Clustering Defaulted Loans: Small Business Administration

Dillan Gump

Agenda

- Context and Goal of SBA
- The Data
- Cluster Analysis
- Discussion and Next Steps

Context: SBA Loan Program

- Small Business Administration (SBA) Loan Guarantee
  - Helps business get approved by reducing potential loss by lender
- Expect some loans to default
  - Risky to begin with- Needed the SBA to get approved
- For a service designed to eat cost, there has to be a diet
  - The more we know the better
- Goal of Analysis: Gain Insight on Defaulted Loans

Why Defaulted Loans?

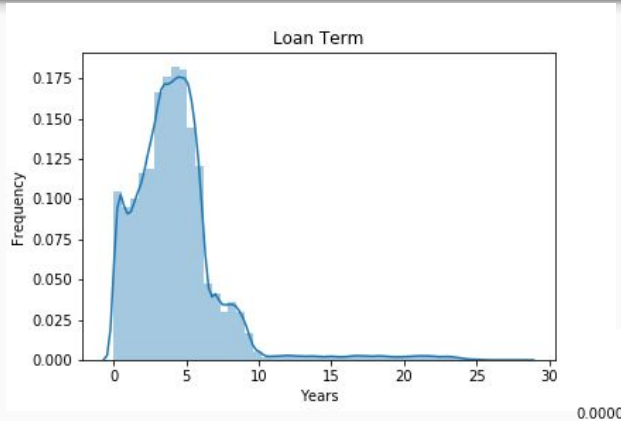
- Budgeting
  - Characterize Risk
  - Predict Losses
- Connect Businesses with Appropriate Resources
  - Businesses that default may need extra help
  - Different Businesses have different needs
    - A large franchise vs a mom-and-pop vs a start-up
- Question: What can we learn about defaulted loans to improve SBA and business outcomes?

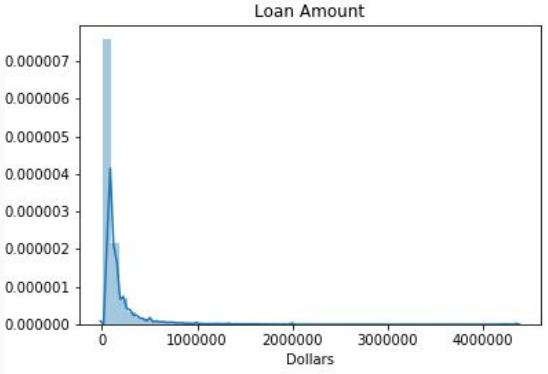
The Data

- Only Defaulted Loans
- Years: 2000-2014
- Important Features:
  - Loan Amount and Term
  - Bank Size and State
  - Number of Employees and Jobs Created
  - New or Existing Business

Features

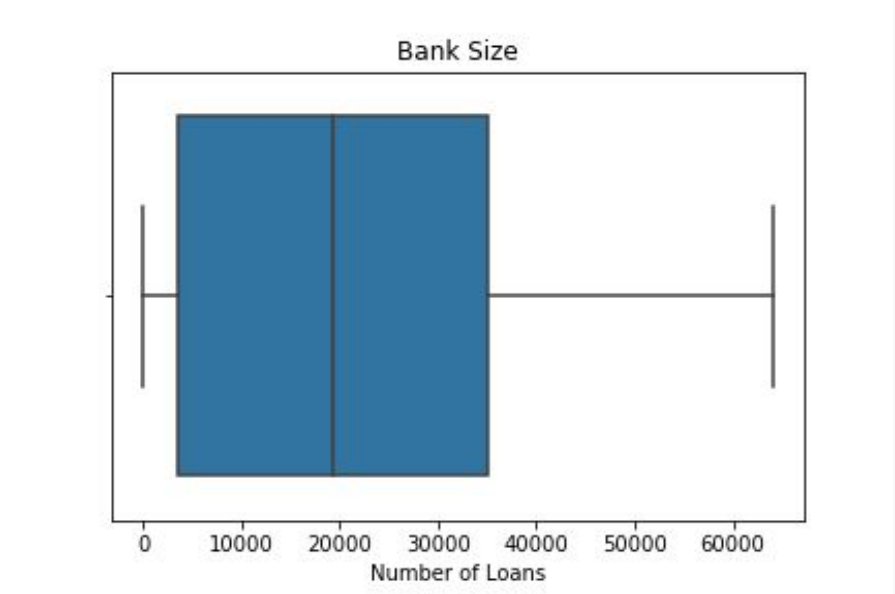
- Loan Amount
  - $4,000 - $4,000,000,
  - Avg: $100,000
- Loan Term
  - 0 - 28 years
  - Avg: 4 years
- Existing or New Business
  - 71% Existing, 29%New
- Bank In or Out of State
  - 74% In-State, 26% Out-of-State





Features (Cont'd)

- Number of Jobs Created
  - 0 - 1,618 jobs
  - Avg: 2 jobs
  - Extremely Skewed
- Bank Size (Number of Loans)
  - 1 - 64,000 loans
  - Avg: 22,000

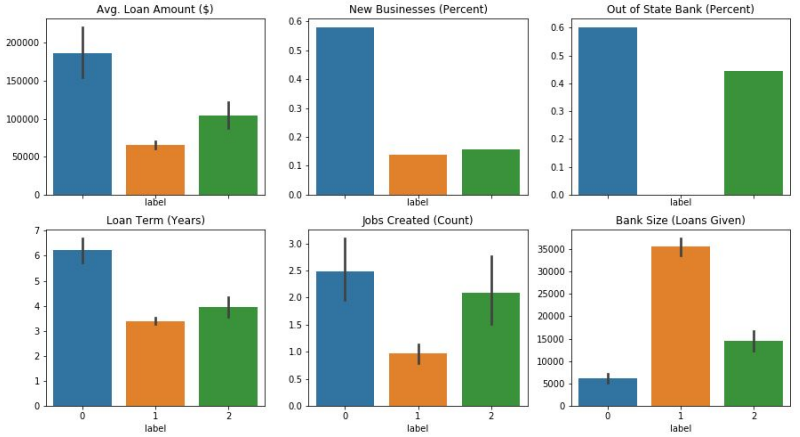


Clustering: Methodology

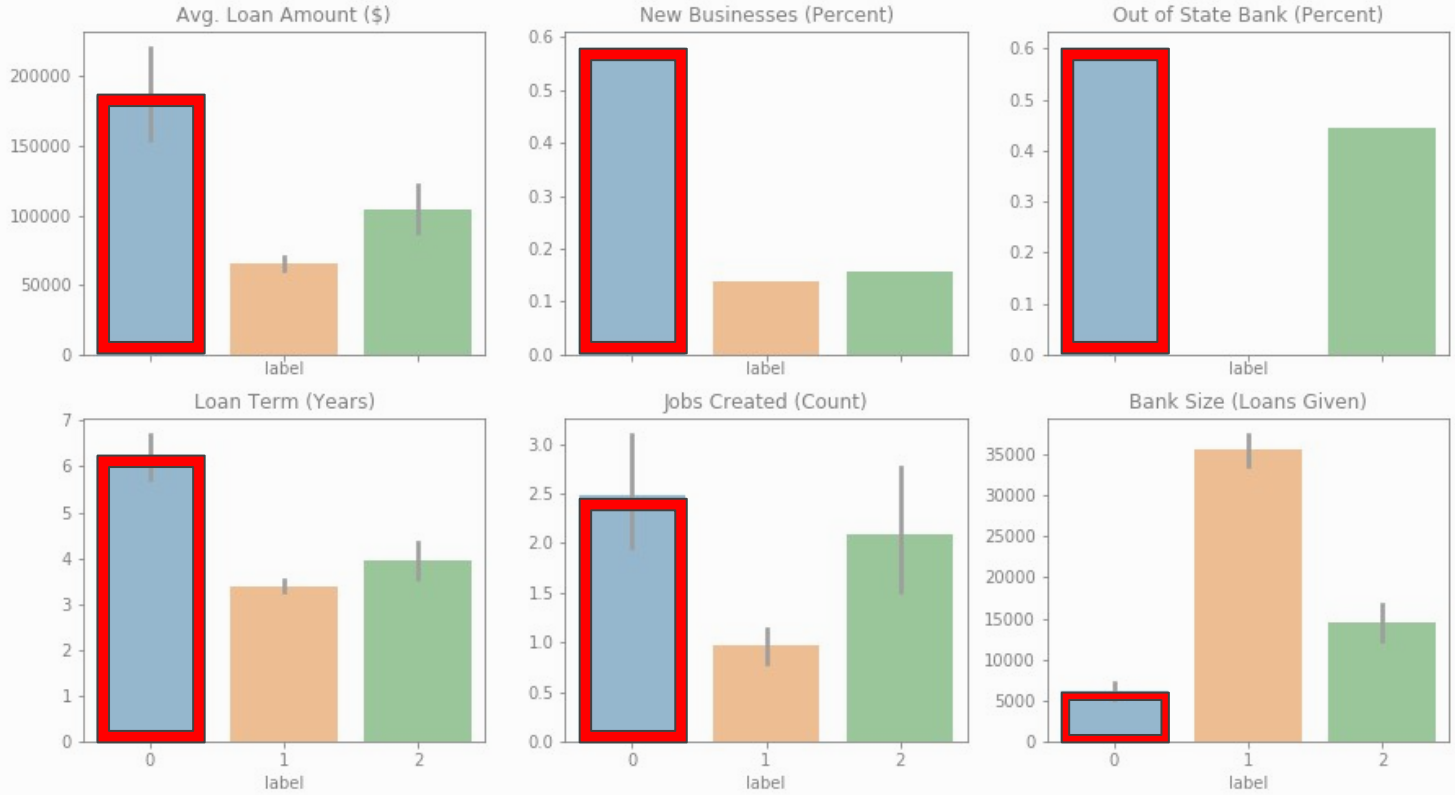
- Concerning Distributions
  - Lots of skewed features
  - Should not affect “low” vs “high” comparisons
  - Need some business context
    - What is a loan term of zero years? Common value.
- Mixed Data
  - Gower Distance
  - Factor Analysis of Mixed Data
- Sampled 1,000 loans from original 126,000 defaulted

Clustering: Methodology (Cont'd)

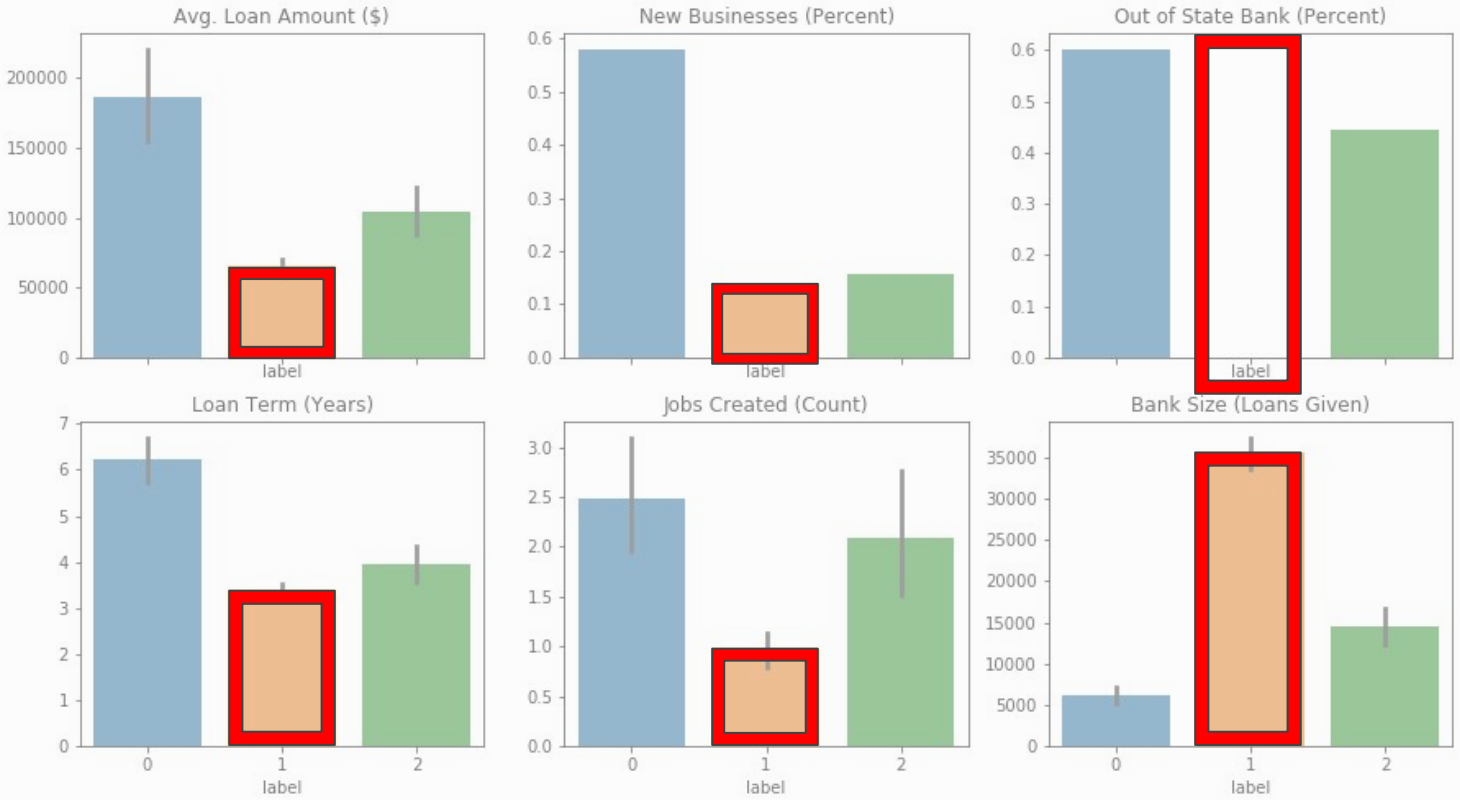
- K-medoids
  - Using a median based method will help handle skew
  - Categorical similarity
- 3 Clusters
  - Maximum number of clusters with discernable differences



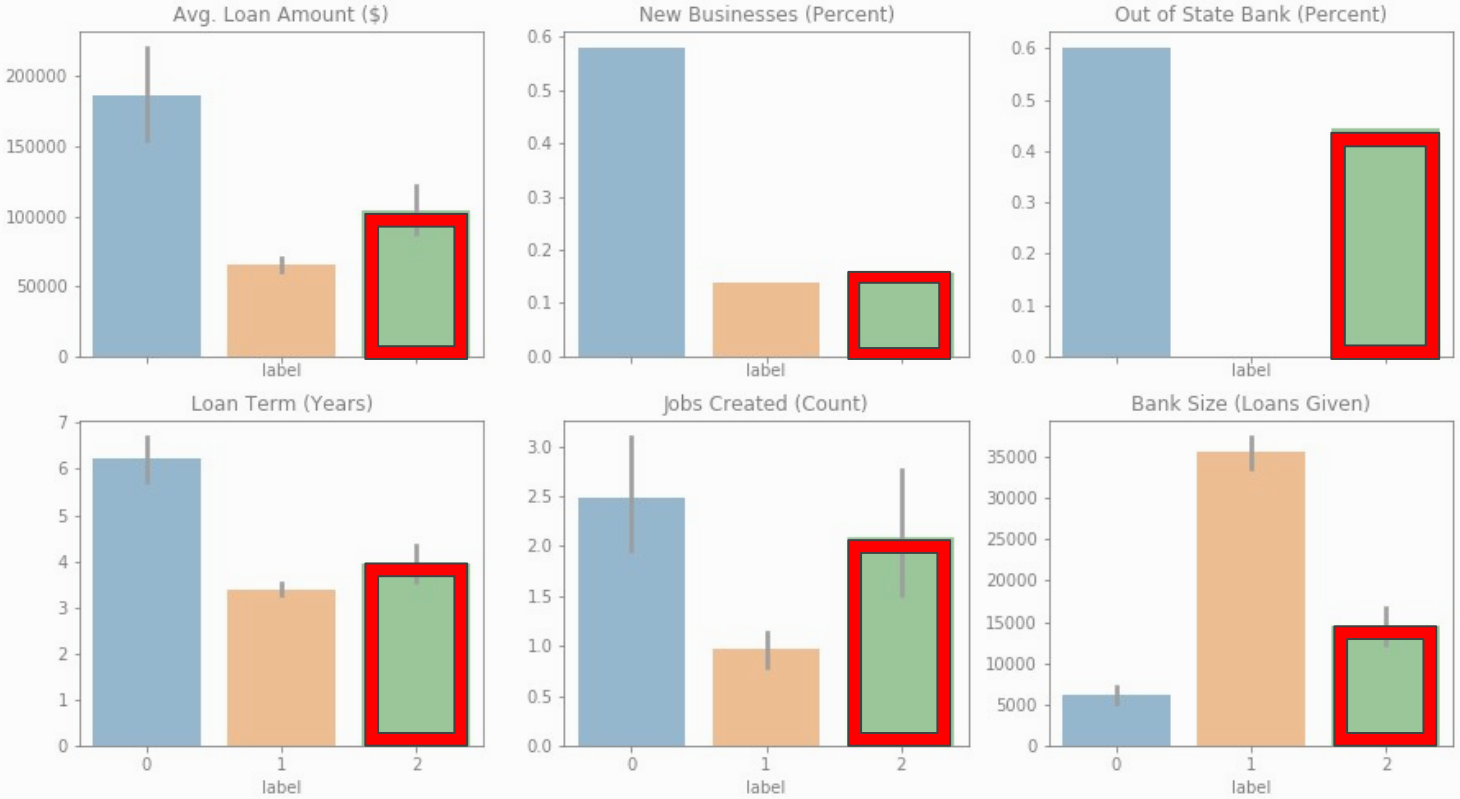
Cluster 1: Failure to Launch



Cluster 2: Business Burnouts



Cluster 3: Average Janes



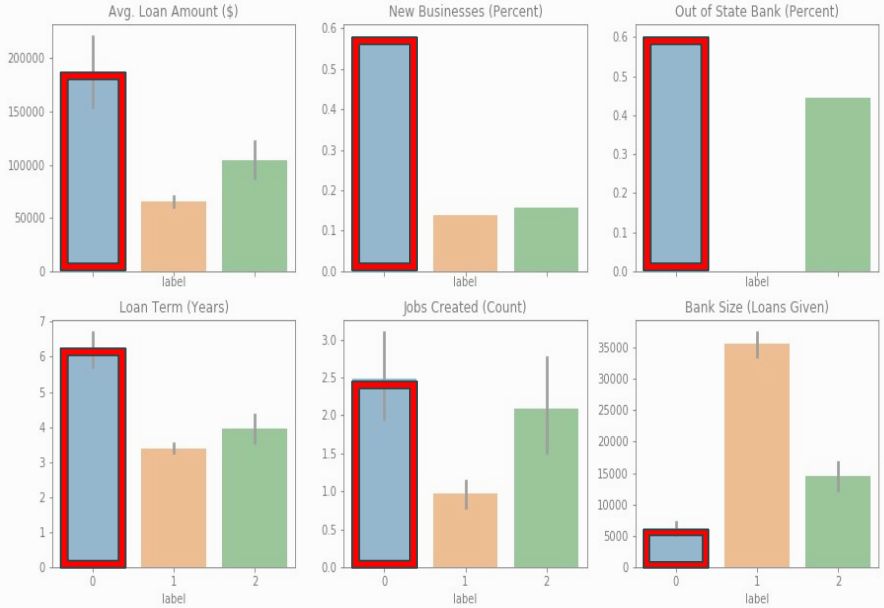
Cluster 1: Failure to Launch

Features:

- Highest Percentage of New Businesses
- Biggest Loans, Longest Average Loan Term
- Smallest Banks, but most out-of-state
- Created the most jobs

Promising new businesses that faltered

- Consider for a second chance
- May need guidance and adjustments
- Connect with more conservative loans



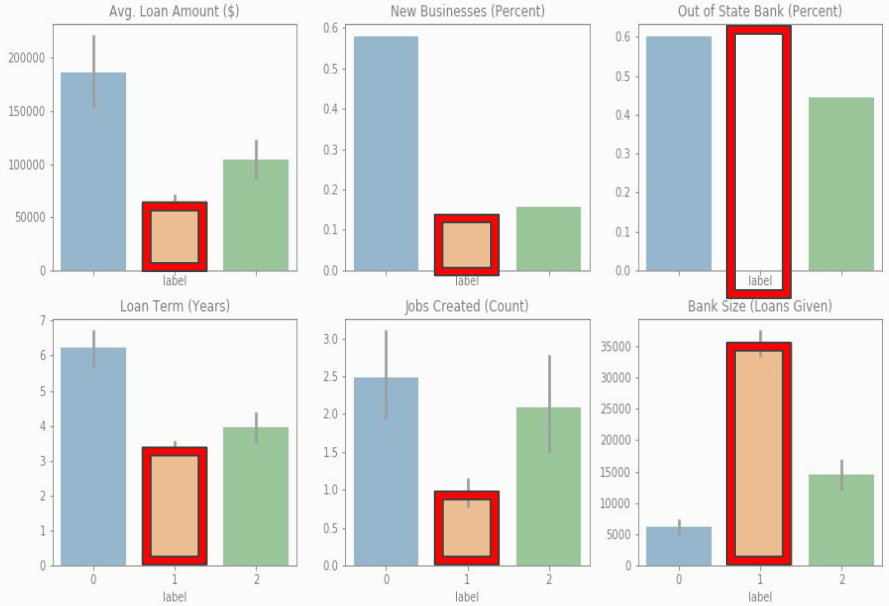
Cluster 2: Business Burnouts

Features:

- High percent of In-State Banks, Biggest Banks
- Highest percent of Existing Businesses
- Shortest Average Loan Term
- Created the Least Jobs
- Biggest Cluster (~500 loans)

Exhausted local credit options and slowing growth

- Connect with new resources, out-of-state banks
- Help modernize



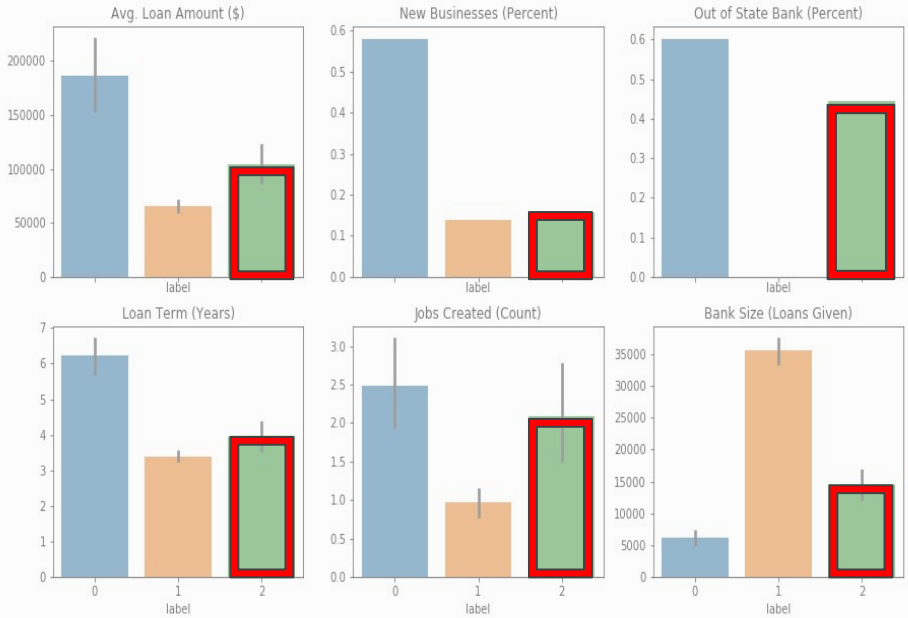
Cluster 3: Average Janes

Features:

- Similar businesses to Group 2
- In-between loan sizes and terms
- Medium sized, in-state banks

Businesses on their way to burning out

- Connect with new resources, bigger banks
- May have features of the first 2 groups
- Help overcome the obstacles that result in Cluster 2



Limitations

- Interpretation of Bank Size and State
  - Some businesses may go out of state for lower rates
  - Interpreted here as struggling to find credit
- Skewed Data
  - Data transformations can help
  - SBA definition of atypical loans for fringe cases
- Would like to know if these businesses have sought SBA loans before
  - Also business age, not just new or existing

Taking Action

- New businesses and established businesses have different needs
  - New Businesses:
    - Marketing, Operations, and Vision
    - Help get a foothold
  - Established Businesses:
    - Assist with upscaling
    - Connect with contractors for new responsibilities
  - Advise businesses on credit opportunities based on what has already been utilized

Going Further

- Internal SBA classifications can inform cluster interpretation
  - Did the clusters challenge trends, or describe known patterns?
- More data engineering
  - Refine Bank Size feature
  - Effect of Industry, State Economy, and Year
- A/B testing to see if suggested further assistance is effective