Data Memo for Arknights Operators Dataset

Chuyu Zhuang

2024-09-15

Overview of the Dataset

Dataset Description

The dataset I have chosen for my final project is titled "Arknights Operators," sourced from Kaggle. It includes detailed information about the operators (characters) in the mobile game "Arknights," including their stats, skills, and other attributes.

Source and Access

The dataset is publicly available on Kaggle and can be accessed via this link. I downloaded the dataset directly from the Kaggle website.

Size and Composition

- **Observations**: The dataset contains a comprehensive list of operators, with 225 rows, each representing an individual operator.
- **Predictors**: The dataset has about 63 columns, including numerical variables (such as attack, defense, HP) and categorical variables (such as operator class, rarity, gender, etc.).
- **Variable Types**: The variables in the dataset include numerical values (such as stats) and categorical data (such as operator class and rarity).

Missing Data

- **Missing Data**: There are some missing values in certain columns, particularly in some of the more advanced stats fields. Approximately 5-10% of the data may be missing in these columns.
- **Handling Missing Data**: I plan to explore multiple methods for handling the missing data, such as imputation with median values or more advanced techniques depending on the extent of the missingness.

Research Questions

Prediction Goals

The main goal of this project is to predict the future statistics of new operators based on existing data. I will build a corresponding prediction model by linking each operator's rarity (i.e. star), class and branch with their corresponding defense, hp, and maximum attack.

Response/Outcome Variables

The response variables will be the key stats for the operators: attack, defense, HP, etc. These are continuous variables, and the goal is to accurately predict their values for new operators.

Analytical Approach

The main goal of this project is to predict the future statistics of new operators based on existing data. I will build a corresponding prediction model by linking each operator's rarity (i.e. star), class and branch with their corresponding defense, hp, maximum attack, etc.

Important Predictors

- **Key Predictors**: I anticipate that predictors such as the operator class, rarity, and initial stats will be particularly useful in predicting the future stats of new operators.
- **Modeling Goals**: The goal of my model is both predictive (to accurately forecast future stats) and inferential (to understand which factors most strongly influence operator stats).

Proposed Project Timeline

- **Data Loading and Cleaning**: Complete by Week 2.
- **Exploratory Data Analysis (EDA)**: Start in Week 2 and complete by Week 3.
- **Model Building and Evaluation**: Begin in Week 4 and finish by Week 6.
- **Finalizing the Project**: Week 7 to Week 8, including model refinement and preparation of the final report.

Dataset Show

```r
# Example of loading the dataset
library(readr)
data <- read_csv("data.csv")
```

```
## Rows: 235 Columns: 63
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (44): name, class, branch, faction, stars, position, tags, trait, availa...
## dbl (19): base_hp, elite_1_hp, trust_hp, base_atk, elite_1_atk, trust_atk, b...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Display the first few rows of the dataset
head(data)
```

```
## # A tibble: 6 x 63
##   name        class branch faction stars position tags  trait availability icon
##   <chr>       <chr> <chr>  <chr>   <chr> <chr>     <chr> <chr> <chr>        <chr>
## 1 "Castle-3"  Guard Dread~ Rhodes~ 1-st~ Melee     ['Su~ Bloc~ Recruitment  http~
## 2 "\"Justice~ Snip~ Marks~ Pinus ~ 1-st~ Ranged    ['Su~ Atta~ Recruitment  http~
```

```
## 3 "Lancet-2"  Medic Medic  Rhodes~ 1-st~ Ranged    ['He~ Rest~ Recruitment~ http~
## 4 "THRM-EX"   Spec~ Execu~ Rhodes~ 1-st~ Melee     ['Nu~ Does~ Recruitment~ http~
## 5 "12F"       Cast~ Splash Rhodes~ 2-st~ Ranged    ['St~ Deal~ Recruitment~ http~
## 6 "Durin"     Cast~ Core   Rhodes~ 2-st~ Ranged    ['St~ Deal~ Recruitment~ http~
## # i 53 more variables: description <chr>, phrase <chr>, file_no. <chr>,
## #   gender <chr>, height <chr>, mobility <chr>, endurance <chr>,
## #   internal_name <chr>, base_hp <dbl>, elite_1_hp <dbl>, elite_2_hp <chr>,
## #   max_hp <chr>, trust_hp <dbl>, base_atk <dbl>, elite_1_atk <dbl>,
## #   elite_2_atk <chr>, max_atk <chr>, trust_atk <dbl>, base_def <dbl>,
## #   elite_1_def <dbl>, elite_2_def <chr>, max_def <chr>, trust_def <dbl>,
## #   base_res <dbl>, elite_1_res <dbl>, elite_2_res <chr>, max_res <chr>, ...
```