Kpop Idols Visual Analysis

Crystal Chen CSE412 AA

Data

This dataset includes information about Kpop idols before 2023. It contains details such as Idols Stage name, Korean Name, Group Name, height, Weight, Place and date of Birth etc.

I decided to select this dataset because this dataset provides a comprehensive information about Kpop idols, allowing me to analyze trends in demographics and group distribution. This dataset is useful for common characteristics such as names and birth dates as well as geographic distribution of their birth place.

Link to the data source
https://drive.google.com/file/d/1LxxT95gUIglPEj2ACBfyas7Zq6UjBYNc/view?usp=sharing

This dataset contains 13 variables, including stage name, full name, Korean name, Stage Name of idol in korean, group name, gender, birthplace, and country of origin shown as strings. Then date of birth and debut date are identified as date in the format of mm/dd/yyyy. Number types are presented in attributes of height, weight, and instagram account.

Data includes both male and female kpop idols, and it includes kpop group idols, solo idols, as well as people who were one of the group before but left the group after.

There exists some missing values for some attributes. Many blocks under Instagram and birthplace are blank, declaring a lack of information collected. Group name of solo idols are blank, meaning they are not in any groups. Zeros under Height and Weight attributes might lead to some calculation errors where I need to fix.

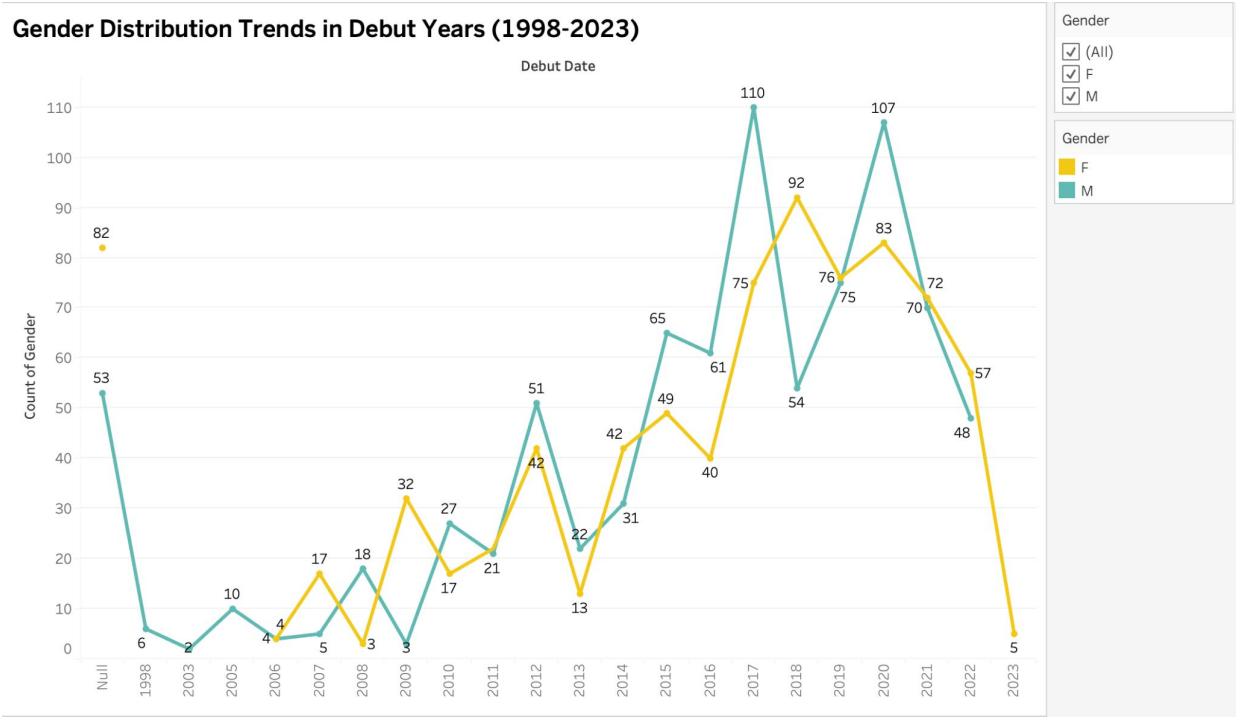
The date of birth for idols in the same group are evenly distributed and relatively in the same range. Some idols are not assigned to any groups, indicating the current status compared to the kpop group idols. While the main presence of Kpop idols are in South Korea, there are still other 10% distribution in other countries, conveying the globalization of Kpop culture.

Questions

1. What is the gender distribution among Kpop idols in the same debut years?
2. How are K-pop idols geographically distributed by birth countries?
3. What are the age ranges of Kpop idols based on their birth years?

Visualization 1:

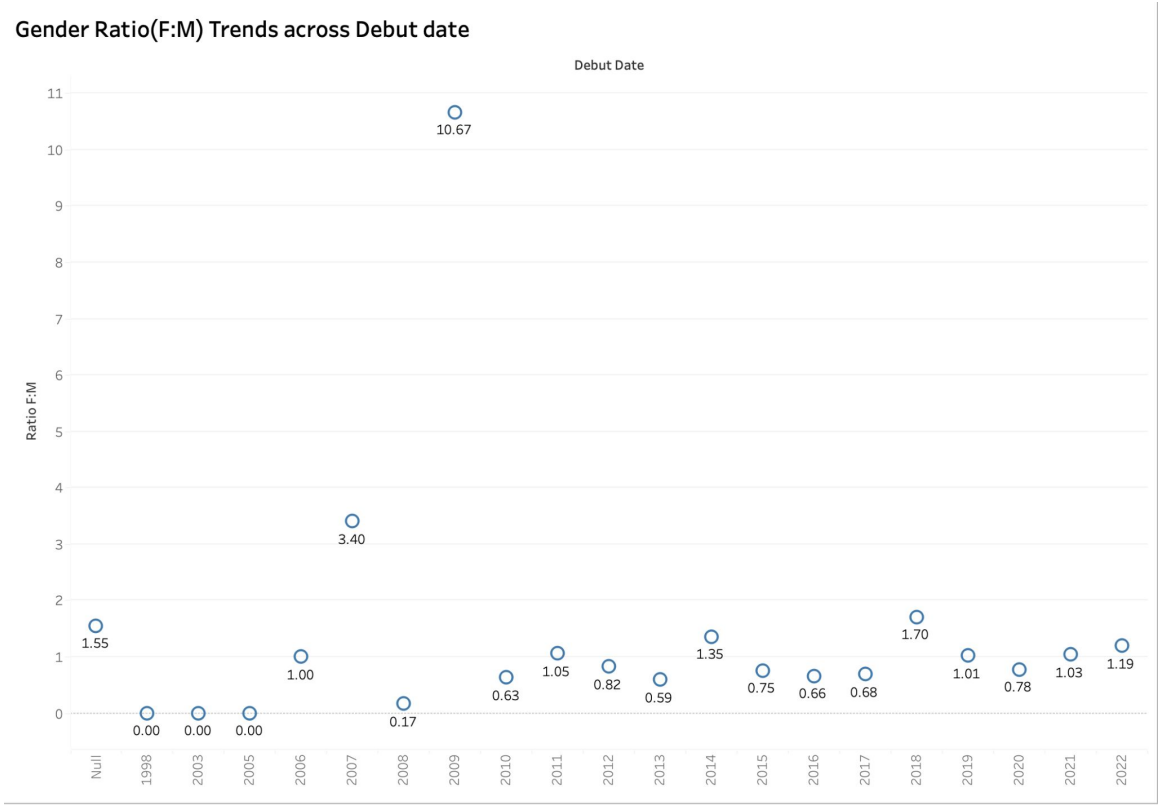
This is a line plot including gender distribution over debut years. The number of female and male over each debut year is fluctuated and unpredictable. Number of 53 and 82 show above Null column, where some kpop idols don't appear to have debut date in the data set. This line plots clearly showed the change of number of gender for each debut year.
This plot intends to answer the first question about gender distribution across debut years.



Visualization 2:

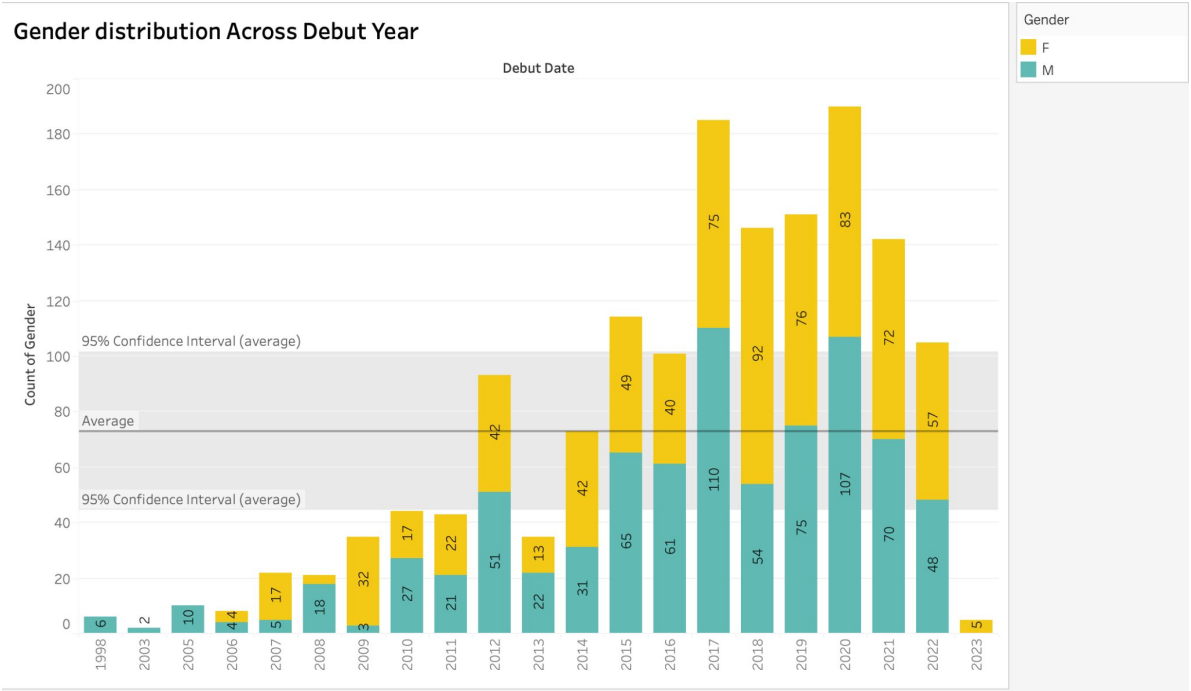
This visualization shows gender ratio (Female:Male) across debut date, where ratio > 1 meaning Kpop girl idols are more than Kpop boy idols active from that debut year, Kpop boy idols are fewer when ratio < 1, and number of them are the same when ratio = 1. This plot gives an direct view of gender distribution through decimal ratio.

It solves the problem of gender distribution over debut years.



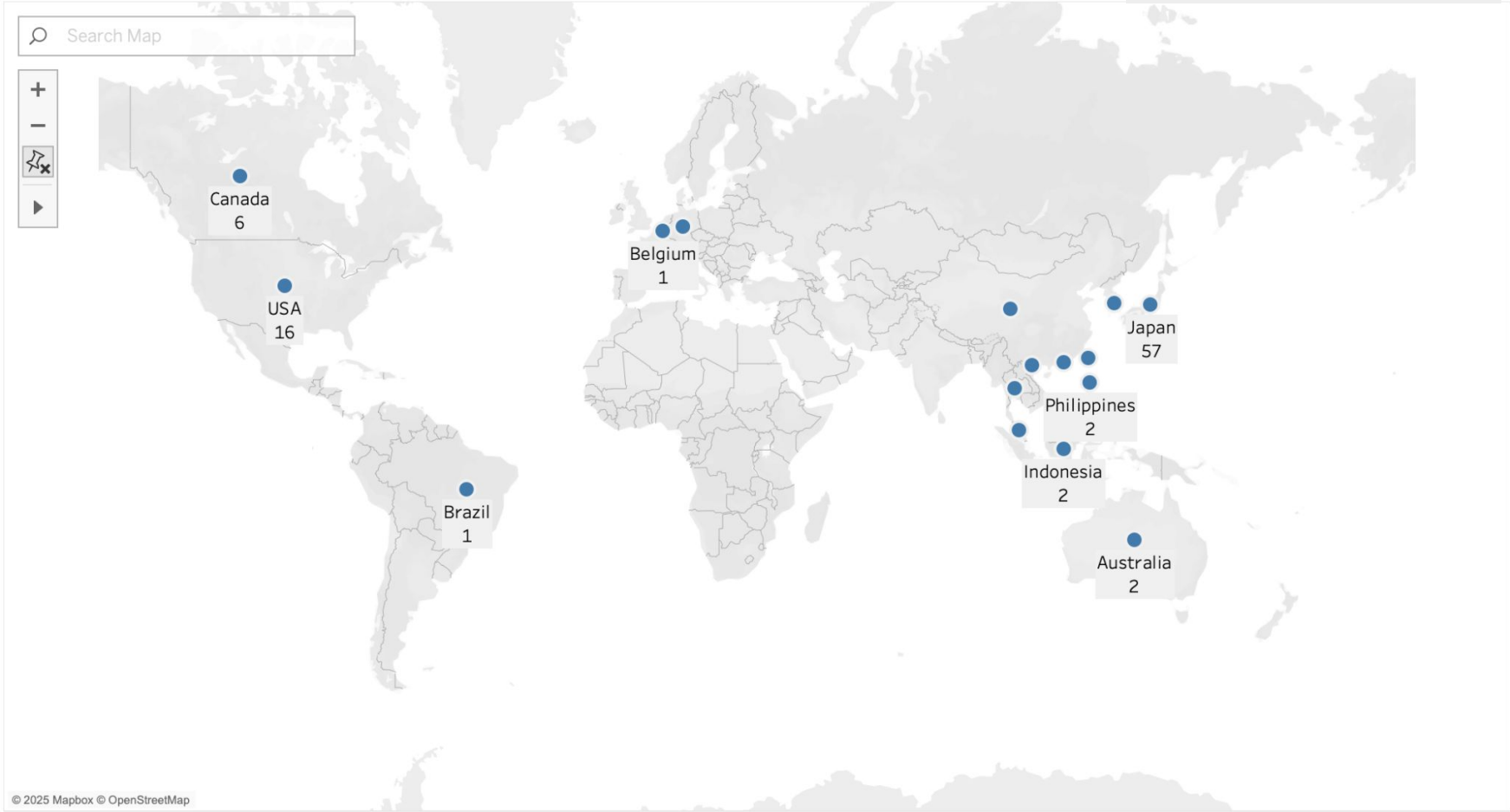
Visualization 3:

This is a bar plot but each bar separates in two colors where yellow presents the number of Female being active from debut year and number of male for that debut year are colored in blue. The length of bar corresponds to the amount of idols in different genders. This bar plot provides comparison between count of gender over debut years with color and length encoding. It also answers the first question about distribution of gender across debut year.



Visualization 4

Geographic distribution of birth

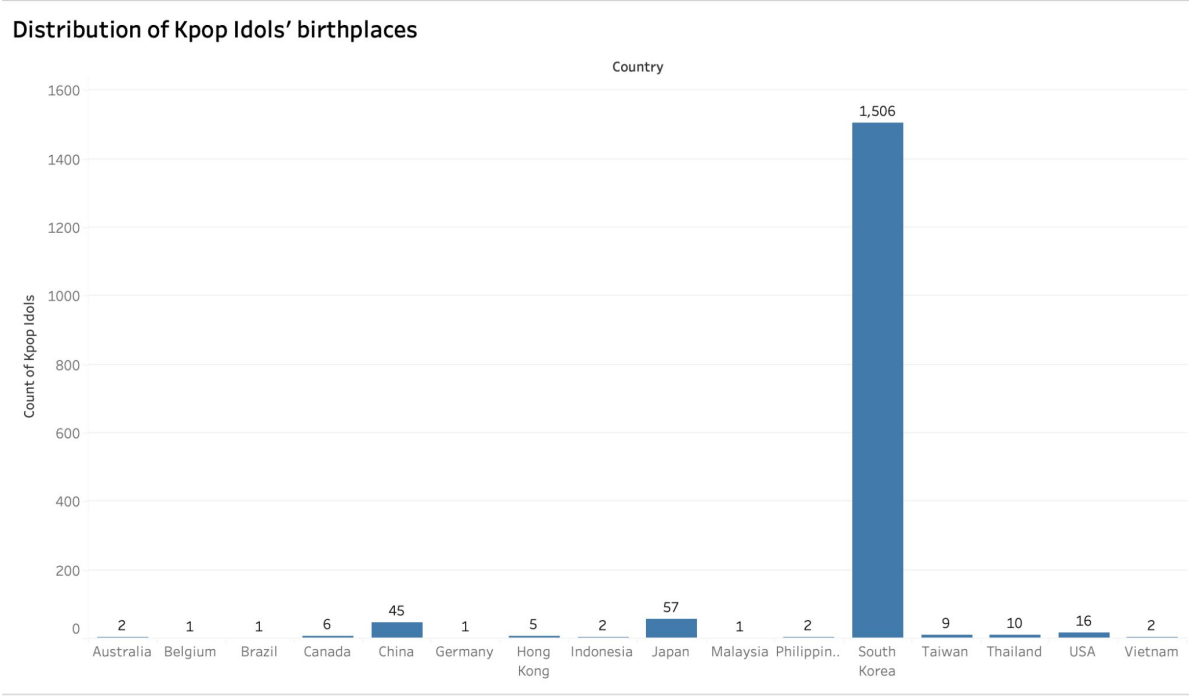


Visualization 4:

This is a geographic map showing kpop idols' birthplace with numbers of kpop idols born in those countries. It implies that kpop groups grow to be increasingly globalization and international from all over the world even though 90% of kpop idols are from South Korea. Idols originate from diverse regions beyond South Korea, countries such as Canada 6, USA 16, Japan 57 etc. The map answers the second question about geographical distribution of kpop idols by their birth places.

Visualization 5:

Distribution of kpop idols' birth place with respect to the counts of idols. The highest bar with 1,506 of kpop idols from Korea indicates the majority of kpop idols were born in South Korea. However, 45 people in China and 57 people in Japan also highlights the international influence from Kpop culture. This addresses the 2nd question about geographic distribution on birthplaces.



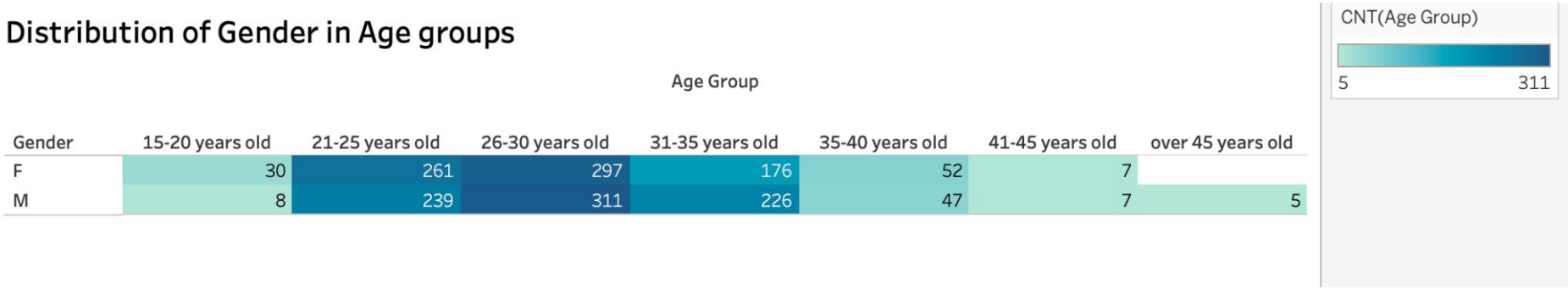
Visualization 6:

Stage Name with Birthplace

Stage Name	Country	
2Soul	South Korea	Abc
A-min	South Korea	Abc
A.M	South Korea	Abc
Ace	South Korea	Abc
Aeji	South Korea	Abc
AhIn	South Korea	Abc
Ahra	South Korea	Abc
Ahyoon	South Korea	Abc
Ahyoung	South Korea	Abc
Ahyung	South Korea	Abc
Ailee	South Korea	Abc
Aini	South Korea	Abc
Aisha	South Korea	Abc
Aki	South Korea	Abc
Alex	Germany	Abc
AleXa	USA	Abc
Alice	South Korea	Abc
Allen	USA	Abc
Amber	USA	Abc
Andy	Hong Kong	Abc
	South Korea	Abc
Anne	South Korea	Abc
Anthony	China	Abc
Ara	South Korea	Abc
Arang	South Korea	Abc
Areum	South Korea	Abc

This vertical long table presents the stage names of K-pop idols alongside their birth countries. Two attributes are string types where one is stage name and the other is birth country. The table provide clear visual cues for navigating through the data, and it can be sorted as desired. This addresses the second question.

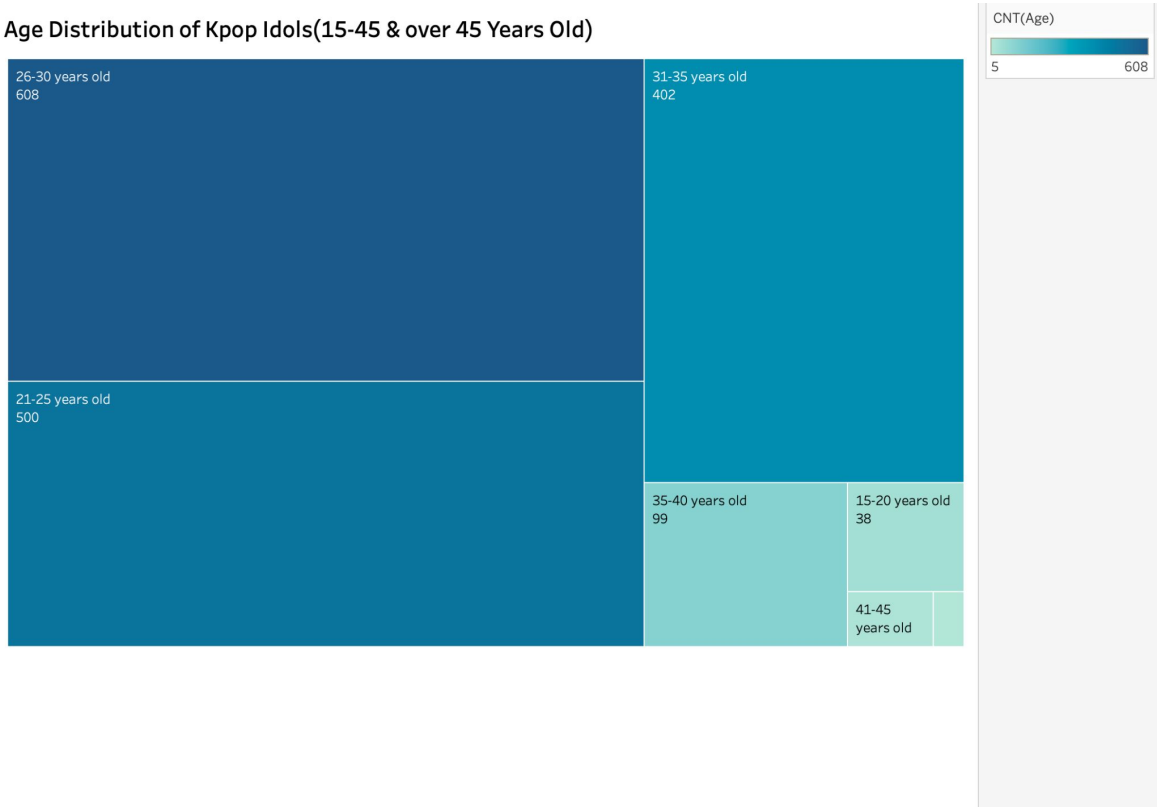
Visualization 7



The heat map illustrates the distribution of gender across different age groups in the dataset. The majority of individuals fall within the 21-30 age range, with a noticeable peak in both the 26-30 age group for males (311) and females (297). The color of heatmap reinforces the trends, where darker blocks indicate higher counts. This visualization addresses the first and third question about gender amd age group distribution.

Visualization 8:

This is a tree map of age distribution of Kpop idols five years in a group where the majority of kpop idols are in the age of 26 - 30 from the original dataset. The darker the block is, the more counts of kpop idols are in these age ranges. The size of each segment also corresponds to the number of people. This treemap addresses the last question about age ranges.



Learning Reflection

For this assignment, the thing I struggled with the most was to handle and deal with the data in my dataset, where I wanted to add one more column about kpop idols' debut date. And when I check my dataset, there exist some blank blocks where I need to pay attention to as I made the visualization. I really loved the time when the plot appeared with expected legend and form, which inspired me to make deeper analysis on tableau. The design of tableau is intelligent that it provides generated data column, including geographical longitude and latitude. I learned how to make a desired data by adding a calculated function in the panel. And methods of showing different dataset are presented by different visualization type to give a clear understanding of the relationship between two or more attributes.