Project aim and outline

Aim: To learn more about web scraping.

Outline:

-> Scrape game titles from the IGN website.

-> Scrape corresponding game data from Steam platform.

-> Explore data using jupyter notebook and Pandas.

Web scraping with Selenium

**Selenium is web browser automation tool. It can handle dynamic web page elements e.g., infinites scrolling.**

1. Open IGN webpage and scroll. Scrape game titles.

2. Clean broken game tiles.

3. Open Steam & scrape game data from titles.

4. Clean data in jupyter notebook and pandas.

Exploring game data: Distributions

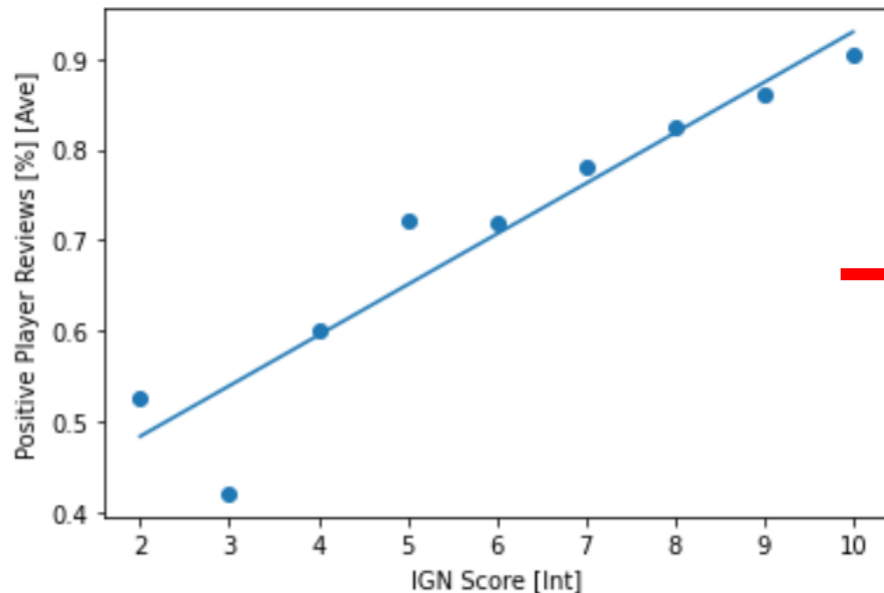**Histograms of game data: Price [£], IGN Score, Positive Steam reviews [%], & No. of reviews.**
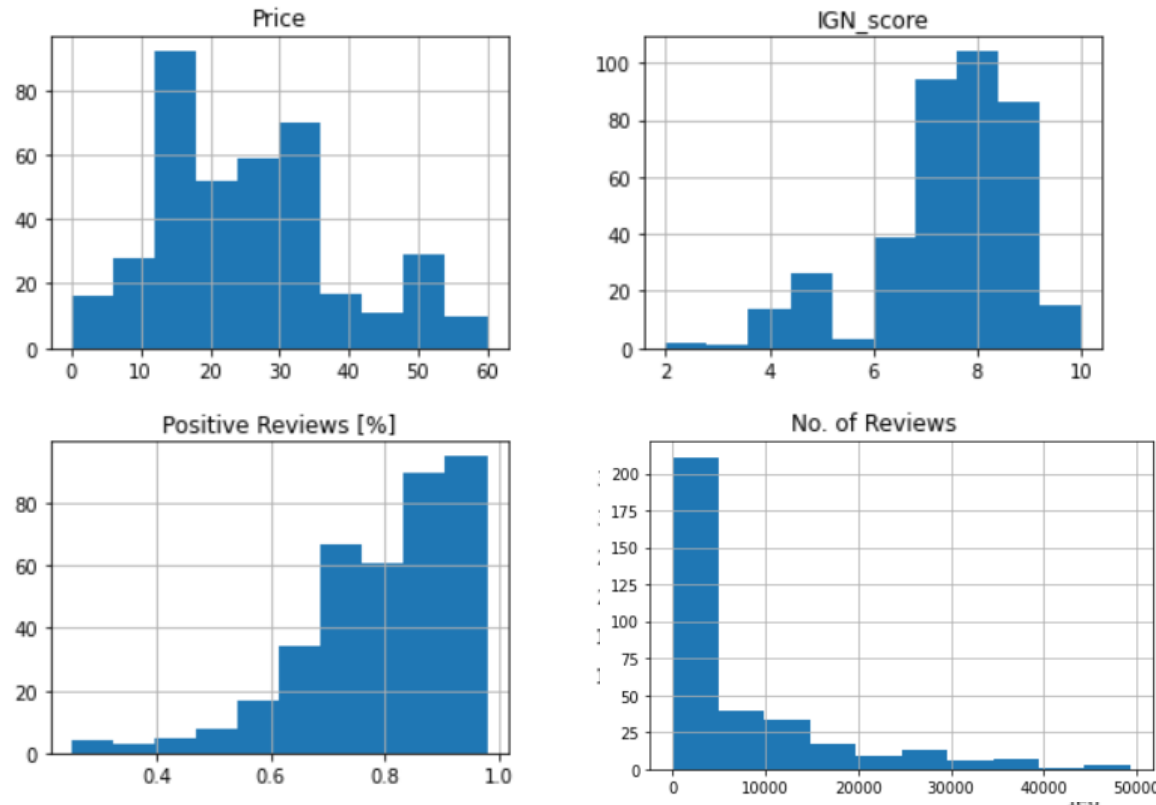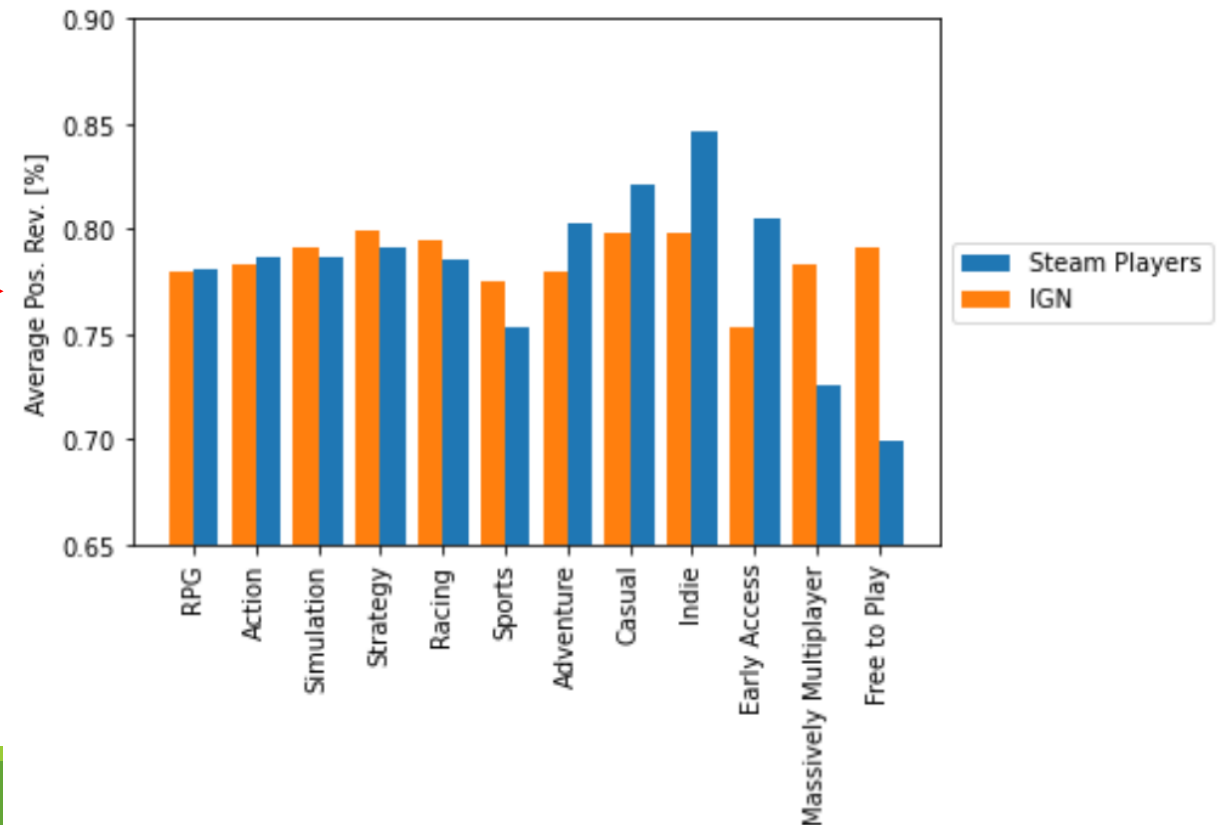
Frequency of game genres
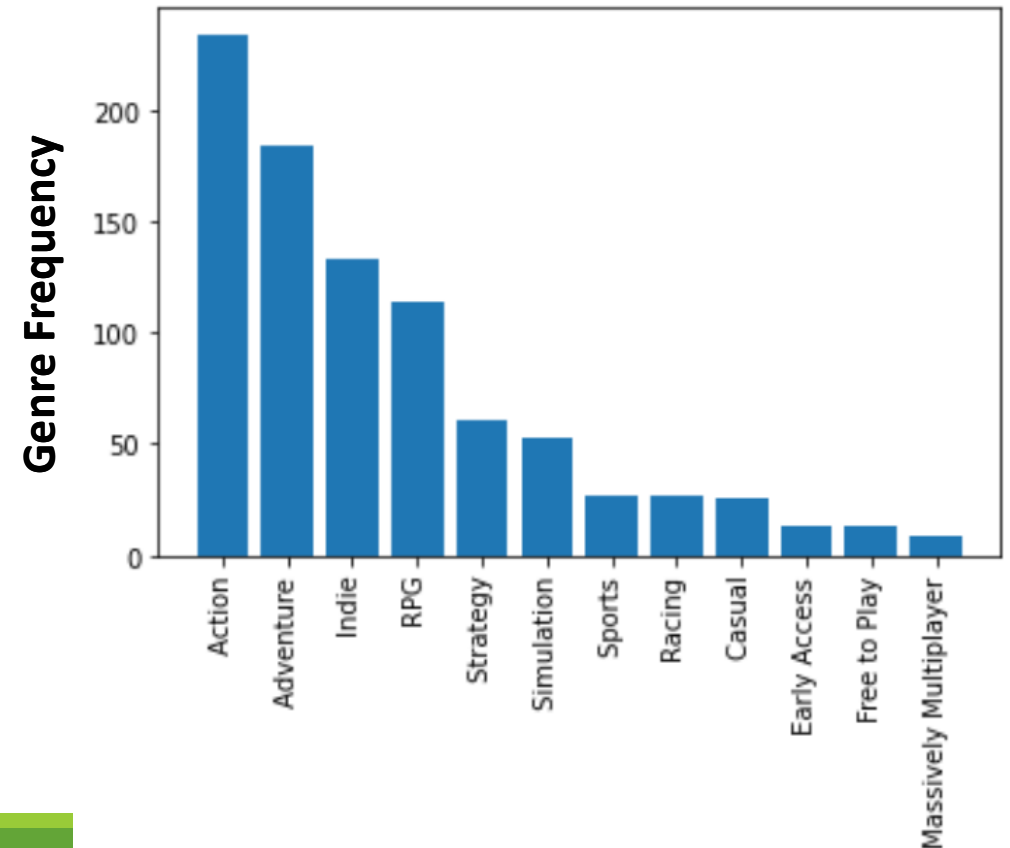
Exploring game data: Players VS IGN

Linear regression relation between positive player review [%] and IGN review score

Scaled Comparison between Steam player and IGN positive review [%]

Conclusions overview

-> Selenium can be used to scrape data from webpages with dynamic elements.

-> It is very helpful to know what your using the data for before you scrape it.

-> Scaped data will require lots of cleaning. Manual cleaning can take up a lot of time.

-> Keep as much data as possible. Likely I could have kept some of the dropped data.

Please see Jupyter notebooks for detailed explanation of how data was scraped cleaned and, explored.

This was a brief summary.