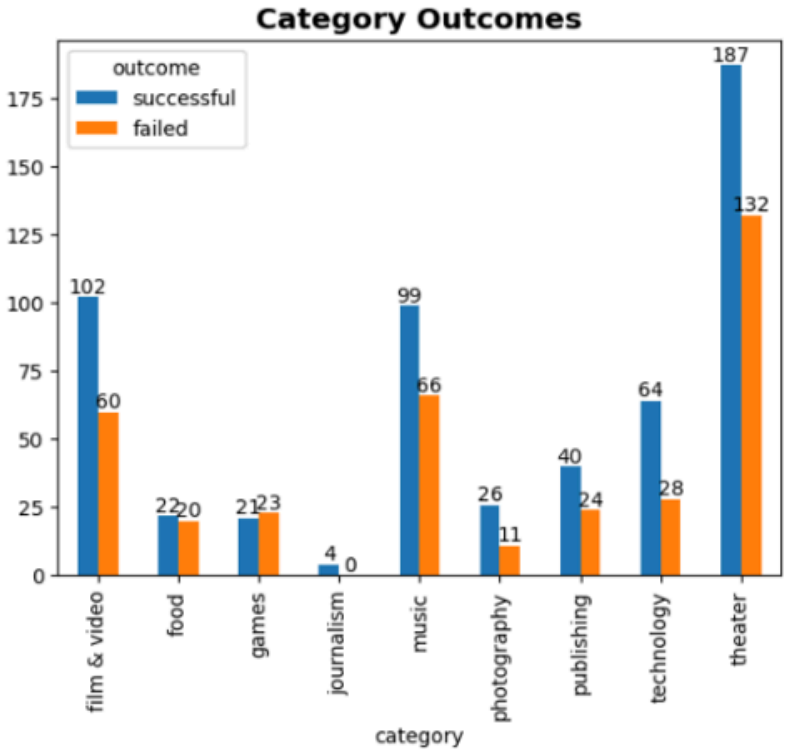
Crowdfunding Writeup

Our dataset for our project contains information from companies about crowdfunding campaigns. Important information includes specific companies, their contact information and individual IDs, descriptions of their companies, the amounts donated and the number of backers, which companies were successful in reaching their goals, nine different categories, and twenty-four subcategories to help organize and categorize the companies for a better overall understanding.

The purpose of this project is to extract data from a source, transform and manipulate the data into a usable format, load it into a DataBase, and finally to do an EDA using our new DataBase.

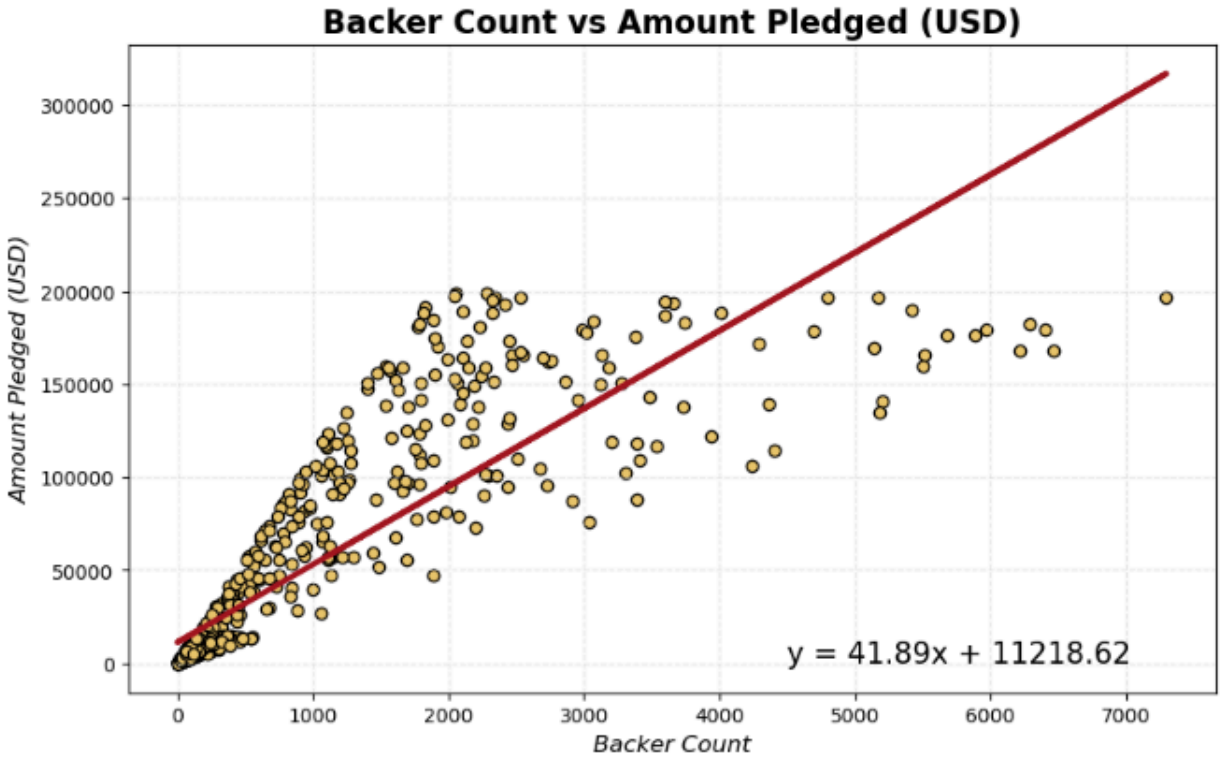
This brings into question, what crowdfunding campaign category is the most successful?



The top four most successful crowdfunding campaign categories are theater, film and video, music, and technology. Based on this information, it can be assumed that a campaign falling under any of these categories has a high chance of reaching its goal. To reach this conclusion, we created a grouped bar chart by running a query to pull the outcomes and category column values from the campaign and category tables. We then performed an inner join to combine the values into the same row, bringing over values from the campaign table and combining them with the category table using the

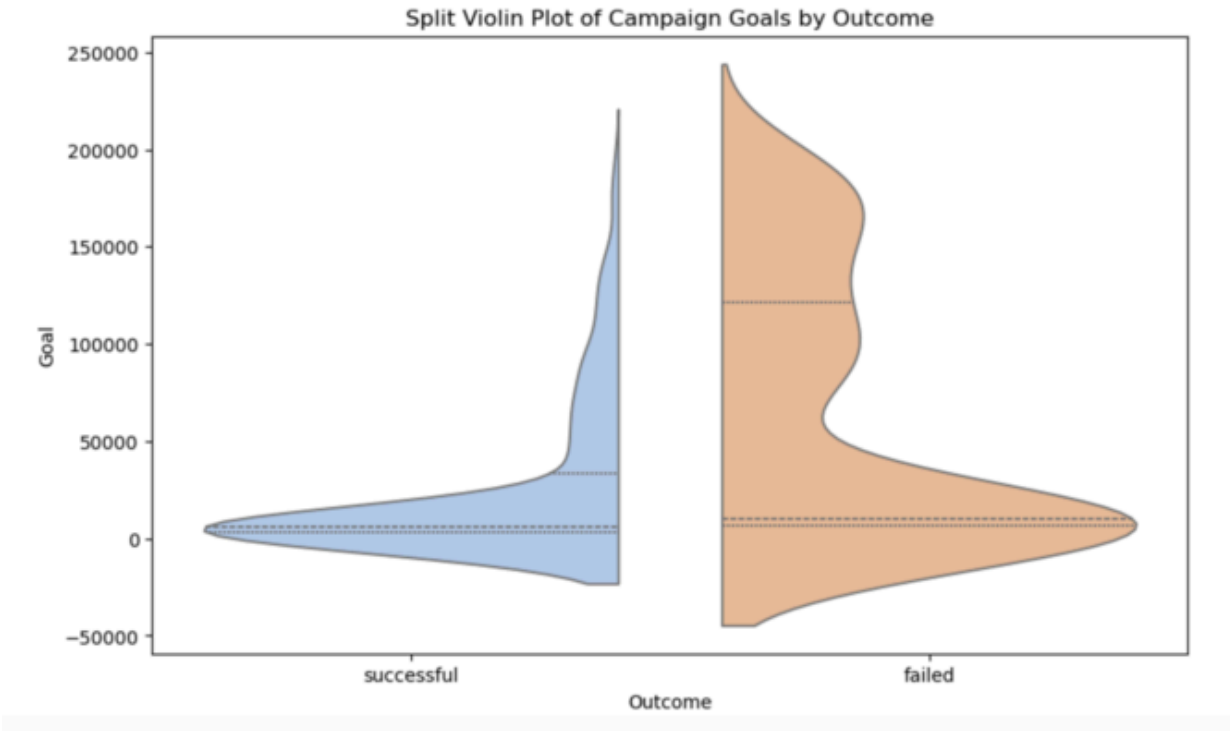
established foreign key in the schema. Finally, we plotted the data into a bar chart using Pyplot and Matplotlib.

What is the relationship between the number of backers and the amount pledged in USD? Do companies need a large number of backers to reach a sufficient amount of pledges, or is there no correlation between the number of backers and the amount pledged?



According to our regression analysis of the scatter plot, which has an R-squared value of 0.75, there is a positive correlation between the amount pledged and the number of backers. This indicates that a larger number of backers results in a higher amount pledged. To get this information we ran a query that selected the backers count, the pledged amount, the currency from the campaign table for values in USD. We then made a linear regression using the backers count as the x value and the pledged amount as the y values using SciPy.

What is the relation of achieving the outcome based on the specified goal amount?



By analyzing the violin plot it is clear that the higher the set goal is, the less likely it will be successful. Additionally, it is shown that at a lower set goal, there is a more balanced amount of successful and failed outcomes. To create this chart, we executed a query to select the outcome values that were either "successful" or "failed," and goal values from the campaign table that were in USD. We then created a split violin plot to visualize the outcomes based on the goal amount.

The data gathered in the dataset is gathered from 2020 through 2022 thus making it potentially outdated. Moreover, the majority of this data is sourced from the United States, which means it does not provide an accurate representation of global trends and patterns.The disproportionate quantity of campaigns for theater compared to other types of campaigns could lead to a misrepresentation of the overall data.

In closing, by gathering the data from the dataset and manipulating it, we discovered that the top leading campaign category is theater, there is a positive correlation between the amount pledged and the amount of backers, and finally we can assume the higher the goal will lead to a lower chance of success.