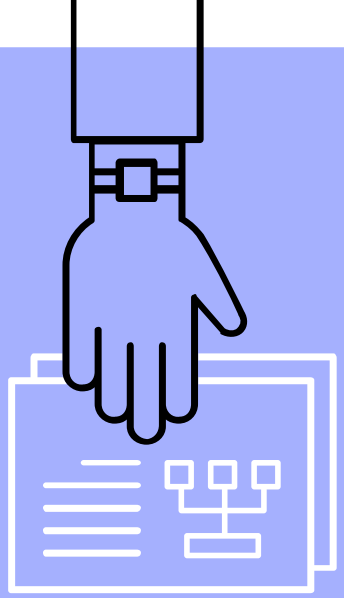
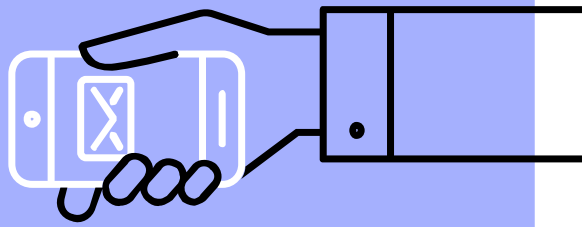
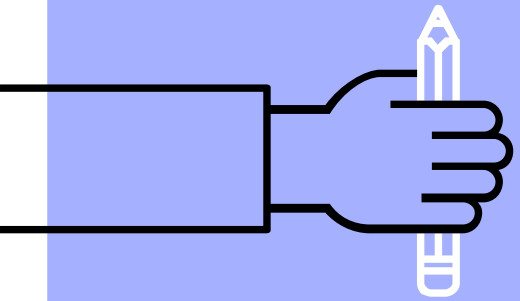
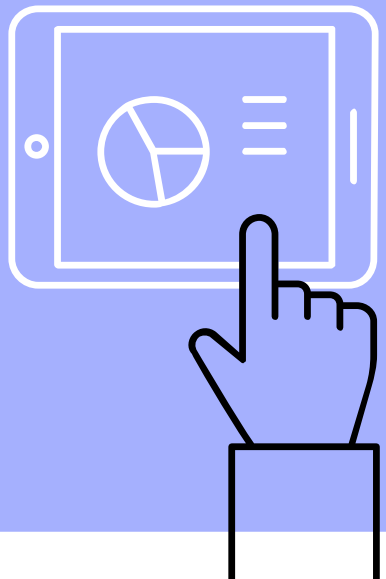




Saving Your Credit Score: Analysis of Small Business Loans





By Stephanie, Morgan, Marisol, Katalina

Kaggle:
Should This Loan be Approved or Denied?

- A large and rich dataset from the U.S. Small Business Administration (SBA)

- SBA loans from 1970-2014

Dataset Size

Rows: 899164 Columns: 27

27 Columns

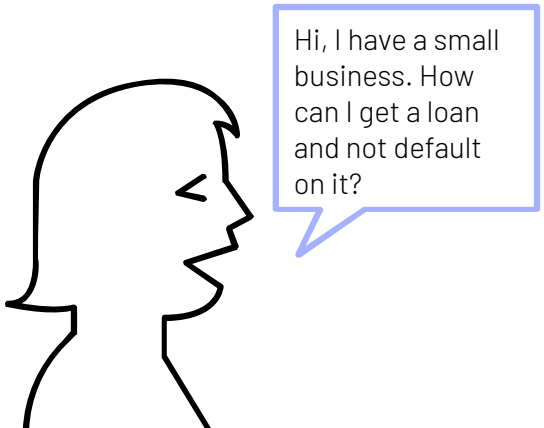
Name of business, bank, state, gross amount, term, approval date, default date...

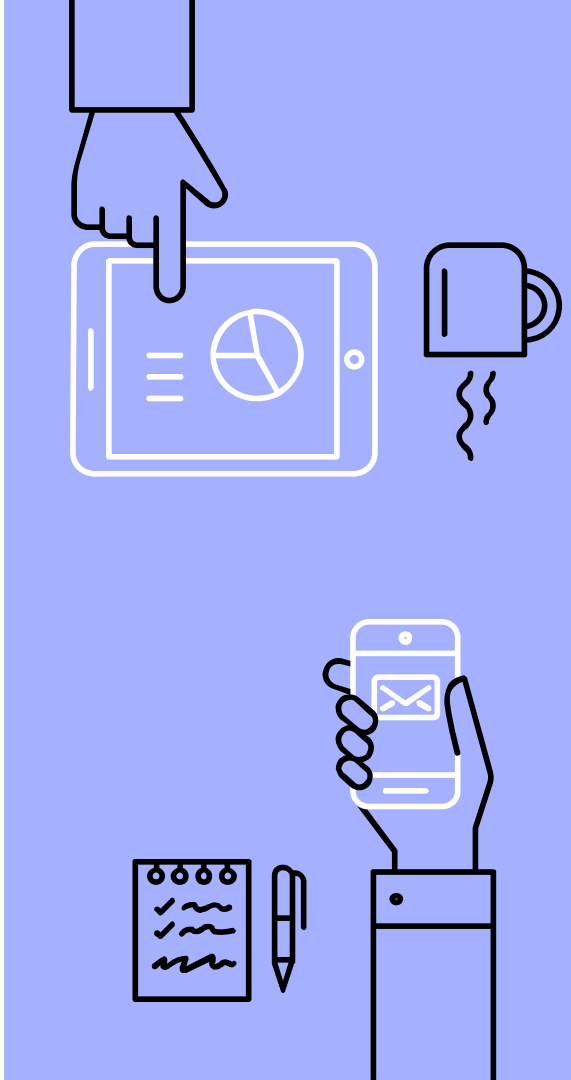
Data Set

	loannr_chkdgt	name	city	state	zip	bank	bankstate	approvaldate	approvalfy	term	noemp	newexist	createjob
0	1000014003	ABC HOBBYCRAFT	EVANSVILLE	IN	47711	FIFTH THIRD BANK	OH	1997-02-28	1997	84	4	2.0	0
1	1000024006	LANDMARK BAR & GRILLE (THE)	NEW PARIS	IN	46526	1ST SOURCE BANK	IN	1997-02-28	1997	60	2	2.0	0
2	1000034009	WHITLOCK DDS, TODD M.	BLOOMINGTON	IN	47401	GRANT COUNTY STATE BANK	IN	1997-02-28	1997	180	7	1.0	0
3	1000044001	BIG BUCKS PAWN & JEWELRY, LLC	BROKEN ARROW	OK	74012	1ST NATL BK & TR CO OF BROKEN	OK	1997-02-28	1997	60	2	1.0	0
4	1000054004	ANASTASIA CONFECTIONS, INC.	ORLANDO	FL	32801	FLORIDA BUS. DEVEL CORP	FL	1997-02-28	1997	240	14	1.0	7

Goals

- What factors tend to cause a loan to be defaulted on.
- Create a model to predict which approved loan applications will default on their loan.





What is SBA ?
Requirements and Tips

The Small Business Administration

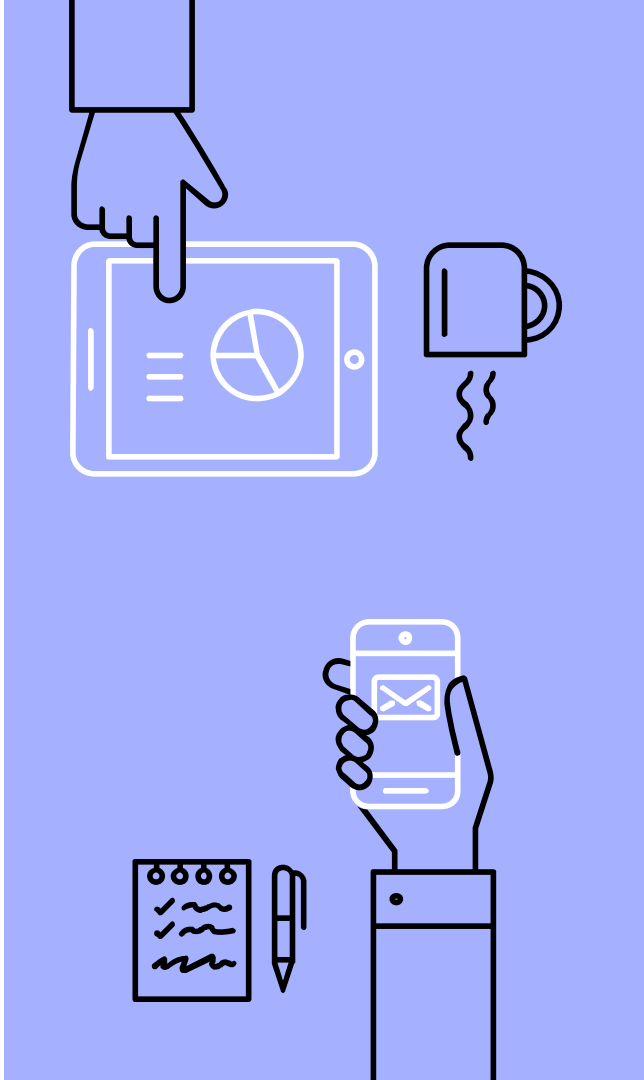
- Provides support to small businesses and entrepreneurs.
- Connects businesses with lenders to help grow their business.

Requirements

- Must be a legally operating business in the USA.
- Cannot be funded by another lender.

Deciding factors

- How you get your income
- Business location
- Business size
- New or existing



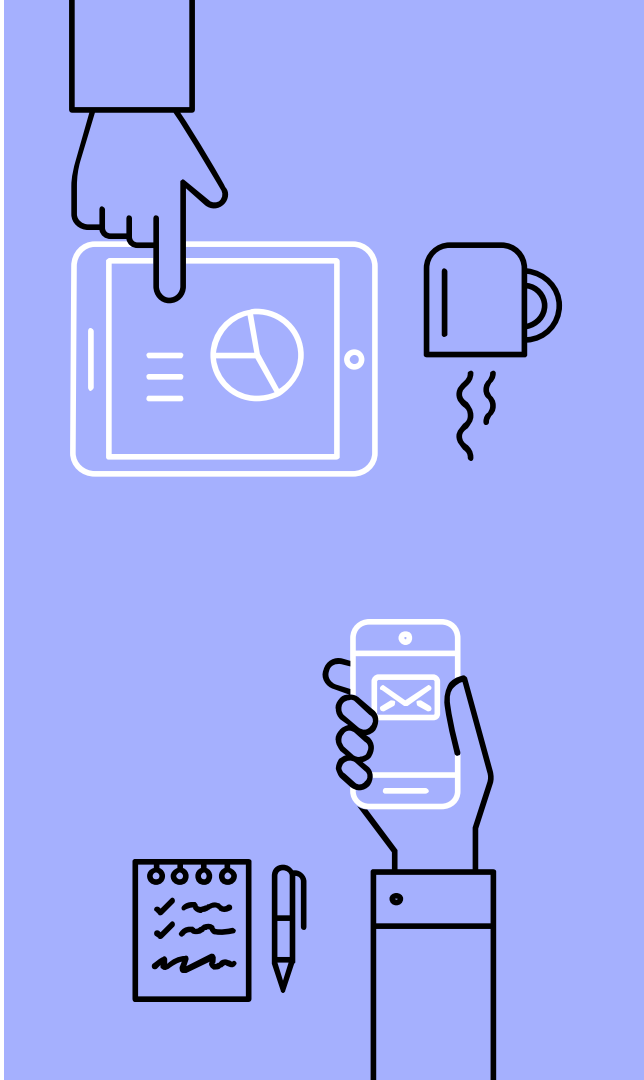
How to get a loan?

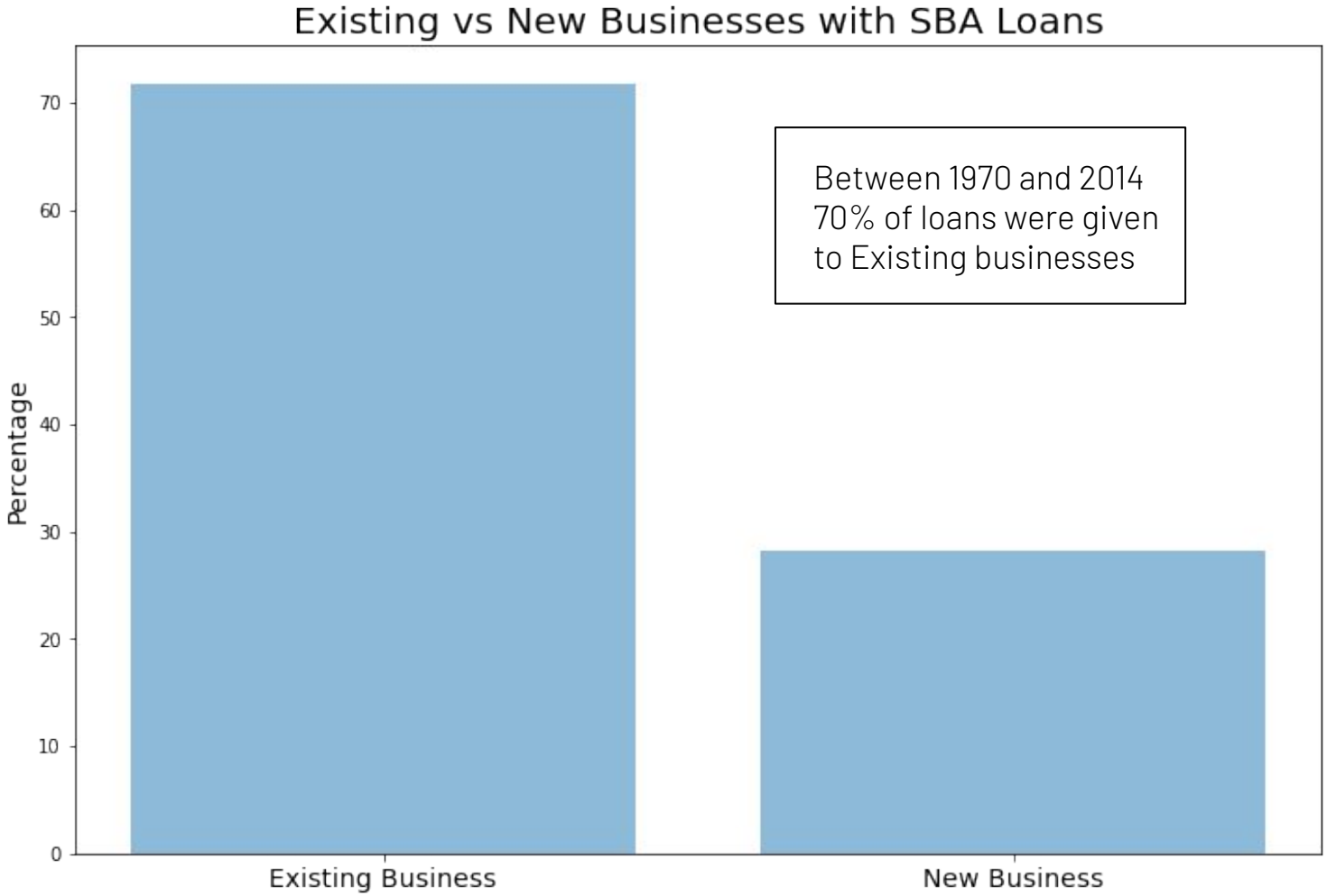
Personal

- Credit Score
- Interview with the banks
- Personal Record

Business

- Existing or New
- Bank
- Industry
- Loan Terms
- Loan Amounts

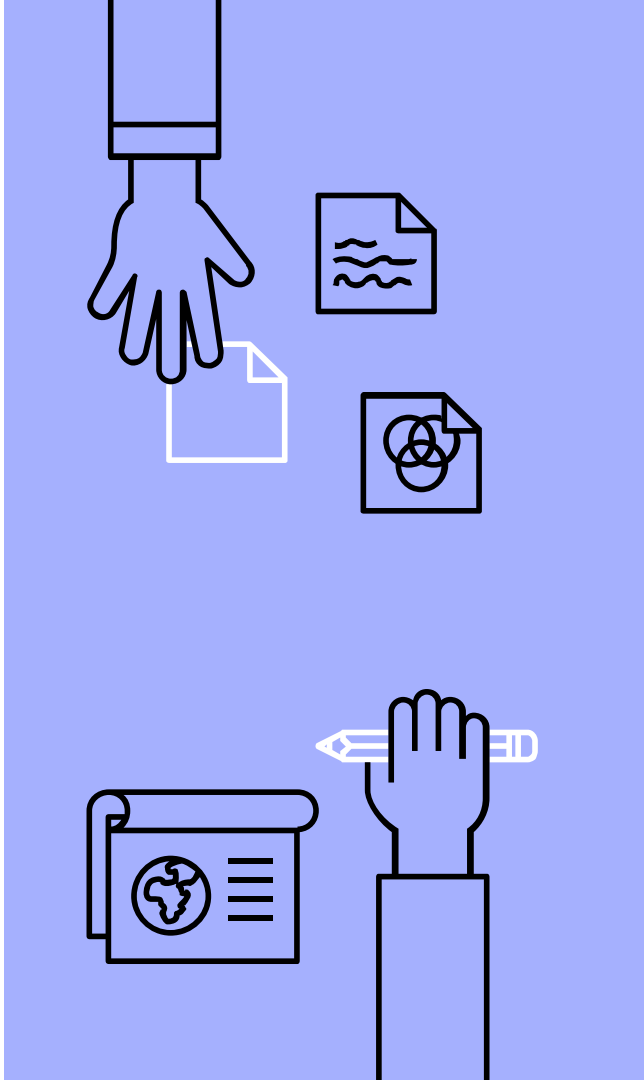


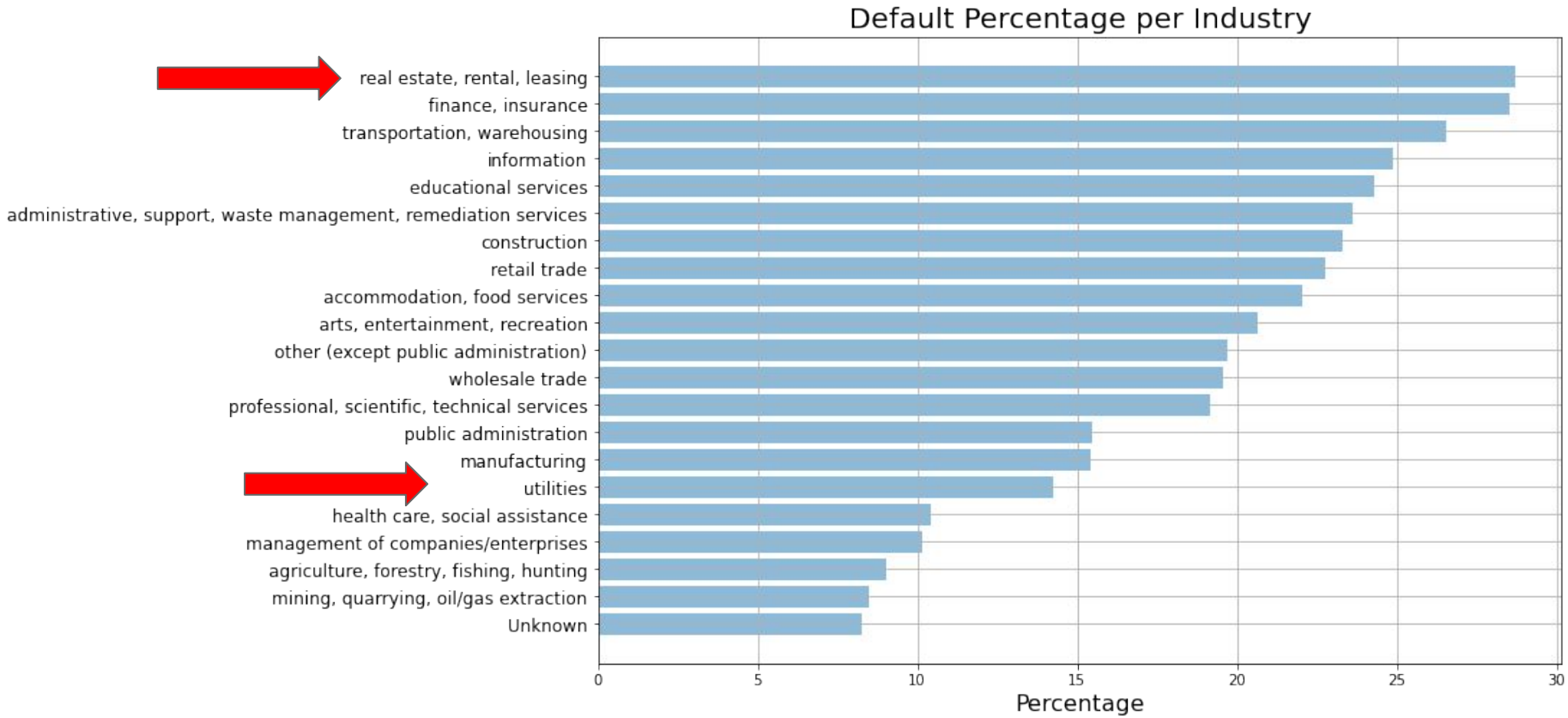

Existing vs New Businesses with SBA Loans
Percentage
0
10
20
30
40
50
60
70
Existing Business
New Business
Between 1970 and 2014
70% of loans were given
to Existing businesses

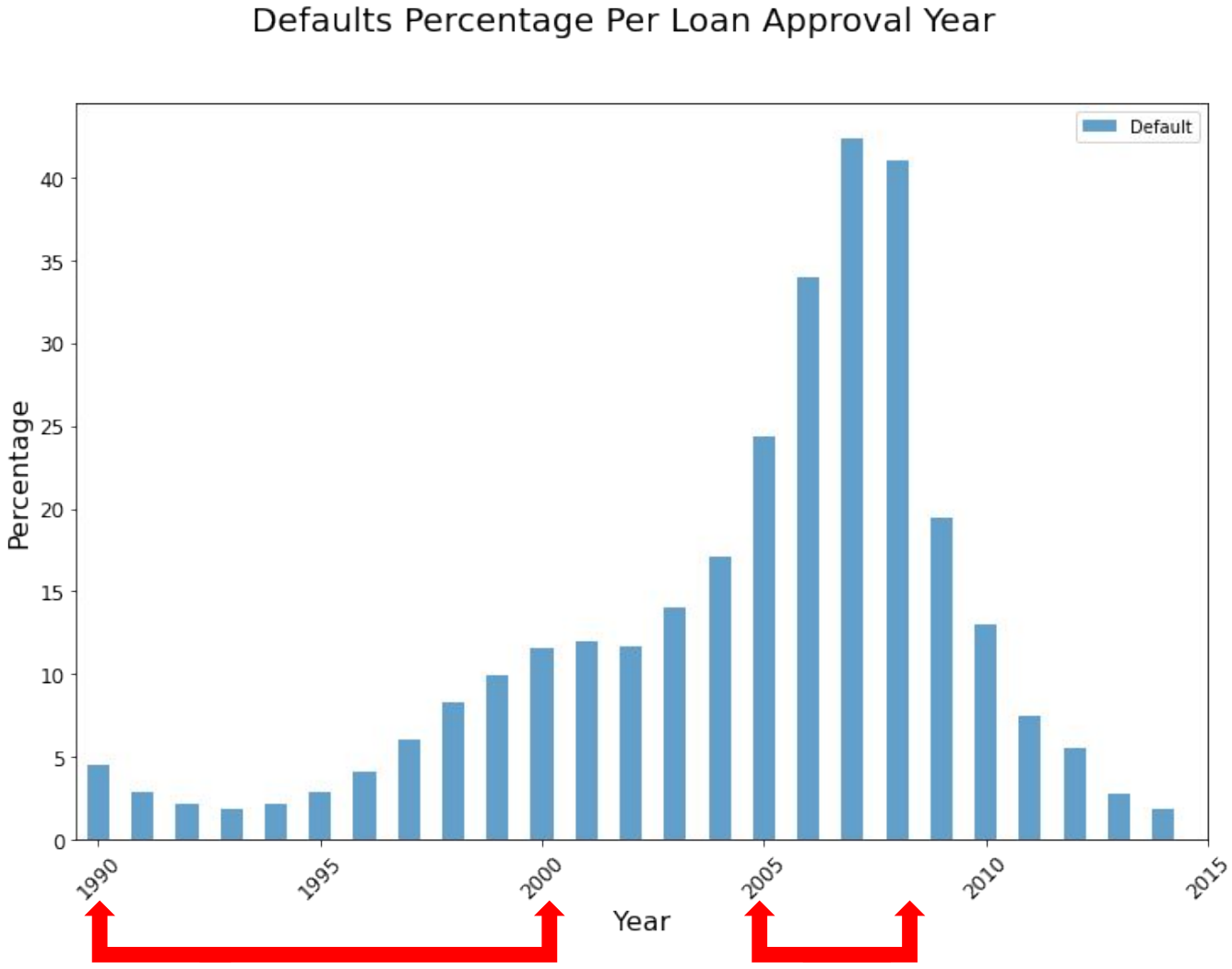
Factors that play a role in defaultness

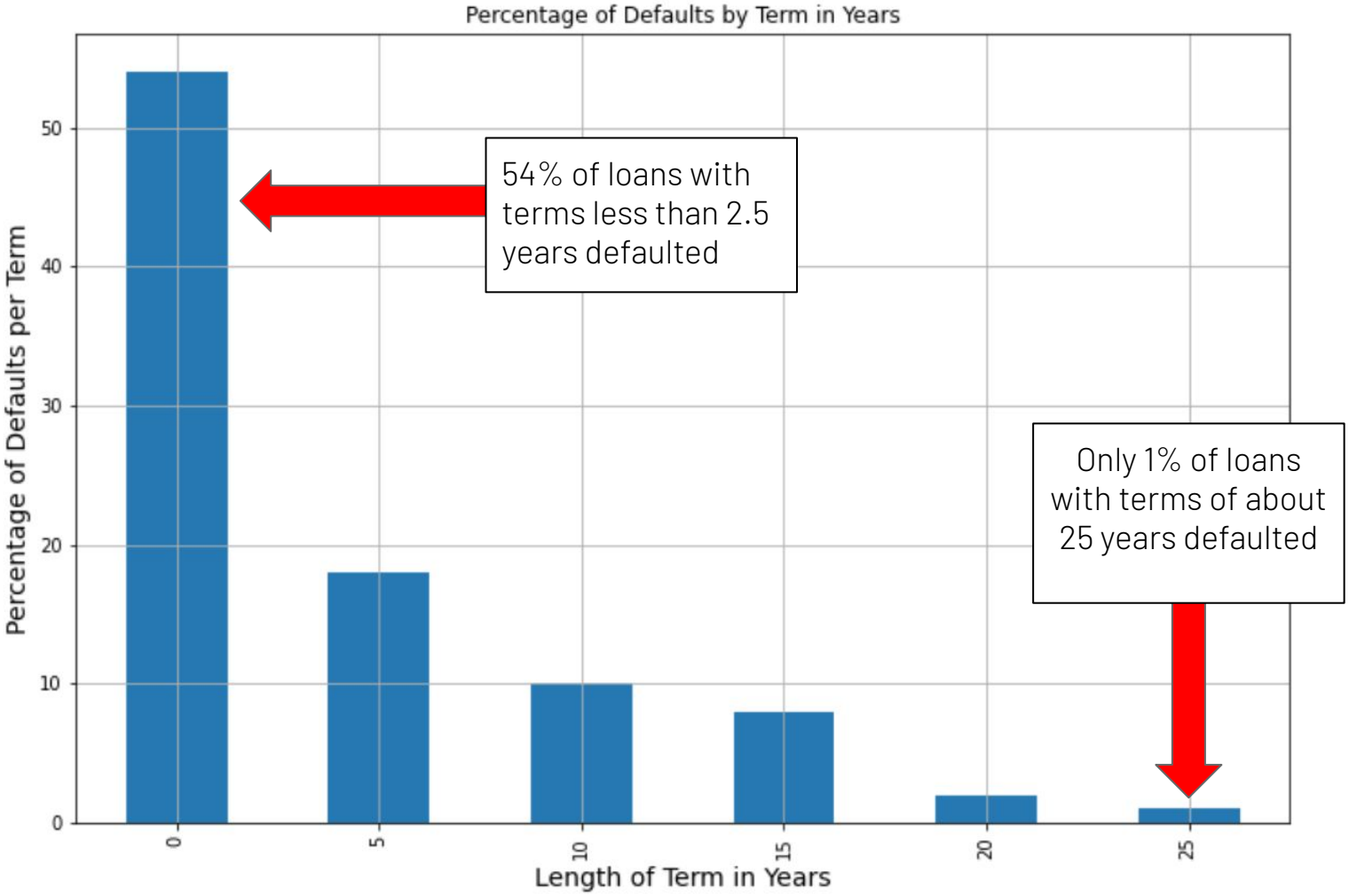
- Identified 5 factors from the SBA website as well as Business.org
- We analyzed these factors on the SBA Loan Dataset to see if they were true

Features
Industry
Year
Loan Term
Loan Amount
Bank



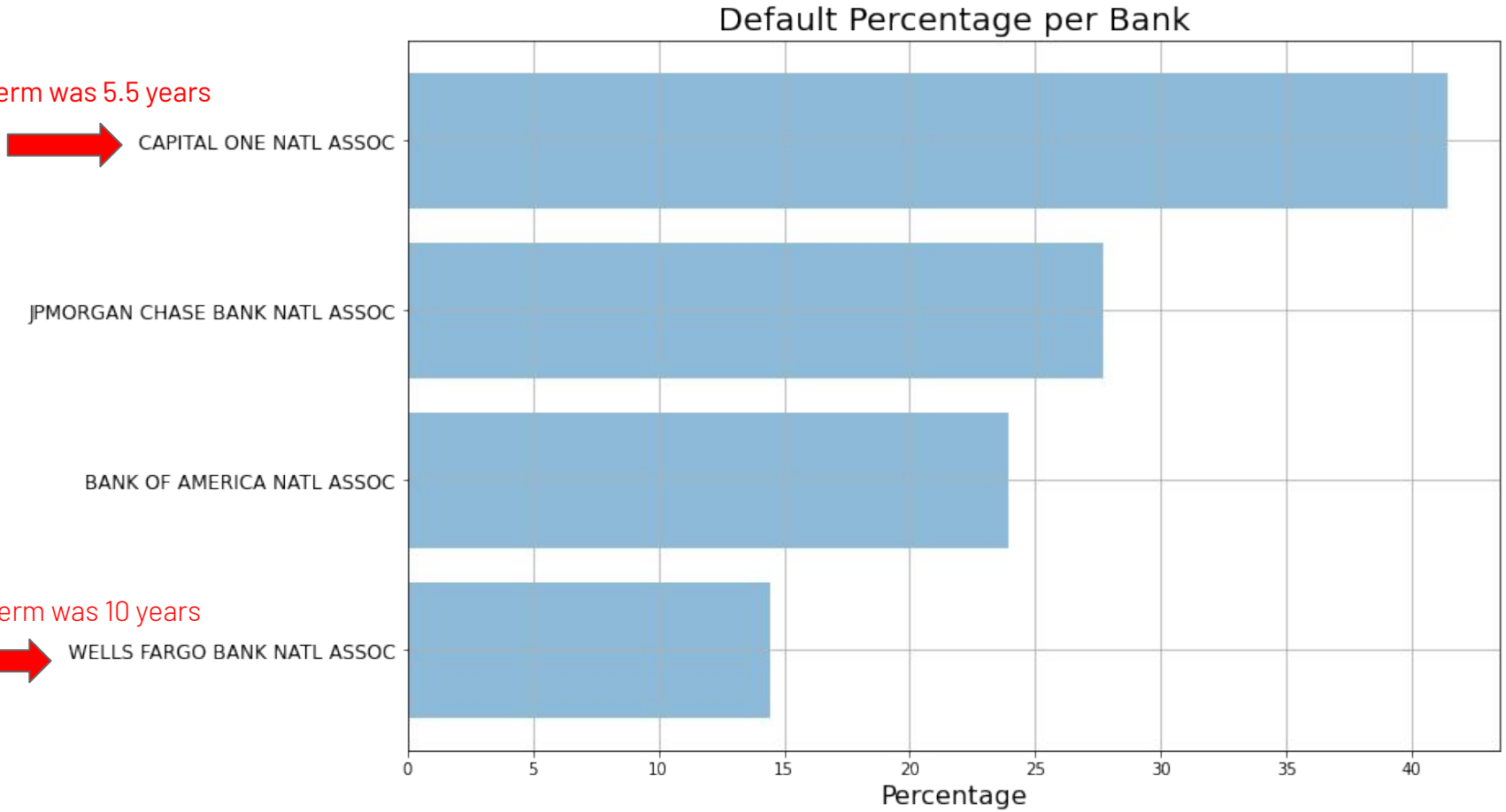

Default Percentage per Industry
real estate, rental, leasing
finance, insurance
transportation, warehousing
information
educational services
administrative, support, waste management, remediation services
construction
retail trade
accommodation, food services
arts, entertainment, recreation
other (except public administration)
wholesale trade
professional, scientific, technical services
public administration
manufacturing
utilities
health care, social assistance
management of companies/enterprises
agriculture, forestry, fishing, hunting
mining, quarrying, oil/gas extraction
Unknown
0
5
10
15
20
25
30
Percentage


Defaults Percentage Per Loan Approval Year
Default
Percentage
0
5
10
15
20
25
30
35
40
1990
1995
2000
2005
2010
2015
Year


Percentage of Defaults by Term in Years
Percentage of Defaults per Term
0
10
20
30
40
50
0
5
10
15
20
25
Length of Term in Years
54% of loans with terms less than 2.5 years defaulted
Only 1% of loans with terms of about 25 years defaulted

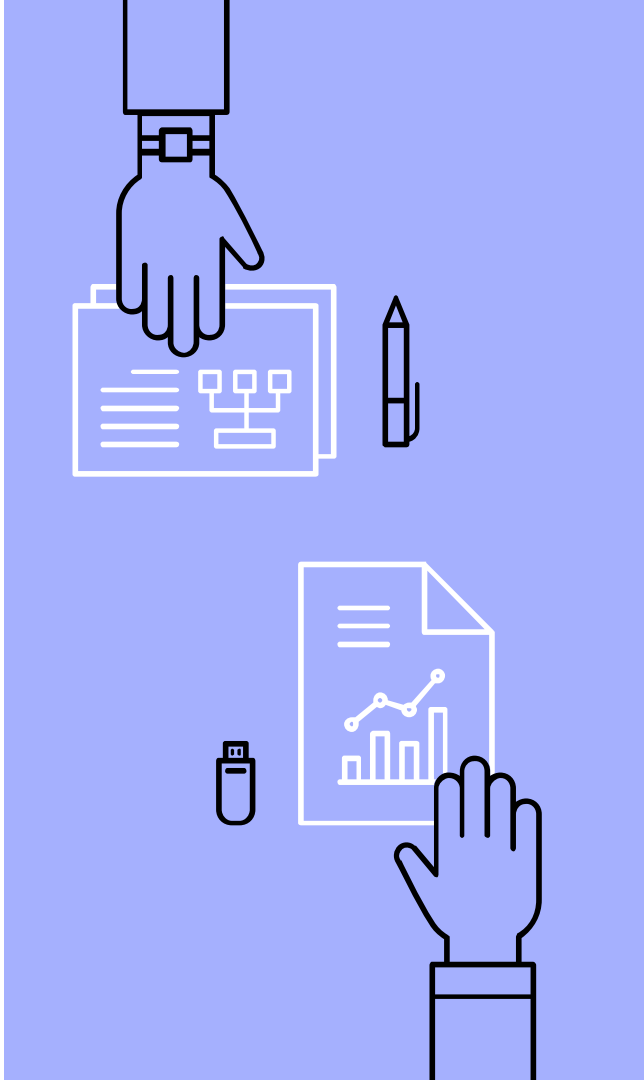
Average term was 5.5 years

Average term was 10 years



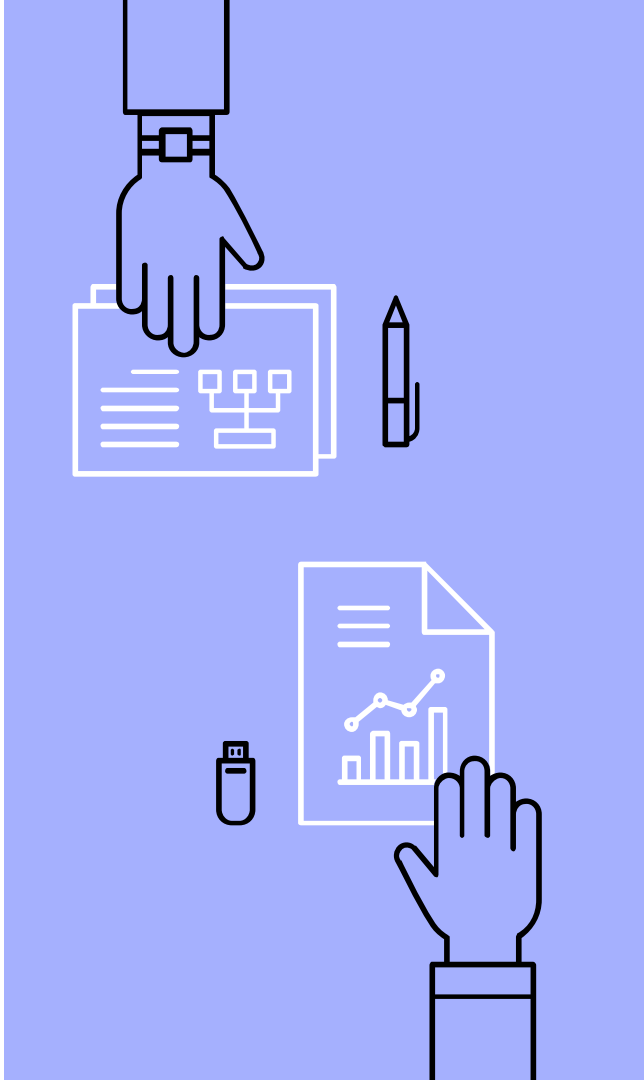
Model (XGBoost)

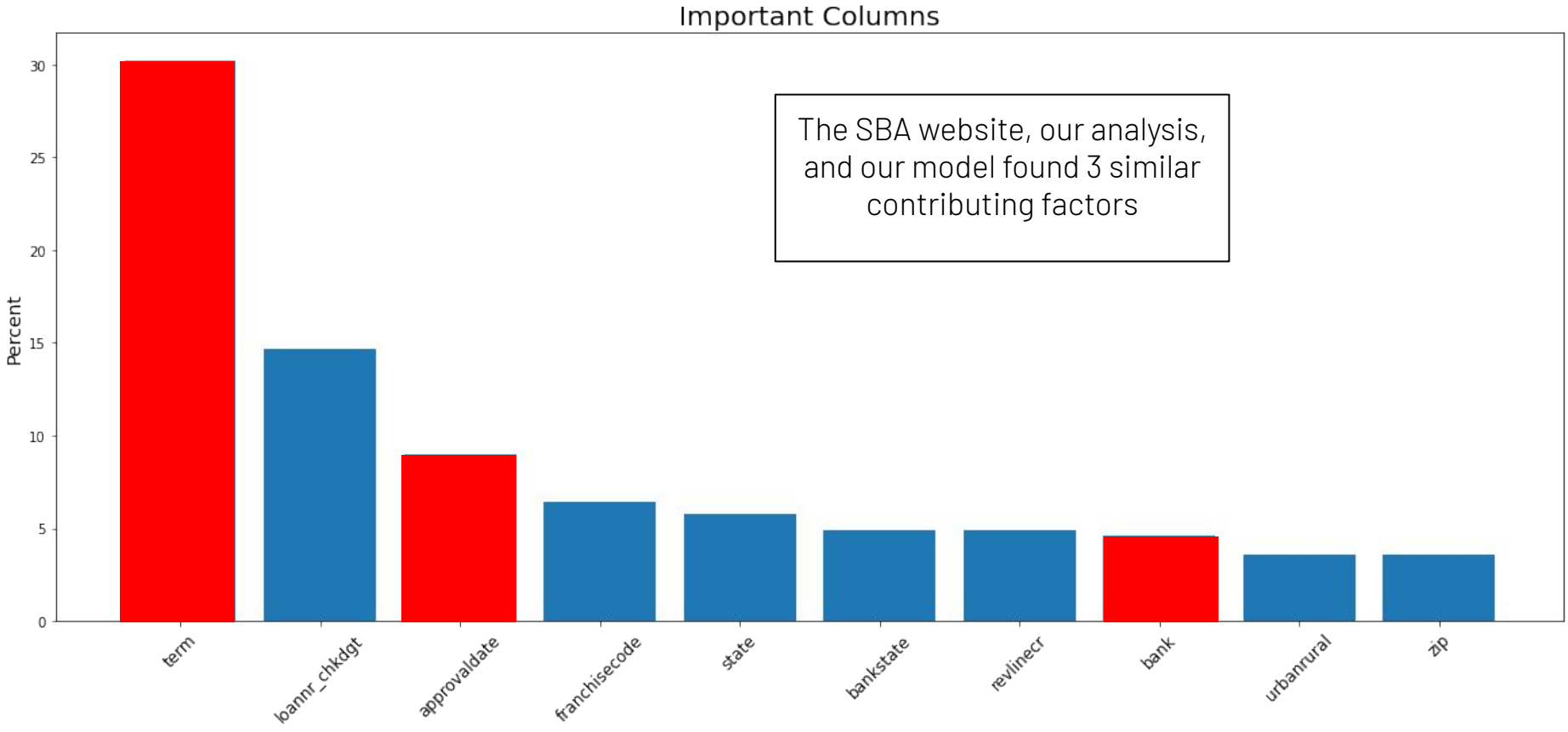
- Decision tree-based model
- Decision trees can capture non-linearity in data
- In XGBoost, multiple decision trees are built subsequently (called “boosting”)
- Builds current decision tree based on where previous decision tree did poorly



Model (XGBoost)

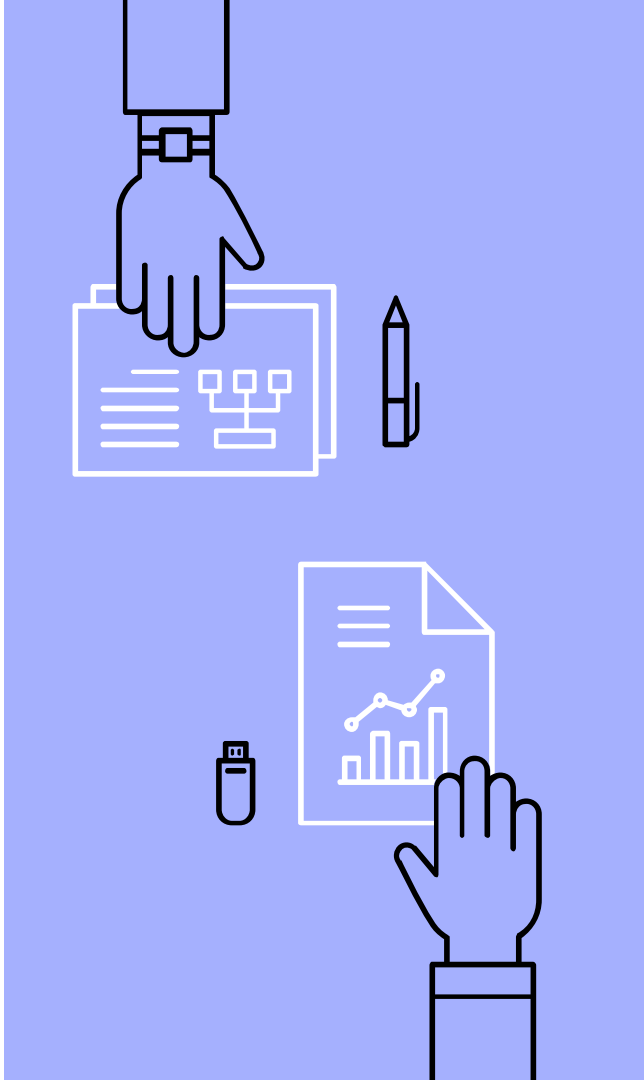
- Predicts whether a loan will default or not based on the SBA Loan dataset from 1970-2014
- Model will help US economy come out of the recession by making sure loans given out won't default
- 90% precision on predicting defaulted loans




Important Columns
Percent
0
5
10
15
20
25
30
term
loannr_chkdgt
approvaldate
franchisecode
state
bankstate
revlinecr
bank
urbanrural
zip
The SBA website, our analysis, and our model found 3 similar contributing factors

In summary, if you're looking to get a loan:

- Be aware that some industries are riskier than others
- The current pandemic can increase your likelihood of defaulting
- Go for a longer term loan (ideally 20-25 years) if possible
- Wells Fargo is the safest bet out of popular banks



$200,000
Average loan amount given by the SBA

20,000
Average amount of SBA loans given yearly

$400,000,000
Saved using our model

THANKS!

Any questions?