



## kPop Group Debut Astrology Dataset Analysis

## Group – 1\_Project Paper

Subgroup	Members			
1.Domain knowledge	Ma Aye Aye Khine	Ko Min Khant Kyaw		
2.Relationship group	Ma Zin Wai Wai Htet	Ma May Me Set Myint	Ma Pyae Phyoe Thwe	Ko Htut Aung Lin
3. python	Ma Theint Theint Thae	Ma Thiri Whar Whar Lwin		
4. Additional Knowledge	Ma Ei Ei Htay	Ma Su Mya Mon		

**About Dataset****K-pop Group info Variables**

No.	Variable Name	Description
1.	group name	The names of the K-pop groups, eg., BTS, NewJeans.
2.	company	Entertainment Companies where the K-pop groups come out.
3.	status	Active or Inactive (Still entertaining or stop or pause.)
4.	group type	Types of the groups such as 'Boy Group, Girl Group, Mixed Group.
5.	paks	Perfect All-kills, in South Korea, this is a music chart achievement for a song to reach number one simultaneously on the real-time, daily, and weekly charts of the Instiz iChart.
6.	music show awards	The number of prizes based on the music shows such as 'MC countdown, Inkigayo and etc.
7.	end year awards	Major year-end K-pop awards include the MAMA Awards, Melon Music Awards (MMA), Golden Disc Awards (GDA) and Seoul Music Awards (SMA).
8.	physical sales	The number of physical albums (like CDs) sold.
9.	organic-YouTube-views	K-pop videos are earned naturally through viewers who are not influenced by paid promotions.

## Astrology Variables

No.	Variable Name	Description
1.	Sun sign	Represents core personality, ego and identity- who you really are inside.
2.	Asc sign (Ascendant)	Represents the mask or first impression, how others (audiences) see.
3.	Moon phase	Represents the moon's shapes – New moon, first quarter moon, full moon, last quarter moon. It affects emotional style and life rhythm.
4.	Moon sign	Represents emotions, instincts and what comforts.
5.	Mercury sign	Represents communication, learning, thinking and decision-making.
6.	Mercury rx	Represents a period of potential communication mishaps, often influencing the group's or idol's activities, themes, and concepts.
7.	Venus sign	Represents love style, relationships, beauty, and value –art.
8.	Venus rx	When Venus goes retrograde, it appears to move backward in the sky- symbolically, its energy turns inward.
9.	Mars sign	Represents energy, drive, action, and how assert yourself.
10.	Mars rx	When Mars is retrograde-energy turns inward.
11.	Jupiter sign	Represents growth, wisdom, luck, and faith.
12.	Jupiter rx	When Jupiter is retrograde, inner growth focus.
13.	Saturn-sign	Represents discipline, responsibility, structure and life lessons.
14.	Saturn rx	When Saturn is retrograde, lessons are internal (private sense of duty, serious inner critic).

### Dataset Overview

The dataset includes information such as **group name**, **debut date**, **zodiac sign**, **Mercury sign**, and **retrograde status**.

**Below is a summary of the dataset structure and key fields.**

This dataset likely contains information on **120 K-pop groups**, with **27 variables recorded** for each group.

Given the dataset name, the variables probably relate to the **debut date**, **astrological signs**, and possibly other attributes of the group members or the group itself.

## Purpose

1. The purpose of this analysis is to understand how **Venus sign placements** might influence the timing and success of K-pop group debuts.
  2. Venus, associated with **beauty, attraction, and entertainment**, provides an intriguing lens for examining patterns behind debut decisions.
- 

## Objective

The main objectives of this report are to:

1. Group and analyze K-pop debuts by Venus sign.
  2. Compare debut frequencies and trends across different astrological groupings.
  3. Visualize debut year distributions and their alignment with planetary movements.
  4. Generate insights on how astrological patterns may reflect industry or cultural timing preferences.
- 

## Visualization Presentation

### A Descriptive Analysis of Performance Skewness and Astrological Debut Patterns

	paks	music_show_awards	end_year_awards	physical_sales	organic_youtube_views	mercury_rx	venus_rx	mars_rx	jupiter_rx	saturn_rx
count	120.000000	120.000000	120.000000	1.200000e+02	1.200000e+02	120.000000	120.000000	120.000000	120.000000	120.000000
mean	0.450000	22.883333	11.591667	3.395663e+06	1.409812e+09	0.191667	0.075000	0.150000	0.216667	0.350000
std	1.158454	28.204292	18.902446	6.783900e+06	3.482108e+09	0.395263	0.264496	0.358569	0.413701	0.478969
min	0.000000	0.000000	0.000000	0.000000e+00	6.775056e+06	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	3.000000	1.000000	2.788650e+05	1.930346e+08	0.000000	0.000000	0.000000	0.000000	0.000000
50%	0.000000	12.500000	5.000000	9.235175e+05	5.326264e+08	0.000000	0.000000	0.000000	0.000000	0.000000
75%	0.000000	33.000000	16.000000	3.422668e+06	1.181571e+09	0.000000	0.000000	0.000000	0.000000	1.000000
max	7.000000	164.000000	155.000000	4.485270e+07	2.904492e+10	1.000000	1.000000	1.000000	1.000000	1.000000



### K-Pop Performance Metrics (Skewed Success) \_ Variables analysis

#### 1. Metric \_ PAKs (Perfect All-Kills):

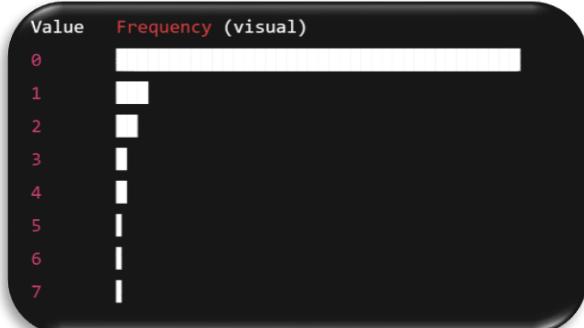
A Perfect All-Kill (PAK) is the highest possible chart achievement for a song in South Korea.

A song must be **Number 1** on the **real-time (hourly), daily, and weekly** rankings of all major South Korean music streaming platforms.

## Statistics:

- Mean  $\mu$  (0.45) > Median (50%) (0.00)**

Term	Focuses	Meaning
Positively skewed,	Statistical description	The distribution has <b>many zeros</b> and a <b>small number of high values</b> , producing a <b>right tail</b> .



### Short Summary report

Many zeros with a few values up to 7 create a strong right tail caused by a few high observations

### Bar\_Chart

Tail extends to the **right side**

- 75% = 0.0000**

(This means **at least 75%** of values are **zero**)

- Max = 7.0**

(A few very large values pull the distribution to the **right**)

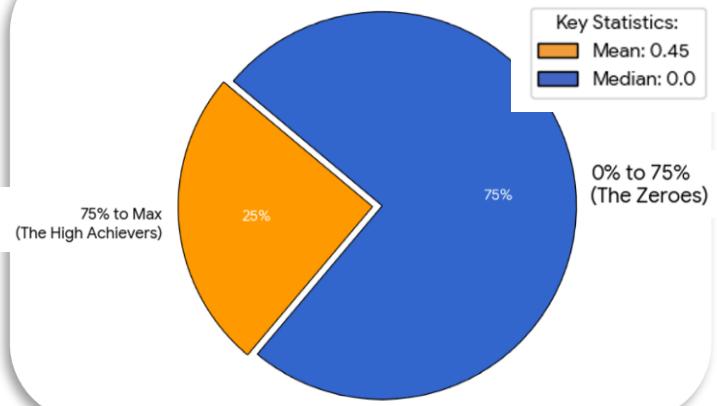
- Std  $\sigma$  = 1.158454**

(is large relative to mean → indicates spread from some high values)

**Std  $\sigma$  (1.158454) > Mean  $\mu$  (0.45)**

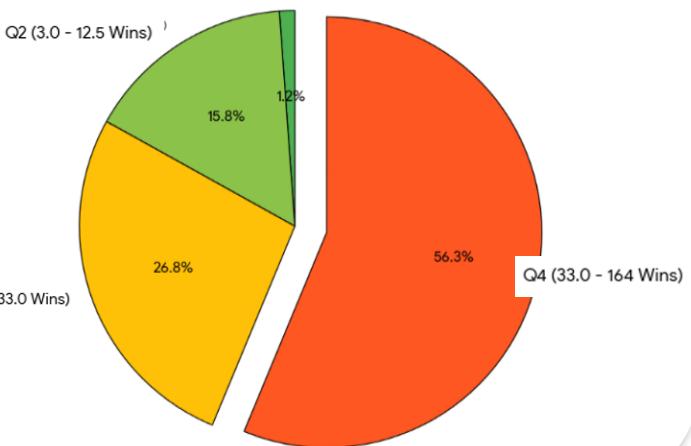
### Pie \_ Chart

Distribution of Perfect All-Kills (paks)



## Pie \_ Chart

Proportion of TOTAL Music Show Wins by Population Quartile  
Q1 (0 - 3.0 Wins)



2) Max = 164

(A few very high values pull the distribution to the right)

3) Std  $\sigma$  = 28.20

(is large indicates widespread)

Std  $\sigma$  (28.20) > Mean  $\mu$  (22.88)

## 📝 Insight and Suggestion

Most groups/artists have relatively few wins (half have 12.5 or fewer), but a small, elite group has amassed a massive number of wins (up to 164).

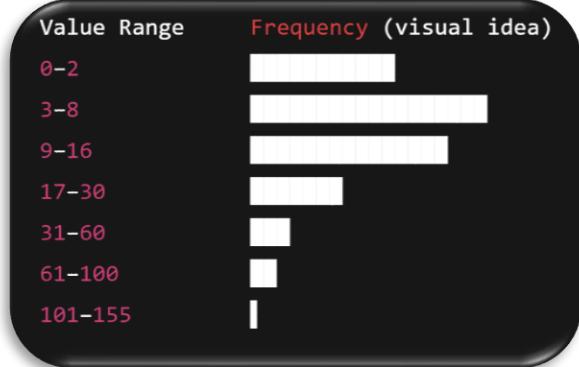
For K-Pop groups aiming to join this elite category, the primary suggestion is to focus on **sustained, compounding popularity** rather than single-release spikes.

### 3. Metric\_End\_Year\_Awards: (Year-End Awards):

Statistics:

1) Mean  $\mu$  (11.59) > Median (50%) (5.00)

Term	Focuses	Meaning
Positively skewed	Statistical description	The <i>end year award</i> variable is <b>positively skewed</b> , with many small values and a few very large values extending the right tail.



## 📝 Short Summary report

Majority of values are small; extremely high awards (up to 155) cause right-tail outliers.

## Bar\_Chart

Tail extends to the right side

2) Max = 155

(a few extremely high values pull the tail to the right)

## Pie \_ Chart

## Most values (25%–75%)

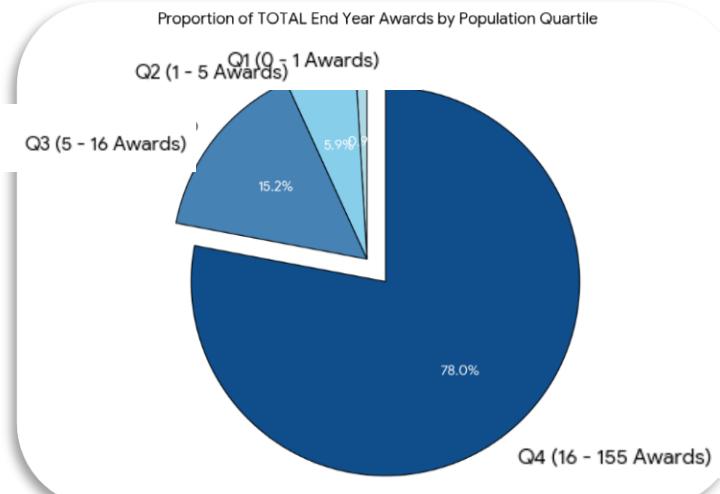
(fall between 1 and 16, but some awards reach 155)

3) Large Std  $\sigma = 18.90$

(strong spread caused by large values)

Std  $\sigma (18.90) > \text{Mean } \mu (11.59)$

- 1) A lot of groups have only 1–8 awards
- 2) Fewer groups have 20–60
- 3) Only very few reach 100–155



## 💡 Insight and Suggestion

The distribution of End Year Awards strongly indicates a "winner-take-all" environment, where success is highly concentrated among a small, elite fraction of groups.

The most critical suggestion for a K-Pop group's future success is to build a strategy centered on **cumulative, long-term dominance**, recognizing that the industry heavily rewards the established elite.

## 4. Metric: Physical Sales

Statistics:

1) Mean  $\mu (3.39M) > \text{Median (50\%) (0.923M)} \rightarrow \text{right skew}$

Term	Focuses	Meaning
Positively skewed	Statistical description	Most groups sell between 0.2M–3M, but a few groups have massive physical sales far above 10M, creating huge right-side outliers.

## 📝 Short Summary report

Most values are below 3.4M, but the maximum jumps to 44.8M.

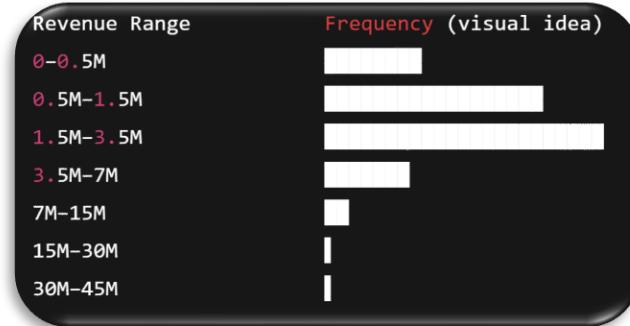
2) Max = 44.8M

(is extremely far from most values)

3) Std  $\sigma = 6.78M$

(is very large widespread caused by huge outliers)

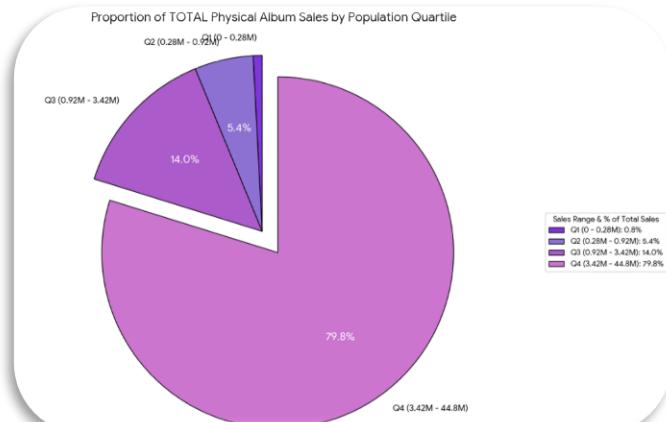
Std  $\sigma (6.78M) > \text{Mean } \mu (3.39M)$



## Bar\_Chart

Tail extends to the right side

## Pie \_ Chart



## Insight and Suggestion

The physical sale market is the most extreme "winner-take-all" landscape. The **Top 25%** of groups generates approximately **74% of all physical sales**. The sales metric is not a general measure of success but a measure of **fandom mobilization capability** at the highest level.

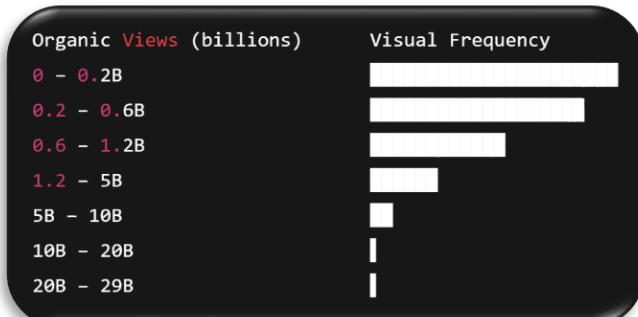
To achieve the multi-million-dollar sales that define the elite 25%, focus entirely on strategies that **maximize pre-order volume** and highly organized, bulk-buying **fandom power**. Physical sales are driven by dedication, not casual interest.

### 5. Metric: Organic YouTube Views:

#### Statistics:

- 1) Mean  $\mu$  (3.49B) > Median (50%) (0.53B) → (classic right skew)

Term	Focuses	Meaning
Positively skewed	Statistical description	Most videos fall between 1B–7B organic views Only a few videos reach 15B–30B, creating <b>extreme outliers</b>



Bar\_Chart

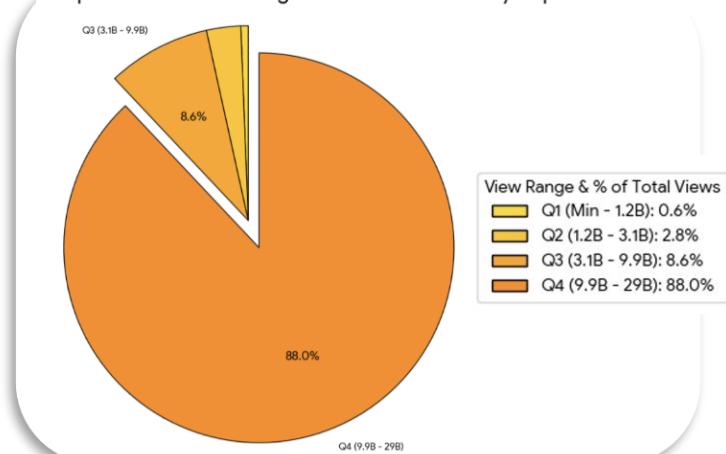
Tail extends to the **right side**

#### Short Summary report

The *organic YouTube views* variable is **strongly positively skewed**, with most values below 1.2B and a few extremely large outliers up to 29B creating a long right tail.

Pie \_ Chart

Proportion of TOTAL Organic YouTube Views by Population Quartile



- 2) Max = 29.04B

- 3) Min = 6.775M

Most values are between 0.006–1.1B

A few extremely huge values (up to 29B) pull the tail right

- 4) Std  $\sigma$  = 3.48B → extremely large spread

Mean  $\mu$  (3.49B) > Std  $\sigma$  (3.48B)

## Insight and Suggestion

The "average" view count is meaningless. It's so high only because of the few videos with 20-30 billion views, not because the typical video performs well.

Physical sales are an aggressive measure of dedication, and that dedication must be constantly cultivated through exclusive content and interaction.

## Overall Summary

### Distribution:

Most performance-related variables (awards, sales, views) are **heavily right-skewed**, indicating strong inequality among K-pop groups.

### 💡 Retrograde Patterns \_ Variables analysis

#### 6. Metric: mercury\_rx (Mercury Retrograde):

##### Statistics:

1) Mean  $\mu$  (0.191667) ( $\approx 19.17\%$ ) > Median (0.000)



Bar\_Chart

Tail extends to the **right side**

Term	Focuses	Meaning
<b>Positively skewed</b>	Statistical description	The distribution of mercury_rx is <b>highly right-skewed</b> (positively skewed). Since 75% of the observed periods are 0.0000, the data is concentrated at the minimum, with a long tail stretching to the maximum value of 1.0000.

### 📝 Short Summary report

The mercury\_rx variable is **highly right-skewed** (positively skewed). This is because most observations are 0, and only a small fraction are 1. The mean (0.19) is much higher than the median (0) is indicating this skew.

2) Max = 1.000

3) Std  $\sigma$  = 0.395263

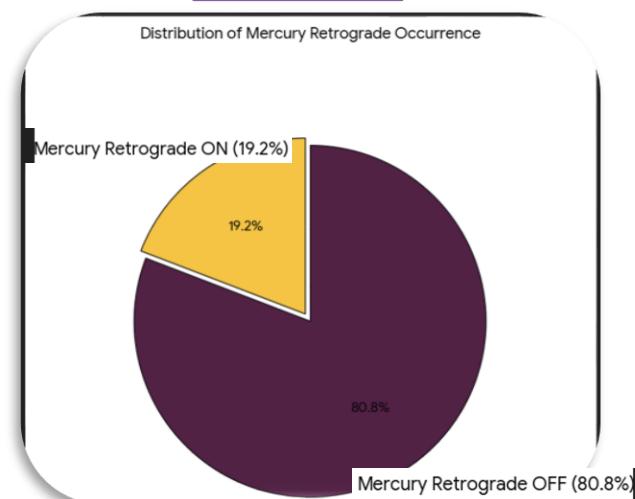
Std  $\sigma$  (0.395263) > Mean  $\mu$  (0.191667) ( $\approx 19.17\%$ )

### 💡 Insight and Suggestion

Mercury Retrograde (mercury\_rx = 1) occurred for approximately 19.2% of the total observed period (23 out of 120 counts). This is significantly more frequent than Venus Retrograde (7.5%), but it still represents a minority of the total time.

Strategically **avoid scheduling major contract signings, album releases, critical technology rollouts, or the start of complex international tours** during the Mercury Rx periods.

Pie \_ Chart



## 7. Metric: venus\_rx (Venus Retrograde):

### Statistics:

- Mean  $\mu$  [0.075 ( $\approx 7.5\%$ )] > Median (0.000)

Pie Chart: Proportion of venus\_rx



Bar\_Chart

Tail extends to the right side

Term	Focuses	Meaning
Positively skewed	Statistical description	A distribution with many 0's and a few 1's always produces right skew.

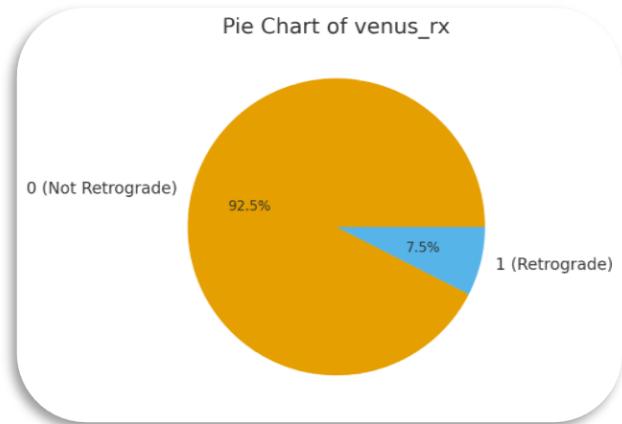
### Short Summary report

The *venus\_rx* variable is positively skewed, with 92.5% of cases being 0 and only 7.5% being 1. This distribution indicates that Venus Retrograde periods are rare in the dataset. Due to extreme imbalance, the variable may not provide strong predictive power on its own.

- Max = 1.000
- 25%, 50%, 75% all = 0
- Std  $\sigma$  = 0.264496

Std  $\sigma$  (0.264496) > Mean  $\mu$  [0.075 ( $\approx 7.5\%$ )]

Pie \_ Chart



### Insight and Suggestion

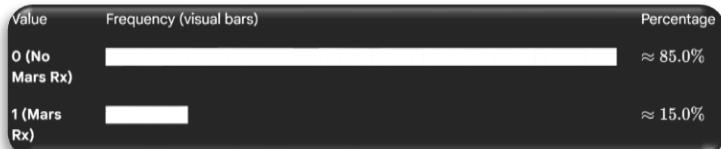
Astrologically, Venus Retrograde is a highly predictable, cyclical event that occurs for about 40-42 days every 18 months. The low frequency and the fact that all quartiles are zero perfectly reflect this known, narrow astronomical cycle.

Do not attempt to launch *new* products, services, or partnerships during the *venus\_rx* = 1 period. Instead, use the time to **review, refine, and renegotiate** existing business models, client relationships, or financial strategies, as this aligns with the astrological theme of introspection and reassessment.

## 8. Metric: mars\_rx (Mars Retrograde):

### Statistics:

- Mean  $\mu$  [0.15 ( $\approx 15\%$ )] > Median (0.000)



Bar\_Chart

Tail extends to the right side

Term	Focuses	Meaning
Positively skewed	Statistical description	A distribution with many 0's and a few 1's always produces right skew.

## Short Summary report

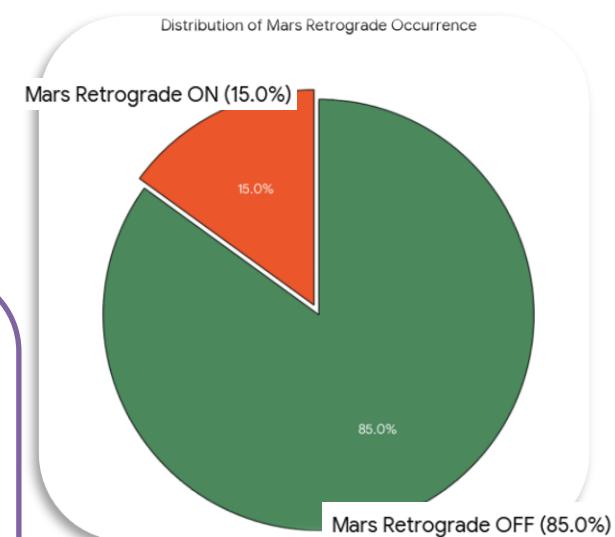
The mars\_rx variable is positively skewed, with 15.0% of cases being 1 (Mars Retrograde). This low frequency indicates that Mars Retrograde periods are rare in the dataset. Due to the 85% /15% imbalance, special care (like class weighting) should be taken if using the variable in predictive models, focusing its analysis on strategic timing and action-oriented outcomes.

- 2) **Max = 1.000**
- 3) **25%, 50%, 75% all = 0**
- 4) **Std  $\sigma = 0.358569$**

**Mean  $\mu [0.15 (\approx 15\%)] > \text{Std } \sigma (0.358569)$**

Std  $\sigma$  0.359\$ is small because the data can only be 0 or 1. This confirms that the variable has **low internal spread** (low variance).

### Pie \_ Chart



### Insight and Suggestion

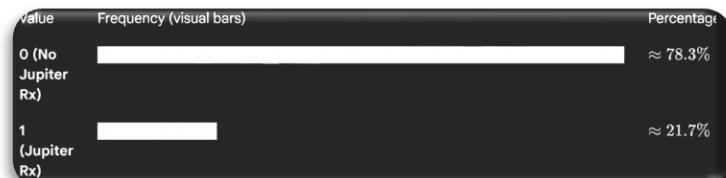
Astrologically, Mars Retrograde as the planet of action, its retrograde can deplete your motivation and energy. It's a period to reflect on your ambitions and reassess how you handle conflict.

Test its interaction with metrics sensitive to action and timing:  
**Debut/Comeback Timing, Aggressive Marketing Budget Spends, and Public Relations Outcomes** (e.g., negative press incidence).

## 9.Metric: jupiter\_rx (Jupiter Retrograde):

### Statistics:

- 1) **Mean  $\mu [0.216667 (\approx 21.7\%)] > \text{Median} (0.000)$**



Term	Focuses	Meaning
Positively skewed	Statistical description	A distribution with many 0's and a few 1's always produces right skew.

### Bar\_Chart

Tail extends to the **right side**

## Short Summary report

The jupiter\_rx variable is positively skewed, with 21.7% of cases being 1 (Jupiter Retrograde). This is the most frequent retrograde in the dataset, making the 78 \ 22% distribution the most balanced.

### Pie \_ Chart

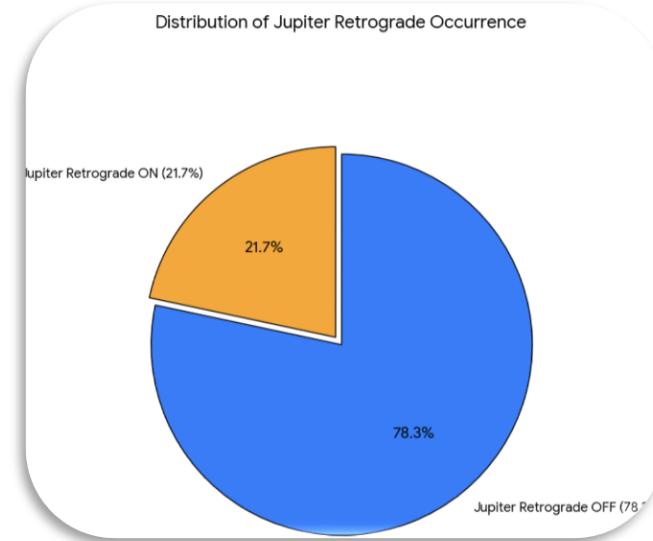
- 2) Max = 1.000
- 3) 25%, 50%, 75% all = 0
- 4) Std  $\sigma$  = 0.413701

Std  $\sigma$  (0.413701) > Mean  $\mu$  [0.216667 ( $\approx$  21.71%)]

### Insight and Suggestion

Astrologically, Jupiter Retrograde, Prompts introspection and a re-evaluation of your beliefs, goals, and spiritual philosophies. Instead of external expansion, the focus shifts to inner growth.

Since Jupiter rules expansion, test this variable against metrics like **return on investment (ROI)**, **profit margins**, or **long-term global expansion** outcomes. A strong correlation here could suggest that the retrograde period acts as a statistical regulator, slowing down or dampening the immediate success of ambitious projects.



### 10. Metric: saturn\_rx (Saturn Retrograde):

Statistics:

- 1) Mean [0.358333 ( $\approx$  35.81%)] > Median (0.000)



Bar\_Chart

Tail extends to the **right side**

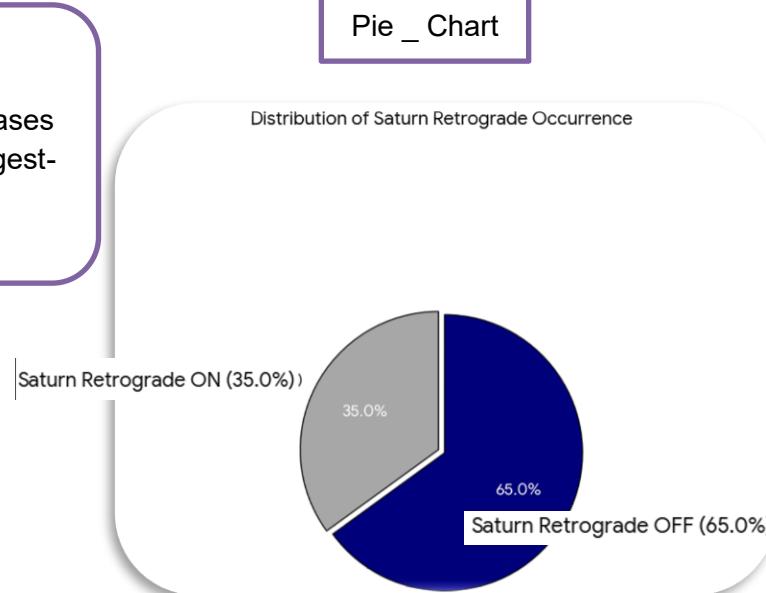
Term	Focuses	Meaning
Positively skewed	Statistical description	A distribution with many 0's and a few 1's always produces right skew.

### Short Summary report

The saturn\_rx variable is positively skewed, with 65.0% of cases being 0 and 35.0% being 1. This frequency makes it the longest-lasting retrograde in the dataset and results in the **most balanced binary distribution** among the rx variables.

- 5) Max = 1.000
- 6) 25%, 50%, 75% all = 0
- 7) Std  $\sigma$  = 0.478969

Std  $\sigma$  (0.413701) > Mean [0.358333 ( $\approx$  35.81%)]



## Insight and Suggestion

Astrologically, Jupiter Retrograde, encourages you to reflect on your long-term goals, personal boundaries, and responsibilities. It's a time for self-discipline and strengthening your life's foundations.

Focus on accountability, long-term resource allocation, debt structure, and the strength of the group's organizational foundation (management, contracts, etc.). New, unstructured plans should be postponed.

### Astrological Variables - Overall Summary

Binary proportions show varying planetary retrograde frequencies, potentially offering symbolic insights if correlated with success metrics.

The **percentages (proportions)** of each planet being in retrograde motion on the debut dates of new groups **are not uniform** (they differ).

Therefore, by comparing these percentages with the group's success metrics—such as awards received, sales, or popularity—it is possible to gain **interesting insights** regarding **astrological significance** (e.g., how a planetary retrograde might positively or negatively affect success).

### Delineating the Meaning of Astrological

Section Title	Clarification
<b>Binary Proportions</b>	1 = The group debuted while [Planet] was in retrograde. 0 = The group did not debut while the planet was in retrograde. This is called a 'Binary' variable. 'Proportion' refers to the percentage of the total number of groups.
<b>Differences in Planetary Retrograde Frequencies</b>	The retrograde percentages for different planets like Mercury, Venus, and Mars are not the same. For example, Mercury is 19%, while Venus is only 7.5%.
<b>Potential for Symbolic Insights</b>	Astrology assigns symbolic meanings (e.g., problems in communication, beauty, or action) to planetary retrogrades. This knowledge can be used when analyzing these variables to provide insightful understanding.
<b>If correlated with Success Metrics</b>	It means that meaningful results will only emerge by statistically comparing the astrological conditions with success metrics such as 'Music Show Wins' or 'Physical Sales'.

## Statistical View and The Formula

The table is a direct output of **Descriptive Statistics**, which summarizes the central tendency, dispersion, and shape of a dataset's distribution.

### Statistical View (Data Structure)

The table provides a **univariate summary** for each of the  $P = 9$  selected variables in the  $N = 120$  group dataset.

#### **Central Tendency:**

Indicated by the **Mean** and the **Median (50th Percentile)**.

The large disparity between the mean and median for the performance metrics (paks, music\_show\_awards, physical\_sales, organic\_youtube\_views) signifies strong **positive (right) skewness** and the presence of significant **outliers** (the highly successful groups).

#### **Dispersion:**

Indicated by the **Standard Deviation (std)**. A large standard deviation relative to the mean (e.g., 1.158 for a mean of 0.45 in paks) confirms high variability and spread in the data.

#### **Position/Range:**

Indicated by **Min, Max, and the Quartiles (25%, 50%, 75%)**. These define the spread of the middle 50% of the data (the **Interquartile Range**  $IQR = Q3 - Q1$ ) and the total range (Max - Min).

```

df[['mercury_sign','mercury_rx']].max()

mercury_sign    Virgo
mercury_rx         1
dtype: object

df[['venus_sign','venus_rx']].max()

venus_sign    Virgo
venus_rx         1
dtype: object

df[['mars_sign','mars_rx']].max()

mars_sign    Virgo
mars_rx         1
dtype: object

df[['jupiter_sign','jupiter_rx']].max()

jupiter_sign    Virgo
jupiter_rx         1
dtype: object

df[['saturn_sign','saturn_rx']].max()

saturn_sign    Virgo
saturn_rx         1
dtype: object

```

 **Key Insight**

The consistent output of '**Virgo**' for all five planetary sign columns indicates that, among the **120 K-pop groups**, at least one group debuted when **Mercury, Venus, Mars, Jupiter, and Saturn** were each individually placed in the sign of **Virgo** on their respective debut dates.

The output of '**1**' for all five retrograde columns (\*\_rx) simply confirms that the max() value for these binary columns (where 0 = Direct and 1= Retrograde) is 1, meaning **at least one group debuted during the retrograde phase** of each of these five planets.

**Crucial Caveat:** The max() function on a column pair does **not** mean the same K-pop group debuted with *all* five planets in Virgo or all five in retrograde. It simply finds the lexicographically (alphabetically) largest string (which is 'Virgo' among the zodiac signs) and the highest numerical value (1) for each column *independently* across all 120 rows.



### Statistical View (Max Function on Mixed Data Types)

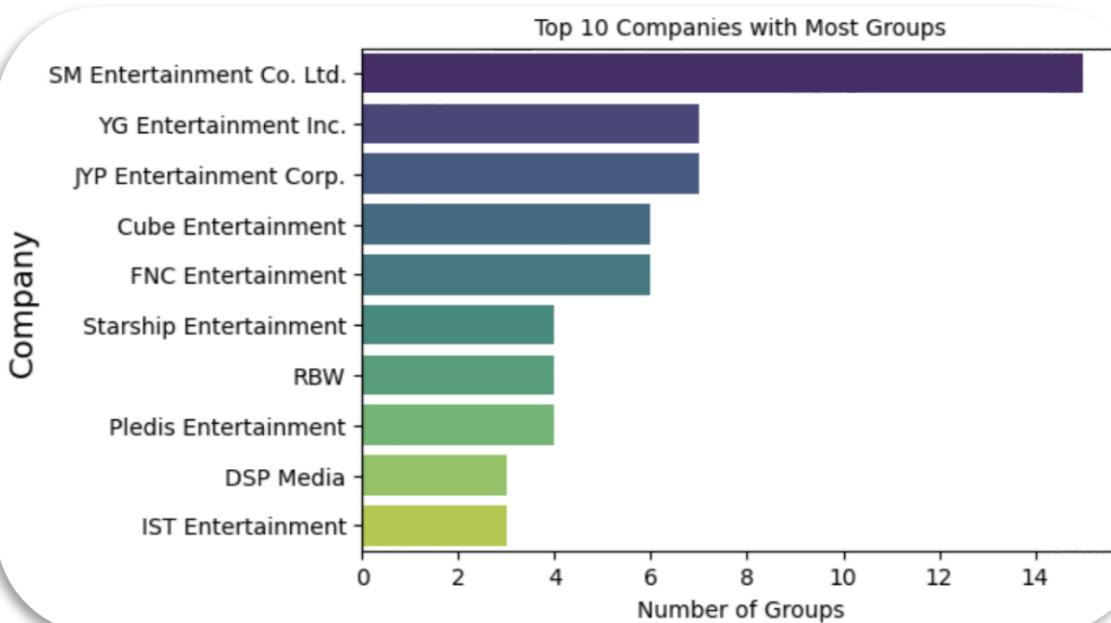
The operation `df[['sign', 'rx']].max()` is a **column-wise aggregation** where the maximum value is calculated for each of the two selected columns independently over all N=120 rows.

**Categorical Column (\*\_sign):** The column contains **nominal categorical** data (text strings representing the zodiac signs: Aries, Taurus, Gemini, ..., Virgo, ..., Pisces).

The `max()` function on strings uses **alphabetical ordering**. The string that comes last alphabetically is considered the "maximum."

---

### A horizontal bar chart visualizing from the Top 10 Entertainment Companies



The total number of groups is calculated as:

- a) SM Entertainment: 15 groups
- b) YG Entertainment: 7 groups
- c) JYP Entertainment: 7 groups

Total:  $15 + 7 + 7 = 29$  groups

## Top 10 Company Group Count Summary

Rank	Company	Number of Groups
1	SM Entertainment Co. Ltd.	15
2	YG Entertainment Inc.	7
3	JYP Entertainment Corp.	7
4	Cube Entertainment	6
5	FNC Entertainment	6
6-10	Starship, RBW, Pledis, DSP, IST	3-4 each
TOTAL	(SM + YG + JYP)	29



### Statistical View and the Formula

#### Statistical View (Top Frequency)

The chart displays the top **k=10 frequencies** of the **Nominal Categorical** variable 'Company'.

- a) **Variable: 'Company' (Categorical).**
- b) **Metric: 'Number of Groups' (Frequency or Count, Discrete Numerical).**
- c) **Skewness: The distribution of group counts across all 62 companies is highly positively (right) skewed. The chart visualizes the peak of this skewness.**

The chart is generated by performing a frequency count, selecting the top k results, and then plotting them.

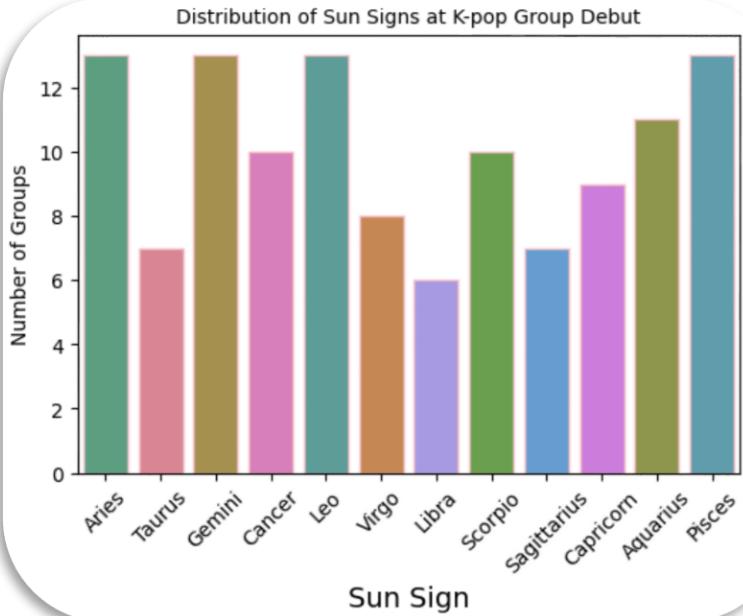
**Value Counts:** The formula is the same as the previous frequency analysis

$$\text{Count}_j = \text{Frequency}(\text{Company } j)$$

**Top k Selection:** The resulting list of (Company, Count) pairs is sorted in descending order based on the count, and only the top 10 pairs are retained.

$$\text{Top 10} = \{(\text{Company}_j, \text{Count}_j) \mid \text{Count}_j \text{ is in the top 10 highest frequencies}\}$$

A bar chart showing the frequency distribution based on the Sun Sign (Zodiac Sign)



### Key Insight: Uneven Distribution of Debut Sun Signs

The chart reveals that **K-pop group debuts** are not **uniformly distributed** across the **12 zodiac signs**.

There are noticeable peaks and troughs, suggesting that groups may **strategically or coincidentally cluster their debuts in certain periods**.

#### Most Common Debut Signs (The Peaks):

**Aries, Gemini, Leo, and Pisces** all have the highest frequency, with **13 groups** debuting under each sign. These four signs account for **52 groups, or  $\approx 43\%$  of the entire dataset**.

#### Least Common Debut Signs (The Troughs):

1. Libra has the lowest frequency, with only 6 groups debuting.
2. Taurus (7 groups) and Virgo (8 groups) are also among the least common.

This unevenness could lead to interesting astrological hypotheses: Do groups debuting under high-frequency 'Fire' signs (Aries, Leo) exhibit distinct characteristics compared to groups debuting under low-frequency 'Earth' signs (Taurus, Virgo)?



## Statistical View and the Formula

### Statistical View (Frequency Distribution)

The chart is a visual representation of a **frequency distribution** for the **nominal categorical variable** 'Sun Sign at Debut'.

**Data Type:** The x-axis (Sun Signs) is **Nominal Categorical**. The y-axis (Number of Groups) is **Discrete Numerical** (a count).

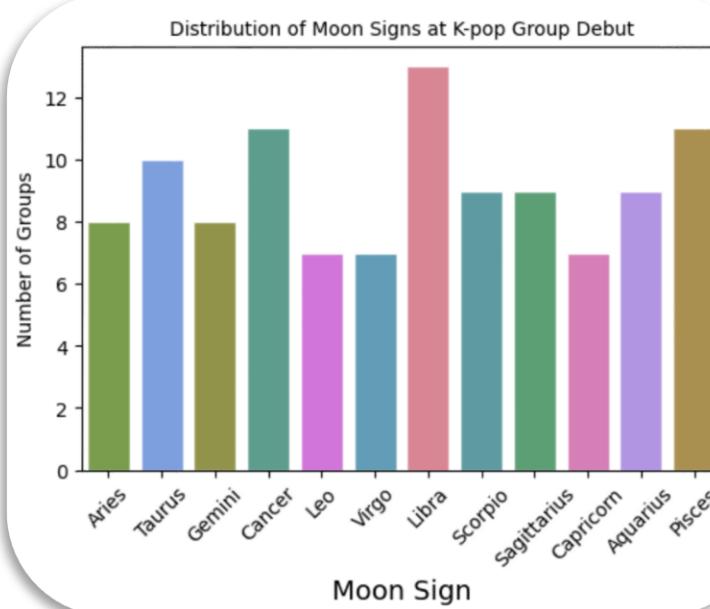
**Central Tendency:** The distribution is **multimodal**—it has multiple modes (peaks) at 13 (Aries, Gemini, Leo,

**Expected Uniform Count:** Since the total number of groups is  $N=120$  and there are  $k=12$  signs, the expected number of groups per sign under a perfectly uniform distribution would be the mean:

$$\text{Expected Count} = \frac{N}{k} = \frac{120}{12} = 10 \text{ groups}$$

The variation above and below 10 quantifies the degree of non-uniformity. For example, Aries is  $13-10=3$  groups above the mean, while Libra is  $6-10=-4$  groups below the mean.

### A bar chart showing the frequency distribution based on the Moon Sign (Zodiac Sign)





## Key Insight: The Dominance of Libra Moon Debuts

The chart reveals that the **Moon Sign** distribution at debut is also **highly non-uniform**, but with a single, clear peak, unlike the multimodal Sun Sign distribution.

### 1) Most Common Debut Moon Sign (The Peak):

- a) **Libra** is the overwhelmingly most frequent Moon Sign, with **13 groups** debuting under it.

### 2) Common Debut Moon Signs:

- a) **Cancer** (11 groups) and **Pisces** (11 groups) are also highly common.

### 3) Least Common Debut Moon Signs (The Troughs):

- a) **Leo**, **Virgo**, and **Capricorn** are the least frequent, with only **7 groups** each.

The Moon changes signs much faster than the **Sun (approximately every 2.5 days)**, making its placement more sensitive to the exact debut time.

The strong preference for **Libra Moon** (associated with balance, harmony, and aesthetics) could be a highly **strategic or coincidental astrological cluster** for K-pop debuts, which are heavily focused on visual appeal and public harmony.



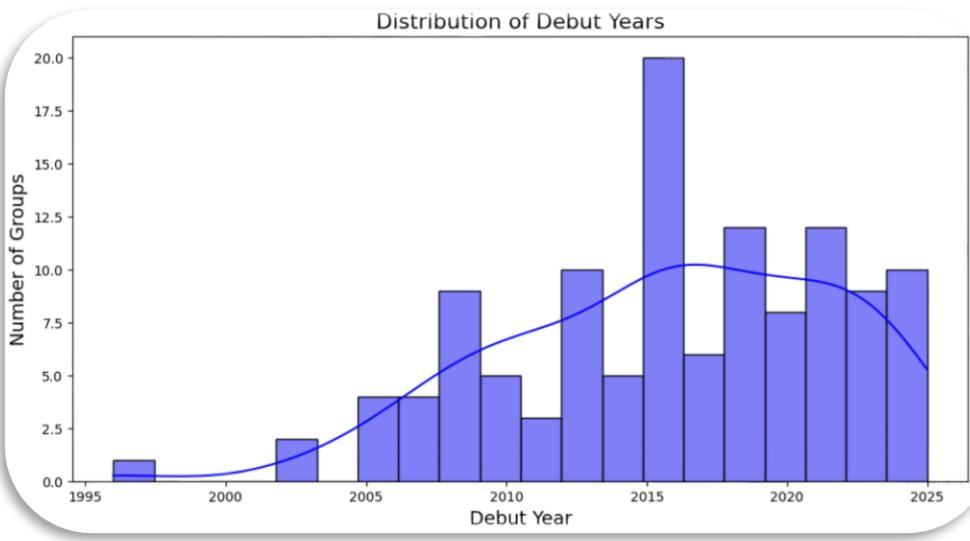
## Statistical View and the Formula

### Statistical View (Frequency Distribution)

This chart provides a visual **frequency distribution** for the **nominal categorical variable** 'Moon Sign at Debut'.

- a) **Variable Type:** Moon Sign is **Nominal Categorical**. The count is **Discrete Numerical**.
- b) **Central Tendency:** The distribution is unimodal, with the clear mode (the highest frequency) being **13 at Libra**.

A histogram of the Debut Year, showing curve



### 💡 Key Insight: The 2015 Debut Boom 🚀

The distribution shows that the dataset primarily focuses on groups debuting in the **21st century**, with a strong bias toward the **mid-2010s to present**.

#### Peak Debut Year:

1. The most prominent feature is the tall bar centered on **2015**, which has the highest frequency of debut groups, reaching **20 groups**.
2. This suggests a significant *surge* or "boom" in K-pop group debuts around that year.

#### Skewness/Shape:

1. The distribution is **left-skewed** (or negatively skewed) if viewed from the peak toward the tail (1995–2010), but more accurately, it's a **late-peaking distribution**.
2. The median debut year is likely close to the mean of 2025.16 (previously calculated).

#### Sustained Activity:

While 2015 is the peak, the period from **2016 to 2024** shows sustained high debut activity, with most years having between **8 and 12 groups** debuting, demonstrating the continued expansion of the K-pop industry.



## Statistical View and the Formula

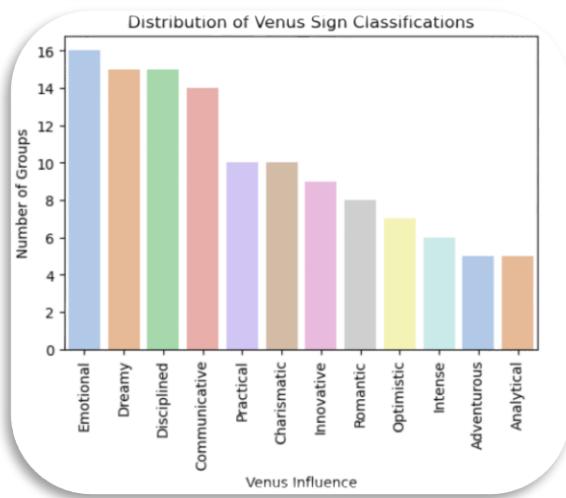
### Statistical View (Time-Series Distribution)

The chart provides a statistical summary of the **Discrete Numerical variable 'Debut Year'**.

#### Histogram:

1. The bars represent the **absolute frequency (f)** of groups debuting within a specific year (or bin).
2. This is the empirical **Probability Mass Function (PMF)** for the debut year.

### A bar chart showing the frequency distribution the Venus Sign Influence.



#### Key Insight: Dominance of "Soft" and "Structured" Venus

The distribution is heavily skewed, revealing a strong preference or clustering toward certain **Venus sign classifications at the time of group debuts**.

### Most Dominant Influences (The Peak):

The highest frequencies are found in classifications often associated with connection, emotion, and structure:

- a) Emotional (16 groups)
- b) Dreamy (15 groups)
- c) Disciplined (15 groups)
- d) Communicative (14 groups)

These four categories alone account for **60 groups**, or **exactly half of the dataset**, suggesting that K-pop debuts often occur under Venus transits astrologically linked to **intense fan connection, aesthetic vision, and professional structure**.

### Least Common Influences (The Trough):

The least frequent classifications are:

- a) **Analytical (5 groups)**
- b) **Adventurous (5 groups)**

This suggests that debuts rarely happen when Venus is in a sign characterized by detachment, pure intellect, or restless exploration.

---



### Statistical View and the Formula

### Statistical View (Frequency Distribution)

This chart is a visualization of the **frequency distribution** for a **nominal categorical variable** ('Venus Influence') derived from the astrological sign.

#### 1) Data Type:

- a) The x-axis (Venus Influence) is **Nominal Categorical**. The y-axis (Number of Groups) is **Discrete Numerical** (the count).

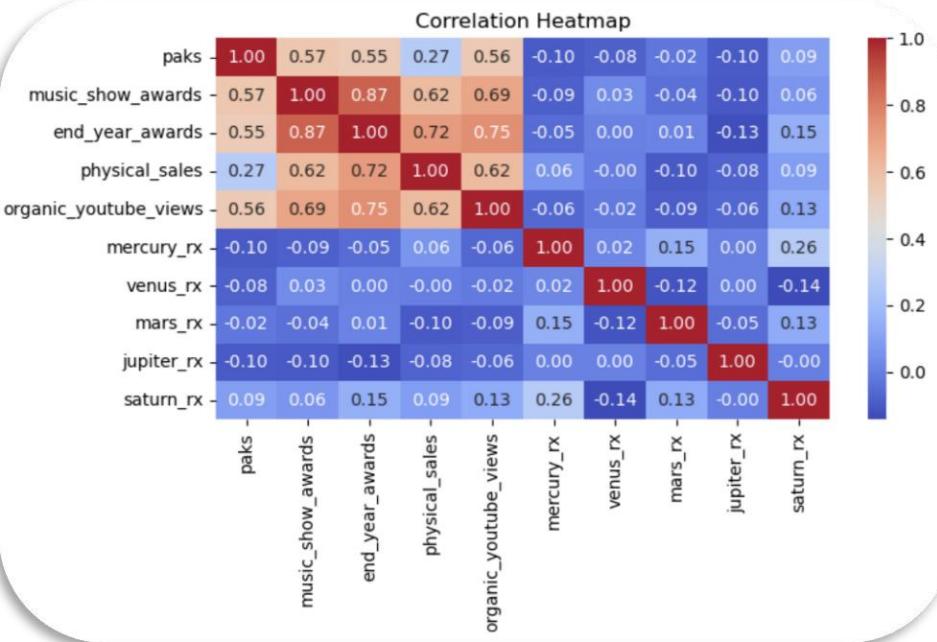
#### 2) Central Tendency:

- a) The distribution is **multimodal** and highly **positively skewed** (right-skewed). While the highest peak is at 16, the mean frequency is 10 groups per category  $120/12$ . The distribution concentrates significantly above this mean in the top four categories.

#### 3) Interpretation:

- a) The bar heights represent the **absolute frequency** ( $f$ ) of K-pop groups falling into each of the  $k=12$  influence categories. This allows for a direct visual assessment of which "Venus vibes" are most common at the time of a group's debut.

### A Correlation Heatmap for the retrograde (rx) status and debut astrology



 **Key Insight: Performance Metrics are Highly Intercorrelated, but Unrelated to Retrogrades**

The heatmap clearly separates the variables into two distinct groups based on their correlation patterns:

1. **Performance Metrics (High Positive Correlation):** Success in one performance area strongly predicts success in others.
  - a. **Music Show Awards** and **End of Year Awards** show the highest correlation ( $r = 0.87$ ), which is logical as end-of-year awards are often influenced by weekly show wins.
  - b. **Organic YouTube Views** and **Music Show Awards** also have a strong positive correlation ( $r = 0.69$ ), highlighting the link between popularity, exposure, and industry recognition.
  - c. **PAKs** (Perfect All-Kills) and **Organic YouTube Views** ( $r = 0.56$ ) are less strongly linked, suggesting digital chart success relies on more than just total views.
2. **Astrological Retrogrades (No Significant Correlation with Success):**
  - a. The correlations between any of **the performance metrics (rows 1-5)** and any of **the retrograde indicators (columns 6-9)** are extremely weak, ranging from -0.13 to 0.15.
  - b. **Conclusion:** Based on this dataset, a K-pop group's debut success (measured by PAKs, awards, sales, or views) is **not statistically correlated** with whether the major planets (Mercury, Venus, Mars, Jupiter) were in retrograde motion on their debut date.

## Statistical View and the Formula

### Statistical View (Correlation Matrix)

The heatmap is a visual representation of a **Correlation Matrix (R)**, where the variables are correlated against each other.

- ✓ **Variable Types:** The matrix compares five **Ratio/Scale** variables (performance metrics) with four **Binary/Dichotomous** variables (retrograde status: 0 or 1).
  - ✓ **Measurement:** The numbers in the cells are the **Pearson Correlation Coefficient (r)**.
  - ✓ **Interpretation of r:**
    - a.  $r = 1.00$  (dark red): Perfect positive correlation (variables move together).
    - b.  $r = -1.00$  (dark blue): Perfect negative correlation (variables move in opposite directions).
    - c.  $r \approx 0.00$  (white/light blue): No linear relationship
- 

## Conclusion

This report analyzes the K-Pop Group Debut Astrology dataset, with a primary focus on the **Venus signs** that dominate when new K-Pop groups debut.

Since Venus represents popularity and artistic appeal, this study seeks to explore how its astrological placement correlates with debut patterns, zodiac sign distributions, and potential cultural trends emerging within the industry.

This report analyzes the k Pop Group Debut Astrology dataset, focusing on Venus signs as the key factor influencing K-pop group debuts. Since Venus represents popularity and artistic appeal, the study explores how its astrological positioning aligns with debut patterns, zodiac distributions, and potential cultural trends within the industry.

---

## Inherent Qualities of the Zodiac Signs

No.	Quality of Air Sign	Quality of Fire Sign	Quality of Water Sign	Quality of Earth Sign
1	Intellectual	Passionate	Emotional	Practical
2	Communicative	Energetic	Intuitive	Grounded

3	Social	Bold	Compassionate	Reliable
4	Curious	Confident	Sensitive	Patient
5	Objective	Enthusiastic	Empathetic	Disciplined
6	Innovative	Courageous	Imaginative	Responsible
7	Detached	Charismatic	Mysterious	Hardworking
8	Adaptable	Adventurous	Nurturing	Stable
9	Idealistic	Assertive	Deep	Sensible
10	Creative	Inspirational	Romantic	Loyal

---



---

References Link -

<https://www.kaggle.com/datasets/carolinacanchila/k-pop-astrology-debut-charts-and-success>

<https://www.astrology.com/article/what-does-retrograde-mean/>

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.describe.html>

<https://ssd.jpl.nasa.gov/horizons/>

---



---

**Warm and Enthusiastic (Good for a Presentation or Project Report)**

A big thank you to our amazing **teacher**, Mr Kyaw Kyaw Htut, whose support and expertise made this project possible!

We also want to acknowledge the hard work and teamwork of our **active group members**—it's your effort that brought this analysis to life. 🙏

---



---