BUSINESS SURVIVAL ANALYSIS SAN FRANCISCO DISTRICT 3

Presented by Abirami

Presentation Held on May 8th 2025

BACKGROUND

Context:

- District 3 includes diverse neighborhoods like Chinatown, North Beach, and parts of the Financial District, each with a unique business landscape.

Executive Summary:

- Goal: Predict administrative closure of businesses in District 3
- Tool: Survival Analysis with Cox Proportional Hazards model
- Outcome: Insights to inform targeted funding decisions

TARGET AUDIENCE

San Francisco District 3 funding team

GOAL

- The Main Goal of this project is to analyze the factors influencing the survival of businesses in San Francisco and for Supervisor District 3. Specifically, aim to predict the likelihood of a business being administratively closed over time using survival analysis techniques.

PROBLEM STATEMENT

Why Model Business Survival in District 3?

- Targeted Resource Allocation
- Informed Funding Strategies
- Performance Measurement
- Prioritization of Investment

DATA COLLECTION

Data Source

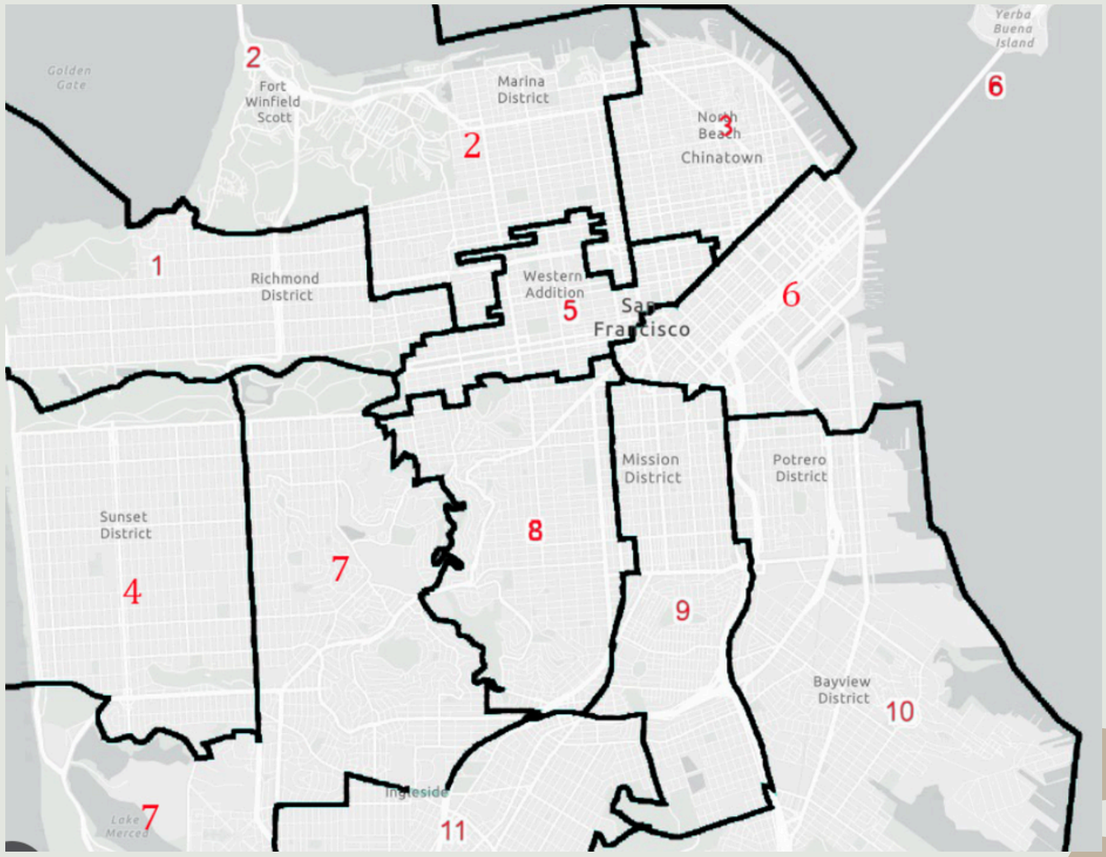
- A Python script collects data on registered businesses from the San Francisco Open Data API and saved in a CSV format for further analysis.

Web API

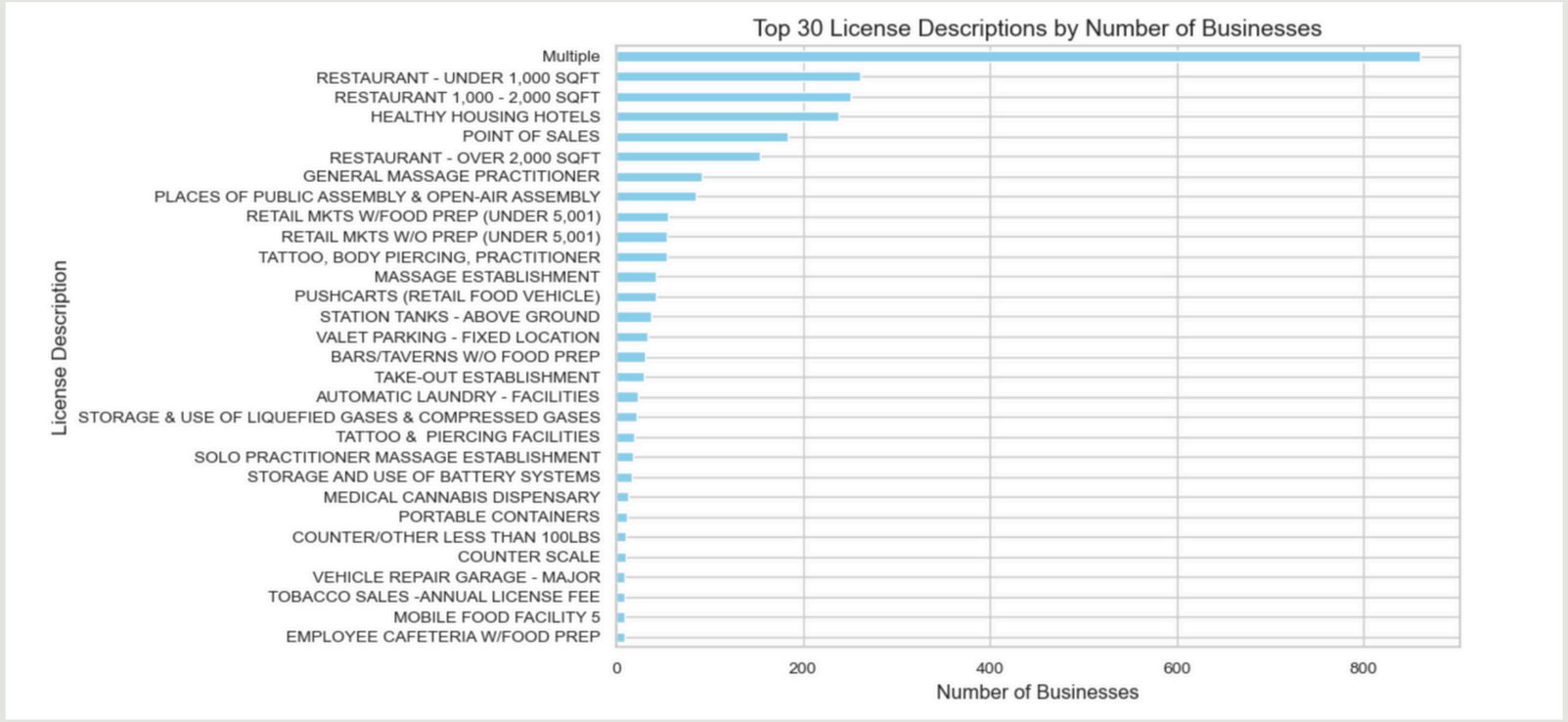
- Endpoint: https://data.sfgov.org/resource/g8m3-pdis.json
- Limit per request: 1,000 records (as per API specifications)

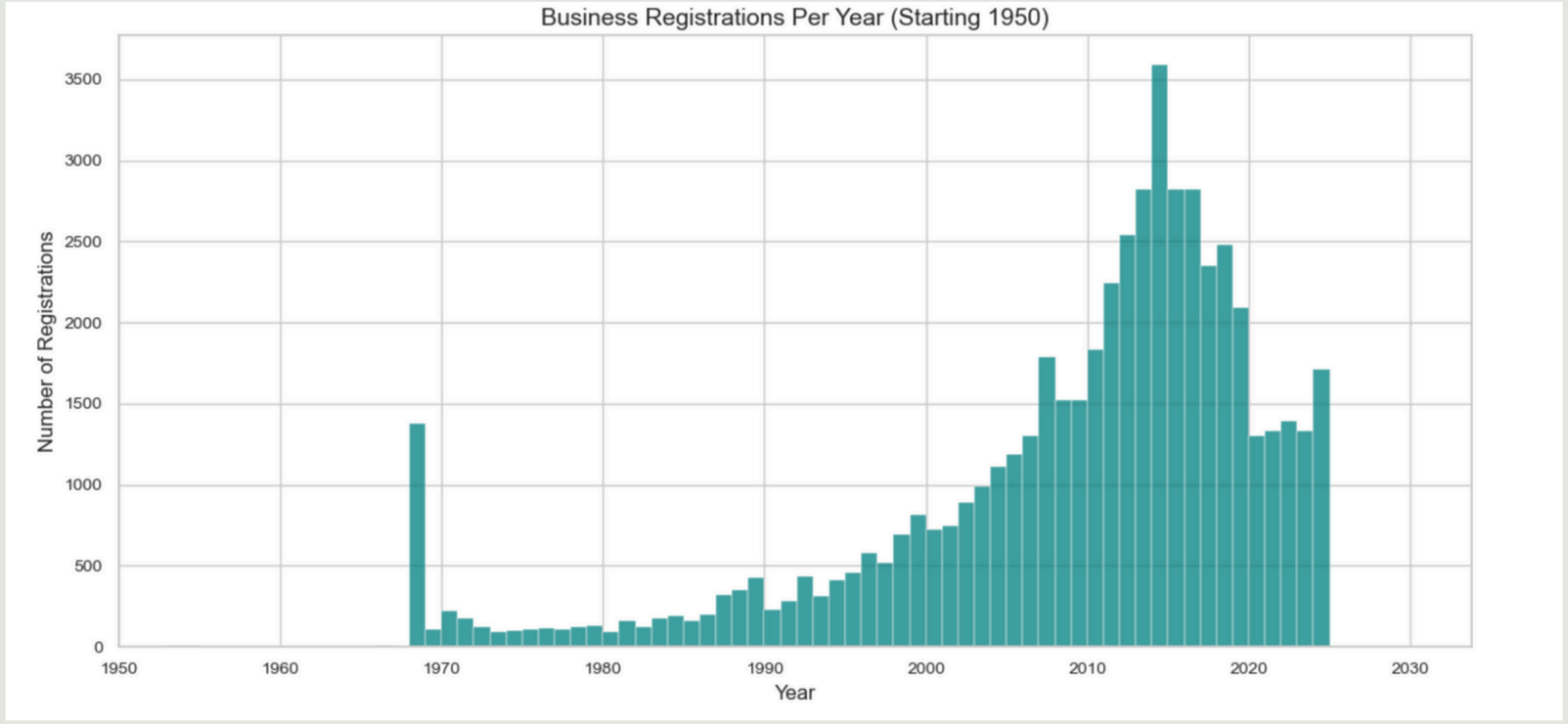
DATA CLEANING AND PREPROCESSING

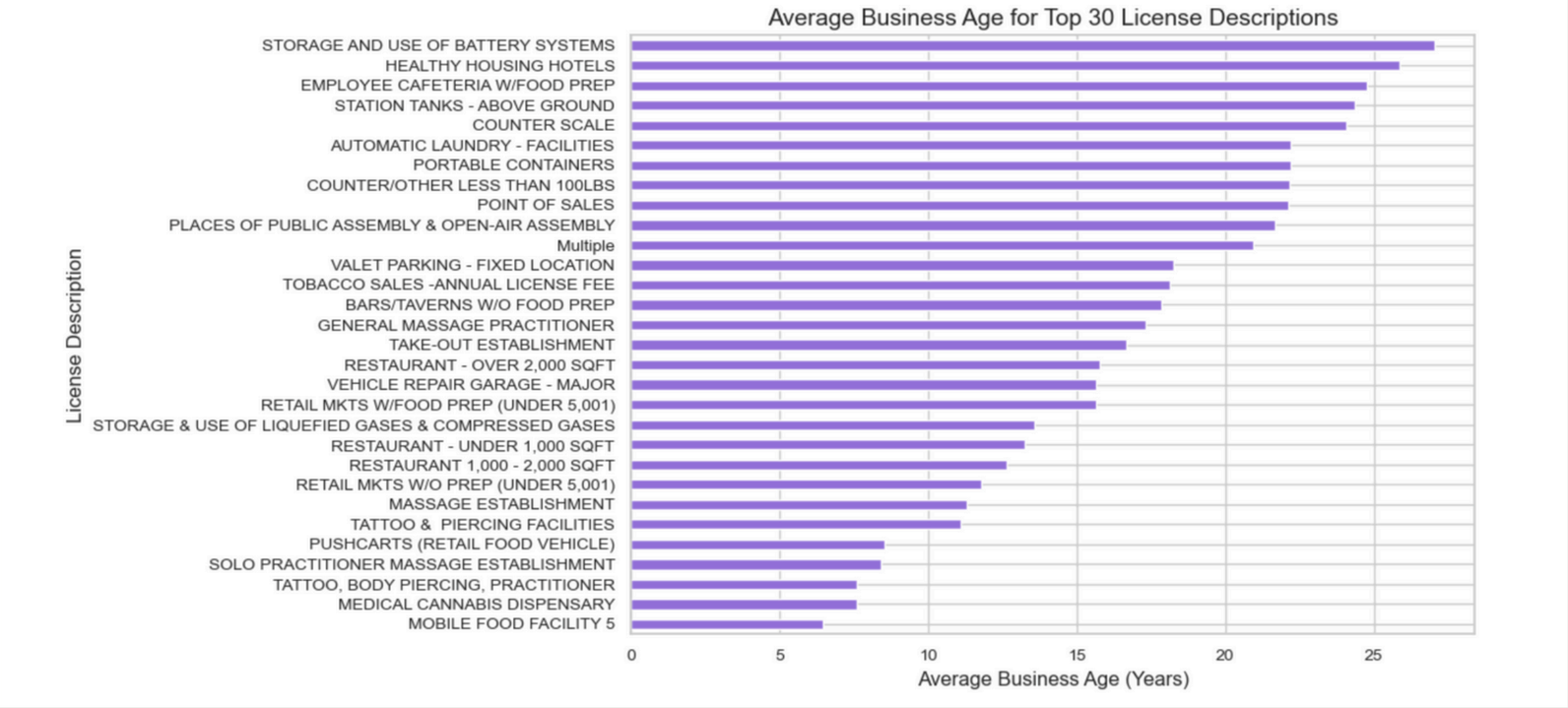
- Standardizing City Names
- Correcting Misspellings of "San Francisco"
- Renaming Columns
- Filtering Data for San Francisco District 3
- Converting Dates and Calculating Business Age
- Duration variable: business_age — represents how long a business has been operating.
- Event indicator: administratively_closed — indicates whether a business was closed (1) or is still operational (0).

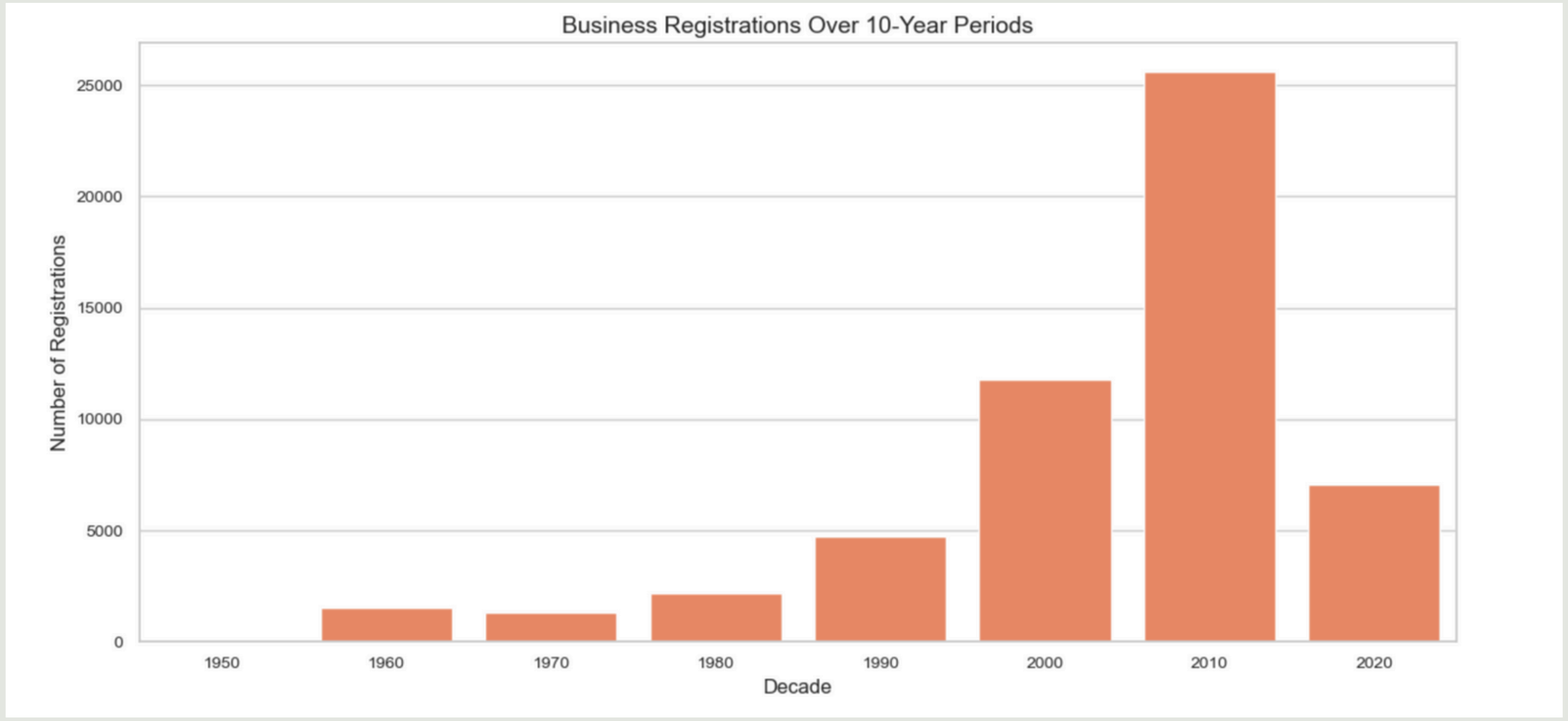


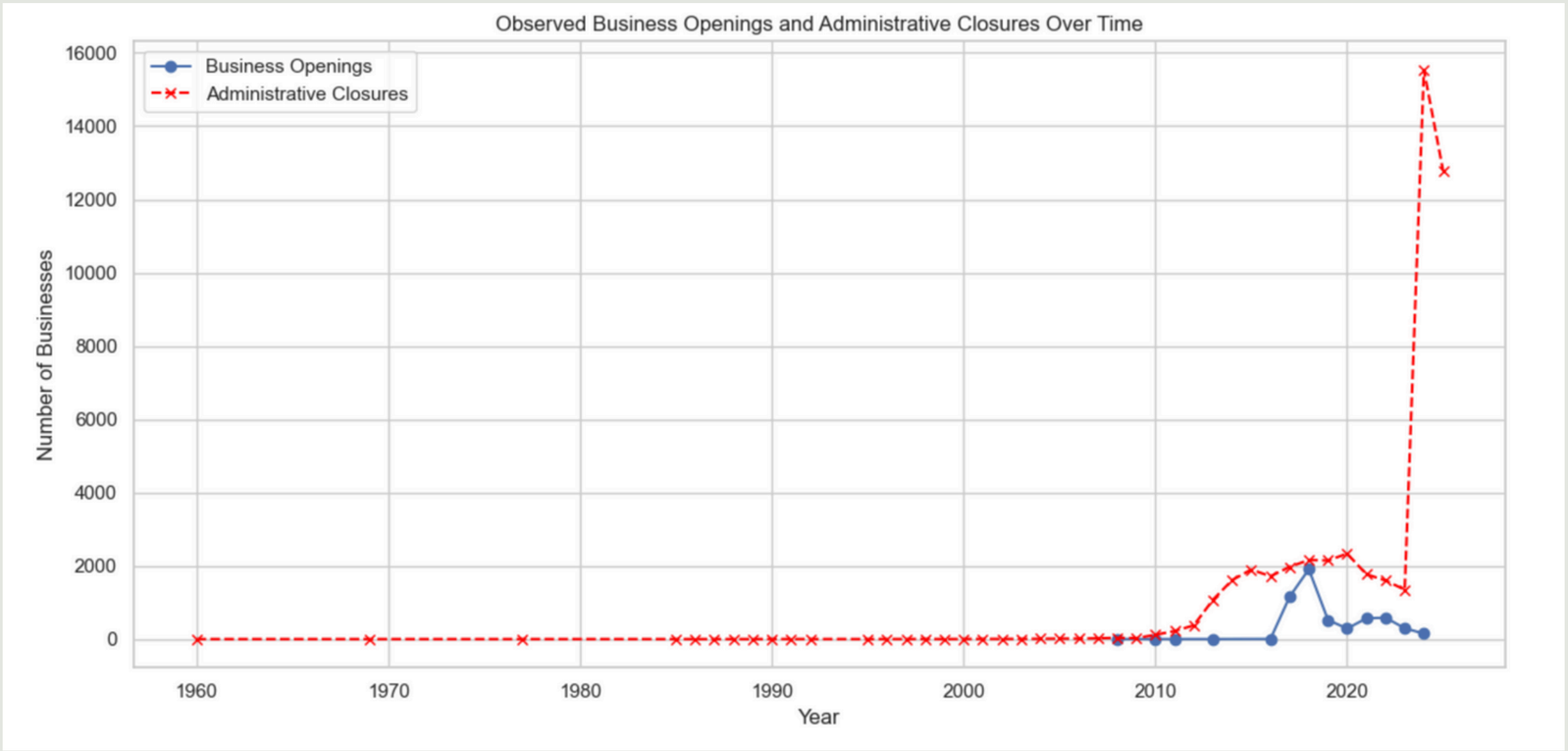
EXPLORATORY DATA ANALYSIS (EDA)


Top 30 License Descriptions by Number of Businesses
License Description
Multiple
RESTAURANT - UNDER 1,000 SQFT
RESTAURANT 1,000 - 2,000 SQFT
HEALTHY HOUSING HOTELS
POINT OF SALES
RESTAURANT - OVER 2,000 SQFT
GENERAL MASSAGE PRACTITIONER
PLACES OF PUBLIC ASSEMBLY & OPEN-AIR ASSEMBLY
RETAIL MKTS W/FOOD PREP (UNDER 5,001)
RETAIL MKTS W/O PREP (UNDER 5,001)
TATTOO, BODY PIERCING, PRACTITIONER
MASSAGE ESTABLISHMENT
PUSHCARTS (RETAIL FOOD VEHICLE)
STATION TANKS - ABOVE GROUND
VALET PARKING - FIXED LOCATION
BARS/TAVERNS W/O FOOD PREP
TAKE-OUT ESTABLISHMENT
AUTOMATIC LAUNDRY - FACILITIES
STORAGE & USE OF LIQUEFIED GASES & COMPRESSED GASES
TATTOO & PIERCING FACILITIES
SOLO PRACTITIONER MASSAGE ESTABLISHMENT
STORAGE AND USE OF BATTERY SYSTEMS
MEDICAL CANNABIS DISPENSARY
PORTABLE CONTAINERS
COUNTER/OTHER LESS THAN 100LBS
COUNTER SCALE
VEHICLE REPAIR GARAGE - MAJOR
TOBACCO SALES -ANNUAL LICENSE FEE
MOBILE FOOD FACILITY 5
EMPLOYEE CAFETERIA W/FOOD PREP
0
200
400
600
800
Number of Businesses


Business Registrations Per Year (Starting 1950)
Number of Registrations
Year
0
500
1000
1500
2000
2500
3000
3500
1950
1960
1970
1980
1990
2000
2010
2020
2030


Average Business Age for Top 30 License Descriptions
License Description
STORAGE AND USE OF BATTERY SYSTEMS
HEALTHY HOUSING HOTELS
EMPLOYEE CAFETERIA W/FOOD PREP
STATION TANKS - ABOVE GROUND
COUNTER SCALE
AUTOMATIC LAUNDRY - FACILITIES
PORTABLE CONTAINERS
COUNTER/OTHER LESS THAN 100LBS
POINT OF SALES
PLACES OF PUBLIC ASSEMBLY & OPEN-AIR ASSEMBLY
Multiple
VALET PARKING - FIXED LOCATION
TOBACCO SALES -ANNUAL LICENSE FEE
BARS/TAVERNS W/O FOOD PREP
GENERAL MASSAGE PRACTITIONER
TAKE-OUT ESTABLISHMENT
RESTAURANT - OVER 2,000 SQFT
VEHICLE REPAIR GARAGE - MAJOR
RETAIL MKTS W/FOOD PREP (UNDER 5,001)
STORAGE & USE OF LIQUEFIED GASES & COMPRESSED GASES
RESTAURANT - UNDER 1,000 SQFT
RESTAURANT 1,000 - 2,000 SQFT
RETAIL MKTS W/O PREP (UNDER 5,001)
MASSAGE ESTABLISHMENT
TATTOO & PIERCING FACILITIES
PUSHCARTS (RETAIL FOOD VEHICLE)
SOLO PRACTITIONER MASSAGE ESTABLISHMENT
TATTOO, BODY PIERCING, PRACTITIONER
MEDICAL CANNABIS DISPENSARY
MOBILE FOOD FACILITY 5
0
5
10
15
20
25
Average Business Age (Years)


Business Registrations Over 10-Year Periods
Number of Registrations
25000
20000
15000
10000
5000
0
1950
1960
1970
1980
1990
2000
2010
2020
Decade


Observed Business Openings and Administrative Closures Over Time
Business Openings
Administrative Closures
Number of Businesses
16000
14000
12000
10000
8000
6000
4000
2000
0
1960
1970
1980
1990
2000
2010
2020
Year

MODELING

- Model: Cox Proportional Hazards
- Features: Business age, zip, tax fields, NAIC codes
- Evaluation: Concordance index
- Findings:
- Newer businesses = higher risk
- Tax codes & zip codes matter

FEATURE IMPORTANCE AND MODEL EVALUATION

Principal Component Analysis:

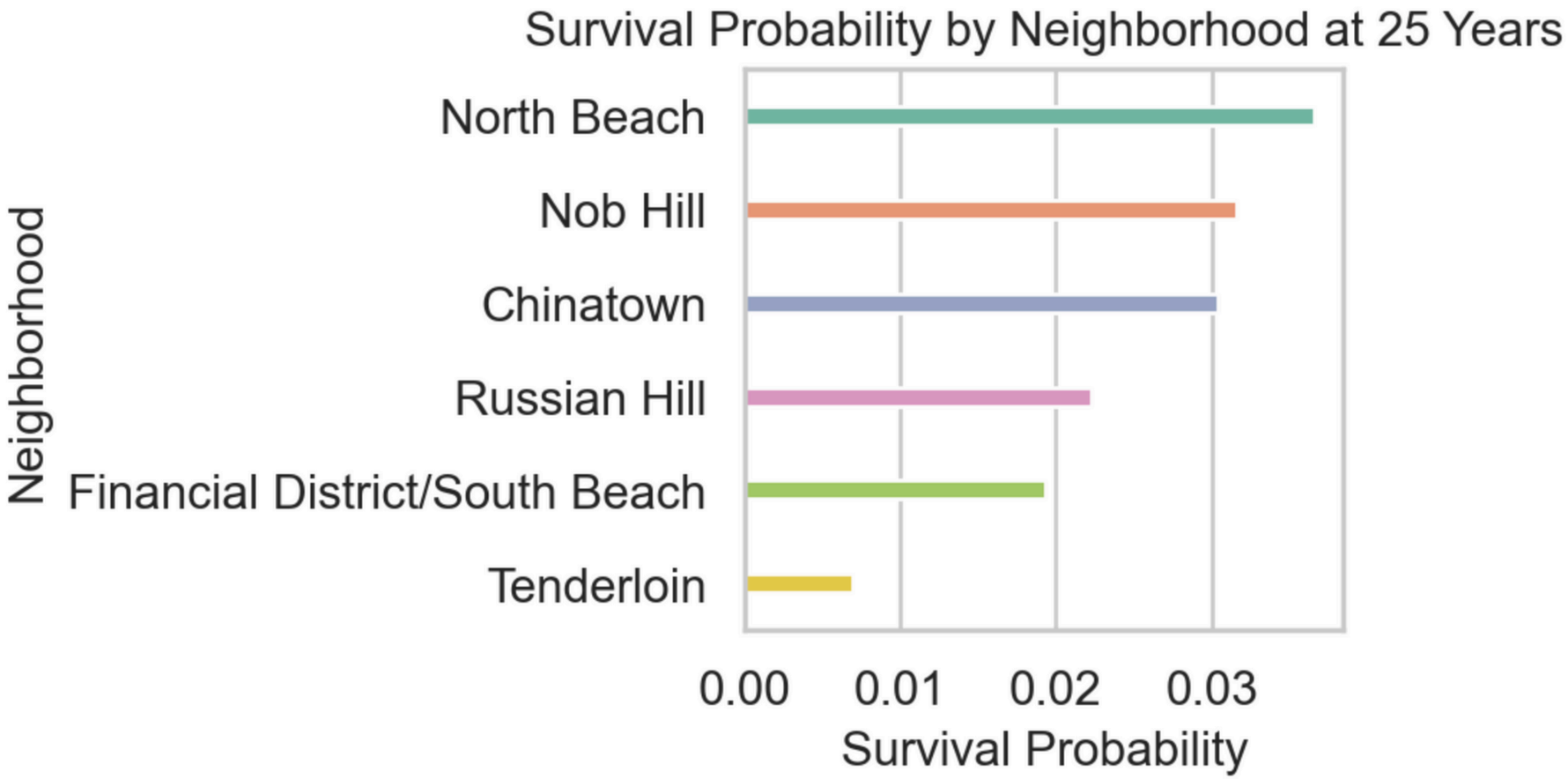
- Applied PCA Principal Component Analysis to find the feature importance to improve model performance

- Model performance: Concordance = 0.74

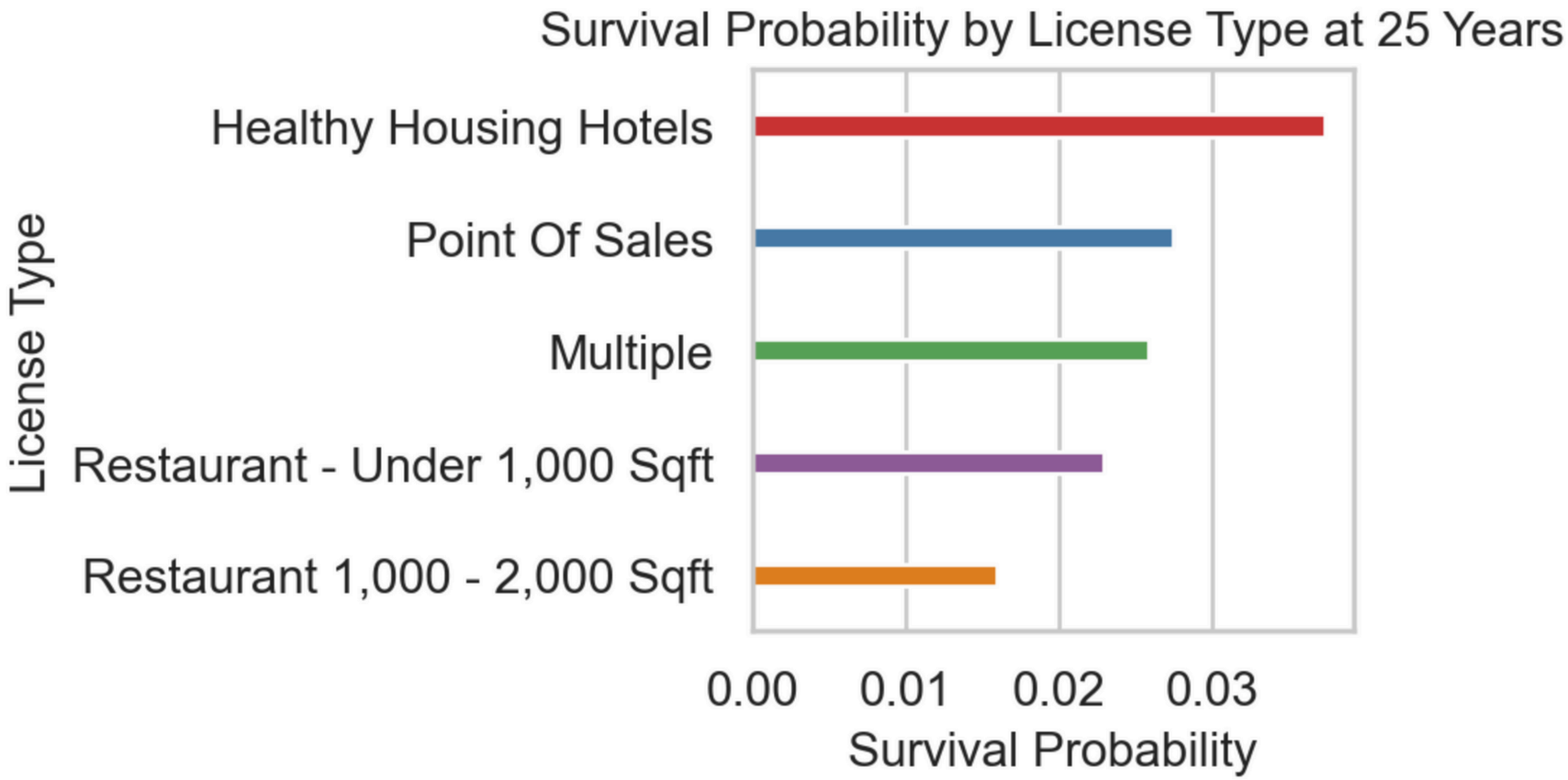
Top Contributing Features:

- Increase in Transient Occupancy Tax/Parking Tax cause higher business Closure risk
- Neighborhood & Zip has a moderate influence in business closure

🔎 **Survival Probability at 25 Years by Neighborhood**



Survival Probability at 25 Years by License Type



SURVIVAL TRENDS

- 5-Year Survival (Avg): 11.03%
- 10-Year Survival (Avg): 7.23%
- Best Neighborhood: Russian Hill (14.81%)
- Best License Type: Restaurant (12.75%)

RECOMMENDATIONS

Neighborhood-Based

- Identify underperforming neighborhoods using survival rates
- Provide targeted grants to support vulnerable areas
- Invest in infrastructure improvements to boost business activity
- Offer mentorship and advisory programs for new or struggling businesses

License-Type-Based

- Analyze top-performing license types to identify best practices
- Use successful categories as models for at-risk business types
- Design targeted support programs for struggling license types
- Consider policy changes and financial incentives to improve survival outcomes

Streamlit

Demo

📚 References

Number Analytics – Kaplan-Meier Survival in 5 Steps

🔗 https://www.numberanalytics.com/blog/kaplan-meier-survival-5-steps

Lifelines Documentation – Survival Regression

🔗 https://lifelines.readthedocs.io/en/latest/Survival%20Regression.html

SF Open Data Portal – DataSF

🔗 https://datasf.org/opendata/

Questions?

Thank you