ETL Project:

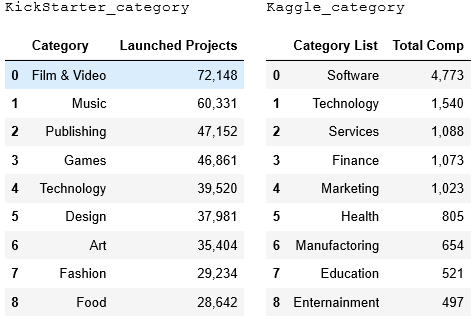
**Extraction**

We used 2 datasets from the public platform Kaggle and Kickstarter. All of our data was based on the amount of money finding the least resistance in the forms of what industries are thriving and where the funding and investing is being poured into. The sources for our dataset are as follows

* Projects and Dollars from Kickstarter
* Startup Analysis Dataset from Kaggle

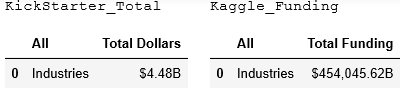
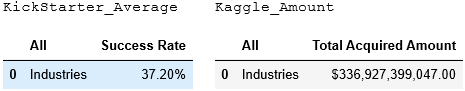
**Transformation**

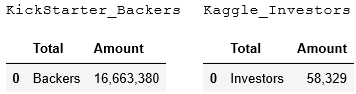
The first step was showing how many Industries are in the Kaggle Dataset and what KIND of industries that are in said dataset, Kick Starter’s dataset already shows us how many and what kind of industries they have. We then compared Kickstarter’s and Kaggle’s top 9 most booming industries. After this was made, we then narrow down the status of all the companies in those industries which came as four categories; Operating, Closed, Acquired and IPO.



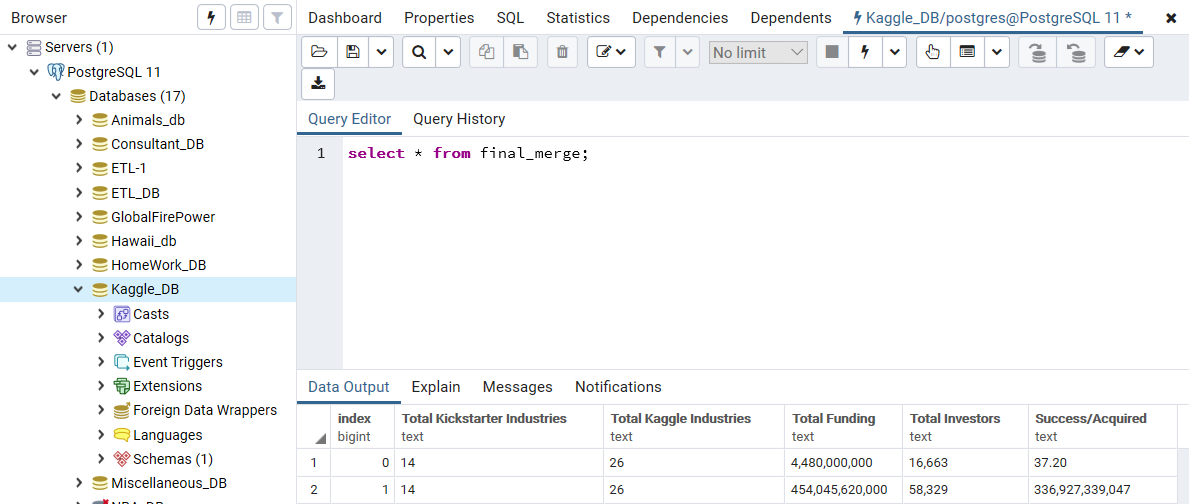
Next, is the total amount of money/funding, success rate/acquired amount and backers/investors in both respective datasets. This was to show where exactly the $$$ is finding the least resistance for it to grow and keep growing as well as the number of investors that were active. Tables are shown as followed;



**Load**

The last step was to transfer our final output into a Database. Since our goal was to compare ALL of the industries in each dataset, we unfortunately were not able to execute the original plan. However, we were able to display that crowdsourcing sites like Kickstarter are beginning to have clot for being a legitimate place for startups to be funded from the people. We performed an ‘to\_sql’ function with ‘con=engine’ and ‘engine.execute’ to manually submit the final merge into postgres.



**Summary**

We used these datasets so we could identify; which industries are thriving, the amount money/investments are being poured in, the status of the businesses that are in those industries, and the amount investors that are part of said industries. The final output shows us from a crowdsource funding perspective and a VC funding perspective.

* Total Amount of Industries
* Total Funding
* Total Investors
* Success Ratio/Acquired

This was not the intended goal of the project, we wanted to show where exactly future investors should be placing their investments and where the money is growing. However, the data instead shows that crowdsourcing sites like Kickstarter are becoming more reliable sources of funding for businesses then for businesses hoping they will be backed by a VC firm. Perhaps, VC’s should consider this an interesting opportunity and start to either partner-up or acquire crowdsourcing sites if the data suggest the upward trend is not stagnant in the near future.