Final Technical Report - Ekans

K-pop Idols Analysis

Selin Cansu Akbaş
Student
Maltepe/Çankaya
191180005
selincansuakbas@gmail.com

Furkan Onur Akman
Student
Maltepe/Çankaya
191180009
akmanfurkanonur@gmail.com

Mert Akgüç
Student
Maltepe/Çankaya
191180006
mert.akgucc@gmail.com

ABSTRACT

In this paper, we describe what we've accomplished about our project, and we write what else we done. This study aimed to analyze about K-pop groups and artists. We analyzed the data and got some results. Our results will be shown in the rest of the report. Our results showed that BTS was the most popular group among international fans, followed by BLACKPINK and EXO. K-pop has become a truly global phenomenon thanks to its distinctive blend of addictive melodies, slick choreography, and production values, and an endless parade of attractive South Korean performers. They spend years in grueling studio systems learning to sing and dance in synchronized perfection. As a music streaming and media services provider, Spotify gives millions of streamers across the globe to listen to this extraordinary music genre. According to Spotify, the topmost Top-streamed K-Pop artists on Spotify include BTS, BLACKPINK, EXO, TWICE, and Red Velvet. K-Pop refers to Korean popular music aimed at the global market, but has been specifically limited to idol music. In the past in 1997, ballads and dance genres accounted for most, whereas K-Pop of various genres is gaining popularity with the public. Along with this diversity of genres, the recent trend of K-Pop shows the strength of idol groups that have undergone various collaboration attempts and systematic development.

1. INTRODUCTION

The specific topic that we plan to work on is K-pop. K-pop (abbreviation of Korean pop) is a genre of popular music originating in South Korea. K-pop idols are groups and artists formed by the various entertainment companies creating catchy Korean popular music and targeting younger audiences. The music groups are formed from a group of people who are all particularly talented in at least one of the following: singing, rapping, and dancing. These idols often enter the entertainment company in their teens and then train hard for years in areas like singing, rapping, dancing, and foreign languages. Then, if they are lucky and talented enough, by the end of their teens they will be picked for an idol group. We chosed K-pop because it became a global phenomenon. So, we thought it would be cool to analyze K-pop using machine learning to explore interesting insights. We do all this using data science.

2. METHODOLOGY

1- Collecting and analyze data on K-pop music and streaming numbers: We use data from music industry sources or online music platforms to investigate trends in K-pop music performance.

2- Analyzing K-pop lyrics and music videos: Music videos are great metrics to analyze because they are usually made for K-pop singers' hit songs, also known as 'debut' and 'comeback'.

3- Analyzing data on K-pop industry trends: As a result of the datasets we use and the results, we can reach general information about K-pop.

4- Using data on K-pop music video views and engagement to study the success of different releases: You could gather data on the number of views and likes that different K-pop music videos receive on YouTube and other platforms, and use this data to study which releases are most popular and successful.

Also we used some libraries and methods:

5- In several situations, such as the estimation of interdependent variables, we used the linear regression method.

6- The Plotly library was used to produce interactive graphics [1].

7- We worked with it to develop the features discovered by the Numpy library.

8- We used the Sklearn library to accomplish tasks like linear regression.

9- We created our statistics graphs using the Seaborn library [2].

3. DATASET

3.1 Kpop Idols

Database of Kpop-Idols [3] (dbkpop) is a dedicated database of most kpop idols. Using this dataset, we made very detailed and highly accurate analyzes. The website gives important information for our questions, such as how many members are in a group, what age are the members when they debut, how many members are in a group, and how many active groups there are currently. There are also other values that may be interesting such as what country idols were born and how many idols were in a former group.

	Stage Name	Full Name	Korean Name	K. Stage Name	Date of Birth	Group	Country	Birthplace	Other Group	Gender
0	A.M	Seong Hyunwoo	성현우	에이엠	1996-12-31	Limitless	South Korea	NaN	NaN	M
1	Ace	Jang Wooyoung	장우영	에이스	1992-08-28	VAV	South Korea	NaN	NaN	M
2	Aeji	Kwon Aeji	권애지	애지	1999-10-25	Hashtag	South Korea	Daegu	NaN	F
3	Ahin	Lee Ahin	이아인	아인	1999-09-27	MOMOLAND	South Korea	Wonju	NaN	F
4	Ahra	Go Ahra	고아라	아라	2001-02-21	Favorite	South Korea	Yeosu	NaN	F
...	...	...	...	...	...	...	...	...	...	...
1305	Ziu	Park Heejun	박희준	지우	1997-06-16	VAV	South Korea	NaN	NaN	M
1306	ZN	Bae Jinye	배진예	지엔	1994-06-09	LABOUM	South Korea	Bucheon	UNI.T	F
1307	Zoa	Cho Hyewon	조혜원	조아	2005-05-31	Weeekly	South Korea	NaN	NaN	F
1308	Zuho	Bae Juho	백주호	주호	1996-07-04	SF9	South Korea	NaN	NaN	M
1309	Zuny	Kim Joomi	김주미	주니	1994-12-08	Ladies' Code	South Korea	Gwangju	NaN	F

Figure 1. Idols Dataset

3.2 Girls Group Dataset

In this dataset [4] , we have accessed the information of girls groups separately. so we had the chance to make a detailed analysis between each girl and boy group.

	Name	Short	Korean Name	Debut	Company	Members	Orig. Memb.	Fanclub Name	Active
0	(G)I-DLE	NaN	(여자)아이들	2018-05-02	Cube	6	6	NaN	Yes
1	15&	NaN	피프틴앤드	2012-10-05	JYP	2	2	NaN	No
2	2EYES	NaN	투아이즈	2013-07-20	SidusHQ	4	5	NaN	No
3	2NE1	NaN	투애니원	2009-05-06	YG	4	4	Blackjack	No
4	3YE	NaN	써드아이	2019-05-21	GH	3	3	NaN	Yes
...	...	...	...	...	...	...	...	...	...
147	We Girls	NaN	위걸스	2018-08-31	Afternoon	8	8	NaN	Yes
148	Weki Meki	WeMe	위키미키	2017-08-08	Fantagio	8	8	NaN	Yes
149	WJSN	Cosmic Girls	우주소녀	2016-02-25	Starship	13	12	Ujung	Yes
150	Wonder Girls	WG	원더걸스	2007-02-13	JYP	4	5	Wonderful	No
151	Yellow Bee	NaN	옐로비	2017-11-23	Addiction	5	5	NaN	Yes

152 rows × 9 columns

Figure 2. Kpop Girls Groups Dataset

3.3 Boys Group Dataset

In this dataset [4], we have accessed the information of girls' groups separately. So, we had the chance to make a detailed analysis between each girl and boy group.

	Name	Short	Korean Name	Debut	Company	Members	Orig. Memb.	Fanclub Name	Active
0	100%	NaN	백퍼센트	2012-09-18	TOP Media	4	7	Perfection	Yes
1	14U	NaN	원포유	2017-04-17	BG	14	14	NaN	Yes
2	1the9	NaN	원더나인	2019-02-09	MBK	9	9	NaN	Yes
3	24K	NaN	투포케이	2012-09-06	Choeun	8	6	24U	Yes
4	2AM	NaN	투에이엠	2008-06-21	JYP, Big Hit	4	4	I Am	No
...	...	...	...	...	...	...	...	...	...
142	VROMANCE	NaN	브로맨스	2016-06-12	RBW	4	4	NaN	Yes
143	Wanna One	NaN	워너원	2017-08-07	YMC	11	11	Wannable	No
144	We In The Zone	NaN	위인더존	2019-05-27	Choon	5	5	NaN	Yes
145	WINNER	NaN	위너	2014-08-12	YG	4	5	Ace	Yes
146	ZE:A	NaN	제국의 아이들	2010-01-07	Star Empire	9	9	ZE:A Style	Hiatus

147 rows × 9 columns

Figure 3. Kpop Boys Groups Dataset

3.4 Youtube Videos Dataset

We have reached the YouTube channel information [5] of these dataset groups separately. There are many other aspects of a K-pop YouTube channel that you could analyze as well. The key is to identify a specific research question or problem that you want to address, and then use data and appropriate methods to try to answer that question or solve that problem. So, we had the chance to make a detailed analysis between each group. Doing this on YouTube worked very well because the number of views, subscriber numbers, video numbers could be displayed very clearly.

	Date	Artist	Song Name	Korean Name	Director	Video	Type	Release
0	2020-05-22	Agust D	Daechwita	대취타	NaN	https://youtu.be/qGjAWJ2zVWl	Boy Solo	Major
1	2020-05-21	Yubin	yaya (Me Time)	냠냠	NaN	https://youtu.be/BJD0arHF_5c	Girl Solo	Major
2	2020-05-21	OnlyOneOf	Angel (prod. GRAY)	NaN	NaN	https://youtu.be/vVTo8p72FxQ	Boy	Major
3	2020-05-20	Ryu Sujeong	Tiger Eyes	NaN	NaN	https://youtu.be/aE6curPGQRY	Girl Solo	Major
4	2020-05-20	Crush	Mayday feat. Joy (Red Velvet)	자나깨나	NaN	https://youtu.be/29ycT6tA-Rs	Boy Solo	Major
...								
3767	1994-08-08	Soo Taiji and Boys	Eternity	영원	NaN	https://youtu.be/74usU0gB7r0	Boy	Minor
3768	1993-06-06	Soo Taiji and Boys	Anyhow Song	하여가	NaN	https://youtu.be/L-AxO7EPU8c	Boy	Major
3769	1993-06-06	Soo Taiji and Boys	To You	너에게	NaN	https://youtu.be/cxlAiAUyvXo	Boy	Minor
3770	1992-03-23	Soo Taiji and Boys	I Know	난 알아요	NaN	https://youtu.be/OEDHEzs5kyk	Boy	Major
3771	1992-03-23	Soo Taiji and Boys	You, In The Fantasy	환상속의 그대	NaN	https://youtu.be/KGcwGoXfUDk	Boy	Minor

3772 rows × 8 columns

Figure 4. Kpop Youtube Channel Dataset

3.5 Spotify Hits Dataset

A very useful dataset [6] for analyzing streaming data of K-pop tracks and albums on Spotify to understand listening trends. We can analyze data on K-pop playlists on Spotify to understand how the genre is curated and consumed with this dataset. It covers the period between 1990 and 2020. In this respect, we can say that it is similar to the k-pop idols dataset. We use this dataset for spotify features.

4. EXPERIMENTS

Based on our experiments and analyses, the following results were obtained. We have some questions about K-pop idols. How is the comparison between male and female idols?
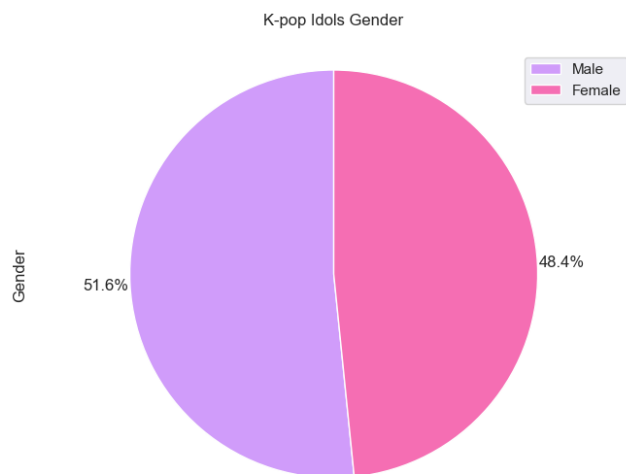

Figure 5. Gender Analysis – Pie Chart

There are 676 male idols and 634 female idols. Here's slight difference between the percentage. So, it can be concluded that the number of female idols is not much different from the number of male idols.
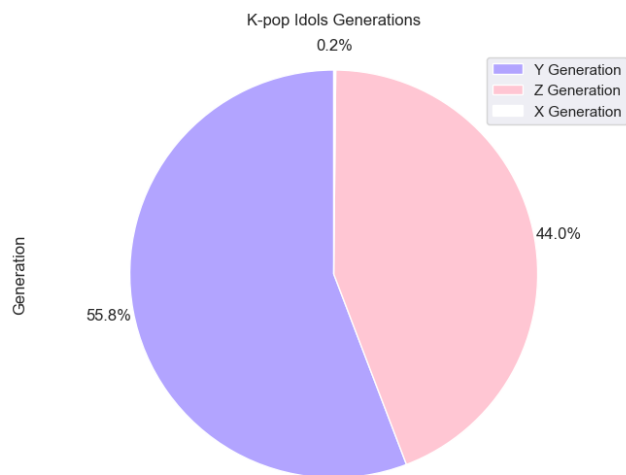

Figure 6. Generation Analysis – Pie Chart

From this pie chart, it can be seen that Millenials (generation Y) is dominating. Until 2020, the K-pop idols mostly born in 1981 – 1996. K-pop's fast development is epitomized by the timeline. Especially, K-pop groups can be divided into generations. There are a lot of opinions about how to divide the generations.
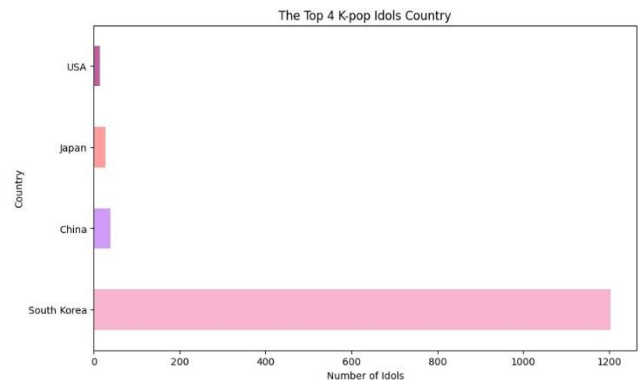
Figure 7. The Top 4 Kpop Idols Country
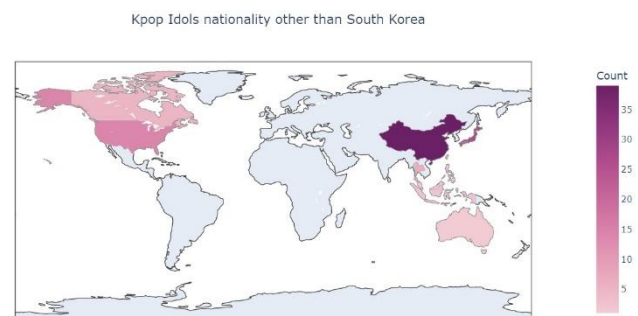
Figure 8. Kpop Idols Nationality Other Than South Korea

K-pop idols came from 12 different countries and most of them from China after South Korea. We obtained country information for all of the idols in our database. We plotted this data on a world map using the plotly library, and found that there are significant correlations between idol users and their respective countries.
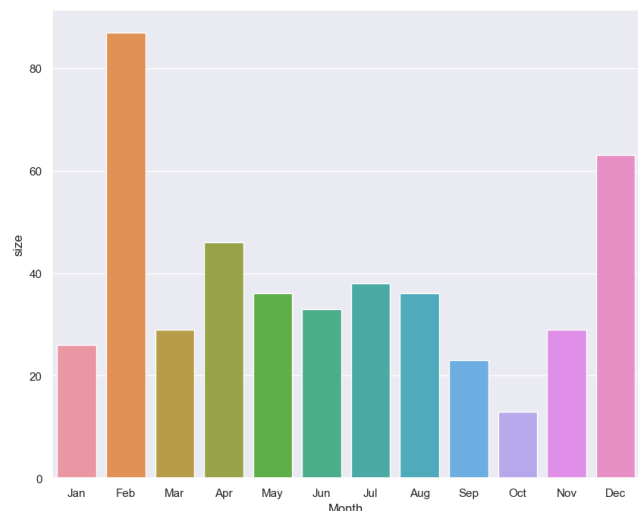
Figure 9. Video Size

As seen in the monthly statistic of video uploads, February is the month in which the most videos are uploaded.
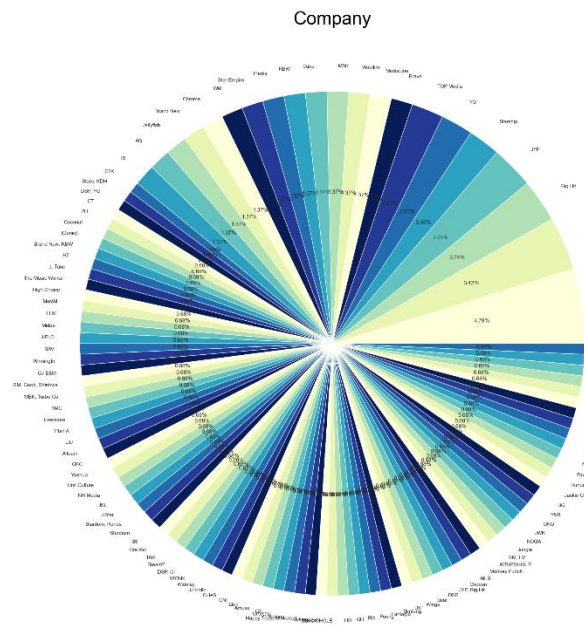
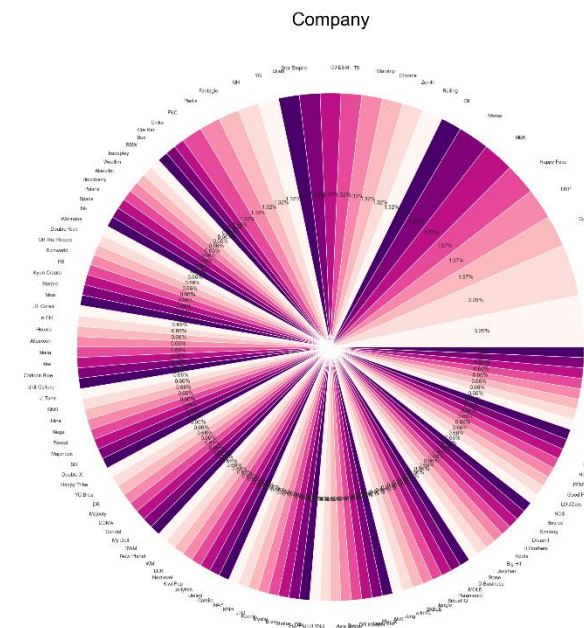Figure 10. Company Distribution Of Boys Groups

Figure 11. Company Distribution Of Girls Groups

K-pop entertainment companies play significant role in this industry. During the last two decades, the K-pop entertainment market was dominated by the "Big 3" K-pop entertainment agencies: SM Entertainment, YG Entertainment, and JYP Entertainment.

We make a Spotify data analysis. There are many ways to analyze data from Spotify. There are some possible questions to answer with data analysis. We look at the factors predict the success of a song danceability, energy, duration etc.
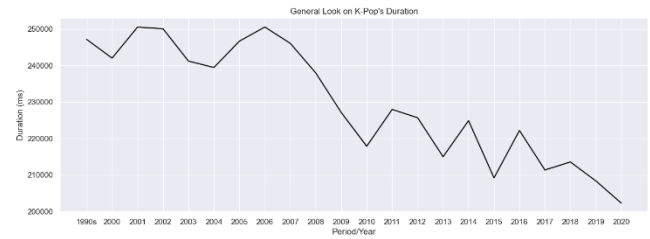
Figure 12. General Look on K-Pop's Duration

The duration of a song is the length of time it takes for the song to play from start to finish. The duration of a song is typically measured in minutes and seconds, and it can vary widely depending on the style and genre of music. Some songs may be only a few minutes long, while others may be much longer.
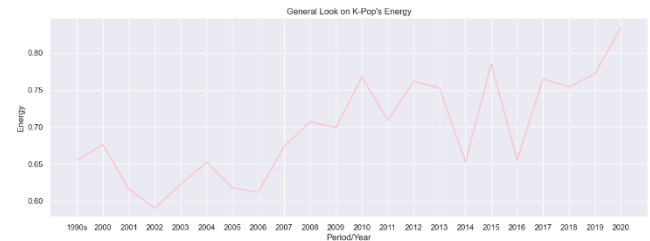

Figure 13. General Look on K-Pop's Energy

Spotify Energy is a feature in the Spotify app that is designed to match the energy level of music to your activity or mood. When you turn on Spotify Energy, the app will automatically select songs with a higher tempo or more upbeat energy to create a playlist that will keep you moving and motivated. You can use Spotify Energy when you're working out, running, or doing any other activity that requires an extra burst of energy. You can also use it to liven up a party or just to lift your mood.
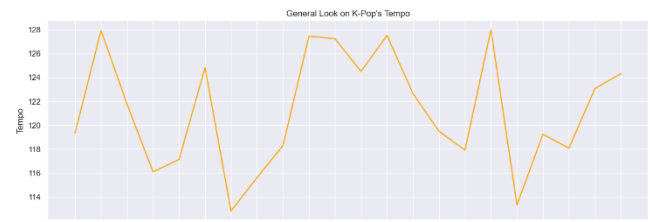

Figure 14. General Look on K-Pop's Tempo

The tempo of a song is a measure of how fast or slow it is. Tempo is typically measured in beats per minute (BPM), and it can have a big impact on the overall feel and energy of a song. A song with a fast tempo is likely to have a more energetic and upbeat feel, while a song with a slow tempo is likely to have a more laid-back and relaxed feel.

Figure 15. Debut Date v Members

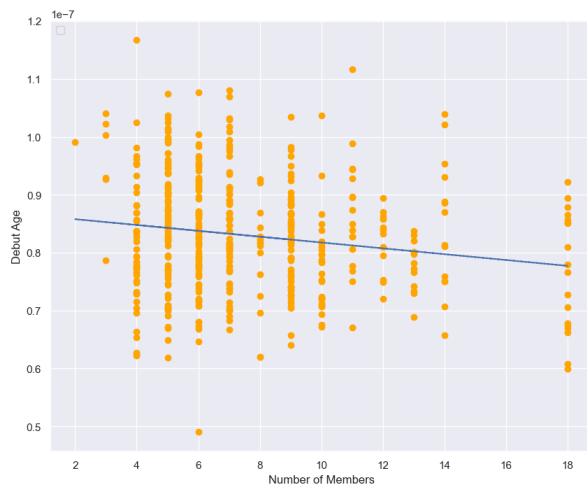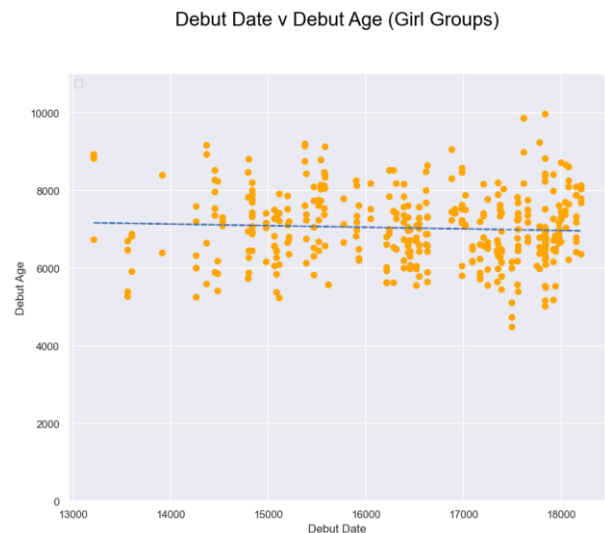
Figure 16. No of Members v Debut Age

Figure 17. Debut Date v Debut Age (boy)
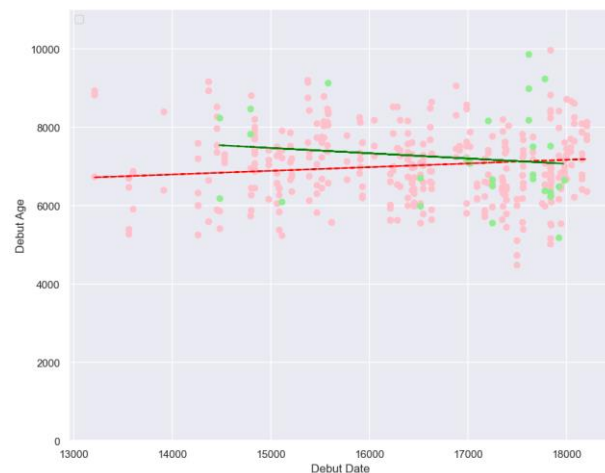

Figure 18. Debut Date v Debut Age (girls)

Figure 19. Debut Date v Debut Age (girl)

We have a linear regression. Linear regression is a statistical method that is used to model the linear relationship between a dependent variable and one or more independent varieables.

5. RESULTS

Our results show us that K-pop groups are still a young industry, but it is growing rapidly and growing internationally. Especially, music videos are great metrics to analyze because they are usually made for K-pop singers' hit songs, also known as 'debut' and 'comeback'. It was impressive to learn the following from Spotify data. Spotify data can provide insight into a wide range of topics related to music and music listening. We can say that In the last two decades, the K-pop industry has become a cultural sensation, continuing to grow in popularity as groups like BTS and Blackpink break barriers to achieve worldwide success. In fact, K-pop is an important phenomenon, because its influence and mass is a big phenomenon.

6. CONCLUSIONS

After these exploration, we can see some insight behind the K-pop idols data from 1992 to 2020. The number of male idols and female idols is slightly different. The idols not only came from South Korea, they came from 12 countries. The top 4 countries: South Korea, China, Japan, and USA. The goal of recruiting foreign idols is to capitalize on K-pop's international growth. There are 208 K-pop group in total. Some of idols are soloist that doesn't belong to any group. The K-pop group also have sub-unit group that may target a different market or audience by exploring different musical influences or promoting in non-Korean languages. About 44.4% of K-pop idols are use stage name than their original name for several reason. Until 2020, Millenials generation is dominating in K-pop industry. We can see that K-pop still a very young industry, but it is growing very fast and is growing internationally.

7. REFERENCES

[1] https://matplotlib.org/

[2] https://seaborn.pydata.org/

[3] https://dbkpop.com/

[4] https://www.kaggle.com/code/mpwolke/kpop-idols-boy-groups/data?select=kpop_idols_girl_groups.csv

[5] https://dbkpop.com/db/k-pop-music-videos/

[6] https://www.kaggle.com/code/sberj127/general-trend-in-kpop/data