Coursera Capstone Project

IBM Applied Data Science

Andriy Zakharko, 2020

The Battle of Neighborhoods: Best place to move in Austin, TX

Contents

Business problem definition

I am a Software Engineer, currently live in Austin Northwest, TX. It is time to move and find a better place to live in the same city, so I decided to attack this problem as a data science challenge. I have been renting a long time and now it is time to buy a house with all my constraint and limitations my family have, such as – price range, school rating, parks proximity and last but not the least – shops, groceries, and restaurants nearby so we could have a good time with a whole big family.

Overall Constraints and Limitations:

- We need Elementary and Middle School with average rating at least 7.0
- At least 2 shops in 5-min-by-car proximity

Crime Rates:

- Average city and suburban's crime rate are low enough to satisfy our needs, so we want focus on this aspect of the problem at all

Price and Payments:

- Based on overall family income we will filter out target dataset so that it will contain only acceptable records and for personal reasons we'd like to keep this information private

School Rates:

- We need Primary School with average rate not less than 7.0
- We need Primary School with average rate not less than 7.0

Miscellaneous:

- At least one park nearby

With all this in place, this is not a search request, even if it looks like this. Those are criteria set we are going to use to cluster Austin Neighborhoods and use this information to decide later on what's the best place to move (or at least what's the options do we have)

Data Needed for the Project

To solve the problem, we will need the following data and figures:

- List of neighborhoods in Austin, TX
- Geo-decoded data (Addresses or Coordinates)
- Venue data (Foursquare API)
- House prices (Zillow website as a data source)

Extracting the Data

- The scrapping of Austin neighborhoods dataset from Wikipedia
- Using Foursquare API to get venue data related to the neighborhoods
- Enrich dataset with houses and schools from Zillow data source

Methodology

- Initial data will be prepared
    - neighborhoods dataset
    - venue data from Foursquare API
    - Zillow data
- Data should be cleaned, filtered and normalized
- Clustering model will be applied (KNN Clustering)
- Results will be reviewed and
    - If Satisfied – final report will be presented
    - If Not Satisfied – data and process will be reviewed and repeated