

## Capstone Proposal

Ravinder Singh

### **Predicting Home Prices (Zillow)**

I found this dataset from kaggle regarding housing market. Zillow's Economic Research Team collects, cleans and publishes housing and economic data from a variety of public and proprietary sources. The data consists of 10 years from 1996 to 2007. It has different csv files depending on what we want to look into such as counties, cities , states etc.

#### **Sample Questions :**

- 1) When does the house prices go up and down? Are they depending on size of house or bedrooms, bathrooms etc?
- 2) How many days, months or time being listed for sale has to do with the price changes?
- 3) If the houses are in rural area? Do we have less cost on bigger houses compare to smaller houses?
- 4) Using machine learning algorithm we can predict house prices in the future and have some sort of insights about changing prices depending on the zipcode or the size of house.

## Problem/Client

- What is the problem you want to solve?
  - There are a lot of buyers in the market. The companies can use ML algorithms to predict the selling price of the house depending in the variables instead of contacting different real estate agents which is very tedious and inconvenient.
- Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?
  - This can be used by buyers, real estate agents, sellers or the web based company (Zillow) to help every different kind of clients. It can also be very useful to analyze the housing market based on last ten years data for research purposes.
- What data are you using? How will you acquire the data?
  - This dataset is available on Kaggle and it was acquired from the Zillow website. It is in CSV format. It can also be downloaded directly from zillow website.
- Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.
  - Finding the market prices of houses depending on the zip-code, nearby schools, neighborhood. First of all, I will explore the data and find some insights about it. After exploring, I will have the data clean and pre-processed and find the important variables. I will use different ML algorithms to have the most accurate model that I can have.
- What are your deliverables? Typically, this includes code, a paper, or a slide deck.
  - Code - (Github)
  - Report - (PDF)

## Data Description

Data Link - <https://www.zillow.com/research/data/>

There are many different kinds of data available on Zillow website. I will be using dataset such as time-series data of cities, states and counties. A few file names of the data are listed below.

cities\_crosswalk.csv

City\_time\_series.csv

County\_time\_series.csv

CountyCrossWalk\_Zillow.csv

DataDictionary.csv

Metro\_time\_series.csv

Neighborhood\_time\_series.csv

State\_time\_series.csv

Zip\_time\_series.csv