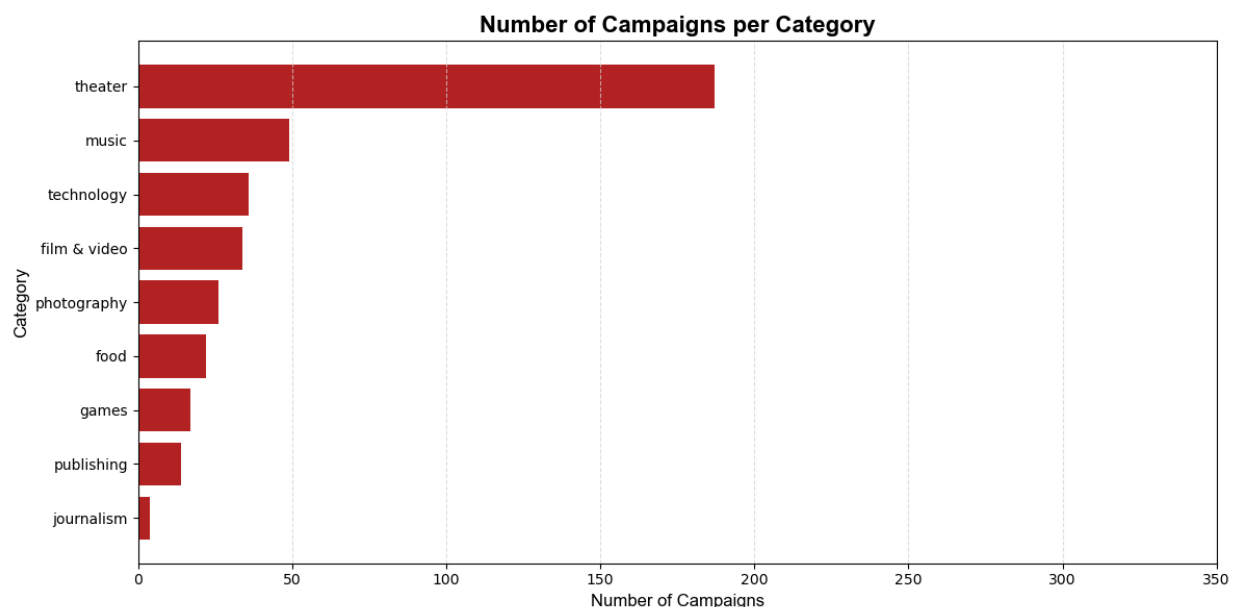Project 2 Group 13 Write Up – Alison McCondichie, Cecilia Rocha, Gina Butler
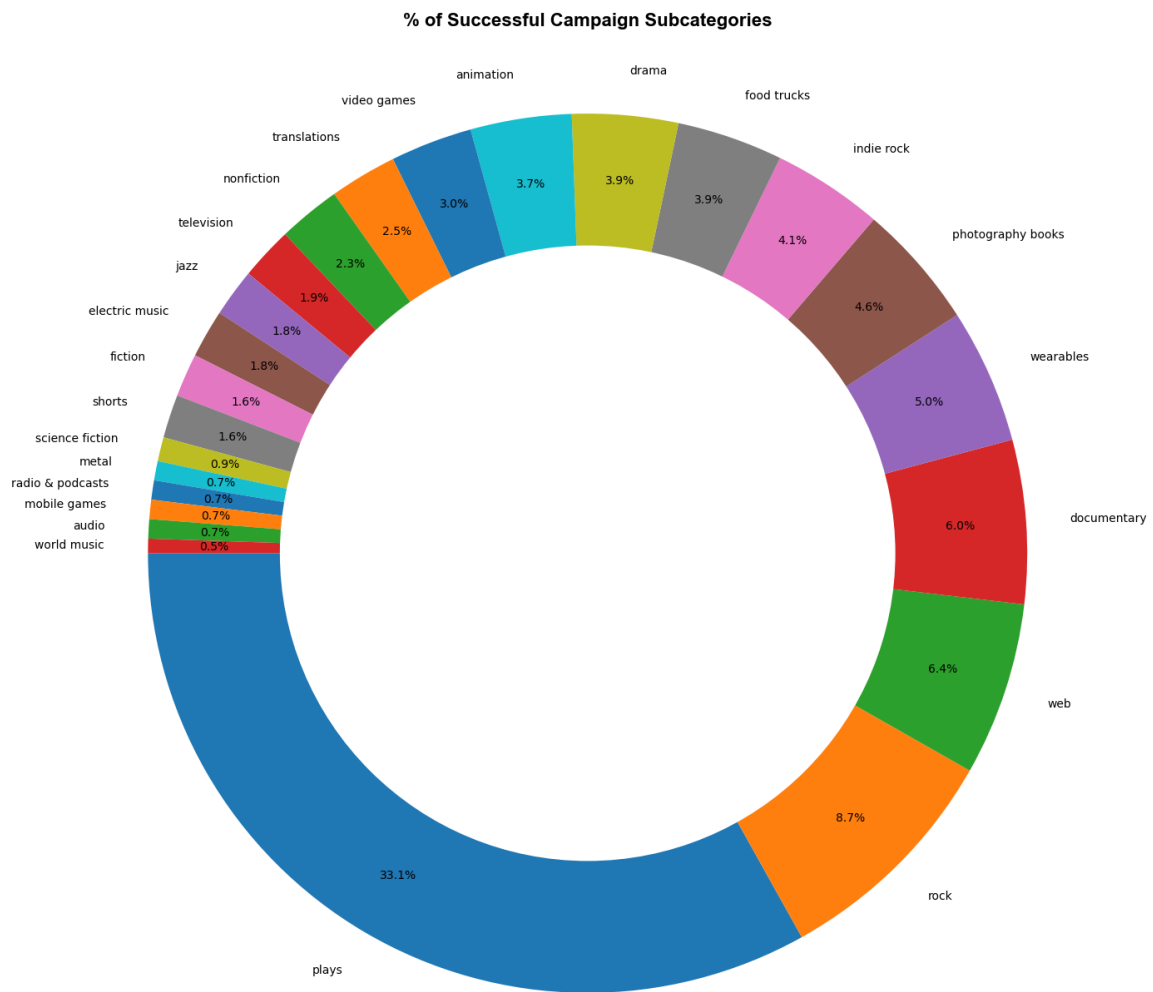
In this project we set out to answer two main questions: Which campaign category has the most successes? Which subcategory has the most successes? We also analyzed the difference between goal and pledged amounts.

We started out by using the crowdfunding.xlsx and contacts.xlsx files. We extracted the data from the xlsx files using pandas function pd.read_excel in a Jupyter Notebook. Within the Jupyter Notebook we added "cat" to the "category_id" values and added "subcat" to the "subcategory_id". In the crowdfunding.xlsx file we renamed it to campaign and cleaned the data by dropping unwanted columns such as "staff_pick". We then exported the files in csv format creating files: campaign.csv, category.csv, subcategory.csv, and contacts.csv.

We designed a database using www.quickdatabasediagrams.com. In the application we created a relational diagram of the csv tables we created. We then generated an ERD snippet .png of the diagram, a text file with the ERD Code, a PDF of the ERD Documentation, and exported a SQL file of the database design. We created the database in PGAdmin (postgres). From there we created a second Jupyter Notebook where we loaded the data to the postgres database using sqlalchemy and pandas. We connected to postgres using sqlalchemy by creating and connecting the engine. Then creating the inspector and connecting it to the engine. After connecting to the engine and inspector we used the pandas function of pd.read_csv to read the csv files that we created in the first notebook, and to_sql function to load the data to SQL.

We then created a third Jupyter Notebook that contained SQL queries that answered our questions. Then we generated two graphs to illustrate our questions.

% of Successful Campaign Subcategories

We found in our queries that theatre is the category that had the most successful campaigns. Specifically, the subcategory of plays within the category of theatre had the most successful campaigns. Our findings from analyzing the difference between the goal and pledged amounts, showed that many campaigns did not come close to goal with their pledged amounts.