Mod 11 Python Project

Members: Kyle Rindlisbacher, Levi Thompson, Xiaomeng Zhang

# Plan:

The project is an automated program that is a to scrape housing information from the website, store in local SQL, and access with API with some functions of aggregation, filtration, and analysis.

For our group project we will be scrapping data of sold houses in cache valley from “Zillow.com” and placing it into a SQL Server database. Before we can place the data into SQL Server we must cleanse the data, this is because the data was created by many different individuals. Once the data is cleansed and imported into SQL Server, we will use a separate code to extract the data and post it as an API. The data will appear in the API as a dictionary list or a list of dictionaries. This is fine because we will later use Juypter Notebooks to import the API as a pandas table to be better able to show the context of the data we have collected.

# Progress:

As of right now, the group has: scrapped the data and cleansed it, imported the data into SQL Server, exported it into an API that shows it as a list of dictionaries, ran a Juypter Notebook that reads the API data into a Pandas data frame.

# Prospect:

The group is currently working on the last step of analyzing the data and setting up different views of the data to present the information we have gathered. Functions will be written in python language, most likely includes group by, stripe, and other aggregate functions.