

Predicting Startup Failure

By Natalya Doris

Business Problem

With the economy at a possible turning point, ABC Ventures hopes we can answer the following questions:

- Given a set of data points describing funding levels, location, market & timing, is it possible to **predict** whether a startup will succeed* or fail, with emphasis on identifying possible **failures**?
- What **factors** are most important in predicting failure?

**Success is defined as a company that has been acquired, as labeled in the data*



Data

- Source: Crunchbase / [Kaggle](#)
- Over 50,000 observations; focused on subset of 6,200
- Features include funding metrics, location, timing & market category



Objectives

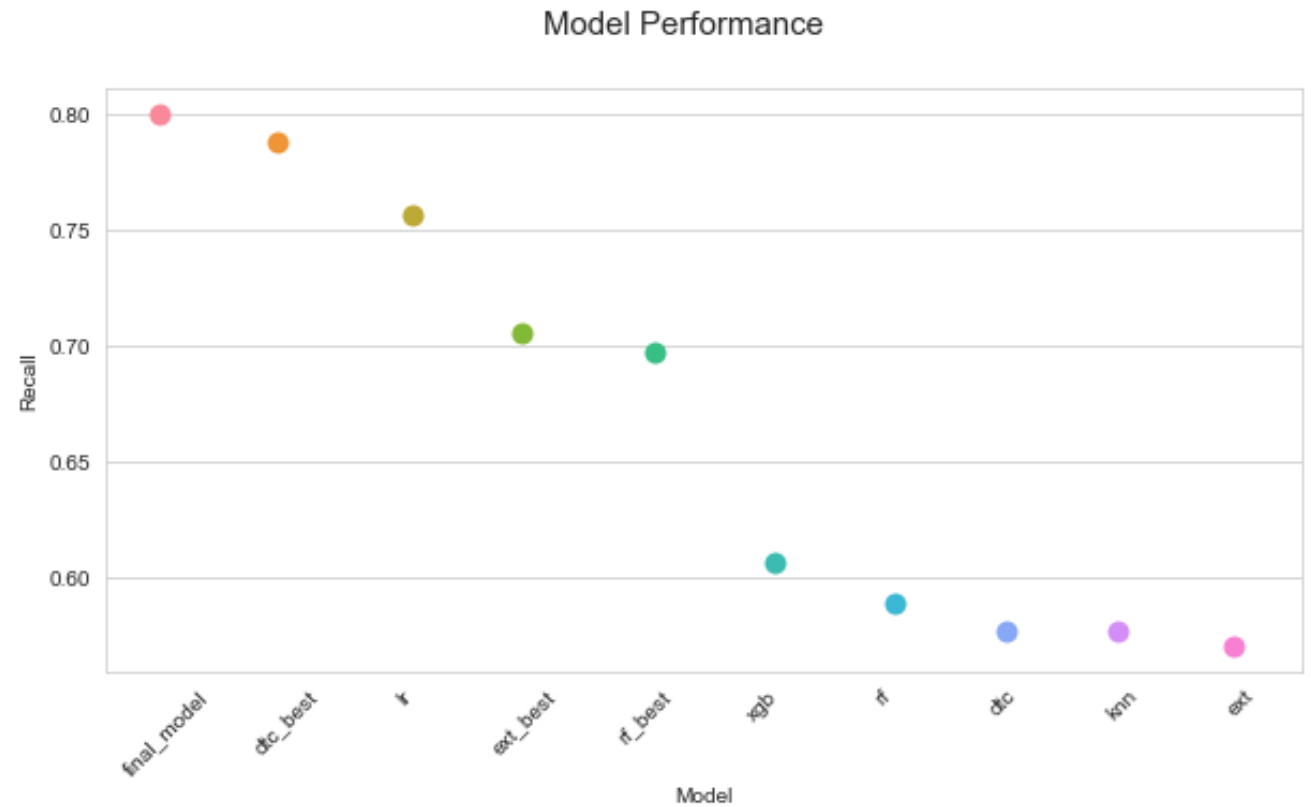
Targeting **status** (company acquired or closed), we would like to:

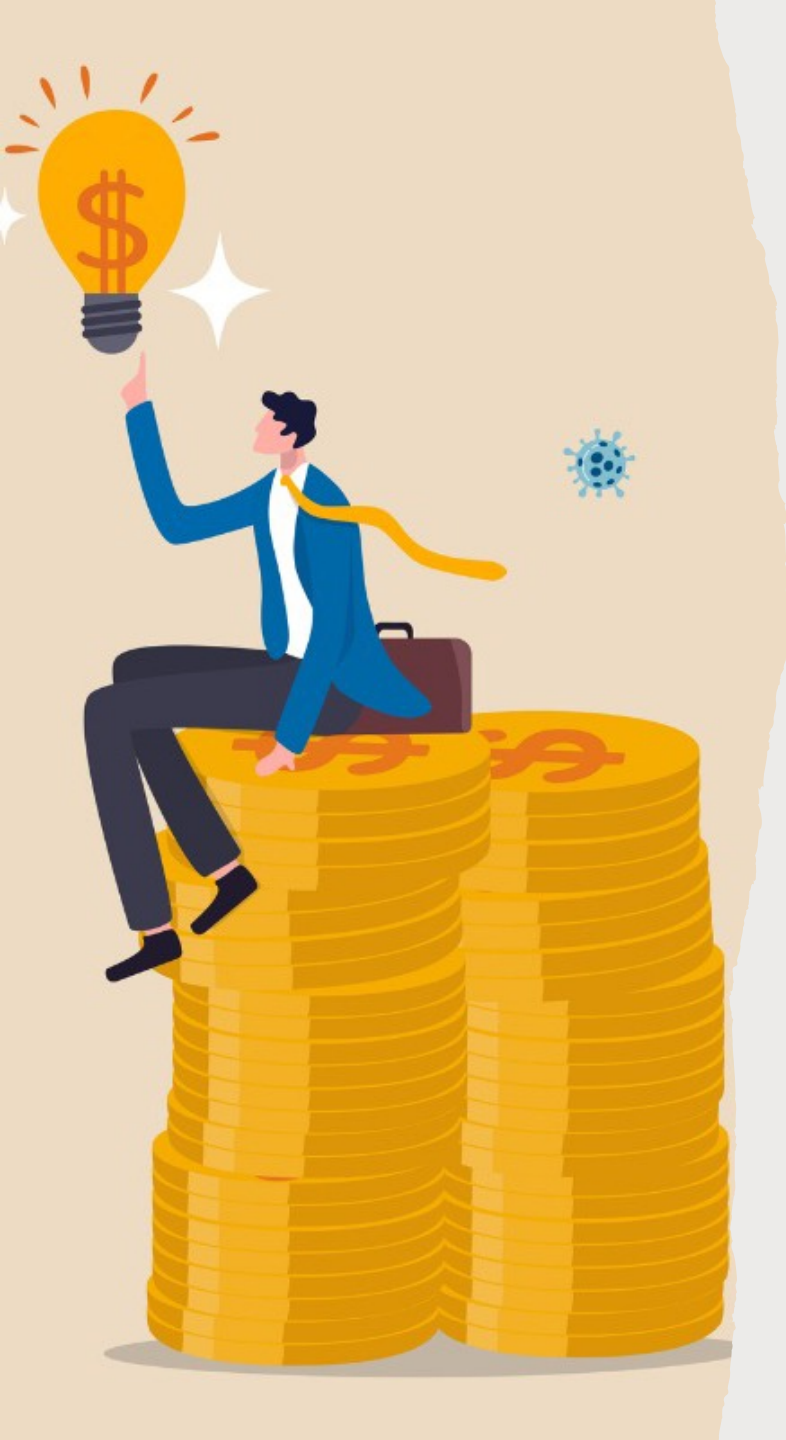
1. Correctly identify the companies that **failed**: recall
2. Keeping in mind **false positives**: ROC AUC score
3. Produce an **accurate** model





Model Selection





Final Model: XG Boost

80% Recall

- Correctly identified 80% of the companies that failed

73% ROC AUC

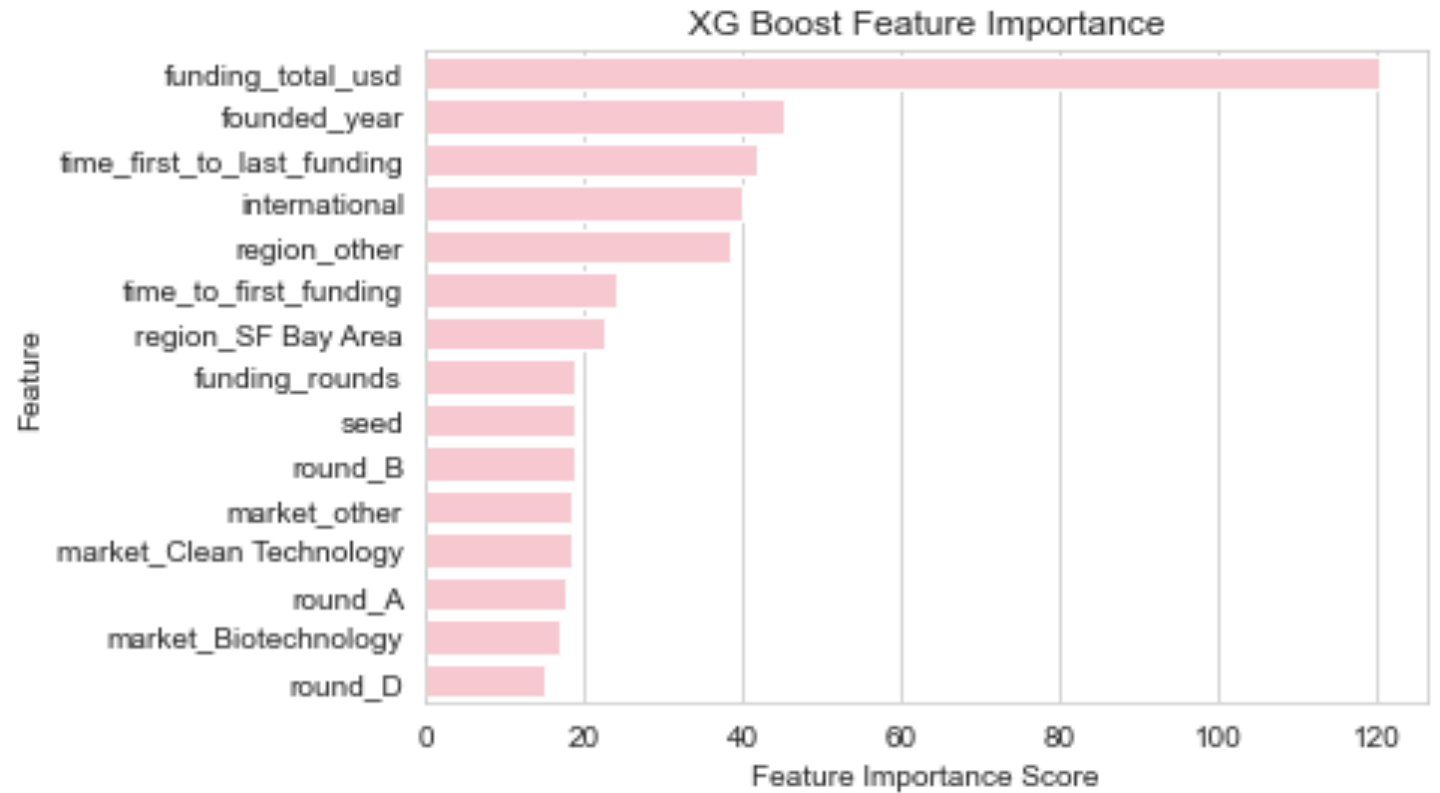
- A measure of how good the model is at distinguishing between success and failure

72% Accuracy

- Achieved 72% Accuracy

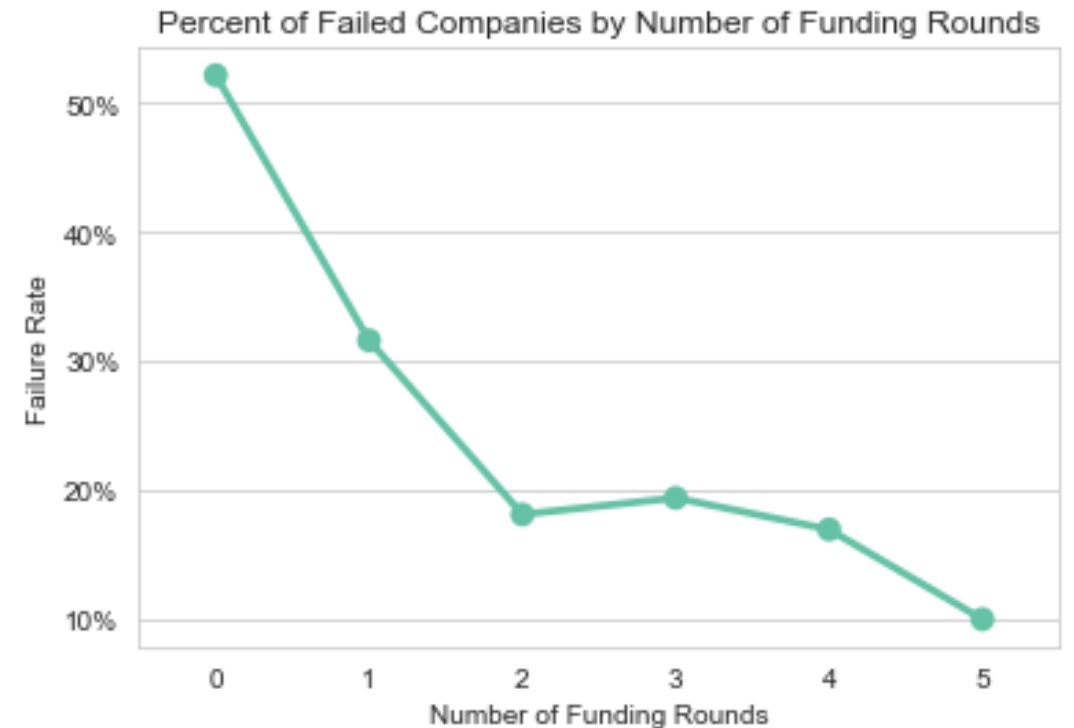
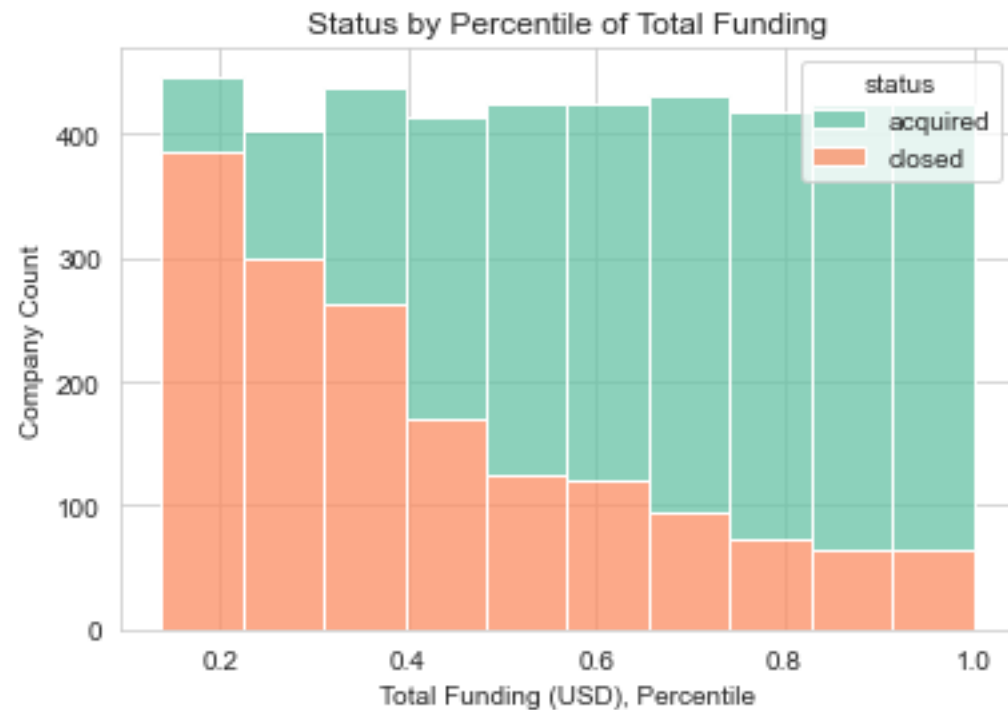


Final Model: XG Boost



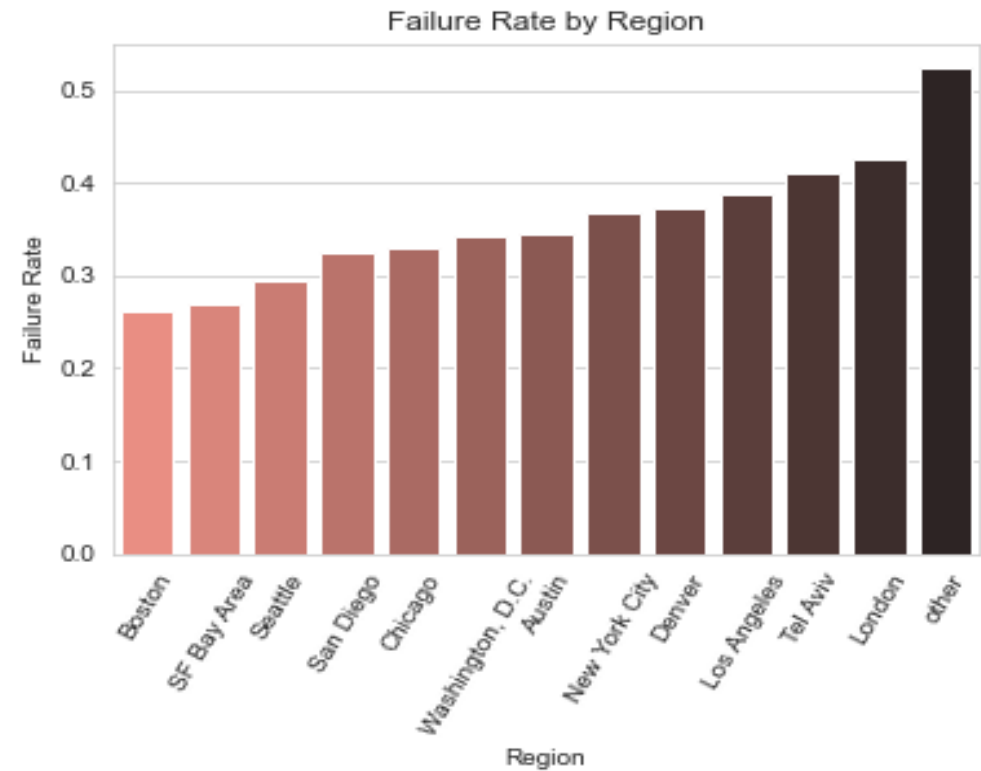
Funding First

Companies with more funding rounds, higher total funding have lowest failure rates



Location

- *U.S. based companies more likely to be acquired*
- *Among cities with greatest concentration of startups, Boston, San Francisco see lowest failure rates*
- *International cities Tel Aviv & London have highest failure rates*
- *Average failure rate across all other regions slightly above 50%*



Age

- *Successful companies survive longer before first funding*
- *Closed companies only last six months after first funding, on average*

Status	Years to First Funding	Years from First to Last Funding	Age
acquired	3.74	1.49	5.23
closed	1.69	0.51	2.20

Conclusion

In identifying startups at greatest risk of failure, ABC Ventures should focus on companies that:

- Have **low overall funding**, or achieve fewer funding rounds
- Are located **outside the U.S.**, or domestically in regions with fewer startups
- Raise funds at an **earlier** age





Next Steps

- *Source data on startups founded in 2015 & later*
- *Separate U.S. and international companies, creating separate models for each category*
- *Do a deep dive on regions & markets with lowest failure rates*

If any questions, you can reach me at ntdoris2@gmail.com. Thank you!