STARTUP SUCCESS CLASSIFIER

Goh Yan Da | DSI 33

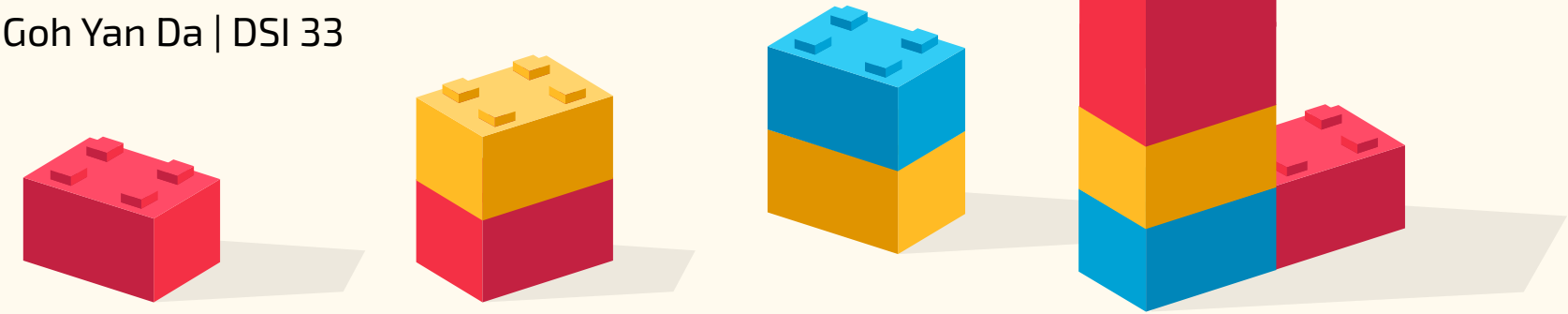


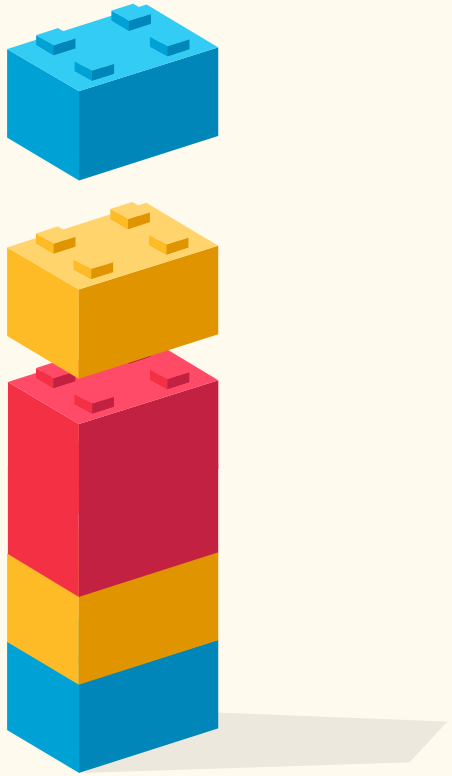
TABLE OF CONTENTS

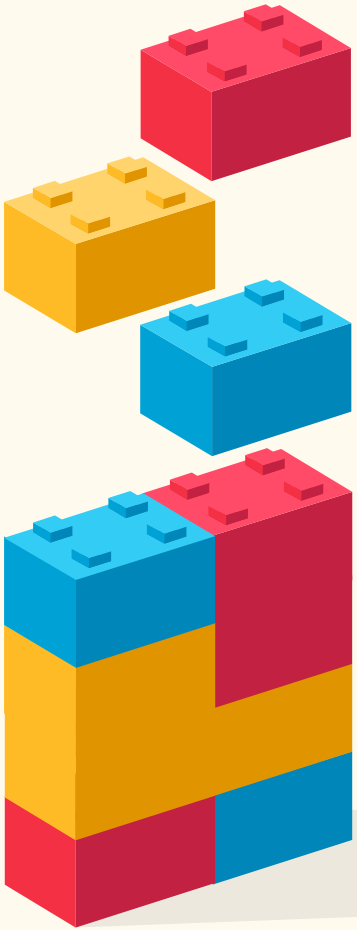
01 BACKGROUND

02 DATASET

03 EDA & MODELING

04 CONCLUSION

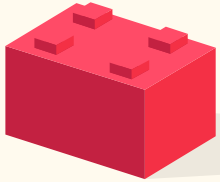




HOWDY!

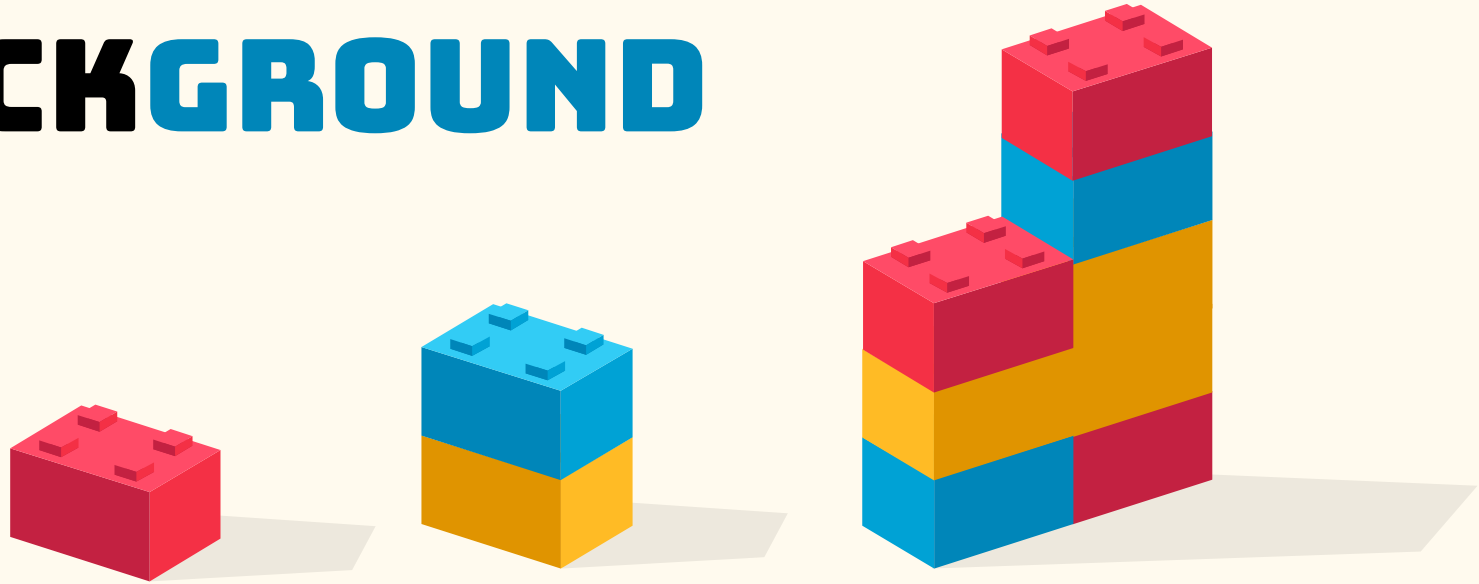
Goh Yan Da | 27
Email: goh_da@yahoo.com.sg
Linkedin:
https://www.linkedin.com/in/yandahere/

Github: https://github.com/Yandahere



01

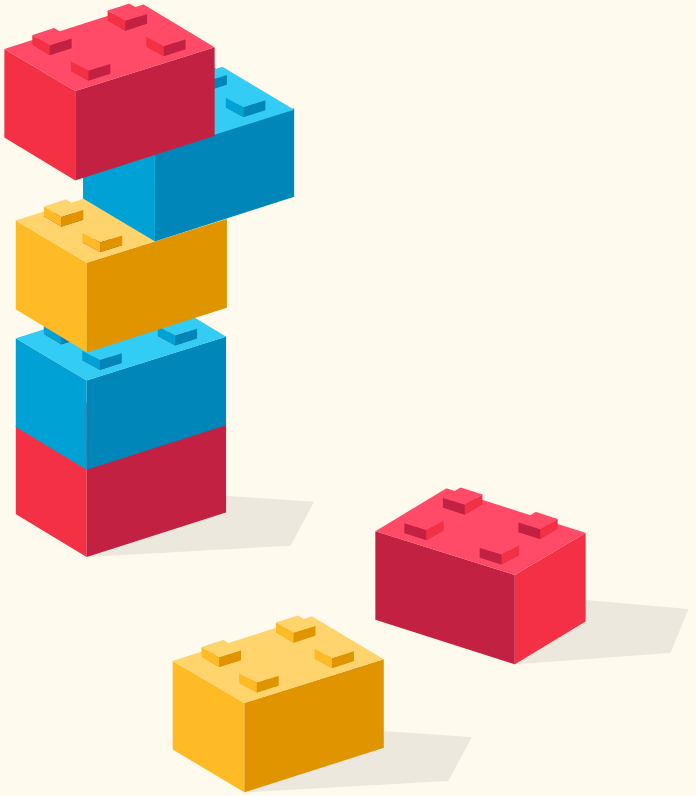
BACKGROUND



WHO ARE WE?

Startup Sg is an online website for new prospective startups and investors to meet one another.

Investors and prospective entrepreneurs are always looking to optimize time and resources.



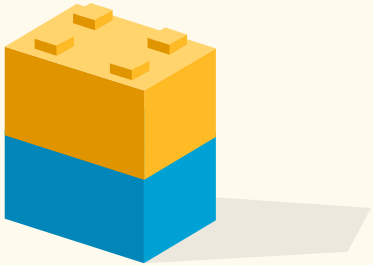
WHY SINGAPORE?

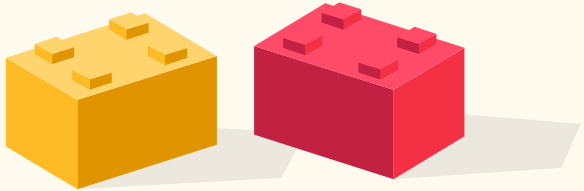
RANKED 1ST

In global startup talent

NURTURING ECOSYSTEM

Government support local startups.

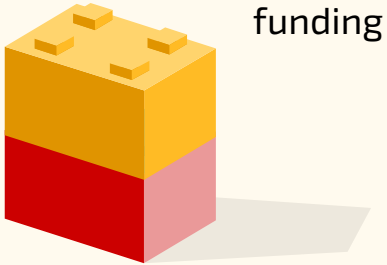




NURTURING ECOSYSTEM

128 COMMERCIAL BANKS

Service regional activities and secure funding



HASSLE FREE SETUP

Register your business on ACRA
Be aware of regulations in Singapore

SUPPORTIVE GOVERNMENT

21 free trade agreements (FTAs) with 27 economies and 76 comprehensive avoidance of double tax agreements (DTAs)

WHAT MAKES A SUCCESSFUL STARTUP?

Branding & Marketing

It's the biggest planet in the Solar System

Ideation

An adaptive business model to accommodate to market demand

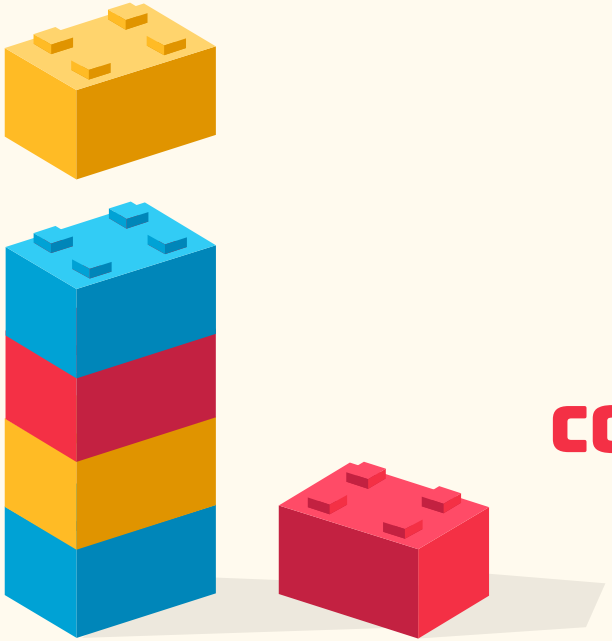
Goals & KPIS

Proposed timeline with reasonable goals

Backing & Funding

Strong connections and investors to back up

COMMON CHALLENGES



ACCESS TO FUNDING

Hesitant investors

TALENT ATTRACTION

Attracting and securing top talent (influencers)

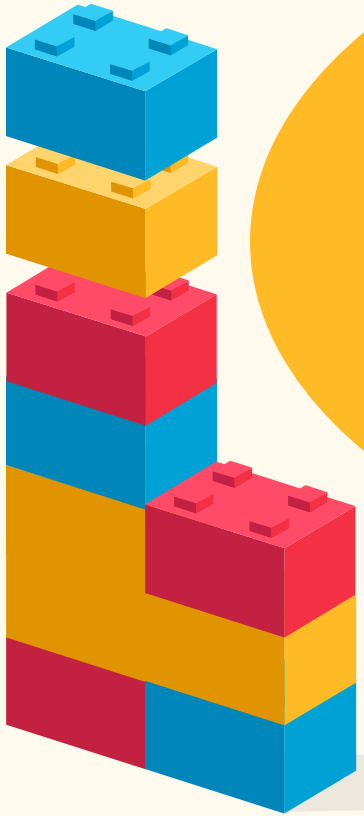
INTENSE COMPETITION

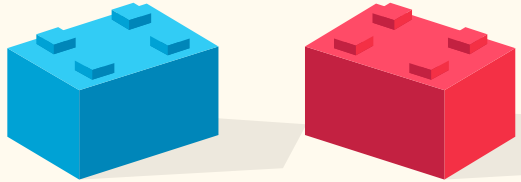
Singapore's home to large amount of start ups

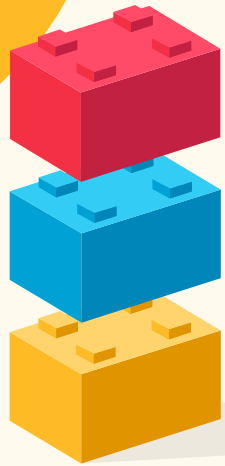
OPERATING COSTS

Singapore high living costs

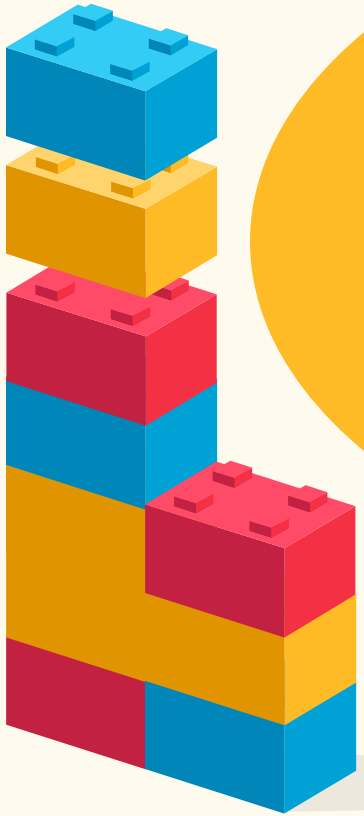
WHAT'S THE PROBLEM?

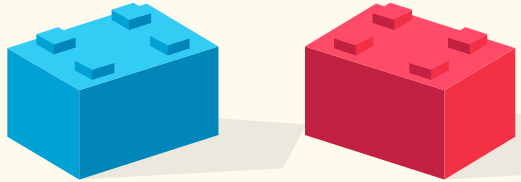


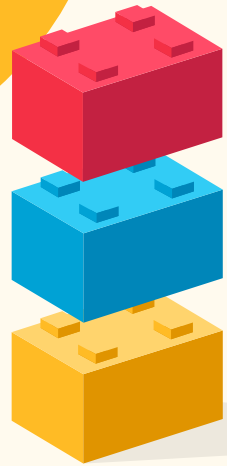




NURTURING THE NEWEST UNICORN IN A SEA OF STARTUPS CAN BE CHALLENGING TO EARLY START UPS AND NEW INVESTORS. WHAT CAN INVESTORS AND ENTREPRENEURS LOOK OUT FOR TO PREDICT A SUCCESS OF A STARTUP?

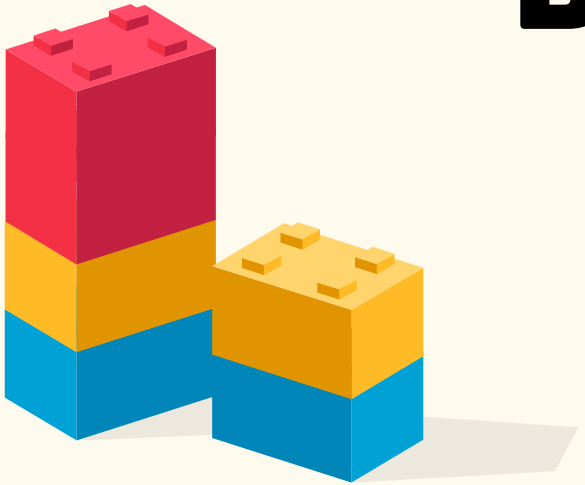


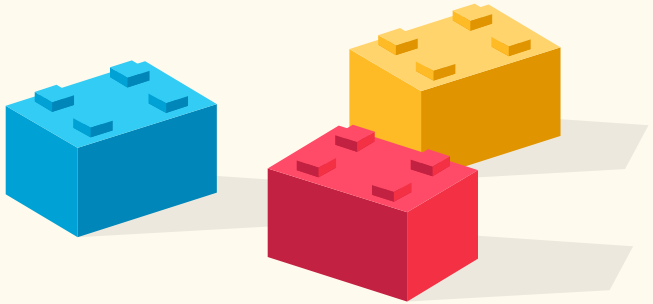




02

DATASET

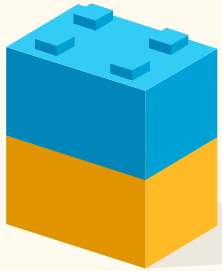




OVERVIEW OF DATASET

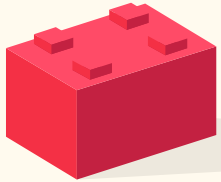
US HISTORICAL DATA

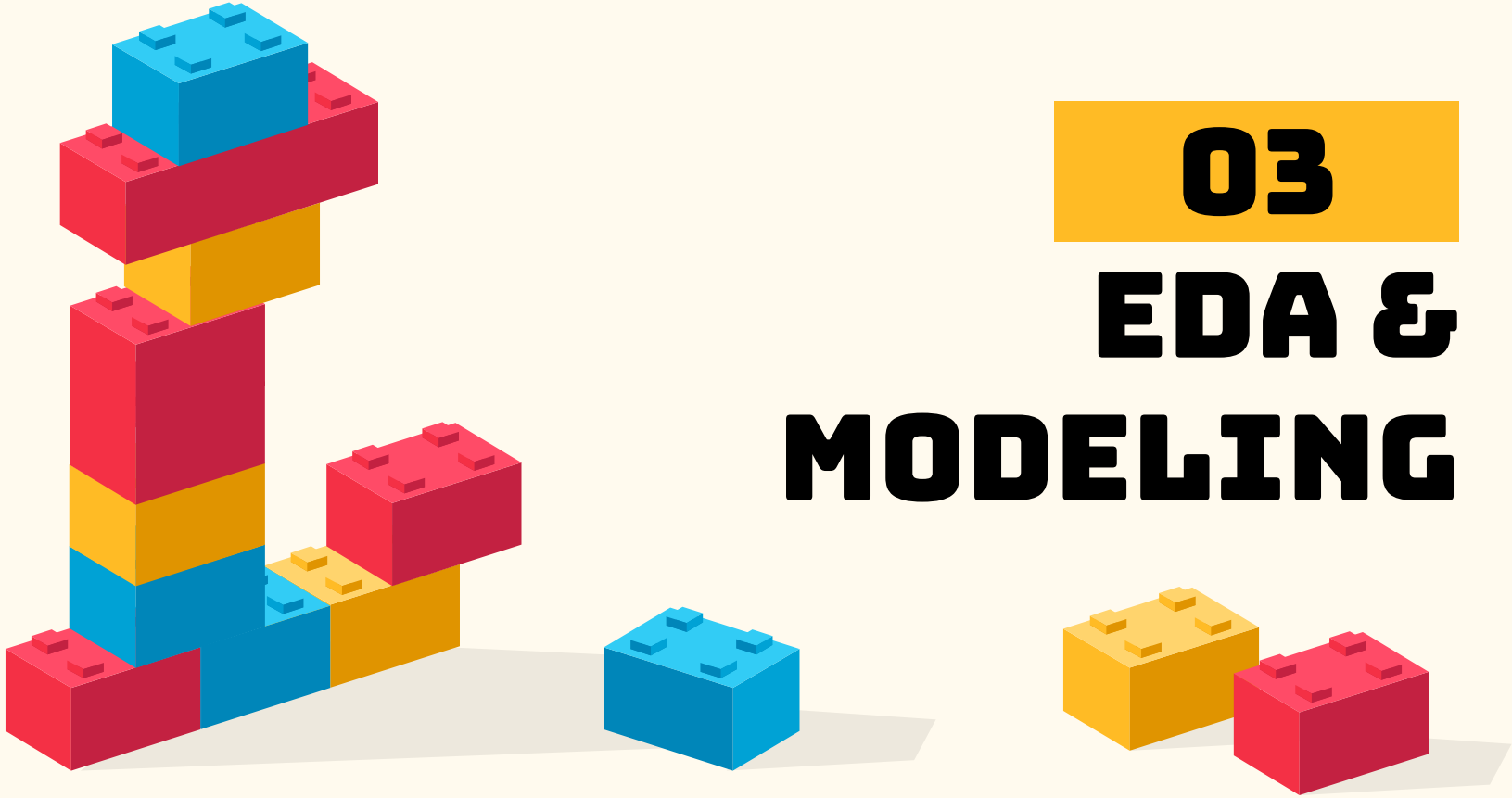
Dataset collected from the US is used as a proof of concept in predicting startup success
Data collected from 2000 to 2010 period



‘STATUS’

If a startup is successful, it is acquired.
A failed startup is closed

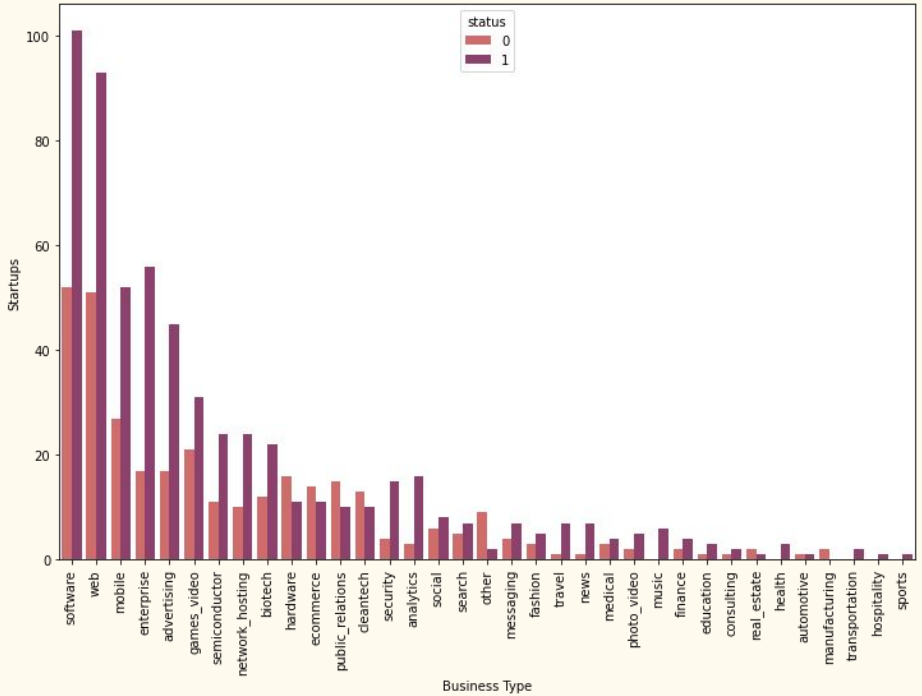




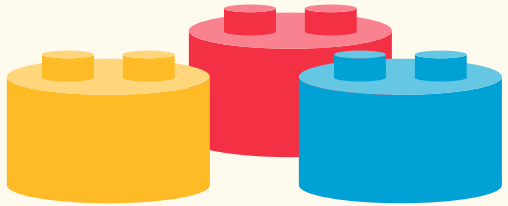
03

EDA & MODELING

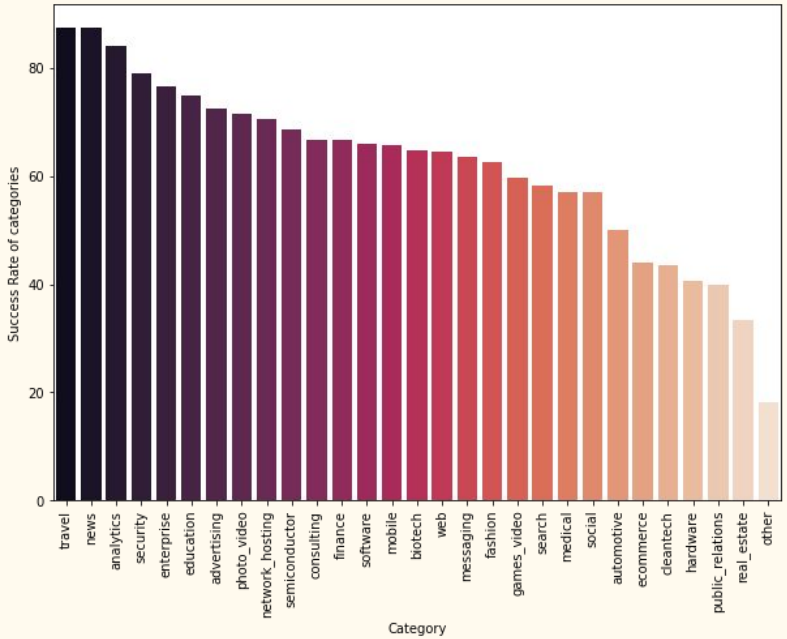
SUCCESS COUNTS BASED ON CATEGORIES



Top 3 most common startup business types are software, mobile and web.

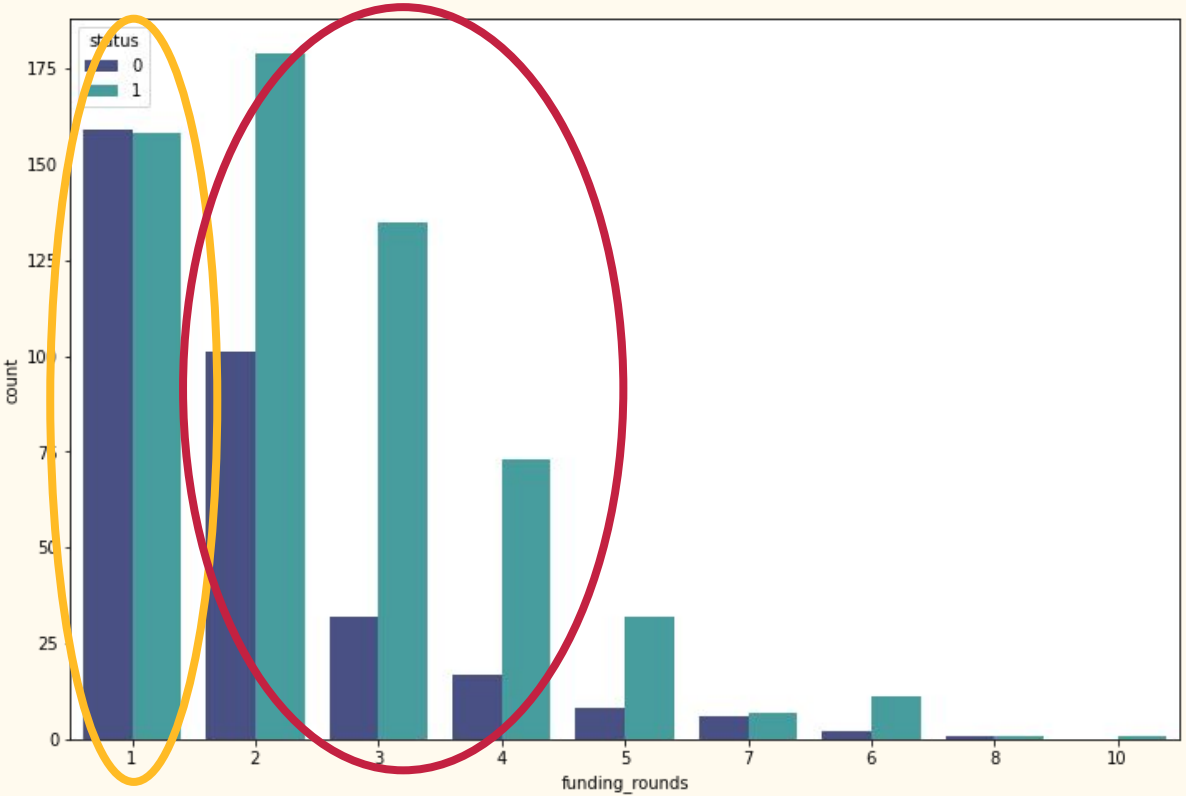


SUCCESS RATES BASED ON CATEGORIES



	category_code	total_success	total_closed	total_startup	success_rate
27	travel	7	1	8	87.50
17	news	7	1	8	87.50
1	analytics	16	3	19	84.21
23	security	15	4	19	78.95
8	enterprise	56	17	73	76.71
7	education	3	1	4	75.00
0	advertising	45	17	62	72.58
19	photo_video	5	2	7	71.43
16	network_hosting	24	10	34	70.59
24	semiconductor	24	11	35	68.57
5	consulting	2	1	3	66.67
10	finance	4	2	6	66.67
26	software	101	52	153	66.01
15	mobile	52	27	79	65.82
3	biotech	22	12	34	64.71

NUMBER OF FUNDING ROUNDS



The success of a startup is quite uniformed with one funding round, however the likelihood of success increases as funding rounds increase from 2 to 5, before it becomes more uniformed in distribution again.

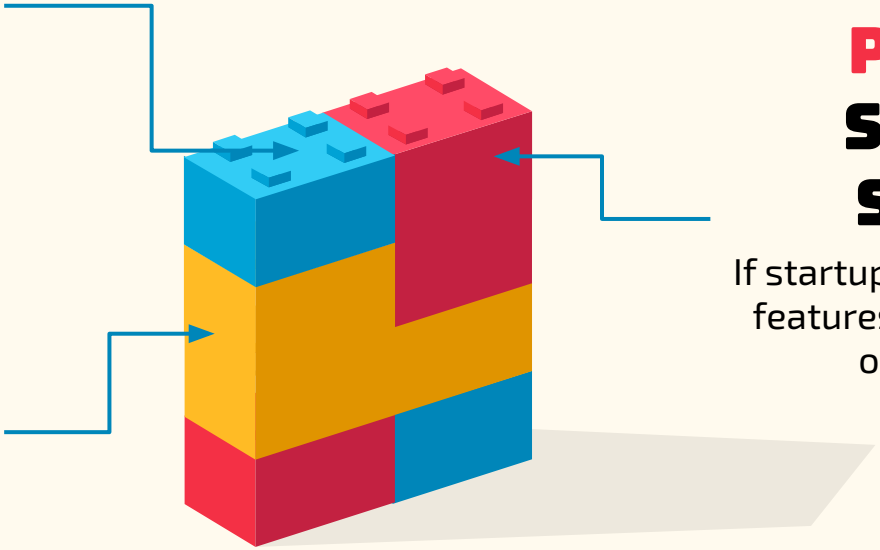
HOW DOES IT WORK?

DEPLOY MODEL

Uses models such as logistic regression and K-nearest neighbor

TRAIN ON DATASET

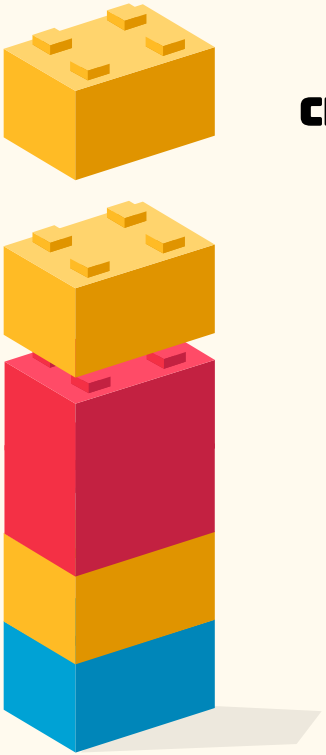
Split the dataset into train/test data



PREDICT STARTUP SUCCESS

If startup has successful features, higher chance of being acquired

HOW DO WE MEASURE?

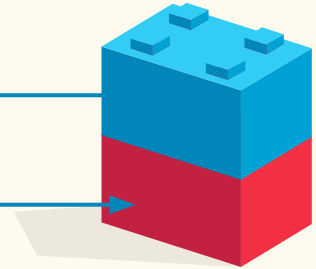


CLASSIFICATION MODELS:

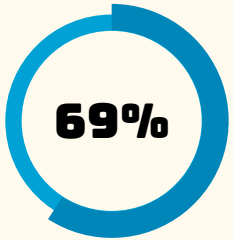
- Logistic Regression
- Naive Bayes (Multinomial, Bernoulli and Gaussian)
- K-NN

WHAT WE'RE LOOKING FOR:

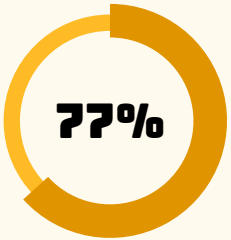
- Higher Accuracy Score
- Higher F1 score



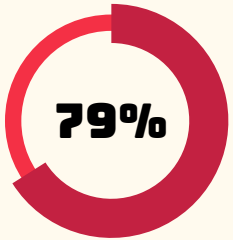
TOP F1 SCORES WITH HYPERPARAMETER TUNING



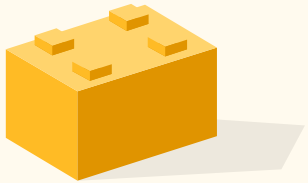
K-NN



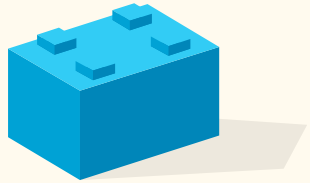
LOGISTIC REGRESSION

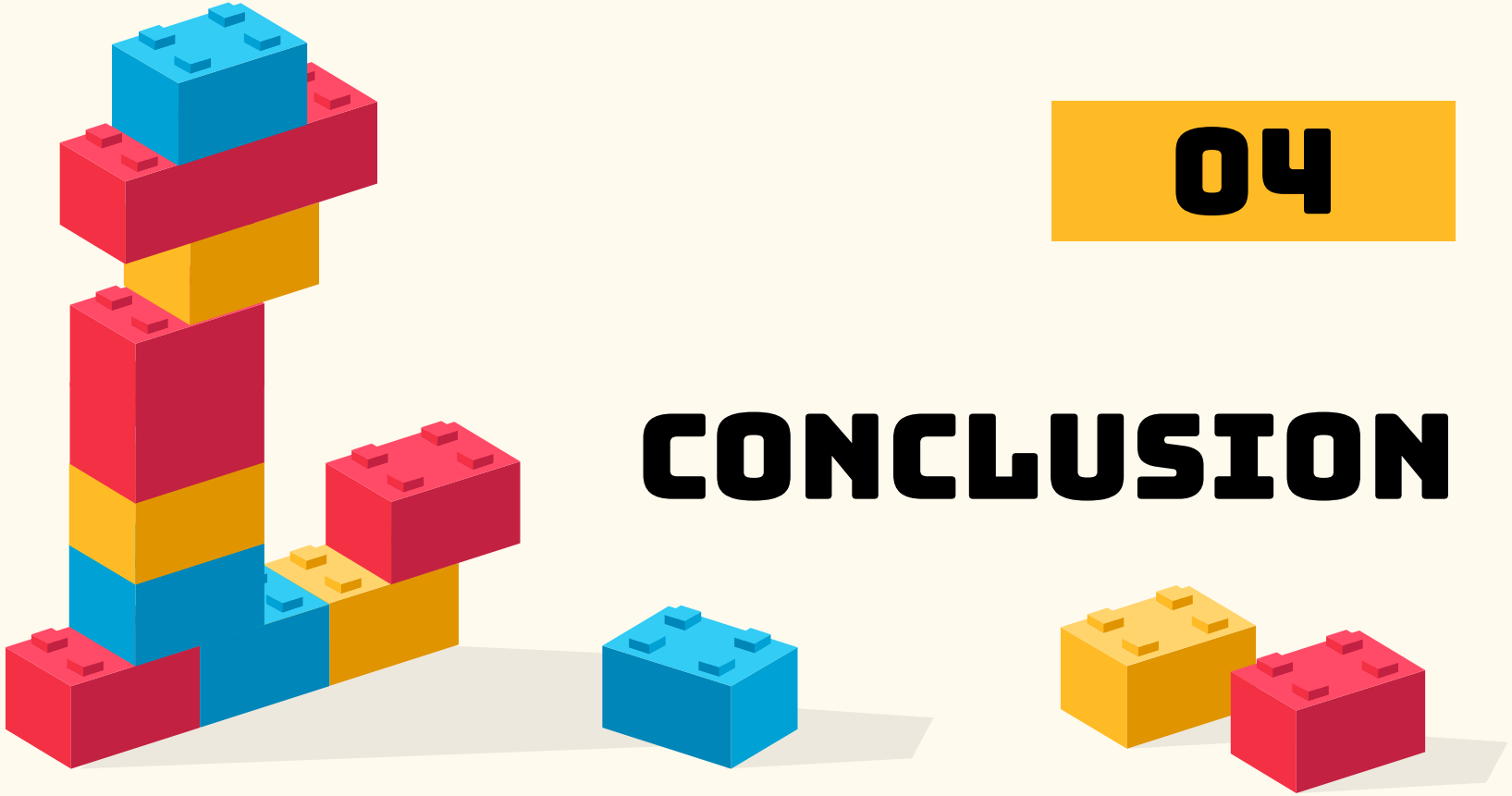


ADABOOST



Model	Train Score	Test SCore	Specificity	F1 Score
KNN	0.85	0.76	0.49	0.69
Logistic Regression	0.84	0.81	0.63	0.77
Adaboost	0.89	0.83	0.65	0.79





04

CONCLUSION

CONCLUSION

WHAT CAN INVESTORS AND ENTREPRENEURS LOOK OUT FOR TO PREDICT A SUCCESS OF A STARTUP?

ADABOOST

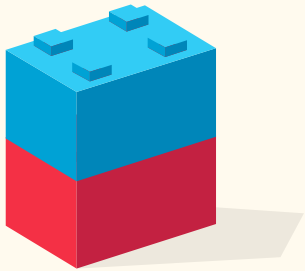
F1 Score: 79%

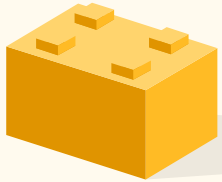
FEATURES

Number of funding rounds & category code

DATASET

Trained on 1000 company results
65% Acquired
35% Closed





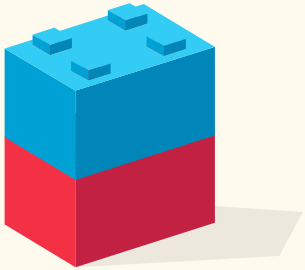
RECOMMENDATIONS

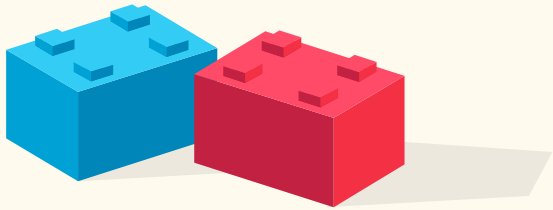
MARKETING AND VISIBILITY

Increase visibility of brand can increase revenue
Collaborations with talents

OFFERS AND BUNDLES DURING LAUNCH

Increase demand and create customer base

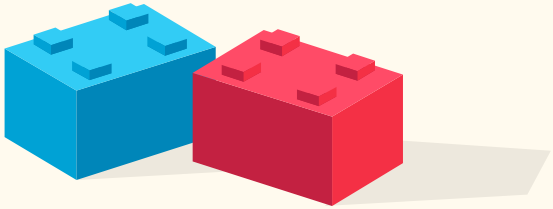




LIMITATIONS AND FUTURE WORK

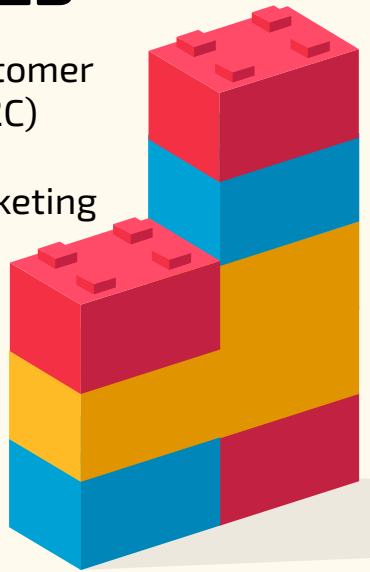
LIMITED DATASET

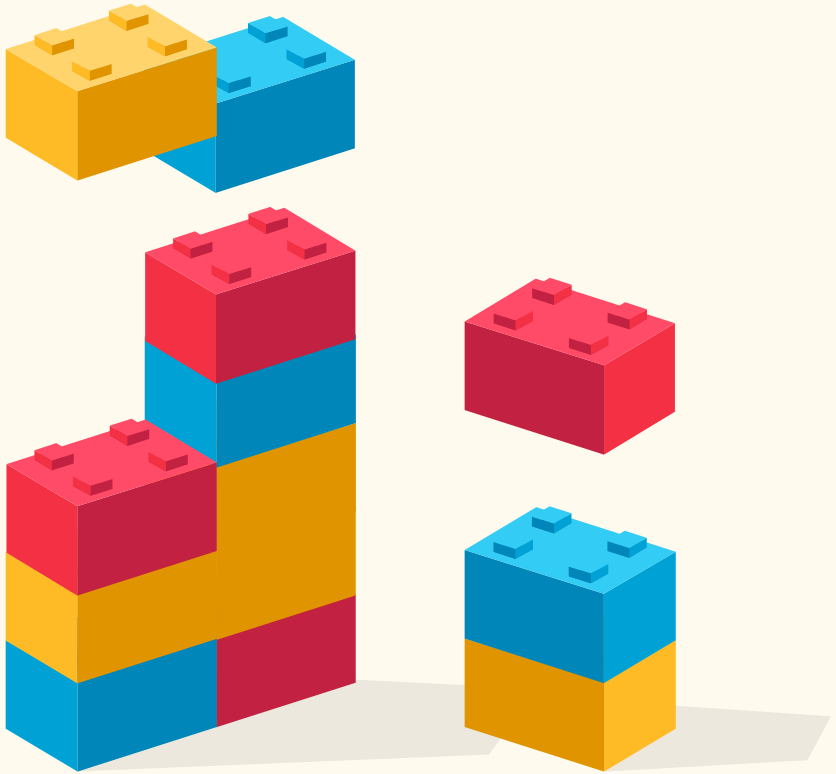
Dataset is specific to the US
Singapore startup can be much more varied
Dataset collected from 2000 to 2010



MORE FEATURES

Business to customer models (B2B, B2C)
Company size
Timing and marketing





THANKS!

Does anyone have any questions?

CREDITS: This presentation template was created by **Slidesgo,** including icons by **Flaticon**, infographics & images by **Freepik**

Please keep this slide for attribution