

STARTUPS ANALYSIS

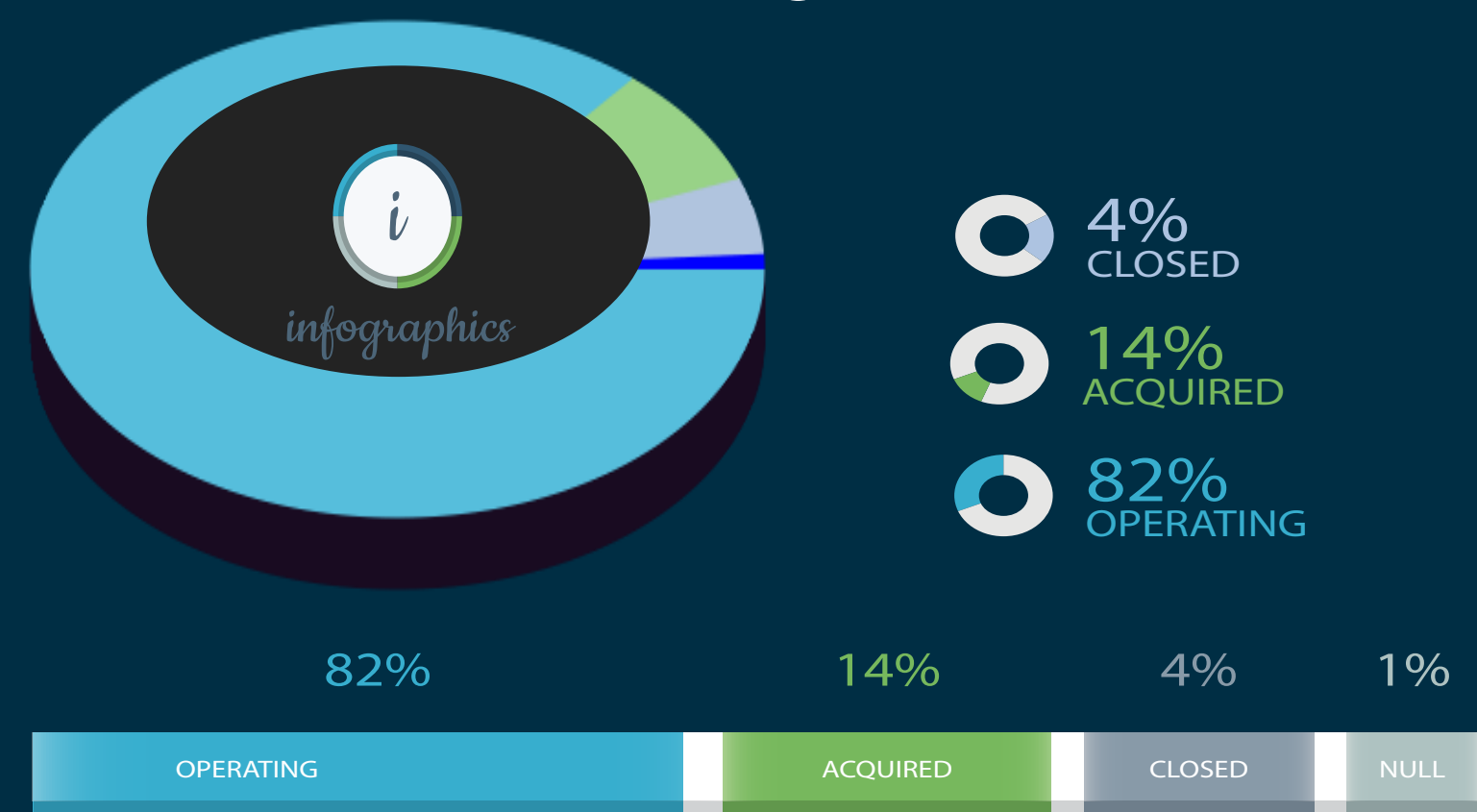
A Project By:
MAXIMILIAN OTT
PRIYA MATNANI
SASCHA HAGEDORN

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

Multiple supervised learning algorithms such as Logistic Regression, Random Forest and Neural Networks are applied after intense data preparation. Validation shows that Neural Networks has the best performance in this case.

With the results of this project existing startups can evaluate their performance in order to discover their probability to be acquired and emerging startups can use the outcomes as a guideline for how to structure their company or which features to emphasize while going down the path of an emerging startup.

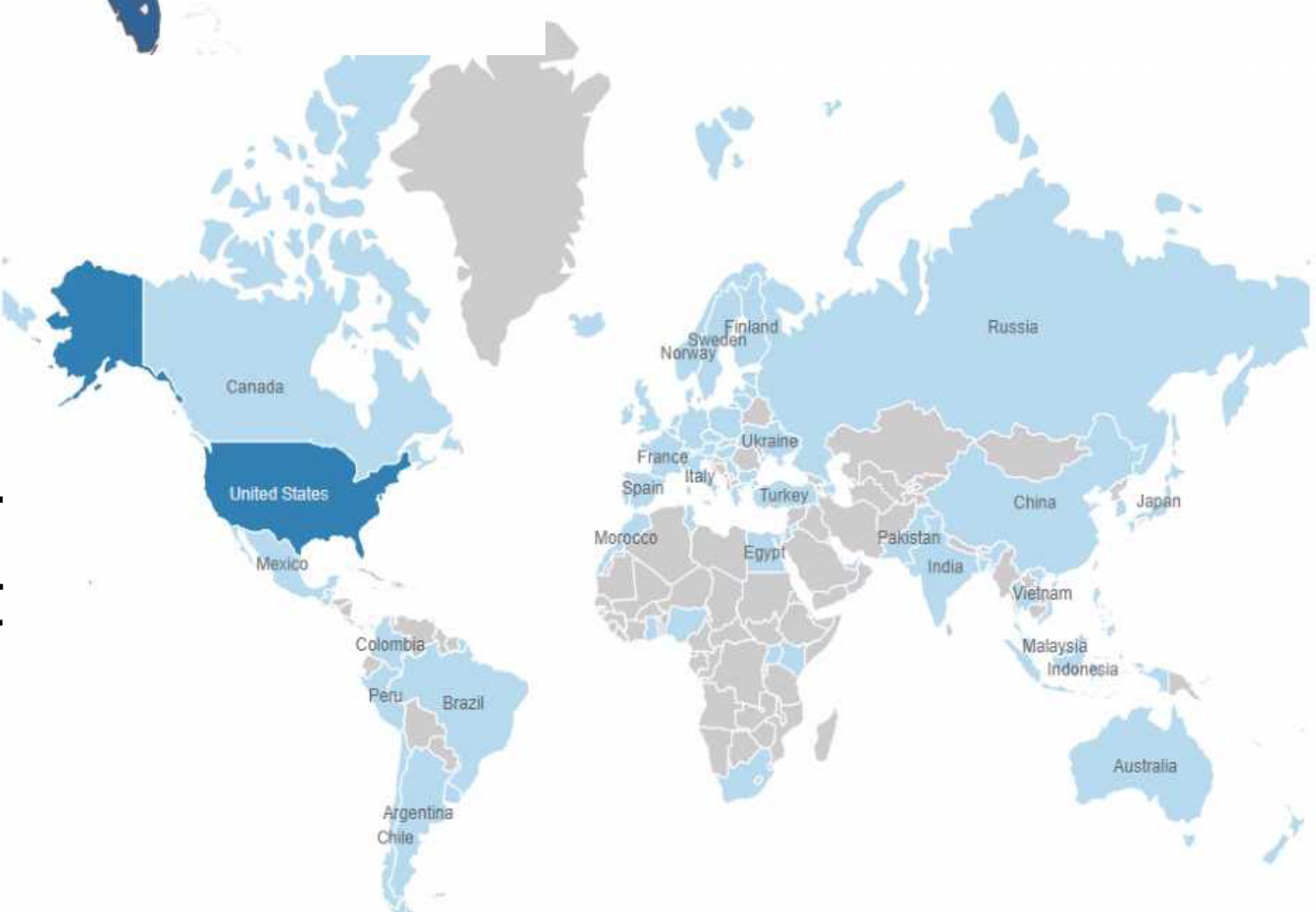
Distribution of Target Variable



GEOGRAPHY

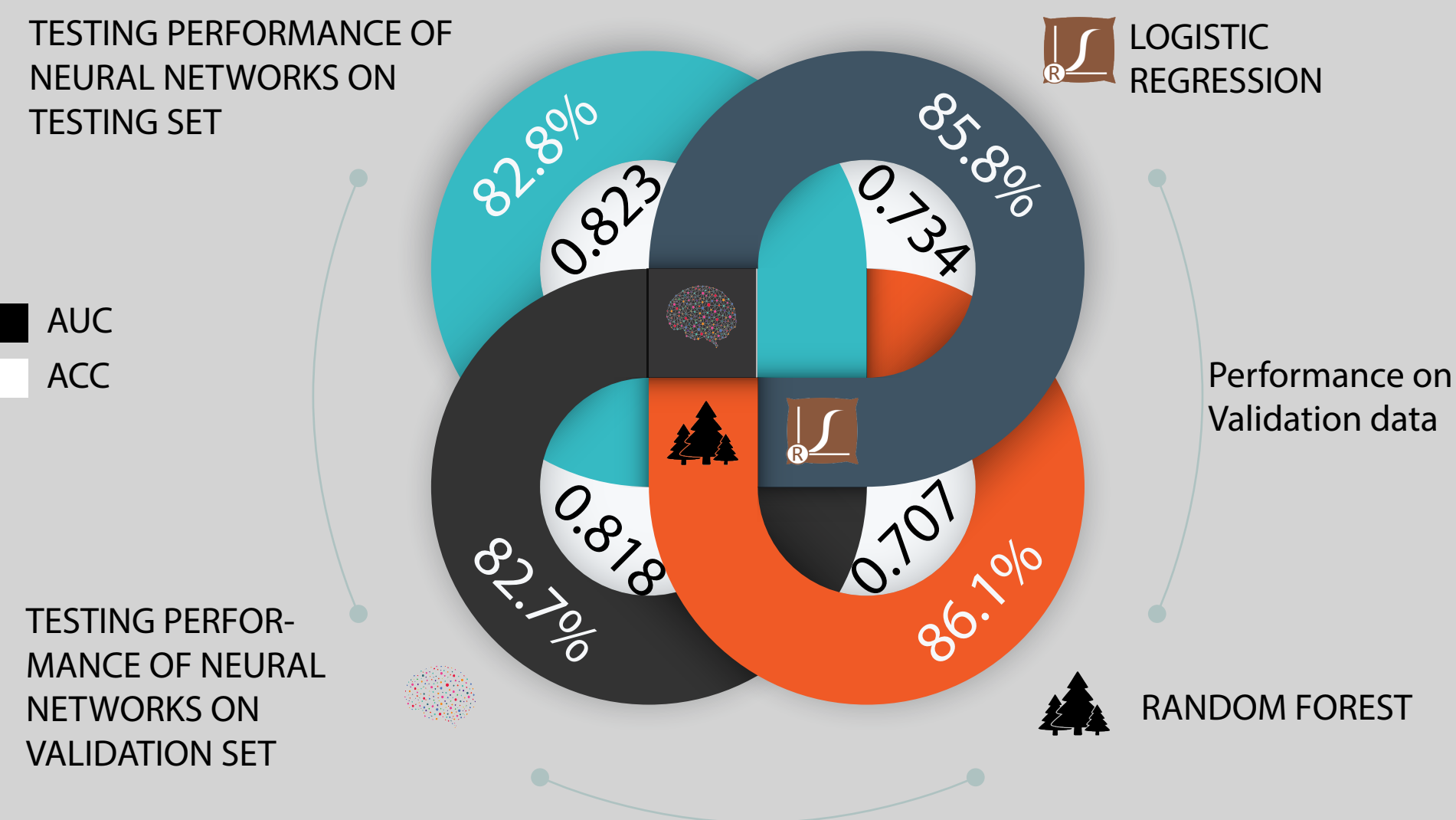


Most of the observations of the data set are about startups in the U.S.

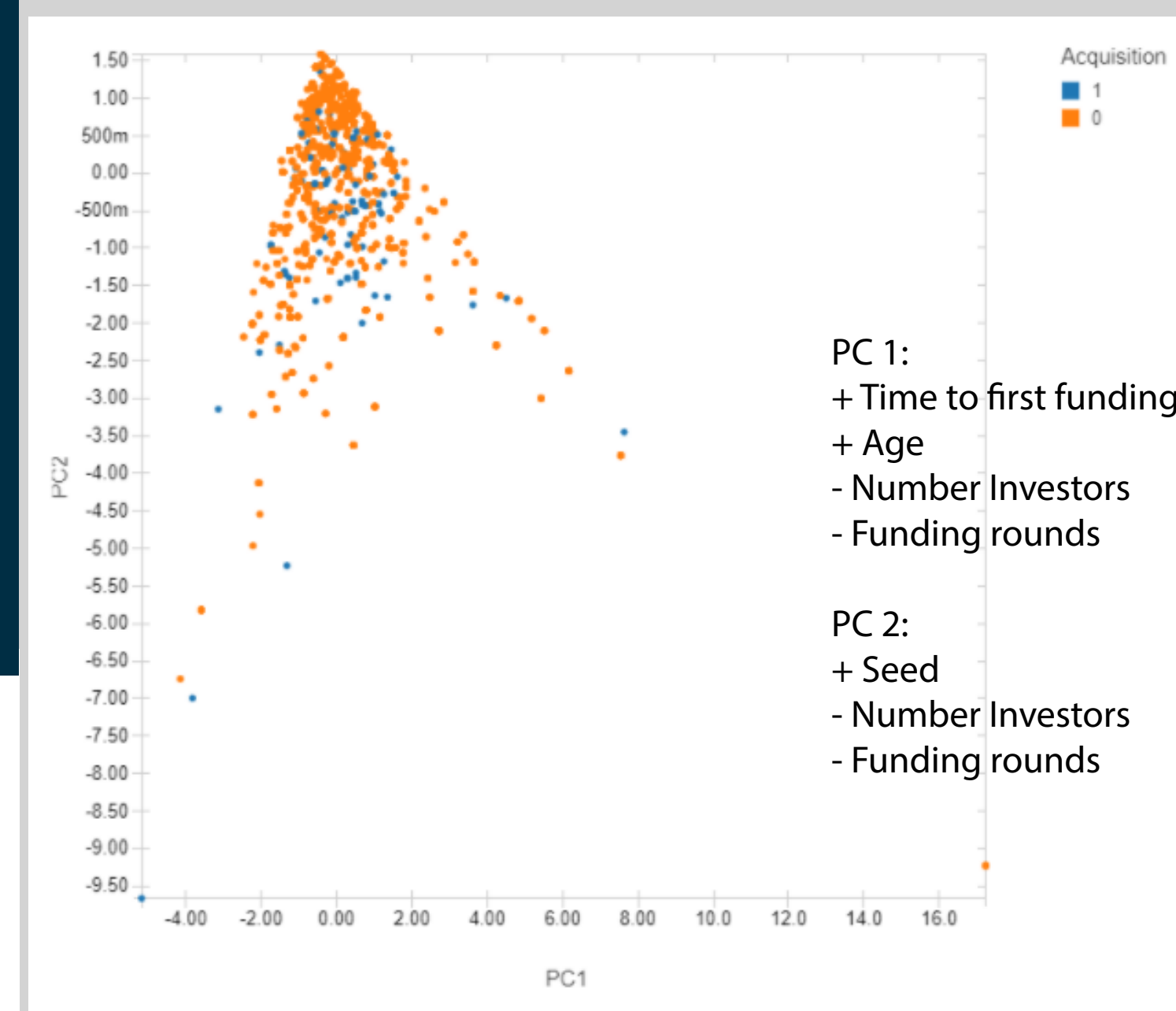


MODEL PERFORMANCE

Models have different performances

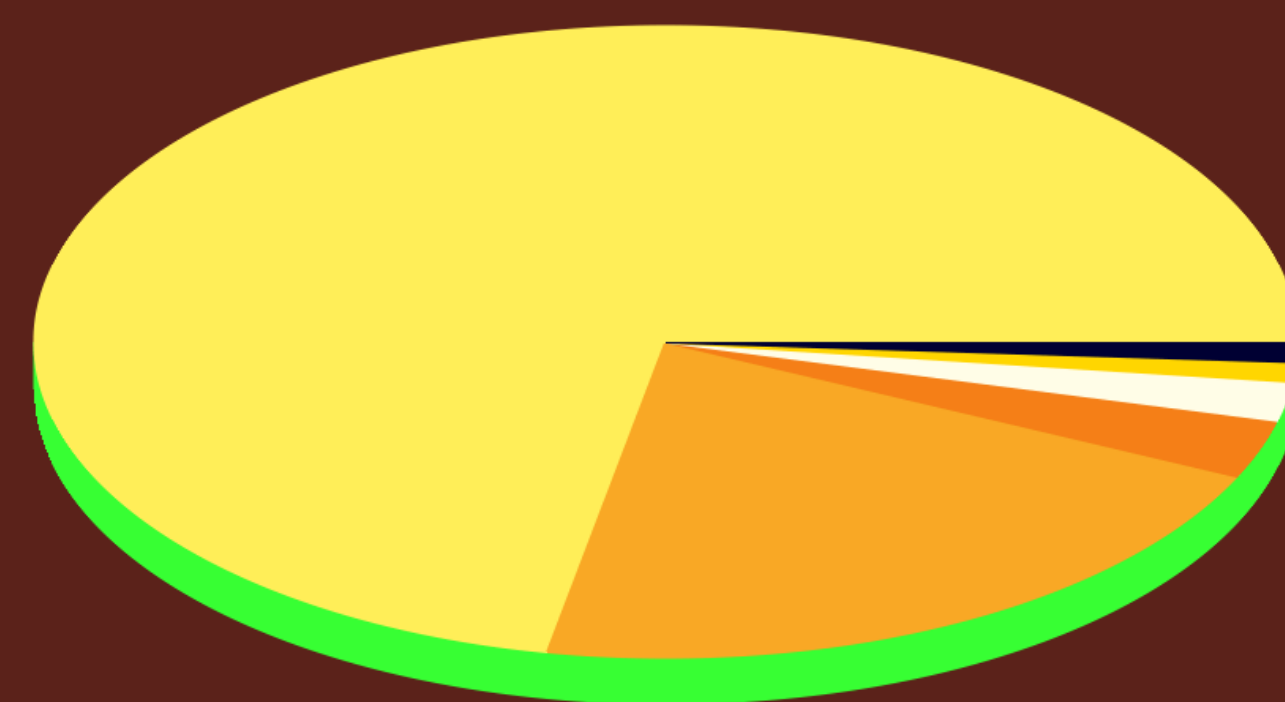
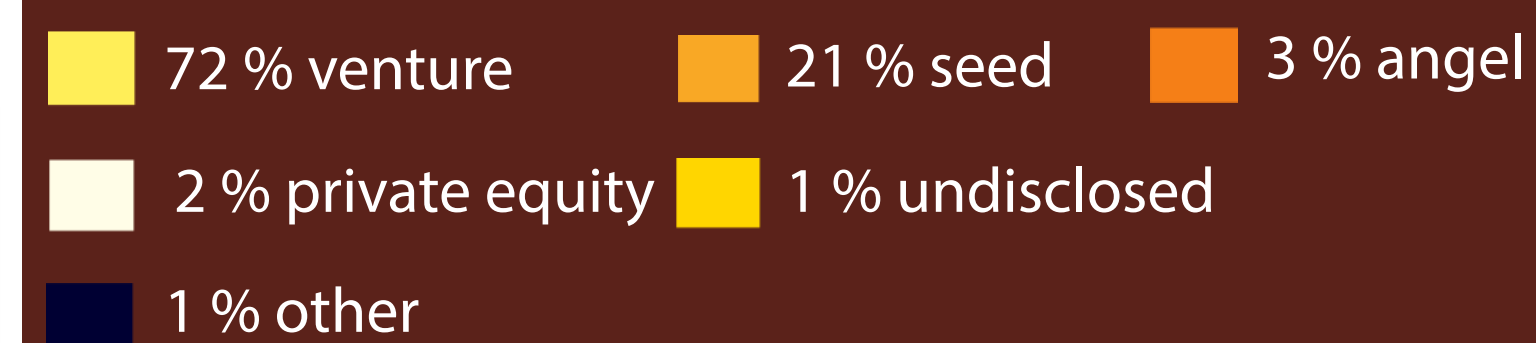


Projection into new Feature Space

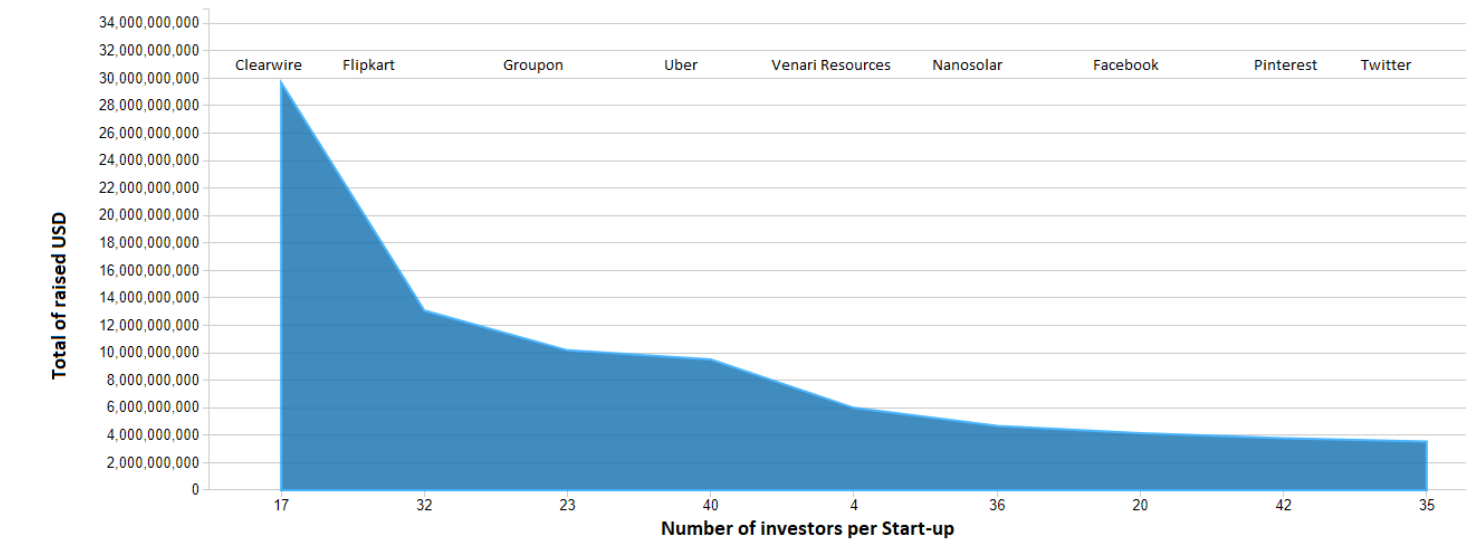


Considering USA, start-ups in California have the highest funding

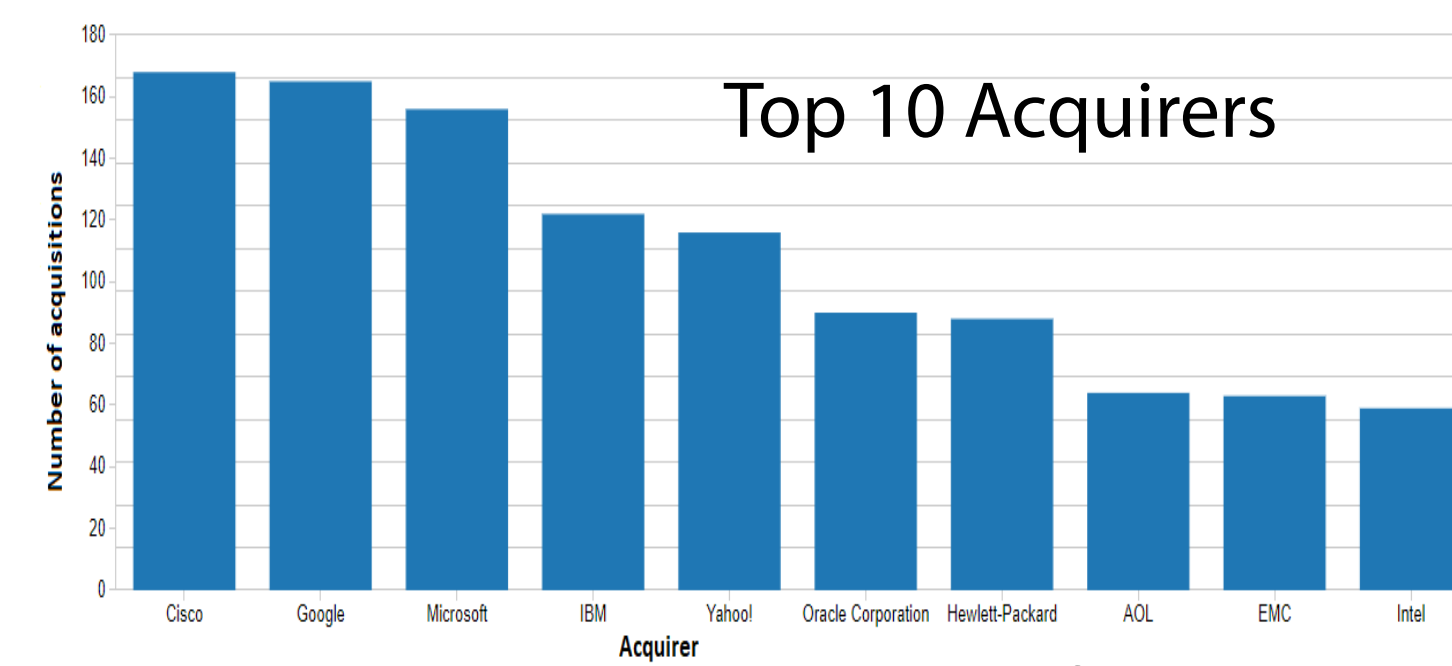
Funding Types



DESCRIPTIVE STATISTICS OF DATA

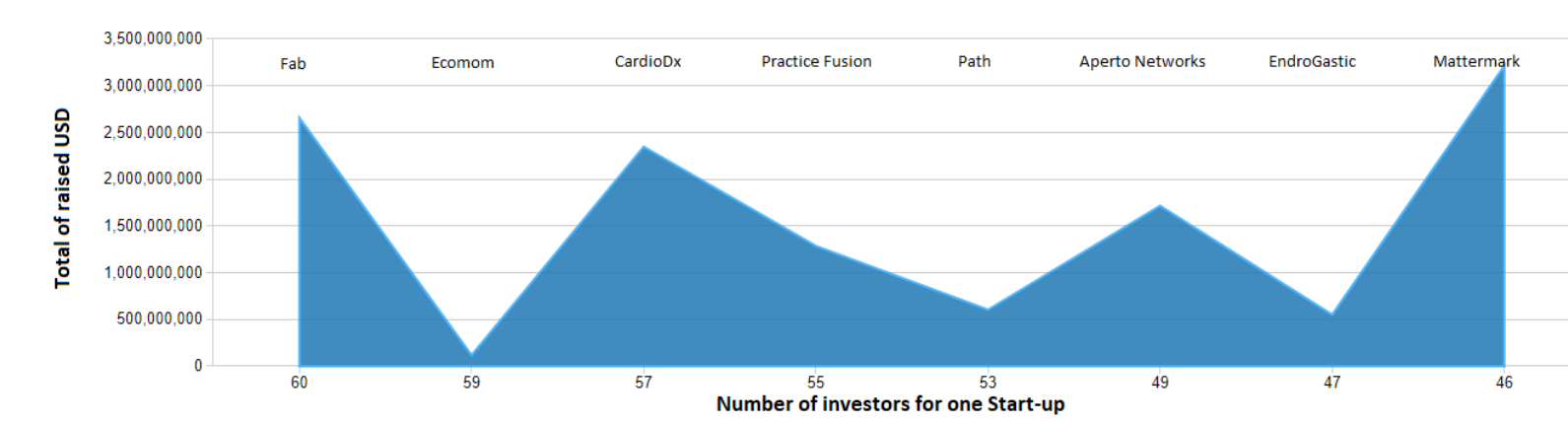
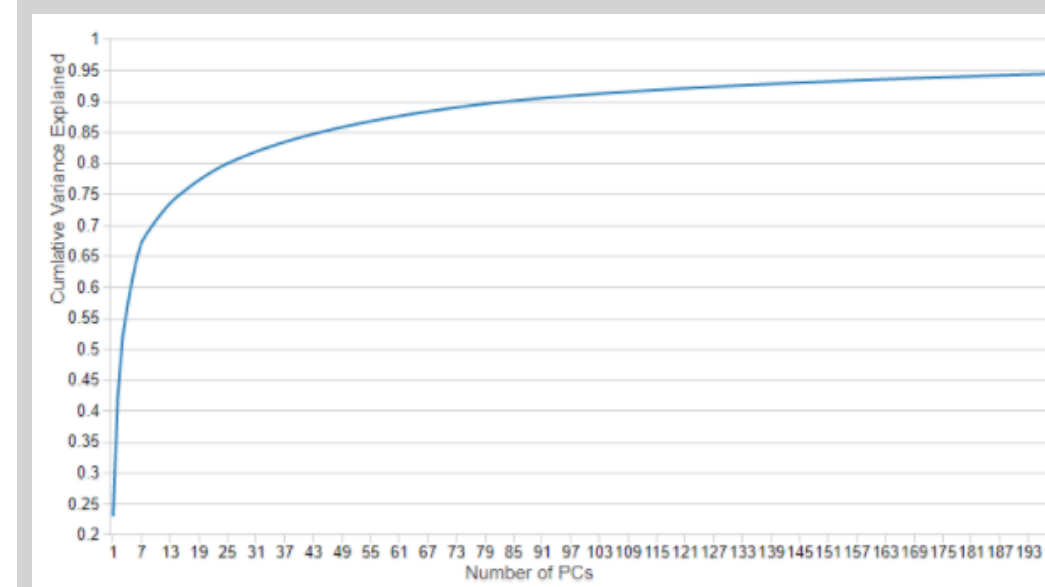


Clearwire raised more \$ than any other company despite not having the most investors



Cisco, Google, Microsoft and IBM have acquired the most startups

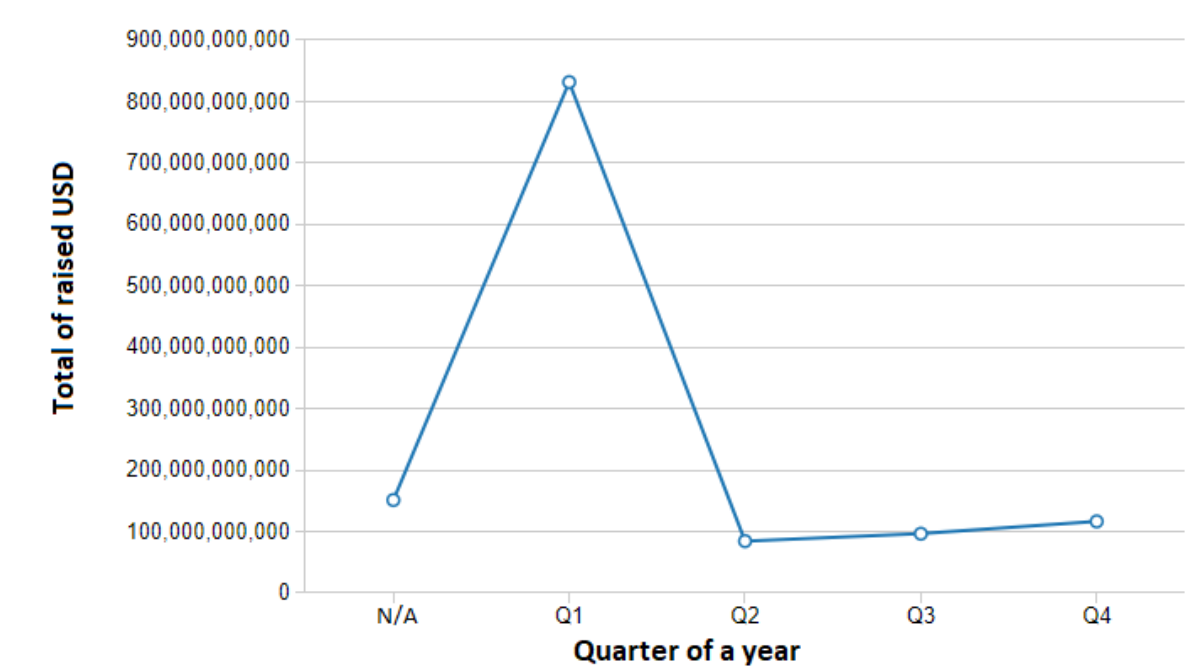
Variance of Principal Components



Fab has with 60 investors the maximum number of investors per company



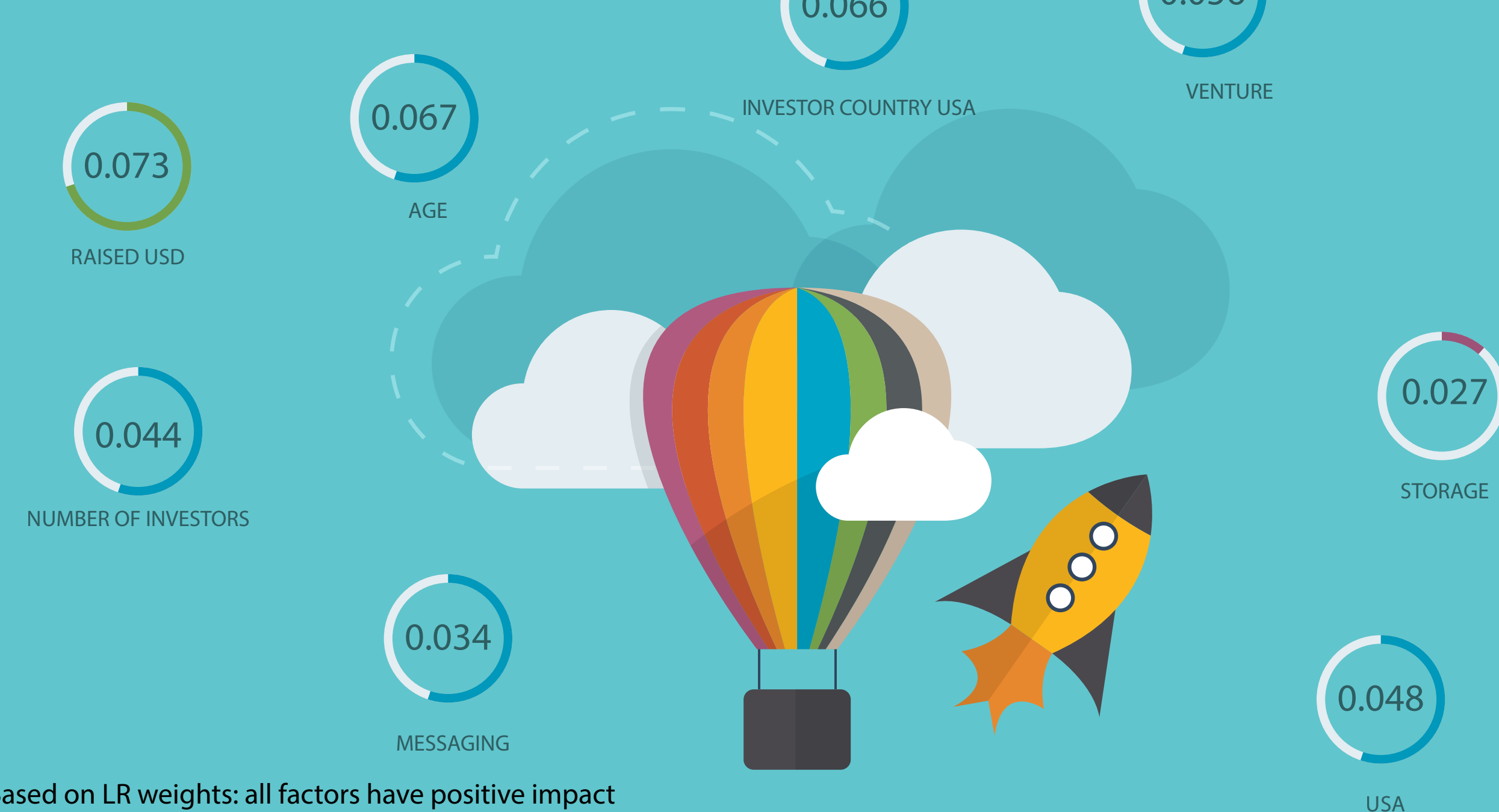
Startups find the most investors in Menlo Park, New York or San Francisco



Companies founded in Q1 tend to get more funds

Highest Contributing Factors

MODEL : RANDOM FOREST



Based on LR weights: all factors have positive impact