

Plum Insights Test Question

Solution by Assel Batu

Question:

What % of titles provide the vast majority of the time viewed?

What is the contribution across the whole catalog?

How does this differ by genre/content area?

In other words, we want to better understand whether a few titles represent a large part of the content that is watched, or whether many titles contribute approximately equally. We want to look at all viewing in total, but also focus on specific areas (such as focusing only on comedies).

Given Data:

Table 1: anonymized_viewings, Table 2: metadata, Table 3: season_metadata

Solution:

- 1) I grouped (joined) two tables (Table1 and Table2) by the *content_id* property.
- 2) Then group by title and wrote a small code, procedure to see how many % of views were generate by each title.

For example:

Title 1: 70% of views

Title 2: 15% of views

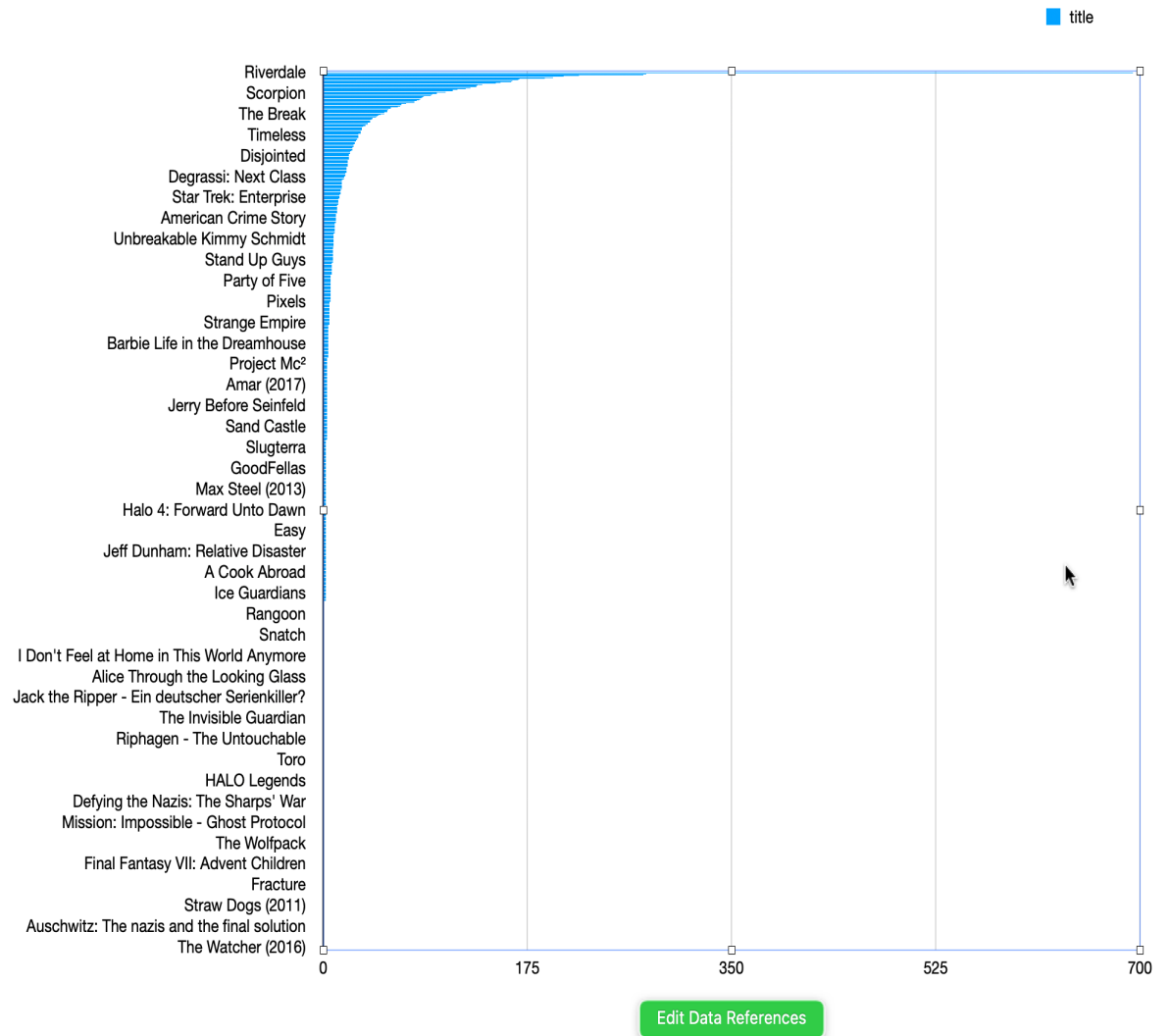
etc...

The Pareto principle states that **for many outcomes, roughly 80% of consequences come from 20% of causes**. In other words, a small percentage of causes have an outsized effect. This concept is important to understand because it can help you identify which initiatives to prioritize so you can make the most impact.

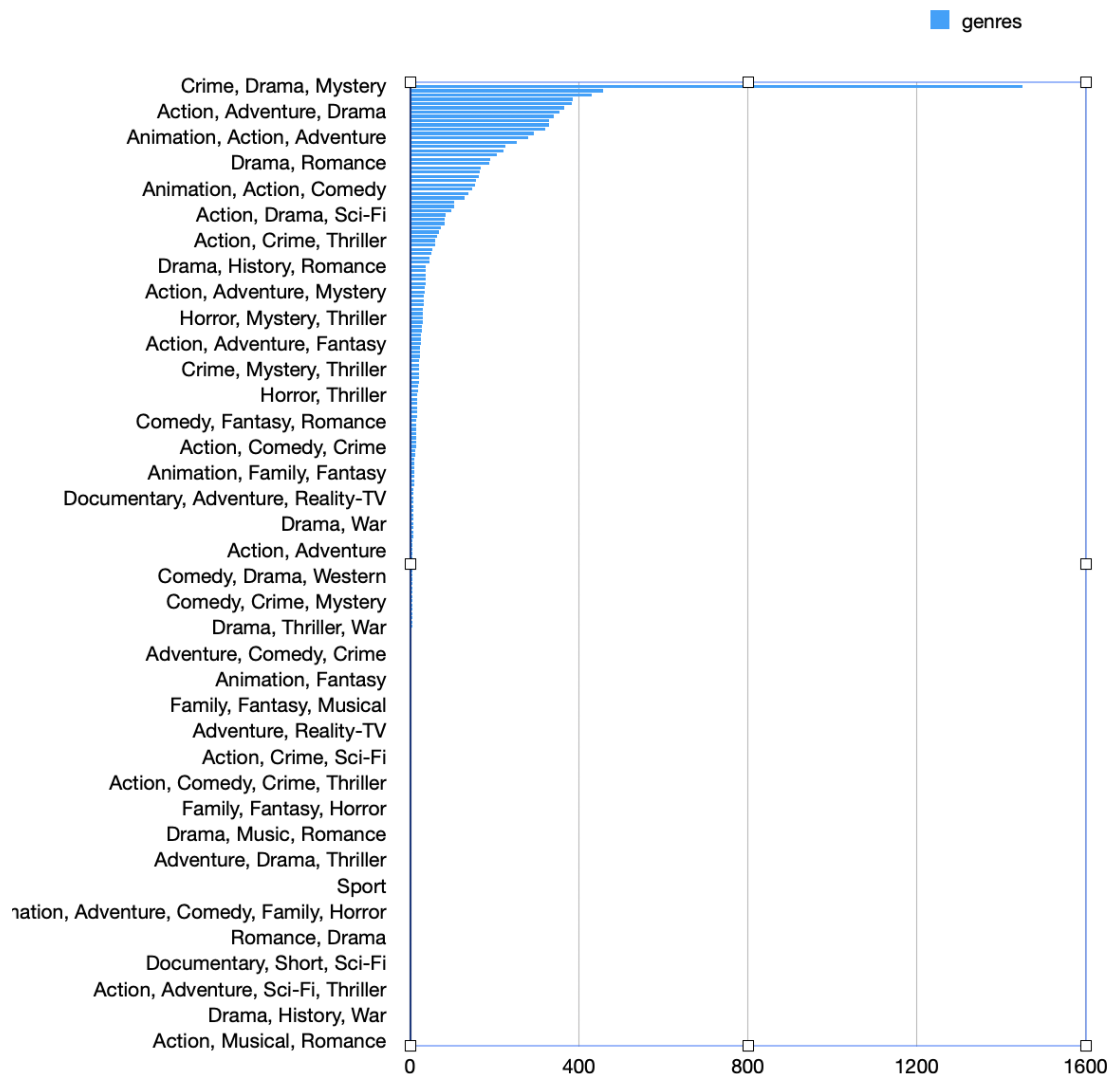
- 3) I have created my task.py file where I run my code and results were written in CSV files: **title_per, genres**.

4) It seems like top 10 titles produce 20% of views, and titles that have small % produce 80% of views.

Titles with small % produce 80% not just individuals, but whole small % titles.



For Genres section:



So I calculated % of titles that are more than >1% out of all titles.

count_big_titles = 2.17 %

count_small_titles = 97.83%

In conclusion: many titles (with small % of titles) represent a large part of the content that is watched. Also looked at all viewing in total, but also focused on specific areas, by the end of calculation, it seems genre “Crime, Drama, Mystery” are more popular than others.