Music genre analysis and prediction using audio features

By Uroš Petković

Can machines listen to the music?

Project Outline

- Getting the Data with Spotify API
- Data Cleaning and Wrangling
- Visual Data Analysis
- Developing Machine Learning Model to predict the genre of given artist

Audio Features are mathematical measurements computed directly from the sound wave

- Danceability
- Energy
- Key
- Loudness
- Mode
- Speechiness
- Instrumentalness
- Liveness
- Valence
- Tempo
- Duration
- Time Signature

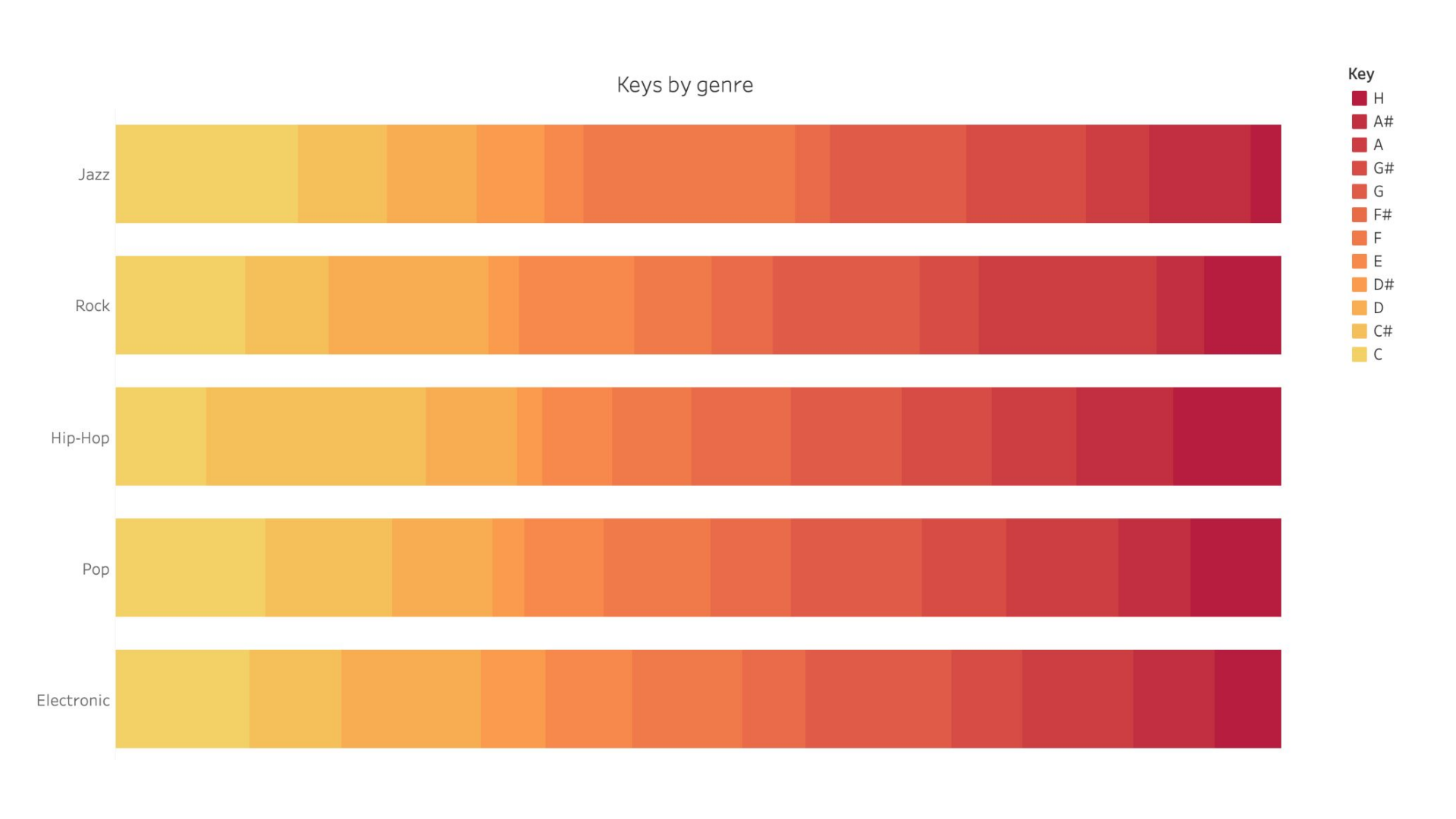
Getting the Data with Spotify API

- Finding playlists based on genre
- Scraping all artists from playlists
- Scraping all songs from artists
- Getting audio features

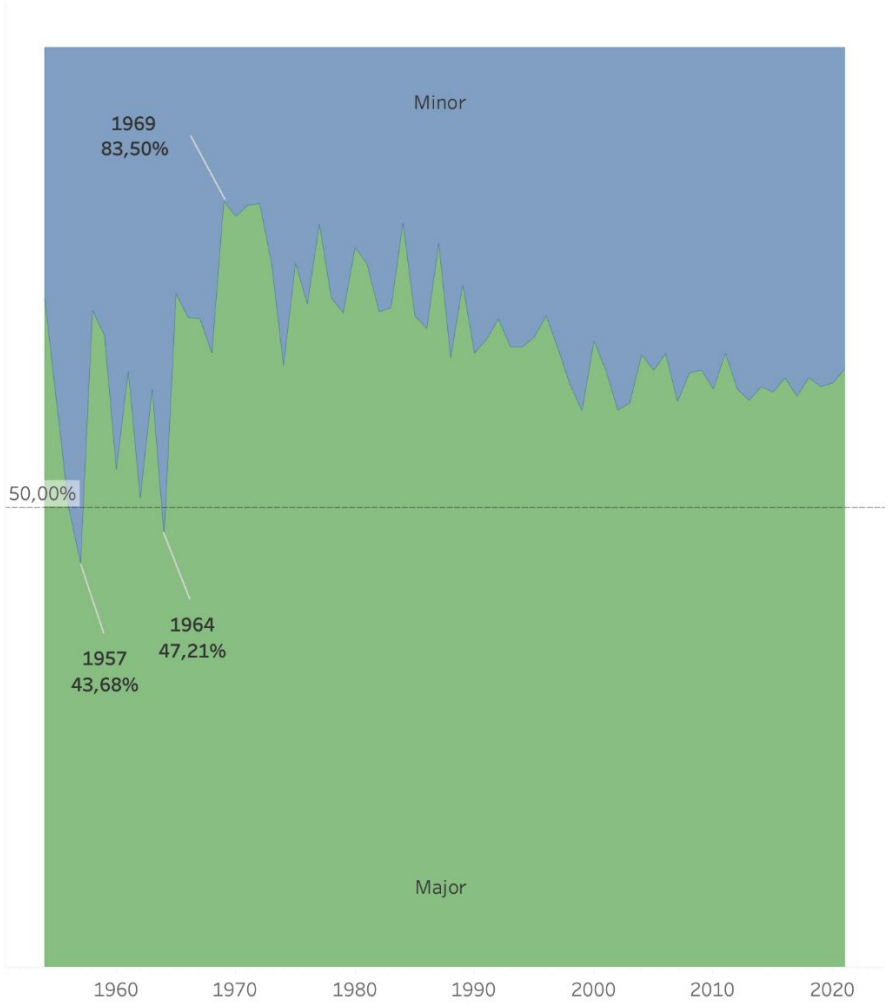
Data Cleaning and Wrangling

- Defining genres
- Adding release years
- Loading all songs in a DataFrame
- Deleting duplicates
- Fixing tempo values range from 60 BPM to 200 BPM

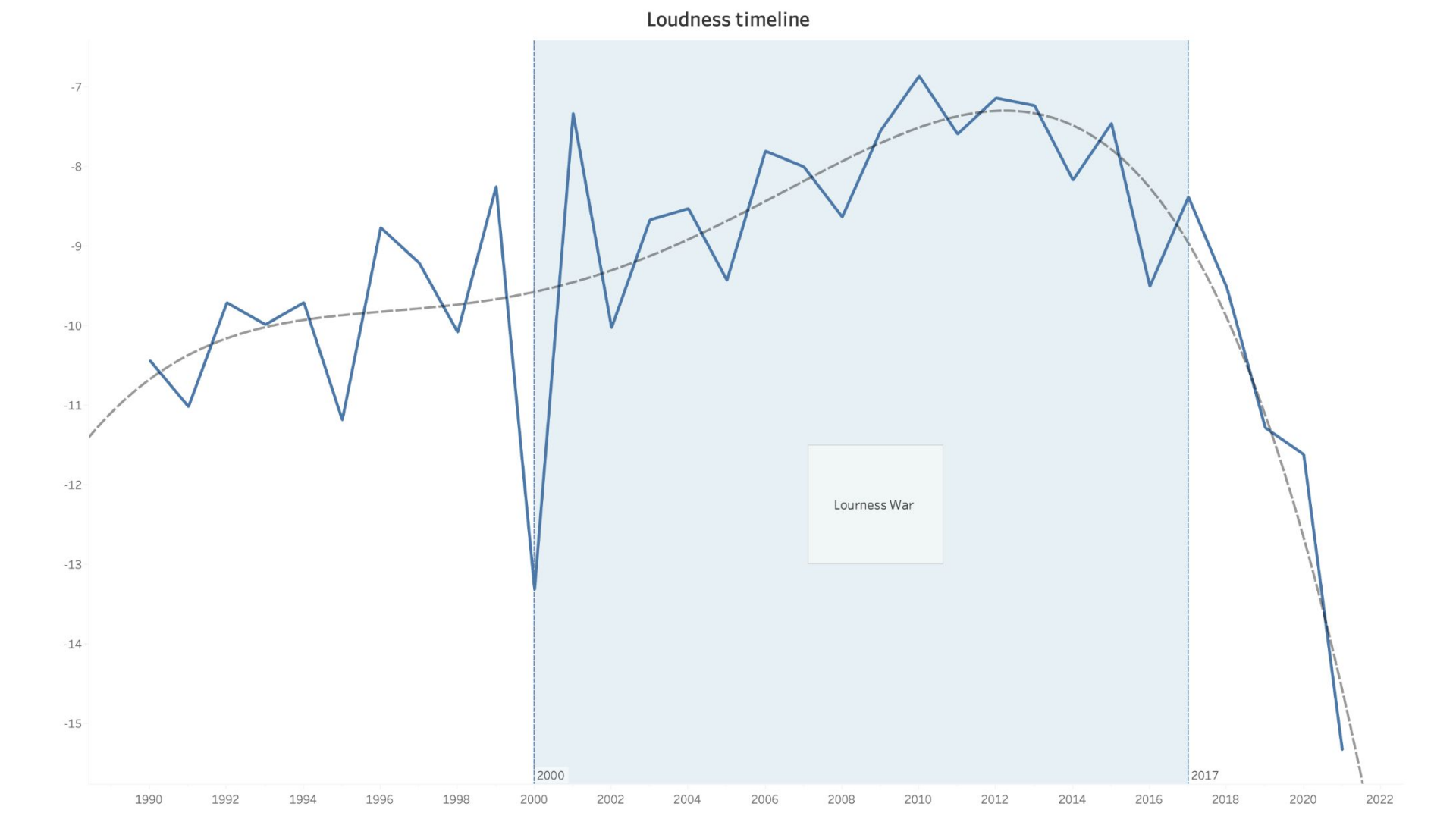
Visual Data Analysis with Tableau

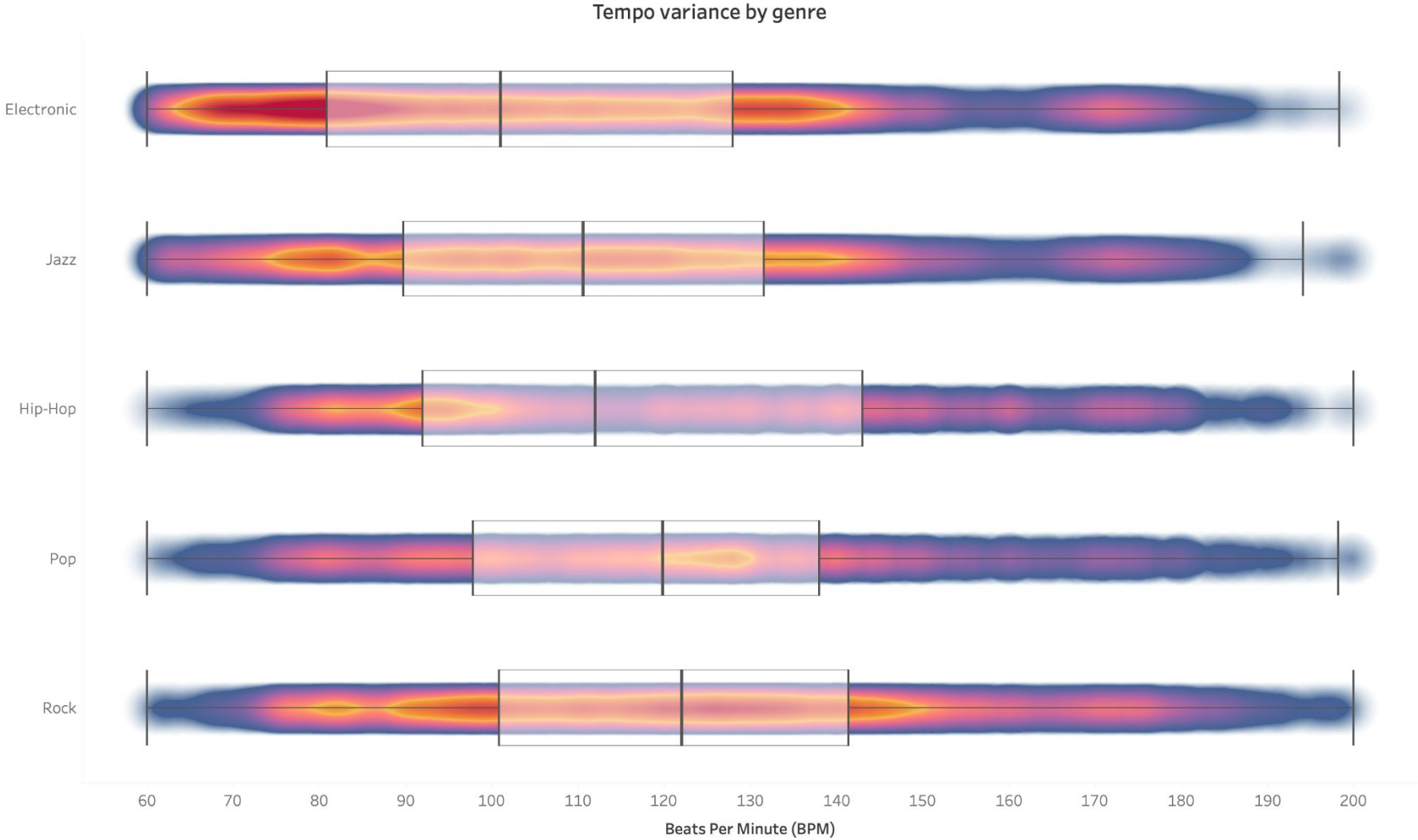

Keys by genre
Jazz
Rock
Hip-Hop
Pop
Electronic
Key
H
A#
A
G#
G
F#
F
E
D#
D
C#
C

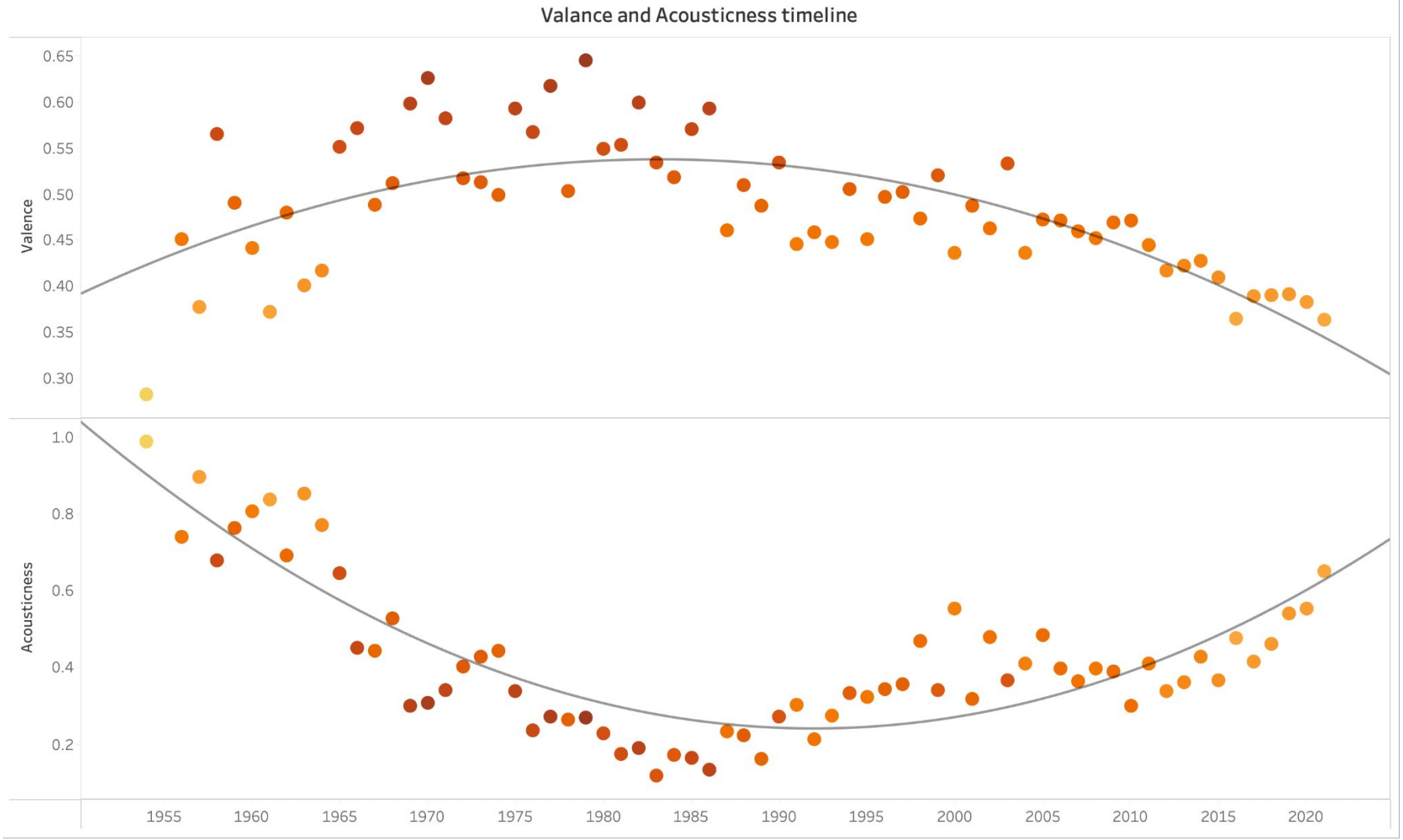
Major and Minor scales timeline



Modes by genre


Loudness timeline
-7
-8
-9
-10
-11
-12
-13
-14
-15
Lourness War
2000
2017
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022

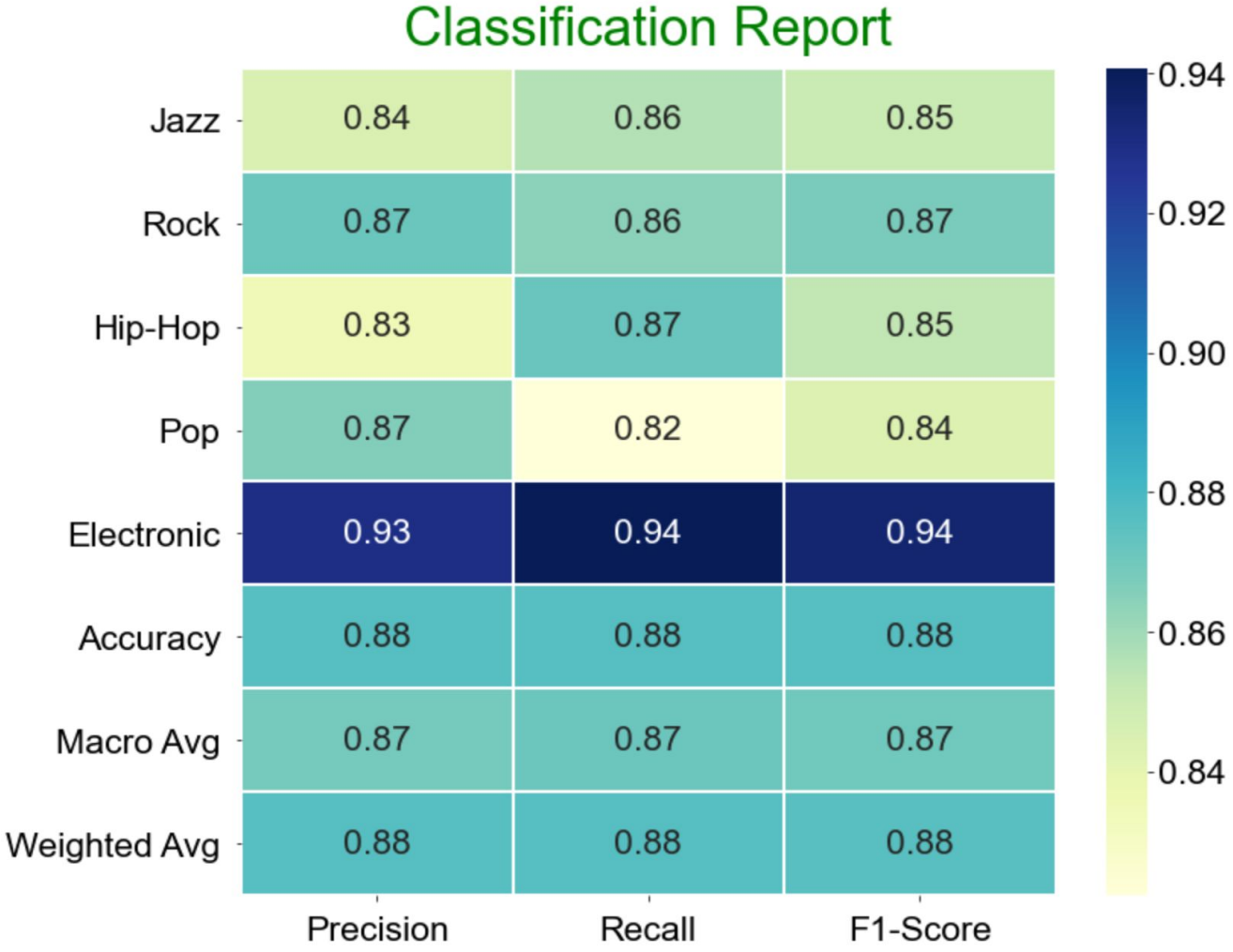

Tempo variance by genre
Electronic
Jazz
Hip-Hop
Pop
Rock
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
Beats Per Minute (BPM)


Valance and Acousticness timeline
Valence
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
Acousticness
1.0
0.8
0.6
0.4
0.2
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020

Developing and Deploying a Model to predict the genre of a given artist

Machine Learning:

- Defining Random Forest Classifier model
- Training with 70% of Data
- Predicting
- Checking Accuracy Score
- Using Randomized and Grid search to tune parameters of the model
- Re-training with new parameters
- Predicting
- Final Accuracy Score 88%


Classification Report
Jazz
Rock
Hip-Hop
Pop
Electronic
Accuracy
Macro Avg
Weighted Avg
0.84
0.86
0.85
0.87
0.86
0.87
0.83
0.87
0.85
0.87
0.82
0.84
0.93
0.94
0.94
0.88
0.88
0.88
0.87
0.87
0.87
0.88
0.88
0.88
Precision
Recall
F1-Score
0.94
0.92
0.90
0.88
0.86
0.84