**Spotify predict which features make a song popular?**

Heather Lemon

Department of Computer Science, University of Denver

COMP 4449 Capstone Midterm

Dr. Claudio Delrieux

March 30, 2023

**Spotify predict which features make a song popular by Heather Lemon**

<intro about spotify & vocabulary >

**Description of the Dataset**

< describe dfs>

**Data Preprocessing**

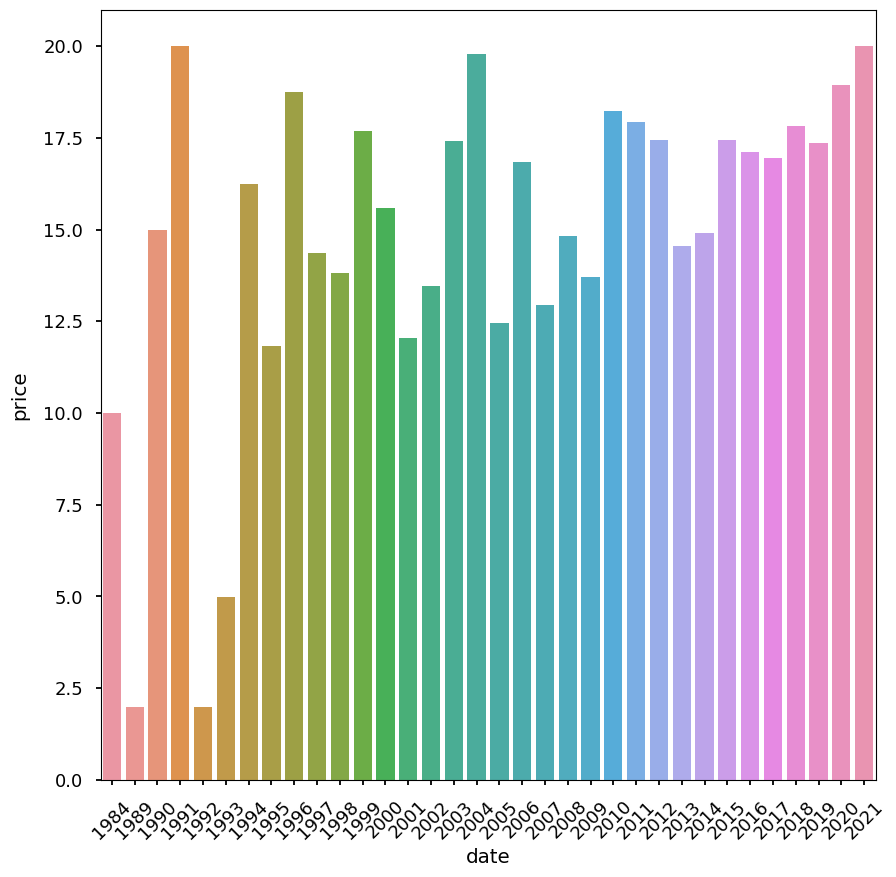
<data preprocessing step>

**Exploratory Data Analysis**

**<EDA>**

**Figure 1**

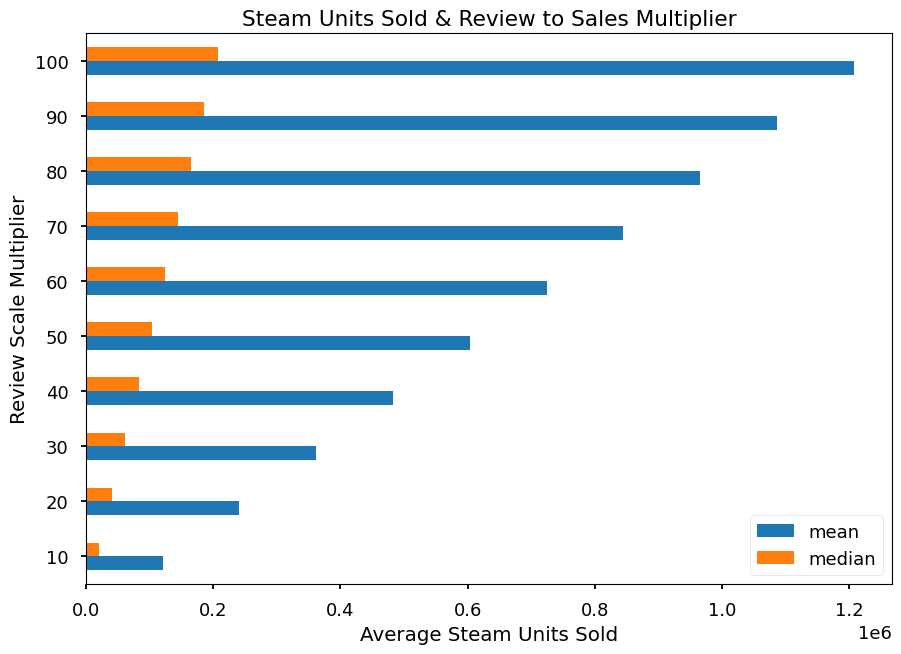
*Yearly Average Price of a video game sold on Steam*



*\*Note: price change, there could have been a steam sale that week when the data was initially pulled and when you go look at that game today, the price can be cheaper, or the game completely removed.*

**Figure 2**

*Estimated Steam Units Sold and Review to Sales Multiplier*



*\*Note: we see a linear relationship with the scaling factor.*

**Figure 3**

*OLS Summary Report*

**Text

Description automatically generated**

**Figure 4**

*Target Distribution Rate*

Chart, histogram

Description automatically generated

**Figure 5**

*Binned Price per Group*

*Chart, box and whisker chart

Description automatically generated*

**Model Building and Evaluation**

<model selection and evaluation>

From this figure shown of each of the models and their accuracy score.

**Figure 4** *Model Selection*

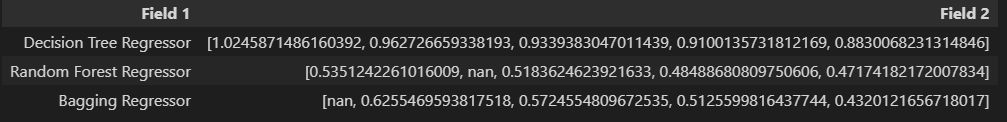
Application

Description automatically generated with medium confidence

*Left: Model, Right: Accuracy Score*

We can see that the Decision Tree Regressor is our highest performing model with 100% accuracy which means we are overfitting. Followed by Random Forest Regressor and Bagging Regressor models. I selected the top three performing models and begun hyperparameter tuning phase. Where we start with running a cross validation score on each of the base models and printing the scores.

**Figure 5** *Cross Validation Scores*



The Decision Tree Regressor is sill the best performing model out of the top three.

After running the GridSearchCV and tuning the hyperparameters that looks like this

**Figure 6** *Hyperparameter Tuning*

Text

Description automatically generated

*Hyperparameter Tuning Decision Tree Regressor*

**Model Selection**

Model selection

**Conclusion**

In answering the question which features best predict what makes a song popular.

**Lessons Learned**

<lesson learned>

**Future Improvements**

Although we were unable to use any of the textual data,

**References**

[0] Valve Inc., <no more information was found>

<https://www.valvesoftware.com/en/about>

[1] (Aug 1, 2021) Video Game Insights,

<https://vginsights.com/insights/article/how-to-estimate-steam-video-game-sales>

[2] Simon Carless (Aug 3, 2020) How that game sold on Steam, using the 'NB number'.

<https://newsletter.gamediscover.co/p/how-that-game-sold-on-steam-using>

[3] Evgeny Obedkov (Nov 15, 2021) Steam reviews/sales ratio approaching 60x for new hit games, so how do we estimate sales in 2022?

<https://gameworldobserver.com/2022/11/15/how-to-count-game-sales-steam-2022-review-multiplier>

[4] Raka Mahesa (Oct 2, 2017) The Difference Between Working in Indie and AAA Game Development

<https://hub.packtpub.com/difference-between-working-indie-and-aaa-game-development/>

[5] SteamDBInfo, <no more information was found>

<https://steamdb.info/app/990080/charts/>