Group Assignment 1: Project Characterization

Interactive-Visual Data Analysis, Fall 2023

Lara Fried, 15-729-627
Mark Woolley, 19-745-223
Matija Zver, 21-705-405
Zejie Guo, 21-749-767

November 4, 2023

What

1. Our project domain is gaming and our task is to create a Pokémon recommender system. Our visualization tool should help players build teams which are good in regards to some objective metric and a user specified goal but still provide enough freedom to also make sure that the player would actually want to be using the recommended Pokémon.
2. Our data has a roughly even split between quantitative and categorical attributes, and only a few ordinal attributes. The most important ones will be: the Pokémon's base stats, which there are 6 of and all are quantitative (see Figure 1); and the name and type(s) of the Pokémon which are all categorical. Most of the other features will only be used to narrow down the pool of Pokémon the player wants our recommender system to choose from (see Figure 2).
3. The quality of the provided data is very high, pre-processing of it will be restricted to dropping columns. As one can see in the Figures we have a constant count of 721 except for "Pr_Male" for which it makes sense that some Pokémon do not have this attribute. Regarding the categorical and ordinal attributes, all missing values make semantically sense. We will extend the data used with the interaction of the types using the GitHub repository "mwooll/Pokemon" [1] and the images of the Pokémon from kaggle "Pokemon Images Dataset" [2]. And to provide helpful filters, we will also parse data listing the available Pokémon in a given game or region, for example for "Hoenn" [3].

	Total	HP	Attack	Defense	Sp_Atk	Sp_Def	Speed
count	721.000000	721.000000	721.000000	721.000000	721.000000	721.000000	721.000000
mean	417.945908	68.380028	75.013870	70.808599	68.737864	69.291262	65.714286
std	109.663671	25.848272	28.984475	29.296558	28.788005	27.015860	27.277920
min	180.000000	1.000000	5.000000	5.000000	10.000000	20.000000	5.000000
25%	320.000000	50.000000	53.000000	50.000000	45.000000	50.000000	45.000000
50%	424.000000	65.000000	74.000000	65.000000	65.000000	65.000000	65.000000
75%	499.000000	80.000000	95.000000	85.000000	90.000000	85.000000	85.000000
max	720.000000	255.000000	165.000000	230.000000	154.000000	230.000000	160.000000

Figure 1: Base Stats

	Generation	Pr_Male	Height_m	Weight_kg	Catch_Rate
count	721.000000	644.000000	721.000000	721.000000	721.000000
mean	3.323162	0.553377	1.144979	56.773370	100.246879
std	1.669873	0.199969	1.044369	89.095667	76.573513
min	1.000000	0.000000	0.100000	0.100000	3.000000
25%	2.000000	0.500000	0.610000	9.400000	45.000000
50%	3.000000	0.500000	0.990000	28.000000	65.000000
75%	5.000000	0.500000	1.400000	61.000000	180.000000
max	6.000000	1.000000	14.500000	950.000000	255.000000

Figure 2: Filters

Why

1. The target users for our project are Pokémon game players, who are interested in building effective teams for their gameplay. We will primarily focus on the playthrough aspect of the game, which will offer us a range of possibilities when it comes to visualization.
2. Firstly, the user will be able to *identify* the strengths and weaknesses of each Pokémon on its own. Our recommendation system will give a personalized recommendation for the current team, helping the user *discovering* new Pokémon. For each recommendation we give an insight into how it was made, and allow the user to change it. The user can *compare* different recommendations and choose the best fitting one. Once the user has completed the team, an overall *summary* of the team will be offered.
3. Why do we feel the tasks above are important?
   → **Identify:** During the game, it is important to understand your current Pokémon and their advantages/disadvantages. We will use their Type 1 and Type 2 attribute, as well as their attack/defense rating.
   → **Discover:** Given the high number of Pokémon, we want to make sure that the user chooses the best team out of all the Pokémon available, even if they are still unknown to them. We will try and incorporate as many attributes as possible.
   → **Compare:** In order to make the best decision the user has to have some freedom in the recommendation system.
   → **Summarize:** After picking the team, the user should be able to have an overview of the whole team, so we will provide a summary. All of the attributes available to us will be included.

How

1.
   - We use a combination of dropdown menus and check boxes for the user to identify the goals of his team, then filter and recommend the Pokémon based on this input
   - For each recommended Pokémon we present the different attributes and how much they influenced the recommendation in a spidermap, this allows the user an insight into how the recommender works.
   - For each recommended Pokémon there is a swap button, that shows the user the 3 closest Pokémon to the selected one, this allows the user to further personalize the team.
   - There is a Recalculate button, if the user does not like the recommended team, this will generate a new recommendation.

2. Sketch

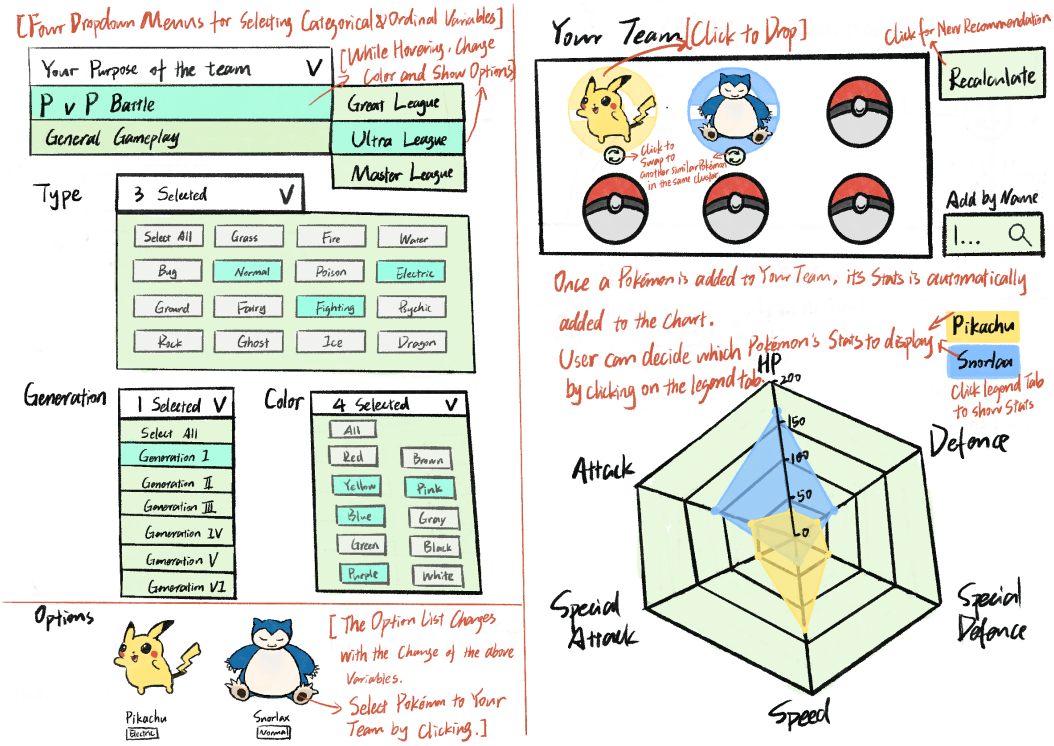

Figure 3: First Draft

3. Modeling method For the recommendation system we plan on using scikit-surprise [4]. We plan on calculating the ideal stats for a recommedation and rank the Pokémon by similarity [5]. Another tutorial using scikit-surprise builds a collaborative filtering recommender system, we might use this as well [6].

Group Dynamics

With our group we have a variety of skills and interests. After filling out the questionaire and discussing our results, we ended up with the following task distribution. Mark is our data steward and domain expert, he has detailed knowledge of all things Pokémon, so he will be in charge of preparing the dataset, and decide which attributes of the provided data set are relevant for our project. Matija as the ML/AI Engineer and Lara as the data Engineer will set up the recommendation system in the backend, and prepare the frontend Zejie as a Generalist and Technical analyst will focus on Analysis. She will work on the details of the frontend, and together with Mark will work out the optimal stats for the different filters, so the recommendation system has an ideal to compare to. Communication will be mostly done by Mark and Lara our Evangelists.

References

[1] M. Woolley, "Pokemon," 2023, https://github.com/mwooll/Pokemon/tree/master, Accessed on 2023-11-02.

[2] kvpratama, "Pokemon images dataset," 2020, https://www.kaggle.com/datasets/kvpratama/pokemon-images-dataset/, Accessed on 2023-11-02.

[3] Bulbapedia, "List of pokémon by hoenn pokédex number," 2023, https://bulbapedia.bulbagarden.net/wiki/List_of_Pok%C3%A9mon_by_Hoenn_Pok%C3%A9dex_number_(Generation_III), Accessed on 2023-11-02.

[4] N. Hug, "Choosing a model," 2015, https://surprise.readthedocs.io/en/stable/building_custom_algo.html, Accessed on 2023-11-02.

[5] ——, "Surprise similarity module," 2015, https://surprise.readthedocs.io/en/stable/similarities.html, Accessed on 2023-11-02.

[6] DOR, "Using scikit-surprise to create a simple recipe collaborative filtering recommender system." 2022, https://www.alldatascience.com/recommender-systems/simple-recipe-recommender-system-with-scikit-surprise/, Accessed on 2023-11-02.