For this project, I sought to find out which types of businesses in New York City saw a lot of growth post-pandemic, and which ones did not. This data project will eventually connect to a story I am pitching in my business-related class.

I retrieved this data from New York City’s Open Data website. I chose to use business license data because, otherwise, it would be difficult to quantify business growth in the city, particularly among small, private businesses. Additionally, the dataset is quite expansive, meaning that I could analyze trends even further back in time than the pandemic.

My central reporting question – which business types are thriving or declining in New York City post-pandemic – is quite simple. Nonetheless, the analysis is complicated by a few things:

* The original dataset doesn’t include any unique identifiers for individual businesses, meaning that it would’ve been difficult to pinpoint particular businesses with the original data alone. I believe this is due to chain stores and businesses having multiple sites.
* Similarly, even individual businesses had several entries for the same license, even though issuance and expiration dates for those licenses differed. I’m not sure why this is, and this is something that I’d have to address through additional reporting.

As a result, I chose to create a unique identifier by combining the business name and location. Then, I used the unique identifier that I made as a basis for the “drop duplicates” function, which removed all duplicate identifiers while keeping the last.

The caveat here, however, is that I no longer have any idea which licenses are active, expired, ready for renewal and so on. I chose to ignore this for the purpose of this particular project, because I just need to know that a business was issued a license at some point — not necessarily that business's current status. And for the same reason as stated above, some individual businesses have multiple entries for the same license, which makes keeping that metric in the analysis all the more confusing.

I found some interesting things once the analysis was complete. For instance, the percentage change between 2018 and 2024 for businesses selling products for people with disabilities, pedicab drivers and scrap metal processors is over 100%. Process server individuals have also increased, despite process serve agencies decreasing.