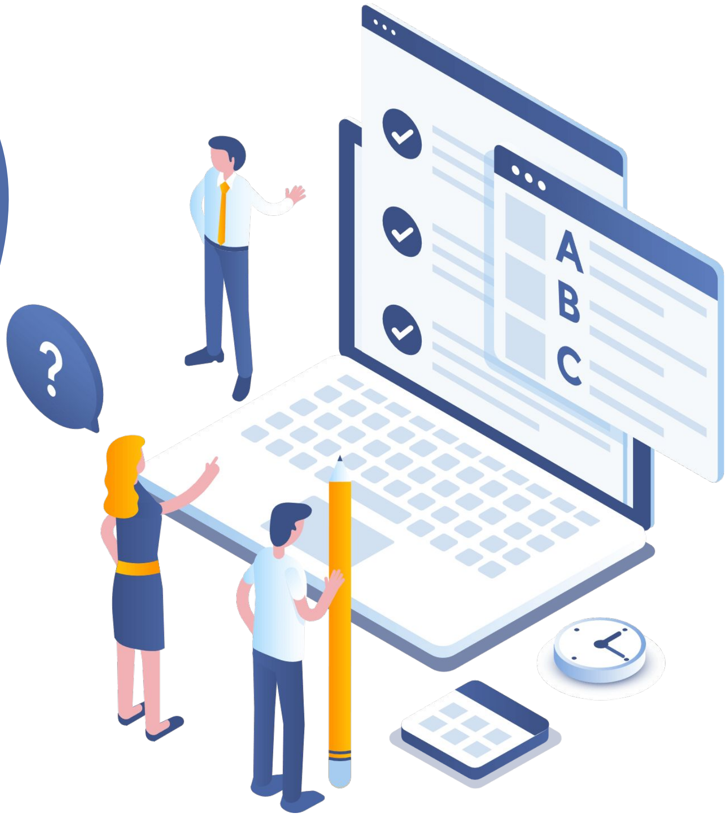
FI Consulting Case Study

Analyzing Climate Hazard Risk to a Loan Portfolio

Hrolfur Sveinsson, Peijia Wu, Yixuan Yang

Dec 13, 2021



Contents

PART 01

General View

- Problem Understanding & Methodology
- Data Understanding
- Results & Conclusions

Introduction -- Problem Understanding & Methodology

Have historical climate hazards had a significant impact on loan risk? If yes, to what extent?

- Determine which data sets and what part of proposed data sets are needed
- Determine metrics to measure credit risk & climate hazard risk
- Integrate climate dataset into loan risk dataset
- Explore the portfolio’s exposure to climate hazards -- Visualizations
- Explore the relationship between the loan portfolio and climate risk

Data Understanding

Loan Data

Field Name	Definition
BorrName	Borrower name
BorrStreet	Borrower street address
BorrCity	Borrower city
BorrState	Borrower state
BorrZip	Borrower zip code
GrossApproval	Total loan amount
GrossChargeOffAmount	Total loan balance charged off (includes guaranteed and non-guaranteed portion of loan)
ApprovalFiscalYear	Fiscal year the loan was approved
InitialInterestRate	Initial interest rate - total interest rate (base rate plus spread) at time loan was approved
NaicsCode	North American Industry Classification System (NAICS) code
ProjectCounty	County where project occurs
ProjectState	State where project occurs

Climate Data

OID_	Hazard	Prefix	Start	End_
1	Avalanche	AVLN	1960	2019
2	Coastal Flooding	CFLD	N/A	N/A
3	Cold Wave	CWAV	2005	2017
4	Drought	DRGT	2000	2017
5	Earthquake	ERQK	2017	2017
6	Hail	HAIL	1986	2017
7	Heat Wave	HWAV	2005	2017
8	Hurricane	HRCN	1851/1949	2017
9	Ice Storm	ISTM	1946	2014
10	Landslide	LNDS	2010	2019
11	Lightning	LTNG	1991	2012
12	Riverine Flooding	RFLD	1996	2019
13	Strong Wind	SWND	1986	2017
14	Tornado	TRND	1986	2019
15	Tsunami	TSUN	1800	2018
16	Volcanic Activity	VLCN	9310BC	2018
17	Wildfire	WFIR	2016	2016
18	Winter Weather	WNTW	2005	2017

- **Number of climate hazard events**
- **Risk scores & Expected annual loss (total value & building & agriculture)**

Results & Conclusions

Data Cleaning and Integration

State-county FIPS code: 5-digit code

	R^2		RMSE	
Selected Model	Initial Interest Rate	Loan Loss Ratio	Initial Interest Rate	Loan Loss Ratio
Ridge Regression	0.0198	0.2632	6.7998	0.1737
ElasticNet Regression	-2.908e-10	-2.908e-10	1.6322	0.2024
Gradient Boosting Machine	0.0379	0.9932	1.60096	0.0167
XG Boost	0.047552	0.9994	1.592915	0.0051
Random Forest	-0.0895	0.99993	1.7037	0.0017

PART 02

Visualizations

Visualizations

02 Data Visualization Agenda

- Which states and counties are exposed to climate hazard risk?
  - Tornado in Kentucky
    - Map Chart Visualization
- Is there an obvious macroeconomic relationship between climate hazard risk and credit risk?
  - Average Initial Interest Rate
    - Scatter Plot Visualization
- Let’s jump into Tableau!

PART 06

Challenges & Improvements

Challenges

1. Insufficient and Unspecific Variables in the Input Data

2. Lack a Better Measurement for Credit Risk

3. Data Granularity

4. Data Processing and Modeling

Challenges

1. Insufficient and Unspecific Variables in the Input Data

- Lack extent of damage to factories, physical injuries to people from climate hazards
- Lack time about when a specific climate hazard happens in the climate data
- Lack companies’ financial data in the loan data

Improvements (Next Steps)

1. Data! Data! Data!

- Focus on several companies on the specific regions (south/west);
- Collect detailed financial data including the changes in cash flows and balance sheets, etc.
- Collect detailed climate data including occurrence time, duration, changes of climate hazards, etc.

Challenges

2. Lack a Better Measurement for Credit Risk

- Initial interest rate
- Loan loss ratio = Gross Charge-Off Amount / Gross Approval

Improvements (Next Steps)

2. Find a Better Metric for Credit Risk

- Credit Ratings/scores
- Default or not, (for classification)
- Amount of liability to the bank

Challenges

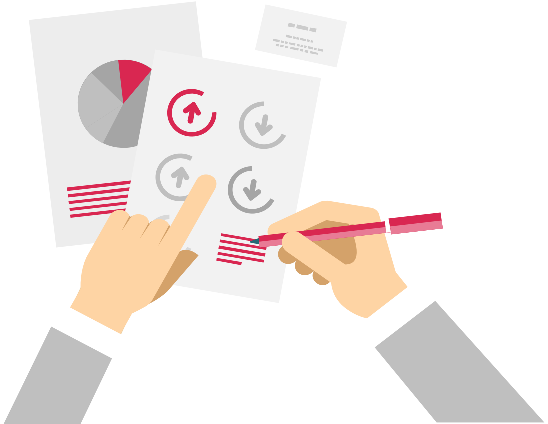
3 Data Granularity

- County-level is not rigorous

Improvements（Next Steps）

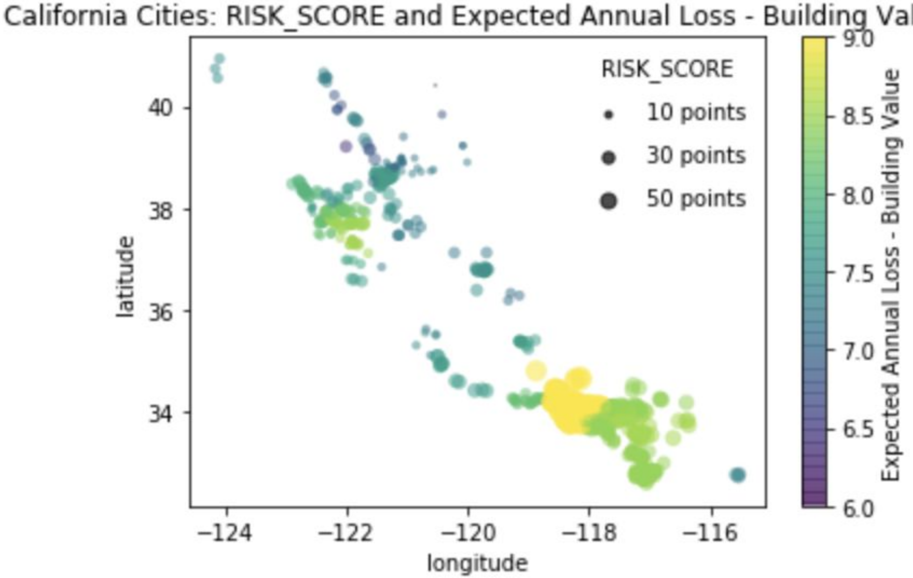
3.Find the Right Level of Data Granularity

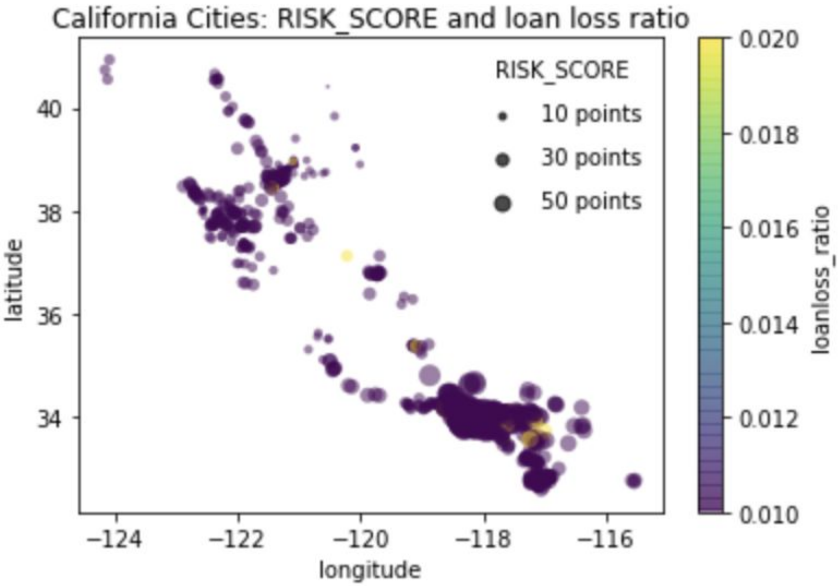
- Census-tract level/ Firm-level(coordinates)



Case Study_Smaller Granularity

California Construction Companies Year 2018





- Google API; Google Sheet Extension: Geocode by Awesome Table;
- US Census Geocoder API (package:censusgeocode) 10,000 rows max https://geocoding.geo.census.gov/geocoder/geographies/coordinates?form
- Third-party Platform: https://www.geocod.io/upload/
- Others: https://andrewpwheeler.com/tag/geocoding/

Challenges

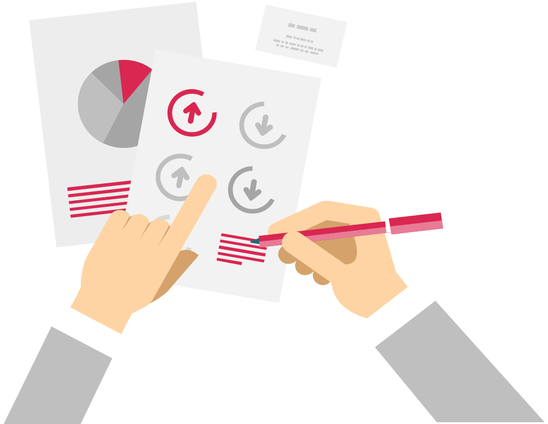
4. Data Processing and Modeling

- Missing values in Initial interest rate, NAICS code
- Removing the outliers of two variables only

Improvements (Next Steps)

4. More Rigorous Data Processing and Models

- Utilize more rigorous methods to do data cleaning and outlier removing
- Typo/ special symbols checking in county name/ street address
- Multicollinearity issues
- Advanced machine learning techniques such as SVM, polynomial regression



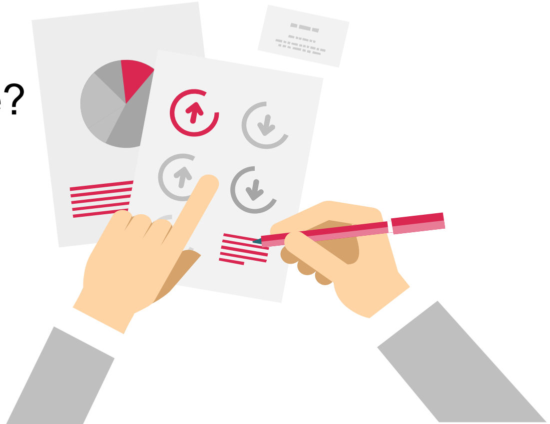
Improvements（Next Steps）

5.Multi-angle Analysis of the impact of Physical Risk on the Credit Risk

- Direct/ Indirect/ <u>Macroeconomic impact</u>

6.Do Scenario Analysis & Predictive Analysis

- Financial impact of physical risk in different scenarios
- How climate hazards impact on the credit risk in the future?



Improvements（Next Steps）

1.Data! Data! Data!

2.Find a Better Metric for Credit Risk

3.Find the Right Level of Data Granularity

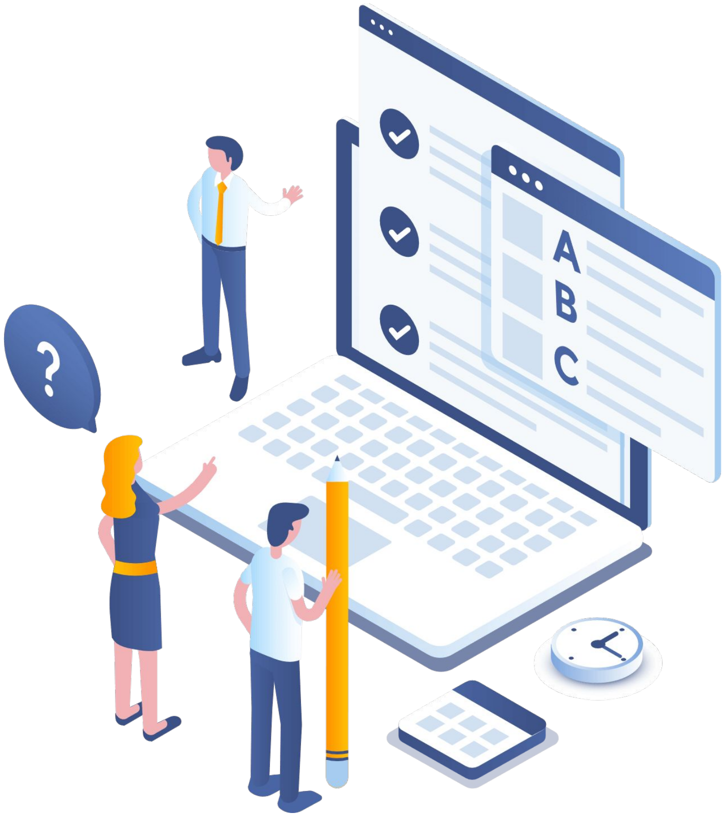
4.More Rigorous Data Processing and Models

5.Multi-angle Analysis of the impact of Physical Risk on Credit Risk

6.Do Scenario Analysis & Predictive Analysis

Questions?

Thank You!



Data Preparation

01 Data Cleaning and Integration

State-county FIPS code: 5-digit code

02 Data Exploration

- **Import Industry sector and title**
- **Import Regions & Divisions**
- **Relationship/ correlation exploration**

Initial Interest Rate

	Pearson Correlation Coefficient
DRGT_RISKS	0.017505
ISTM_EALP	0.016083
RFLD_RISKS	0.015303
TRND_EALT	0.013108
CFLD_RISKS	0.013099
TRND_RISKS	0.012869
SOVI_SCORE	0.012861
TRND_EALP	0.012573
HRCN_EXPP	0.012448
HRCN_EXPT	0.012442