Milestone 1

Datasets

Here you will find the datasets that we will mainly use for this project. Please find the accompanying exploratory data analysis notebook to see the results of our findings: - Harry Potter Collection contains a lot of information about characters, potions, spells and books. - We enhance the book dataset (in the previous bullet) with more information about sales and translations - Harry Potter network describing the network of interactiosn between characters. - Harry Potter Books content containing the texts of all the Harry Potter books. - Harry Potter movies containing information about the movies as well as the transcript that can be useful for advanced analysis. - Top 40 favourite Harry Potter characters based on this article .

Problematic

The general topic of our project consists in visualising the Harry Potter world by analysing the characters, their specificities, as well as their relationships. Our goal is to retrieve a great amount of information from our datasets to better understand the series and eventually retrieve the main differences between the movies and the books.

Exploring characters:

- *Houses*: The concept of Houses is extremely relevant. Hogwarts is divided into 4 houses. Before beginning their schooling, each student must be sorted into their houses by the Sorting Hat. Given our list of characters, we are interested to study their distribution among the 4 houses.
- *Wands*: The study of the history and the magical properties of wands is called *wandlore*. Each wand consists of a type of wood surrounding a core of magical substance. Although the wand cores might come from the same creature, or the wood might come from the same tree, no two existing wands were exactly alike. One idea of interactive visualisation will be to let the user choose components of a wand and show him/her the closest character in terms of wand components.
- *Blood Types*: One of the underlying problems in the wizarding world is intolerance and prejudice based on the “purity” of a person’s wizarding blood. We will study the distribution of blood types, dstinguished between purebloods, half-blood and Muggle-borns, among all the characters.

- *Spells*: We have at our disposal a complete dataset collecting a lot of spells that can be cast by wizards. For each spell it is interesting to display the pronunciation and the description.
- *Association between characters*: We are interested in retrieving the existence of an eventual association between two characters.

Movie Datasets:

These datasets contain a lot of information about the movie. We start by observing the budget granted to each movie and see how much it produces at the box-office. We also check the length of the movies to see if there is a trend for the movies to last longer or to be shorter.

The target audience:

This project targets a large set of people. On the one hand, people that are new to the Harry Potter world, if they wish to get to know more about the series or are just curious, they can have a clear overview of the set-up of the story. On the other hand, people that are already familiar with the wizarding world, will hopefully enjoy going through this.

Related work

With an intellectual property this big and impactful, there has been many vizualizations of Harry Potter's universe. Here are some examples found on different websites: - *Pinterest*: A lot of the works found on Pinterest related to Harry Potter are made by graphic designers, but there is some real data visualization behind it. Such as The Wand-o-graphic found on wizardingworld.com which showcases interesting facts about the magic wands used throughout the Harry Potter universe, based on observational studies which is part of what we'll do. Even though we might not be able to recreate these visually aesthetic pieces, a lot of them could serve as inspirations in the way they showcase interesting information in an eye-catching way. - *Tableau*: Another very good source for data vizzes in general is the Tableau public repository. This vizualization of spell types or this one showcase exactly what we'd like to do with the spells in the Harry Potter universe. - *Accio data*: As an additional Harry Potter-specific source we have Accio data, a work done by a graphic designer during her senior year, which she describes as a data-driven visual guide to all things Harry Potter. Since it's a book, these data vizzes are static but showcase a lot of what we'd like to show in our project in a simple yet visually appealing way. For example, the info we'd like to show of characters and their blood type is displayed here as a periodic table, with each blood type represented by a different color. - *Explore Harry Potter via a dynamic social network of characters*: The final source of information we find useful is a network representation of the relations in the Harry Potter saga. It is a good source of inspiration since we already plan to do a visualization involving connections/relations between characters. Our approach will be

different in the sense that we won't explore the relations between the characters at the chapter level, but we will rather see if they interact at any time during the whole book series. This means that we will lose the temporal component in the first place but in the other hand we will end up with a bigger network and maybe more insightful network.