

# The Mood of Music

Youn Hee Pernling Frödin



# Business Case

A black and white photograph of a stage microphone on a stand, illuminated by a spotlight against a dark background with smoke. The microphone is positioned on the left side of the frame, angled towards the right. A bright spotlight beam shines down on the microphone from the upper right. In the background, there are other light sources, including a pair of small, bright lights at the top center and a larger, dimmer light source on the right. The overall atmosphere is dramatic and moody, typical of a live music performance.

In which of the top 10 music markets should a record label release a new artist?

Based on the mood of the songs an artist is performing and what music that is popular in what country.

# Data

The Kaggle logo, featuring the word "kaggle" in a light blue, lowercase, sans-serif font, set against a dark gray rectangular background.The Genius logo, featuring the word "GENIUS" in a black, uppercase, sans-serif font, set against a bright yellow rectangular background.

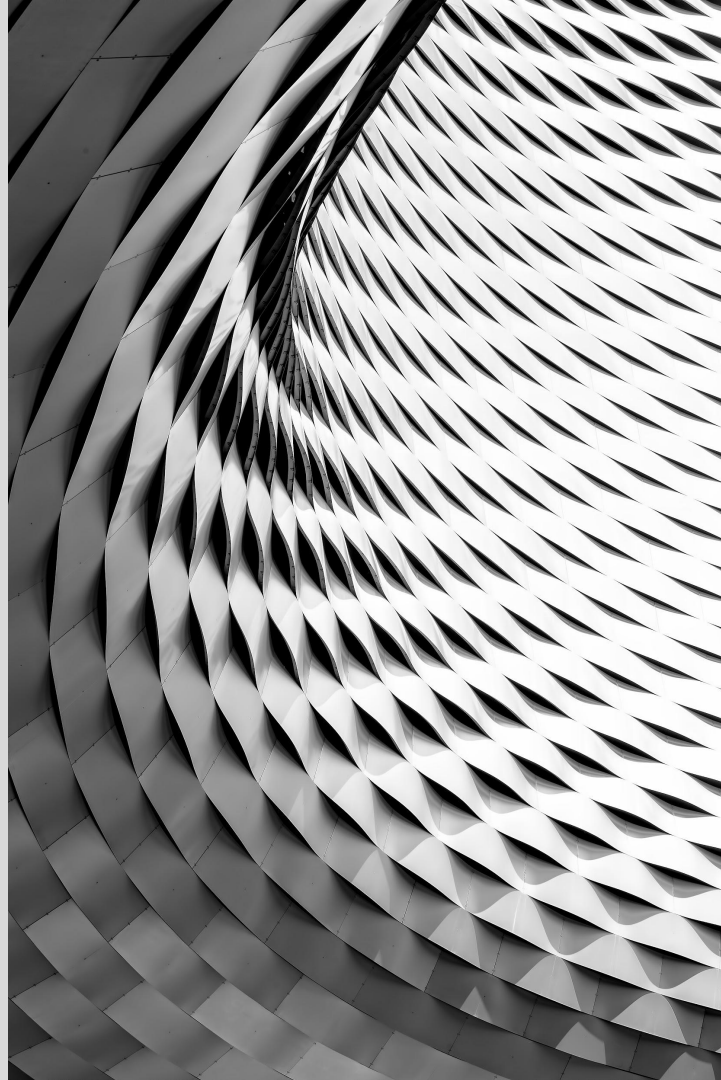
# Getting the Data

- 2 data sets from Kaggle (lyrics & daily top 200)
- Spotify API (additional features)
- Genius API + web scraping (lyrics)
- Manually added lyrics from Google search

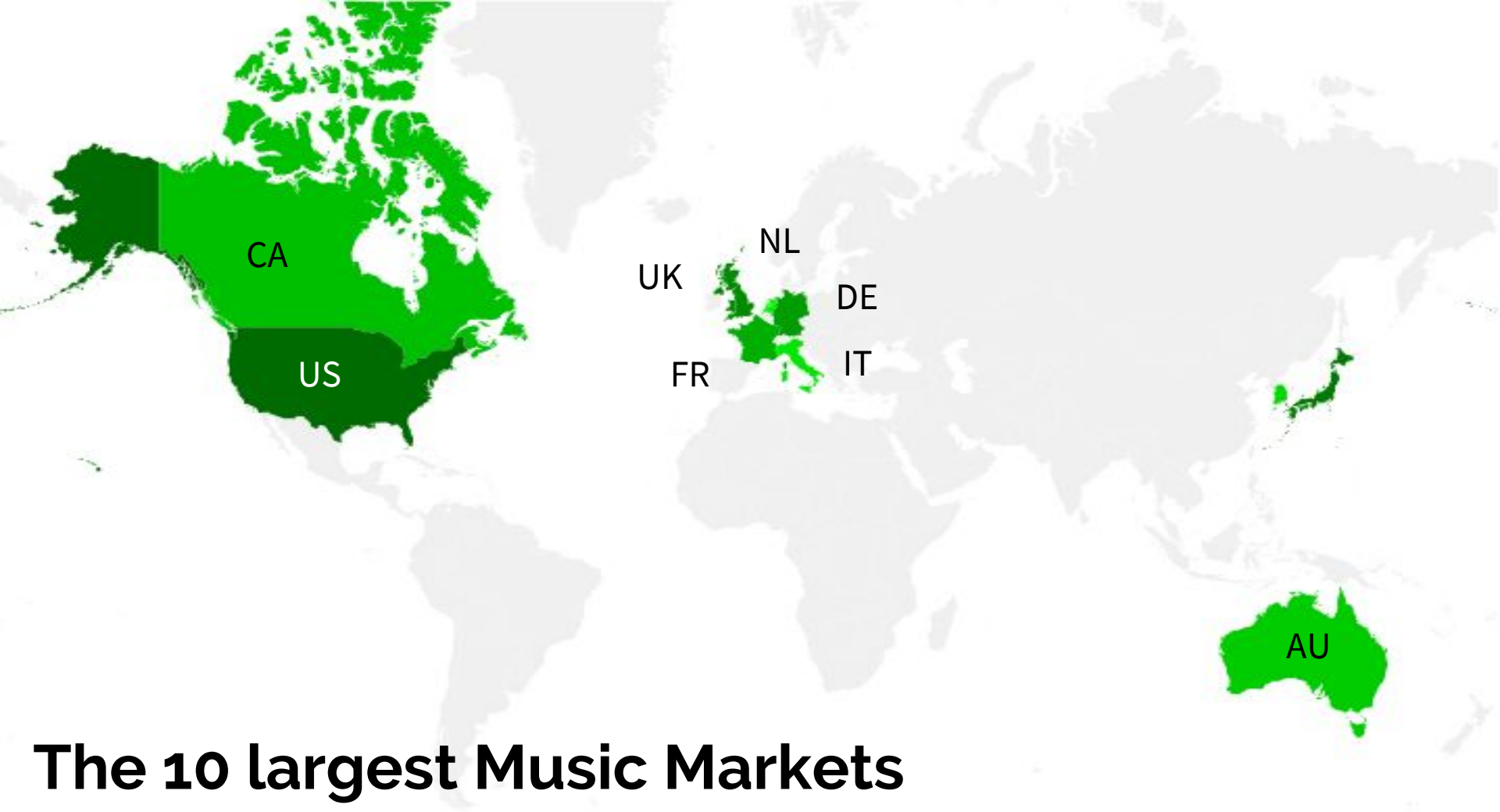
Lyrics in total and per market.

**68.46%**

82.43% France // 59.17% Netherlands

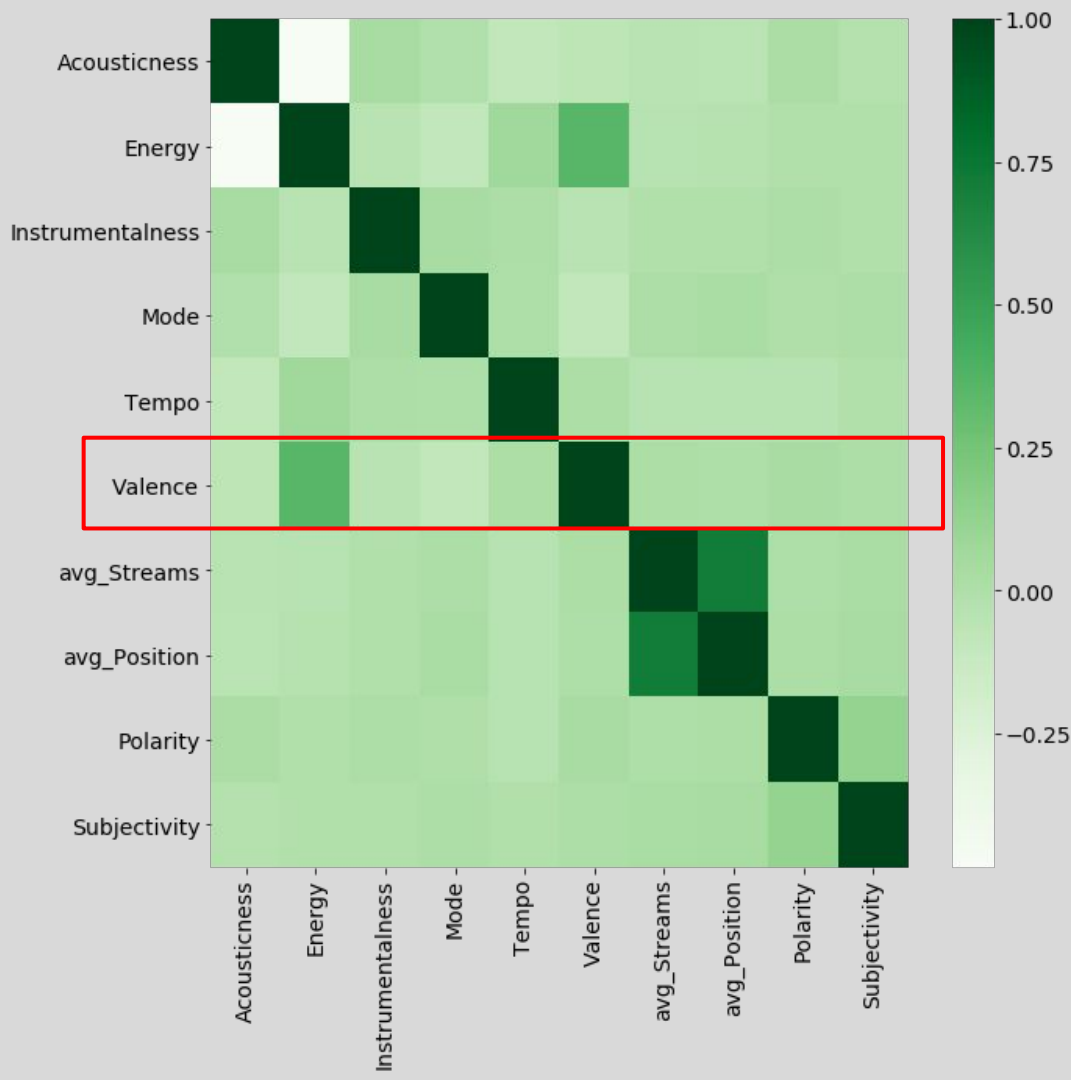






## The 10 largest Music Markets

# Correlation with the target (Valence)





# Process

- Collect data + EDA
- Train/Test-split
- MVP: total
- Good model: total + per market
- Best model: total + per market

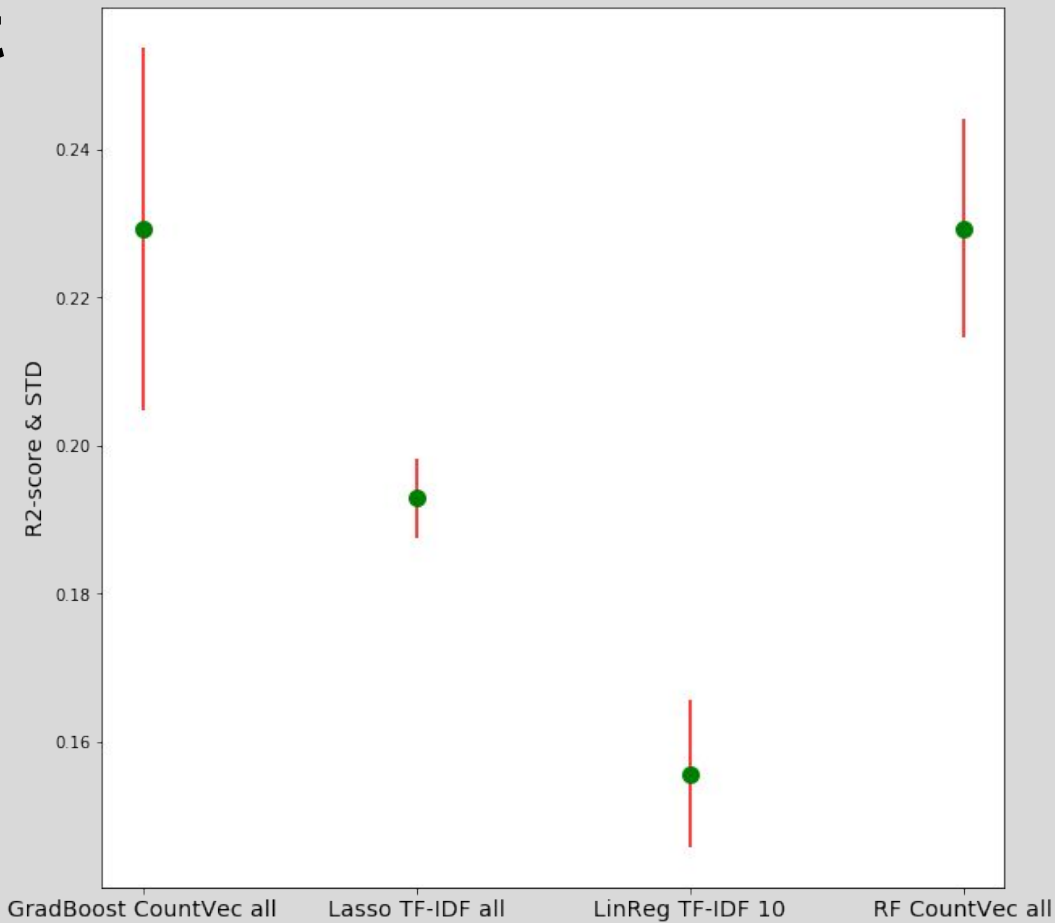
# Models

	NLP*	Scaling	Regressors**	Tuning HP***	Features	Comment
<b>MVP</b>	TB	x	LR, RF	GS	All, 10, 3	
<b>Good M</b>	TB, CV, TF-IDF	x	LR, Lasso, RF	GS, RS	All, 10, 3	
<b>Good M market</b>	TB, CV, TF-IDF	x	Lasso	GS	All, 10, 3	
<b>Best M</b>	TB, CV, TF-IDF	x	LR, Lasso, RF, Ada, Gradient	RS****	All, 10, 3	Dummies market. Avg position & streams.
<b>Best M market</b>	TB, CV, TF-IDF	-	RF	RS	All, 10, 3	Avg position & streams.

*\*Stop words: eng, spa, fre, ger, \*\*RF do not need scaling, \*\*\*LR have no HP to tune, \*\*\*\*Lasso also GS*



# Comparing the best models from each type of model



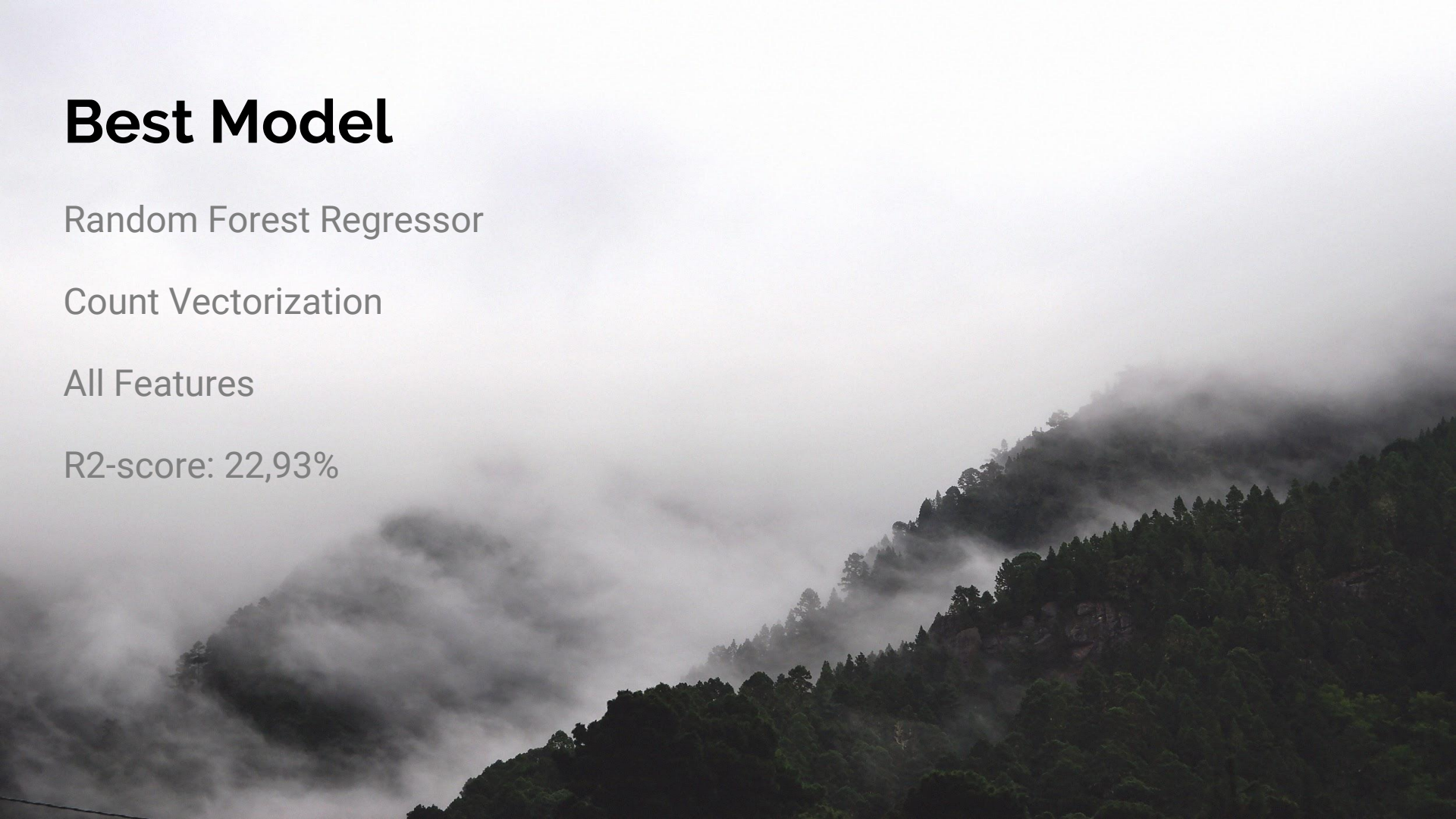
# Best Model

Random Forest Regressor

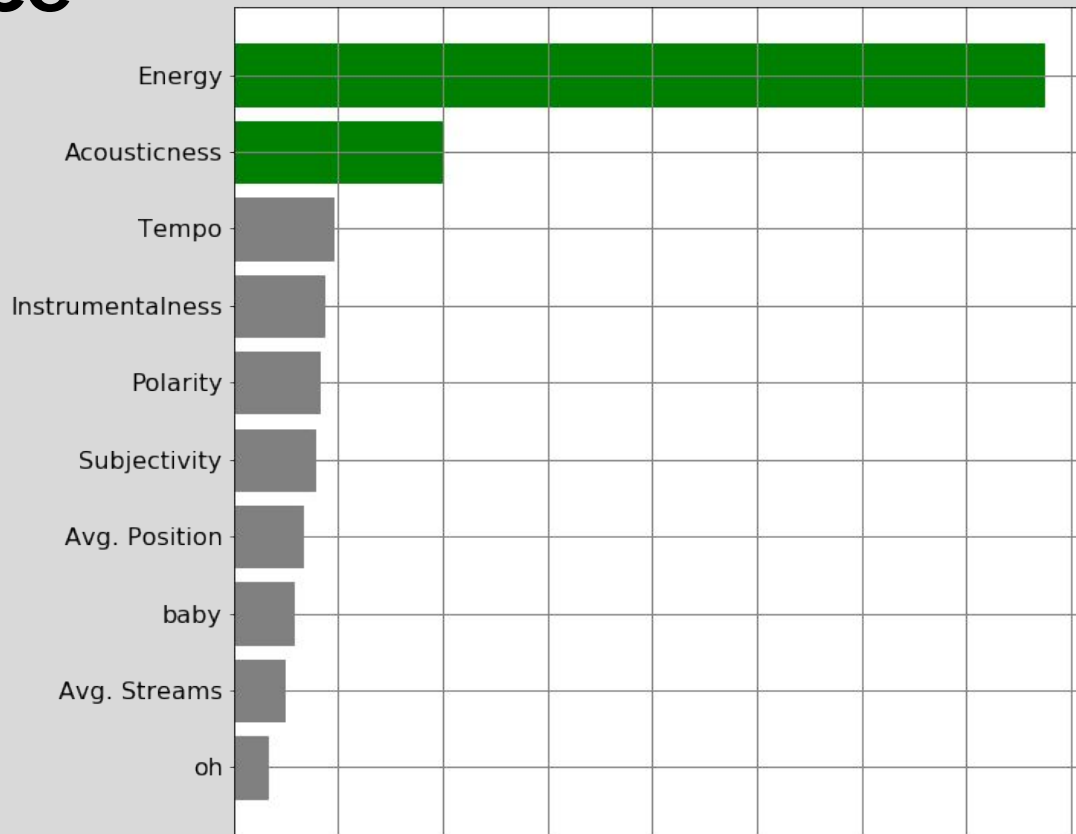
Count Vectorization

All Features

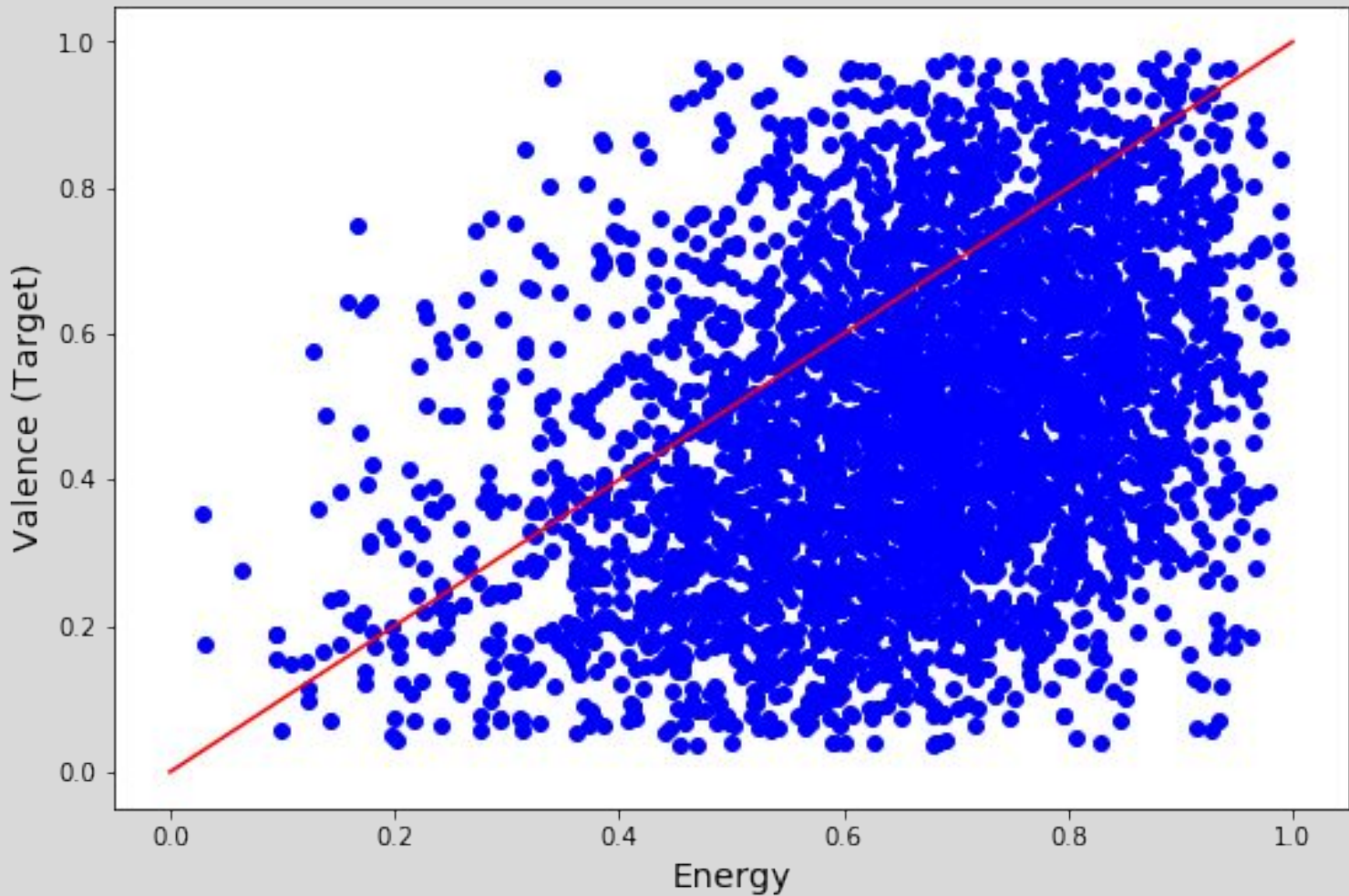
R2-score: 22,93%



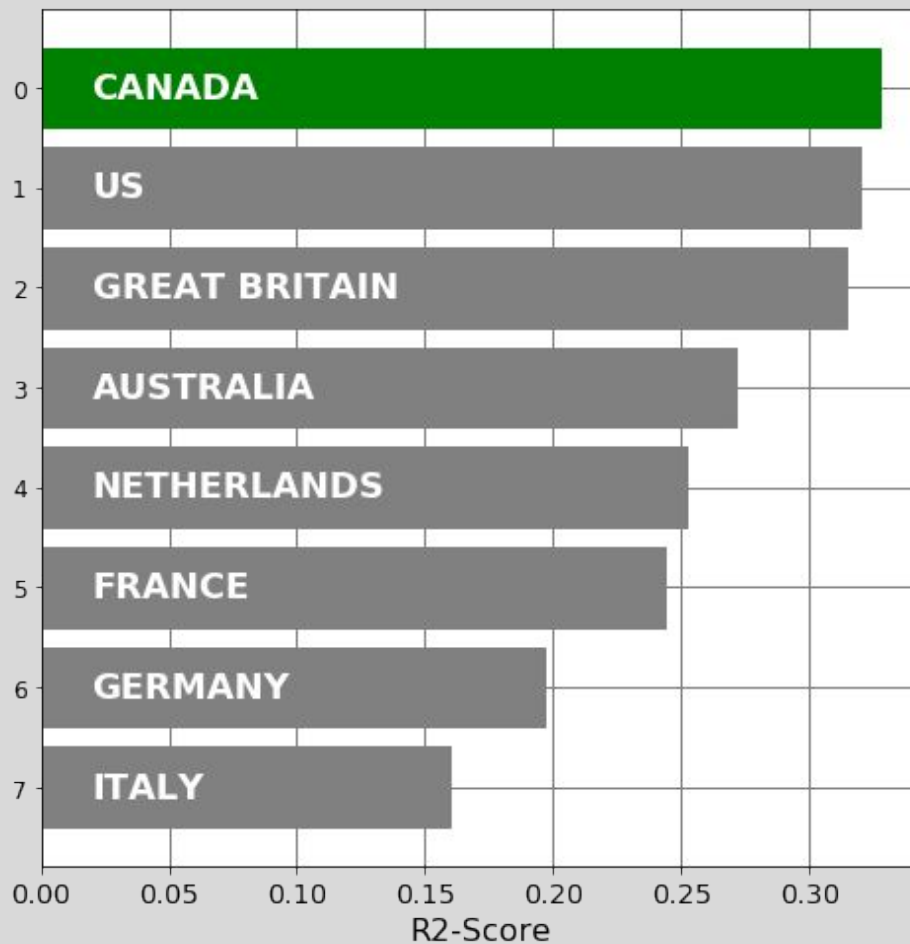
# Feature importance of the best model



# Correlation: Target and Most important feature

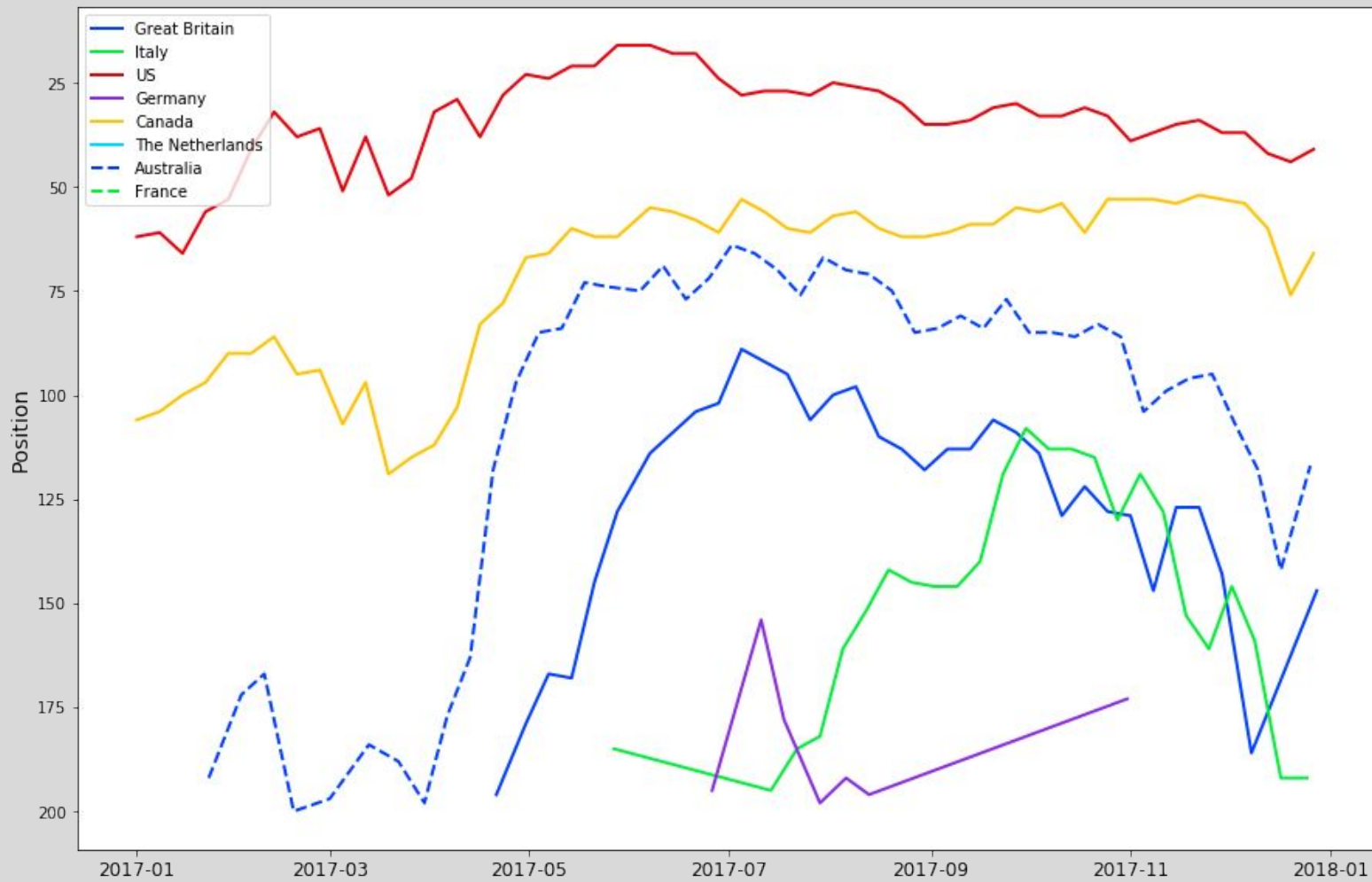


**In which market did  
the model perform  
the best?**





# Top song US 2017: Travis Scott - Goosebumps



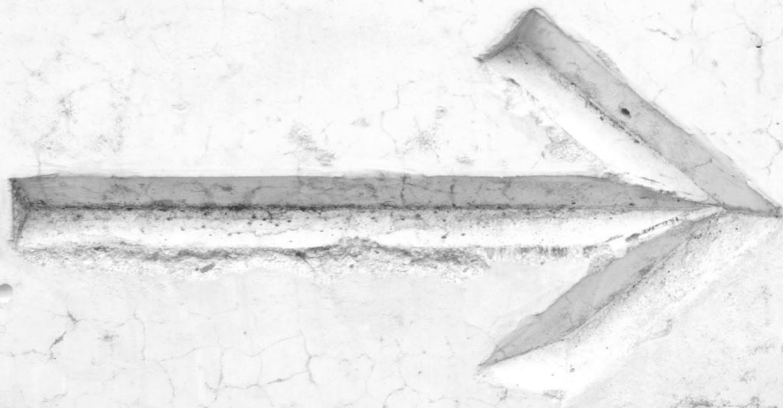


# Takeaways

It turned out that predicting the mood of a song is hard.

I learnt a lot about music popularity in the largest music markets and I want to focus more on the difference between the markets and how you best can make good use of that knowledge as a company connected to the music industry.

# Questions?



[yonipernling@gmail.com](mailto:yonipernling@gmail.com)

[github.com/yhpf](https://github.com/yhpf)

[medium.com/@yhpf](https://medium.com/@yhpf)

[www.yhpf.se](http://www.yhpf.se)

Marcus Frödin


Adam Blomberg

Penelope Leach

All fellow students of DSI+ NYC

Zigmund Kermish

Thank  
you!



# Prediction about the future of music that were completely wrong

*"Copyright will no longer exist"*

- David Bowie

*"Groups of guitars are on the way out"*

- Brian Epstein

*"No one will ever buy your stuff on CD"*

- Unnamed EMI executive

*"The wireless music box has no imaginable commercial value"*

- David Sarnoff