San diego housing team

# Capstone report 1

**Team Member: Wen Yan, Mengting Wang, Salah Ahmad, Xia Song**

**SECTION 1 Challenge**

The business challenge that we are trying to solve in this capstone project is to evaluate and project housing value in San Diego area by county. Followed by a couple of years of decline since 2006, San Diego housing price is climbing back up again – we are curious to know that when the next peak hits and when the next bubble breaks, which area within San Diego holds the strongest housing value compare to others; in other words, which area within San Diego has the housing value relatively less vulnerable against bigger market.

To achieve this goal, our direction of approach will be evaluating numerous factors that differentiate house values between areas and quantifying appreciation or depreciation of house values overtime. It will involve specific periods such as housing booms and bubble breaks, risks of default in taxation for properties under macro-economic environment.

The research will eventually produce insights and guidance to home owners, investors and government sectors, hence benefit the public. We’ll deploy various data analysis and machine learning algorithms to answer the questions from difference perspectives. Eventually the project will be set up as prototype pipeline to be able to keep track of new data and identify valuable and interesting attributes and findings.

**SECTION 2 Opportunities as a set of questions**

As the general challenge and business questions laid out, we specified the bigger picture into a set of approachable questions, listed as follows:

* What are major factors that impact the housing prices in San Diego county?
* In short-term uptrend/downtrend market, what type of houses by location, profile (attributes) will appreciate/depreciate faster than others?
* In long-term (several decades), will certain houses or areas appreciate more than others?
* Can we predict the house price based on current average price movement in its area?
* Can we identify the area where the house conditions being the major factor that differentiate house prices (in other words, if we flip the house with enhancement, can we make a big profit)?
* If bubble breaks, in which area/what condition will most likely have the mortgage default? (Which area is more vulnerable to the economic environment?)
* Is there any relationship between the house rental price and house market price? If yes, what relationship would that be?
* How to incorporate different data analysis skills and machine learning algorithms to address above questions?
* How to build automatic system that keep tracking on the market and identify any hotspot that are most interesting to different parties?

**SECTION 3 Data Sources**

We will incorporate two sets of data for this analysis for the purpose of cross checking and observe any data errors.

Main data source:

San Diego County data – initial data obtained; more snapshots will follow

* Many characteristics of property in history
* Transactions, ownership, tax & mortgage history for each property

Backup data source:

Redfin scraped data – obtained; one snapshot (Oct. 2017) only

Redfin page of each property for about 700K properties in San Diego County

**SECTION 4 Approach**

**SECTION 5 Team Roles and Responsibilities**

|  |  |
| --- | --- |
| Project manager/coordinator | Mengting |
| Lead programmer | Wen |
| Database manager | Salah |
| Solution Architecture | Salah |
| Analytics & Machine Learning | Xia/Wen |
| Visualization | Mengting |
| Paper write-up | Xia/Mengting |

**SECTION 6 Project Coordination and Communication Plan**

Outside of campus, our team members hold default team meetings through google hangouts 3 times a week on Tuesdays, Thursdays and Saturdays for minimum of an hour.

During each meeting, we communicate current progress, calibrate goals and summarize achievements, and prepare for bi-weekly meeting with advisors.

Other than meetings with Professor Ilkay, our meeting with advisor Professor Volkan is currently set up as bi-weekly, on Fridays after class.

In order to record progress, we set up Trello as watching board to keep track of assigned tasks, upcoming meetings and planned assignments.

**SECTION 7 Bullets for each team member’s individual contributions in Step 1**

Salah: communicating with San Diego county CTO office and various personnel to obtain SD county data

Wen: scrapped Redfin data; set up research goals and questions;

Mengting: set up research goals and questions; prepare proposals and PowerPoints slides; set up and manage Trello boards for the team

Xia: