**HOUSING SALES PRICE PREDICTION OF AMES, IOWA**

VISHNU SIVADAS

16-05-2019

# **INTRODUCTION**

### **Background**

**Ames** is a city in Story County, Iowa, United States, located approximately 30 miles (48 km) north of Des Moines in central Iowa. It is best known as the home of Iowa State University (ISU), with leading agriculture, design, engineering, and veterinary medicine colleges. Housing prices of this area depends on a lot of factors. For the people who are looking for buying a house or somebody who wants to sell a house, making a wild guess is difficult and often results in bad business decisions. In this project a model is created to tackle the same.

### **Business Problem**

When we ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. There are a lot of features to be considered before one can set the price or start negotiating. The project aims in creating a model for predicting housing sales price for Ames, Iowa considering all the important features including the neighbourhood venues

### **Target Audience**

* House aspirants who can roughly estimate the value of a house based on its features and the average price.
* Real estate people and city planners who can decide what kind of venues to put around their products to maximize selling price.
* House sellers who can optimize their advertisements.

# **DATA DESCRIPTION**

## **Data Sources**

Data sets are prepared from the following sources:

* The Ames Housing dataset is taken from Kaggle.com which was compiled by Dean De Cock for use in data science education. It consists of 79 explanatory variables describing various aspect of residential homes in Ames, Iowa.
* **Foursquare API** is used to get the most common venues of Ames, Iowa. There is a categorical variable ‘Neighborhood’ in Ames housing dataset. Using this variable and ‘geopy’ library in python, latitude and longitude of neighbourhoods’ is found which in turn is used for finding nearby venues using Foursquare API.

## **Data Cleaning**

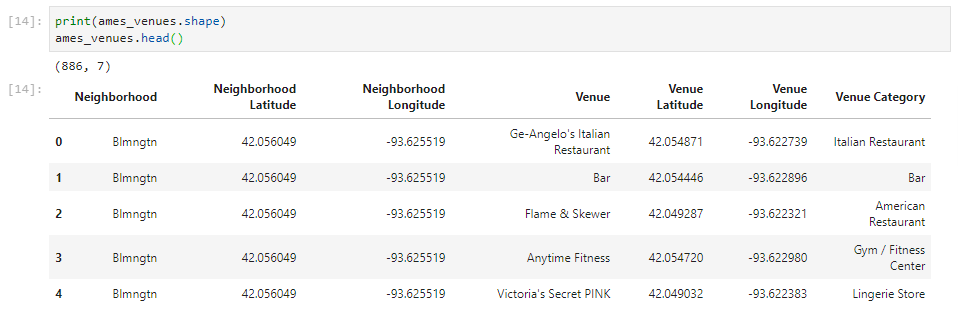
In the Ames housing dataset, each neighborhood is given a code, for example 'Blmngtn' was given for 'Bloomington Rd'. Using the code directly ‘geocode’ could not translate it onto the required latitude and longitude. Further data was given in ‘kaggle.com’ describing the neighborhood code into neighborhood name. Same was extracted and made into a Data frame which was passed into ‘geocoder’ for translation after concatenating ‘Ames,Iowa’

Still there were some locations whose translation could not be run by ‘geocoder’. These were searched in the web and following actions were taken:

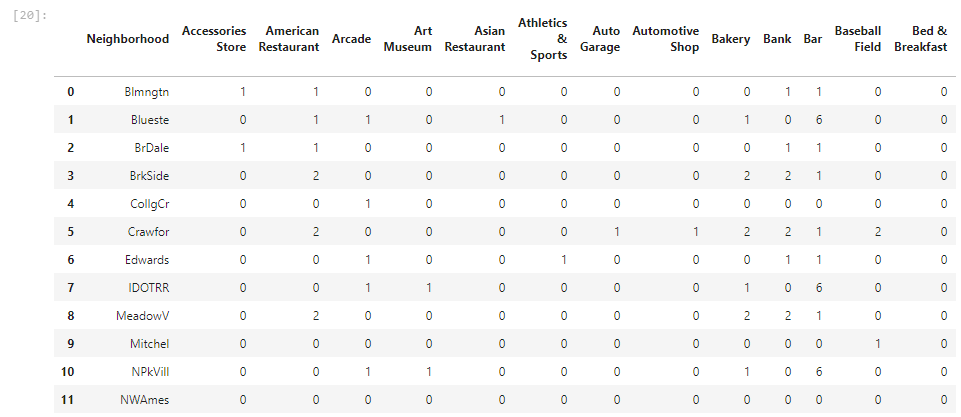
* If the name is different, decide which one to use after searching on the internet.
* If the neighborhood is missing from the geo data frame, add it's coordinate.
* If the neighborhood is made up, combine them into the larger neighborhood which exist in the geo data frame.



The Foursquare API is used to explore the neighbourhoods’ and segment them. The limit was set as 100 venue and the radius 1500 meter for each neighborhood from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API.

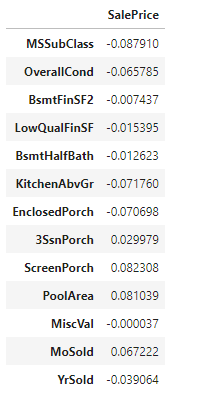


‘One hot encoding’ was done the ‘Venue Category’ and grouped by ‘Neighborhood ’to make the required data set.



## **Feature Selection**

Pearson correlation was done on metric variables on Ames housing dataset and following variables were dropped which were found to have poor correlation:



The Ames Housing data set and venue data are merged into a single one after cleaning on which explanatory data analysis is done. .Further training is done on these datasets using Machine Learning (Regression) algorithms to make the required housing sales price prediction model.