

Daniel Sterling

Edward Yang

Ugur Cetin

Project 2 Writeup

Introduction

This analysis focuses on companies in a variety of countries attempting to gain funds for their crowdfunding campaign. We began by taking our initial excel sheets and extracting the data from them to highlight the important information and create our CSV files. Following the creation of our CSV files, we used them to make a SQL database to load and analyze the data.

Extract and Transform

Initially we were given two excel sheets, crowdfunding and contacts, that we needed to transform into CSV files. To do so we used Jupyter Lab. The two main things we needed to extract were the categories and subcategories, which we needed to create data frames from. We then created data frames for the campaign and contacts as well. We created columns for category and subcategory, as well as dropped unnecessary columns like staff pick and spotlight. Changed certain data types as well as merged others. We also labeled our categories and subcategories to organize them.

Database Design

Using Quick DBD, we created a layout of our database and drew connections from our individual CSV files. We connected our contacts, category, and subcategory ids with our campaign database to show how they are related to each other. This also establishes the datatypes that have for our SQL database.

Loading our Data

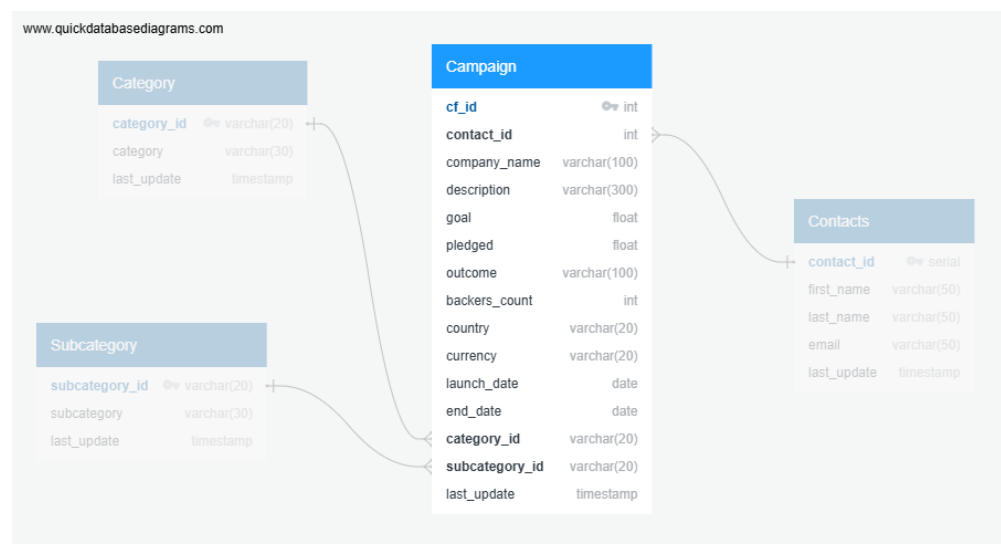
Once we had a design for our database, we used PgAdmin to create an SQL database to test queries that we selected from the data. Instead of manually loading each CSV, we used jupyter lab to create an engine that connected to SQL so we could write lines of code that connected and loaded our data into our crowdfunding database for us.

Analysis of our Data

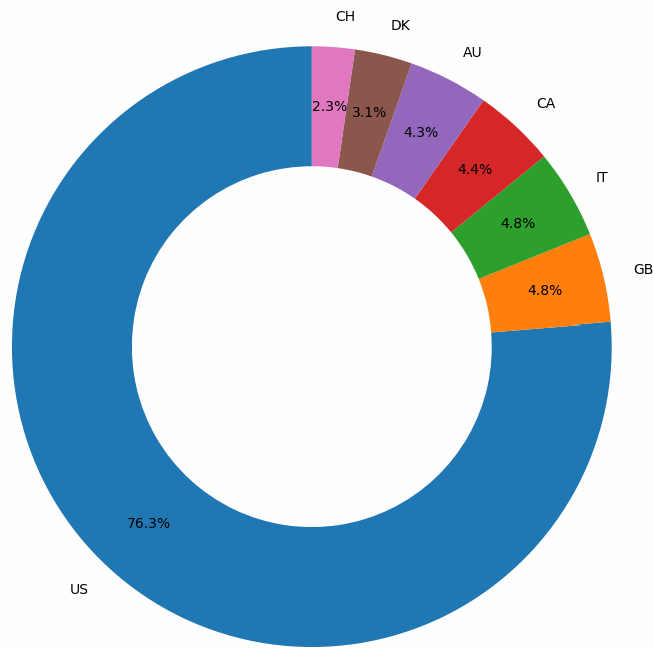
We tested our database by running four select queries that had varied outcomes. From the four queries, we selected two that we created visualizations from. First was our donut chart that represented the number of campaigns per country, which shows the US was responsible for 76.3% of the campaigns. Second was our bar chart that represented the total backers based on currency, which shows the US has the highest number of backers at over 500,000. The majority of campaigns were founded and funded in the US.

Conclusion

The US dominated other countries in regards to crowdfunding campaigns. Instead of involving the US, other countries should be looked at so that we have less biased data. We feel the data doesn't represent crowdfunding on a global scale, and the information was gathered mostly from the US.



Number of Campaigns by Country



Total Backers by Currency

