# Project Proposal - Analyze Real Estate Trends During Pandemic by Utilizing Web Crawler

10/22/2022

Team Name: Team-Chuan Jiang
Member: Chuan Jiang (NetID: chuanj2)

## Project Motivation

I'm on the market for a new home in the southern California area, however I did notice that the home sale price has climbed sharply during the pandemic in the U.S. so, I would like to have an application to help me gather real estate automatically. In the lecture 5.4-Web Search, when professor ChengXiang Zhai introduced the crawler with its complications and strategies, I gained the interest to build my first crawler. Since I have never built a web crawler before, it is a good time to build one that focuses on real estate to serve my home search data needs. Also, it is a good opportunity to dig into the web crawler complications that the professor mentioned in the lecure.

## Project Description

A modern web crawler will be built to grab home listing/sales related data from major real estate websites and/or Twitter. By analyzing the real estate dataset that was extracted from the internet, I will provide a summary report of the real estate trends during the pandemic. The planned tools used in this project will be Python, MS SQL Server, and Power BI. And the estimated hours to complete this project will be 32-56 core hours.

## Planned Tools:

Python: This will be the primary programming language that is used to build the web crawler.

MS SQL server and/or Python: This tool will be used for data cleaning, data validation, and data analysis of the dataset that web crawler downloaded.

Power BI or other data viz tools: This tool will be used to build the final report that I discovered from the dataset.

Other tools: Some other tools may also be involved during the development, however it remains undecided at this time.

## Workloads Distribution:

| Topics | Estimated Hours |
|---|---|
| Research web crawler techniques and complications | 4-8 |
| Build web crawler and testing | 12-24 |
| Data Cleaning and Analysis | 4-8 |
| Final project writeup, data visualization report, and presentation | 12-16 |

| | |
|---|---|
| **Total estimated hours spend** | **32-56** |

**Expected outcome and Evaluation:**
I will be focusing on developing the web crawler, so most of my time will be spent on building the web crawler. This project should deliver a good working web crawler that can extract real estate data from the internet. Then it should resolve the crawler complications to some extents that were mentioned in the Web Search lecture. In addition, a final report will also be delivered to the users to show my findings.