Capstone 3 Presentation

Ashley Owings

April
2021

Problem Statement

The focus of this project is to identify the key factors that will impact the number of streams a song will receive in a week and define what position the song will sit on the Spotify global charts and create an algorithm that will define the genre based on those factors.

In this project, we will use Spotify data. First, we will download the top 200 global charts for the last 53 weeks. With the track ID's we will pull the associated track features to check for features that may predict the stream count in a given week.

Machine Learning Type:

Supervised Regression

Recommendations and Key Findings

The analysis found that the stream counts were too random to identify any clear patterns. We look at the data using all positions over 53 weeks using linear regression and found no correlation with all positions. We then tried using the songs in the first position only. While the training data showed some promise with an r squared of .25, the test data using this algorithm showed an r squared of negative .31 indicating that the algorithm was overfitted to the data provided in the training set.

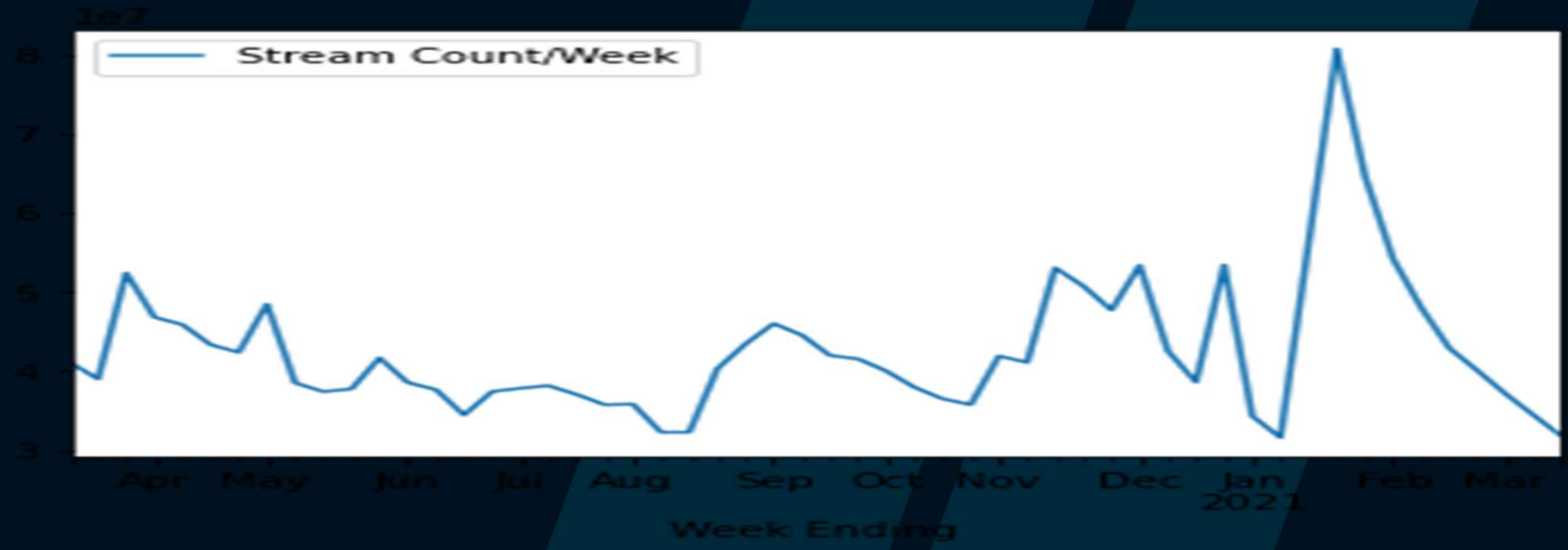
We then took the first position data and ran it through an ARIMA model for loop that tried all possible variations of the ARIMA models with no string fit to the data.

More research and new features will need to be identified to continue this analysis.

LINEAR REGRESSION MODEL

Measure	All songs	Position 1
Explained Error	0.04	-0.20
Mean Squared Log Error	0.23	0.40
R squared	0.04	-0.32
Mean Absolute Error	3,839,079.81	6,429,035.22
Mean Squared Error	31,413,852,182,776.15	84,886,230,174,460.88
Root Mean Squared Error	5,604,806.17	9,213,372.36

Position 1 Stream Counts



Sample Output for ARIMA (1,1,1)

Dep. Variable:	Stream Count/Week	No. Observations:	53
Model:	ARIMA(1, 1, 1)	Log Likelihood	-903.606
Date:	Wed, 07 Apr 2021	AIC	1813.211
Time:	08:45:51	BIC	1819.065
Sample:	03-15-2020	HQIC	1815.456
	- 03-21-2021		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.5472	0.137	3.985	0.000	0.278	0.816
ma.L1	-0.9416	0.105	-8.945	0.000	-1.148	-0.735
sigma2	8.007e+13	2.13e-15	3.75e+28	0.000	8.01e+13	8.01e+13

Ljung-Box (L1) (Q):	0.10	Jarque-Bera (JB):	438.16
Prob(Q):	0.76	Prob(JB):	0.00
Heteroskedasticity (H):	4.50	Skew:	2.91
Prob(H) (two-sided):	0.00	Kurtosis:	15.97


THANK
YOU!