

EREZ BEN HORIN

DS18



Key Challenges in the Project

- (1) Machine Learning is a new subject that combines several fields of study
- (2) The size of the dataset
- (3) Going through all the stages until prediction
- (4) Creating connections between different features
- (5) Repairs backwards and forwards throughout the project



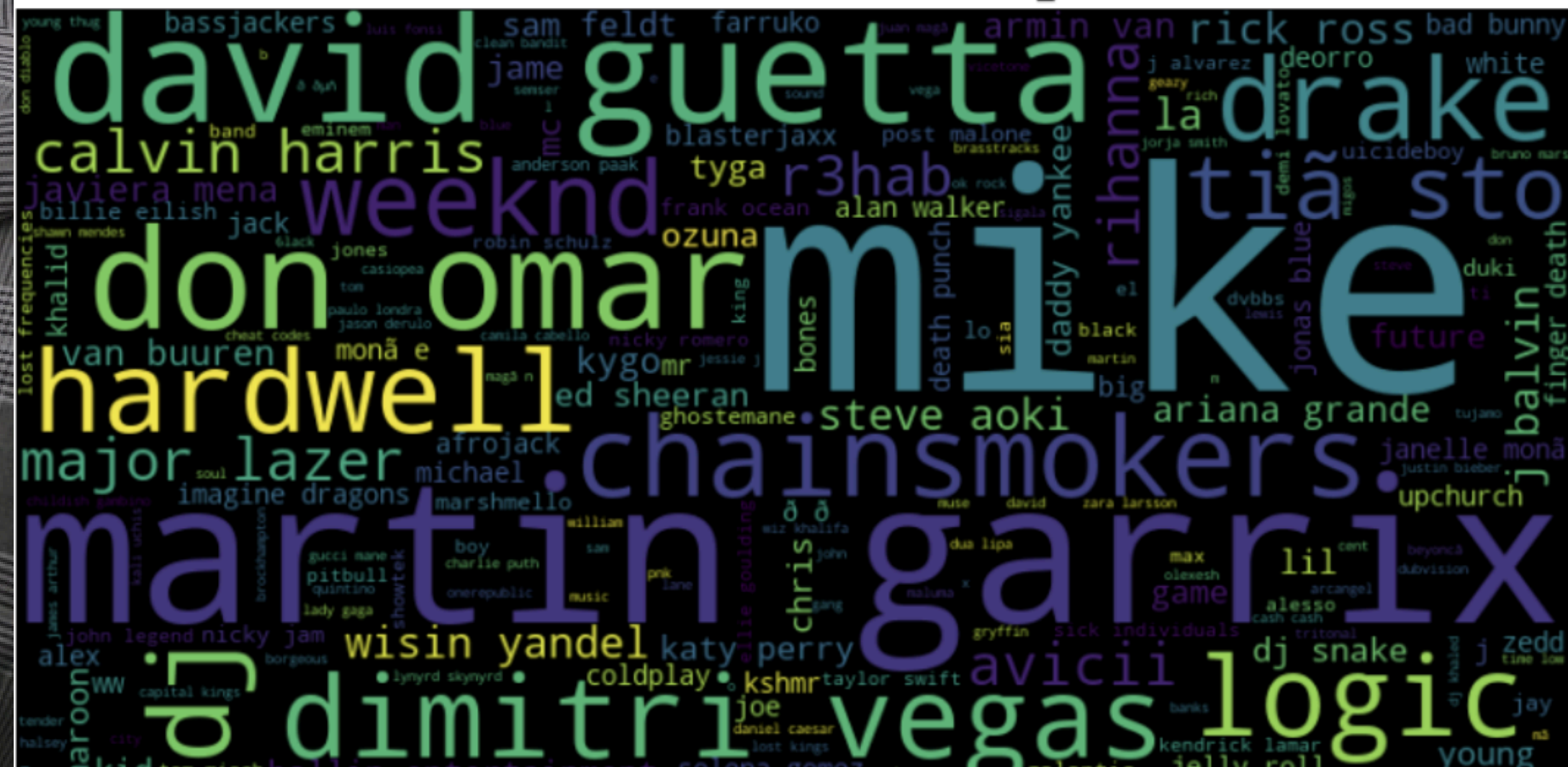
The goal of Spotify is to provide a personalized music streaming experience, offering users access to a vast library of songs, podcasts, and other audio content while fostering discovery and enjoyment.



UNIQUE FEATURES:

- (1) PLAYLIST_GENRE_ENCODED 03
- (2) ARTIST_TYPE_ENCODED
- (3) TOP 5 WORDS IN TRACK
- (4) TOP 5 WORDS IN ARTIST
- (5) ENGLISH_TRACK_NAME
- (6) TRACK_ALBUM_RELEASE_YEAR 2006
- (7) PLAYLIST_COUNT_PER_TRACK

Word Cloud from DataFrame: track_artist




key	loudness	mode	speechiness	acousticness	instrumentalness	...	tempo	playlist_count_per_track	track_album_release_year
5	-6.242	1	0.0347	0.065100	0.000000	...	150.863	1	2018
9	-4.739	1	0.0442	0.011700	0.009940	...	135.225	1	2017
10	-5.644	0	0.0540	0.000761	0.132000	...	128.041	1	2015
1	-6.300	1	0.0499	0.114000	0.000697	...	129.884	1	2018
7	-4.247	0	0.0873	0.187000	0.000000	...	94.033	1	2018

track_album_release_month	duration_minutes	english_track_name	Top 5 words in artist	Top 5 words in track	artist_type_encoded	playlist_genre_encoded
1	3.288100	0	0	0	2	3
11	6.225200	0	0	0	2	5
7	3.809417	0	0	0	0	2
11	3.938467	0	0	0	2	2
1	3.627550	0	0	0	2	1

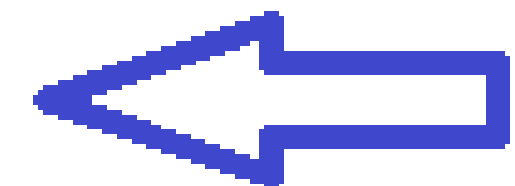
WHAT ELSE I COULD HAVE IMPROVED BUT DIDN'T HAVE TIME TO:

- (1) RUNNING CODE TO TRANSLATE ALL FOREIGN SONGS INTO ENGLISH – CAUSED THE COMPUTER TO CRASH
- (2) ADDING EXTERNAL DATA/TABLES TO SUPPORT THE MODEL (DATA FROM RECENT YEARS...)
- (3) IMPROVED OUTLIERS TREATMENT – WHICH COULD HAVE RESULTED IN IMPROVED RESULTS
- (4) ADDING VISUALIZATION



Results of the Prediction Model Popularity of a Song

	MSE Test	RMSE Test	MAE Test	R ² Test
Linear Regression	442.461175	21.034761	17.525174	0.192095
Random Forest	414.809007	20.366861	16.668437	0.242586
AdaBoost	451.781913	21.255162	18.059642	0.175076
Gradient Boosting	410.741973	20.266770	16.799900	0.250012
XGBoost	430.851704	20.756968	16.844327	0.213293



Personal experience from the project (I'm not in the field):

- (1) Importance of teamwork for brainstorming
 - (2) Learning and using project-supporting technological tools (AI, code completion, presentation preparation, data saving, and analysis)
 - (3) Time is a valuable resource – it's very important to invest in data research before running with the code
-

Spotify

Thank you very much for listening

