

Will Parker, Neyda Morales

Professor Alexander Booth

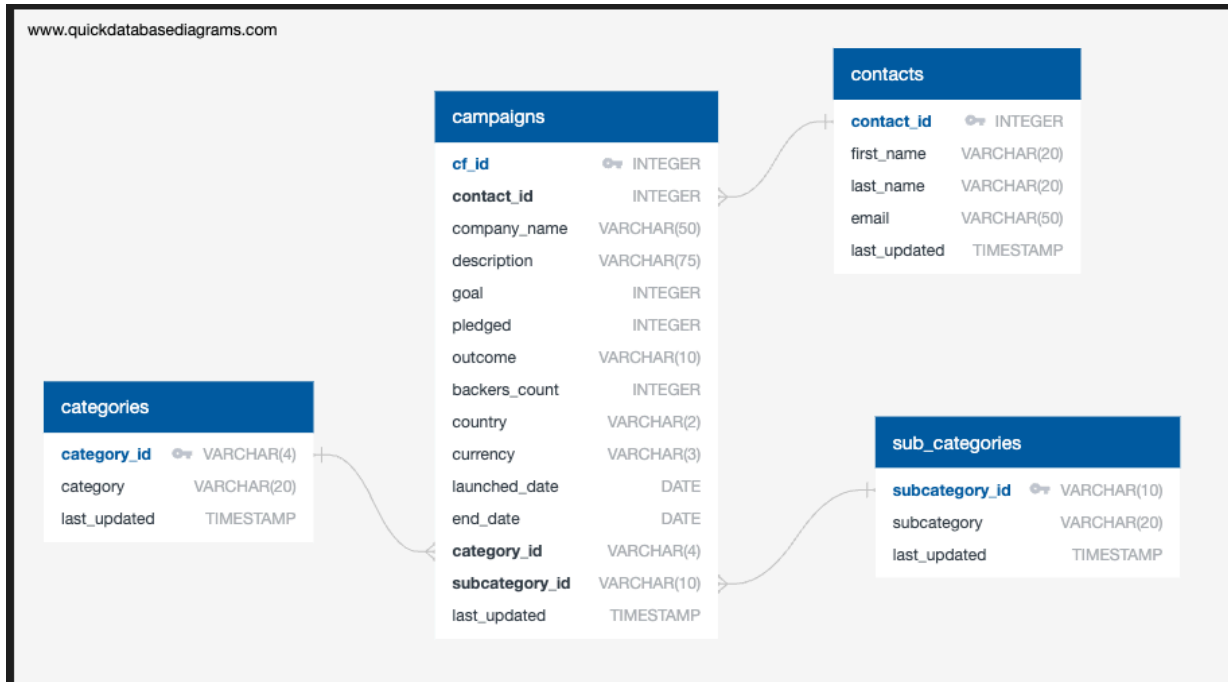
Data Analytics Bootcamp (Project 2)

July 24, 2024

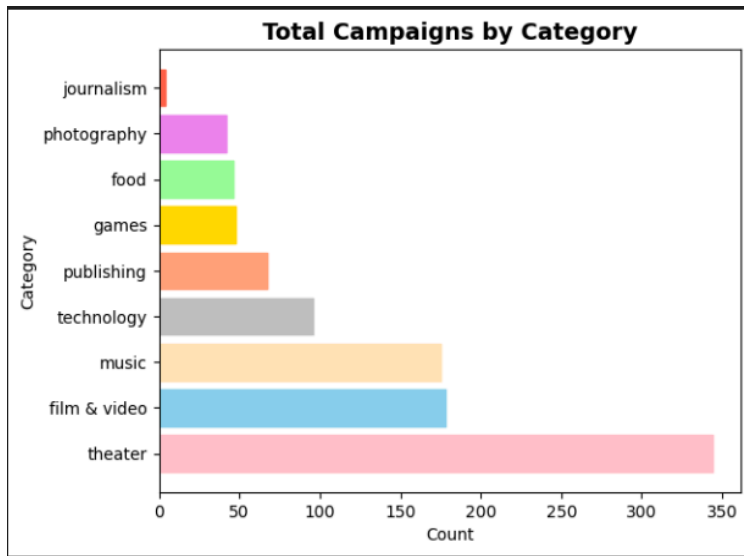
### Crowdfunding Database and Analysis

The purpose of this project was to design a relational database utilizing best practices and implementing an Extract, Transform, Load pipeline. The data for this project compiled 1,000 companies and their crowdfunding campaign targets. The dataset included a brief description of what the campaign was for, the company's crowdfunding goal, how much was pledged to their campaign, how many individuals contributed, and the categories each company belonged to. The companies were located across 7 countries: US, Canada, Australis, China, Denmark, Great Britain, and Italy, and the time frame was between 2020 - 2022.

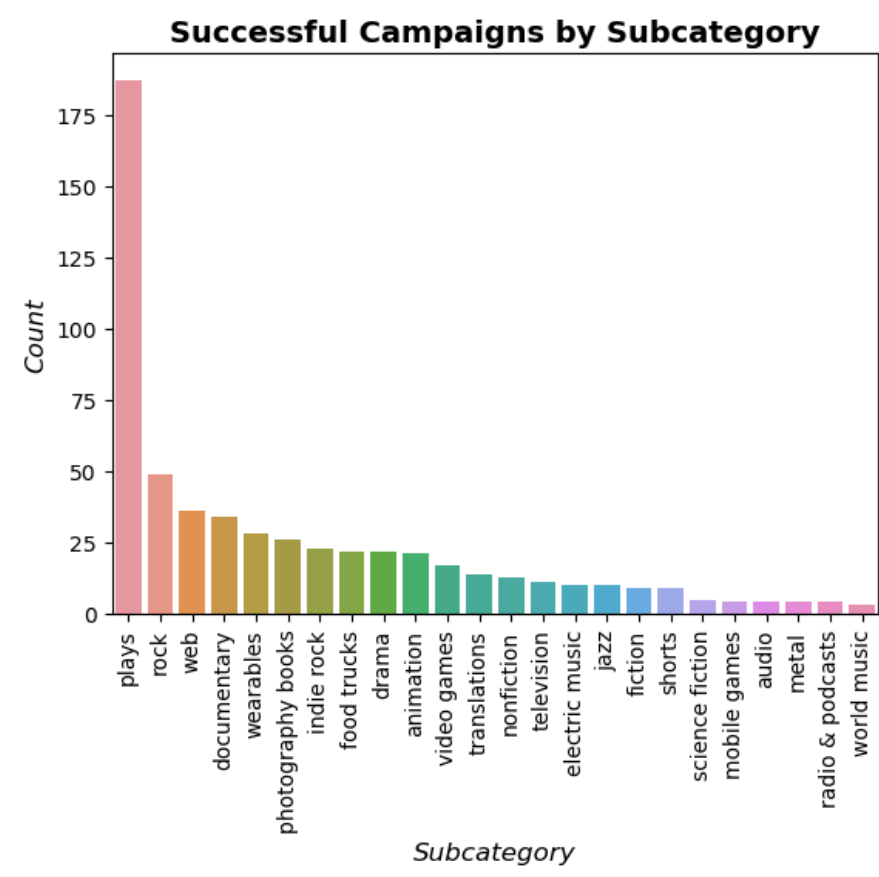
We began by Extracting the data into Python data-frames. After reviewing the data types, we converted the goal and pledged columns into floats, adjusted the launch date and end date columns to datetime format, and constructed the contact info from its original form into separate columns for Contact ID, First Name, Last Name, and Email. We decided to create separate tables for category, sub-category, and contacts, so the crowdfunding table could reference those instead of having the same repeated text. Thus, the data was Transformed into 4 tables to fit the database schema: Category, Subcategory, Campaign, and Contact Info. We annotated data types as required for database schema constraints. Lastly, we Loaded the cleaned database schemas into pgAdmin for analysis. The analysis incorporated an Entity Relationship Diagram and several queries and visualizations.



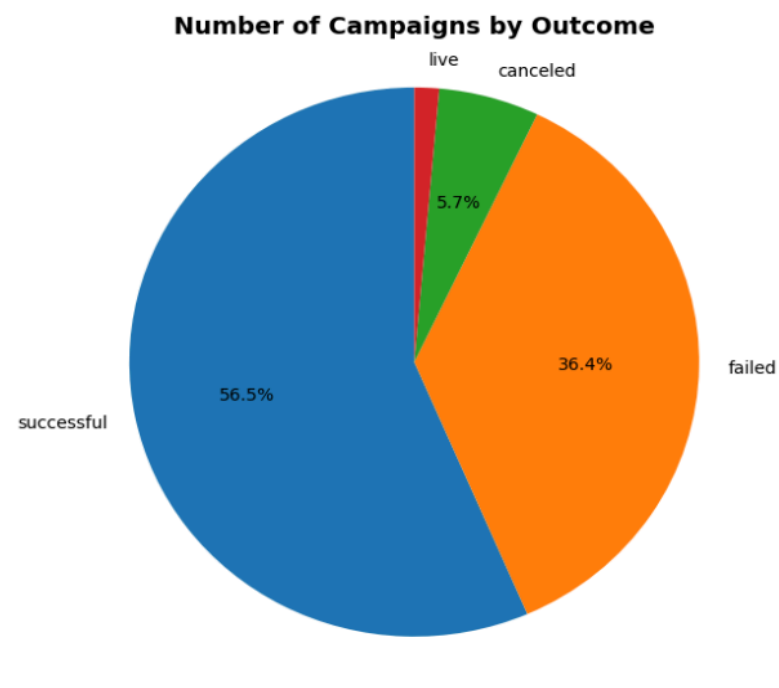
The first query we ran looked at the total number of campaigns by category. The category with the most campaigns was theater followed by film/video, music, technology, and publishing. Those being the top 5. We noticed there was a major difference between theater with 344 campaigns and journalism with 4 campaigns. Film/video and music had about the same number of campaigns with 178 and 175 respectively while Games, Food, and Photography held the same number of campaigns at around 45.



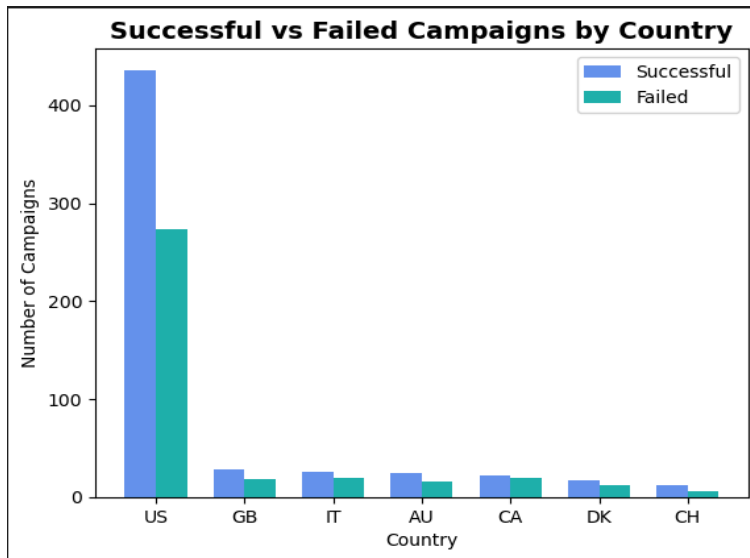
Our second query looked at the number of successful campaigns by **subcategory**. Plays had the highest number of successful campaigns with 187. The runner up was Rock, a subcategory of music, with 49 successful campaigns followed by Web, a technology subcategory, Documentary, a film/video subcategory, and Wearables, another technology subcategory. The bottom 5 subcategories fell within the Games, Journalism, Music, and Publishing categories. The subcategories being mobile games, audio, metal music, radio/podcasts, and world music.



Our third query looked at the number of campaigns by outcome. There was a total of 565 successful campaigns across all categories compared to the total of 364 failed campaigns. The total number of campaigns that were canceled was 57 and the total number of ongoing (or live) campaigns is 14. Successful campaigns were 20% higher than failed campaigns.



Finally, the last query we ran looked at the number of successful and failed campaigns by country. The US tremendously outweighed the other countries in both successful and failed campaigns. The US had 436 successful campaigns and 274 failed campaigns. Great Britain, Italy, Australia, and Canada had between 22 – 28 successful campaigns and 16 - 19 failed campaigns. Denmark and China had the lowest numbers with 17 and 12 successful campaigns respectively and 12 and 6 failed campaigns respectively.



In conclusion, "Theater" had the highest number of campaigns with 344 whereas "Journalism" had the lowest number of campaigns with 4. Within Theater, "Plays" had the highest success with campaigns followed by "Rock" music. The subcategories with the lowest success rate were mobile games, audio, metal music, radio/podcasts, and world music. Additionally, when looking across all the categories, there was a total of 565 successful campaigns compared to the total of 364 failed campaigns showing that there were 20% more successful campaigns than there were failed campaigns. Ultimately, both successful and failed campaigns in the US tremendously outweighed successful and failed campaigns in other countries. The US had 436 successful campaigns and 274 failed campaigns while in contrast China had 12 successful campaigns and 6 failed campaigns. This could be due to biases in regulatory environments where tighter regulations in countries such as Denmark and China may prevent companies from engaging in crowdfunding practices.