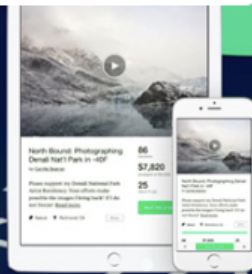


Explore projects, everywhere

Check out Kickstarter for iPhone and iPad — now with Apple Pay!

Download it now



KICKSTARTER

PREDICTING WHETHER A PROJECT IS FUNDED AND THE AMOUNT OF FUTURE FUNDING

Kim Kraunz

The year in

Kickstarter

—2014—

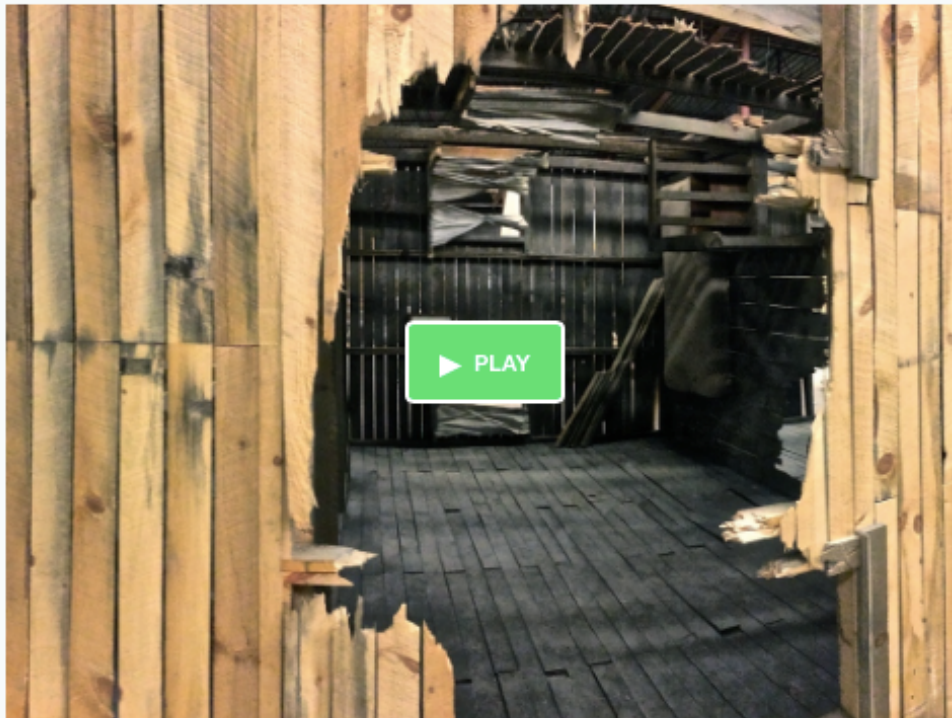
3.3 million people
from nearly every country on Earth

Pledged more than half a billion dollars

(that's \$1,000 a minute)

WISE BLOOD

by The Soap Factory



A classical composer, an artist, a brass band, a punk rocker and a folk singer collaborate in a joyously dark, immersive new opera.

📍 Minneapolis, MN

🎭 Performance Art

Share this project

169

backers

\$18,673

pledged of \$20,000 goal

3

days to go

Back This Project

★ Remind me

This project will only be funded if at least \$20,000 is pledged by Sat, Apr 11 2015 10:56 AM EDT.

The Soap Factory

🕒 First created | 0 backed

🌐 soapfactory.org

[See full bio](#) [Contact me](#)



SCOPE OF PROJECT

- Can we predict, using creator entered parameters, whether a project will be funded or not?
 - Can we predict the amount of pledged funding for each main category for future years?
-

PROJECT OUTLINE

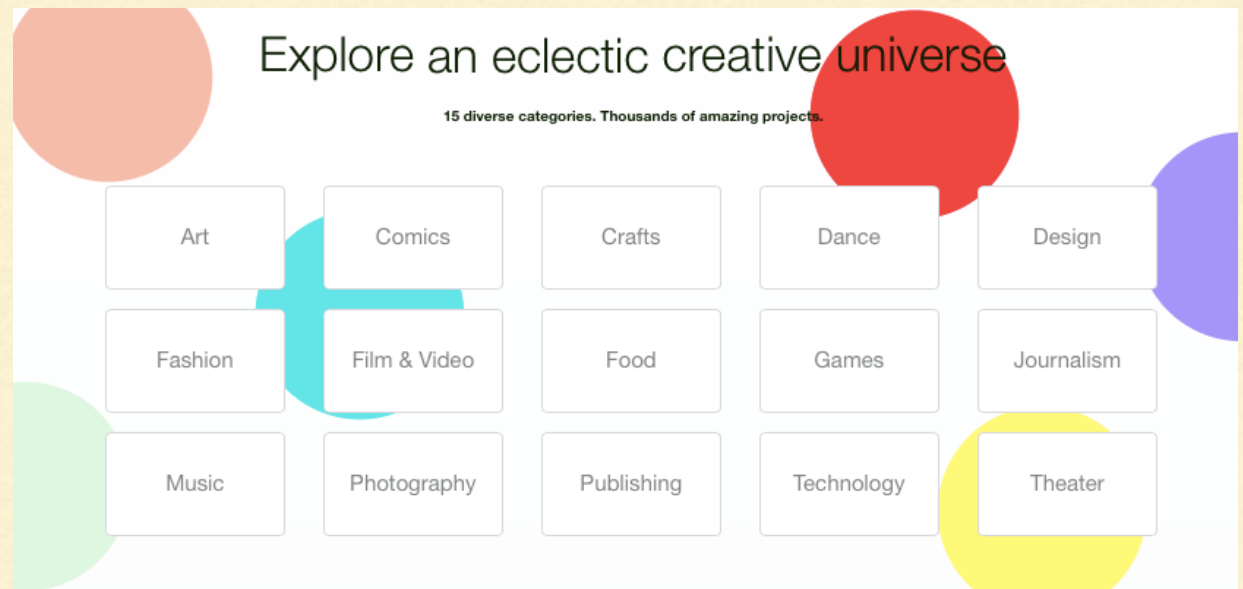
- Data Scrubbing
 - Data Manipulation
 - Data Exploration
 - Part I: Predicting whether or not a project is funded
 - Part II: Predicting how much will be pledged in the future
 - Next Steps
-

THE DATA

- Obtained from someone who had scraped the Kickstarter website

- Variables used:

- Goal
- Pledged
- Deadline
- Main category
- Sub category
- Number of project backers
- Currency type
- Blurb (text)



DATA SCRUBBING & MANIPULATION

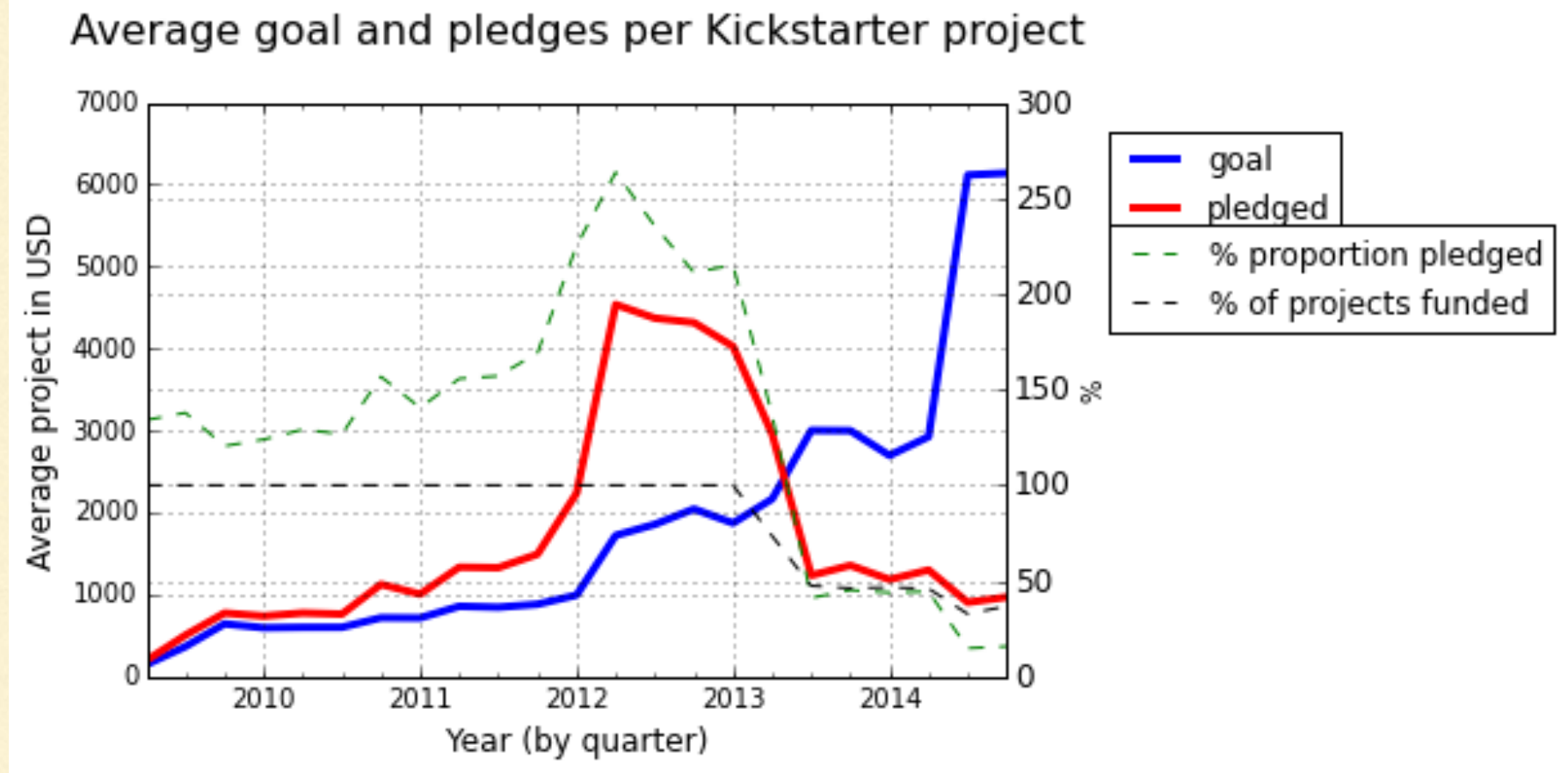
- Data scrubbing

- Fixed data alignment
- Fixed main categories and sub categories
- Fixed data types

- Data manipulation

- Converted pledged and goal values to USD
 - Converted dates to quarters using `.resample()`
 - Created a time series lag component based on proportion funded in previous quarter
 - Created a polarity and subjectivity rating using `.TextBlob()`
-

DATA EXPLORATION

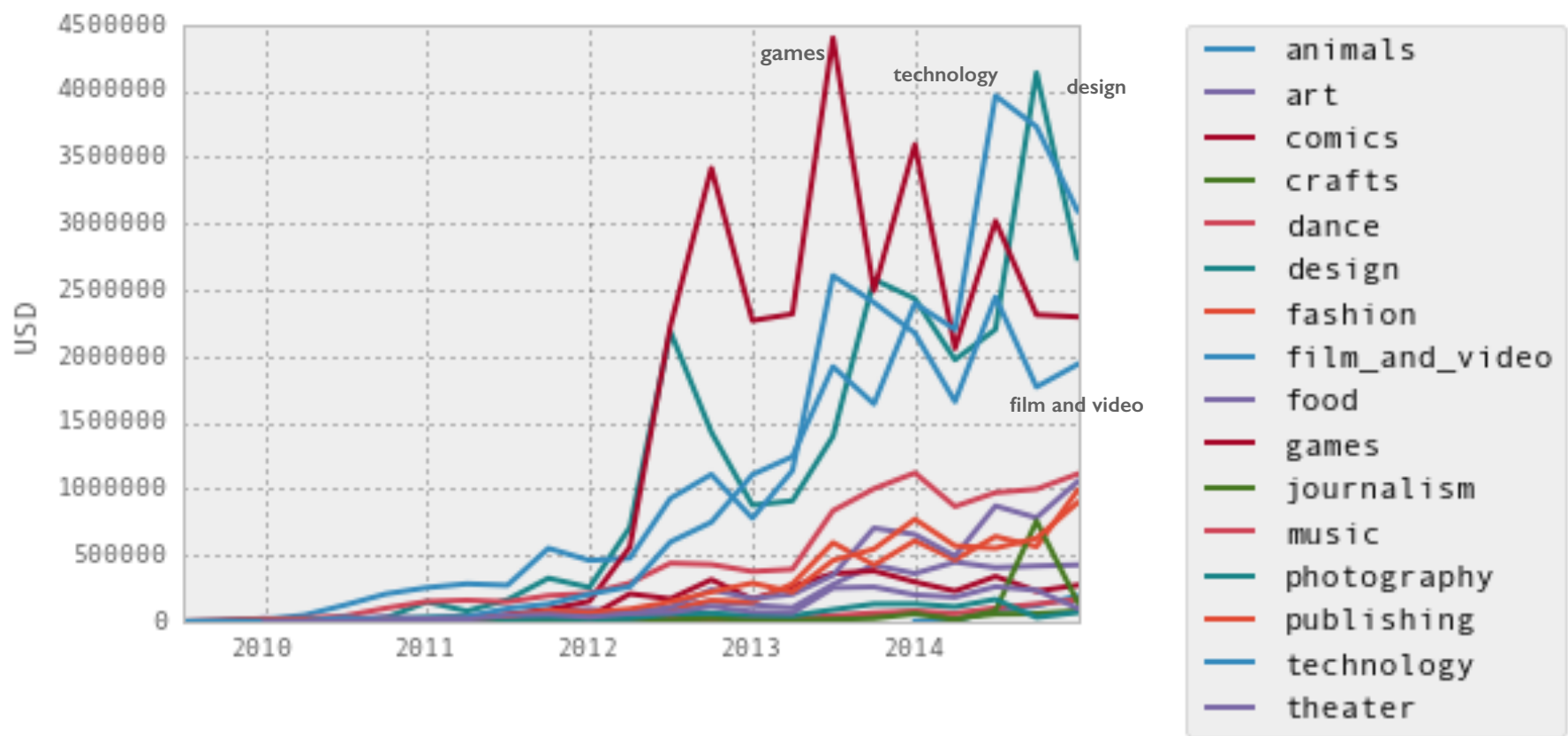


DATA EXPLORATION

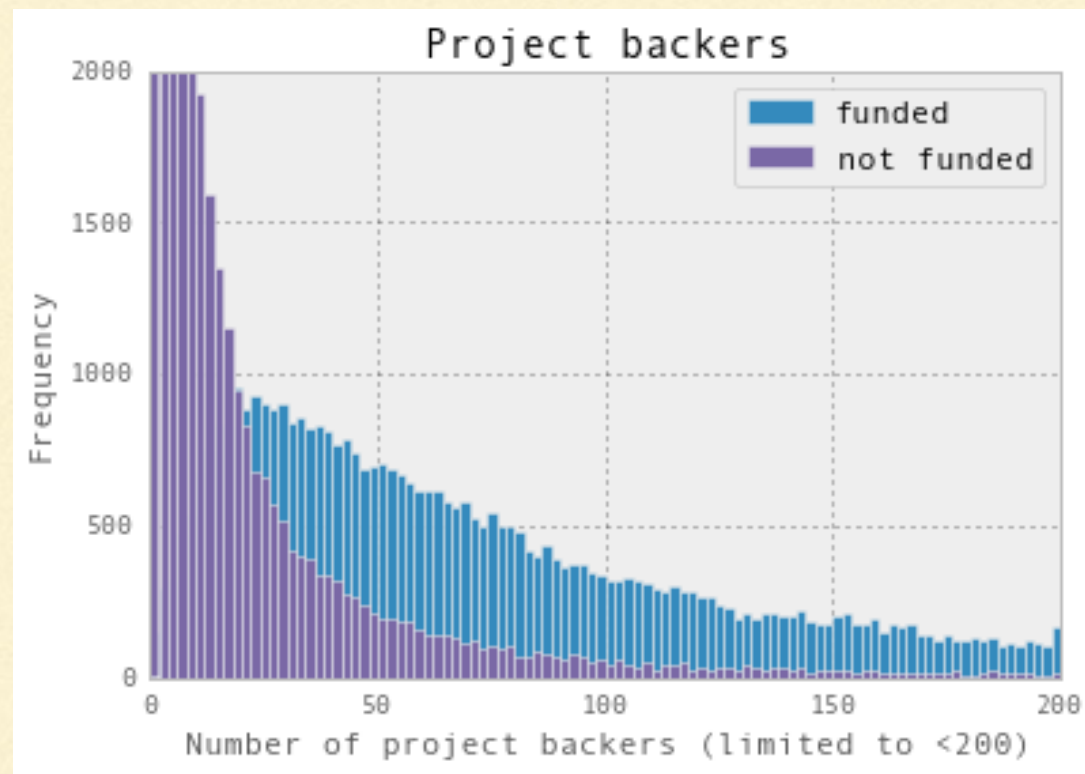


DATA EXPLORATION

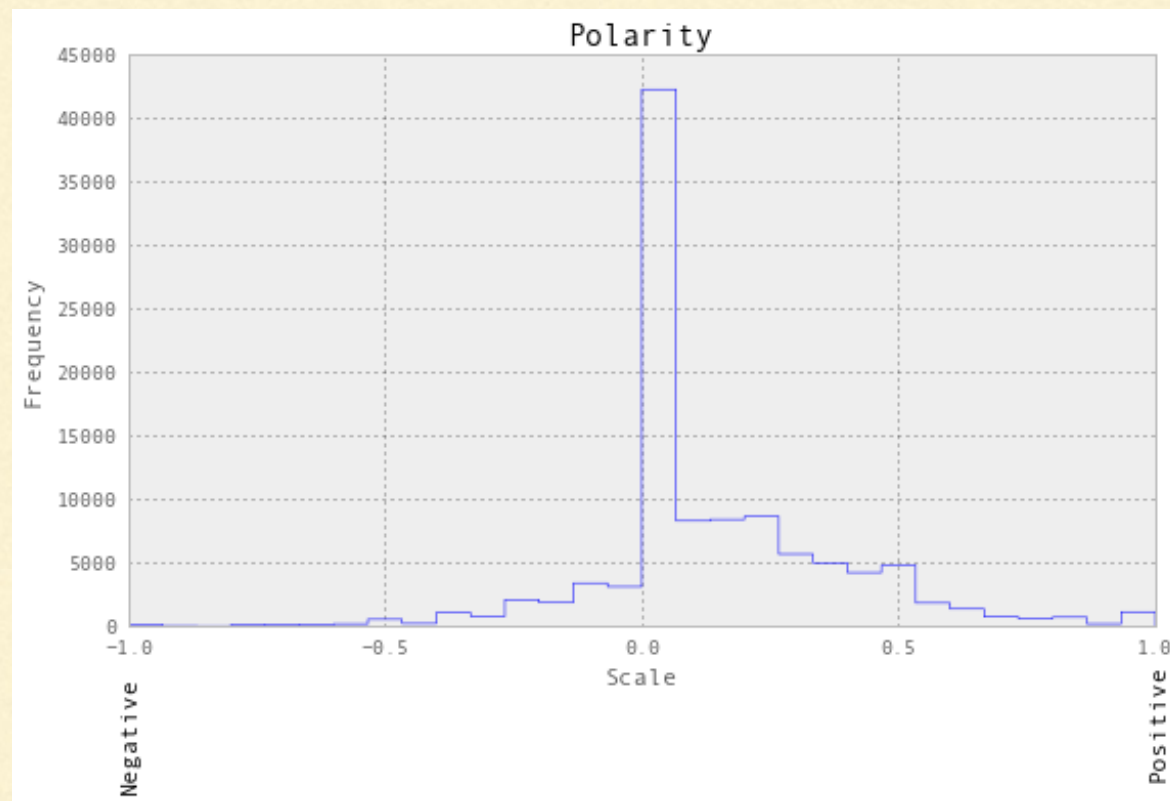
Sum of Kickstarter pledges in USD by main category



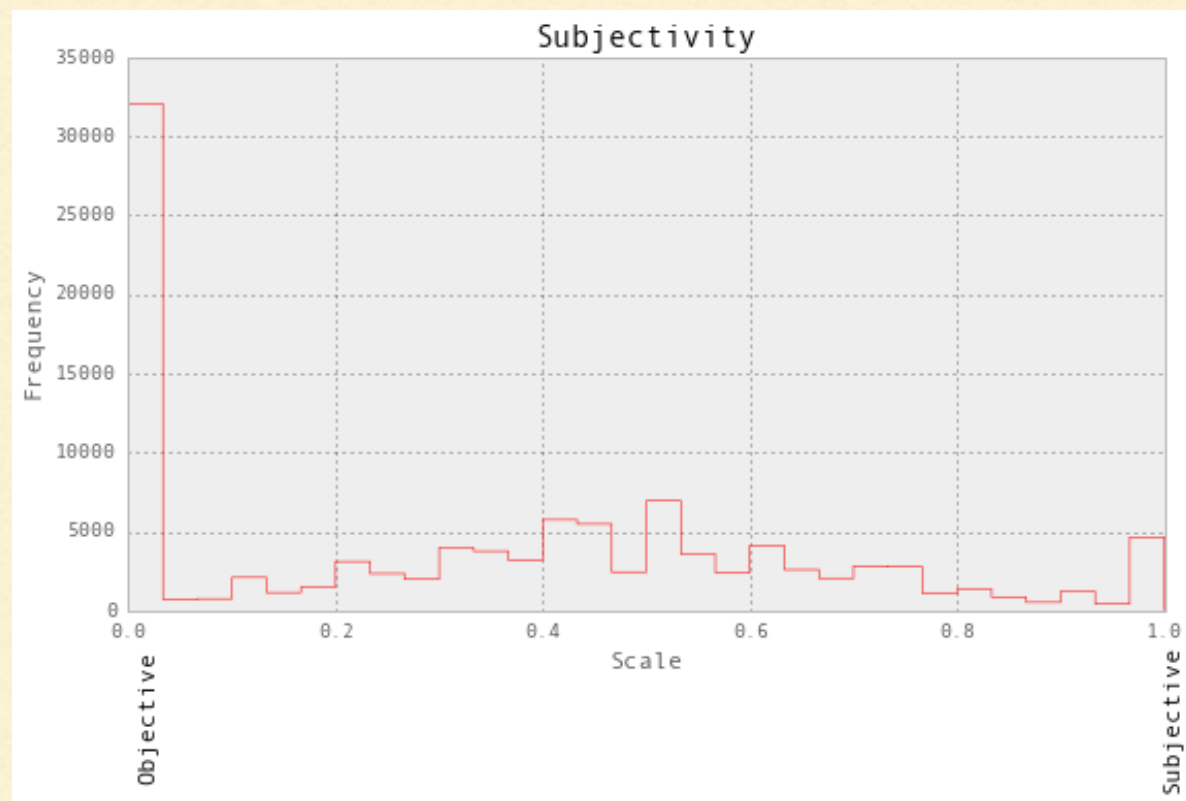
DATA EXPLORATION



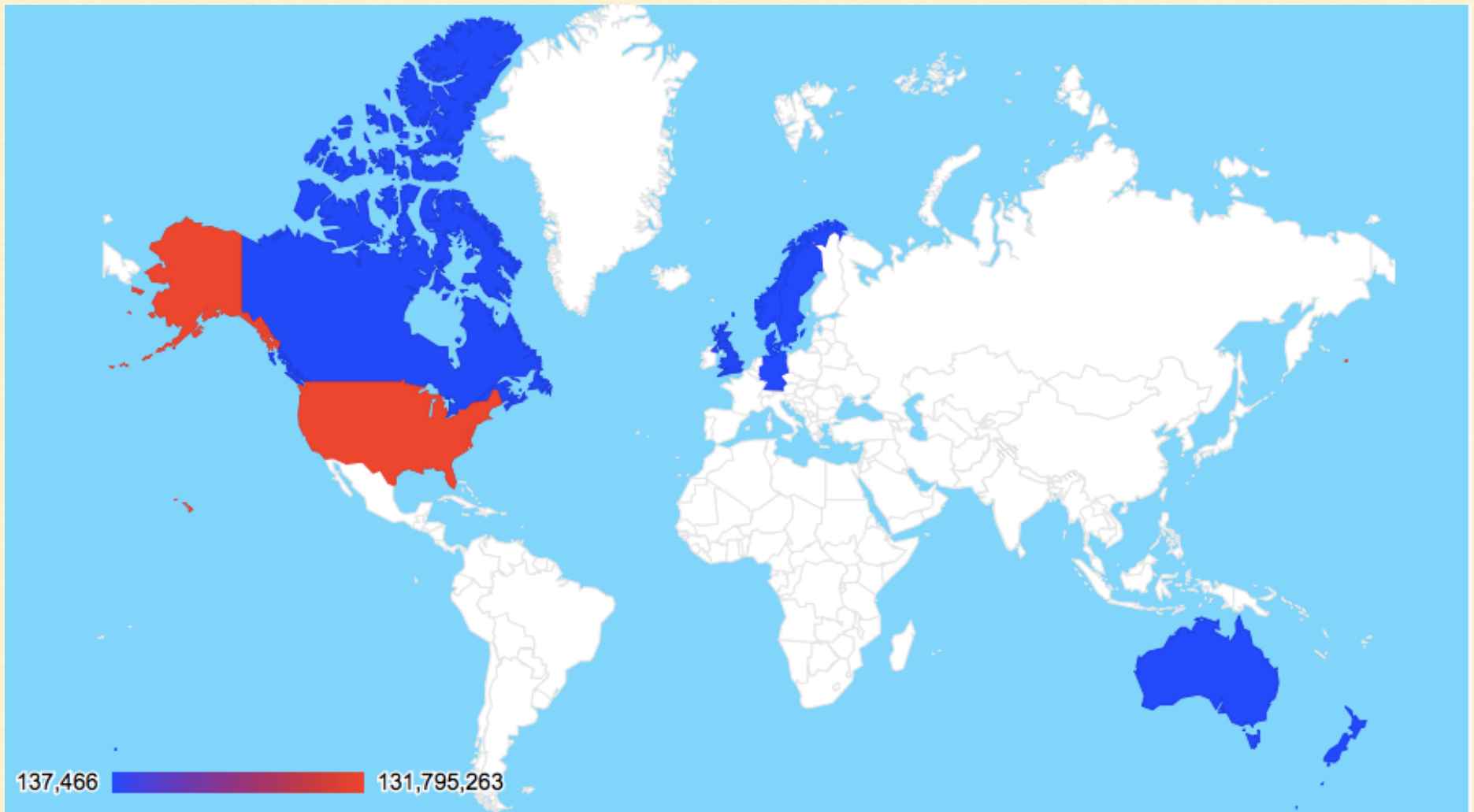
DATA EXPLORATION



DATA EXPLORATION



KICKSTARTER PLEDGES (USD) BY COUNTRY CURRENCY



WILL THIS PROJECT BE FUNDED?

- Tested the Decision Tree, Random Forest, and Extra Forest models for accuracy to predict funding.
 - Used projects with a deadline from Q2-2009 to Q4-2014
 - Random Forest performed the best- 91.0%
 - Tested the Random Forest model on Q1-2015 data- 92.0%
 - The most important features were: # of project backers, project goal (USD), proportion of projects funded in the previous quarter, currency(USD), currency(unknown), main category (music)
-

WILL THIS PROJECT BE FUNDED?

This felt like cheating!



Can we make a prediction without using the
number of backers?

(especially since I found using RFE that I could be 96% accurate with
just knowing the goal(USD) & number of backers)

WILL THIS PROJECT BE FUNDED?

- Tested the Decision Tree, Random Forest, and Extra Forest models for accuracy to predict funding.
 - Included the sentiment and polarity ratings and a lagged time series of proportion of projects funded in the previous quarter
 - Random Forest performed the best - 60.0%
 - Performed RFE and increased accuracy to- 93.0%
 - Tested the Random Forest model on Q1-2015 data- 63.4%
 - The most important features were: goal (USD), polarity, & subjectivity
-

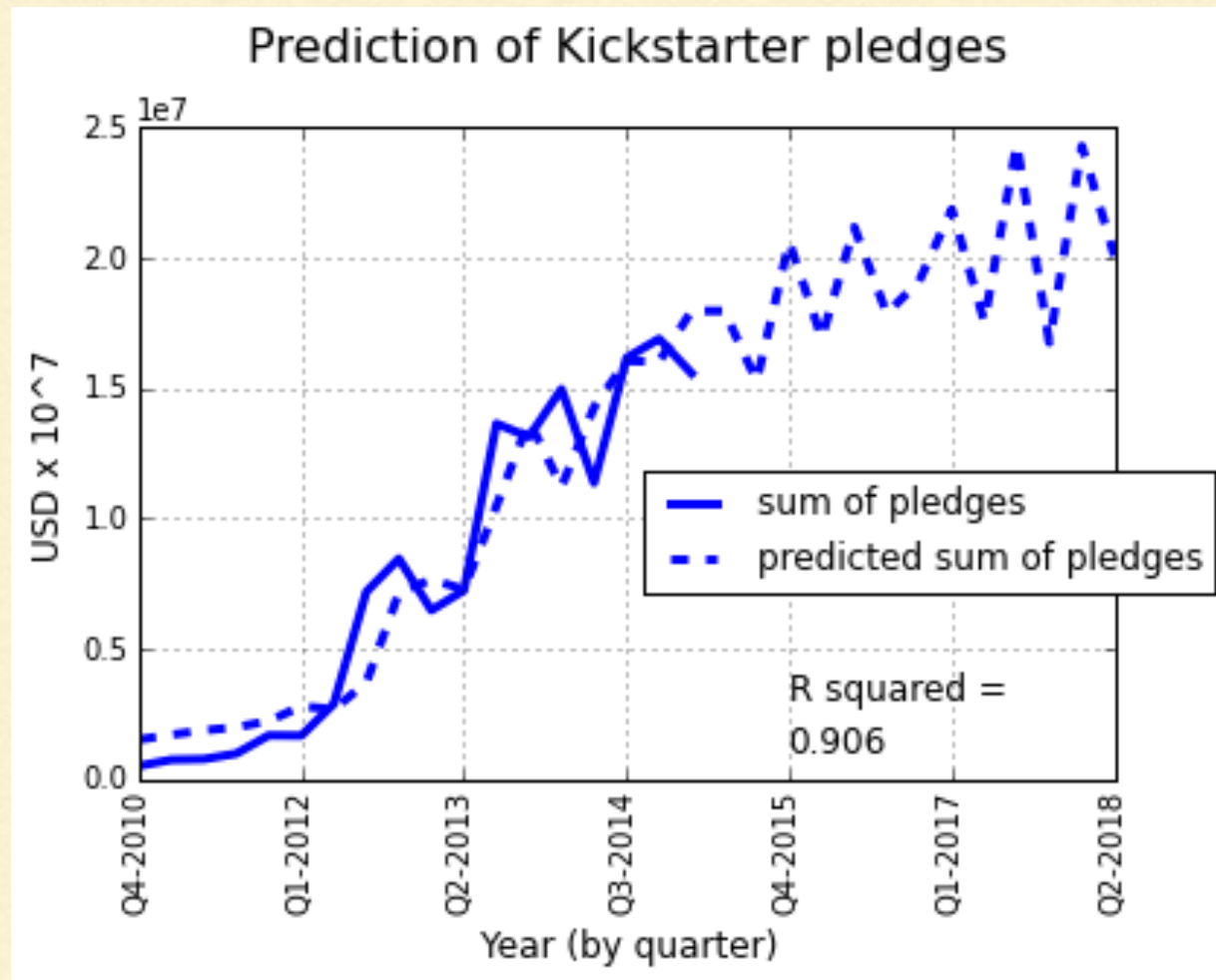
HOW MUCH WILL BE PLEDGED IN THE FUTURE?

- Created a linear regression model using the sum of pledges from the previous four - six quarters to predict funding for the current quarter.

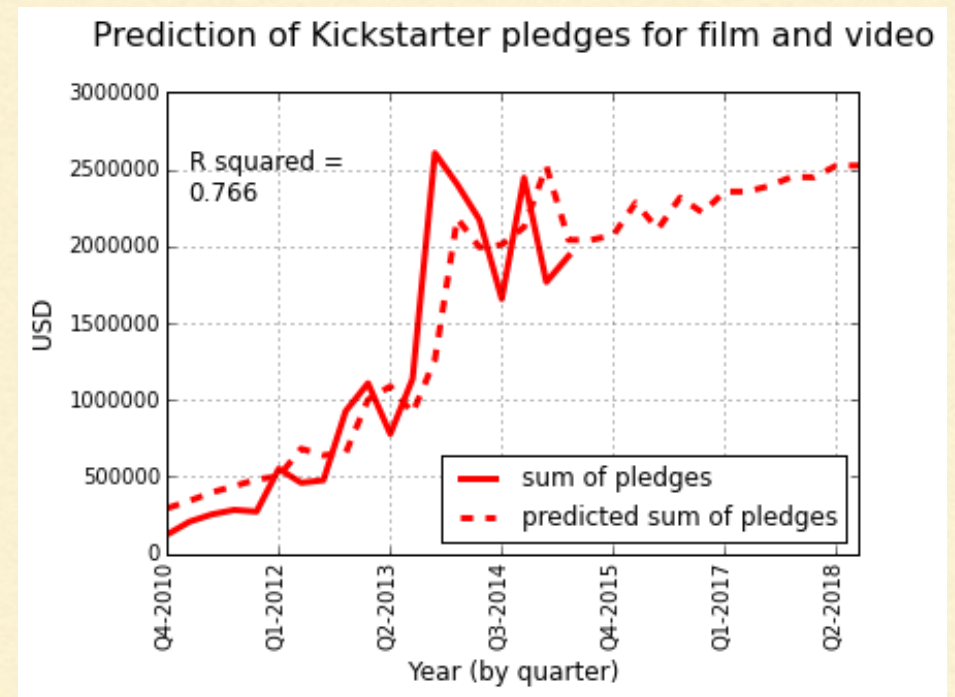
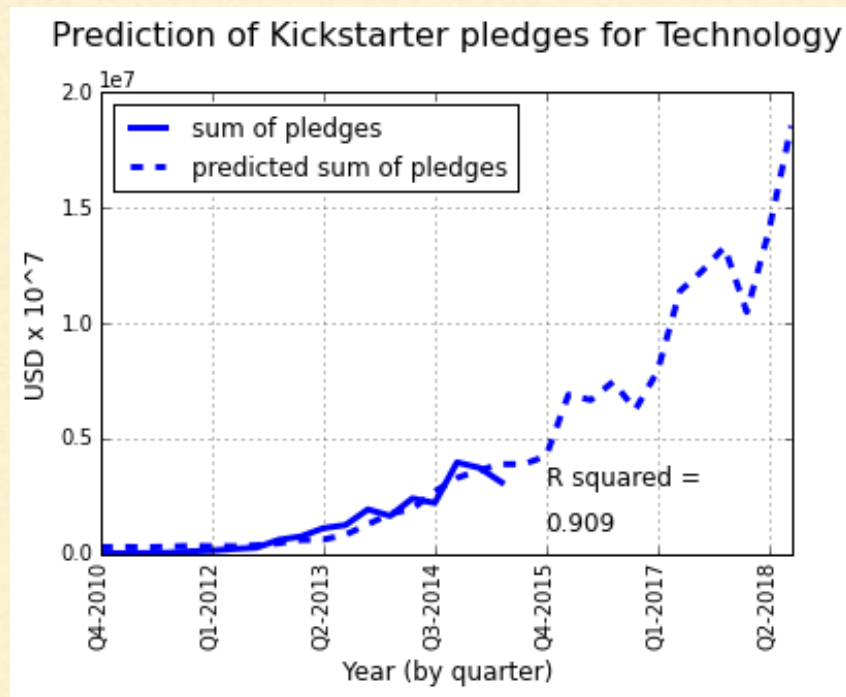
- Example:

DATE	X1	X2	X3	X4	X5	X6	Y
Q2-2009	150	200	175	300	280	340	375
Q3-2009	200	175	300	280	340	375	365
Q4-2009	175	300	280	340	375	365	395
Q1-2010	300	280	340	375	365	395	410

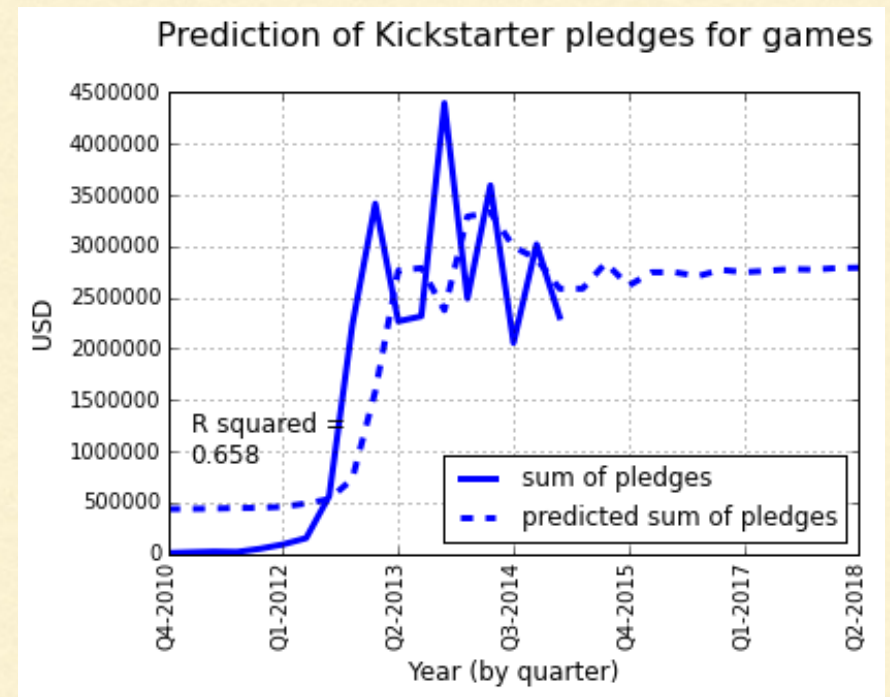
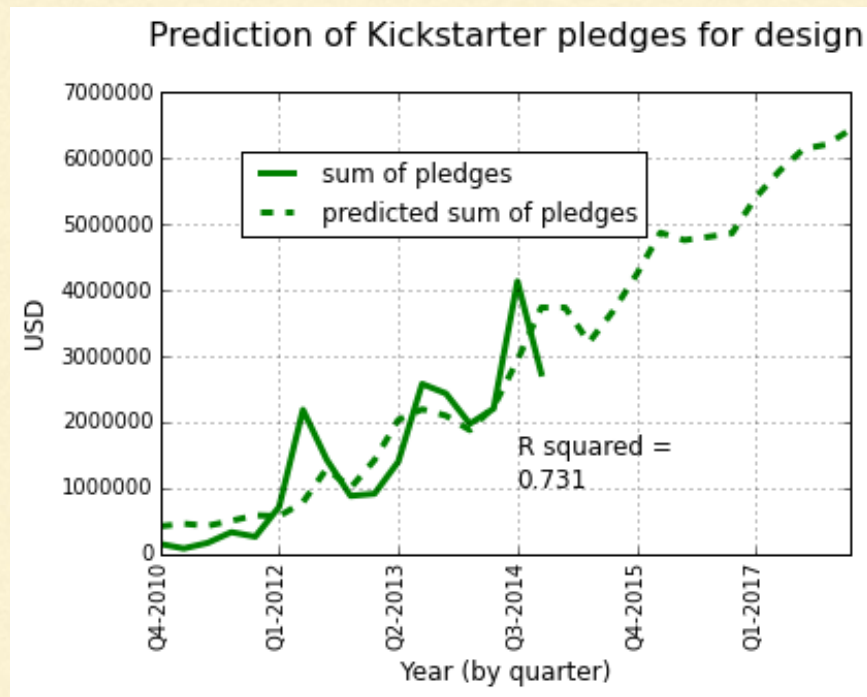
HOW MUCH WILL BE PLEDGED IN THE FUTURE?



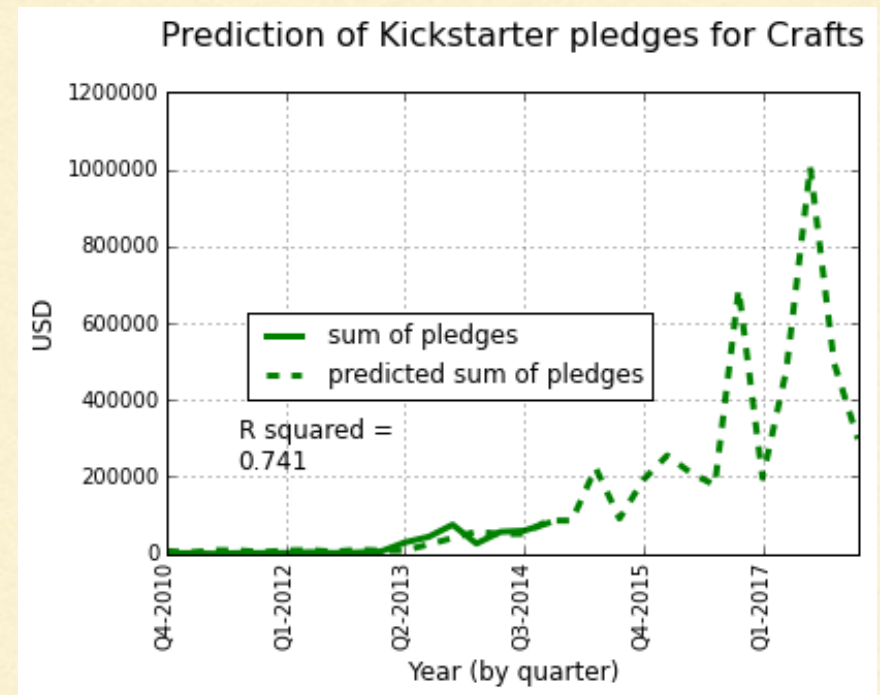
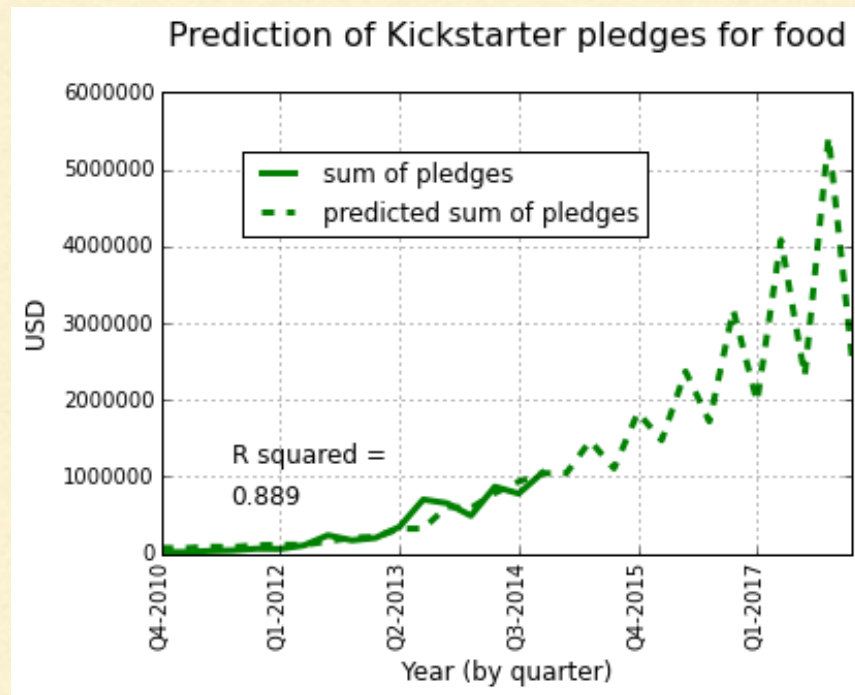
HOW MUCH WILL BE PLEDGED IN THE FUTURE?



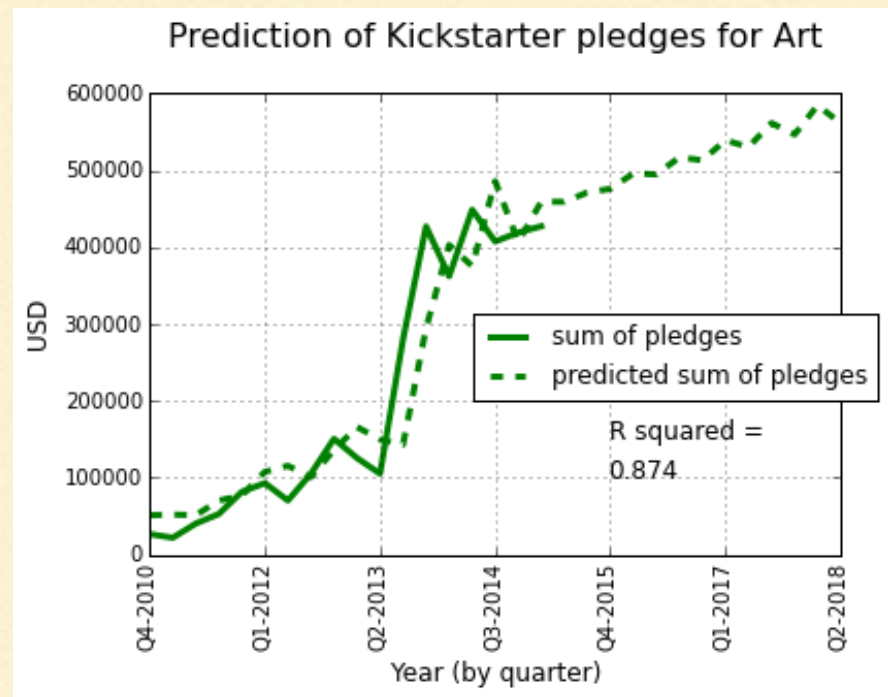
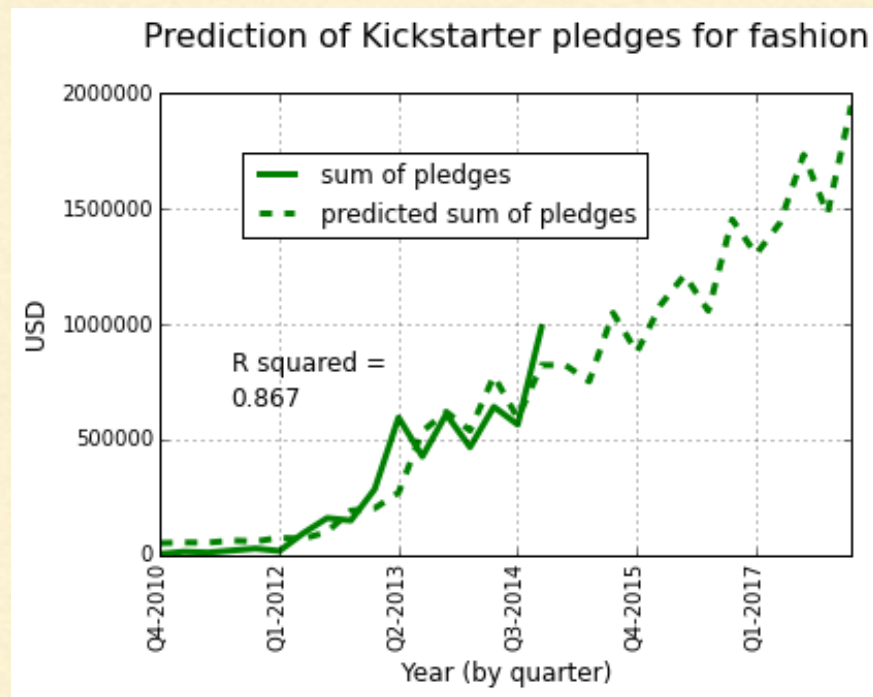
HOW MUCH WILL BE PLEDGED IN THE FUTURE?



HOW MUCH WILL BE PLEDGED IN THE FUTURE?



HOW MUCH WILL BE PLEDGED IN THE FUTURE?



NEXT STEPS

- **Data manipulation:** use funding \$'s instead of pledged \$'s
 - **Additional features (within this dataset):**
 - % previously funded by category
 - **Creators:** analyze for success on previous projects
 - **Photographs:** analyze for features
 - **Duration of project posting**
 - **Additional features (with expanded dataset):**
 - **Location of creator**
 - **Rewards:** do incentives increase probability of funding?
 - **Twitter:** do appeals in the first week increase probability of funding?
-