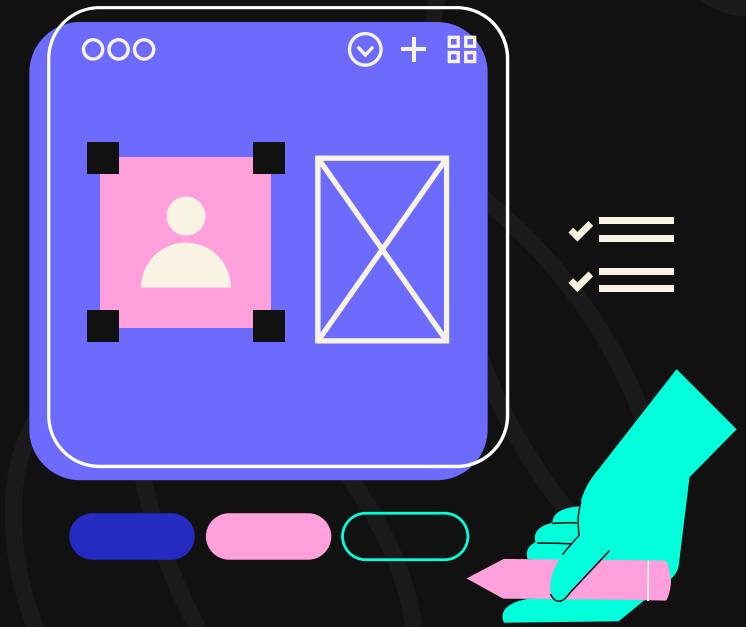
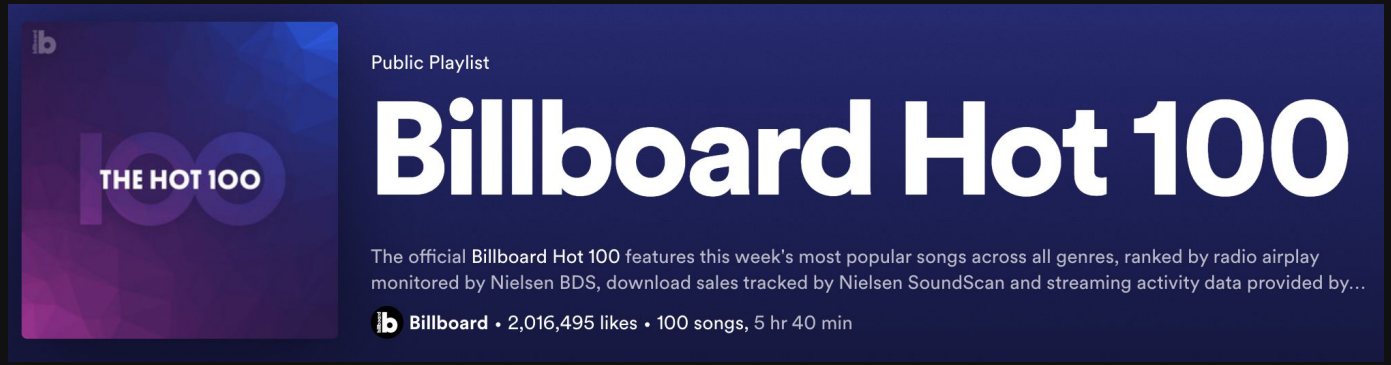
Predictability of music trends in time

Maxim Velli, Eva Koskova



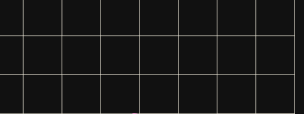
CSS

Dataset



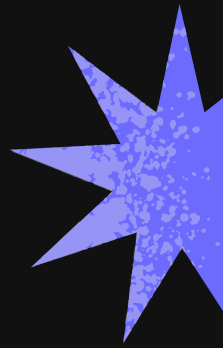
The Spotify Hit Predictor Dataset (1960-2019)
Over 40,000+ Tracks labeled hit or flop, with their features.

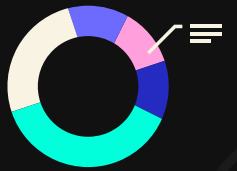
Farooq Ansari (2019)



Dataset - features

- danceability
- energy
- key
- loudness
- mode
- speechiness
- acousticness
- instrumentalness
- liveness
- valence
- tempo
- duration_ms
- time_signature
- chorus_hit
- sections
- target





Farooq Ansari (2019)

Motivation

Music taste development in society

Predictability of popular music

Musical features that make a hit

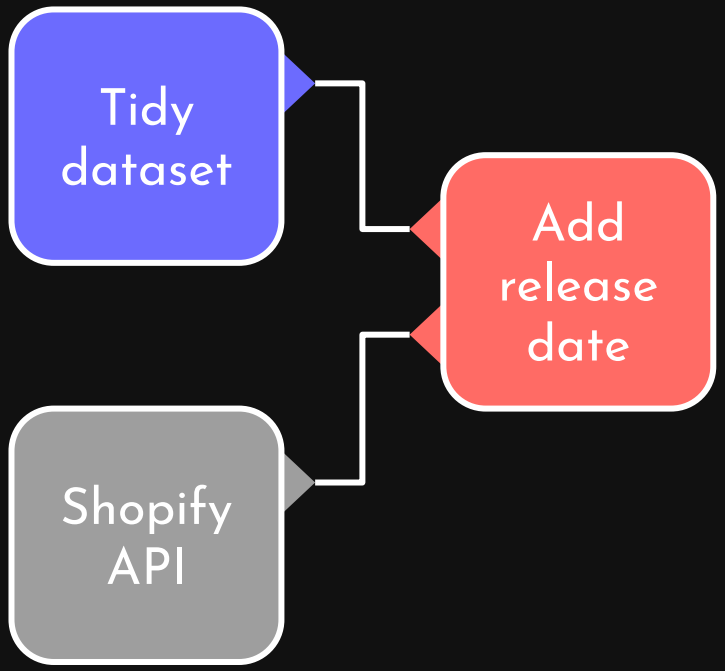
Research questions

Cross-timeframe hit prediction (10/5/3 years)

Feature importance

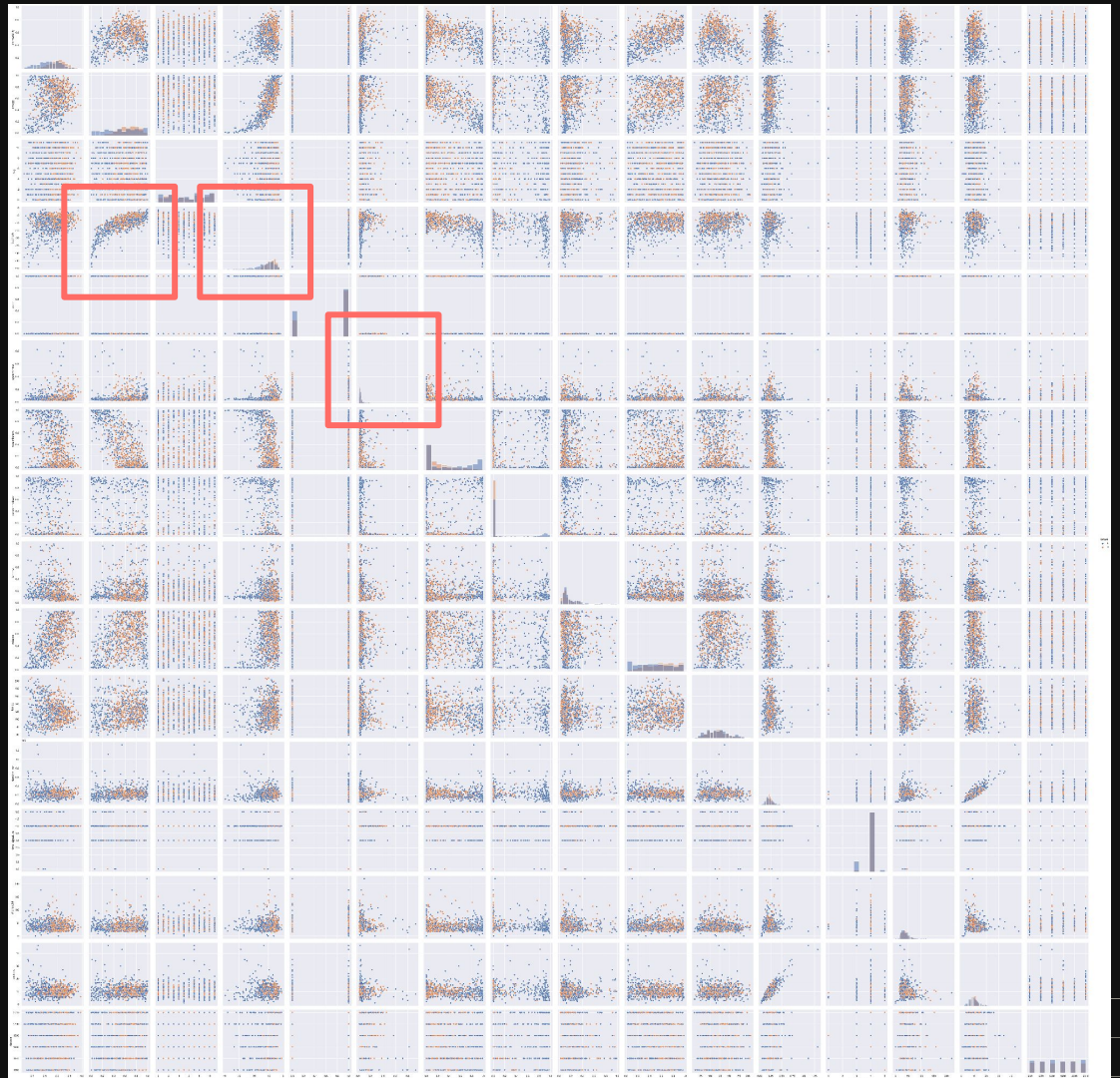
Artists’ adaptivity

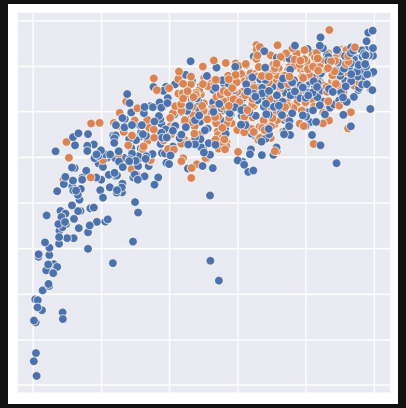
Data extraction & cleaning

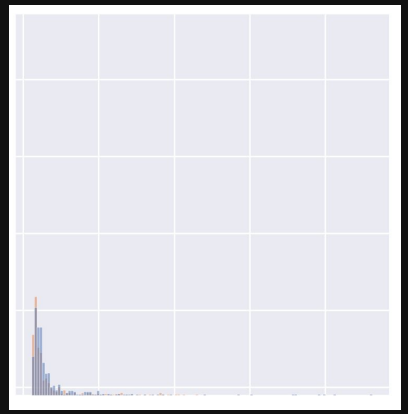


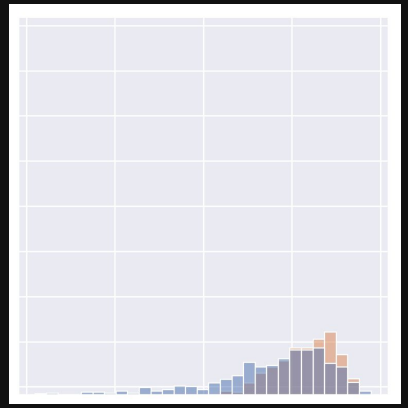
What can we expect?

Histogram-correlation for 150 randomly sampled songs from every decade









Some redundancy

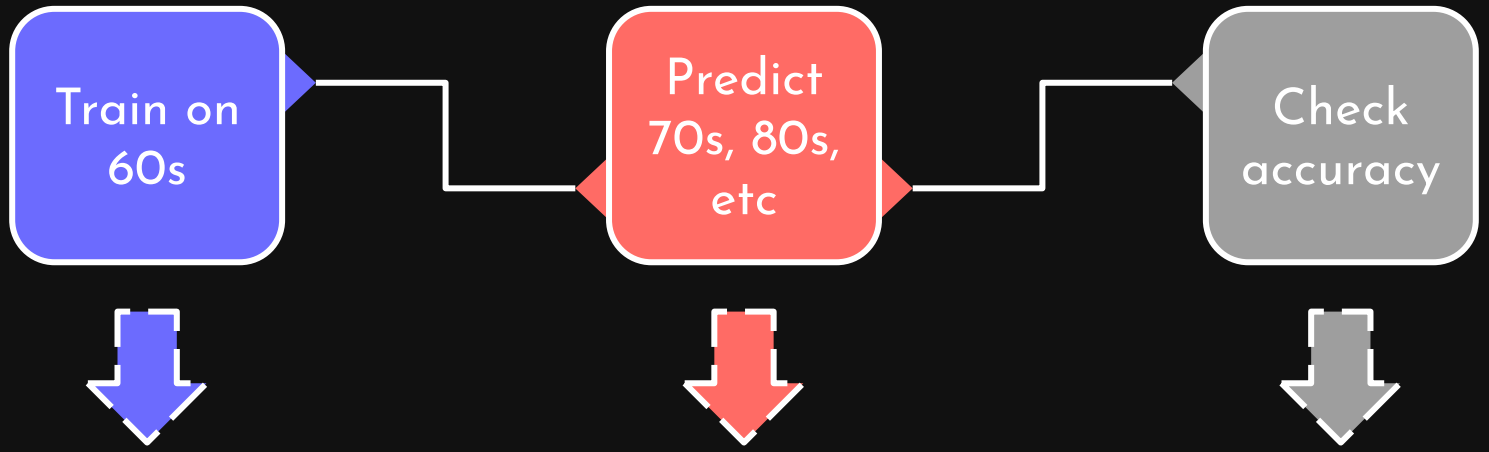
Normalisation

Characteristics of hits

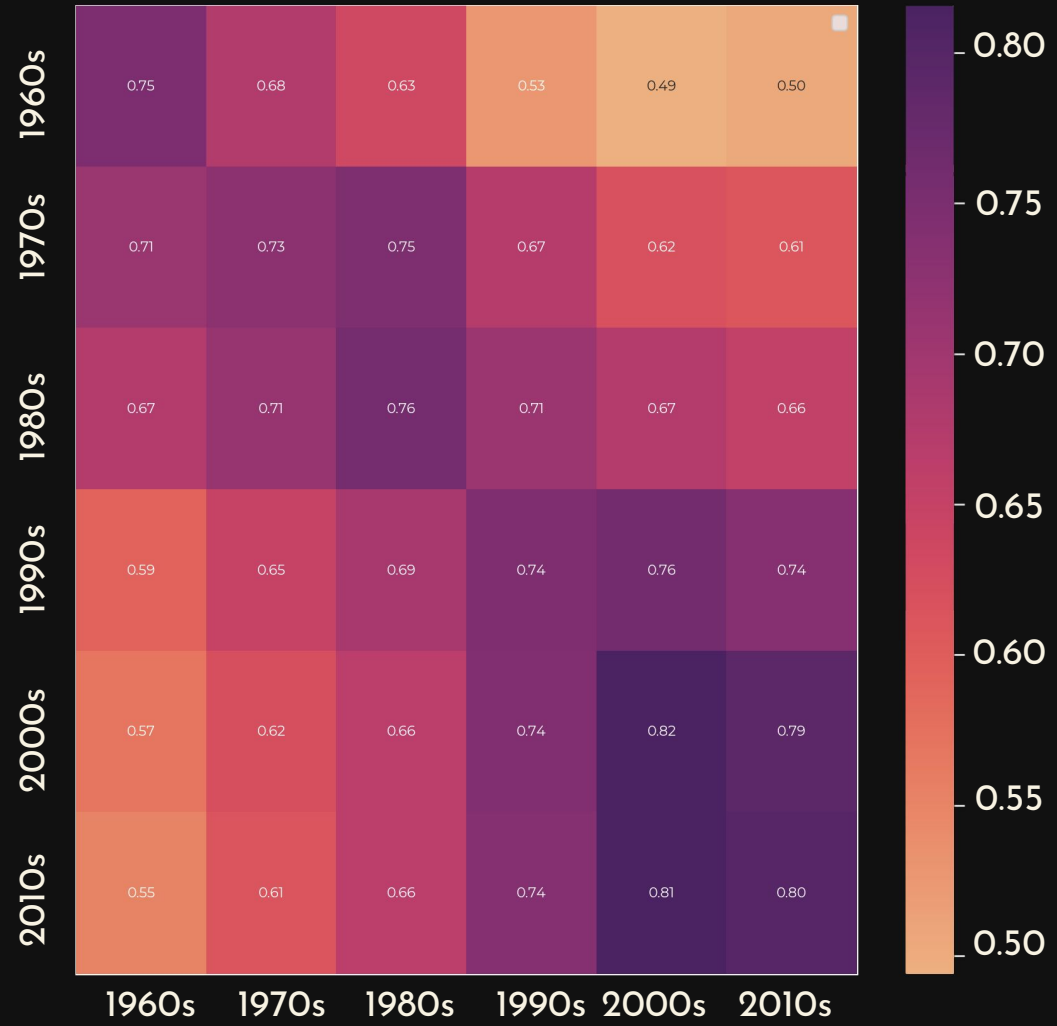
= hit

= not hit

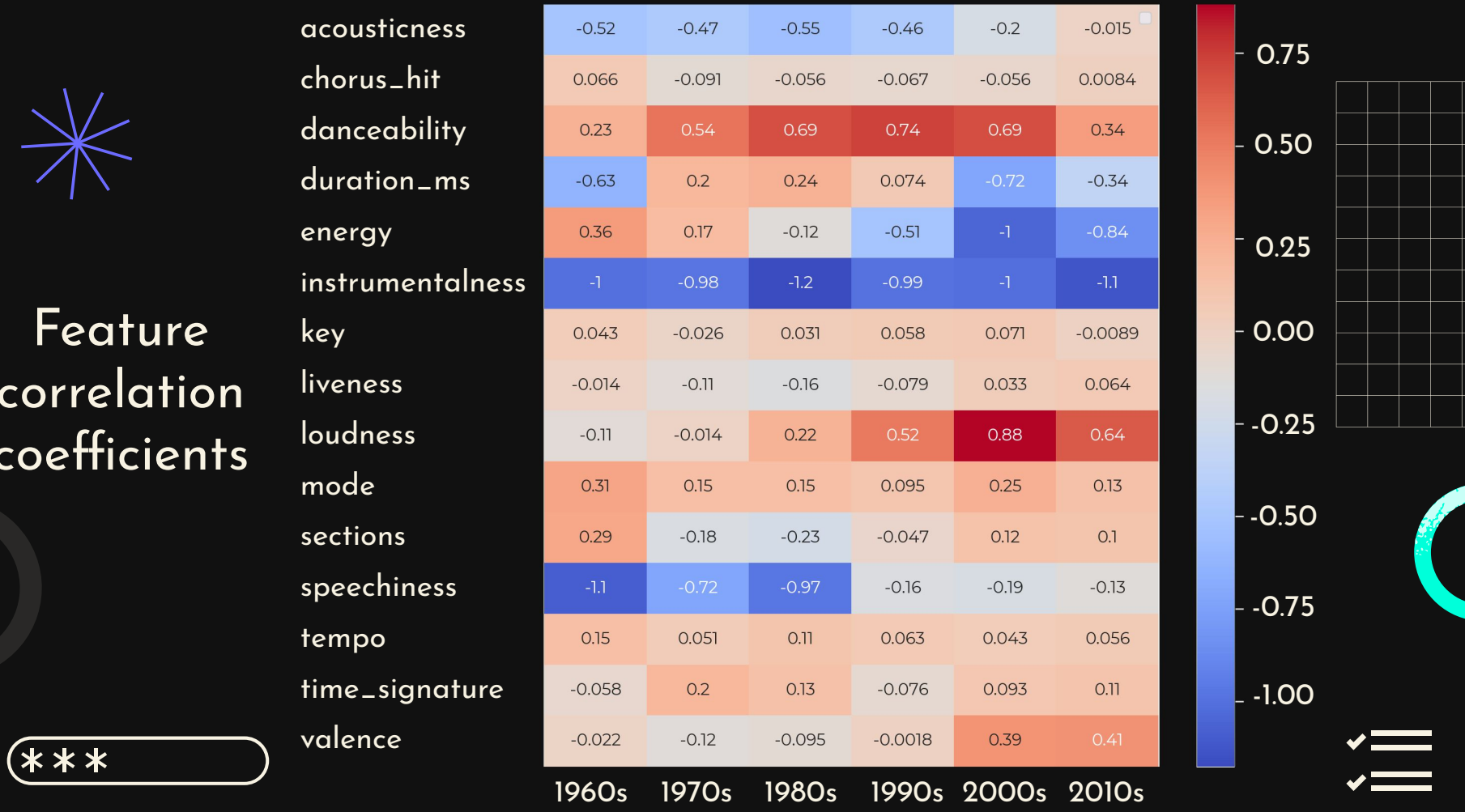
Predicting hits across decades



Prediction accuracy heatmap



Feature correlation coefficients

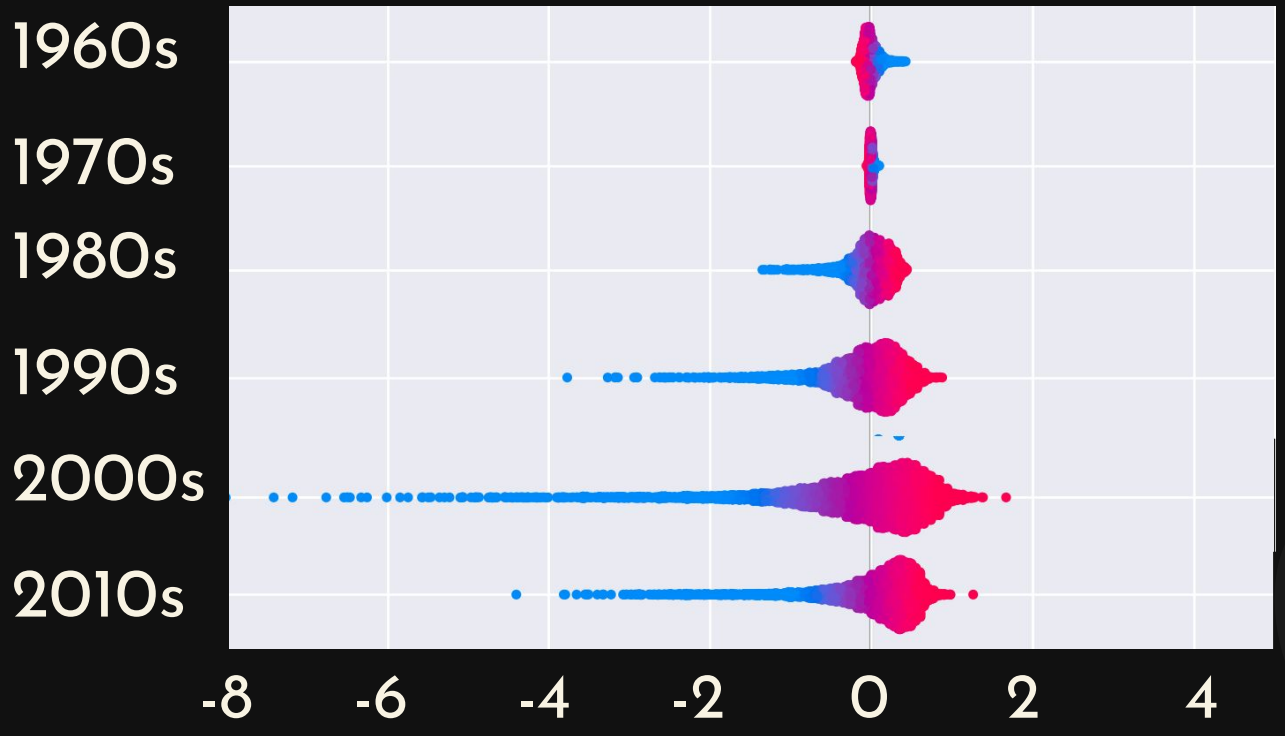


SHAP Values: Loudness

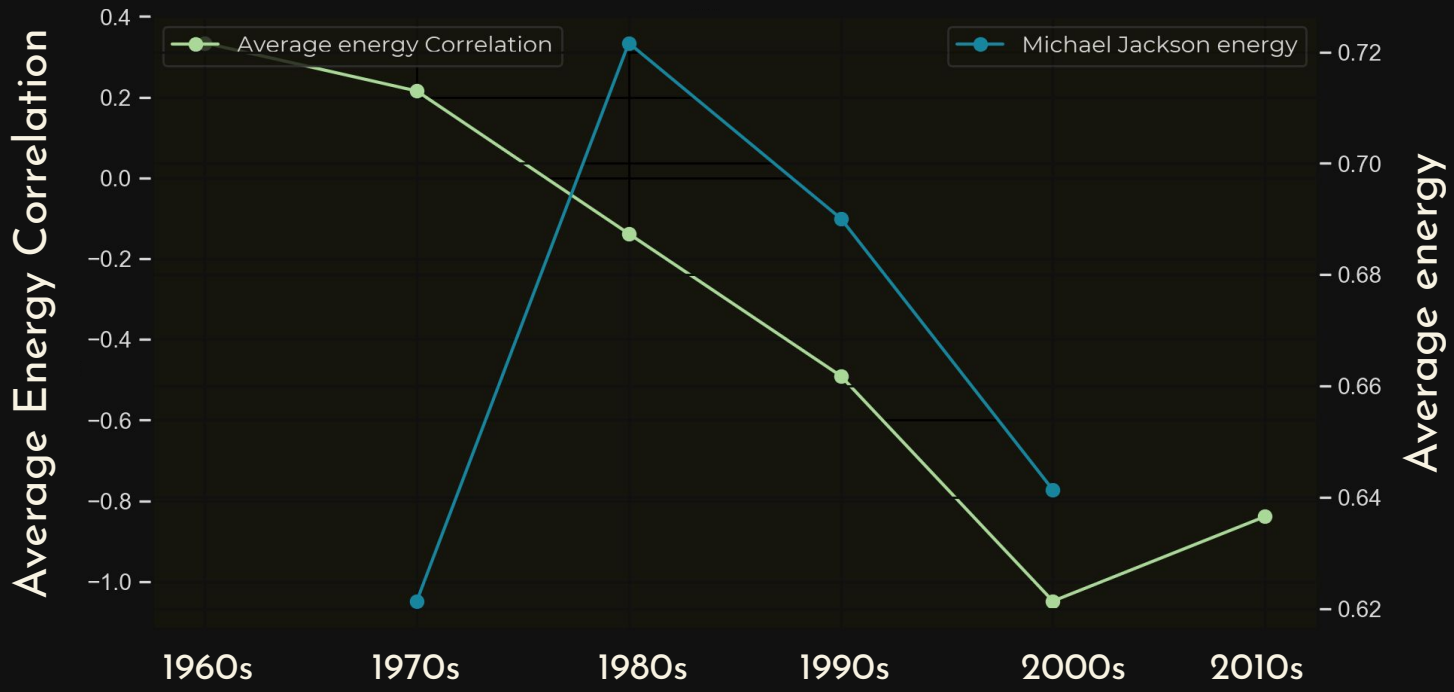
is a visual representation of the contribution of each feature in the model to the prediction for each instance in the dataset

- loudness is the amplitude of the wave
- loud tracks have more positive impact on the accuracy of the predictor (after 1980)
- quiet tracks negatively affect the accuracy of the predictor

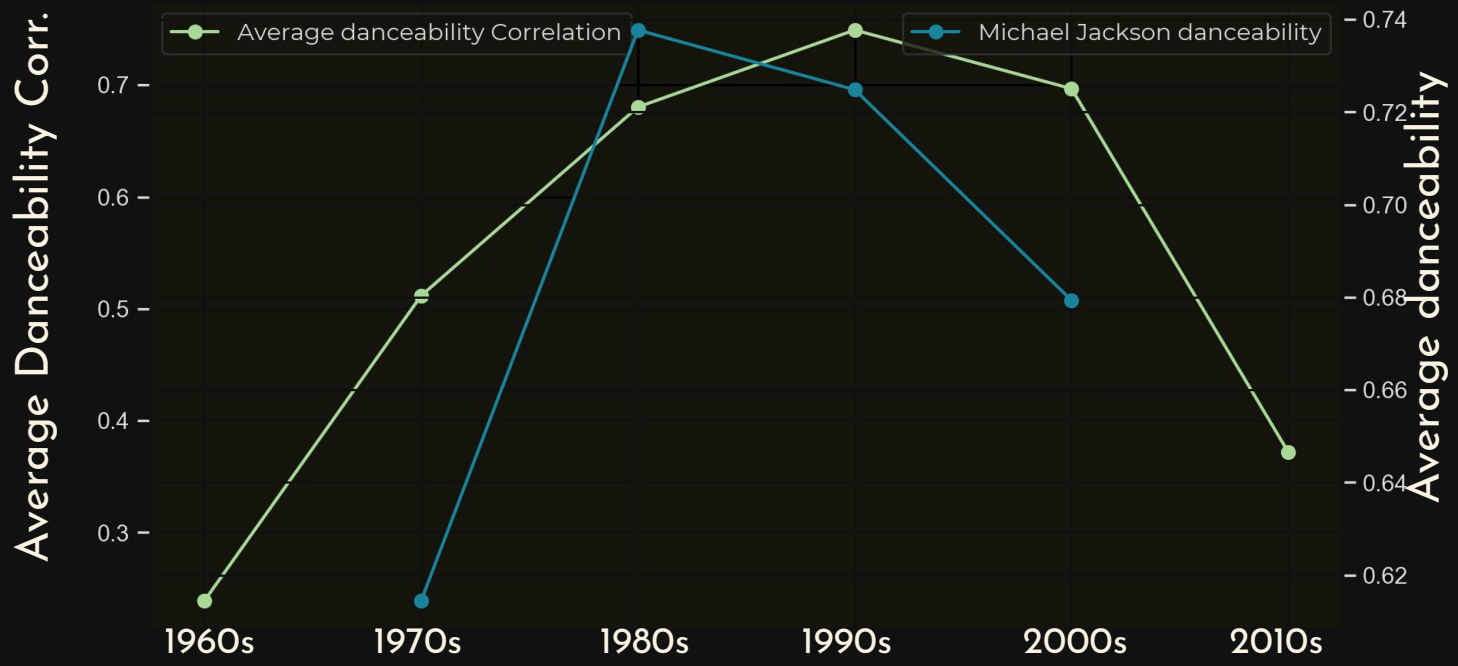
SHAP Values: Loudness



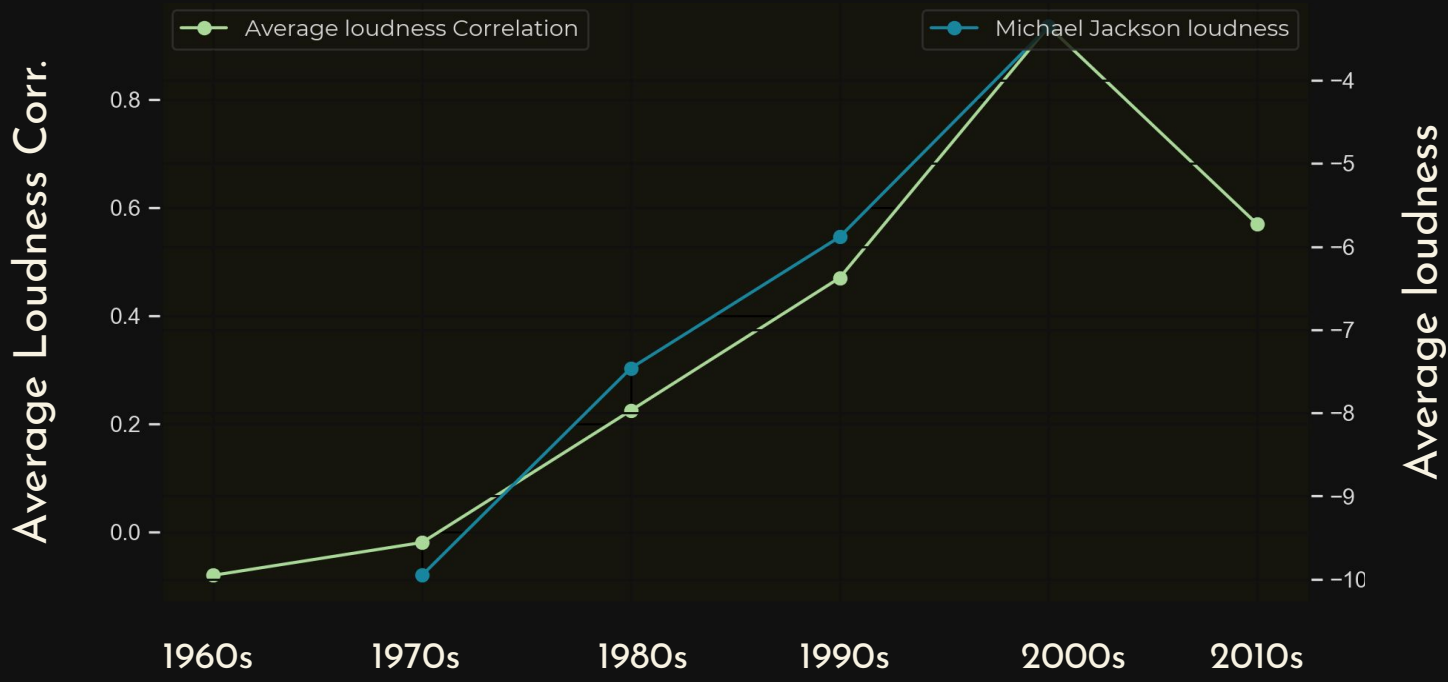
Following trends: Michael Jackson



Danceability



Loudness


Average loudness Correlation
Michael Jackson loudness
Average Loudness Corr.
Average loudness
0.8
0.6
0.4
0.2
0.0
−4
−5
−6
−7
−8
−9
−10
1960s
1970s
1980s
1990s
2000s
2010s

Let's check out hit musicians!

Prince

```
Prince  -  1999 ( 1982 )
1960: not a hit
1970: hit
1980: hit
1990: hit
2000: hit
2010: hit
```

Drake

```
Drake  -  Hotline Bling ( 2016 )
1960: hit
1970: hit
1980: hit
1990: hit
2000: hit
2010: hit
```

Conclusion

Popular music is becoming less variable

Popular music is reasonably predictable

Artists follow some features

Discussion

Confidence intervals for correlation coeff.

Alternatives to Logistic Regression

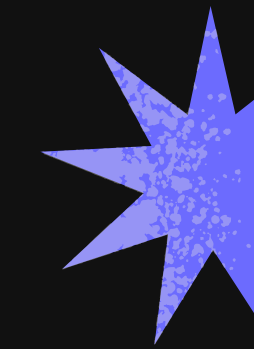
Reducing dimensionality

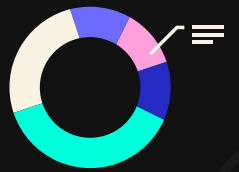
Attributions

Maxim: Brainstorming, Data Extraction, Predictive Accuracy & Feature Correlation Analysis, Data Visualisation, Hit Predictor

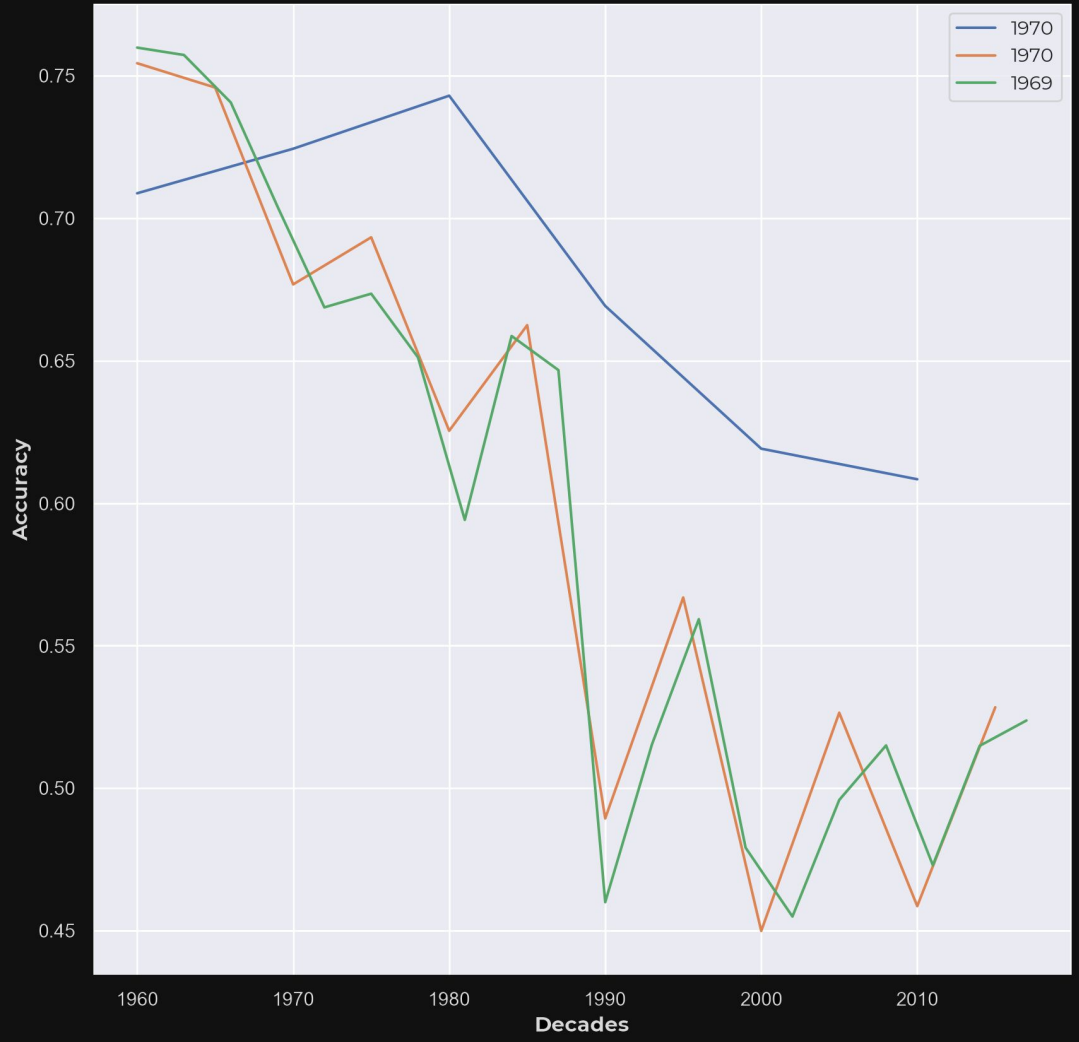
Eva: Brainstorming, Data Cleaning, SHAP Feature Importance Analysis, Artist Feature Trend Analysis, Significance Report (attempted)

Marc: Brainstorming, consultation

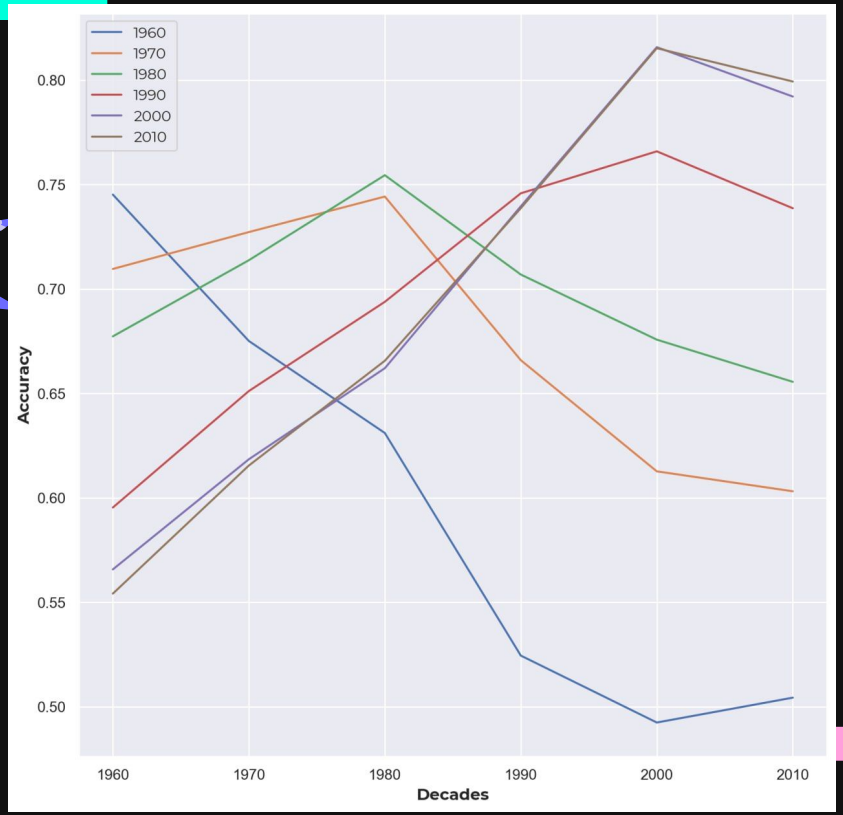


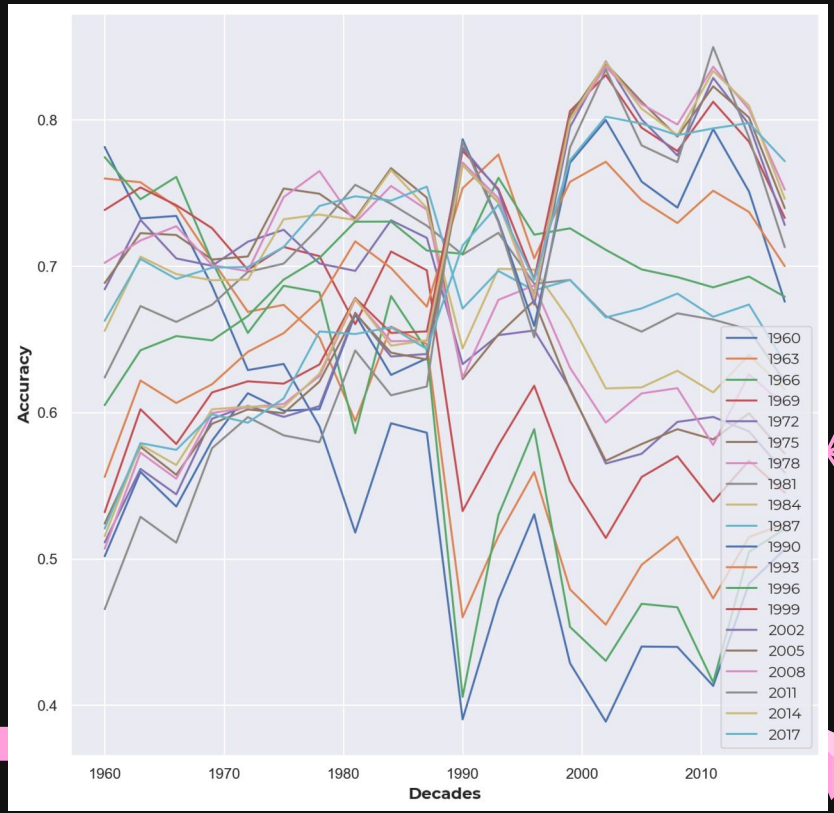


Discussion


1970
1970
1969
0.75
0.70
0.65
0.60
0.55
0.50
0.45
Accuracy
1960
1970
1980
1990
2000
2010
Decades

Discussion


1960
1970
1980
1990
2000
2010
0.80
0.75
0.70
0.65
0.60
0.55
0.50
Accuracy
1960
1970
1980
1990
2000
2010
Decades


0.8
0.7
0.6
0.5
0.4
Accuracy
1960
1970
1980
1990
2000
2010
Decades
1960
1963
1966
1969
1972
1975
1978
1981
1984
1987
1990
1993
1996
1999
2002
2005
2008
2011
2014
2017