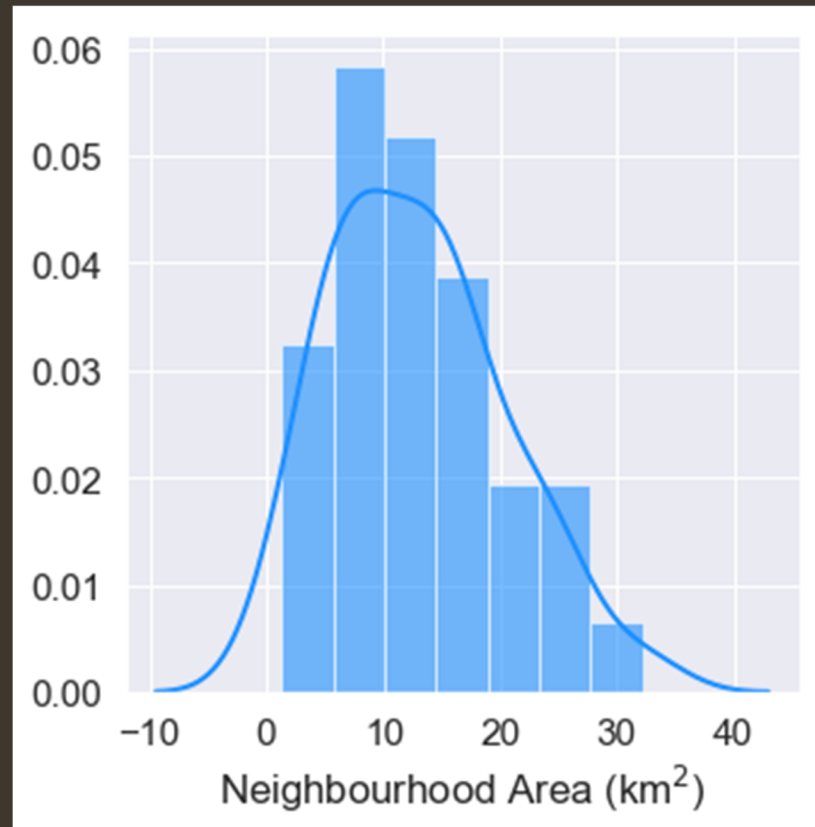# Calgary's Battle of Neighbourhoods

CAPSTONE PROJECT

Calgary's Neighbourhoods

- Calgary attracts thousands of migrants every year

- Information about each neighbourhood of the city will be very helpful for newcomers

- Project goals:
  - Provide insight information about neighbourhoods
  - Cluster neighbourhoods
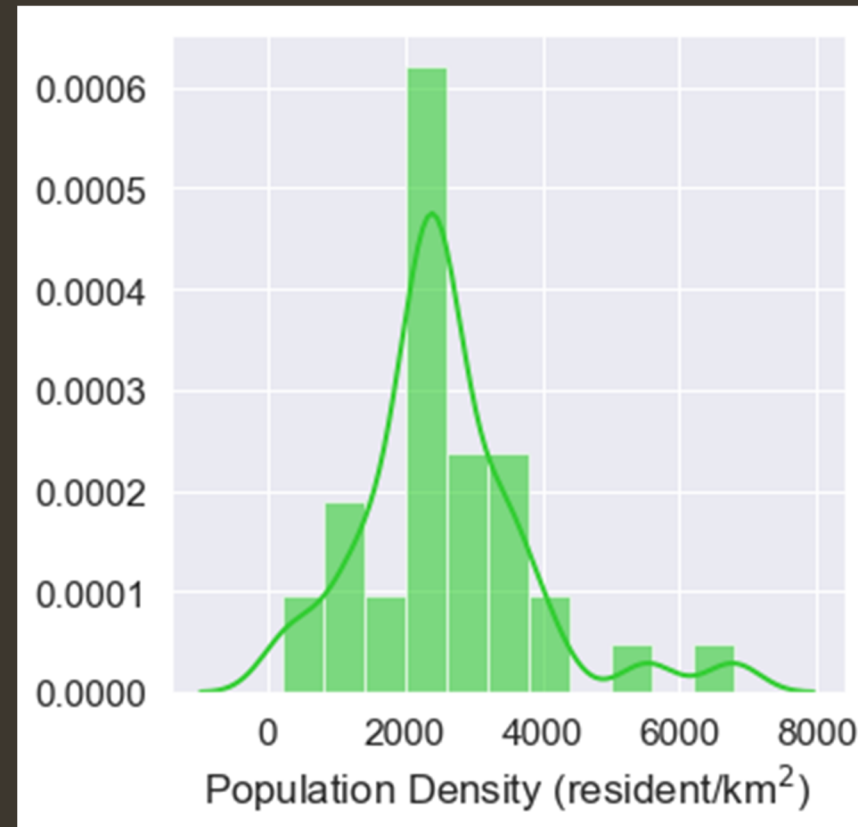 - Identify variables most influential on house price

Data

- Postal codes and coordinates: GeoNames and Mapawi

- Venue: Foursquare

- Community boundaries: Statistics Canada

- Community population and area: Wikipedia

- House price data: The Globe and Mail

- Crime Rates: Calgary Police Statistical Reports

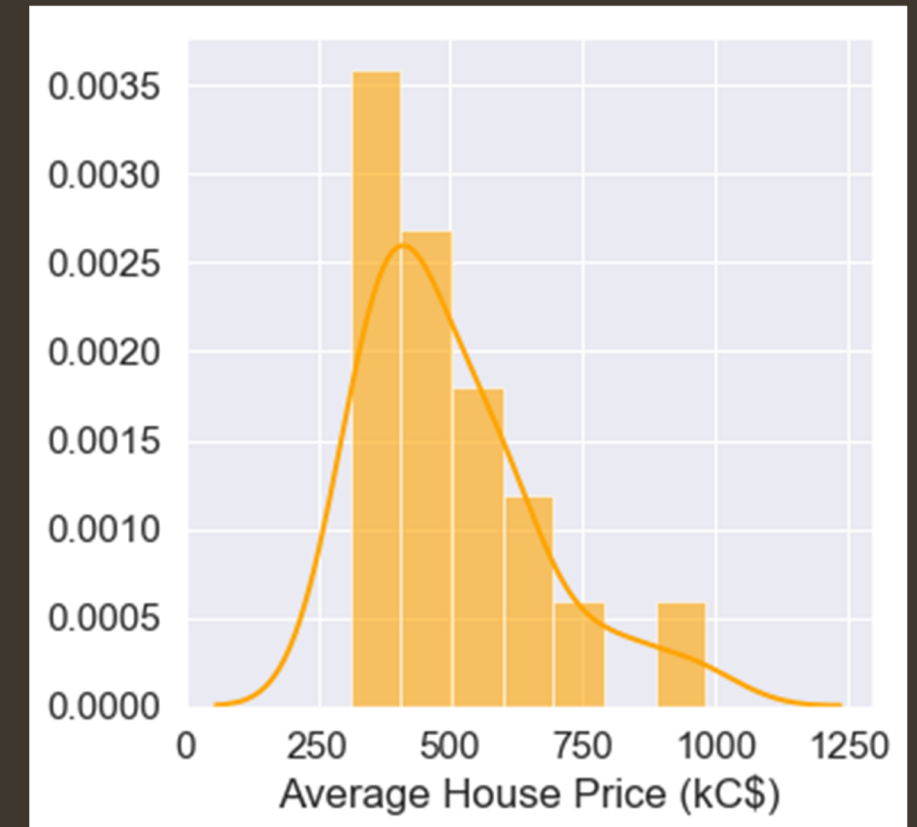- School ratings: Fraser Institute School Rankings

Exploratory Data Analysis
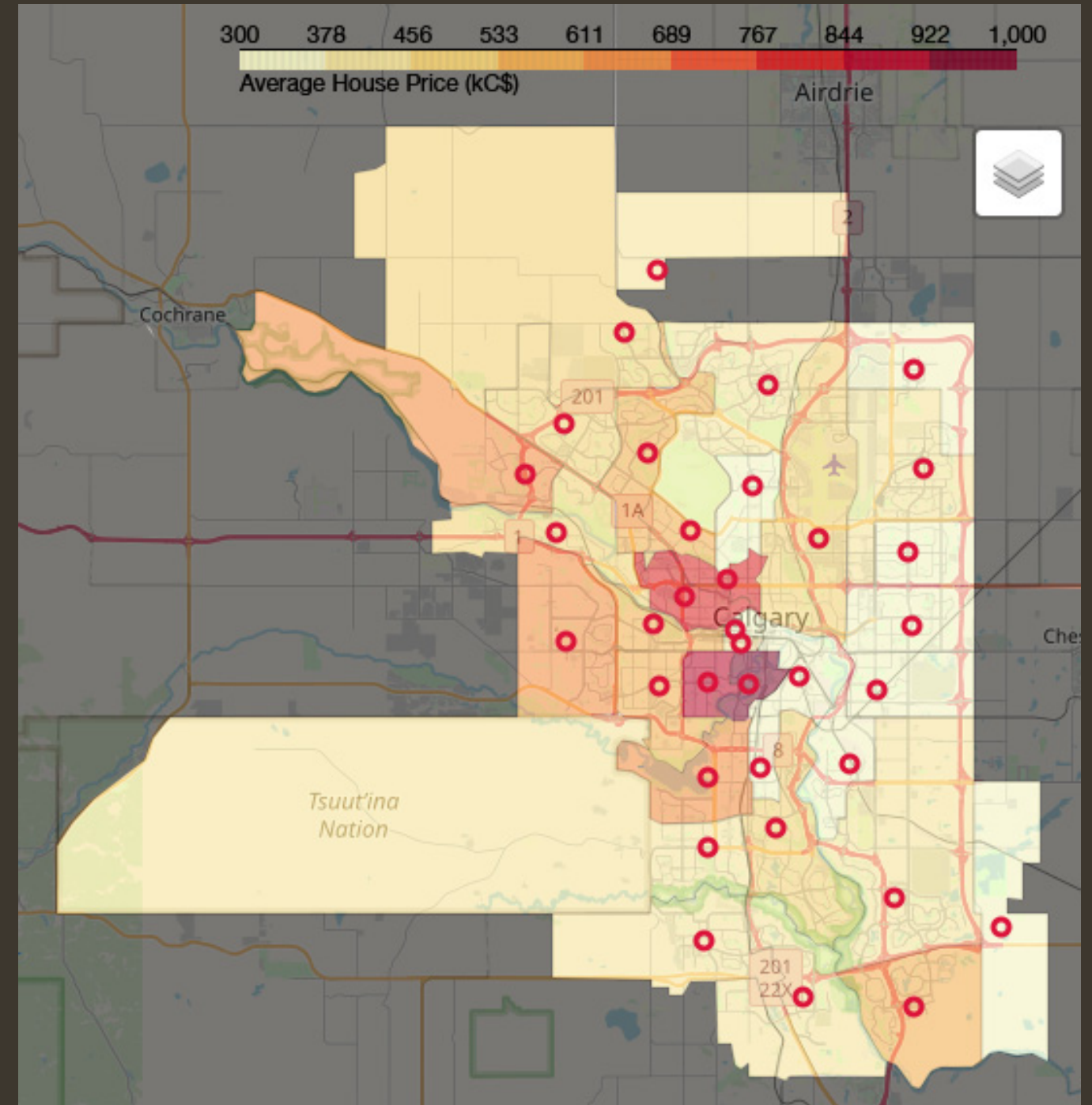

Wide distribution of neighbourhood area

Average density: 2000 residents/km²
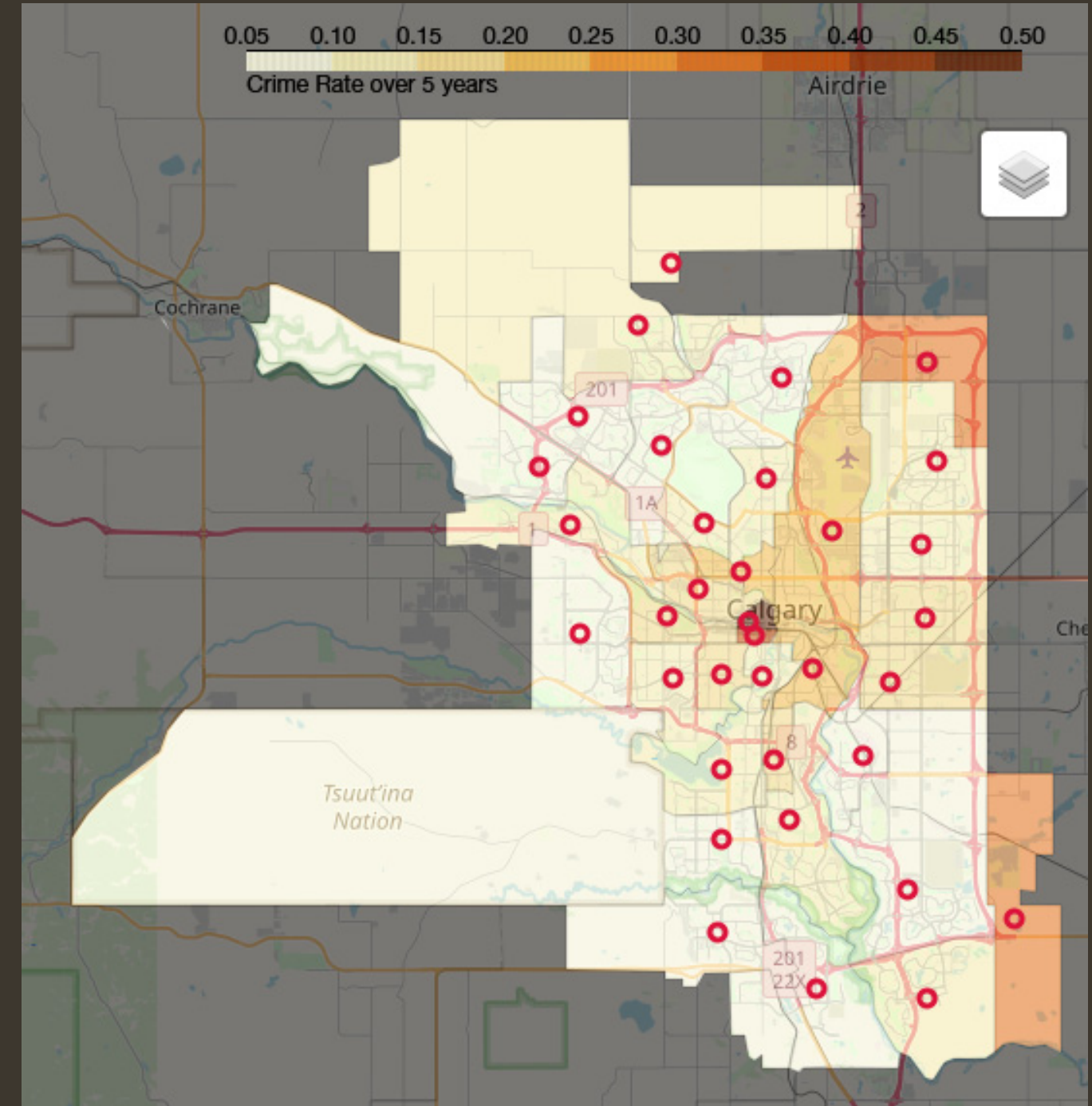
Most common house price: 300-400 kC$

House Price

- Surrounding downtown are the most expensive

- Northwest and Southwest are sought-after areas

- Downtown is moderately priced
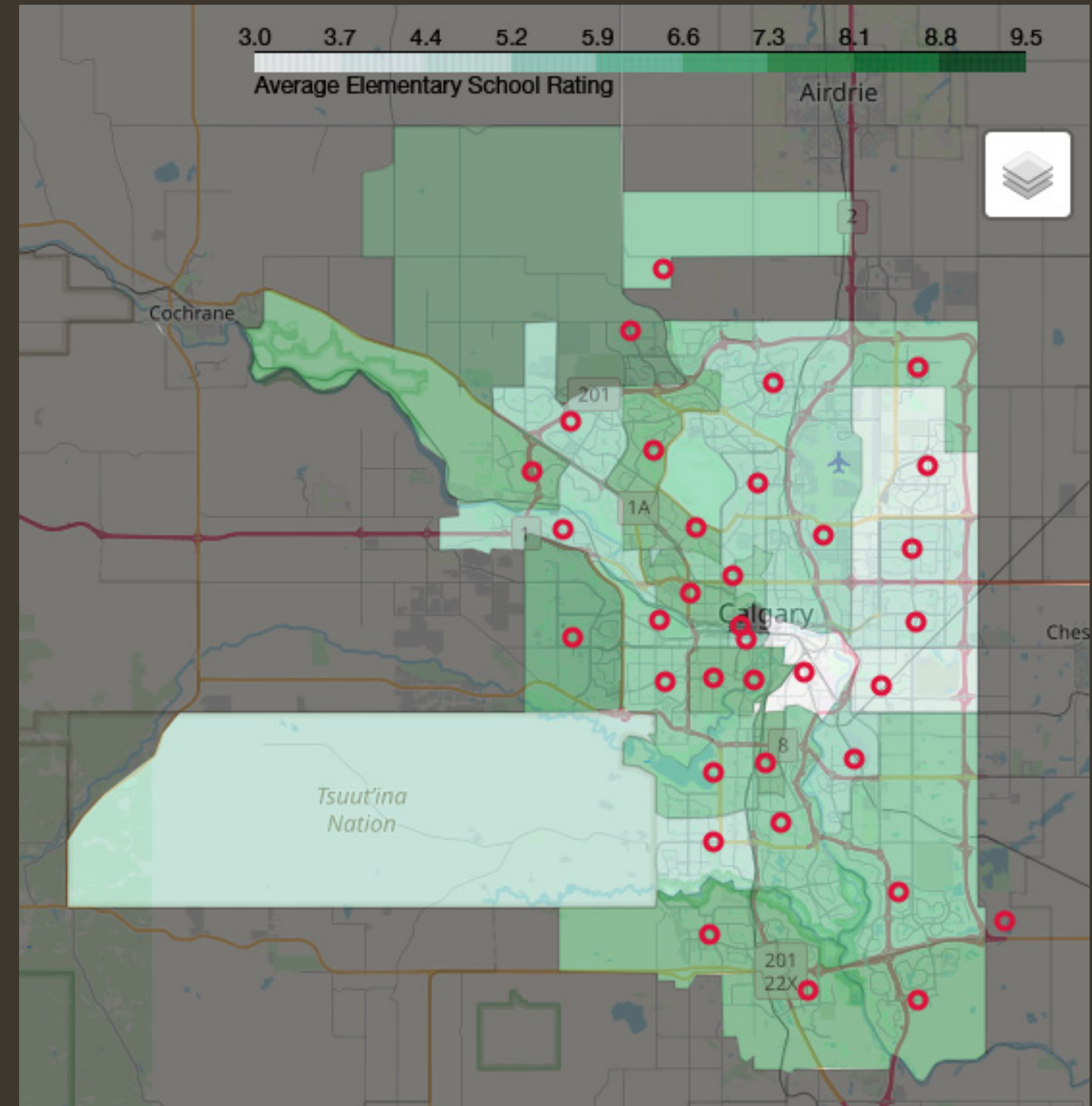
- Northeast are lower than most

Crime Rate

- Northeast has the highest number of crime incidents

- Downtown has the highest crime rate
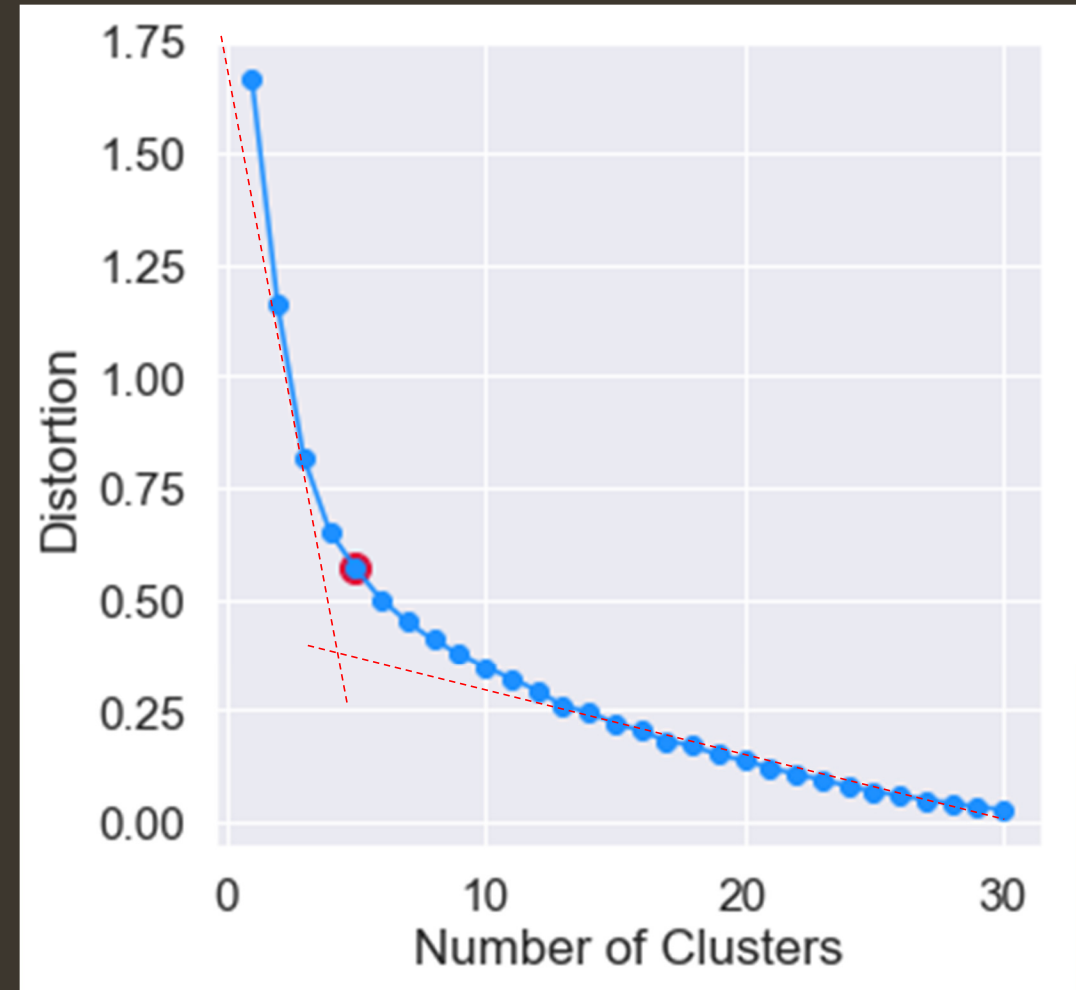
- Average crime rate: 0.024 incident/resident/year

School Ratings

- Private school are expensive

- Public school ratings vary wildly

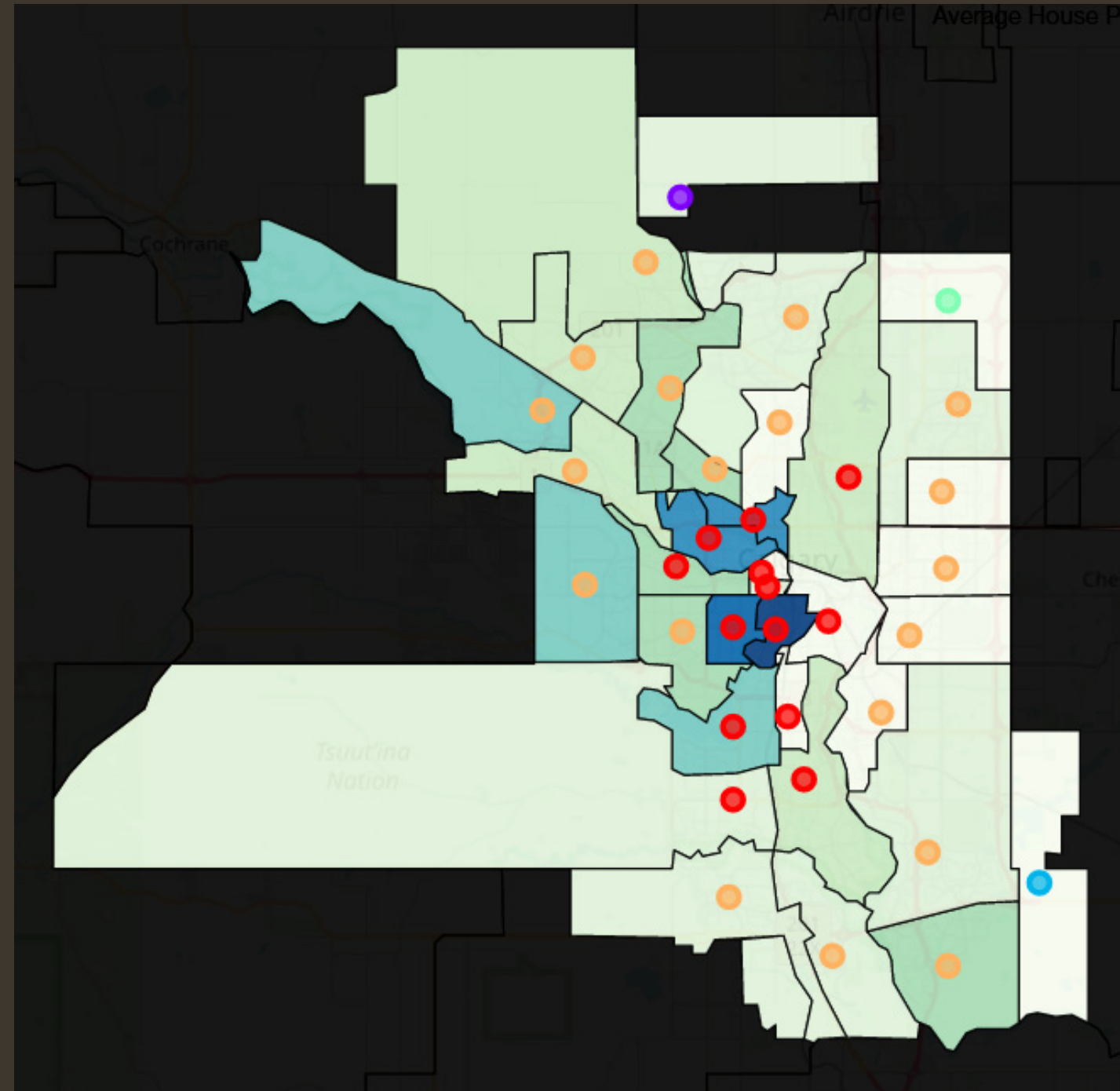- Northeast areas ratings are lower than others

Foursquare Venues

- 199 unique categories of venues

- Foursquare limit caps 10 neighbourhoods

- K-Mean clustering optimal number of clusters: 5
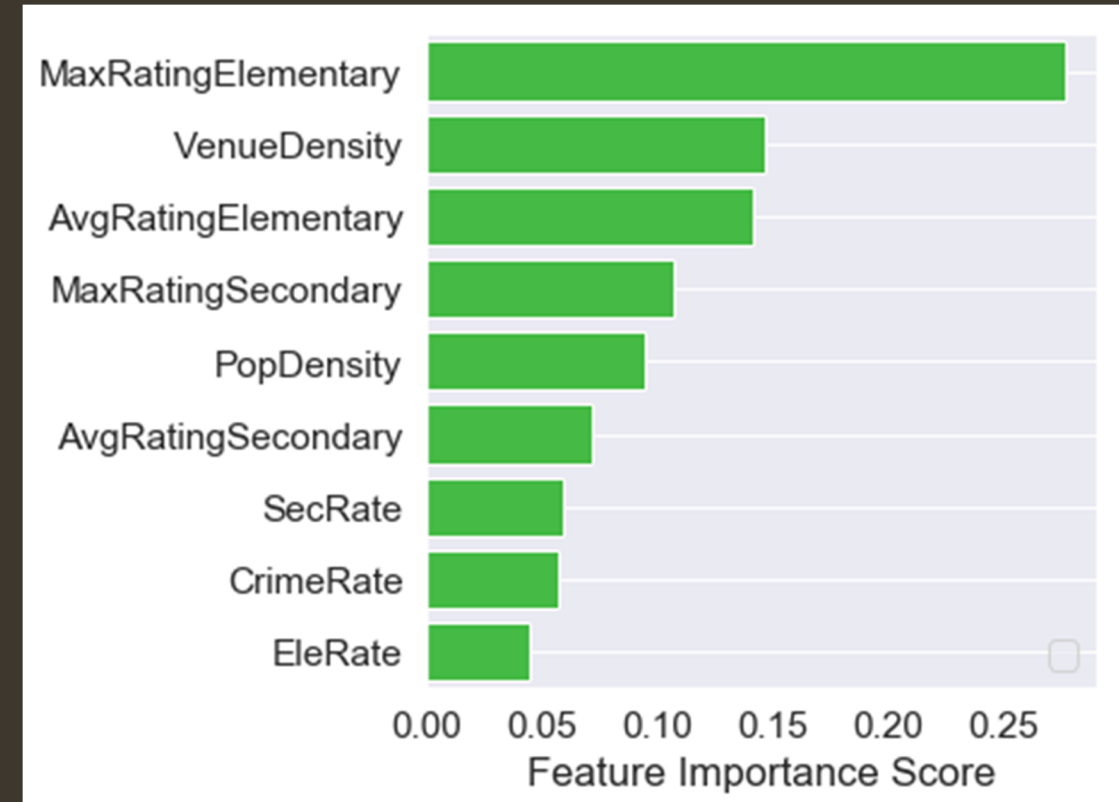
Neighbourhood Clustering

- Red cluster
  - Around downtown
  - Dominated by café, coffee shop, pub, bar, restaurants etc.
- Orange cluster
  - Further from downtown
  - Dominated by grocery store, pharmacy, shopping mall etc.
- Other clusters
  - Remote areas

House Price Reasoning with Random Forest

- Model performance

- Venue density & school ratings seem to drive house price

Discussions

- Coordinates dictate how venues are queried in Foursquare.
  - GeoNames and Mapawi data are unreliable occasionally
  - An algorithm to calculate center coordinates from boundary file will yield better data.

- Better clustering can be achieved if:
  - Foursquare limit of 100 is raised
  - Venue ratings is accessible.

- Demographic data available on Statistics Canada can be included.