Finding over-weighted osu!mania maps with Score & Star Rating

Evening

# Goal

If Star Rating (SR) perfectly represents difficulty, we would have a perfect correlation between SR and Score achieved by players. However, this is not true, and we would like to find out what beatmaps are so-called “overrated” and “underrated”.

# Procedures

* Retrieval of **players to analyze**
* Retrieving player scores of **maps to analyze**
* Flesh out data and look at correlation

## Retrieval of players to analyze

### Extracting Maps

Firstly, who we want to analyze are players that currently have scores in **ranked and loved maps**. This is to ensure that we don’t analyze “dead accounts”.

This has proven to be hard to get a **random sample** as it’s hard to avoid “dead accounts” even with recent plays (the API only retrieves a 24h backlog), this method will trim out too many players.

To get the best possible **sample**, we will look at players that have been playing the recent maps, from there, we will generate a **purposive sample**.

We retrieve all maps starting from 2015 and retrieve scores from each of these maps for a total of **4621 maps**.

Code is provided in Annex

### Getting an effective random sample of maps

Next, we run the data through KNIME for this task

#### Filtering out unnecessary data

To cut down on API retrieval load we:

* Remove all maps below 3.0 Star Rating (Non-productive data)
* Remove all maps above 10.0 Star Rating (Outlier data)

We end up with **2282 maps** to analyze.

#### Random sampling of remaining data

We will randomly take 20% of the data to analyze, leaving us with **456 maps** to analyze.

# Annex



