

# DS3 Final Project Presentation

Predicting Investment Crowdfunding Success

Jason Fitts

August 11, 2016

# Predicting Investment Crowdfunding Success: Outline

- Project Summary
- Modeling Insight
- Modeling Approach
- Results
- Conclusion
- Next Steps

# Predicting Investment Crowdfunding Success:

## Project Summary

- Traditional donation-based crowdfunding is hard and results have been poor. Investment crowdfunding is return-based and no less difficult for those raising capital.
- The JOBS Act (2012) gave rise to Title II 506(c) offerings where accredited investors ('wealthy') can fund issuing companies with no maximum capital raise limitations via debt, equity or other investment vehicles.
- Crowdfunding Platforms (CFPs) can participate in the process and rely on Form D (SEC document) for data about these transactions listed on their portals.

# Predicting Investment Crowdfunding Success:

## Project Summary

- There are thousands of CFPs in various verticals which make it difficult for issuers to determine the best place to launch a round. As such, the question of whether a CFP's strength impacts issuer success was intriguing.
- Crowdfunding providers, advisors and other stakeholders are working to improve the ecosystem. CrowdWatch is market data on private Tier II active and closed listings in the U.S.
- Prior to this project I purchased a dataset with details on closed deals from October 2013 to November 2015 which had 3,020 rows with 79 columns in structured .csv format.

# Predicting Investment Crowdfunding Success: Project Summary

- It became apparent there was a significant amount of missing data. Not immediately evident was the realization ‘traditional’ predictors (i.e. number of contributors/investors) was not included.
- Below is a very small sample of the mostly categorical variables in the original dataset. As such, Success and CFP\_Strength variables were created for all records and ultimately only two continuous variables were used in the creation of a logistic regression model.

CFIIN	PIPR_Name	Portal_Name	Sector	Subsector	Industry	City
US1000000006	Connxys Technology, Inc.	Angellist	Technology	Computer Software	Location-Based Services	Santa Barbara
US100000000F	Style Struck	Angellist	Services	E-Commerce	E-Commerce, Apparel & Accessories	Boynton Beach
US3000000010	CREATE.IT	Angellist	Services	Marketing	Marketing, General	Los Angeles

# Predicting Investment Crowdfunding Success: Modeling Insights

- Initial EDA confirms hypothesis that investment crowdfunding success is poor and belief CFP strength was a primary driver of these results:

```
#### Exploratory Analysis:
```

```
#### Overall Crowdfunding Success  
df.Success.value_counts()
```

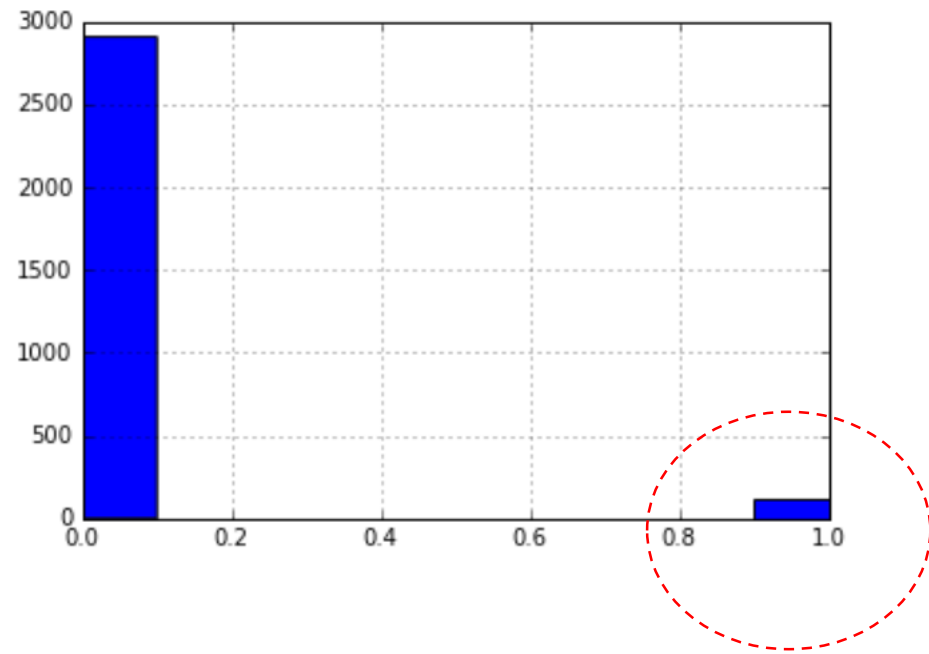
```
0    2906  
1     114  
Name: Success, dtype: int64
```

```
114 / (114.0 + 2906.0)  
0.03774834437086093
```

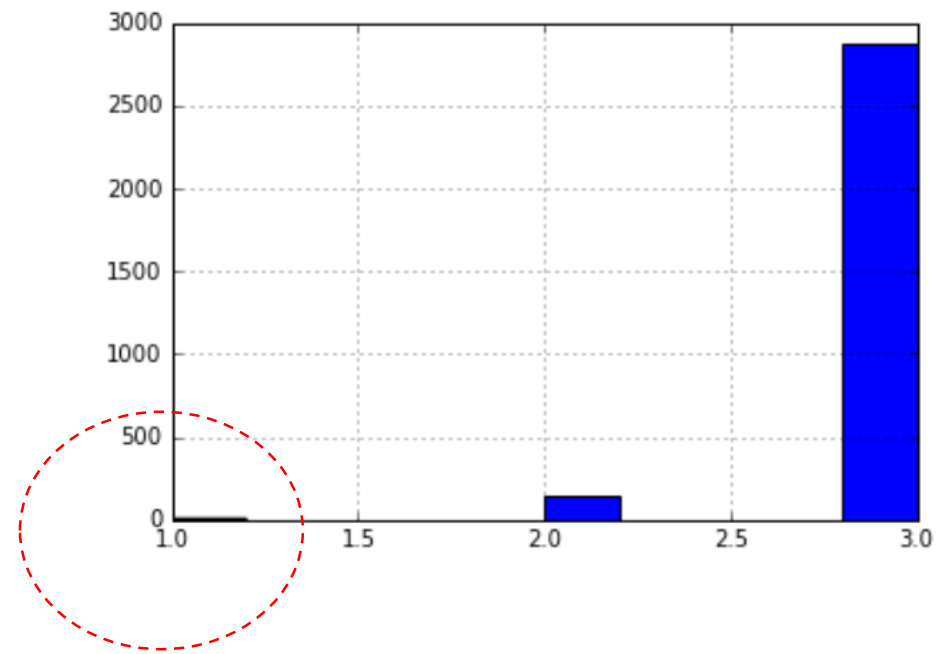
	Success	CFP_Strength
count	3020.000000	3020.000000
mean	0.037748	2.948675
std	0.190619	0.232392
min	0.000000	1.000000
25%	0.000000	3.000000
50%	0.000000	3.000000
75%	0.000000	3.000000
max	1.000000	3.000000

# Predicting Investment Crowdfunding Success: Modeling Insights

```
df['Success'].hist()  
plt.show()
```



```
df['CFP_Strength'].hist()  
plt.show()
```

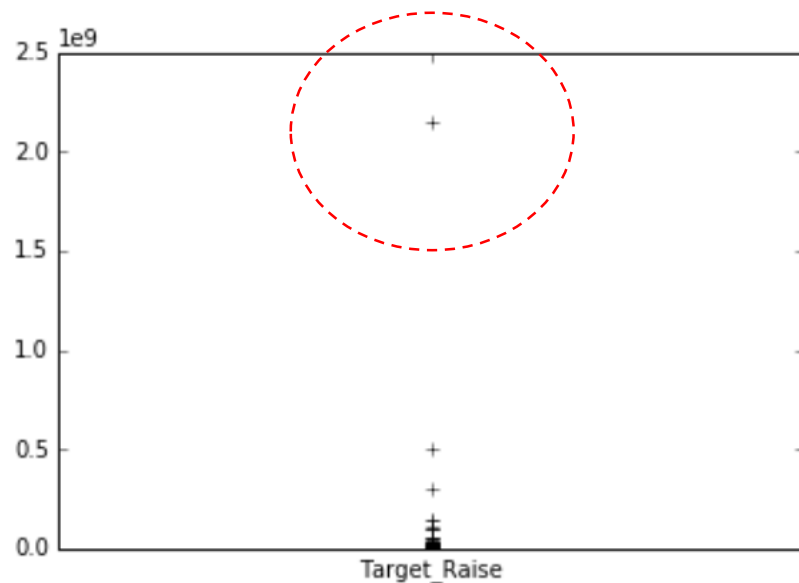


# Predicting Investment Crowdfunding Success: Modeling Insights

- Along with CFP strength I speculated that the issuers target raise amount and their pre issue market cap could predict future crowdfunding success.

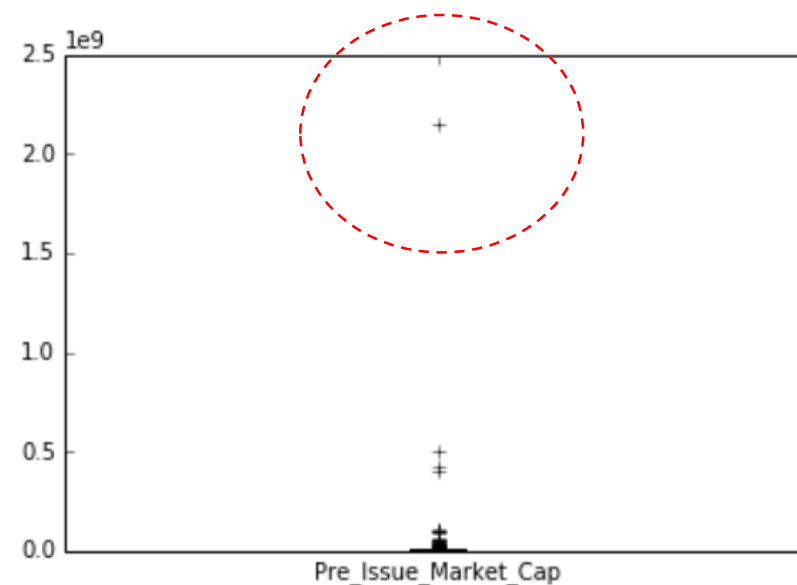
```
df['Target_Raise'].plot(kind='box')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0xacecf60>
```



```
df['Pre_Issue_Market_Cap'].plot(kind='box')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0xd8c8f98>
```





# Predicting Investment Crowdfunding Success: Modeling Approach

- Initially, an odds ratio on CFP\_Strength\_1 ('high strength') was done which suggested those issuers using a high strength platform have 81x the odds of success than those companies who do not.

```
odds_exposed = 6.0/2.0  
odds_exposed
```

```
3.0
```

```
#### Calculate odds of success if company used an industry leading platform
```

```
odd_unexp = 108.0/2904.0  
odd_unexp
```

```
0.0371900826446281
```

```
#### Calcualte the odds ratio
```

```
OR = odds_exposed/odd_unexp  
OR
```

```
80.66666666666667
```

# Predicting Investment Crowdfunding Success: Modeling Approach

- A logistic regression model was used given desire to predict Success with CFP\_Strength, Target\_Raise and Pre\_Issue\_Market\_Cap as predictors using Python logit function:

```
logit = sm.Logit(data['Success'], data[train_cols])  
result = logit.fit()
```

```
Optimization terminated successfully.  
    Current function value: 0.129138  
    Iterations 8
```

## Logit Regression Results

```
=====
```

Dep. Variable:	Success	No. Observations:	3020
Model:	Logit	Df Residuals:	3015
Method:	MLE	Df Model:	4
Date:	Thu, 11 Aug 2016	Pseudo R-squ.:	0.1965
Time:	13:13:51	Log-Likelihood:	-390.00
converged:	True	LL-Null:	-485.38
		LLR p-value:	3.639e-40

```
=====
```

# Predicting Investment Crowdfunding Success: Results

- The logistic regression predictive model results confirms the problem using a derived continuous variable (CFP\_Strength) and only two continuous variables which range from 0 to more than 3,000,000.

```
#### Make predictions on the enumerated dataset
```

```
combos['Success_pred'] = result.predict(combos[train_cols])  
print combos.tail()  
len(combos)
```

Success\_pred

0.999756

0.999964

0.999799

0.996156

0.999783

- When Total\_Raise and Pre\_Issue\_Money\_Cap are the same there is essentially no difference in investment crowdfunding success between high and low strength CFPs; despite OR calculated.

# Predicting Investment Crowdfunding Success:

## Conclusion

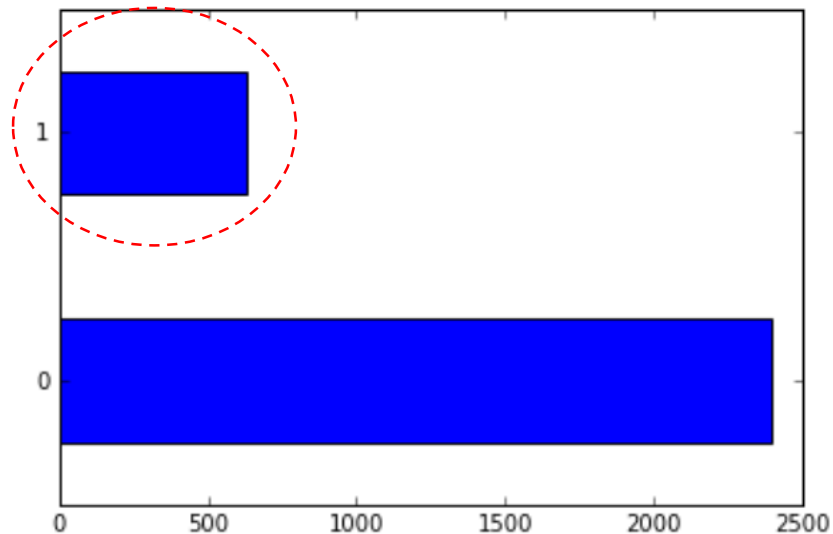
- Based upon the above the results a key conclusion is that the CFP\_Strength variable should not be created using the Total\_Raised vs. Target\_Raise variables (should be uncorrelated metric.)
- When the CFP\_Strength variable was then used in conjunction with the Target\_Raise variable the results were unexpected and not consistent with expectations or deemed credible.
- Going forward, the categorical variables would need to be implemented and transformed and combined with continuous variables of reasonable measure for use in the logistic regression model.

# Predicting Investment Crowdfunding Success: Next Steps

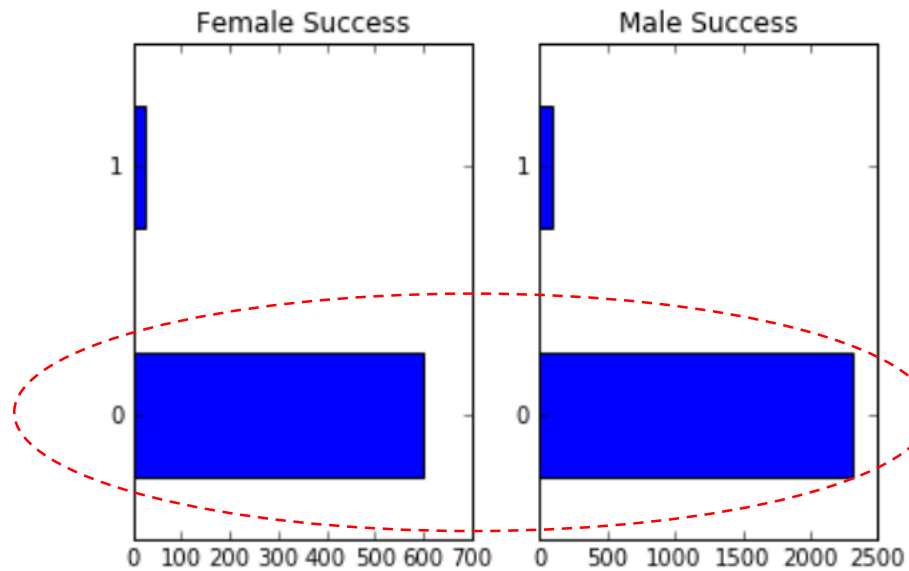
- In addition to including more variables including the number of investors I would like to examine the impact of gender and other demographics. Similarly, using decision trees or SVM could be useful.

```
df.Gender.value_counts().plot(kind='barh')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0xc720198>
```



```
<matplotlib.axes._subplots.AxesSubplot at 0xc295e48>
```



# DS3 Final Project Presentation

Questions?

# DS3 Final Project Presentation

Thank You

Dataset used was obtained for a fee from CrowdWatch Market Data

<https://www.crowdwatch.co/crowdwatch/>