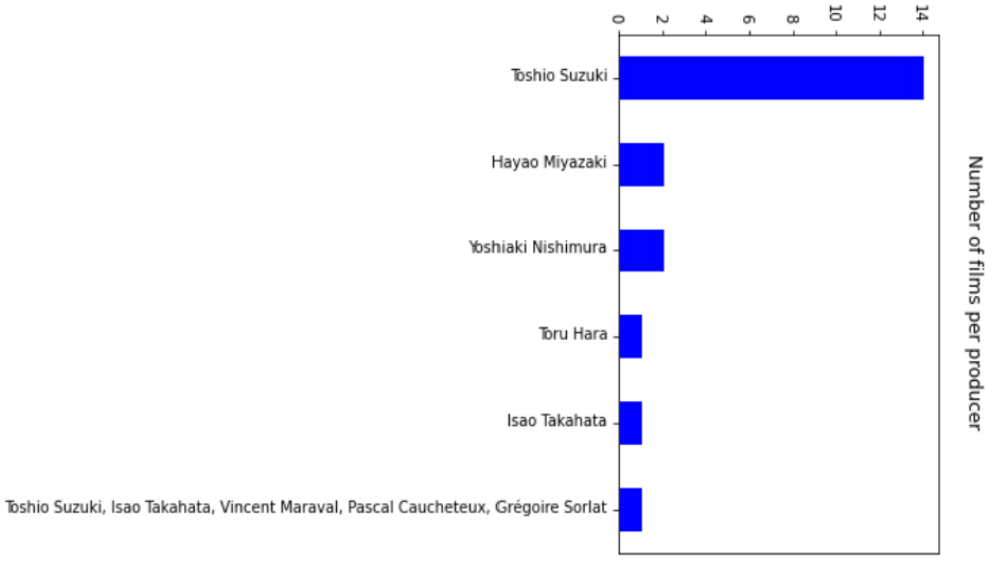
Documentation Python project
Task 2

API: Studio Ghibli
Group members: Olimpia Sannucci, Hanna Carucci Viterbi, Giulia Di Martino,

The second task of the project aimed at visualizing and manipulating data. The .csv files generated in the first part of the project were used as a dataset for the second task.
Thus, an analysis has been carried out on Python based on the Studio Ghibli API. In a first stance, matplotlib and numpy have been imported in Python. Matplotlib is a plotting library for the Python programming language as a component of NumPy, a big data numerical handling resource.
Then, for each of the 5 dataframes, a data visualization and manipulation analysis was carried out.

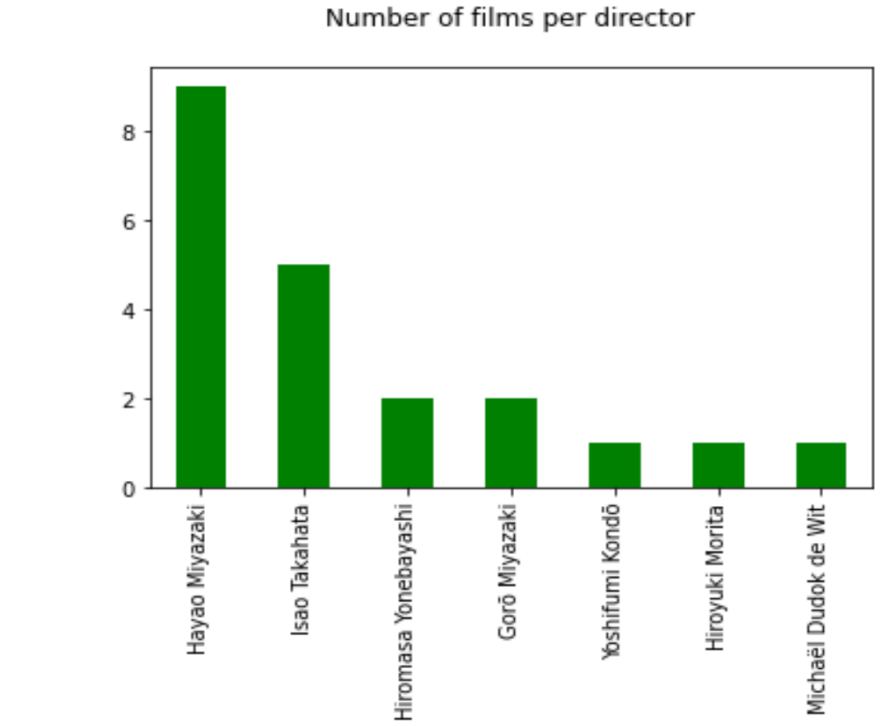
Dataframe 1: films
The first dataframe contains information regarding films. Indeed, it provides particular insights on the films supplied by Studio Ghibli. In order to better visualize this information, the non-significant variables were removed: the "id", "url", "original_title" ,"description" columns were dropped from the dataset. Following, graphs related to the Studio Ghibli films were generated.
The first graph displays the number of films supplied by each producer.

Graph 1 - Number of films by each producer



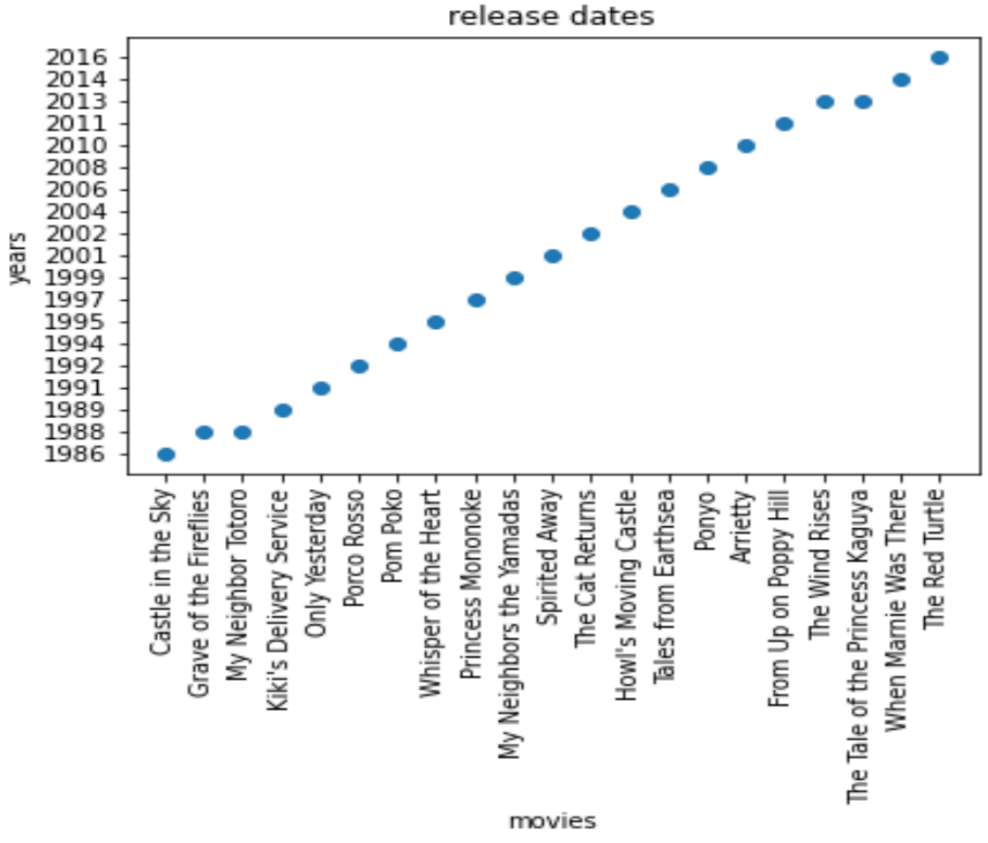
The second graph outlines the number of films produced by each director.

Graph 2 - Number of films by each director



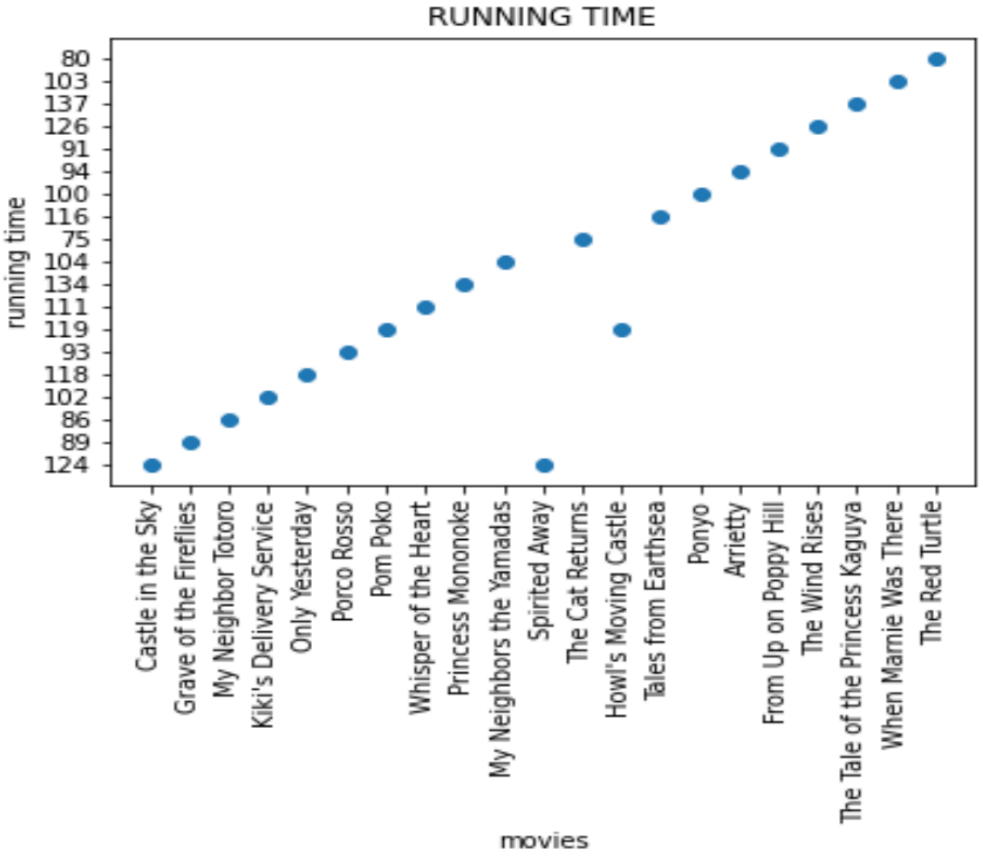
The third graph is a scatter plot representing all the movies produced and the year in which they were released.

Graph 3 - Movies produced and the year of release



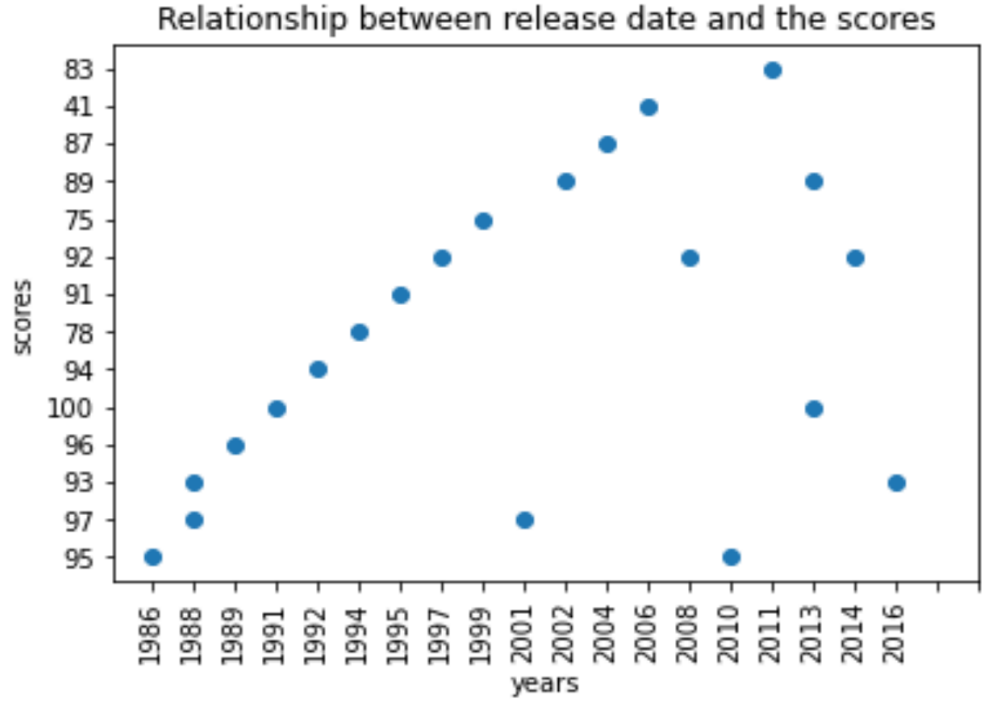
The fourth graph is a scatter plot displaying all the movies provided by Studio Ghibli and their respective running time.

Graph 4 - Movies and running time



The last graph is a scatter plot concerning the score and the release date of Studio Ghibli's films.

Graph 5 - Score and release date

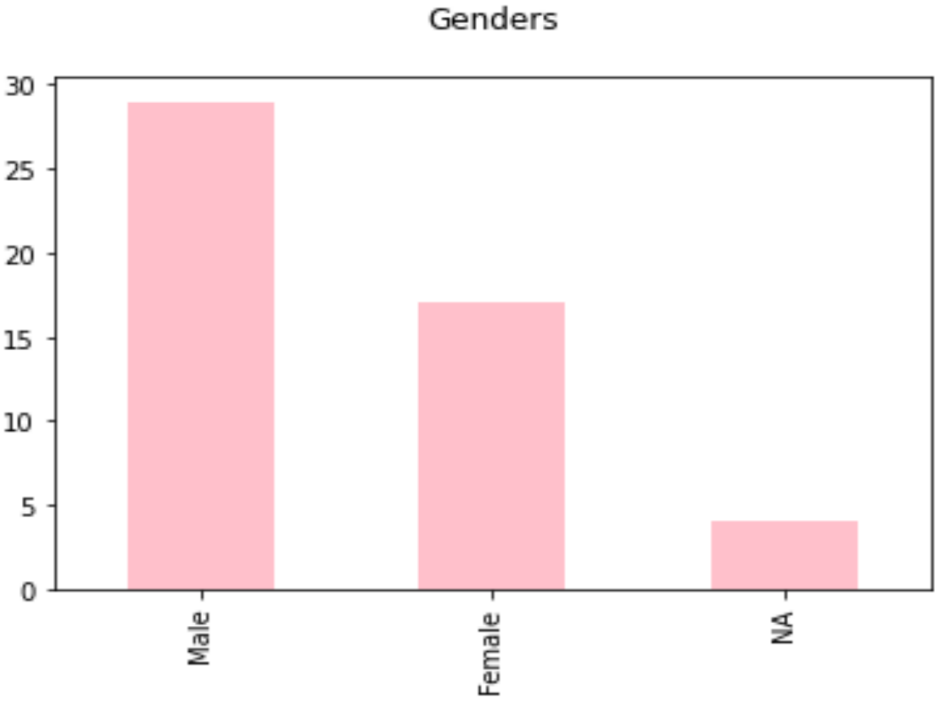


Dataframe 2: people

The second dataframe concerns people. In detail, it provides information about the characters of each of Studio Ghibli's films. With the aim of visualizing data, two irrelevant columns for the purpose of the analysis have been dropped: "id" and "url". Consequently, the distribution of each gender in the dataframe "people" has been found.

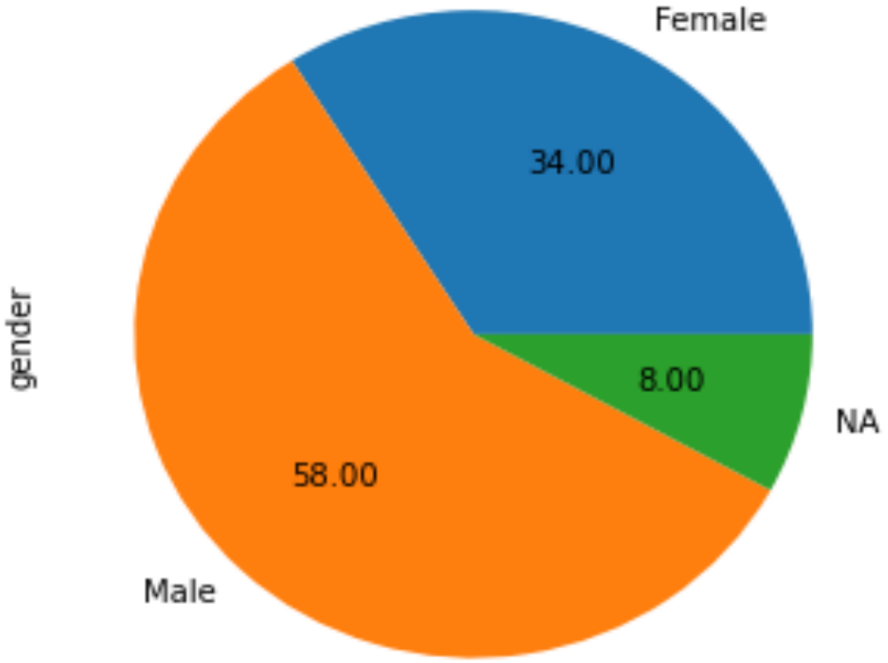
Firstly, gender bar plot and pie plot have been generated.

Graph 5 - Gender bar plot



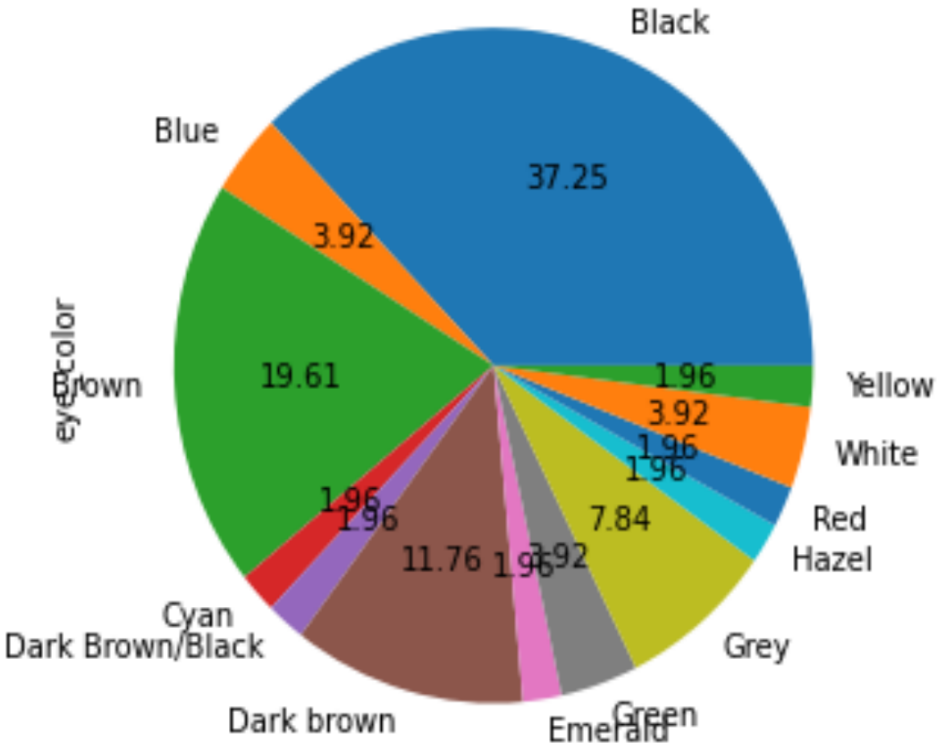
The function count() has been used in order to transform non numerical variables for generating the gender pie plot.

Graph 6 - Gender pie plot

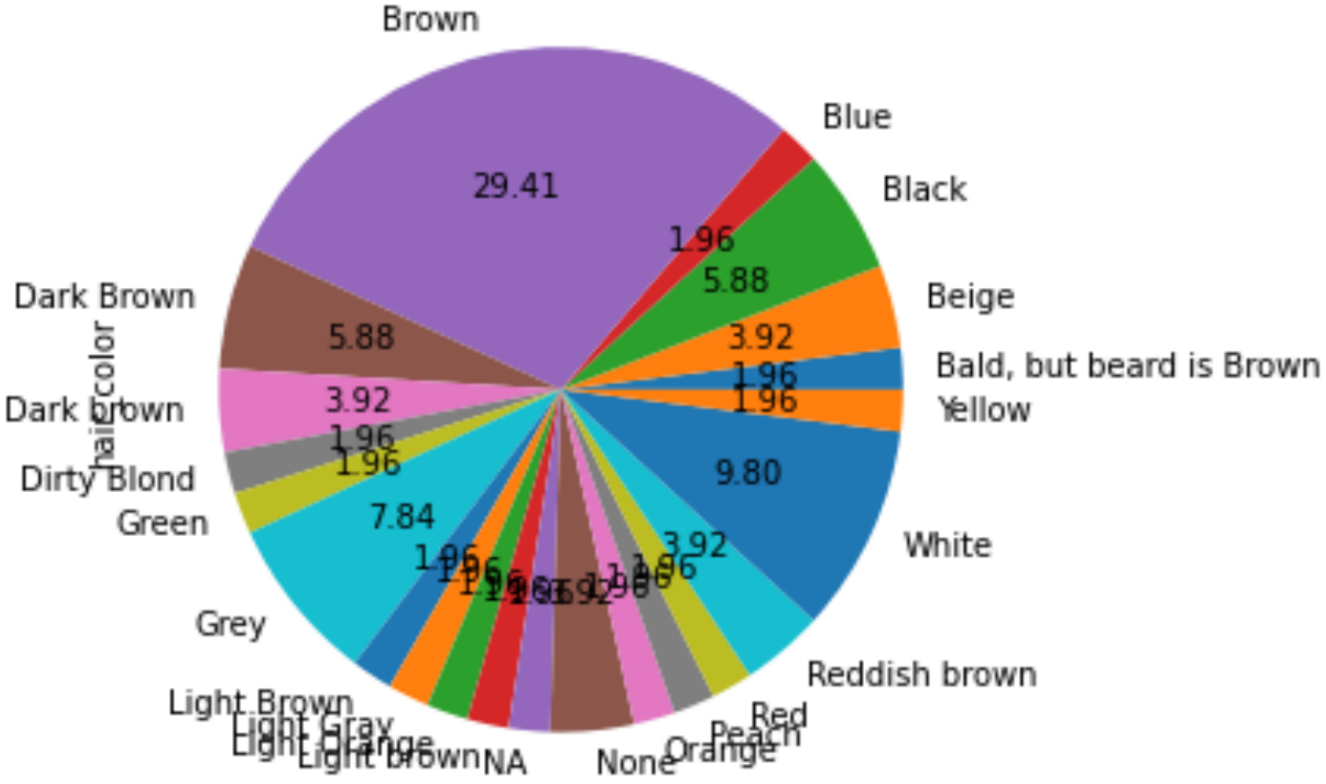


Then, pie plots regarding the eye and hair color of these film's characters have been created.

Graph 7 - Eye color pie plot



Graph 8 - Hair color pie plot

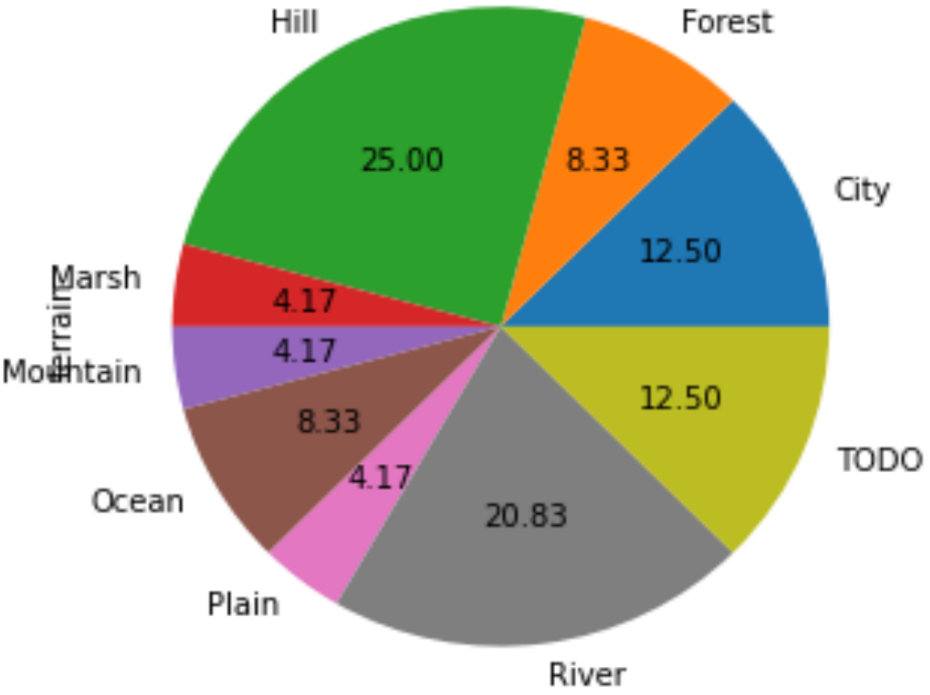


Dataframe 3: locations

The third dataframe outlines the names of the locations in which the films are set and their climate properties. Indeed, two variables that were not necessary for the analysis have been dropped from the dataset: “id” and “url”.

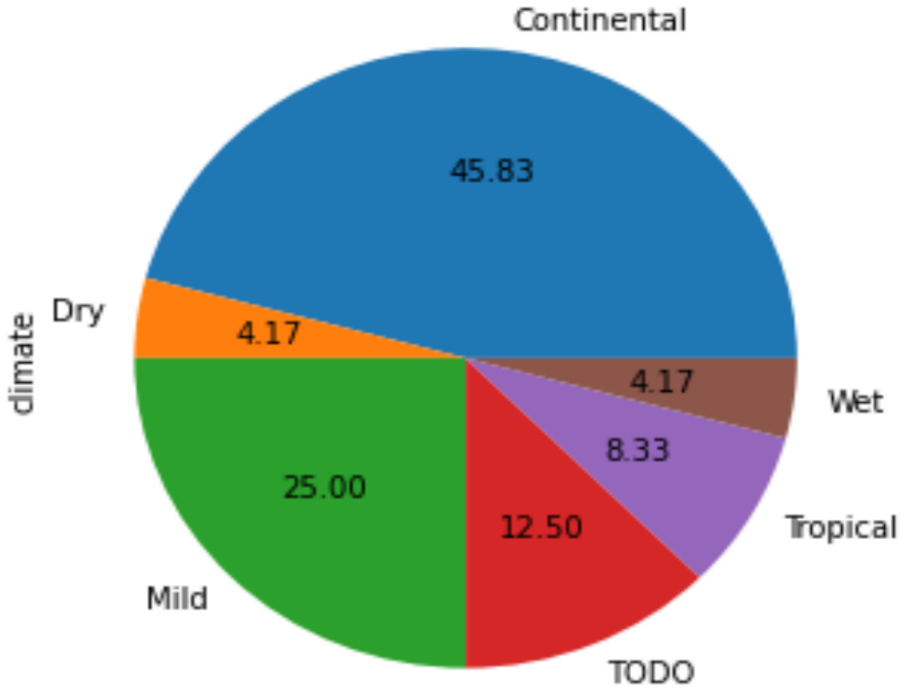
A pie plot has been generated for displaying the terrain of the locations.

Graph 9 - Terrain pie plot



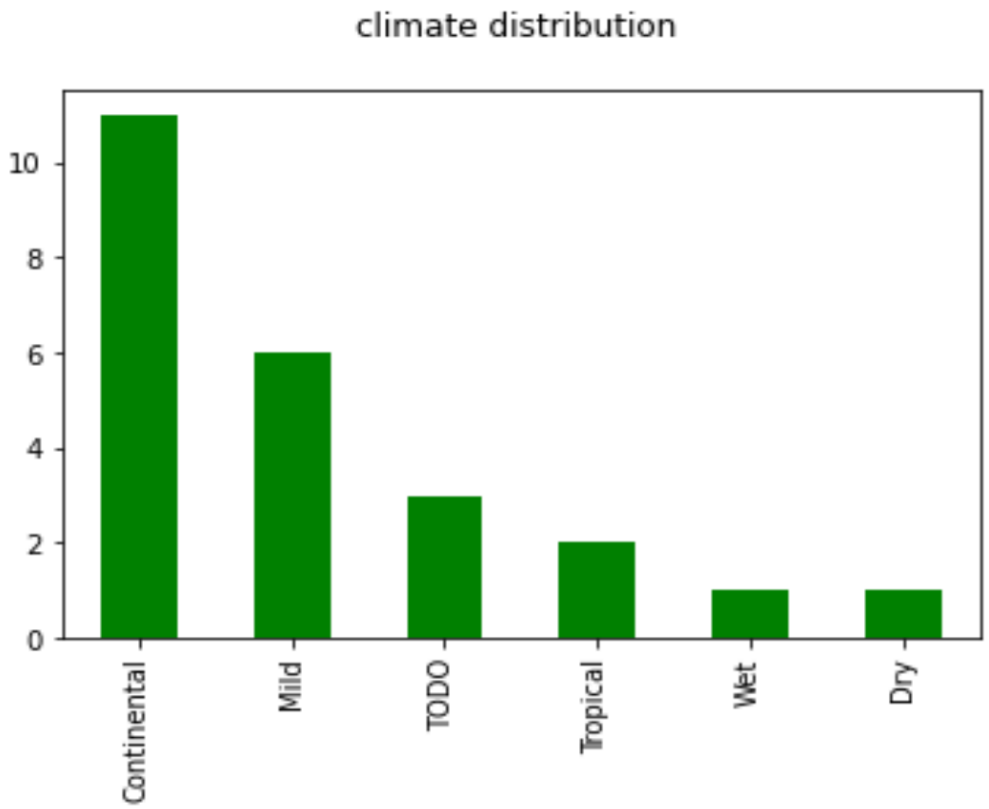
Then, in order to graph the climate of the locations, a pie plot has been constructed.

Graph 10 - Climate pie plot



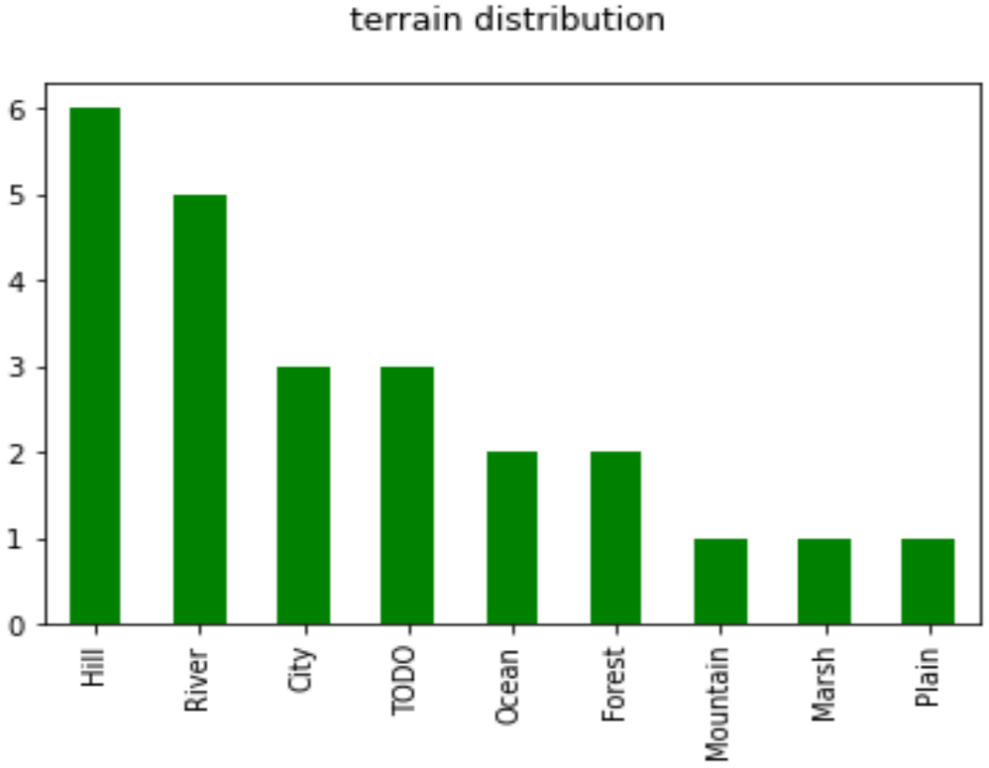
Furthermore, a climate distribution was graphed.

Graph 11 - Climate distribution



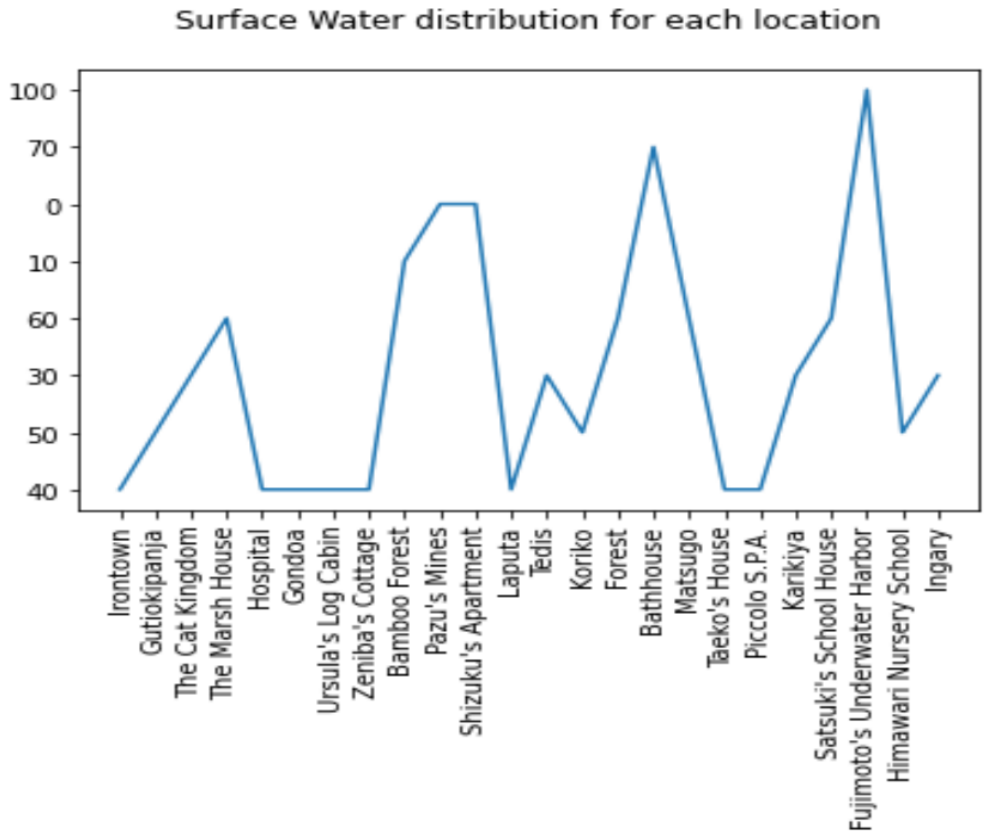
Finally, a representation of terrain distribution for each location was provided.

Graph 12 - Terrain Distribution



Finally, a water distribution for each location was graphed.

Graph 13 - Water distribution for each location

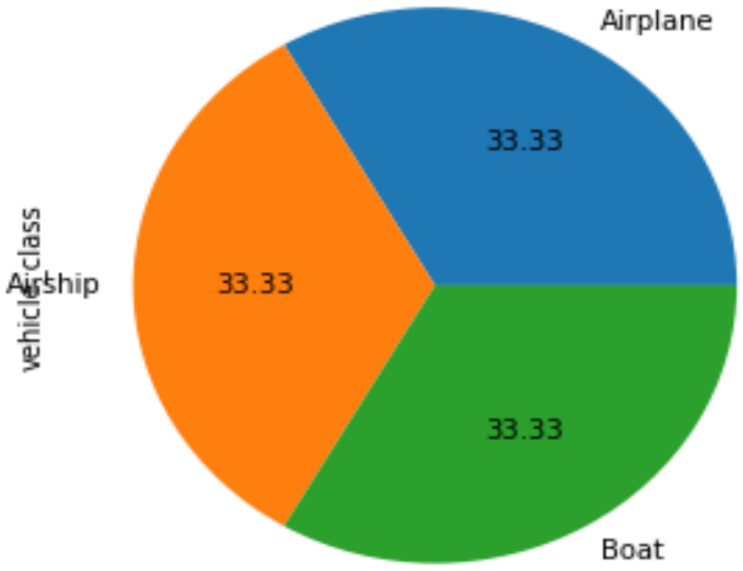


Dataframe 4: vehicles

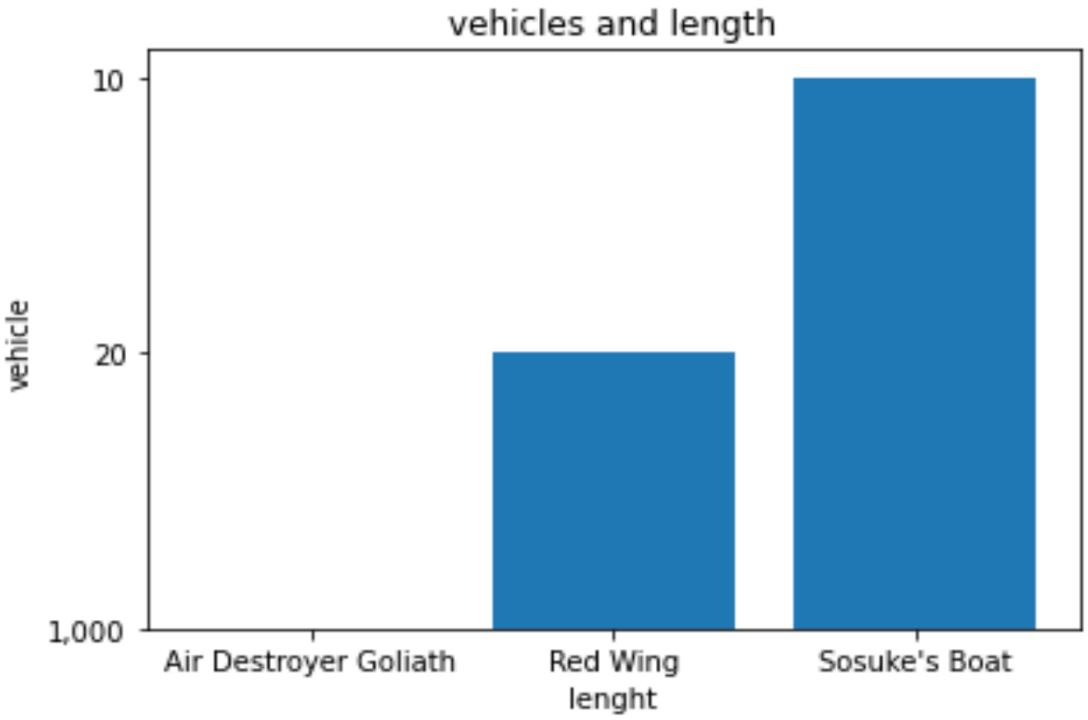
The fourth dataframe outlines the 3 main types of vehicles employed in the films. Two variables from the dataset have been removed, since "id" and "url" were irrelevant for visualizing this dataframe.

A pie and bar plot of the vehicles employed in Studio Ghibli's films have been generated.

Graph 14 - Vehicles pie plot



Graph 15 - Vehicles barplot



Dataframe 5: species

The fifth dataframe defines the living species present in the films, addressing attention to their imaginary nature. The columns “id” and “url” have been removed from the dataset because they are irrelevant for the data visualization.

Finally, a pie plot displaying the species present on the Studio Ghibli’s films has been built.

Graph 16 - Pie Spices