Regional Music Popularity Prediction

Team 32: Snigdha Verma, Sanjana Date, Chloe Saleh, Dennis Frank, Ananya Sharma

GOAL

- **Predict song popularity and the most important acoustic features for every region around the world.**
- **Design an interactive visual component to display the information.**

DATASET

Spotify API - over 220,000 songs & acoustic features.

INTUITION

Existing models **lack precision** in predicting song popularity. **Integrating predictions into interactive maps** offers a new way to **visualize regional music preferences.**

MODEL SELECTION

Ran and evaluated 6 models

- K Neighbors
- Decision Trees
- Logistic Regression
- Random Forests
- Linear SVC
- XGBoost (Extreme Gradient Boosting)

Best Model: Random Forest

	Model	Accuracy
1	RandomForestClassifier	0.947904
3	DecisionTreeClassifier	0.888340
5	XGBClassifier	0.864707
0	LogisticRegression	0.862961
2	KNeighborsClassifier	0.841966
4	LinearSVC	0.781000

	Model	AUC
1	RandomForestClassifier	0.811065
3	DecisionTreeClassifier	0.804608
2	KNeighborsClassifier	0.577932
4	LinearSVC	0.576775
5	XGBClassifier	0.506714
0	LogisticRegression	0.500000

DATA CLEANING & FEATURE ENGINEERING

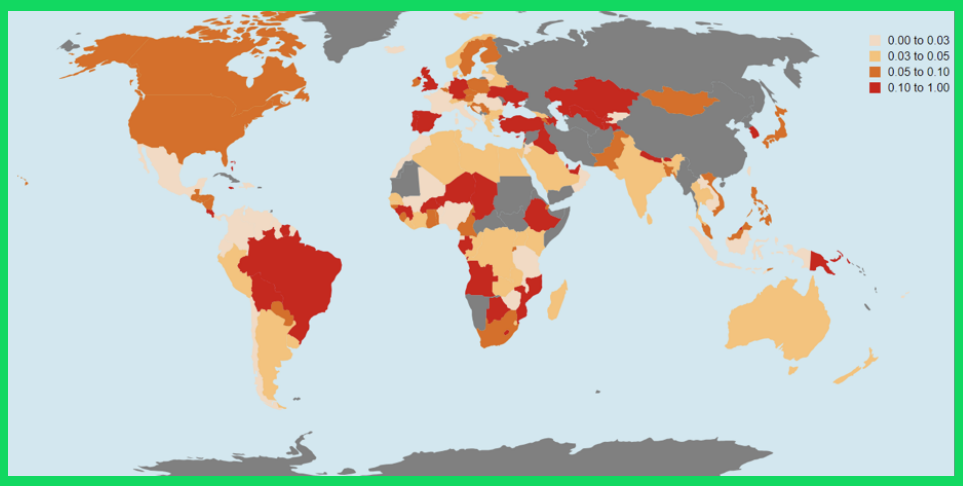
Dropped records with missing details, & chose key features: **acousticness**, **energy**, **danceability**, **tempo**, **duration**, **loudness**, **liveness**, **valence**, **speechiness**, & **instrumentalness**

FINAL MODELLING

Fitted and trained a Random Forest Regressor model for every region. Determined which features were most important in predicting song popularity.

Average Random Forest Statistics

MSE	146.18951
MAE	9.39969
R^2	0.63478



VISUALIZATIONS USING D3

- **Choropleth** (song popularity by country)
- **Heat map** (correlation between importance of acoustic features)
- **Cleveland Dot Plot** (compares acoustic features of actual & predicted song)

VISUALIZATION EVALUATION

- Evaluated with **user tests**
- Combination of choropleth with heatmap and dot plot **engaging for users, provide useful insights**
- Users able to **detect trends** over larger **geospatial regions** with choropleth

INNOVATIONS

- **Predicting song popularity by region**
- **Evaluation of XGBoost Model**
- **Correlation of acoustic features**
- **Heat map and dot plot visualizations**

WHO CARES and WHY?

- **Producers** can invest in hit songs
- **Artists** can refine songs based on popular metrics
- **Music platforms** can curate playlists based on regional preferences
- **Content creators** can select popular songs for wider reach