

Datathon Information Session

February 22, 2020

Monthly Zip9 Homebuyer Data

Demographic Data

- For zip9 from credit bureau data
- Averages of age, aggregates of person/home count

Credit Bureau Data

- US consumers w/ credit activity
- Rolled up to zip9 level
- Zip5 provided for visualization
- Basic Credit Data
- Five most recent MTG/HEQ
- Presence of certain accounts (auto/student loan, etc.)
- Covers April - September 2019



Transaction Information



Will people in a zip9 buy homes?

Personal Information



House Purchase Data

- For zip9 from credit bureau data
- Number of people in each zip9 who bought a house
- Number of people in each zip9 who bought their first house
- Covers all US house purchases from October - December 2019

Where are prospective homebuyers?

Data Summary

- Data provided is grouped by zip9 and gives monthly shots from April - September 2019
- Some home purchaser information is provided

Primary Objectives

- Predict:
 - How many people in each zip9 will buy a house from October - December 2019
 - How many of the above people in each zip9 will be buying their first home

Other Objectives

- Find/present data trends
- Which fields were good/bad predictors? Why?
- Discuss how you handled the exploration of moderately large data
- Prioritize readability, process and results should be coherent

Recommendations

- While zip9 values are coded, the zip5 correspond to real regions - use it to visualize
- Have at least a baseline understanding of the data before you start

Questions

- We will be around for a few hours to answer questions!



More Specifics on the Dataset

File Name	# Rows	# Fields	Summary of Fields
zip9_coded_20190X_pv	6009259	53	Snapshot of credit bureau data for month X
zip9_demographics_coded_pv	6009259	7	Demographic, homebuyer data

Examples of variables in zip9_coded_20190X_pv

Variable	Type	Description
zip9_code	Num	Numerical key matching to a zip9 region
bankcard_trades	Num	Average number of open bank cards per person in the zip9 region
mortgage1_limit	Num	Average balance for most recent mortgage per mortgage owner in the zip9 region
total_homeequity_limit	Num	Average sum of home equity limits per person in the zip9 region
studentloan_open	Num	Proportion of people who have at least one student loan in the zip9 region

- Potential data noise and biases
 - Data covers adults with credit history, delay in culling out death records means age may skew high
 - Averages are only on non-null values, if someone is missing a quality, they are not counted in the zip9 average
- We have withheld about 10% of zip9s for testing purposes - avoid overfitting!
- Expect many fields to be unpopulated or have unexpected values; raw data is often noisy and imperfect in unpredictable ways
 - But feel free to ask us about fields and their values