

## README

### Group 34 Quadruple Z

#### Group Members

Wendy Zhang	wanyingz@andrew.cmu.edu
Kelly Zhang	yizhang4@andrew.cmu.edu
Yi Zhang	yizhang5@andrew.cmu.edu
Yuhe Zhu	yuhezhu@andrew.cmu.edu

Youtube : <https://www.youtube.com/watch?v=6R6YwJdbNWQ>

#### ● What can it do?

- Our product can decide whether a district in US is a good place for families to make property investment, as well as to let their children study here. The information we provide includes pricing changes, predicted returns, schools, race distribution, etc.
- We need users to input a valid zip code, as well as whether they want us to do live-scraping for the data.
- Then we go web scrapping, collect and analyze the data, and show below information to the users:
  - ◆ The information of schools located in this area. (Scrap from publicschoollreview.com)
  - ◆ "School quality score" of every school, we derive that according to its attributes.
  - ◆ The information of houses located in this area. (Scrap from trulia.com)
  - ◆ "house quality score" of every house, we derive that according to its attributes.
  - ◆ A pie-chart representing the distribution of # people with different races. Its score will be higher if it has more variety.
  - ◆ A time-series plot of the average monthly house price in the past 10 years in this district. An auto-arima model is implemented to forecast the house price for future 6 months using ARIMA model.
  - ◆ Our summary (quantitative analysis) of the value for the families of this district

#### ● How to install - Prerequisite tools for this program

To execute selenium web driver, you may need to install **Google Chrome**. This program does not require other external API or applications to be run. However, you may still need to do install some common packages that not be included in Anaconda. Here are these packages and ways to install them.

##### ■ matplotlib

We use matplotlib package to plot the pie-chart of race distribution and time series of house price. You can simply install it by executing this on your console:

**pip install matplotlib**

##### ■ pmdarima

We use pmdarima package to create time series for monthly house price, and use 'autoarima' function to automatically create an ARIMA model and make forecasting. You can simply install it by executing this on your console:

**pip install pmdarima**

##### ■ bs4

According to our class about web scraping, we use bs4 to get the plain text from a website, as well as search for the elements we want by identifying different kinds of http labels. You can simply install it by executing this on your console:

**pip install bs4**

## ■ selenium

Some websites are really fancy and hard to be scrapped, for instance, housing agents like Zillow or Trulia have Flash-like UI and really hard to find a text element. So we use selenium to simulate human action to these websites. You can simply install it by executing this on your console:

**pip install selenium**

(This package may automatically install a webdriver when initializing)

## ■ tkinter

This is our tool for the frontend. We use tkinter to create windows and diagrams for the users to input their query and see the result. You can simply install it by executing this on your console:

**pip install tkinter**

## ■ PIL

We use PIL to load and show image on our frontend. You can simply install it by executing this on your console:

**pip install PIL**

## \*\*\*\*NOTE on the UI- The front-end

Please refer to [frontend.py](#) for the UI we designed particularly for our product. We have finalized the interface, the buttons, and the logic links for each search operation. We have encountered some problems regarding extracting user input as values and pass on to our backend codes to generate expected results to be displayed in our UI. We have tried converting get() to global variable, StringVar(master = window), and breaking down the values and the function etc.. and more, and we approached you (TAs and Professor Oustlund) for some relevant tips, due to time constraints and perceived complexity, we still haven't managed to move the results perfectly on our UI. But we still include it in our file for your reference to understand our product logic.

To show our efforts, you can run the frontend and load some existing local files.

# PRODUCT DEMO

## Multi-faceted Information

825 Morewood Ave #G-12

3546 Melwood Ave

1045 Devon Rd

4803 Ellsworth Ave

3721 Melwood Ave

825 Morewood Ave #G4

107 Holmes Pl

307 S Dithridge St #406

3229 Parkview Ave

45 Lawn St

Zip Code

House Price

# Bedrooms

# Bathrooms

Area in sqft

Address

City

State

Show house Info

15213

299900

2

2

1176

825 Morewood Ave #G4

Pittsburgh

PA

Show School Info

15213

Pittsburgh Liberty K-5

32

52

K-5

601 Filbert St Pittsburgh, PA 15232

1.3

408

(412) 529-8450

https://www.publicschoolreview.com/pittsburgh-lit

Pittsburgh Science And Tech

Pittsburgh Millones 6-12

Pittsburgh Well K-8

Pittsburgh Greenfield K-8

Pittsburgh Liberty K-5

Pittsburgh Woolslair K-5

Hill House Passport Academy

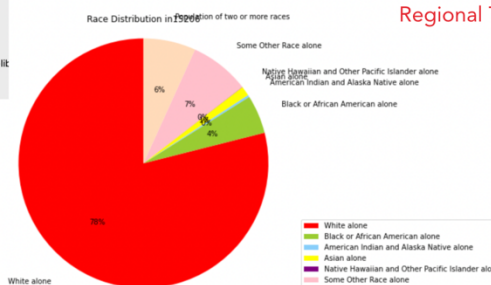
Pittsburgh Miller K-5

Propel Cs - Hazelwood

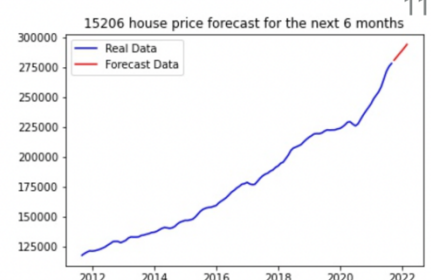
Pittsburgh Arsenal K-5

## Neighborhood School Information

## Neighborhood Demographics



## Regional Time-series Housing Price




## Housing Information

## Group 34 Quadruple Z

### README

---

Welcome to Quadruple Z



**To Empower People To Make Smart Investment Decisions!!**

Enter the Zip Code for house listings:

OR

Choose the Zip Code from our database

### Sample Search Result

Search Completed

[Click here to see house results](#)

[Click here to see housing price changes results](#)

[Click here to see school results](#)

[Click here to see ethnicity results](#)

Search completed!

Found 79 school results, sorted by SCORE:

Found 40 house results, sorted by SCORE:

Race distribution can be found in the race page

Forecast average price for the next 6 months:  
422687.10 430126.44 437635.68 445214.84 452863.91 460582.89  
The estimated annualized rate of return is 18.73%

The school quality score of 15219 is 40.52 out of 100  
The house quality score of 15219 is 44.16 out of 100

Imagary, it is not a good place for your children to study, as well as for you