The general topic of our project consists in visualising and recreating the Harry Potter world by analysing the characters and 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**:

* **Gender**: The distribution of genders in the Harry Potter series (between Male and Female)
* **Houses**: The concept of Houses in Harry Potter world is extremely relevant. Hogwarts is divided into four houses, each bearing the last name of its founder: Gryffindor, Ravenclaw, Hufflepuff and Slytherin. Before beginning their schooling, each student must first be sorted into their correct houses by the magical Sorting Hat.

Given our list of characters, we are interested to study their distribution among the 4 houses.

* **Wands**: Wands are unique and tells a lot about its owner. The study of the history and the magical properties of wands was called wandlore. Each wand consisted of a specific type of **wood** that surrounded 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 of our ideas of interactive visualisation will be to ask to the user to choose components of the wand and display to him/her the closest character in terms of wand components. From our dataset, only 27 wands given our 707 characters are known.

* **Physical Traits**: Another interesting feature available from this dataset is the physical characteristics such as the eye and hair color for some characters. This will help us build and display their profiles.
* **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. Wizards distinguish between **purebloods** (no Muggle ancestry), **half-blood** (at least one wizarding parent but at least one Muggle parent or grandparent) and **Muggle-borns** (both parents were Muggles).

We therefore study the distribution of blood types among our characters.

* **Species**: A part from humans, we can distinguish and explore several other creatures in Harry Potter world, such as Ghosts, Goblins, House-elves …
* **Spells**: We have at our disposal a complete dataset collecting a lot of spells that can be cast by Harry Potter world wizards. We collect about 300 spells, that we sort into categories. 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 such as the budget, box-offices or title and length of 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. It might be interesting to compare the length of the movie with the corresponding book length.

The target audience: This project targets a large set of people. On the one hand, people that are new to 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 through our project.

On the other hand, people that are already familiar with this whole wizarding world, will hopefully enjoy going through this.