions:

- compareCharacter(Character1, Character2) → Compares the stats of both characters
- orundumCalculator(current, days) → Calculates the amount of in-game currency you will have
- banner_release_date(bannerName) → Returns the data when a specific banner released
- role_stat_compare(role) → returns the best character within the given role
- getCharacter(characterName) → returns all the values a character has
- shouldYouPull(orundum, banner) → simulates pulling on a banner and gives you the odds
- dayCalculator(orundum, orundumGoal) → returns the amount of days to reach the amount
- popularCharacters( ) → returns the current popular characters
- roleCharacters(role) → returns all the characters from that assigned role
- statCharacters(stat, value, operator) → returns all the character with the filtered stats
- daysTillBanner(bannerName) → returns x amount of days until the desired banner
- roleSize(role) → returns the amount of characters within a role as a int
- averageStatValue(stat) → returns the average value of a specific stat

Note: *more functions will most likely be created that weren't listed here*

Name: Andrew Angulo
Data Systems Final Semester Proposal

Entities

Given the information we will collect from the given websites, we will primarily store the data within a Dictionary of Lists due to us being able to store the headers as keys within the dictionary and retrieving the data from the websites using list comprehensions and XPath which will return all of our data in a respected list. Some issue when retrieving the data is that the HTML file will contain multiple <td> nodes without stating its class or another node without any information to distinguish it from another node. Most likely to overcome this problem we would have to use the position() function so we can retrieve the data from nodes that possibly don't specify a specific class.

Dataframes

Once we finally retrieve the data from both of the respected websites we will store all of the information in their respected Dataframe. The reason why we won't merge both of them is that one website gives us detailed information on each operator while the other gives us information about the upcoming banners. Naturally, these two datasets won't mix well as they have their own independent and dependent variables. However, once we get these into their data frames, we can then pivot and make them tidy so we can better understand and visualize our data. A benefit of having multiple data frames means that we can better represent our data within SQL, this will further provide the foundation for accurately representing our data and answering our central question.

Name: Andrew Angulo
Data Systems Final Semester Proposal