A geographical analysis of Culture

This subject tackles the question of cultural diffusion and availability. We are everyday confronted to cultural ideas and concepts. In our globalized world, who creates the cultural content on which a population is confronted? Is it uniform over the globe or are there disparities? Using the movie dataset, we can try to dive into a snippet of this broad question by looking at culture revolving around movies.

Who produces the films? On what geographical area are the movies available to the public? Is there a monopoly of a few country (US films for example) or is the movie production more geographically distributed? We could extend the research to actor's nationalities as well as producer's. Are actors only playing in movie studios of their own nationality or is there crosspollination? This analysis would use the language in which a film is translated, as well as potential additional databases such as the Wikipedia pages of actors to know about their nationality, or info about cinemas frequentations and visas (https://www.data.gouv.fr/fr/organizations/centre-national-du-cinema-et-de-l-image-animee/)

History of Memories

The way historic events are vulgarized and popularized evolves with time and space. By analysing the plot summaries of movies revolving around historic events, we could try to understand how those historic periods are portrayed. Is it an action movie? A drama? On what element is the movie focused? On what type of characters is it focused? Heroic players, villains, nobles and elites or more popular social classes? This character analysis could be based upon already existing researches such as the one presented in the Learning Latent Personas paper.

Interesting periods to be used for this analysis could be World War I and II, the Cold War, the Colonies. In some cases, it may be interesting to expand the dataset with more recent movies in order to observe evolution over the 2000-2020s.

How has a specific genre evolved over the years?

The dataset contains movies that were filmed over a more-than-60-years period. Ranging from the 1950s until the 2010s, the dataset usually presents at least 200 movies per year. Most of them include genre categorisation.

By leveraging the plot summaries available in the dataset, we could try to analyze how a specific genre (for example horror or science-fiction) has evolved over the years. How many characters are displayed? Are the personas presented similar to each other or is there an evolution trend? What kind of events happen in the plot? We could relate the evolution in plots and subjects with technical improvements in the filming industry (cameras, digitalisation, etc.) and see whether technological progress influences plots in any manner.

Guess the story's genre

In a reverse way, we could also try to extract a movie genre from a textual plot summary and fill in the gaps in the dataset. By analysing the vocabulary used, the types of situations happening in the plot,

we could create a model that would attribute a genre to a textual plot by aggregating keywords to specific categories. The method could be tested using the data already provided about genre.