CGT 270 Data Visualization

Module 1

Week 3

**Lab 3: Mining Data**

The goal of this lab is to identify and implement techniques for mining data. In this lab you will identify patterns, extreme and subtle feature about data. You will identify basic descriptors for the data, and categorize data according to the specifications defined in the Parse Worksheet you completed in Week 2. After completing this lab, you will:

1. List at least three (3) questions you feel you can answer with the data sets you have acquired (Week 1) and parsed (Week 2).
2. Your questions must incorporate ALL three (3) of the data sets you’ve acquired from Lab 1: Tableau Dataset, Additional Dataset #1, and Additional Dataset #2
3. List any assumptions you are making in this stage of the data visualization process.

**What you should be able to do (at the end of this lab):**

|  |  |
| --- | --- |
| Understand | ***Describe*** the type of techniques to be used to better understand the data. |
| Apply | ***Execute*** techniques and methods (statistical methods) on the data. |
| Evaluate | ***Examine*** the resulting data and determine if it enables you to answer the question being solved. |
| Analysis | ***Identify*** patterns, extreme and subtle features about the data. |
| Create | ***Determine*** if the data can support the question to be answered. |

In the table below list each variable in the Tableau dataset, its data type (parsing) and a basic statistical or mining technique that can be applied to better understand the variable.

**Part I: Tableau Data set:** *Pokemon Data Set*

1. **Basic Descriptors**

List the **variables** from Week 2’s parsing lab and provide basic mining procedures.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Table** | **Variable** | **Data type** | **Basic mining procedure** |
| 1 | Pokemon | # | Float | Average, max, min |
| 2 | Pokemon | Name | Alphanumeric | Character length |
| 3 | Pokemon | Type | String | String length |
| 4 | Pokemon | HP | Integer | Average, max, min |
| 5 | Pokemon | Attack | Integer | Average, max, min |
| 6 | Pokemon | Defense | Integer | Average, max, min |
| 7 | Pokemon | Special Attack | Integer | Average, max, min |
| 8 | Pokemon | Special Defense | Integer | Average, max, min |
| 9 | Pokemon | Speed | Integer | Average, max, min |
| 10 | Moves | Name | String | String length |
| 11 | Moves | Type | String | String length |
| 12 | Moves | Cat. | String | String length |
| 13 | Moves | Power | Integer | Average, max, min |
| 14 | Moves | Acc. | Integer | Average, max, min |
| 15 | Moves | PP | Integer | Average, max, min |
| 16 | Moves | TM | Alphanumeric | Character length |
| 17 | Moves | Effect | Alphanumeric | Character length |
| 18 | Moves | Prob. (%) | Integer | Average, max, min |
| 19 | Evolution | Evolving from | String | String length |
| 20 | Evolution | Evolving to | String | String length |
| 21 | Evolution | Level | Integer | Average, max, min |
| 22 | Evolution | Condition | Alphanumeric | Character length |
| 23 | Evolution | Evolution Type | String | String length |
| 24 | TypeChart | Attack | String | String length |
| 25 | TypeChart | Defense | String | String length |
| 26 | TypeChart | Effectiveness | String | String length |
| 27 | TypeChart | Multiplier | Float | Average, max, min |

Add more rows to the table above as needed.

1. **Categorize**

Consider what variables are similar and what variables are different. This will help you to categorize the data. Are the data normal, ordinal or ratio? Take a look at this webpage and video: <https://www.graphpad.com/support/faq/what-is-the-difference-between-ordinal-interval-and-ratio-variables-why-should-i-care/>

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Table** | **Variable** | **Category** |
| 1 | Pokemon | # | Ratio |
| 2 | Pokemon | Name | Nominal |
| 3 | Pokemon | Type | Nominal |
| 4 | Pokemon | HP | Ratio |
| 5 | Pokemon | Attack | Ratio |
| 6 | Pokemon | Defense | Ratio |
| 7 | Pokemon | Special Attack | Ratio |
| 8 | Pokemon | Special Defense | Ratio |
| 9 | Pokemon | Speed | Ratio |
| 10 | Moves | Name | Nominal |
| 11 | Moves | Type | Nominal |
| 12 | Moves | Cat. | Nominal |
| 13 | Moves | Power | Ratio |
| 14 | Moves | Acc. | Ratio |
| 15 | Moves | PP | Ratio |
| 16 | Moves | TM | Ratio |
| 17 | Moves | Effect | Nominal |
| 18 | Moves | Prob. (%) | Ratio |
| 19 | Evolution | Evolving from | Nominal |
| 20 | Evolution | Evolving to | Nominal |
| 21 | Evolution | Level | Ratio |
| 22 | Evolution | Condition | Nominal |
| 23 | Evolution | Evolution Type | Nominal |
| 24 | TypeChart | Attack | Nominal |
| 25 | TypeChart | Defense | Nominal |
| 26 | TypeChart | Effectiveness | Nominal |
| 27 | TypeChart | Multiplier | Interval |

Review the different types of data and indicate the data types in your variables table:

<https://www.centralriversaea.org/wp-content/uploads/2017/03/F_Four-Types-of-Data-Revised-5.10.17.pdf>

1. **Temporal**

Is the data temporal (represent time, over several years, in years, days, minutes, seconds)?

No since the table does not have any sort of temporal identification besides missing recent data.

1. **Range and Distribution**

What is the distribution of the data? Few values, small size, evenly spread, sparse or dense? Explain.

The data is very dense in the Pokemon and Moves tables since there is 1168 rows and 9 columns of data in the Pokemon table and 608 rows and 9 columns in the Moves table. In the Evolution and TypeChart tables, the data is relatively sparse with 391 rows and 5 columns in Evolution and 325 rows and 5 columns in TypeChart. In the Moves table, it should also be mentioned that there are fewer values than apparent by its rows and columns since there are many null values.

**Part II: First (1st) additional data set:** Kaggle Pokemon

1. **Basic Descriptors**

List the variables from Week 2’s parsing lab and provide basic mining procedures.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Variable** | **Data type** | **Basic mining procedure** |
| 1 | abilities | String | String length |
| 2 | against\_bug | Float | Average, max, min |
| 3 | against\_dark | Float | Average, max, min |
| 4 | against\_dragon | Float | Average, max, min |
| 5 | against\_electric | Float | Average, max, min |
| 6 | against\_fairy | Float | Average, max, min |
| 7 | against\_fight | Float | Average, max, min |
| 8 | against\_fire | Float | Average, max, min |
| 9 | against\_flying | Float | Average, max, min |
| 10 | against\_ghost | Float | Average, max, min |
| 11 | against\_grass | Float | Average, max, min |
| 12 | against\_ground | Float | Average, max, min |
| 13 | against\_ice | Float | Average, max, min |
| 14 | against\_normal | Float | Average, max, min |
| 15 | against\_poison | Float | Average, max, min |
| 16 | against\_psychic | Float | Average, max, min |
| 17 | against\_rock | Float | Average, max, min |
| 18 | against\_steel | Float | Average, max, min |
| 19 | against\_water | Float | Average, max, min |
| 20 | attack | Integer | Average, max, min |
| 21 | base\_egg\_steps | Integer | Average, max, min |
| 22 | base\_happiness | Integer | Average, max, min |
| 23 | base\_total | Integer | Average, max, min |
| 24 | capture\_rate | Integer | Average, max, min |
| 25 | classfication | String | String length |
| 26 | defense | Integer | Average, max, min |
| 27 | experience\_growth | Integer | Average, max, min |
| 28 | height\_m | Float | Average, max, min |
| 29 | hp | Integer | Average, max, min |
| 30 | japanese\_name | Alphanumeric | Character length |
| 31 | name | Alphanumeric | Character length |
| 32 | percentage\_male | Float | Average, max, min |
| 33 | pokedex\_number | Integer | Average, max, min |
| 34 | sp\_attack | Integer | Average, max, min |
| 35 | sp\_defense | Integer | Average, max, min |
| 36 | speed | Integer | Average, max, min |
| 37 | type1 | String | String length |
| 38 | type2 | String | String length |
| 39 | weight\_kg | Float | Average, max, min |
| 40 | generation | Integer | Average, max, min |
| 41 | is\_legendary | Boolean | Mode |

Add more rows to the table above as needed.

**Part III: Second (2nd) additional data set:** Kaggle FirstGenPokemon

1. **Basic Descriptors**

List the variables from Week 2’s parsing lab and provide basic mining procedures.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Variable** | **Data type** | **Basic mining procedure** |
| 1 | Number | Integer | Average, max, min |
| 2 | Name | String | String length |
| 3 | Types | Integer | Average, max, min |
| 4 | Type1 | String | String length |
| 5 | Type2 | String | String length |
| 6 | Height(m) | Float | Average, max, min |
| 7 | Weight(kg) | Float | Average, max, min |
| 8 | Male\_Pct | Float | Average, max, min |
| 9 | Female\_Pct | Float | Average, max, min |
| 10 | Capt\_Rate | Integer | Average, max, min |
| 11 | Exp\_Points | Integer | Average, max, min |
| 12 | Exp\_Speed | String | String length |
| 13 | Base\_Total | Integer | Average, max, min |
| 14 | HP | Integer | Average, max, min |
| 15 | Attack | Integer | Average, max, min |
| 16 | Defense | Integer | Average, max, min |
| 17 | Special | Integer | Average, max, min |
| 18 | Speed | Integer | Average, max, min |
| 19 | Normal\_Dmg | Float | Average, max, min |
| 20 | Fire\_Dmg | Float | Average, max, min |
| 21 | Water\_Dmg | Float | Average, max, min |
| 22 | Eletric\_Dmg | Float | Average, max, min |
| 23 | Grass\_Dmg | Float | Average, max, min |
| 24 | Ice\_Dmg | Float | Average, max, min |
| 25 | Fight\_Dmg | Float | Average, max, min |
| 26 | Poison\_Dmg | Float | Average, max, min |
| 27 | Ground\_Dmg | Float | Average, max, min |
| 28 | Flying\_Dmg | Float | Average, max, min |
| 29 | Psychic\_Dmg | Float | Average, max, min |
| 30 | Bug\_Dmg | Float | Average, max, min |
| 31 | Rock\_Dmg | Float | Average, max, min |
| 32 | Ghost\_Dmg | Float | Average, max, min |
| 33 | Dragon\_Dmg | Float | Average, max, min |
| 34 | Evolutions | Float | Average, max, min |
| 35 | Legendary | Boolean | Mode |

Add more rows to the table above as needed.

**Part IV: Questions and Assumptions**

List at least three (3) questions you feel you can answer using the datasets you have acquired and mined. You MUST use complete sentences. Your questions must incorporate ALL three (3) of the data sets you’ve acquired.

Q1: What is the average value of all stats across all 1st generation Pokemon?

Q2: What is the most commonly occurring type in Pokemon?

Q3: What is the mean value of the HP stat of Pokemon?

**List 3 assumptions you are making in this stage of the data visualization process:**

1. **Assumption #1**

There are no errors with the given data.

1. **Assumption #2**

The providers of data are not purposefully leaving out data from the set.

1. **Assumption #3**

The providers of data have given the data with no ulterior motive.