**Outline:**

Objective: Predict probability that a project will be ‘exciting’

**Dataset:**

1\_projects.csv: project info - train/test

2\_essays.csv: project text - train/test

3\_resources.csv: resource info - train/test

4\_donations.csv: donation info - train

5\_outcomes.csv: outcome info - train

There are several criteria for ‘is\_exciting’

To see patterns for the criteria, we plotted them out.

Observations:

1. Exciting projects only occur after 2010-04-14 (Presumably some of the criteria were not kept in track before that)
2. Has seasonality (When looking at them by months) -> Model should take seasonality into account.
3. Should validate on the most recent data as the test data is after 2014-01-01. (Use 2013 data for validation)

Preprocess:

Issues:

- There are empty fields (Find appropriate method to fill them in)

-

Projects:

Essays: