
KPOP Music Group Analysis

Content

Analysis Aims

Data Gathering & Cleaning

Feature Engineering

Exploratory Data Analysis

Modeling Process

Evaluating Predictions

Conclusion and Recommendation

Analysis Aims

Analysis aims

- Implement a model to determine the lifespan of a music group
- Determine factors which affects their lifespan
- Provide insights to music entertainment groups and its current trend.

Data Gathering & Cleaning

Data Gathering & Cleaning

- GoogleTrend
 - Keyword “Kpop” trend from 2004 to 2021
- Data gathered and compiled through various sites
 - Kpop.fandom
 - Dbkpop
- Dataset is checked for errors
 - Year within range
 - Mission Value

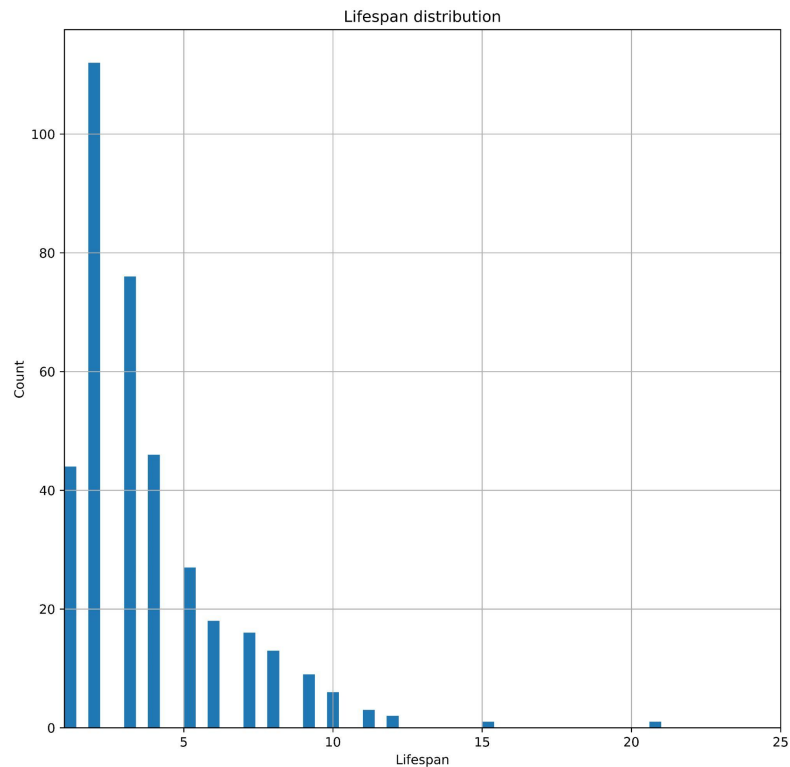
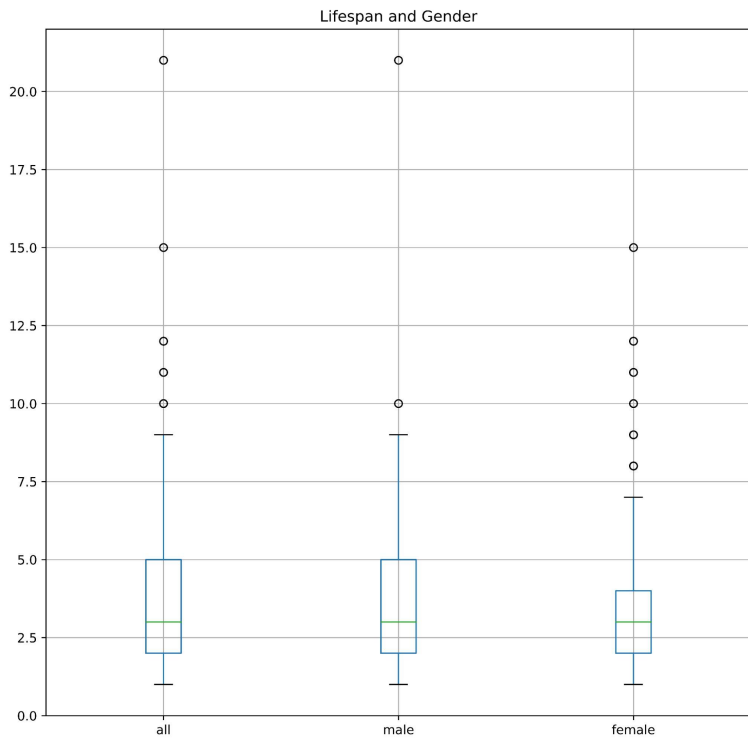
Feature Engineering

Feature Engineering

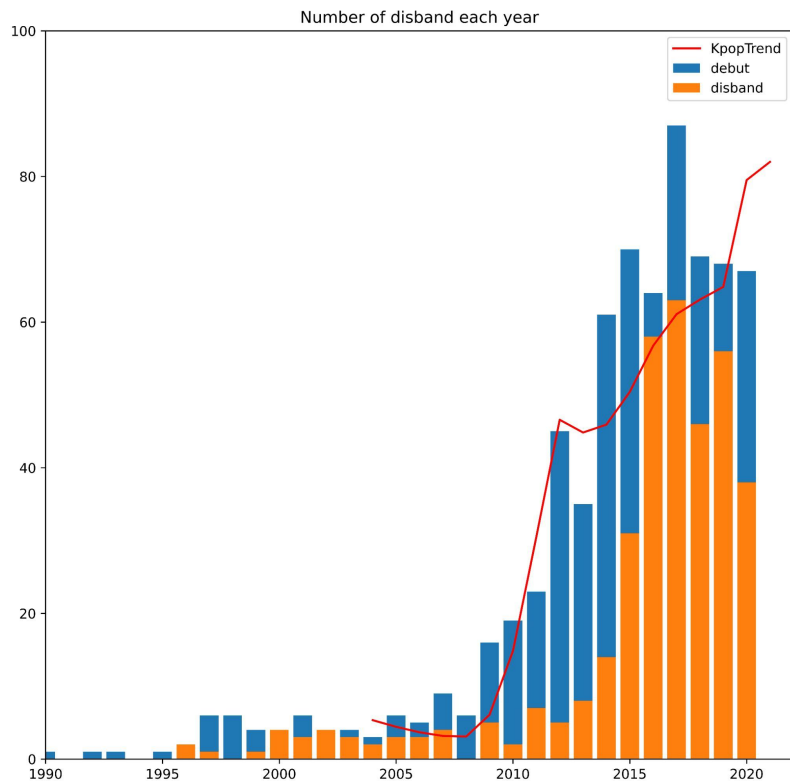
- Splitting dataset into train and testing dataset
 - Disband/Active datasets
- Calculating the lifespan of each music group
- New Features/Removing features
 - Total Releases
 - Disband Year
 - Last Production Year

Exploratory Data Analysis

Exploratory Data Analysis : Lifespan and Gender

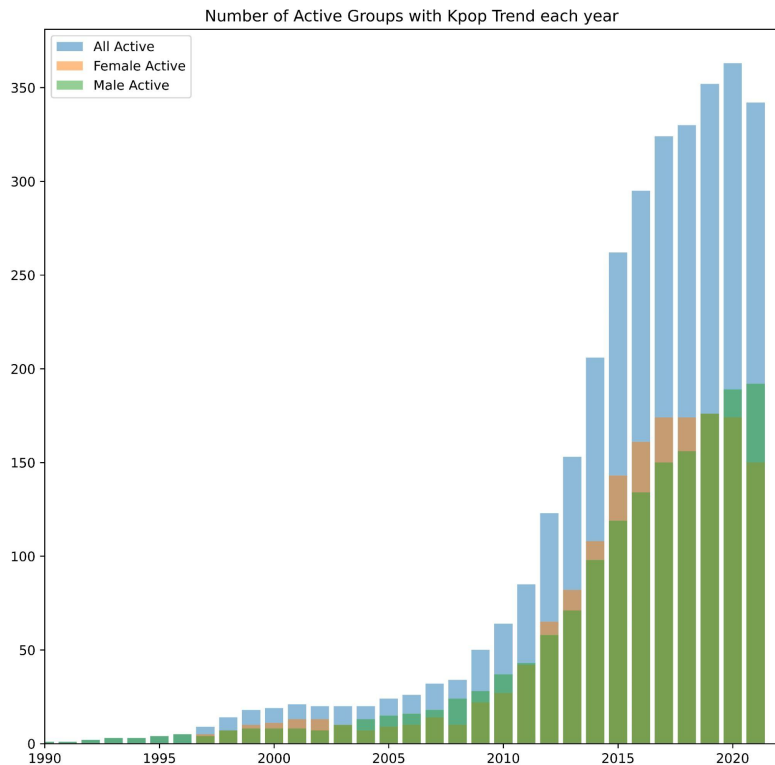


Exploratory Data Analysis : Debuts and Disbands



- Number of debuts increases with the google trend.
- High volume of debut and disbands
- New group debuts plateauing from 2014 onwards

Exploratory Data Analysis : Yearly Active



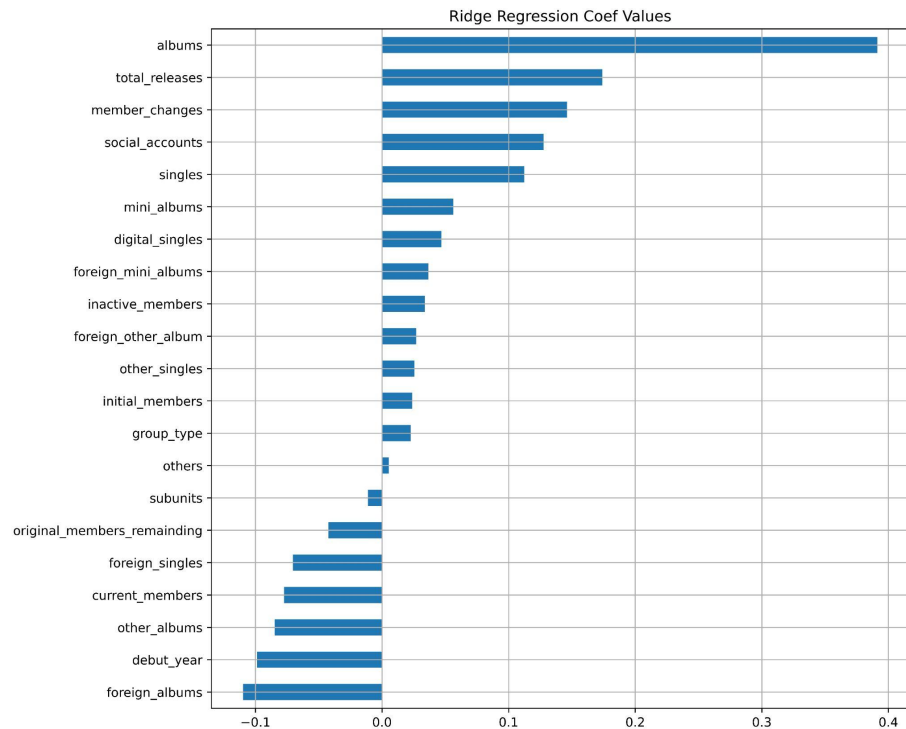
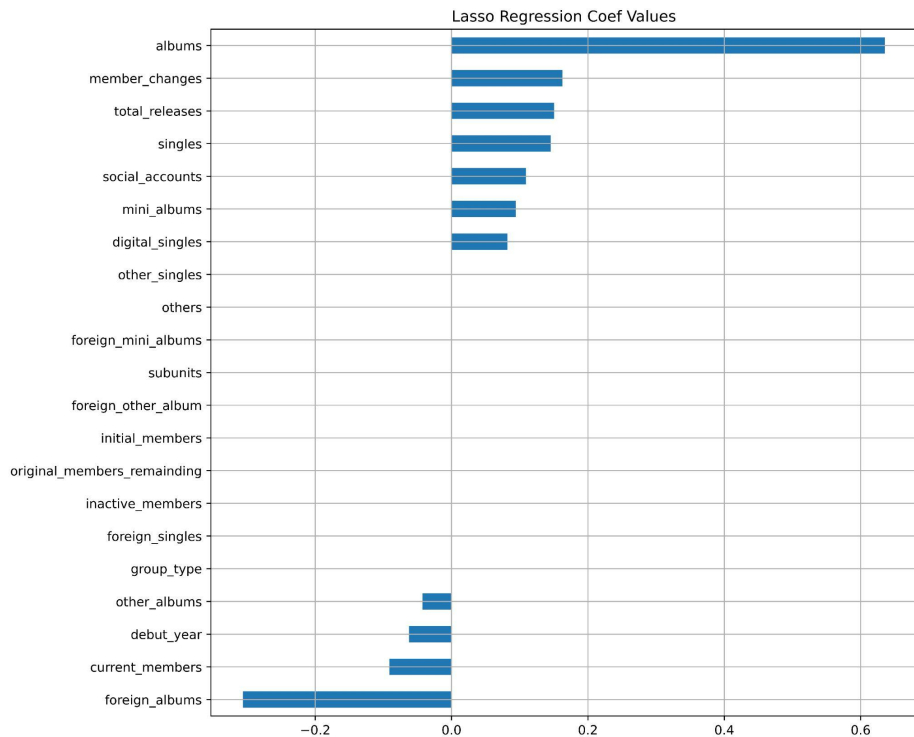
- Number of actives plateauing.
- Showing an increase in male group lasting longer.
- More female groups losing popularity?



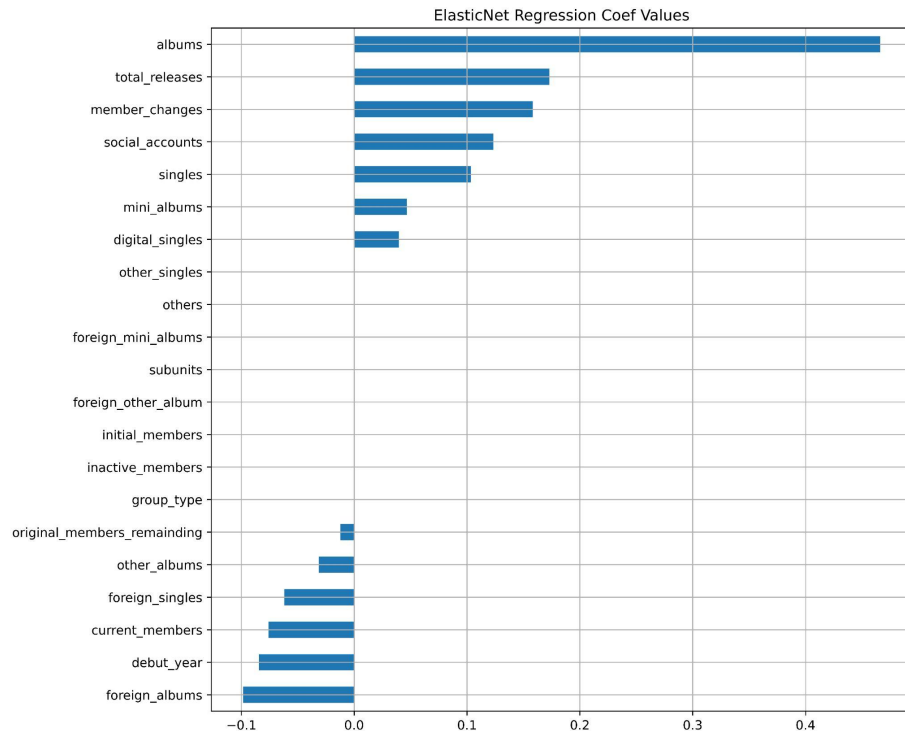
Modeling



Modeling : Lasso & Ridge



Modeling : Enet Regression



L1_ratio	Score
0.2	0.681988
0.1	0.681514
0.3	0.680618
0.5	0.672416
0.4	0.672278

Modeling Results

Model	Train R ² Score	Val R ² Score	Train RMSE	Val RMSE
OLS	0.792	0.348	1.168	2.046
Lasso	0.736	0.703	1.314	1.380
Ridge	0.726	0.702	1.342	1.385
Enet (best)	0.727	0.701	1.338	1.385

Evaluating Predictions

Evaluation Predictions : Longest Lifespan

	group_name	company	debut_year	disband_year	lifespan
303	TVXQ!	sm_entertainment	2003	2022.0	19.792361
53	Gavy NJ	good_fellas_entertainment	2005	2022.0	17.929408
284	SUPER JUNIOR	sm_entertainment	2005	2020.0	15.865277
278	SHINHWA	sm_entertainment	1998	2013.0	15.634963
252	NELL	space_bohemian	2001	2015.0	15.481491
149	4men	major9	1998	2012.0	15.321407
305	U-KISS	nh_media	2008	2022.0	14.966312
57	Girls' Generation	sm_entertainment	2007	2020.0	13.803270
208	FTISLAND	fnc_entertainment	2007	2020.0	13.770035
145	015b	the015b	2006	2018.0	13.457528

Evaluation Predictions : Disband Year

	group_name	company	debut_year	disband_year	lifespan
131	TWICE	jyp_entertainment	2015	2026.0	11.841408
242	MONSTA X	starship_entertainment	2015	2024.0	10.297438
304	TXT	bighit_music	2019	2024.0	5.670203
267	PENTAGON	cube_entertainment	2016	2024.0	9.136587
212	GOT7	jyp_entertainment	2014	2024.0	10.635437
282	Stray Kids	jyp_entertainment	2018	2024.0	7.069113
191	DAY6	colabs	2015	2024.0	9.625902
89	MAMAMOO	rbw	2014	2024.0	11.147397
176	BTS	bighit_music	2013	2024.0	11.973411
0	(G)I-DLE	cube_entertainment	2018	2023.0	6.280251

Conclusion and Recommendation

Conclusion & Recommendations

- Albums releases has the largest impact on the lifespan of a music group.
- Current trends indicate that number of active music group has peaked.
- Improvement of predicting lifespan of music group.
- But more data collection is required.