Collecting Data Final Project Gr. 4

Exploratory Data Analysis and Data Visualization of Steam Games Genres

The goal of the project: Research:

Research question/hypothesis: How do Steam game genres show different trends in popularity over time

Description of the data to be collected:

In terms of data sources, the official Steam Games Genres website provides a lot of first-hand data, which can be queried by category, region, etc., and also by monthly/weekly charts. It has the advantage of a large data sample size and is easy to query. We will mainly collect the game name, release date, producer, game platform, tags, positive--ratings, negative--ratings, average_ playtime, price, game description and other data.

Preliminary ideas for the 3 sections:

Introduction and Background: Since its debut in 2003, Steam, a sizable digital distribution service for video games, has drawn a lot of attention. It is currently the largest digital distribution platform for computer games worldwide. The Steam platform held approximately 75% of the market share in 2013. In 2019, Steam saw over 34,000 games on Steam's shelves and over 95 million monthly active users, while Steam is available in all major global gaming markets. As such, the data on the platform regarding video games is very representative. We believe it provides a more representative picture of the changing trends in video game themes and genres over the past 10 years.

The goal of the project is to collect, collate, analyse and visualise Steam game data to produce a report on global trends in Steam games, which will include changes in the popularity of games by category, which genres have become more popular over the last ten years and, where possible, the most popular categories of games by country and region, etc.

Tutorial:

In addition to using data that already exists, Steam also has a relatively well-established data system (SteamDB) to support our project. In addition to this, we will also use python requests and beautifulsoup modules to do some scraping during the data collection phase.

In the collation phase, we use the pandas module to organise and clean the dataset, and in the analysis and visualisation phase, we use the seaborn/matplotlib modules to visualise the data.

This section will be clearly presented in the Jupyter Notebook.

Active Learning Exercises:

In this section, we plan to guide the audience to analyze our charts and discover some visible trends. For example, try to identify the most popular game genres in China in 2012 and analyze what you think are the main reasons for this trend. How does the number of different platforms change in the game? Or which tag of the game is more popular with the public?

Output chosen:

In addition to the Jupyter Notebook and the DMP, we are going to create a data story using Flourish. This is to show vividly how the different game genre (tag) has changed during this decade. If possible, a Flourish visualisation of the popular game genres in different countries will also be presented.