

Project 2

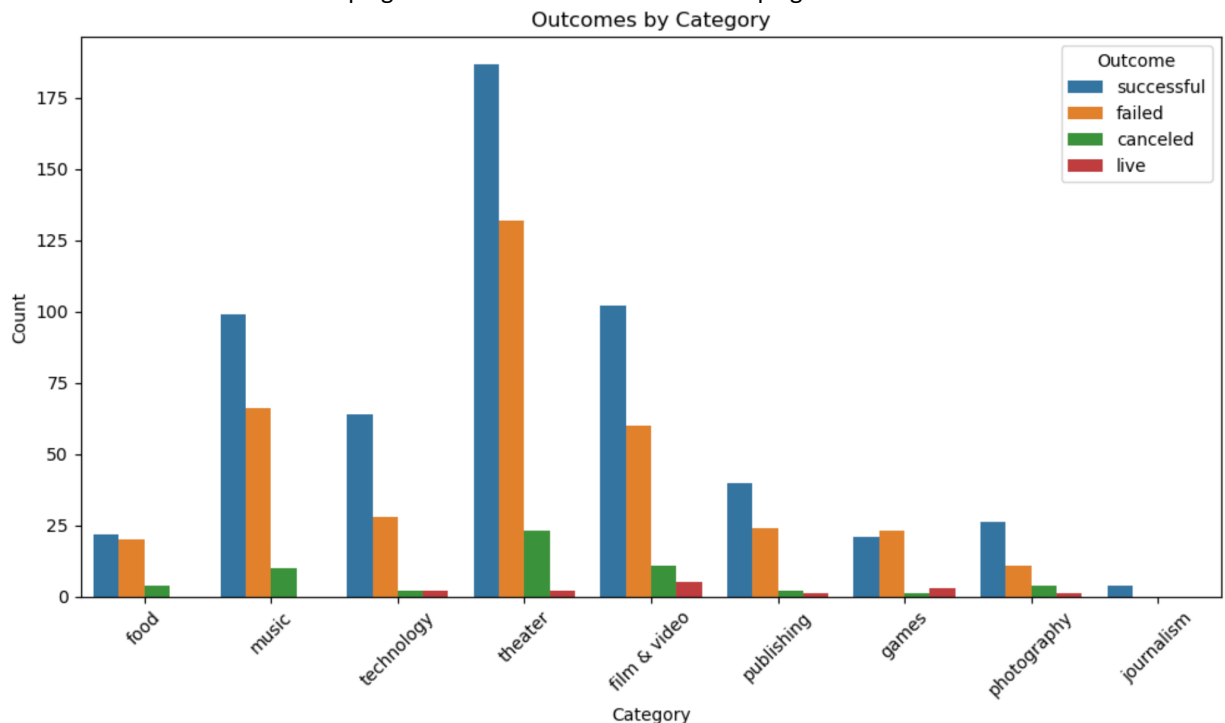
Group 2: Kevin Carney, Fatima Rodriguez, Herbert Dennis

Project 2 Crowdfunding Summary

Over the course of Project 2 our group worked with crowdfunding datasets with the purpose of extracting the data, transforming it to a database schema, and loading it into a database where three queries were run. The crowdfunding dataset that was used had factors such as contact id, first and last names, email, company name, goals, pledged amounts, outcomes, country, and category, and subcategory. Using some of these factors we hoped to create a database that could represent certain patterns in the crowdfunding that may lead to successful campaigns.

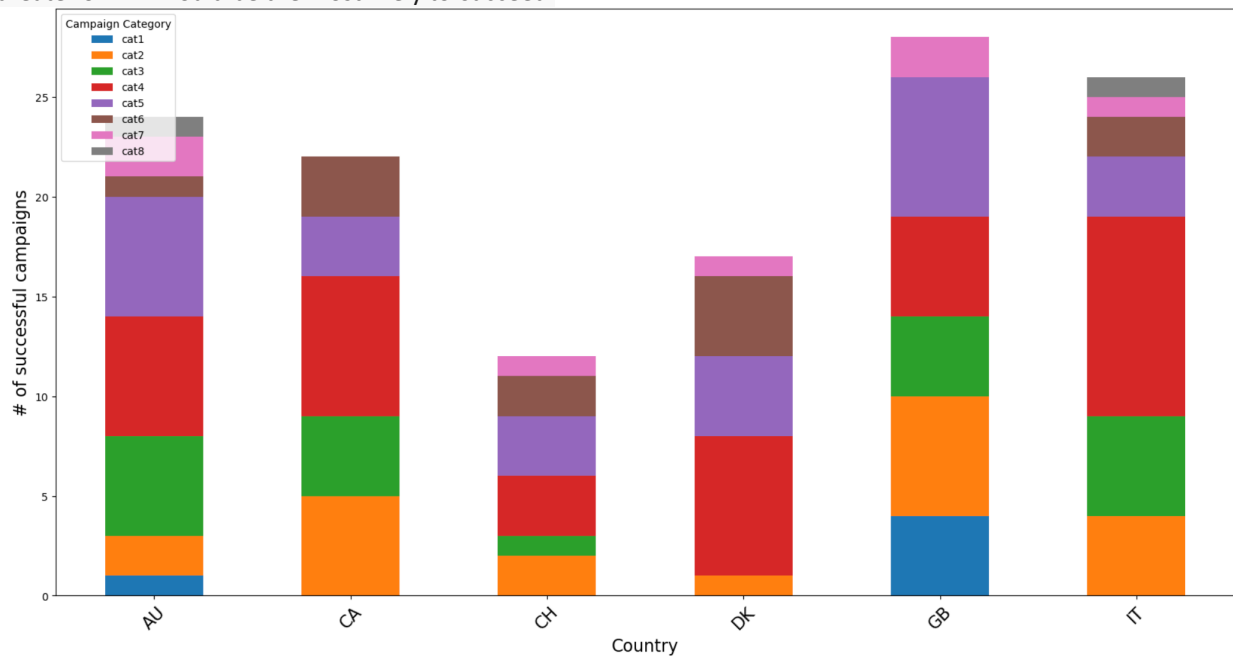
The focus on the design of the database was to create a database with four tables being “campaign”, “contacts”, “category”, and “subcategory”. Within these tables valuable information from the crowdfunding datasets were sorted in a way that is easily accessible to draw better conclusions from. The first step in moving towards this was to extract the data from the datasets in the “contacts” and “crowdfunding” Excel files. This was done in Pandas through a Jupyter Notebook and being put into various data frames of “campaign”, “contacts”, “category”, and “subcategory”. The transformation was complete once the data frames were exported to CSVs. Once the CSVs were obtained, they were loaded into the database once the tables were created after the schema was put into Postgres.

After the database was completed, our group created three queries based on the data. The first query was to show the outcomes of the crowdfunding campaigns based on the categories to have a representation of the success level of different categories. The category of “theater” led all categories in successful campaigns and the overall number of campaigns showing a significant interest in the category. It was followed by “music” and “film & video” in terms of successful campaigns and the overall number of campaigns.



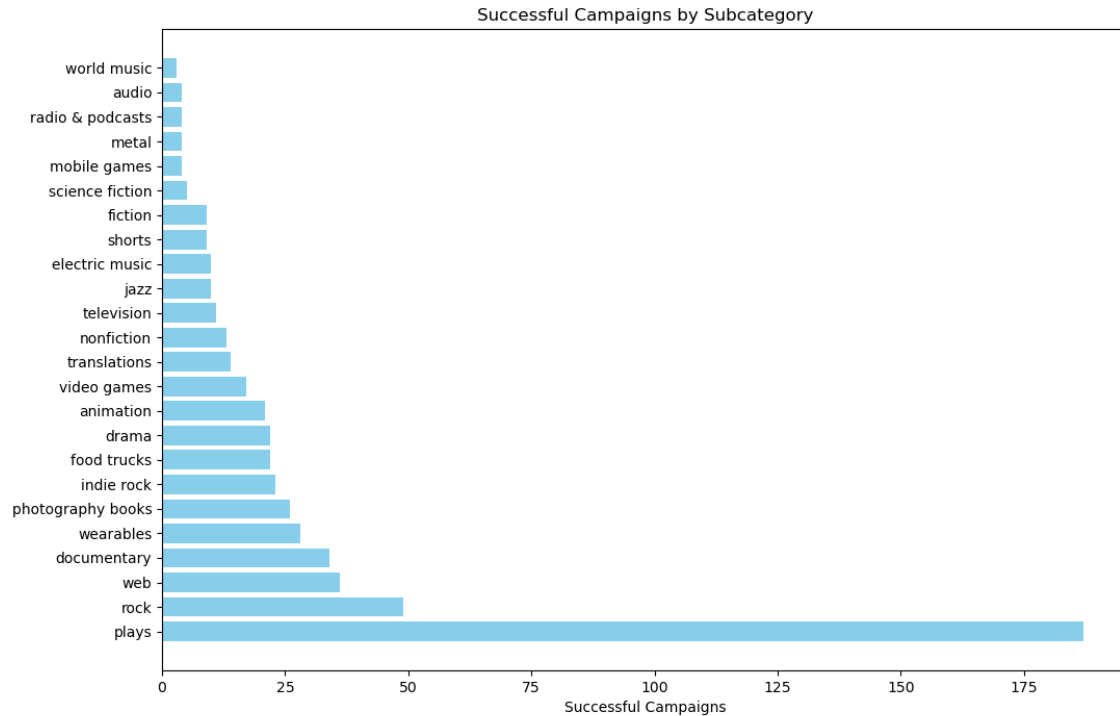
The second query was reviewing the number of successful campaigns from different countries. When we were reviewing our data, it was noted that in terms of number of campaigns the US is the outlier in this dataset,

making up 77% of all the successful campaigns. With that in mind, a question was raised on what the breakdown of the successful campaigns would be if the US was not included in the query. Out of all the remaining datapoints, the country with the most successful campaigns was Great Britain, coming in with 28 campaigns. Italy came in second with 26 successful campaigns, and Australia came in third with 24 campaigns. Of the category types for these campaigns, the category with the most successful campaigns was category 4 which is tied to theater. 38 of the 129 successful campaigns for non-US countries fell into this category, equating to about 30% of the dataset. Category 5, which was film and video, ranked 2nd at 26 of 129 or 20% of the dataset. With these results in mind, were a potential creator of a Kickstarter project trying to devise a campaign for outside of the US, the data suggests that theater or film would be the most likely to succeed.



The final query looked at the success of the subcategories. There were in total 24 sub-categories. As shown in the graph, the most popular sub-categories were plays, with a count of 187 successful campaigns followed by rock and web with 49 and 36 successful campaigns respectively. The least popular sub-categories were

world music, audio and metal with less than 4 successful campaigns.



Some of the bias in the data is that some of the categories can have overlapping characteristics, so it would be interesting to see the process on how they were grouped to have a better idea of which category to invest in for a better success rate. For example, “theater”, “music”, and “film and video” were the top three categories, but are music videos apart of “music” or “film & video”. Another example is for theater, if a play is to be filmed and put into theaters is it considered “theater” or “film & video”. These were not represented in the subcategories.

In conclusion, the process of extract, transform, and load proved to be a good option in the transformation of the two original datasets as a finished product to a database. Once the tables were created, we were able to do actions such as use joins, selects, and wheres for table transformation. This in turn allowed us to create queries where we got a clearer picture of what types of categories and subcategories were successful as well as the countries that they originated from. We can share the code of our database so other users can query other information they might need moving forward for a campaign seeking funding.