

Project Name: Ups and Downs with Kardashians & Jenners

The Sisters and Google Trends

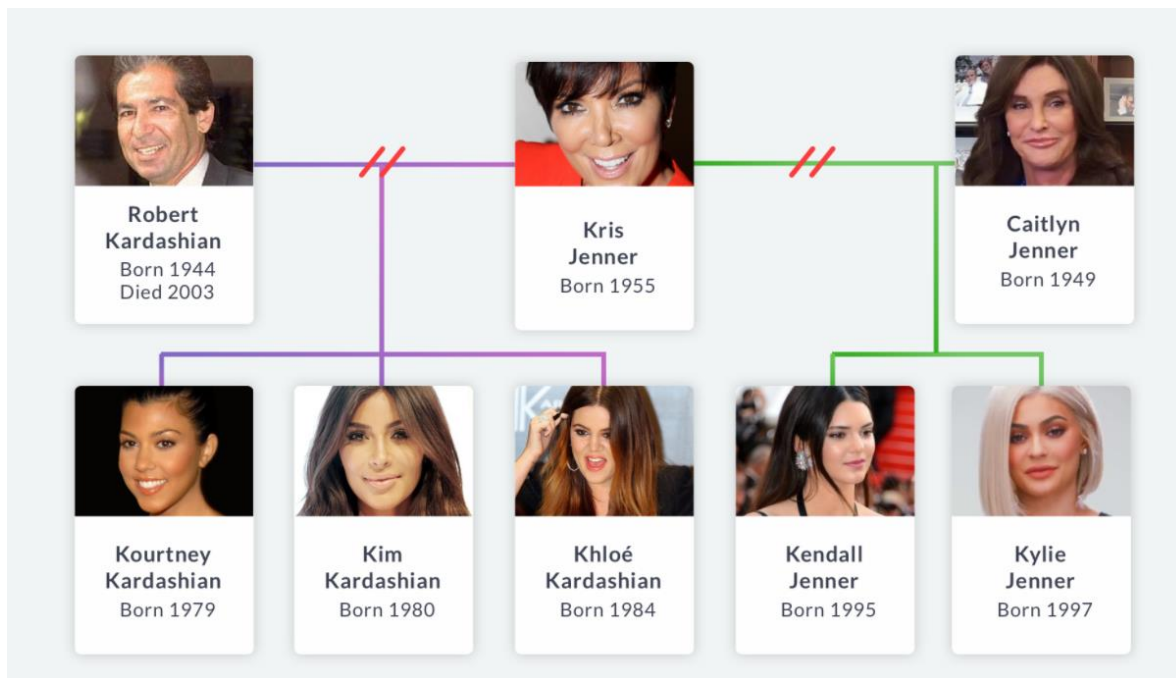
While I'm not a fan nor a hater of the **Kardashians and Jenners**, the polarizing family intrigues me. Why? Their marketing prowess. Say what you will about them and what they stand for, they are great at the hype game. Everything they touch turns to content.

The sisters in particular over the past decade have been especially productive in this regard. Let's get some facts straight. I consider the "sisters" to be the following daughters of **Kris Jenner**. Three from her first marriage to lawyer **Robert Kardashian**:

Kourtney Kardashian (daughter of Robert Kardashian, born in 1979) **Kim Kardashian** (daughter of Robert Kardashian, born in 1980) **Khloé Kardashian** (daughter of Robert Kardashian, born in 1984) And two from her second marriage to Olympic gold medal-winning decathlete, Caitlyn Jenner (formerly Bruce):

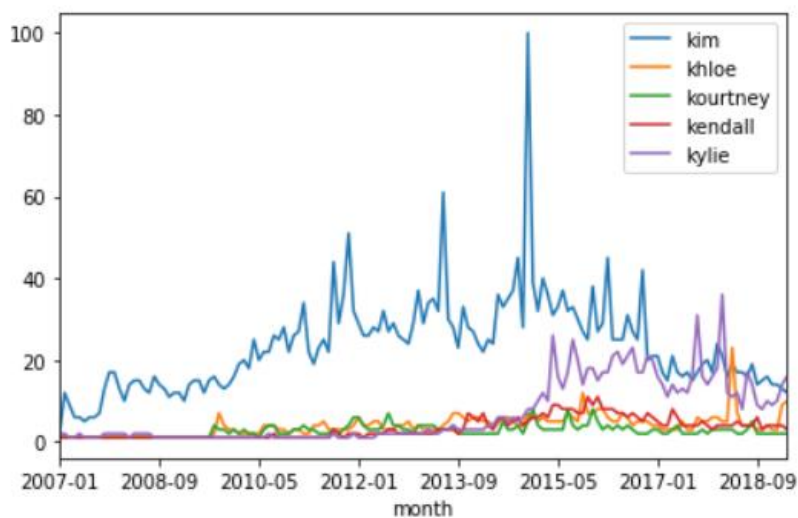
Kendall Jenner (daughter of Caitlyn Jenner, born in 1995) **Kylie Jenner** (daughter of Caitlyn Jenner, born in 1997)

This family tree can be confusing, but we aren't here to explain it. We're here to explore the data underneath the hype, and we'll do it using search interest data from Google Trends. We'll recreate the Google Trends plot to visualize their ups and downs over time, then make a few custom plots of our own. And we'll answer the big question: is Kim even the most famous sister anymore?

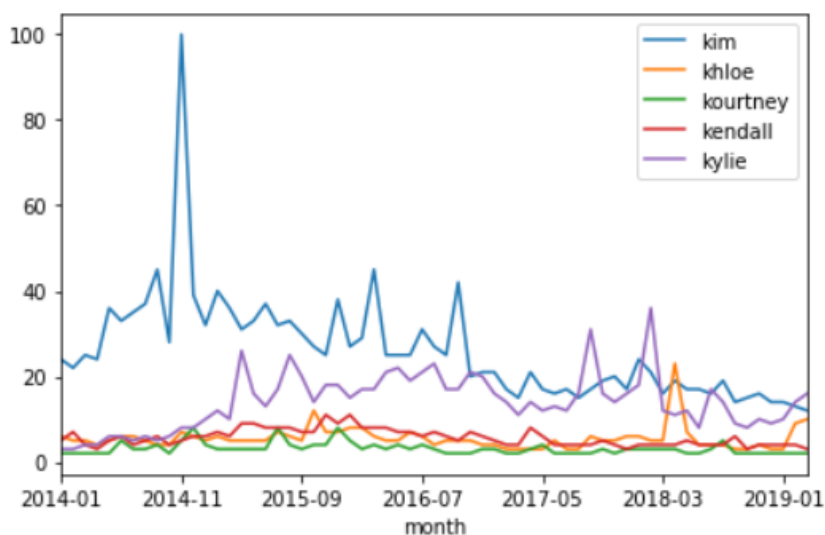


The Early Kim's Hype

- There is so much to make sense of here. Kim's sharp rise in 2007, with the beginning of *Keeping Up with the Kardashians*, among other things. There was no significant search interest for the other four sisters until mid-2009 when Kourtney and Khloé launched the reality television series, *Kourtney and Khloé Take Miami*. Then there was Kim's rise from famous to literally more famous than God in 2011. This *Cosmopolitan* article covers the timeline that includes the launch of music videos, fragrances, iPhone and Android games, another television series, joining Instagram, and more. Then there was Kim's ridiculous spike in December 2014: posing naked on the cover of *Paper Magazine* in a bid to break the internet

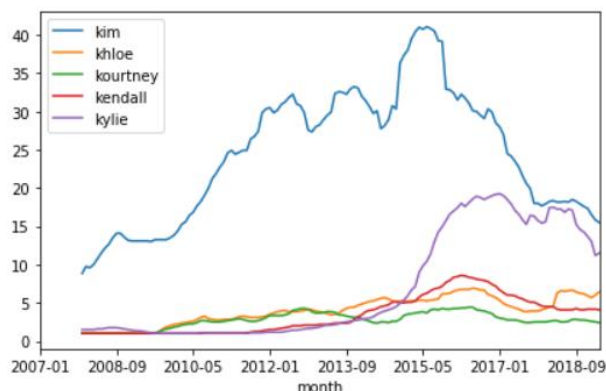


Kylie's Rise

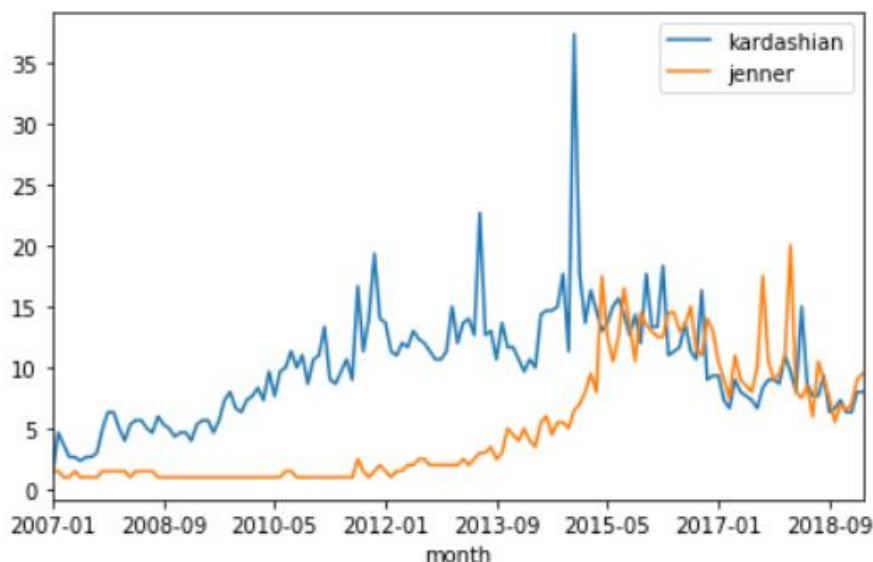


Fluctuation of Sisters

- Whoa, okay! So, by this metric, Kim is still the most famous sister despite Kylie being close and nearly taking her crown. Honestly, the biggest takeaway from this whole exercise might be Kendall not showing up that much. It makes sense, though, despite her wildly successful modeling career. Some have called her "the only normal one in her family" as she tends to shy away from the more dramatic and controversial parts of the media limelight that generate oh so many clicks.



The Kardashians or The Jenners



Steps:

1. Let's load, read and inspect our Google Trends data, which was downloaded in CSV form. The [query](#) parameters: each of the sisters, worldwide search data, 2007 to present day. (2007 was the year Kim became "active" according to Wikipedia.)
2. So, we have a column for each month since January 2007 and a column for the worldwide search interest for each of the sisters each month. Let's make the column names first easier to work with. A column named "Kim Kardashian: (Worldwide)" is not the most usable for coding purposes. Let's shorten those so we can access their values better. Might as well standardize all column formats, too. I like lowercase, short column names.
3. That's better. We don't need to scroll our eyes across the table to read the values anymore since it is much less wide. And seeing five columns that all start with the letter "k" ... the aesthetics ... we should call them "kolumns" now! (Bad joke.)
4. The next thing I see that is going to be an issue is that "<" sign. If *"a score of 0 means there was not enough data for this term,"* "<1" must mean it is between 0 and 1 and Google does not want to give us the fraction from google.trends.com for whatever reason. That's fine, but this "<" sign means we won't be able to analyze or visualize our data right away because those column values aren't going to be represented as numbers in our data structure. Let's confirm that by inspecting our data types.
5. the khloe, kourtney, and kendall columns aren't integers like the kim and kylie columns are. Again, because of the "<" sign that indicates a search interest value between zero and one. Is this an early hint at the hierarchy of sister popularity? We'll see shortly. Before that, we'll need to remove that pesky "<" sign. Then we can change the type of those columns to integer.
6. Now let's convert our month column from type object to datetime to make our date data more accessible.
7. Set the month column as our index to wrap our data cleaning. Having month as index rather than the zero-based row numbers will allow us to write shorter lines of code to create plots, where month will represent our x-axis.
8. Because we cleaned our data, now we can visualize it by making plots.

9. Oh my! There is so much to make sense of here. Kim's sharp rise in 2007, with the beginning of Keeping Up with the Kardashians, among other things. There was no significant search interest for the other four sisters until mid-2009 when Kourtney and Khloé launched the reality television series, Kourtney and Khloé Take Miami. Then there was Kim's rise from famous to literally more famous than God in 2011. This Cosmopolitan article covers the timeline that includes the launch of music videos, fragrances, iPhone and Android games, another television series, joining Instagram, and more. Then there was Kim's ridiculous spike in December 2014: posing naked on the cover of Paper Magazine in a bid to break the internet will do that for you.
10. Smooth out the fluctuations with rolling means. Since late-2016, at various months, Kylie overtakes Kim. Two big spikes where she smashed Kim's search interest: in September 2017 when it was reported that Kylie was expecting her first child with rapper Travis Scott and in February 2018 when she gave birth to her daughter, Stormi Webster. The continued success of Kylie Cosmetics has kept her in the news, not to mention making her the "The Youngest Self-Made Billionaire Ever" according to Forbes. We can use rolling means to smooth out short-term fluctuations in time series data and highlight long-term trends.
11. Let's end this analysis with one last plot. In it, we will plot (pun!) the Kardashian sisters against the Jenner sisters
to see which family line is more popular now. We will use average search interest to make things fair, i.e., total search interest divided by the number of sisters in the family line.