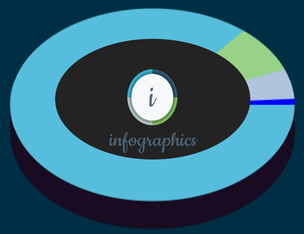
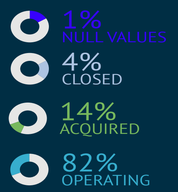
# ANALYSIS OF STARTUPS

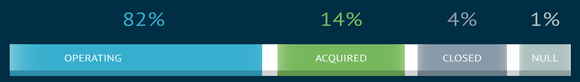
ABSTRACT:
The data was extracted from Crunchbase on February 2014. The dataset provides information about startup companies, investment, and acquisitions via Crunchbase.

Since an acquisition brings a lot of money to the founders and indicates that the startup succeeded in some way, this project explains which features increase the probability of an acquisition and predicts if a so far unseen startup will be acquired or not. To get the best possible results, multiple supervised learning algorithms such as Logistic Regression, Random Forest and Neural Networksvare applied after intense data preparation. Validation shows that Deep Learning has the best performance in this case.

A Project By:
MAXIMILIAN OTT
PRIYA MATNANI
SASCHA HAGEDORN



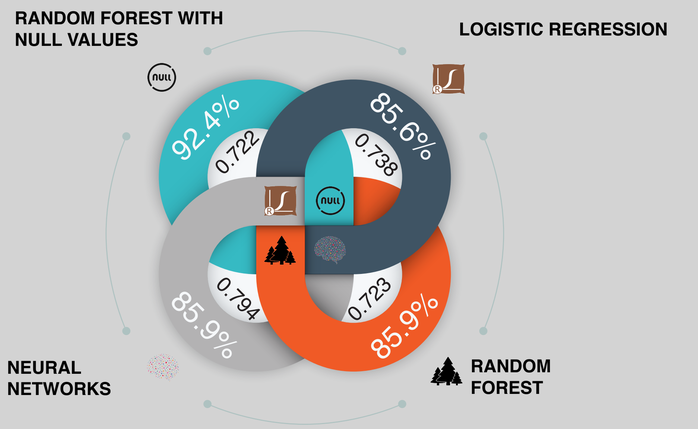



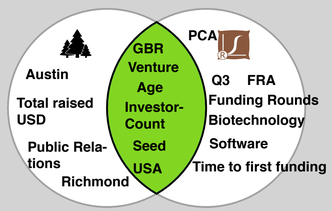

**** The balance observed might be a concern. Thus we also focused on AUC & not only accuracy!**

# ACCURACY IMPROVEMENTS

We used different models and algorithms to predict acquisitions. In this process we found different accuries & areas under the curve. The best performance is displayed below. Eventually we chose to combine the results of Logistic Regression with PCA & Random Forest to find the best contributing features.

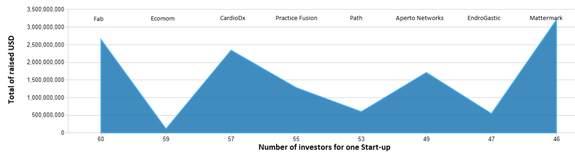


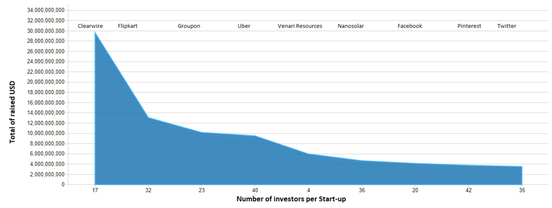

| FEATURE | PC-1 | PC-2 | PC-3 |
|---|---|---|---|
| Q3 | **1.74E-11** | 0.000283819 | -7.52587E-05 |
| FRA | **1.86E-11** | -0.00117669 | 0.002562859 |
| Software | **2.24E-11** | -0.00115355 | 0.002421146 |
| numeric_time_to_first_fun | **2.25E-11** | -0.788022422 | -0.52650934 |
| seed | **9.55E-11** | -0.003261288 | 0.008146963 |
| Biotechnology | -2.76E-11 | **0.001230917** | -0.003046385 |
| venture | -6.54E-11 | **0.004175266** | -0.009528042 |
| GBR | -9.18E-11 | **0.005716466** | -0.013011638 |
| USA | -7.49E-11 | **0.005795564** | -0.013348459 |
| numeric_count_investor | -4.14E-09 | **0.017208184** | -0.517671104 |
| Software | 2.24E-11 | -0.00115355 | **0.00242115** |
| FRA | 1.86E-11 | -0.00117669 | **0.00256286** |
| seed | 9.55E-11 | -0.003261288 | **0.00814696** |
| numeric_age | -7.19E-10 | -0.432248567 | **0.36404787** |
| numeric_funding_rounds | -8.82E-10 | -0.43792418 | **0.56719212** |




# VISUALIZATIONS & FINDINGS

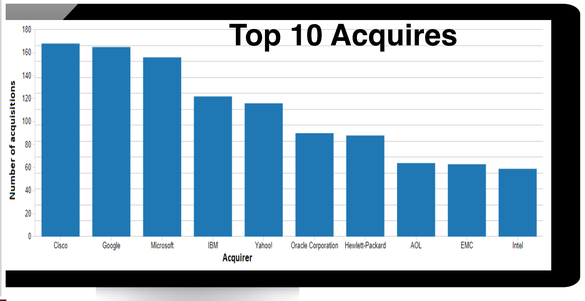
**Highest number of investors does not mean highest raise of USD**
**Fab has with 60 investors the maximum number of investors per company**

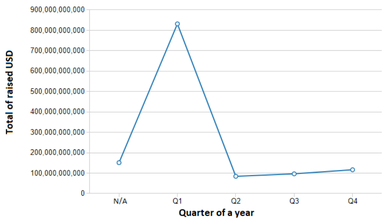



**Clearwire raised with 17 investors more than any other company**
**The number of raised number varies over the number of investors per company**


**Startups find the most investors in Menlo Park, New York or San Francisco**
**In Europe, London is the capital of investors**

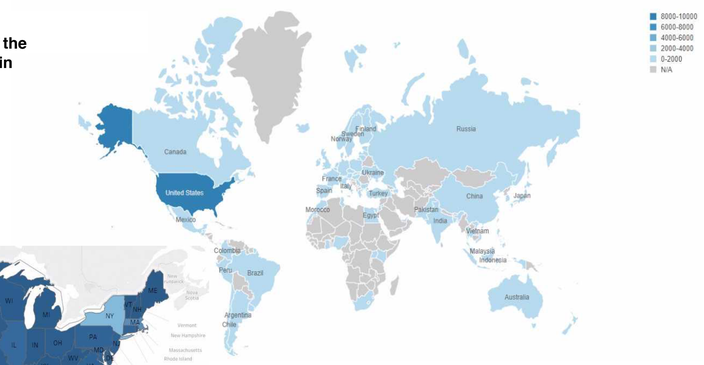


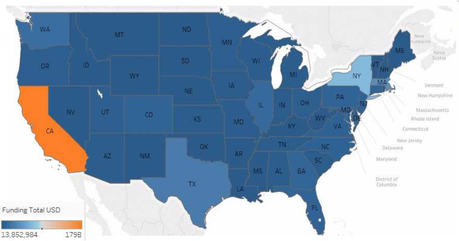




*** According to the data, during Q1 of a year companies raise most money**

## GEOGRAPHIC DISTRIBUTION

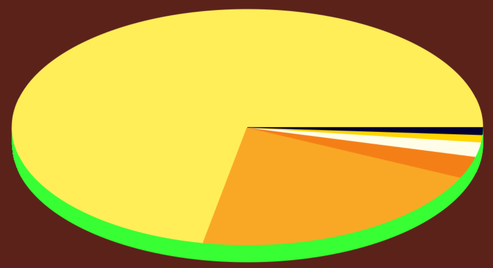
**Most of the observations of the data set are about startups in the U.S.**




**Considering USA startups in California have the highest funding Washington, New York and Texas witness a decent amount**

## Funding Types




## Highest Contributing Factors

**MODEL : RANDOM FOREST**


****Public relations & Seed were contributing factors in other random forest classifications that had**