STRV

Data Engineer Technical Challenge

What results did you achieve?

01 THE CHALLENGE
What was our briefing

02 THE WHOLE PICTURE
A look at all the data

03 THE LAST 50 YEARS
The active generations

04 TRENDS ARE CHANGING
The last 2 two decades in detail

05 THE 2010S ARE WEIRD
The last 5 years challenge the norm

06 CONCLUSIONS & NEXT STEPS

THE CHALLENGE

STRV

What was asked?

What were the requirements?

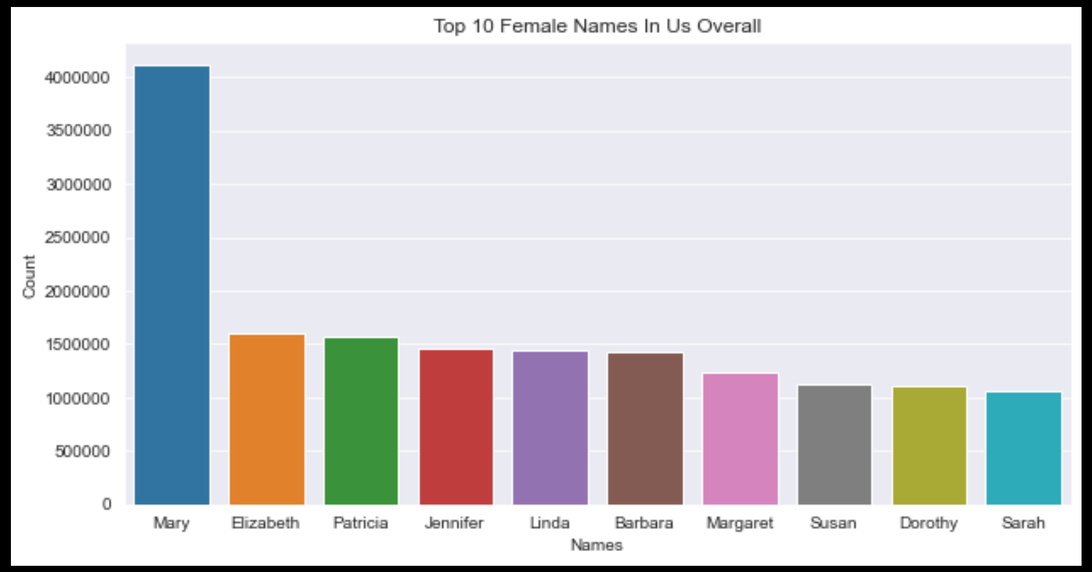
A newly launched fashion startup wants to help families choose names for their newborns — and then sell personalized clothing featuring those names. They haven’t defined their target audience, genders or any specifics yet.

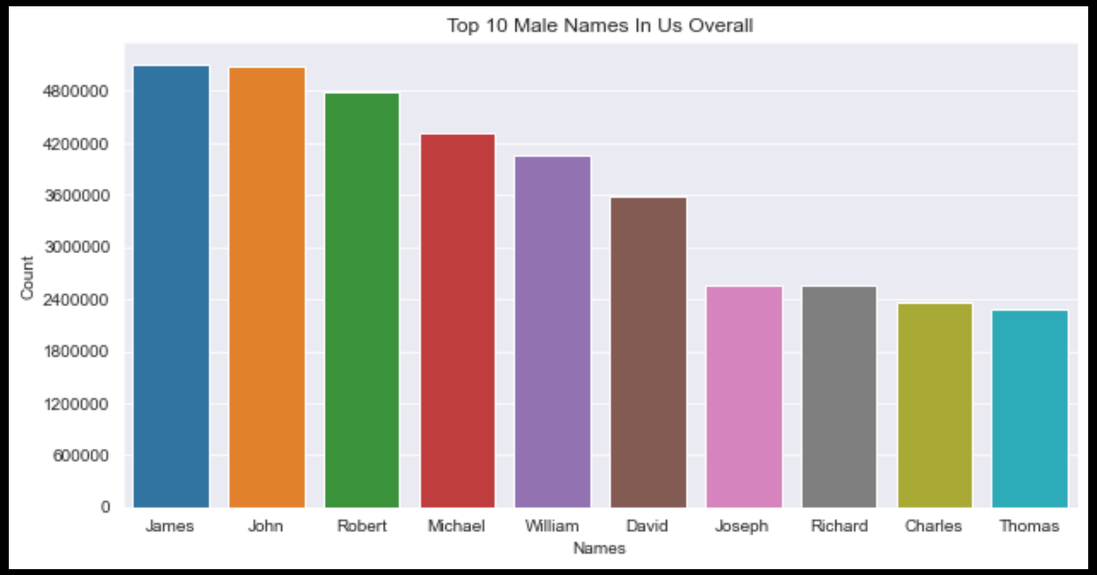
The key is to help parents make data-driven decisions when selecting a name while also supporting the startup's goal of selling customized products.

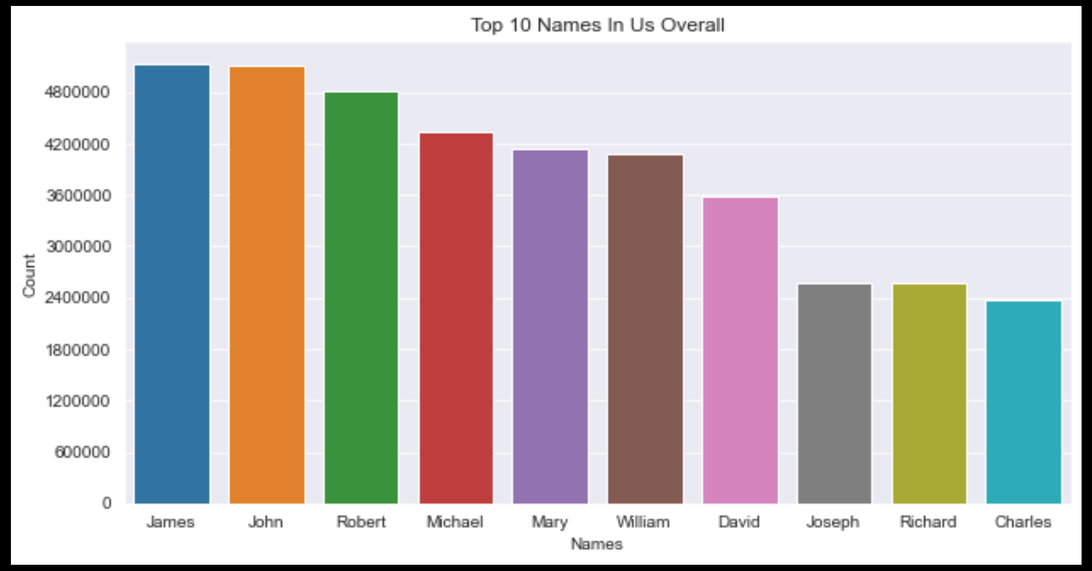
THE WHOLE PICTURE

STRV

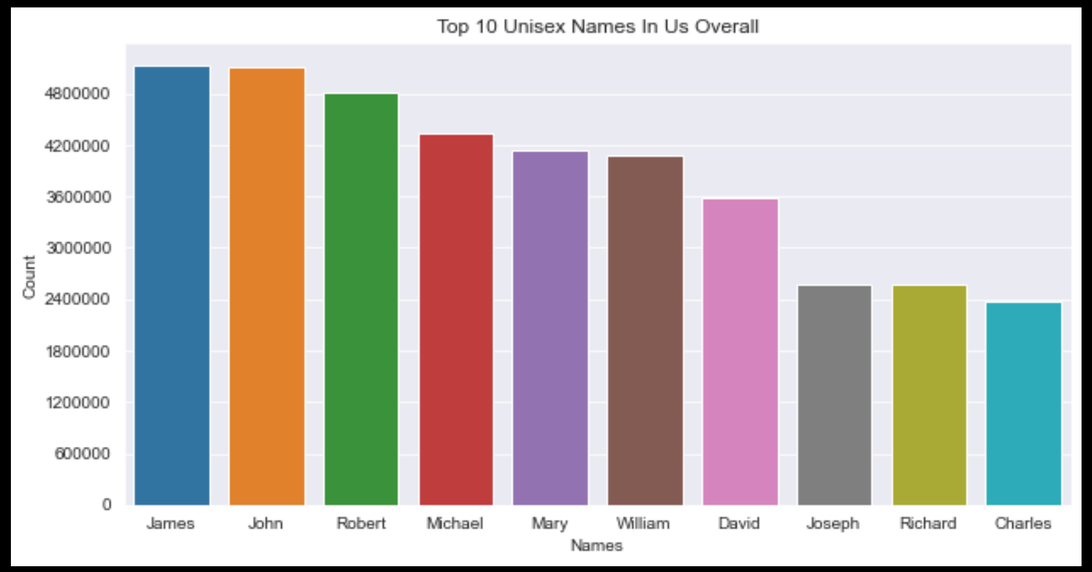
Examining the Data Available



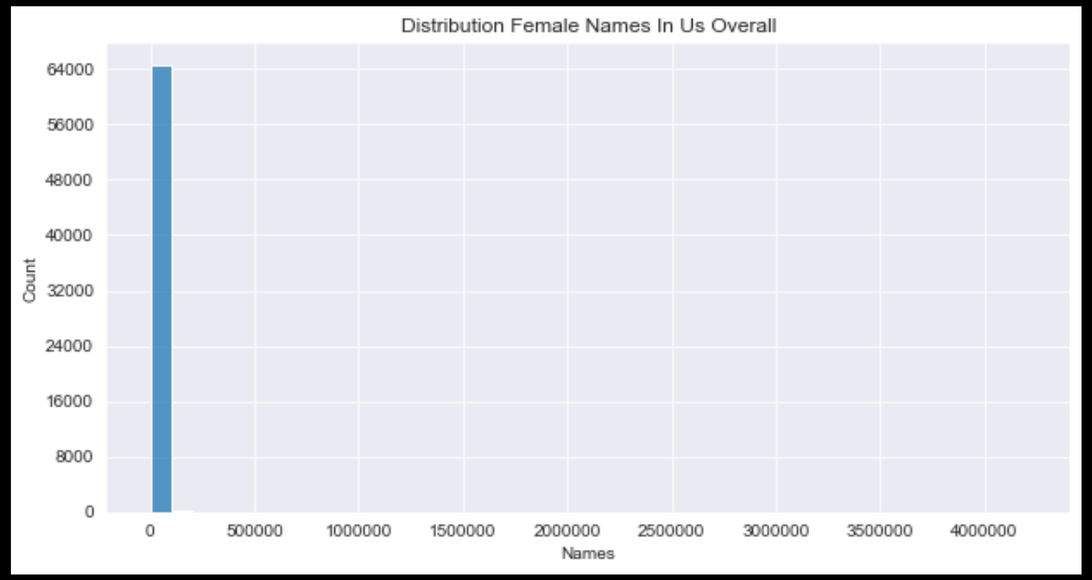


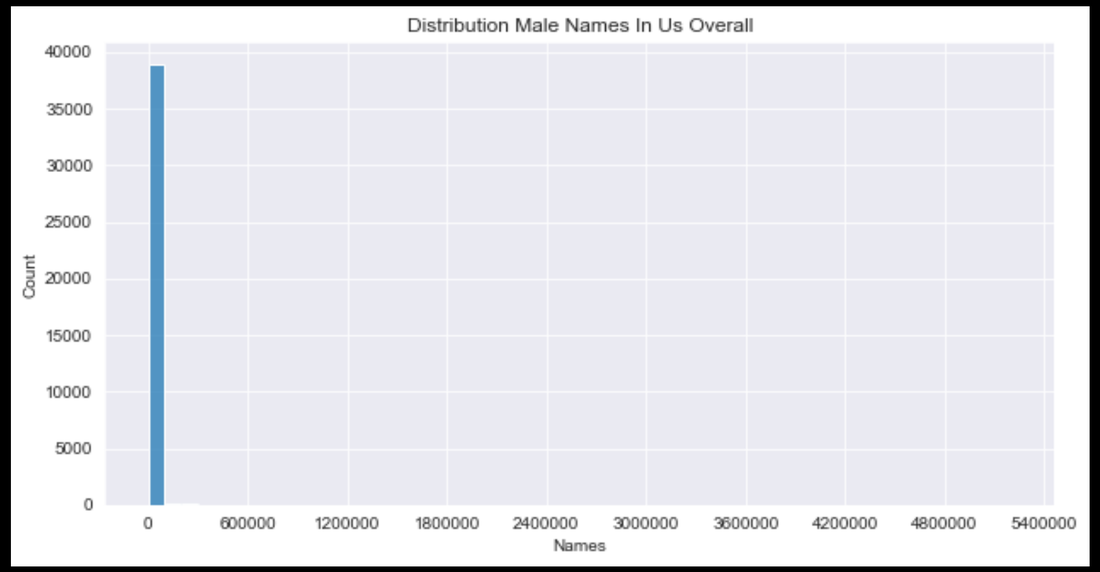


Amongst the top 10 names overall, 9 are typically male. The name Mary is extremely more common than any other Female name. This prompts us to look at its use spread across time! Male names have their counts much more similar, and we can see a gradual dropoff as opposed to a clear leader in occurrences and then a very gentle dropoff.



The top "Unisex" names and the top names are a match. Although these might actually be the most unisex names, by consequence of being the most common names, they're also the names most likely to have been incorrectly inputted in registry systems.



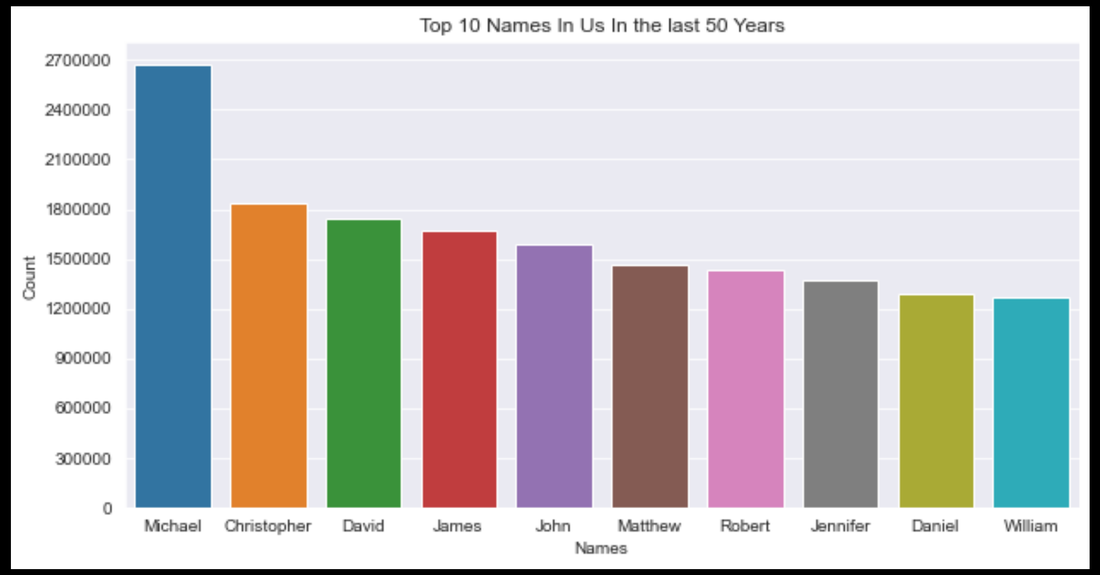


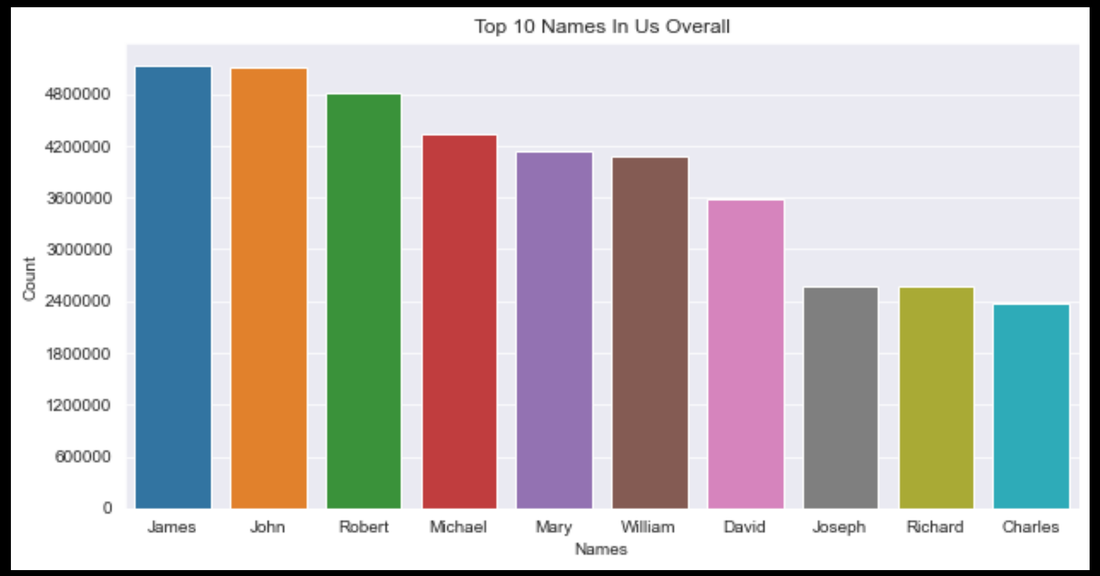
We have confirmed that, overall, Female names have significantly more variety than male names in the US. In both female and male names registered, the vast majority have between 0 and 100000 registers! As hypothesised, the more common male names have more occurrences, shown by the mean and median count of male names being higher, significantly so when it comes to the mean (2573.84 vs 4338.50).

THE LAST 50 YEARS

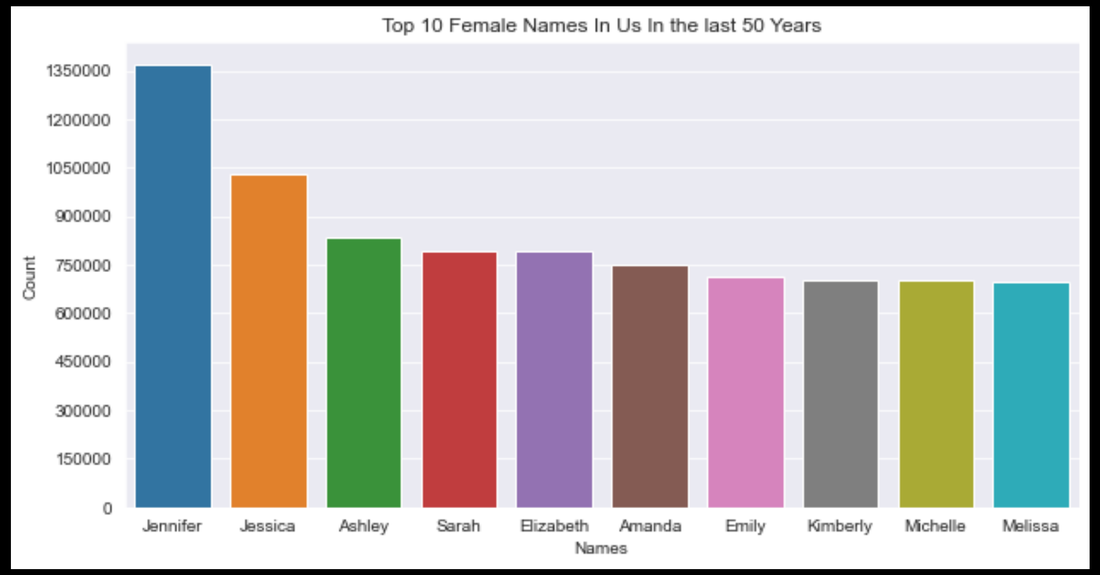
STRV

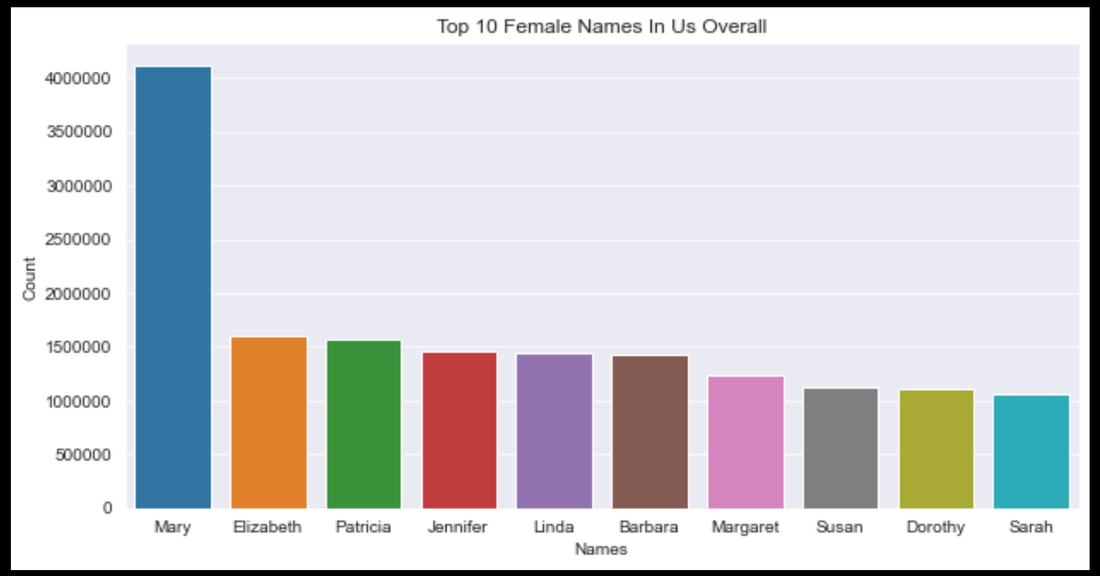
A more focussed prespective



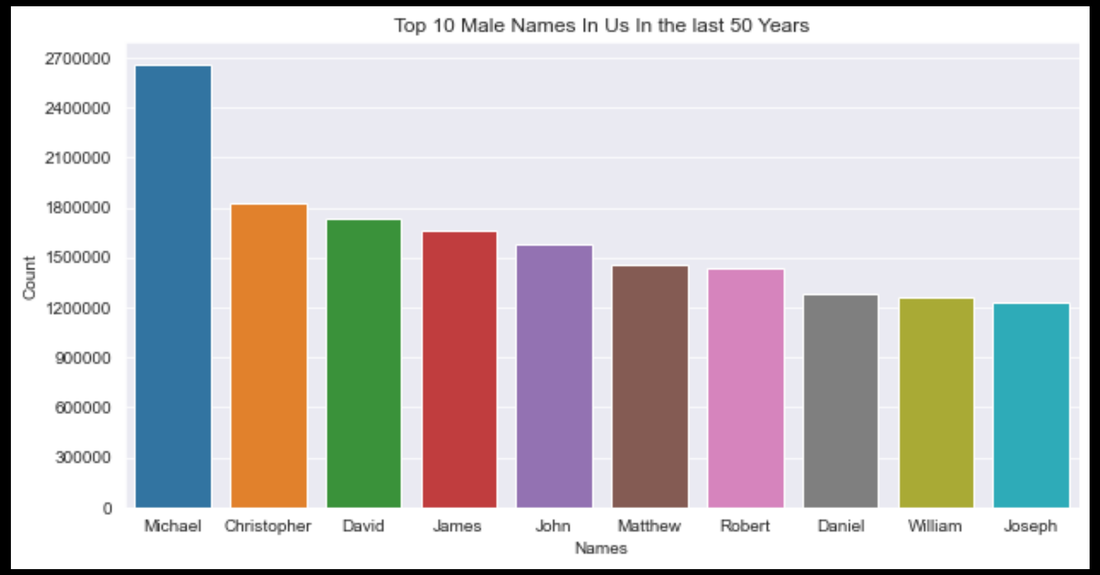


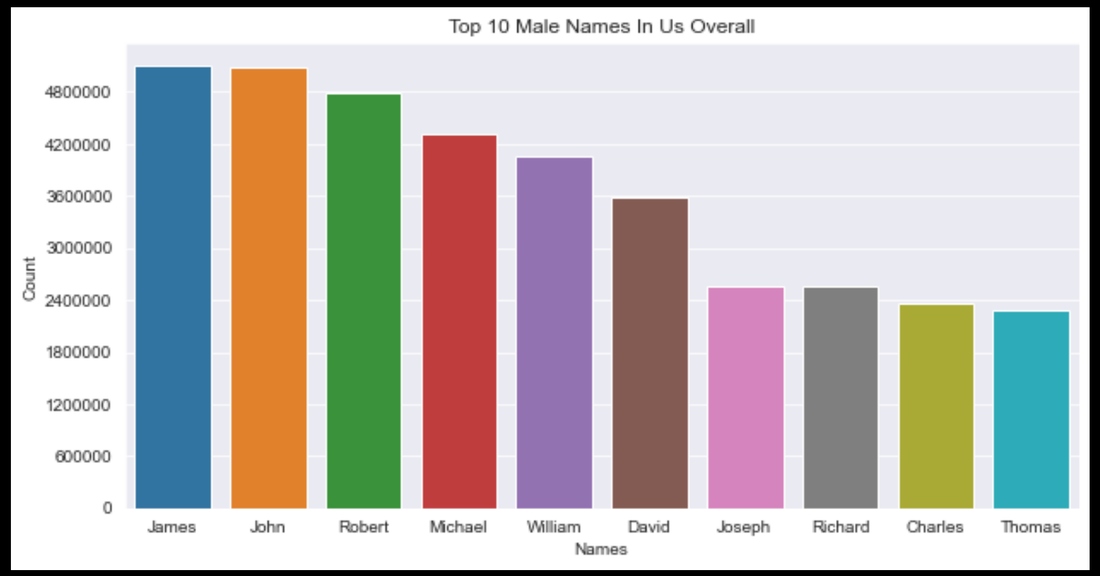
In comparison with the overall data, we can see that in the last 50 years, 6 of the top 10 remain the same (Michael, David, James, John, Robert and William) but none in the same rank. This indicates that, at least for male names, the more common names are consistent.



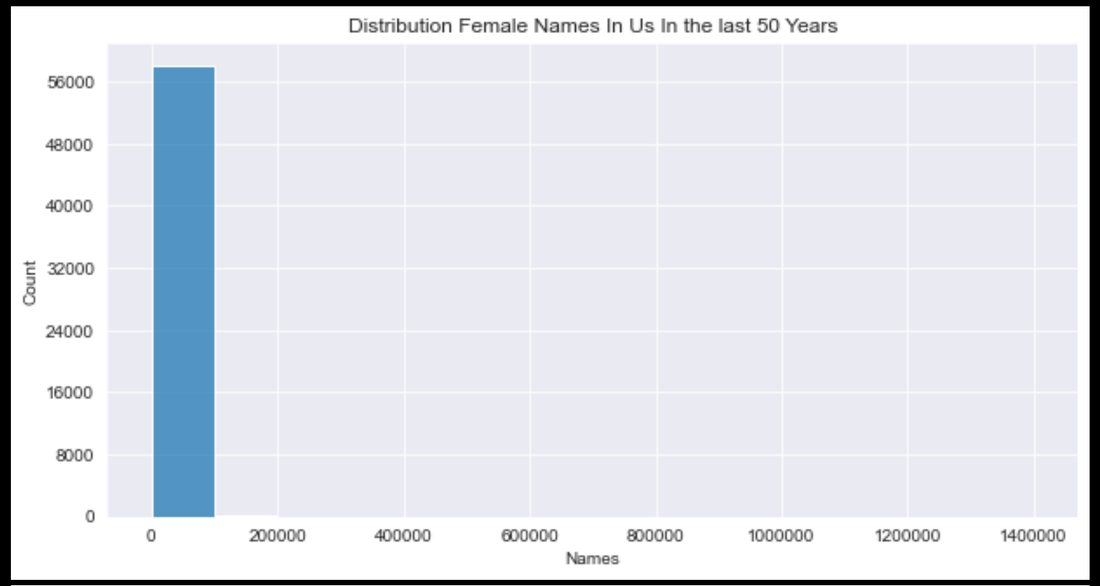


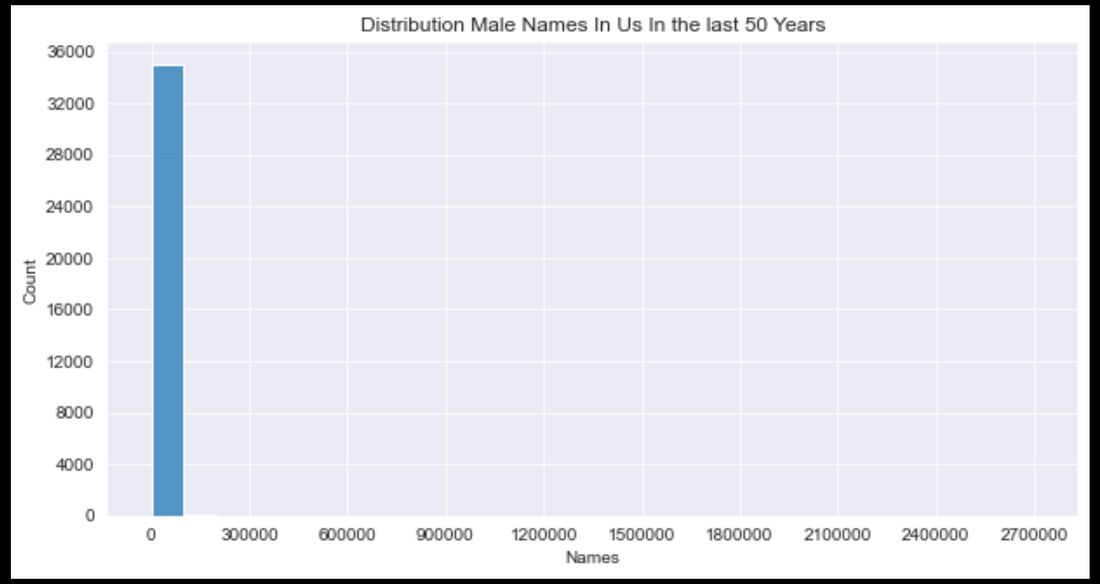
In comparison with the overall female data, we can see that in the last 50 years, only 3 of the top 10 remain the same (Jennifer, Sarah and Elizabeth). This indicates that, for female names, the more common names are significantly trendier than male names. This is also supported by the absence of the overwhelmingly dominant name in the overall data, Mary, in the top 10 of the more recent five decades of data.





In comparison with the overall Male data, we can see that in the last 50 years, 7 of the top 10 remain the same (Michael, David, James, John, Robert, William and Joseph) but none in the same rank. As mentioned before, this indicates that for male names, the more common names are quite consistent.



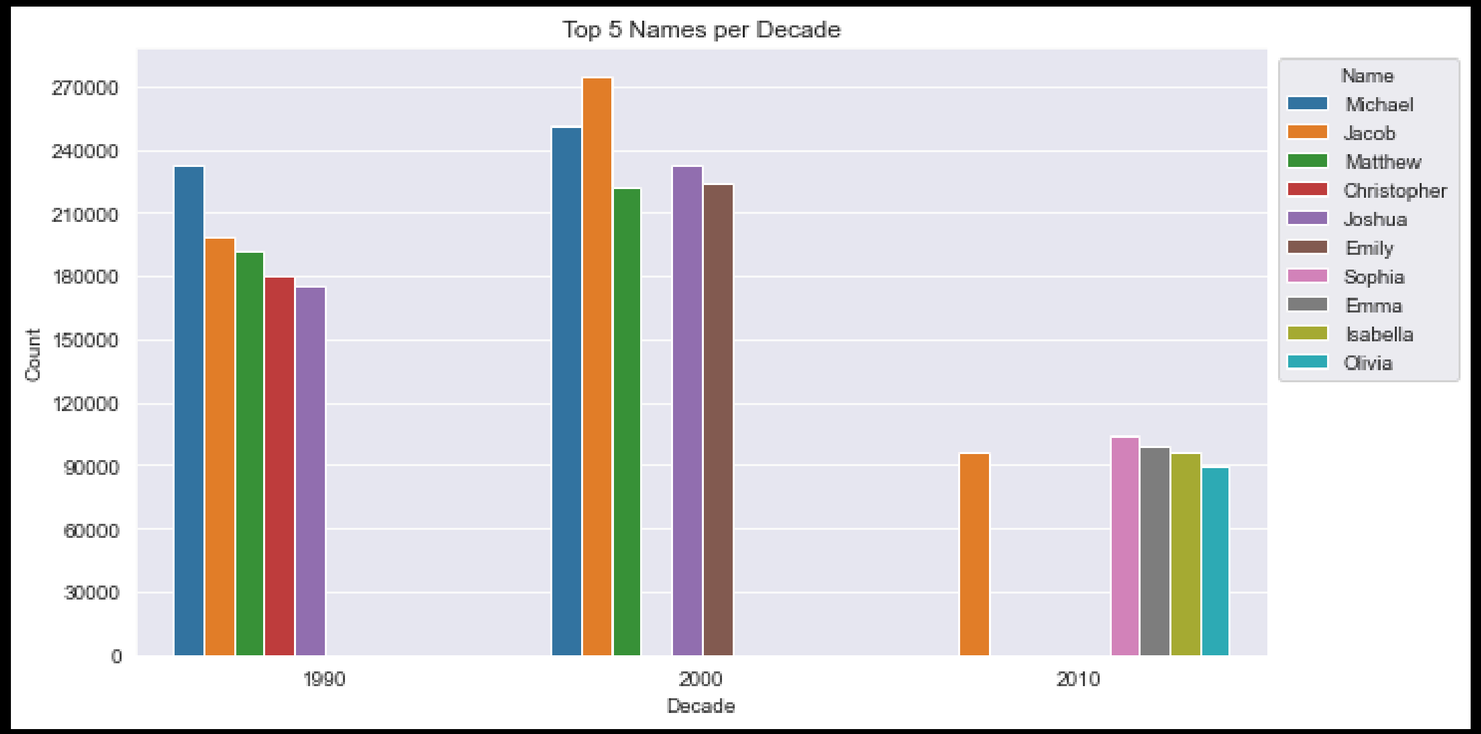


Looking at the distributions for the last 50 years, we can see that the observations made previously on the overall data remain accurate, although the absence of the acute reduction of registered Maries has made the statistics on both groups closer together

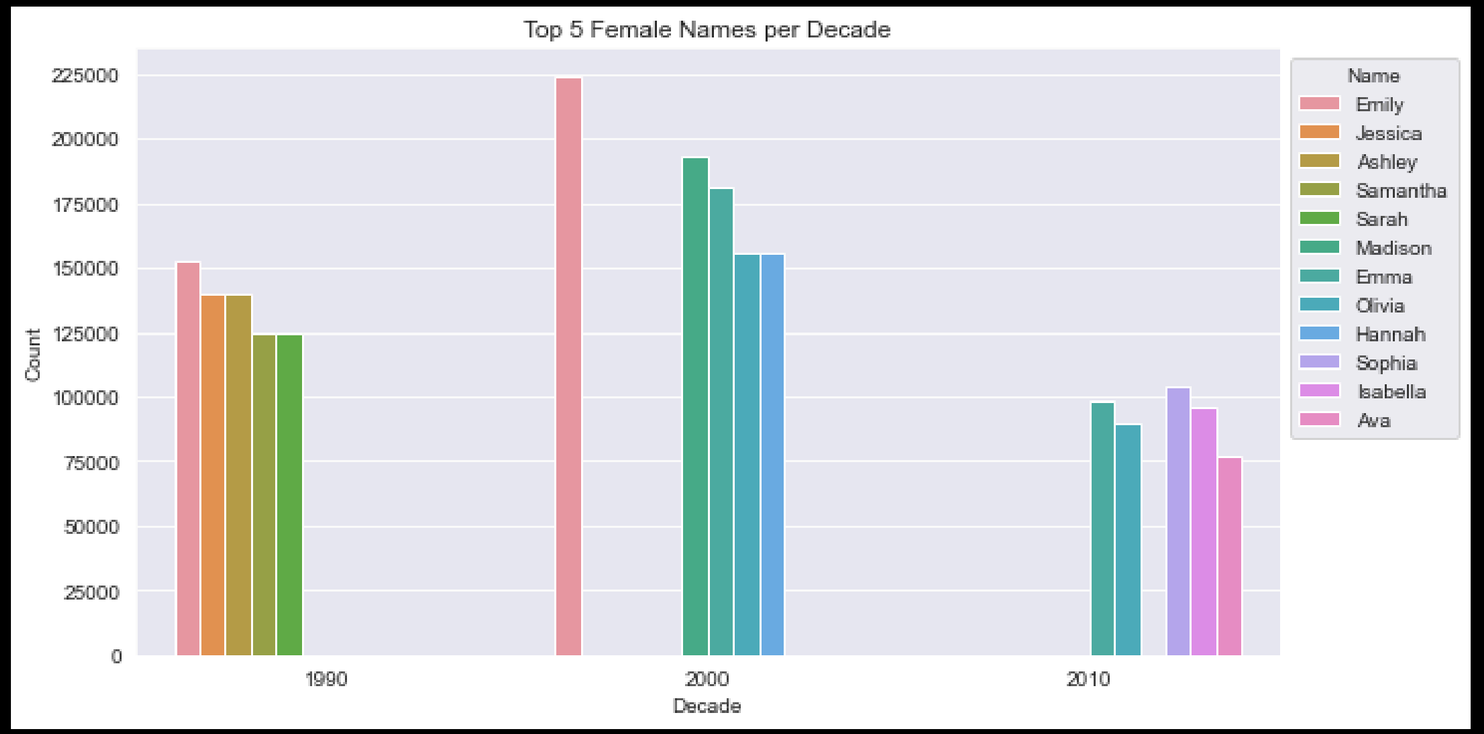
TRENDS ARE CHANGING

STRV

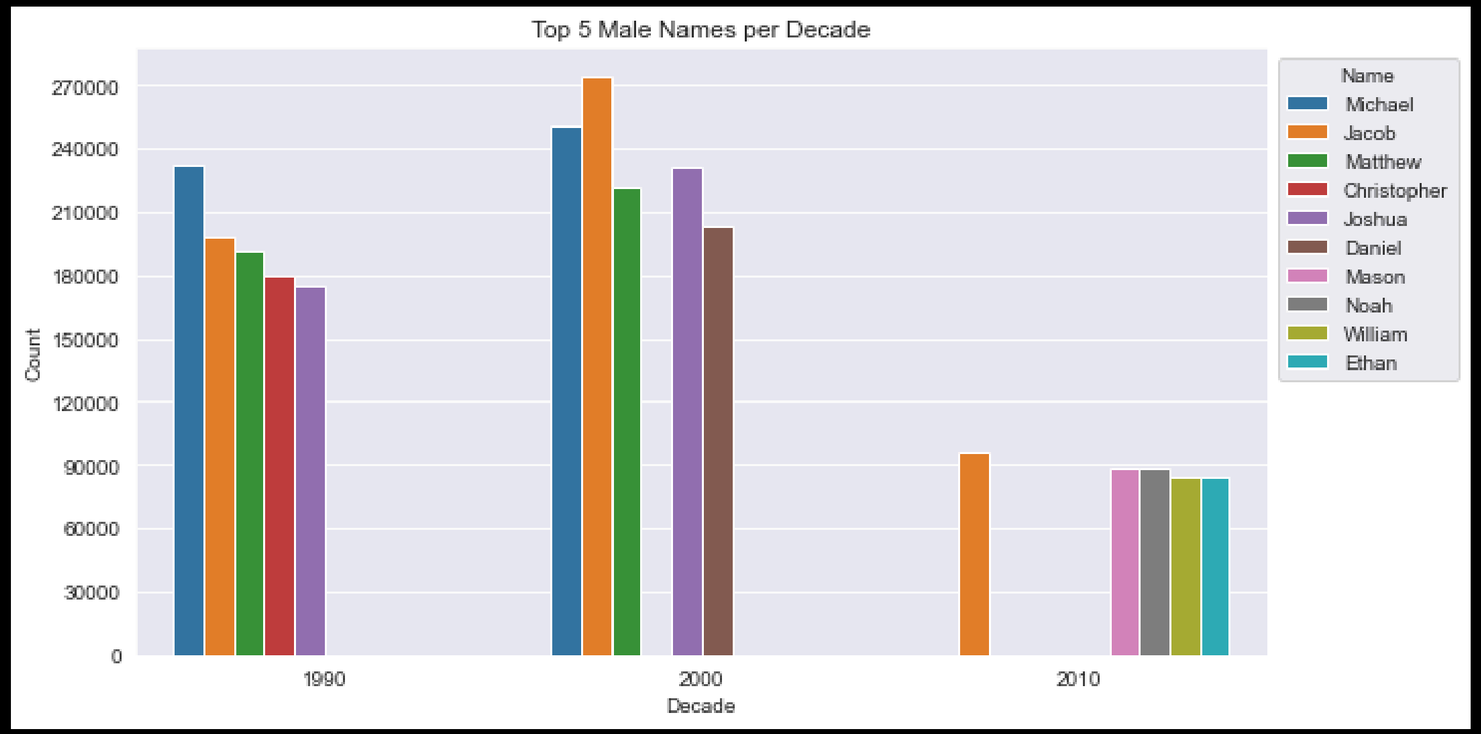
From the mid-90s to the mid-2010s the world evolved



Looking at the top 5 names in the last 20 years of data per decade shows us an interesting trend: In the 2010s, the variety of Female names decreased drastically while the variety of Male names may have increased, being the only period of Data analysed where the top names are mostly female. This can be seen in opposition to the previous 15 years(dataset has data up to 2015, hence 3 "different" decades for only 20 years) where the top 5 registered names are mostly male. This can be affirmed as the proportion of male-registered and female-registered names in the period remains close to 50/50.



While the variety of common names has increased, the trendiness of female names seems to remain, being that this is the only group that includes 11 distinct names out of 15 possibilities and it can be observed that in the late 90's the 5 most common names are completely different from the 2010s. This is less true when looking at the data from 2000 on, where the overlap is more significant.

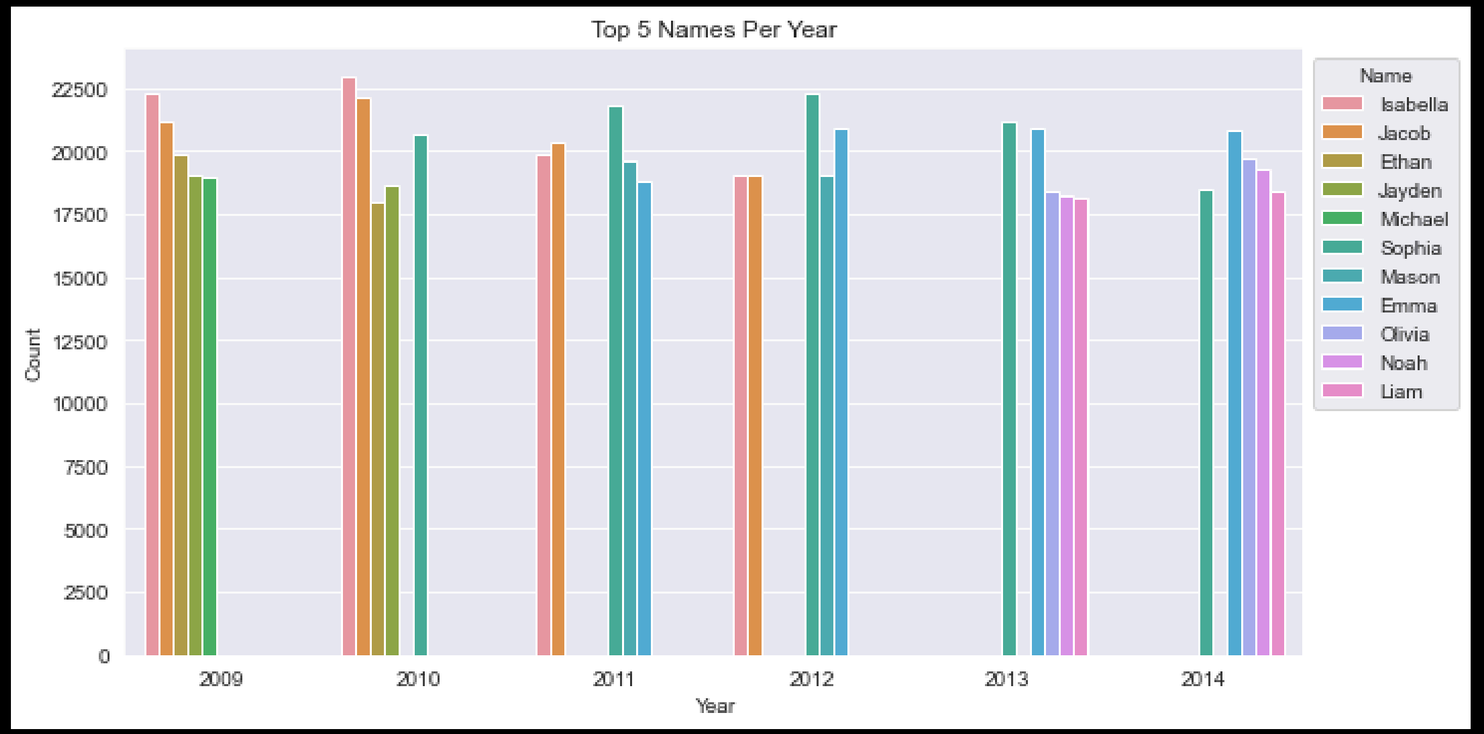


As postulated earlier, the pattern of the most common male names remaining the most common over time appears to be shifting to a more trendy name choice. The fact that we can only see this when looking at the most recent data available, as opposed to the female case (where names appear to go in and out of "fashion" completely in a relatively short timeframe), might indicate that this is a temporary shift, but more recent data would be necessary to confirm this hypothesis.

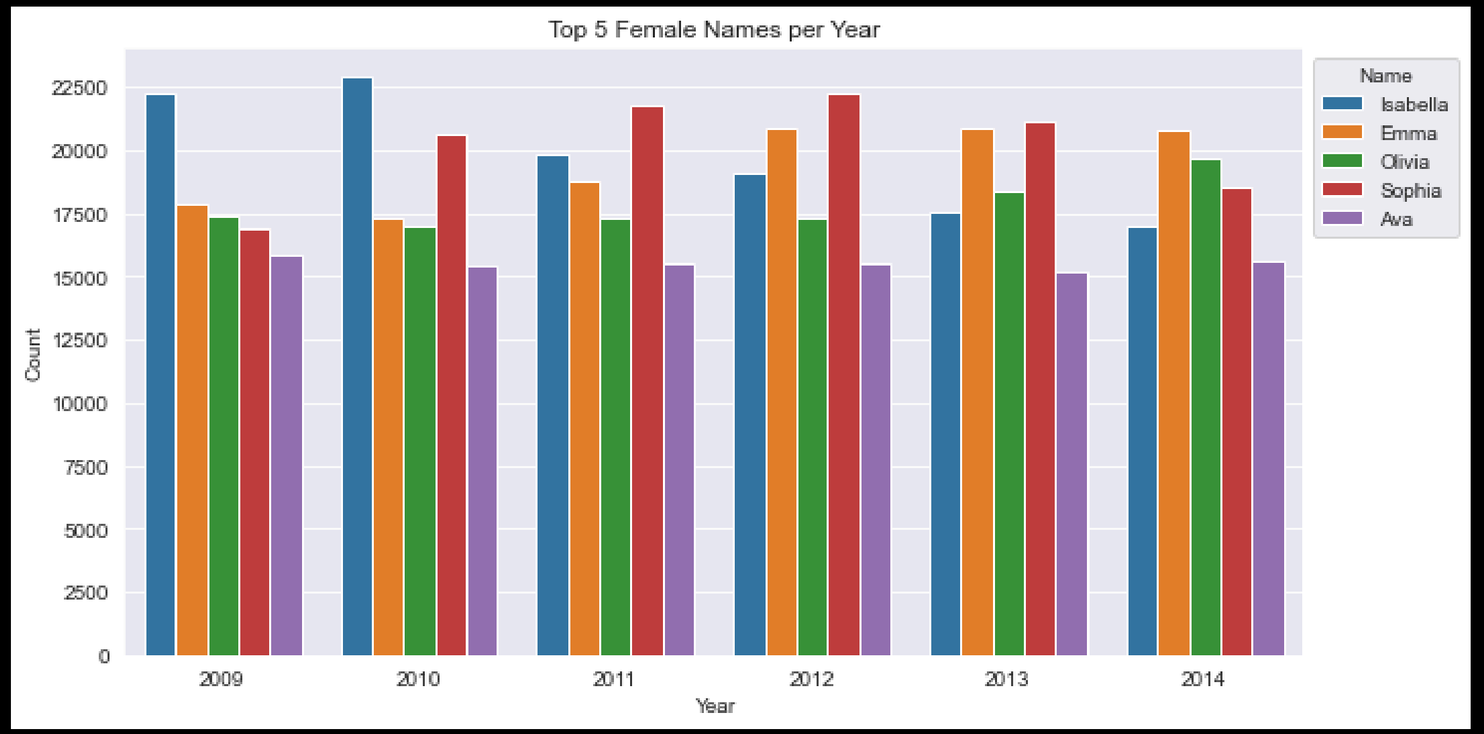
THE 2010S ARE WEIRD

STRV

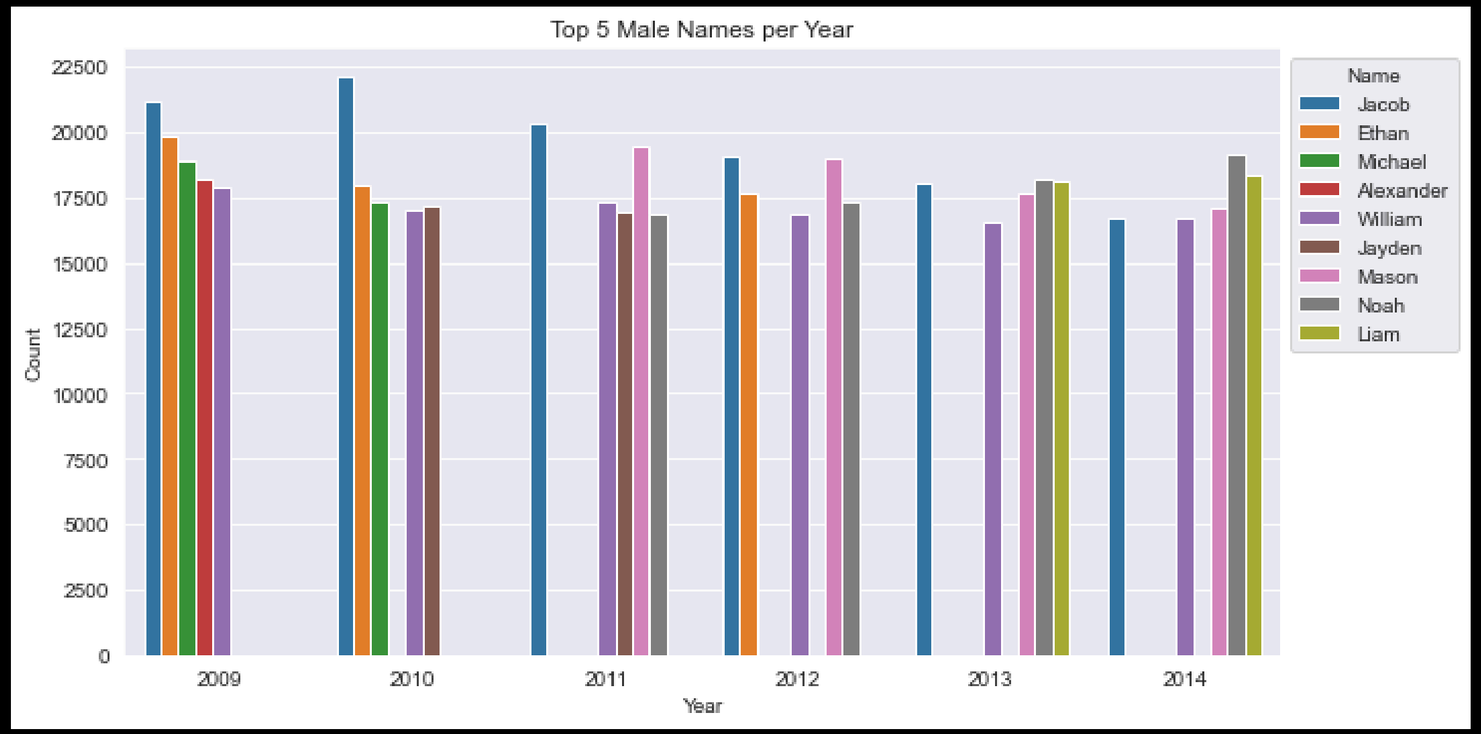
A complete 180



As seen in the previous section, the last 5 years of data show a reversal in the overall, last 50 and possibly last 20 years most common names, where the top ranks were usually occupied by male names and the most common female names exhibited high variance (only 3 names make the top ten both overall and in the last 50 years, as opposed to 7 male names. We also see a sharp decline in absolute occurrences of the most common names in the last 5 years when compared to an equal window of time in the second half of the 90s decade.



Looking closer at the data from the last 5 years, we see a more exaggerated reversal of the trend, having the 5 most common female names remain the same for all years.



Again, when looking at male names in the first half of the 2010s, we're presented with 9 different names in the top 5 most common names for each year, a reversal of the overall and past 50 years trend (where the most common male names showed low variance).

Conclsions & Next Steps

STRV

What we learned and What we can take from this

Insights

The first years of a company are known to be the most crucial times to set up for or prevent future success. As such, the importance of managing the resources allocated is even higher when they are more limited (as is common for launching startups). This is also exacerbated by the nature of the fashion business, a cash and resource intensive industry with high overhead and complicated infrastructure and supply chains. The analysis supports a few key choices that could guide the resource allocation and allow the company to focus on improving manufacturing processes and refining the supply chain while also balancing the delivery of customized products:

Sell to all babies - The population of newborns tends to be approximately 50% male and 50% female; why ignore either half of the possible market?

Pre-poduce for the most common names - As shown in the distributions both for the overall count of names and in the count for the last 50 years, more than 75% of the new babies have one of the most common names, while the 75th percentile of names has at most 250 occurrences and the most common names have counts in the hundreds of thousands of occurrences. Knowing this, investing in “templated” designs for these more common names will ideally ensure that most customers will have to wait next to nothing for their product, while a small amount (with less common name requests) will also wait less, as only their requests demand design from the ground up.

Keep an eye on new trends - While we can’t say for certain if it will continue in this direction, there has been a noticeable shift in the last 5 years of available data. While the top male names appear to change a lot more than in the past and the more common female names a lot less, a switch of trends compared to other periods explored, one thing is consistent in both groups - The more common names are becoming less common in absolute terms. An early suggestion would be to look a larger group of the most common names when pre-producing designs.