

## 1. Dataset Overview

Dataset: DSI Kickstarter Scrape Dataset

### **After data cleaning and preprocessing:**

Total Projects: 42,217 U.S. based campaigns

Features: 124 after one-hot encoding

### **Data Cleaning Performed:**

- Removed rows with missing values (including: location, pledged amounts, reward levels)
- Filtered to include only U.S. locations using state pattern matching
- Removed outliers using winsorization
- One-hot encoded categorical variables (category, subcategory, state)
- Extracted temporal features (month from funded date)

## 2. Success vs. Failure Rates

### **Overall Project Outcomes**

- **Successful Projects:** Projects that met or exceeded their funding goal
- **Failed Projects:** Projects that did not reach their funding goal
- The data shows an imbalance between successful and failed projects, revealing that many Kickstarter campaigns struggle to reach their funding targets.

### **Funding Achievement Distribution**

This distribution reveals that Kickstarter, for many projects, partial funding is common but success requires meeting the full goal.

## 3. Category Analysis

### **Top 5 Most Popular Categories:**

- Film & Video: 11,749 projects (27.8% of total)
- Music: 10,391 projects (24.6% of total)
- Publishing: 4,360 projects (10.3% of total)
- Theater: 2,359 projects (5.6% of total)
- Design: 1,637 projects (3.9% of total)

### **Key Insights**

- Film & Video and Music dominate the platform, accounting for over half of all projects
- Creative categories are significantly more represented than technology or product-based categories
- The platform seems to attract content creators and artists seeking funding for creative works

## 4. Geographic Distribution

The analysis identified the states with the highest number of Kickstarter projects which are states with larger populations and established creative industries.

### **Success Rates by State**

Success rates varied significantly by states, suggesting that:

- Geographic location may influence campaign success
- Communities play a role in funding

Note: Success rate does not directly correlate with project volume- some states with fewer projects show higher success rates, possibly indicating more selective or prepared campaigns.

## 5. Key Metrics Analysis

### **Goal Amounts (After Outlier Removal):**

- Mean Goal: \$8,735
- Median Goal: \$4,000
- Range: \$200 to \$100,000
- Distribution: Right-skewed, with most projects requesting modest amounts

### **Pledged Amounts:**

- Mean Pledged: \$3,764
- Median Pledged: \$1,285
- Range: \$0 to \$45,901
- Average pledged amount is significantly lower than average goal, revealing many projects fall short.

### **Backers:**

- Mean Backers: 52 people
- Median Backers: 23 people
- Range: 0 to 605 backers
- Distribution: Highly right-skewed, most projects have modest backer counts

### **Campaign Duration:**

- Mean Duration: 39.6 days
- Median Duration: 31.5 days
- Range: 10 to 90 days
- Most campaigns run for approximately one month

## 6. Comparing Successful vs. Failed Projects

### Average Goal Amount

- Successful Projects: Lower average goals
- Failed Projects: Higher average goals

### Average Number of Backers

- Successful Projects: Significantly higher backer counts
- Failed Projects: Fewer backers
- Community engagement and reach are critical success factors

### Average Campaign Duration

Successful vs. Failed: Relatively similar durations

Campaign length alone is not a strong predictor of success; quality of campaign and network matter more

## 7. Correlation Insights

- Strong positive correlation expected between pledged amount and number of backers
- Goal amount shows interesting relationships with success outcomes
- Duration appears to have weak correlation with most success metrics
- Funded month (seasonal patterns) may show interesting temporal trends

## 8. Key Takeaways for Predictive Modeling

Based on these initial insights, the following factors appear most relevant for predicting campaign success:

- Goal Amount: Lower, realistic goals correlate with higher success rates
- Category: Certain categories may have inherently different success rates
- Geographic Location: State-level differences exist in success patterns
- Backer Engagement: Number of backers is a strong indicator of success
- Funding Ratio: Understanding how much projects typically exceed or fall short of goals

This initial analysis reveals that success is influenced by goal-setting, category choice, geographic factors, and community engagement. The data shows clear patterns that can inform predictive modeling, with goal amount and backer count being important factors.