# Project 2 Group 5: Crowdfunding Campaign Analysis

**Purpose:**
**The purpose of this project is to analyze crowdfunding campaign data to uncover trends and insights that can help organizations and individuals optimize their campaign strategies. By examining key factors such as campaign goals, pledges, success rates, and country wise performance, the project aims to provide actionable insights to improve future fundraising efforts.**

## Project was divided into 4 parts:

**Part 1 - Extract/Transform**
**Part 2 - Database Design**
**Part 3 - Load Data**
**Part 4 - Analysis**

### Part 1, 2, 3 - Extract/Transform, Database Design and Load Data:

1. **Extract:**
    - **Data was extracted from multiple sources, including campaign details, contact information, and category/subcategory mappings.**
    - **The primary data sources included CSV files obtained from crowdfunding platforms.**
2. **Transform:**
    - **Data cleansing operations were applied, including handling missing values and data type conversions.**
    - **Columns were renamed for consistency, launched_at and deadline was changed to launch_date and end_date with timestamps converted to the appropriate format(UTC). Also blurb was renamed to project_description.**
    - **Foreign key relationships were established in QuickDBD to normalize the data using the provided entity relationship diagram (ERD).**
3. **Load:**
    - **The transformed data was loaded into a PostgreSQL database.**
    - **SQL queries were run to verify data integrity and consistency.**

---

**Context and Results:**

### Part 4 - Exploratory Data Analysis (EDA):

**Several queries and visualizations were performed to analyze campaign success based on different dimensions:**

1. **Campaign Success by Country:**
    - **A bar chart visualized the number of campaigns and their success rates by country.**
    - **Key Insight: The U.S. had the highest number of campaigns, but Canada showed a higher success rate.**
2. **Pledge vs Goal Analysis:**
    - **A pie chart was used to analyze how well campaigns met their goals across different categories.**
    - **Key Insight: Technology and creative projects tend to exceed their goals more frequently than others.**
3. **Campaign Trends Over Time:**
    - **A bar chart depicted the top 3 categories with the highest funding.**

**Embedded Visualizations:**

- **Entity Relationship Diagram (ERD) of the database schema.**
- **Bar chart of campaign outcomes by country.**
- **Bar chart showing top 3 categories with the highest funding.**
- **Pie chart comparing pledges vs goals.**

---

**Conclusions and Future Work:**

**Conclusions:**

- **Countries with higher GDP tend to have more crowdfunding campaigns, but success rates vary significantly based on local market behavior.**
- **Campaigns in creative and tech categories often exceed their goals, indicating strong demand and audience interest.**
- **Timing of campaign launches significantly impacts success rates, suggesting strategic timing as a key factor.**

**Future Work:**

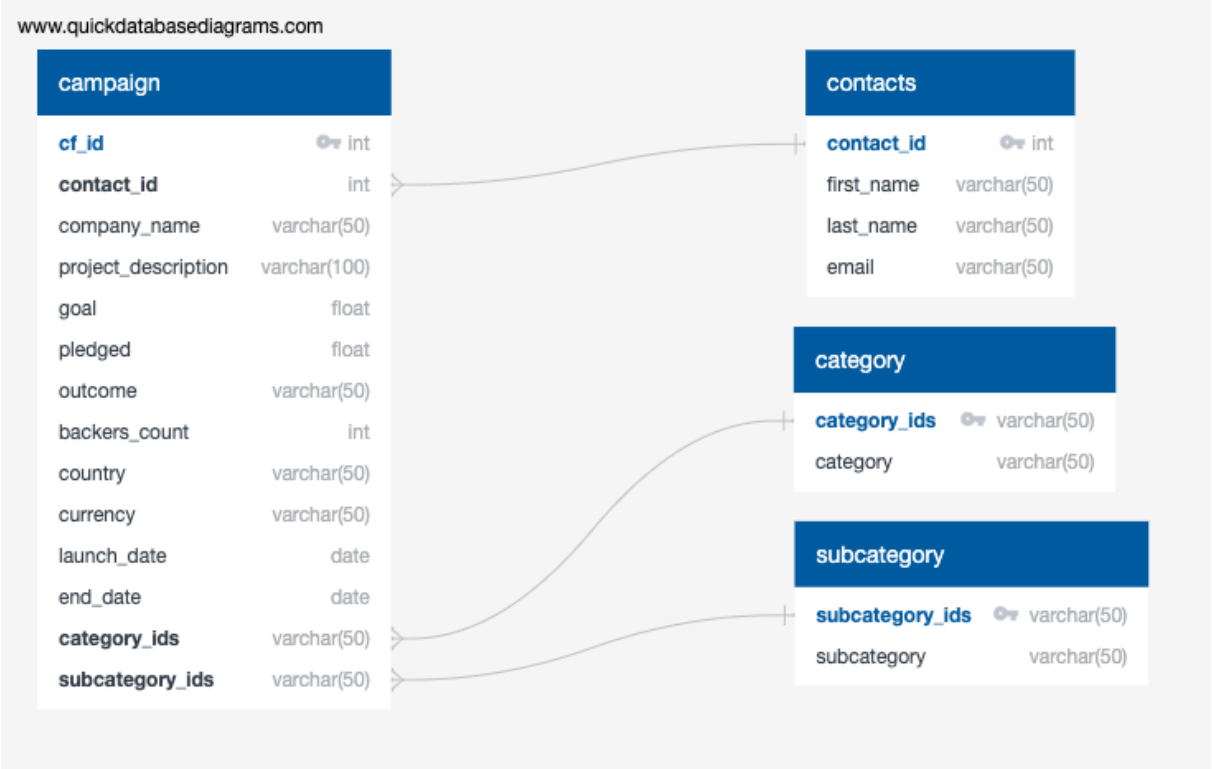
- **Implement machine learning models to predict campaign success based on historical data.**
- **Expand data sources to include social media engagement metrics for more comprehensive analysis.**
- **Enhance the ETL pipeline to automate data updates and processing.**

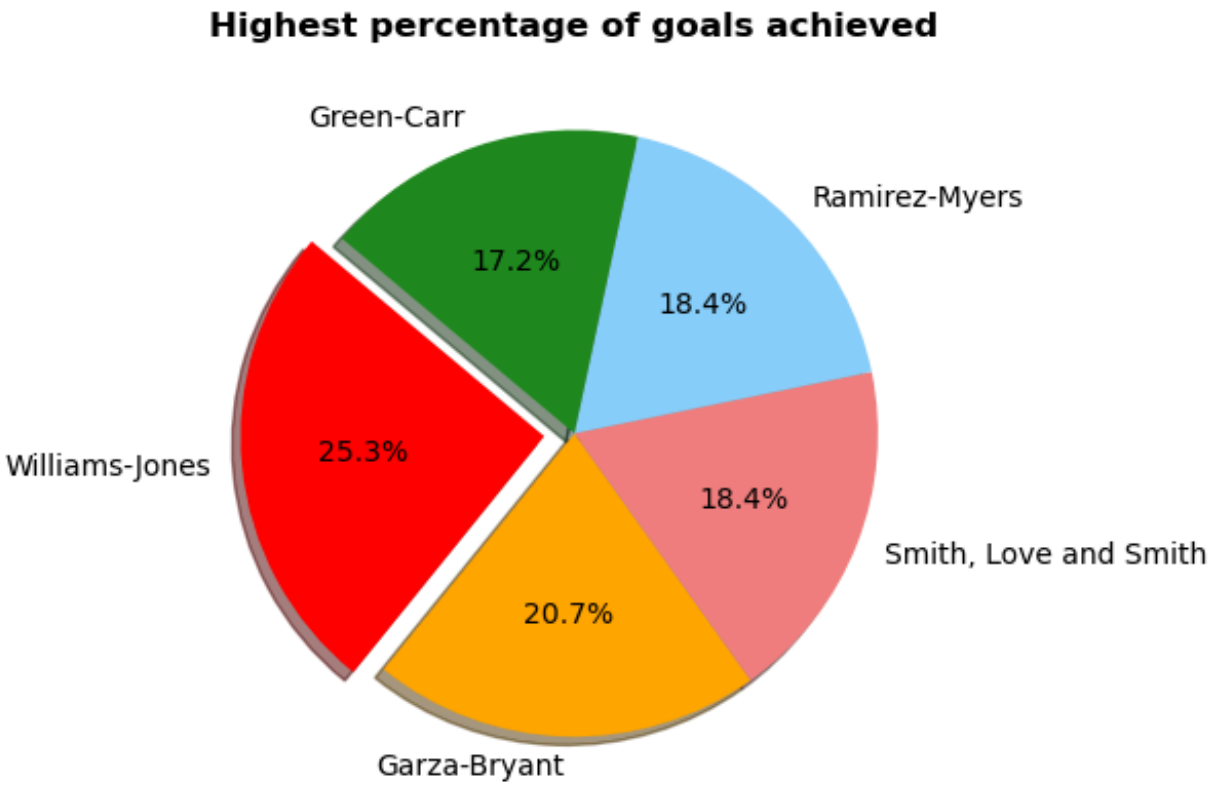
---

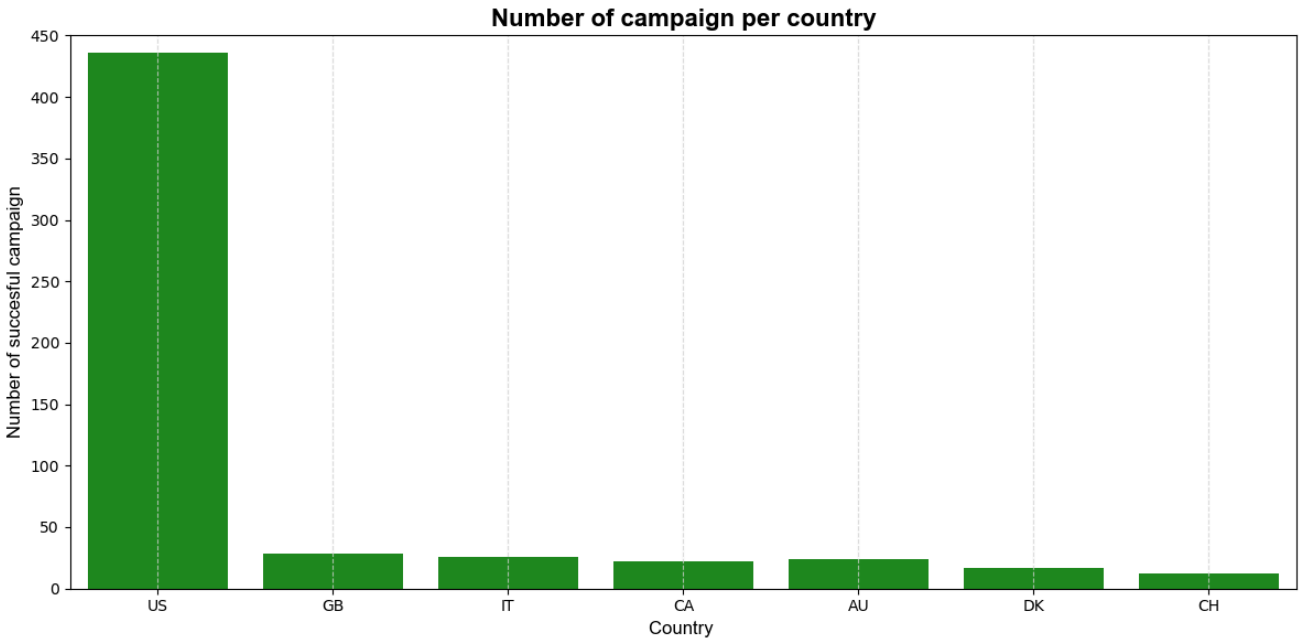
**Productionalization Strategy: To bring this project into a production environment with nightly updates, the following steps are recommended:**

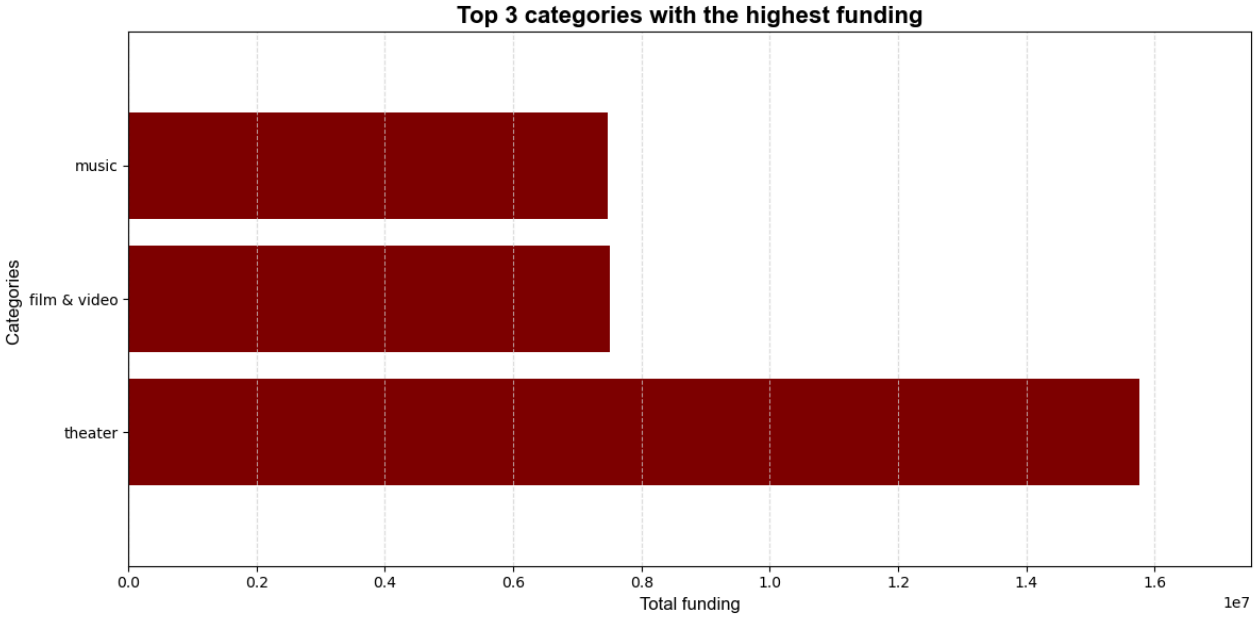
1. **Automated Data Pipeline:**
    - **Use Python scripts with scheduling tools like Apache Airflow to extract and transform new data daily.**
    - **Load updated data into the PostgreSQL database using efficient batch processing.**
2. **Data Quality Monitoring:**
    - **Implement automated checks to ensure data accuracy and completeness.**
3. **Dashboard Integration:**
    - **Develop interactive dashboards using tools like Tableau or Power BI for real-time insights.**
4. **Cloud Deployment:**
    - **Deploy the solution on cloud platforms like AWS or Azure to ensure scalability and accessibility.**

**By implementing these strategies, the project can scale to accommodate ongoing data influx and provide continuous insights for stakeholders.**




Highest percentage of goals achieved
Green-Carr
Ramirez-Myers
17.2%
18.4%
Williams-Jones
25.3%
18.4%
Smith, Love and Smith
20.7%
Garza-Bryant


Number of campaign per country
Number of succesful campaign
0
50
100
150
200
250
300
350
400
450
US
GB
IT
CA
AU
DK
CH
Country


Top 3 categories with the highest funding
Categories
music
film & video
theater
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1e7
Total funding