Predicting Calgary Housing Prices

ENSF 612 Big Data Presentation
Corey Yang-Smith, Mehreen Akmal, Eric Diep, Jenn Bushey, Hao Liu

Strategy

- Identified features and what their impact is on our target variable
- Loaded individual datasets using API into pandas dataframe
- Inspected quality and extracted desired subset
- Explore EDA on individual Datasets

Categorized Features and Mapped to Dataset

Economic Indicators
- Large-scale economic indicators that may affect the housing prices in an economy. The employment rate, a city's wealth, its population and growth, etc
  - Average Home Price
  - Unemployment Rate
  - Housing Starts
  - Calgary Population
  - Interprovincial Immigration
  - Inflation Rate
  - Assessed Year

Group Price Indicators
 - Factors that may influence a house price broadly based on other, location based factors. Ie neighborhood, density, crime rate, distance to downtown, desirability, etc
  - Density by Community
  - Community Name/Code
  - Crime Count
  - Resident Count
  - New Development by Community

Individual Price Indicators
 - Individual factors that may influence a house price, ie land size, type of dwelling
  - Property Type
  - Land Use Designation
  - Dwelling Type
  - Land Size

Challenges

- Discrepancies between datasets, finding common data
- Limited overlapping years
- Combining Datasets (Working with 8 Datasets)

Next Steps

- Combine Datasets
- Perform EDA
- Build Model