

SLICED SEASON 0 EPISODE 3

Data comes from <https://www.kaggle.com/michau96/popularity-of-games-on-steam>

There are two datasets:

- `sliced_data.csv`
- `sliced_holdout_data.csv`

Objectives

Modeling: Predict the `volatile` column with the best `accuracy` (30 points for the predictions with the highest accuracy)

Visualization: Make compelling data visualization (worth up to 30 points)

Golden Features: worth up to 15 points, 5 points for being the first person to discover a Golden Feature

Audience: 10 points for being the most beloved by chat

Target Variable: `volatile`

This is a custom feature that describes whether the gains in avg number of players for a game were within the 25th - 75th percentiles in comparison to other games within a given timeframe. There are three responses: -1, 0, 1. If the volatility is -1 then the avg player gains were below the 25th percentile; if the volatility is 1 then the avg player gains were above the 75th percentile; if the volatility is 0 then the avg player gains were within the 25th and 75th percentiles.

You can think of this variable as two layers: a binary volatile-or-not variable and a directional variable for volatility (above-average or below-average).

Holdout data: Skyrim

The Elder Scrolls v: Skyrim is the holdout dataset. The object is to predict volatility of Skyrim over time.

Holdout columns

The following columns are **NOT** in the holdout dataset:

`volatile`

The above column does not exist in the holdout data!

`sliced_data.csv`

`gamename` : str - video game name

`year` : int - year

`month` : str - month name

`avg` : float - average number of players at the same time

`gain` : float - difference in average compared to the previous month (NA = 1st month)

`peak` : int - highest number of players at the same time
`avg_peak_perc` : str - share of the average in the maximum value (avg / peak) in %
`month_num` : int - month in numeric form
`yearmonth` : str - date in YYYY-MM-DD format. Note there is no actual day, the date formatting defaults to the 1st of the month
`volatile` : int - the volatility of users gained based on all other games, with some adjustments for time context

holdout_sliced_data.csv

All of the columns are the same as in `sliced_data.csv` except for `volatile`.