1. **Introduction**

It is no secret that the K-pop Industry has boomed over the last 10 plus years. Everyone know what it is, or most probably has heard a song or two, or even know of a group or artist. Korea’s reach can be felt in other facets in Philippines as well. Over the past decade or so, K-Dramas have been on the front of Netflix homepages, Korean food, such as noodles, KBBQ and Korean grocery stores are all over the country. Even local commercials, billboards, and other advertisements have Korean celebrities, or are following obvious Korean culture influences. With the prevalence of their impact in our country, this analysis will look into data regarding the industry’s artists, groups, agencies, nationalities, and other pertinent data. The aim here is to explore and see the pattern, trends, and gain a better understanding about the idols over the past 4 decades of the industry. Just some of the questions that will be answered are the following:

1. Is there any predominance in sex (male vs. female) in the industry?
2. Aside from South Korea, which countries do idols originate from?
3. Which companies have the most groups that are currently active?
4. How well do groups retain their original members?
5. **Methodology**

The datasets used were from kaggle.com and dbkpop.com. The data was scraped using Python, specifically with BeatifulSoup and Pandas. Afterwards, it was cleaned then exported to csv files using the Pandas library. Subsequently, the analysis was done by using Microsoft Power BI to visualize the data, thereby showing connections and trends that the data presented. Some assumptions regarding the data had to be made, particularly that the data pertaining to the idols, such as birthdate, height, and names, were correct and accurate.

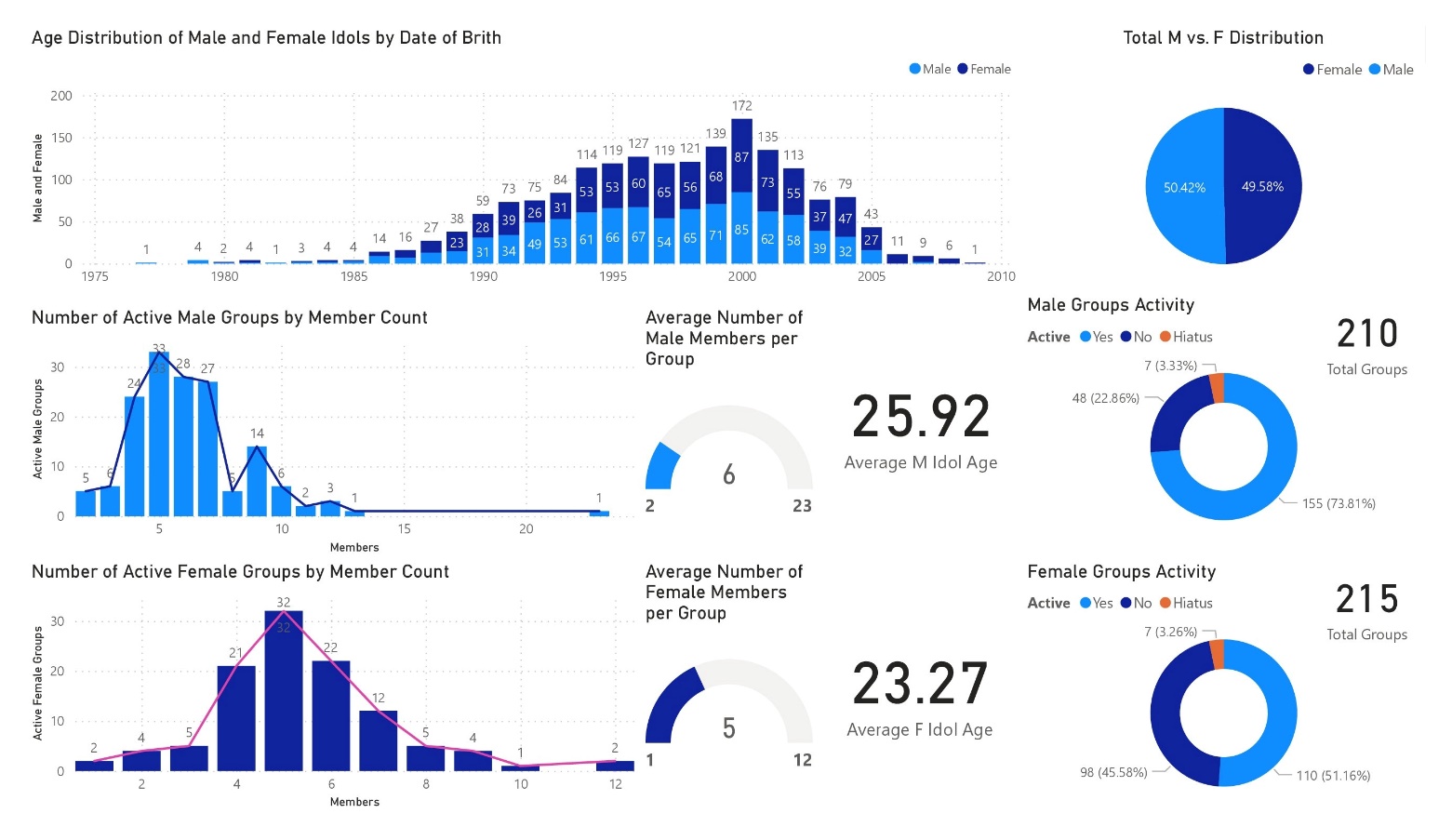
Some limitations of the analysis are regarding the data. The dataset contained blank cells, especially regarding countries and birthplaces (cities). This was addressed by not including the blank information in the visualizations. Another limitation about the data is that these might not contain all the idols and groups to ever exist in the industry. Lastly, the data is only up to early 2023, hence groups to come later in the year will not be included.

1. **About the Data**

The data used for this analysis consisted of three datasets obtained from different sources.

1. The first dataset, sourced from *Kaggle* (all\_kpop\_idols.csv), contained 1793 entries representing idols. The dataset included information such as stage name, full name, Korean name, date of birth, group affiliation, nationality, height, weight, birthplaces, other group associations, former groups, gender, and social media usernames.
2. The second dataset, obtained from *dbkpop* (kpop\_girlgroups.csv), focused exclusively on girl groups. It comprised 215 entries representing different girl groups and included information such as the group name, abbreviation, Korean name, debut date, company, current member count, original member count, fanclub name, and active status.
3. The third dataset, also from *dbkpop* (kpop\_boygroups.csv), focused specifically on boy groups. It consisted of 210 entries representing various boy groups and included information such as the group name, abbreviation, Korean name, debut date, company, current member count, original member count, fanclub name, and active status.

These datasets provided comprehensive information on K-pop idols, including their personal details, group affiliations, debut dates, company associations, and fanclub information. The combination of these datasets enabled a comprehensive analysis of the K-pop industry, its trends, and the characteristics of idols and groups.

1. **Analysis and Findings**
2. Male and Female Ages, Groups and Members Count

*GENDER*. From the gender distribution within the industry, it is evident that there is an almost equal split between males and females. From past to present, approximately 50.42% of idols are male, while 49.58% are female.

*AGES*. Regarding their ages, the most common birth year among K-pop idols is 2000, with a total of 172 individuals born in that year, comprising 85 males and 87 females. The second most common year is 1999, with 71 males and 68 females, totaling 139 idols. Following closely is the year 2001, with 62 males and 73 females, totaling 135 idols.

When factoring in the active status of groups in the industry, the average age of active idols, it is observed that male idols have an average age of 25.92 years, while female idols have an average age of 23.27 years. Consequently, male idols tend to be slightly over 2.5 years older than their female counterparts.

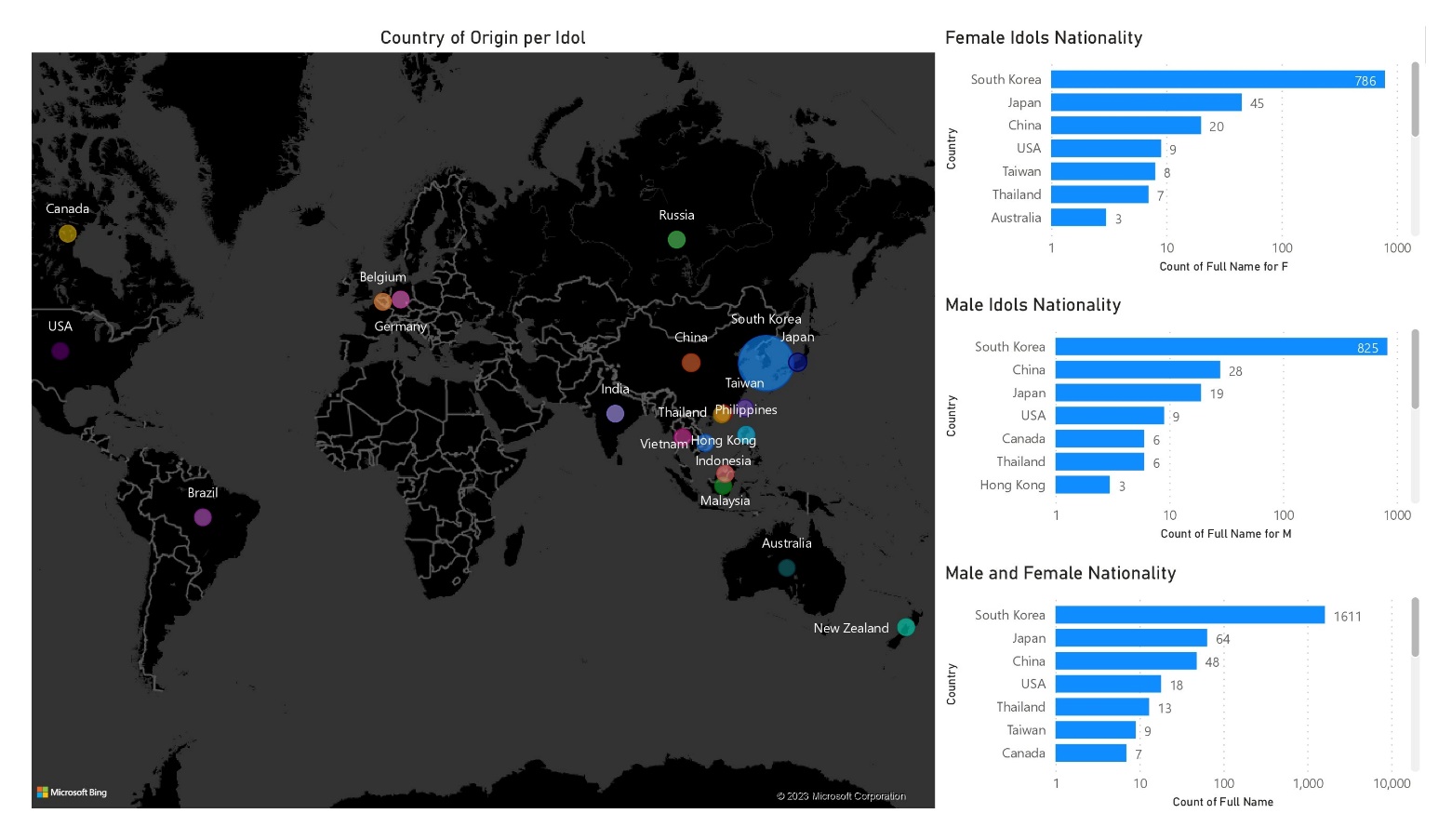
*GROUP COMPOSITION.* From data on the composition of male groups, it is found that the most common format is five-member groups, with 33 active groups falling into this category. Following closely are six-member groups, with 28 active groups, and four-member groups, with 21 active groups.

Similarly, in female groups, five-member groups are the most common format, with 32 active girl groups. The second most common is six-member groups, with 28 groups, followed by seven-member groups with 27 groups. Four-member groups have 24 active groups, while nine-member groups have 14 active groups, showing a significant gap between these two categories.

On average, male groups have approximately six members per group, while female groups have approximately five members per group.

*GROUPS.* The total count of male groups is 210, of which 155 (73.81%) are active, seven (3.33%) are on hiatus, and 48 (22.86%) are inactive. Likewise, the total count of female groups is 215, with 110 (51.16%) active, seven (3.26%) on hiatus, and 98 (45.58%) classified as inactive.

These findings shed light on various aspects of the K-pop industry, providing insights into birth trends, gender distribution, age differences, group composition, and the current status of male and female groups.

1. Nationality Distribution

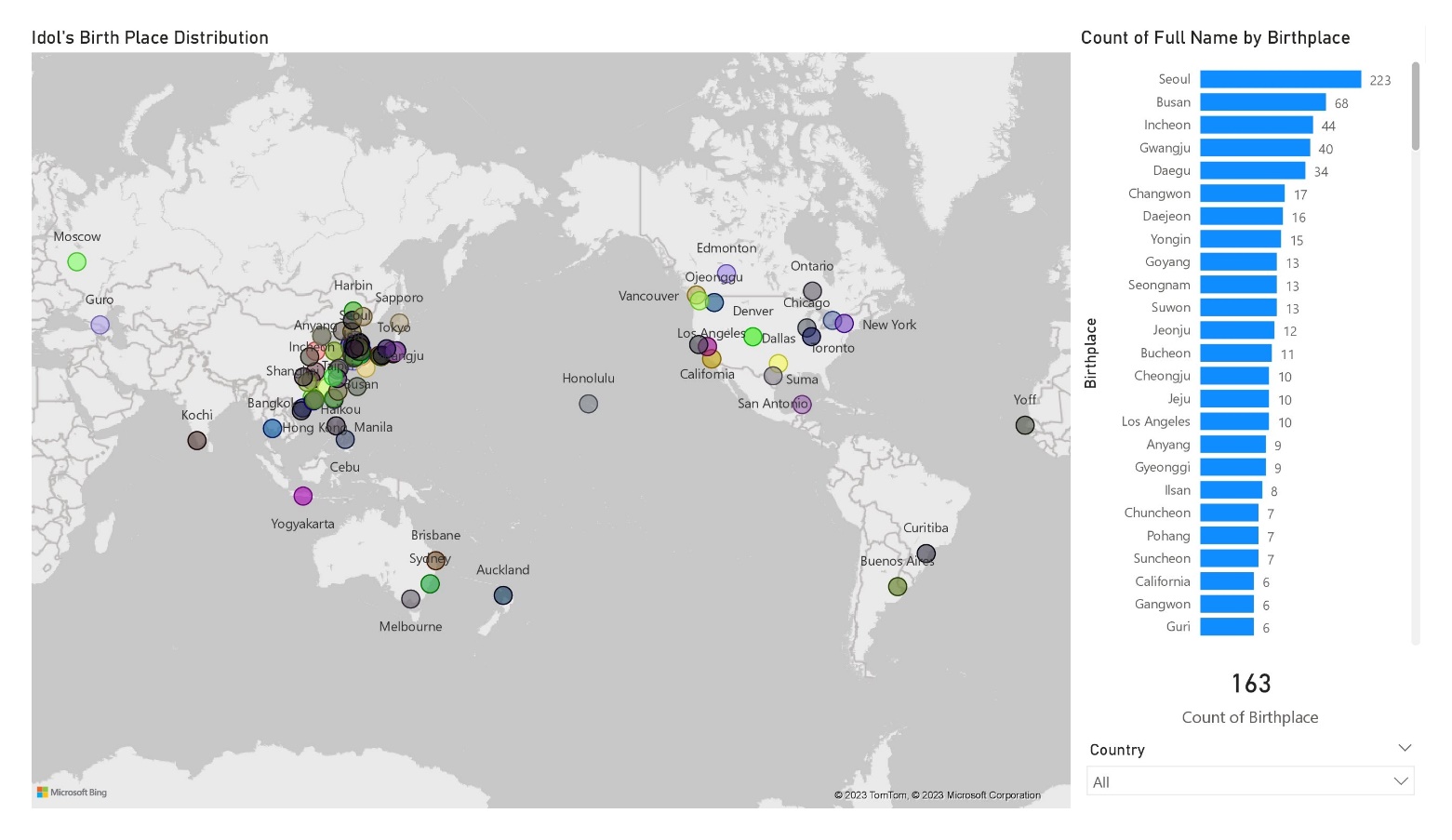
Analyzing the nationality distribution of male and female idols reveals interesting patterns within the K-pop industry. The majority of idols, totaling 1611, hail from South Korea. Among them, 786 are female and 825 are male.

Following South Korea, Japan has the second highest number of idols, with a total of 64. Of these, 45 are female and 19 are male. China holds the third position with 48 idols, including 20 females and 28 males.

Notably, although Japan has a higher total number of idols compared to China, the distribution between male and female idols does not align with the overall order. Japan has more female idols, while China has more male idols.

South Korea, Japan, and China emerge as the top three countries contributing to the K-pop industry in terms of idol count. Notably, the gap between these three countries and the next highest contributor, the United States, is substantial, with the US having a total of only 18 idols.

These findings shed light on the nationality distribution within the K-pop industry, highlighting the dominance of South Korean idols, the significant presence of Japanese idols, and the growing influence of Chinese idols.

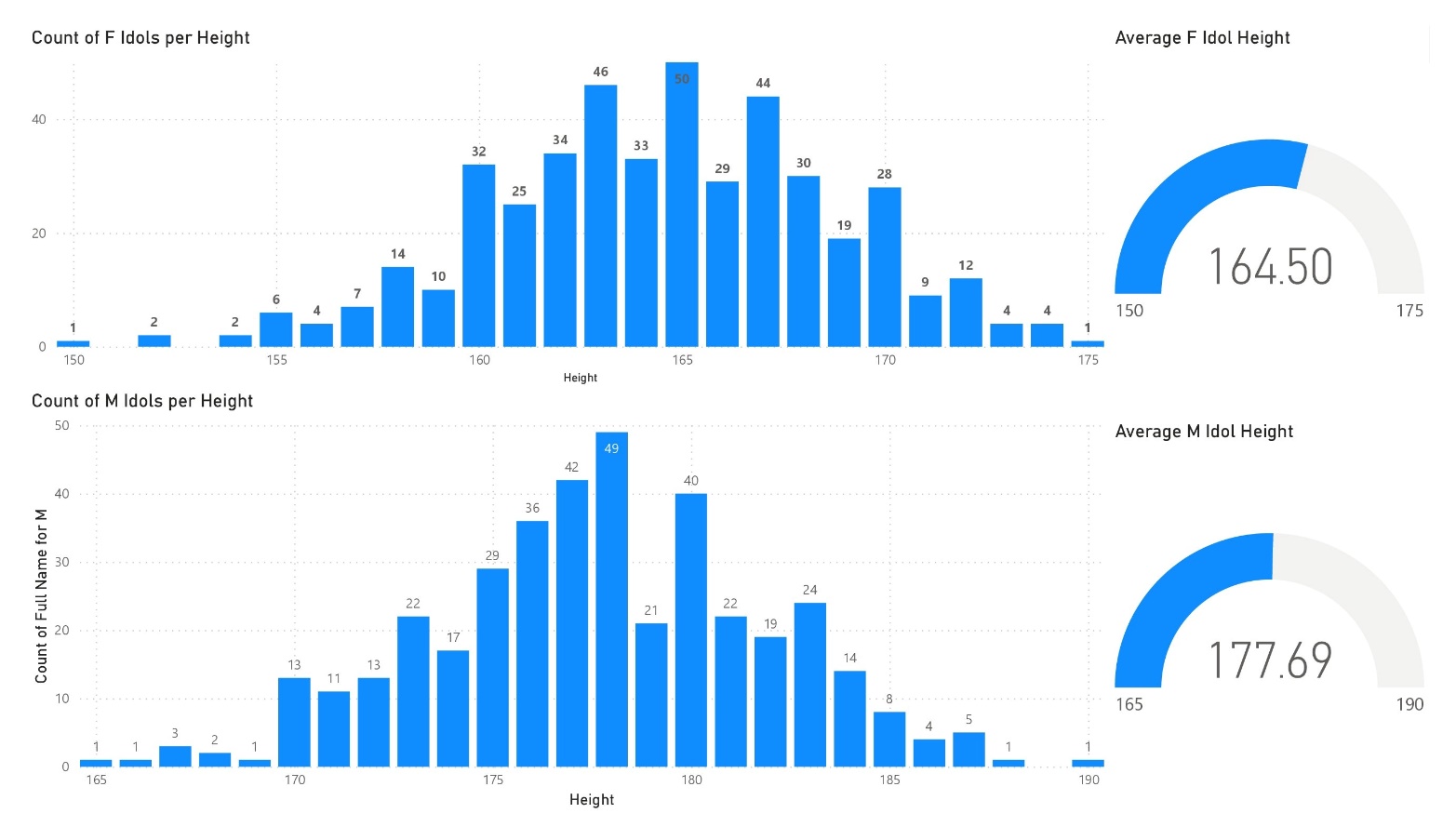
1. Birthplace Distribution

The birthplace distribution of K-pop idols provides insights into the geographical origins of these artists. The most common birthplaces for idols are Seoul, with 223 idols, followed by Busan with 68, Incheon with 44, Gwangju with 40, and Daegu with 34. There is a significant gap between Daegu and the next most common birthplace, Changwon, which is home to 17 idols.

When considering non-Korean birthplaces, Los Angeles stands out with 7 idols, followed by Osaka with 6, and Bangkok with 5. It is interesting to note that there are 104 different places in South Korea from where idols originate before reaching the number of idols from Los Angeles.

However, it is important to mention that not all entries in the dataset had birthplace information available, with a total of 951 blank entries. This indicates that the birthplace information for a significant number of idols is either unknown or not recorded in the dataset.

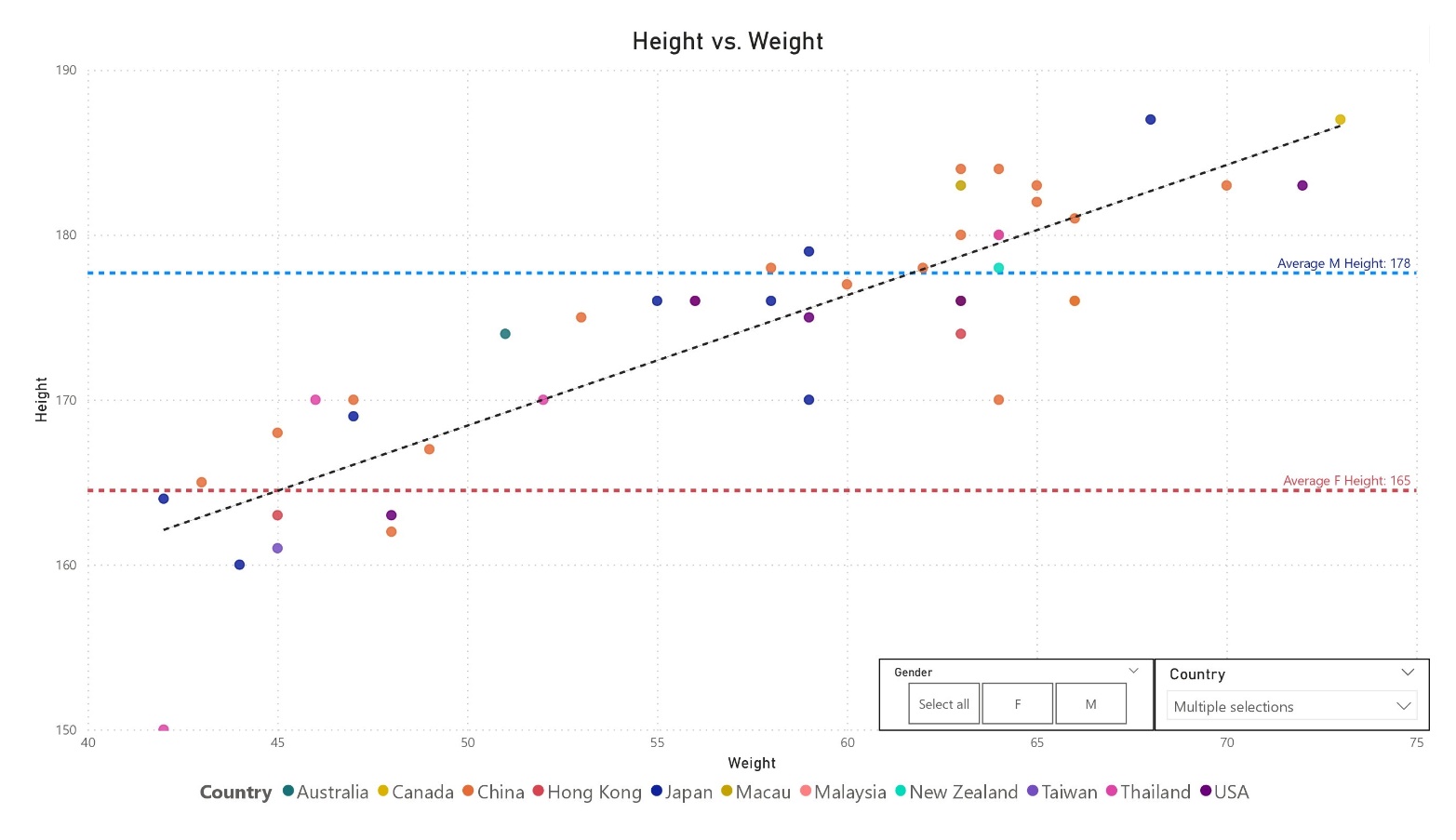
These findings provide an overview of the birthplace distribution among K-pop idols, highlighting the dominance of major Korean cities as the birthplaces of many idols, along with notable non-Korean birthplaces. The presence of a considerable number of missing or unknown birthplace entries emphasizes the limitations of the dataset in capturing comprehensive birthplace information for all idols.

1. Height Averages and Scatterplot

Examining the height distribution of K-pop idols reveals interesting patterns and variations between genders and nationalities.

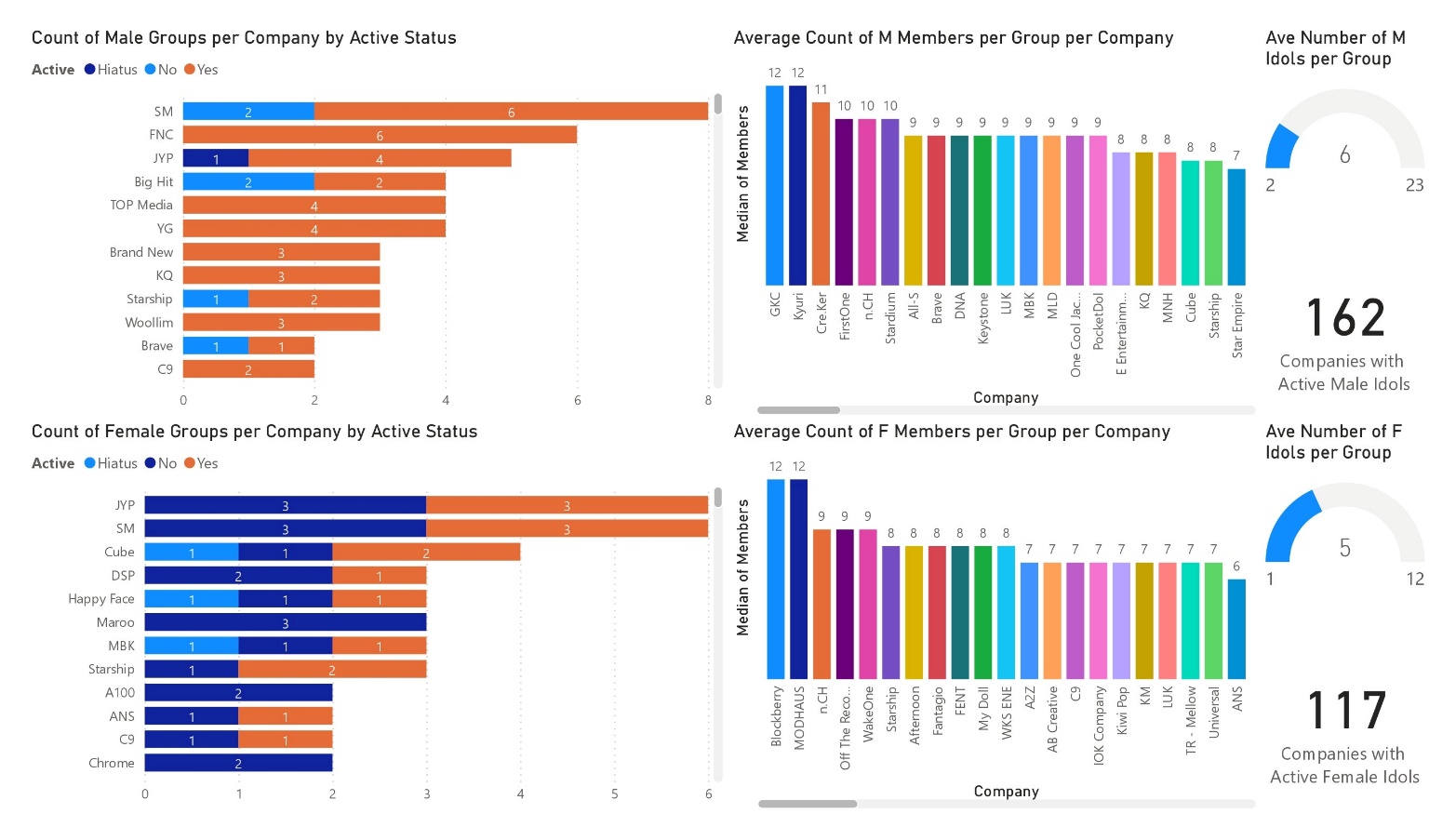
For female idols, the mean average height is 164.50cm. The three most common heights among female idols are 165cm (50 idols), 163cm (46 idols), and 167cm (44 idols). There is a noticeable difference of 10 idols before reaching the next most common height of 162cm, which consists of 34 idols.

On the other hand, the mean average height for male idols is 177.69cm. The most common heights for male idols are 178cm (49 idols), 177cm (42 idols), 180cm (40 idols), and 176cm (36 idols). There is a difference of 7 idols before reaching the next most common height of 175cm.



Analyzing the scatterplot of height to weight among idols per country, it becomes apparent that for non-Korean idols, China has the most male idols who are taller than the average male idol height of 177.69cm, with 9 male idols above the average line and 4 below. Similarly, China also has the most female idols who are taller than the average female idol height, with 4 idols surpassing the 164.5cm average for females and only 1 below the average female line.

These findings illustrate the height distribution among K-pop idols, showcasing the most common heights for both males and females. The scatterplot analysis highlights the variations in height among idols from different countries, particularly showcasing the higher average heights among male idols from China.

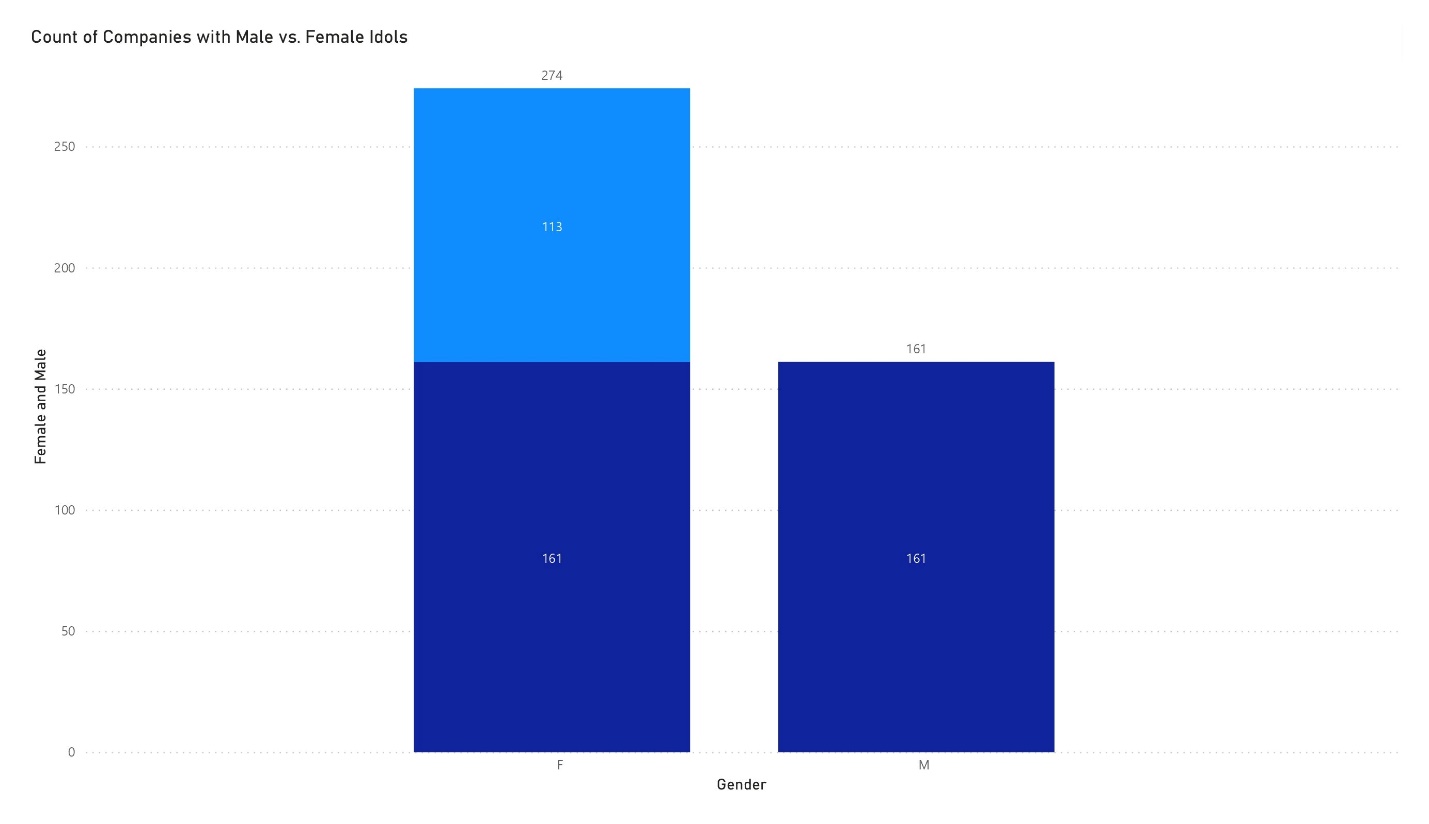
1. Active Groups and Average Member count per Company
2. Male groups per Company by Active Status

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Company | Total | Inactive | Hiatus | Active |
| SM | 8 | 2 |  | 6 |
| FNC | 6 |  |  | 6 |
| JYP | 5 |  | 1 | 4 |
| Big Hit (HYBE) | 4 | 2 |  | 2 |
| YG | 4 |  |  | 4 |
| Cube | 2 |  |  | 2 |

1. Female groups per Company by Active Status

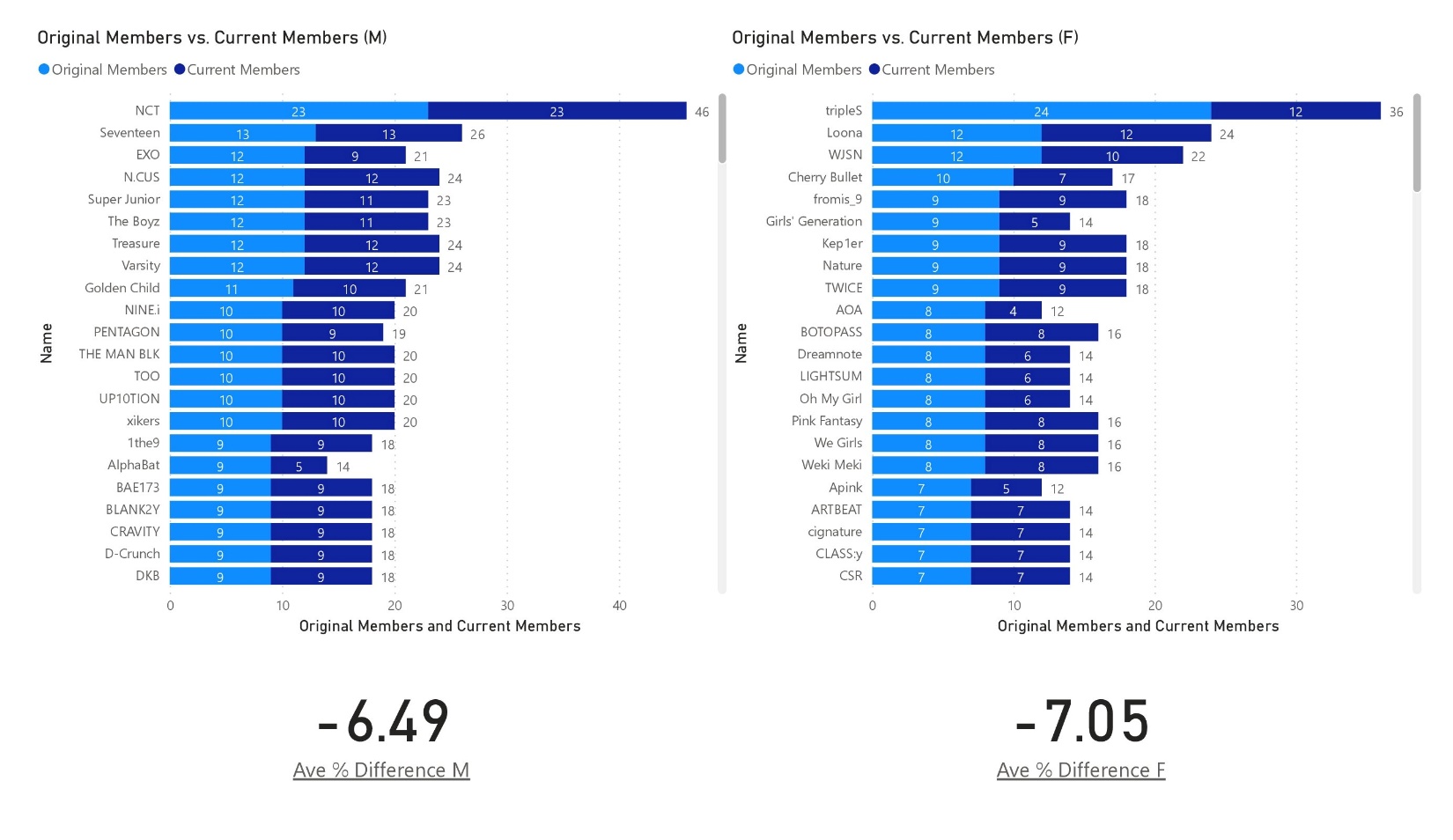
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Company | Total | Inactive | Hiatus | Active |
| JYP | 6 | 3 |  | 3 |
| SM | 6 | 3 |  | 3 |
| Cube | 4 | 1 | 1 | 2 |
| YG | 2 | 1 |  | 1 |
| Big Hit (HYBE) | 1 | 1 |  |  |
| Source (HYBE) | 2 | 1 |  | 1 |
| ADOR (HYBE) | 1 |  |  | 1 |

Examining the active groups and average member count per company provides insights into the K-pop industry's structure and the distribution of male and female idol groups.

In analyzing the data, several key observations can be made regarding the distribution of idol groups among different companies. Firstly, it's worth noting that FNC, despite having the second most male groups, is not considered part of the widely recognized "Big 5" companies. Among these Big 5 companies, SM Entertainment stands out with the highest number of groups, boasting a total of 14. JYP Entertainment follows closely with 11 groups, while HYBE, the parent company of various subsidiaries, has a total of 8 groups. YG Entertainment and Cube Entertainment both have 6 groups each.

Looking at the active groups, there are 162 companies housing male idols, while 117 companies have active female idol groups. However, it's interesting to note that among the 274 companies with female idols, only 113 exclusively focus on female artists. The remaining 161 companies have both male and female idols, indicating a diverse representation of genders within these companies.

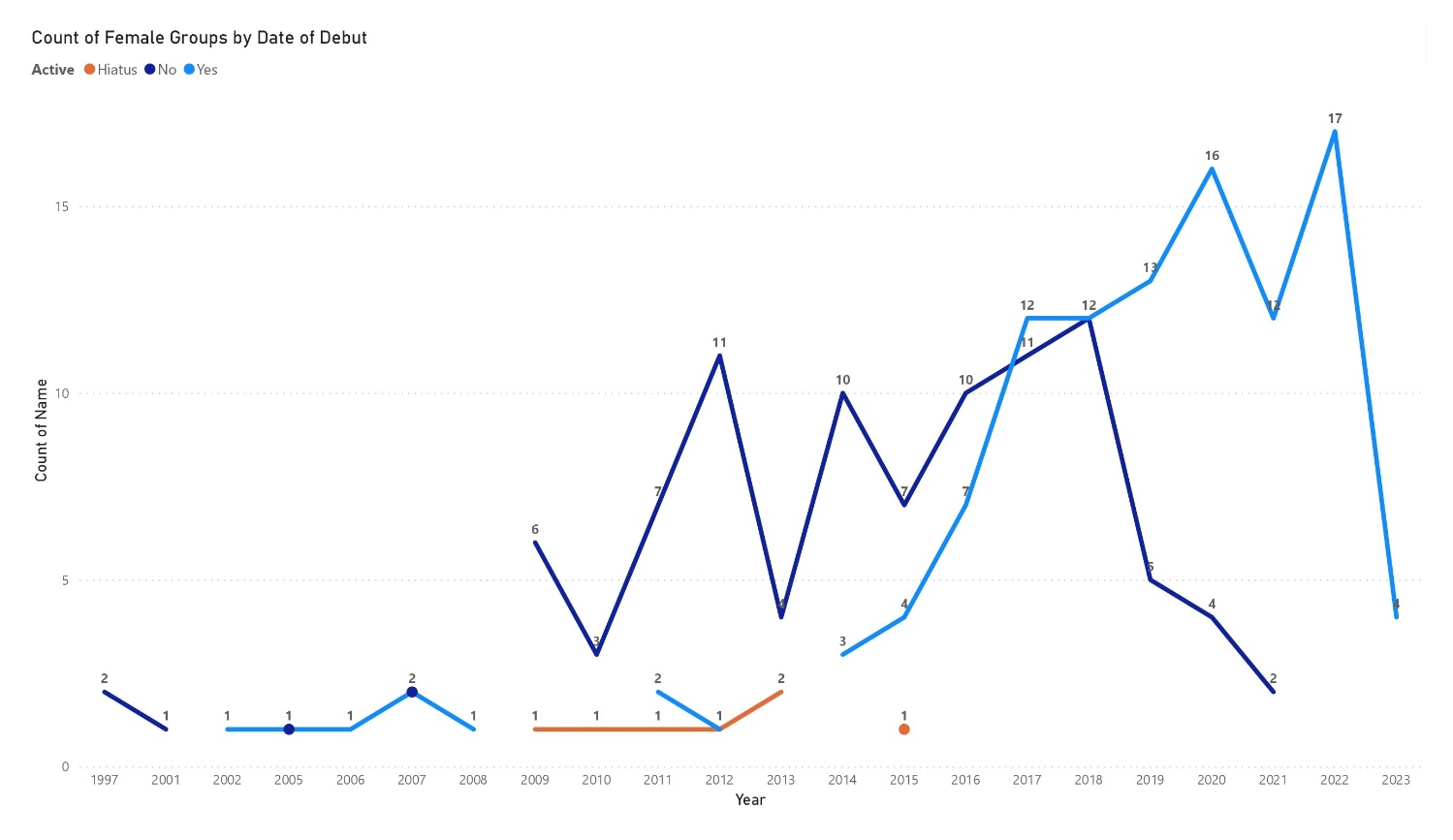
These findings shed light on the dynamics of the K-pop industry, showcasing the dominance of certain companies like SM Entertainment and the existence of mixed-gender groups alongside those exclusively composed of male or female idols.

1. Member Retention Rate

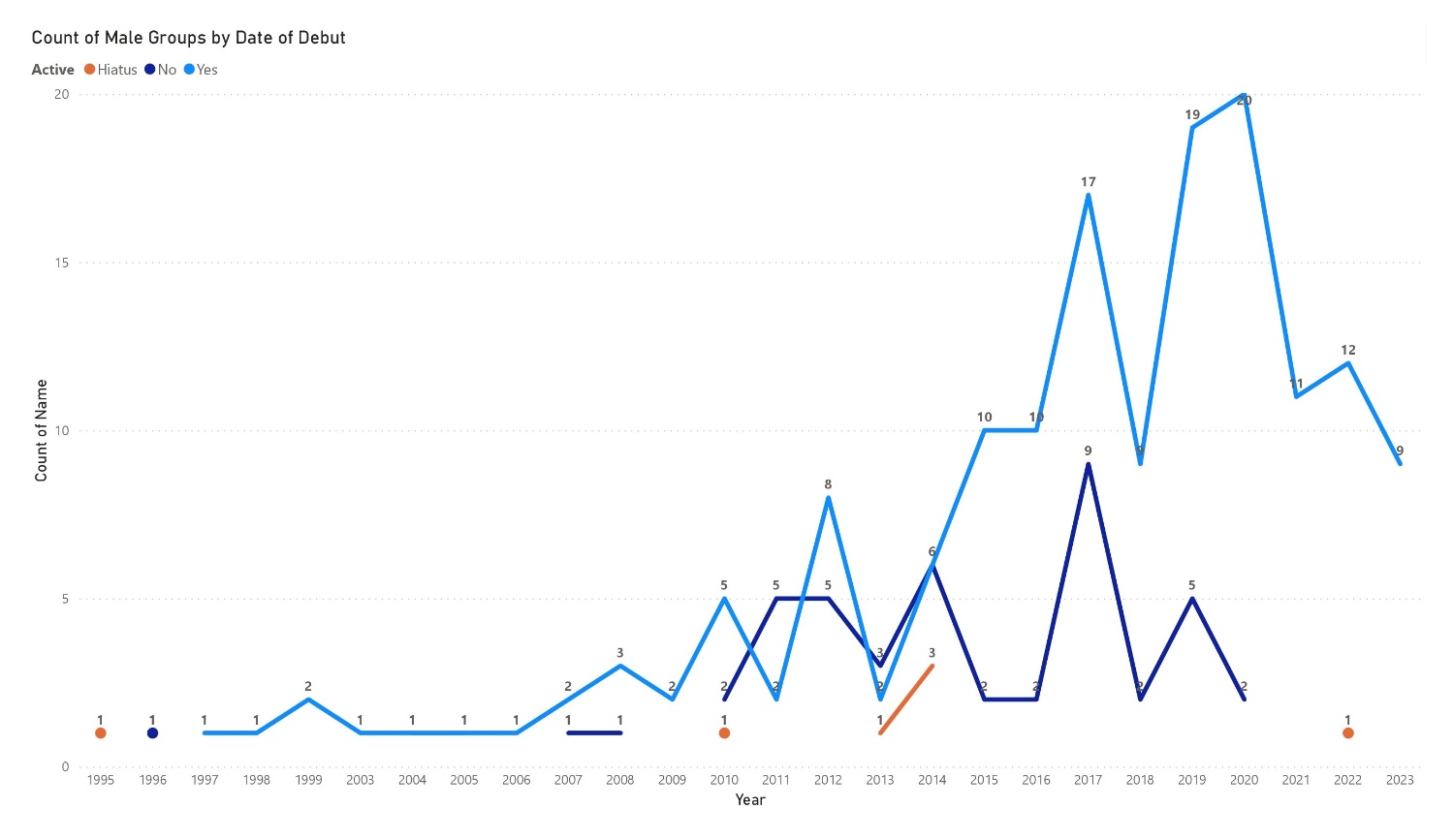
When examining member retention rates in K-pop groups, it becomes evident that there is a notable difference between male and female groups. The retention rate represents the percentage of groups that manage to maintain their original member count over time. In this analysis, it is observed that the male retention rate stands at -6.49%, while the female retention rate is slightly higher at -7.05%.

These figures suggest that female groups experience a higher degree of member turnover compared to their male counterparts. It implies that female groups are more prone to changes in their lineup, whether through departures or additions of new members. On the other hand, male groups tend to have a relatively more stable roster, with a slightly lower retention rate.

This information provides insights into the dynamics of group formation and member changes within the K-pop industry, highlighting the varying degrees of stability and continuity across male and female idol groups.

1. Count of Groups per Debut Year

The count of K-pop groups per debut year provides valuable insights into the trends and patterns of group formations within the industry. When examining female groups, it becomes apparent that the majority of currently active female groups debuted between 2017 and 2022. In contrast, the inactive groups span a wider range of debut years, primarily from 2009 to 2021. Notably, there was a significant surge in the number of female group debuts between 2015 and 2022, indicating a period of increased activity and competition.



Similarly, for male groups, the highest concentration of currently active groups debuted from 2017 to 2020. Inactive male groups, on the other hand, increase primarily between 2010 and 2020. Analogous to the trend observed in female groups, there was a notable increase in male group debuts between 2013 and 2020, reflecting a period of substantial growth and interest in the male idol market.

These findings shed light on the evolving landscape of K-pop group debuts, indicating specific time periods that experienced a surge in new group formations for both male and female artists. Such insights can be valuable for understanding the dynamics of the industry and identifying key moments of change and development.

1. **Interpretation and Discussion**

Korean Fans

Korean fans may find the data on member retention rates and debut trends interesting. They can gauge the stability and longevity of their favorite groups based on member retention rates, and they can anticipate the influx of new groups based on debut trends. This information allows them to follow and support their preferred idols and stay updated on the dynamic nature of the industry.

International Fans

International fans can gain a better understanding of the industry's gender dynamics, height expectations, and regional influences. This knowledge helps them appreciate the cultural nuances and challenges faced by idols from different countries. It also provides insight into the industry's preferences, allowing international fans to engage more meaningfully with their favorite groups and support their idols' endeavors.

Aspiring Idols

Aspiring individuals who want to become idols can analyze the data to make informed decisions about their career paths. They can identify the most common group formats, which can guide their choices when forming or joining a group. Additionally, understanding debut trends and company dynamics can help them strategize and plan their entry into the industry more effectively.

Companies

Companies in the K-pop industry can leverage the data to inform their talent acquisition strategies, group formations, and marketing efforts. The insights on member counts, company representation, and group activity status can help them identify market gaps, target specific demographics, and optimize their roster of talent. It also provides valuable information for companies to assess their competitive position within the industry.

Brand Sponsorships

Brand sponsorships play a crucial role in the K-pop industry. Companies seeking brand partnerships can utilize the data on popular group formats, member counts, and debut trends to identify potential collaborations. They can align their brand image with the preferences of target audiences, leverage the popularity of specific groups, and maximize the reach and impact of their marketing campaigns.

Investors

Investors in the K-pop industry can analyze the data to identify trends, growth opportunities, and potential areas for investment. Insights on company dynamics, debut trends, and group popularity can help investors assess the financial viability and market potential of different agencies and groups. This information allows them to make informed decisions about funding, endorsements, and long-term investments within the K-pop industry.

By understanding the implications of the data for various stakeholders, each group can make more informed decisions, tailor their strategies, and engage more effectively with the dynamic and ever-evolving world of K-pop.

1. **Conclusion**

In conclusion, the analysis of the K-pop industry provides valuable insights into the trends and patterns observed over the past four decades. The industry's impact has been widespread, with K-pop music, dramas, food, and cultural influences becoming increasingly prevalent in various countries, including the Philippines.

The analysis revealed that the industry maintains a relatively balanced representation of male and female idols, with a slight majority of female idols. South Korea remains the primary source of idols, followed by Japan and China. Notably, while Japan has a higher total number of idols, there is a difference in the male-female split, with more Japanese female idols and more Chinese male idols.

In terms of company analysis, the "Big 5" companies, including HYBE, JYP, SM, YG, and Cube, hold a significant presence in the industry. SM Entertainment has the highest number of groups, followed by JYP Entertainment. HYBE, with its subsidiary companies, accounts for a considerable number of active groups. Additionally, there are numerous other companies that contribute to the industry, both in terms of male and female idols.

The analysis also examined member retention rates, which indicate the percentage of groups that retain their original member count. The findings reveal a negative retention rate for both male and female groups, with female groups experiencing a slightly higher loss of members over time.

The data on group debuts by year indicates a surge in debuts from 2015-2022 for female groups and from 2013-2020 for male groups. This highlights a period of significant growth and activity within the industry.

Overall, this analysis provides a comprehensive overview of the K-pop industry, shedding light on various aspects such as sex predominance, nationalities, company dynamics, member retention, and debut trends. These insights can be valuable for Korean fans, international fans, aspiring idols, companies themselves, brand sponsorships, and investors seeking to understand and navigate the dynamic world of K-pop.

**References**

Kaggle Dataset: Kimjihoo. (n.d.). KpopDB. Retrieved from <https://www.kaggle.com/datasets/kimjihoo/kpopdb>

dbkpop Dataset: K-pop Girl Groups. (n.d.). Retrieved from <https://dbkpop.com/db/k-pop-girlgroups/>

dbkpop Dataset: K-pop Boy Groups. (n.d.). Retrieved from <https://dbkpop.com/db/k-pop-boybands/>