Predicting song popularity

Student name: Anastasia Zhivilo

Student ID: SID450608025

Uni key: azhi4513

The Goal and The Data

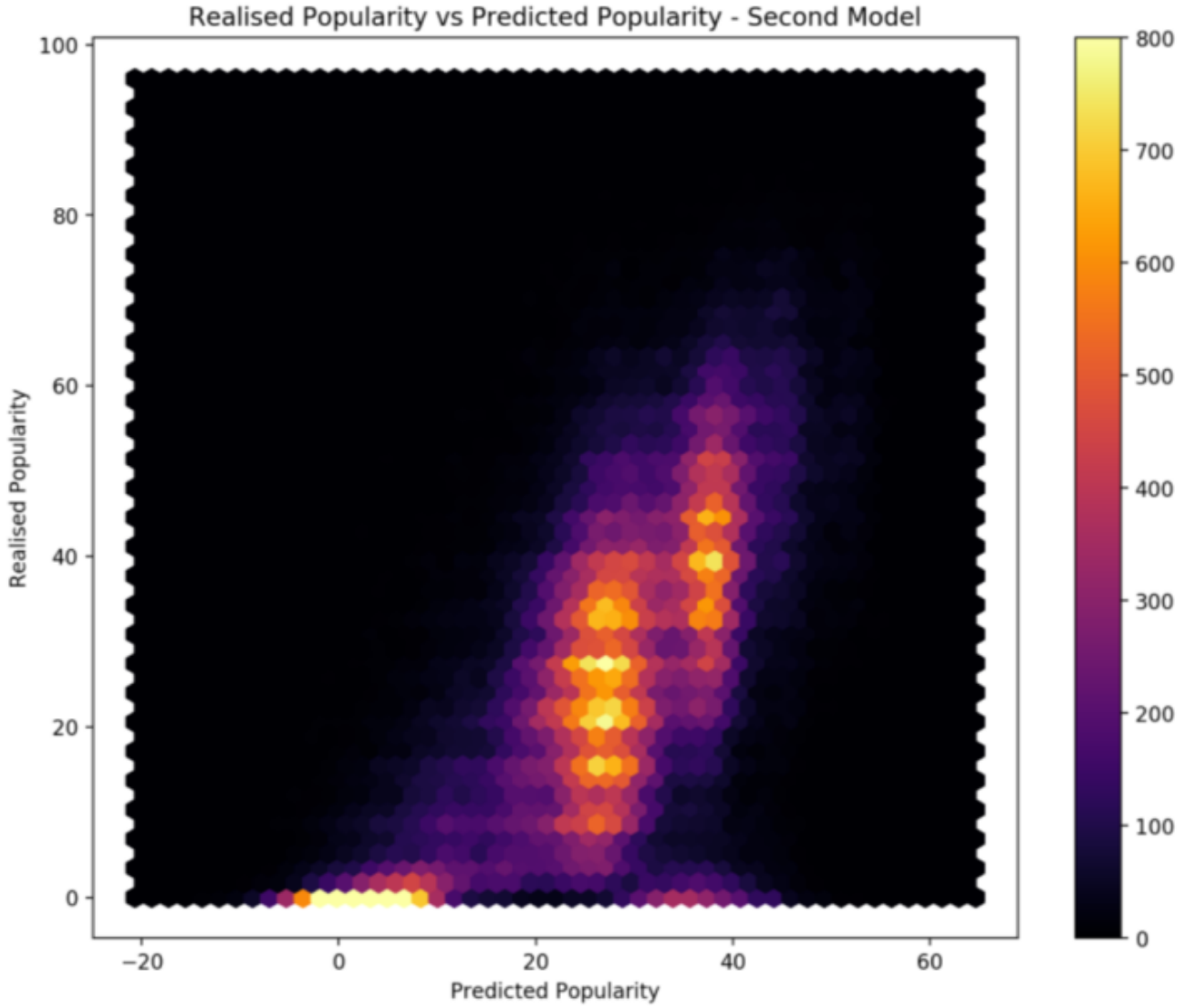
- Creating a model to predict popularity of a song
- Data from Spotify with song attributes and popularity score out of 100
- Data covers songs released from January 1900 to April 2021, with popularity recorded in 2021.

Approach and Evaluation

- Multiple Linear Regression Model used to predict popularity
- R-Squared and p-values were used for evaluaton
- The model was tuned by adding dummy variables, non-linear variables; and removing fields that were not statistically significant
- The 2021 data was split from the rest and the model was built on pre-2021 data. Tuned model was fitted on 2021 data and predictions compared to actual popularity scores.

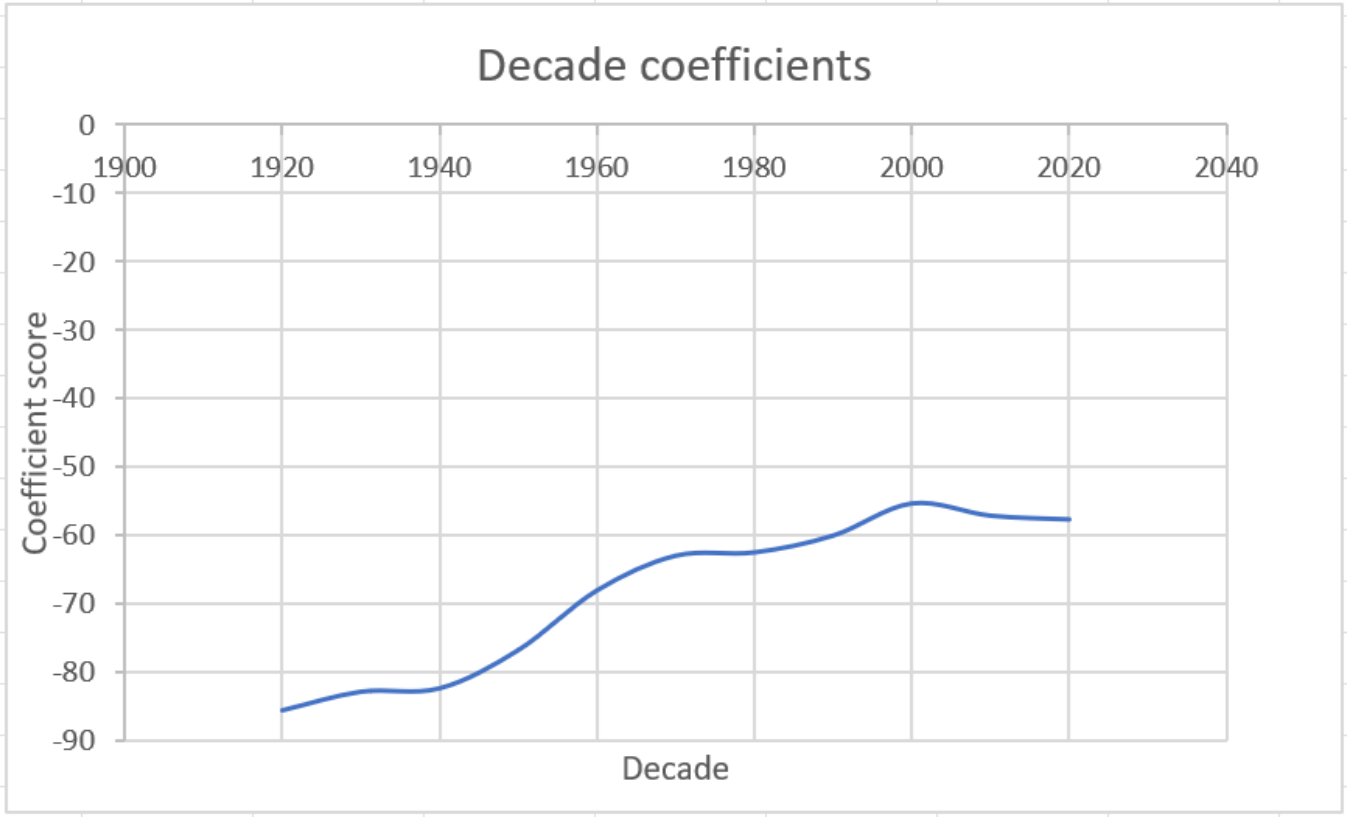
Actual vs Predicted popularity scores (pre-2021)

- Trained model had R-squared of 0.4
- Predicted vs Actual popularity heat map has upwards linear trend.
- Model predicts well unpopular songs but has a few clusters that could further be explored



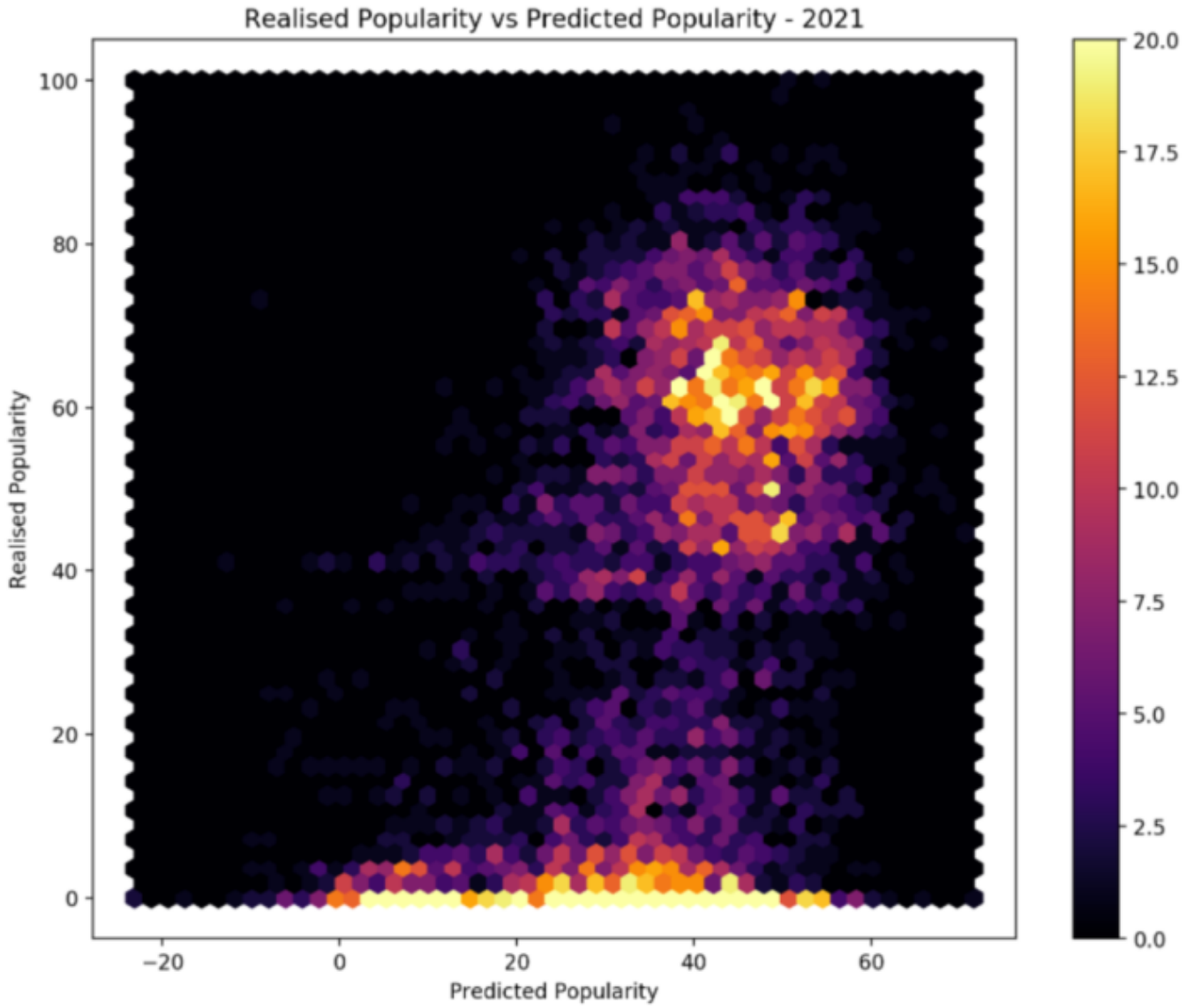
How the decade the song was released in influenced its popularity

- Older decades pushed popularity score down
- Newer songs had a dip in popularity compared to 20 year old songs
- Peak popularity reached at 20 years old



Actual vs Predicted popularity scores (2021 data)

- Modelling on 2021 data also produces general upwards trend
- Cluster of songs where predicted and actual scores were similar.
- Cluster of songs with predicted high popularity but actual 0 popularity, possibly songs with high potential for high popularity in the future.



Further improvements

- Adding interaction variables
- Adding in artist information
- Removing the songs that have a popularity of 0
- Interesting to explore how songs with predicted high popularity score in 20 years time.

Thank you!