ADS Assignment 2

So, The assignment is divided into different parts.

The RawDataEDA.pynb is to find out some coorelations and explore the data.

Following is the link for it.

[https://github.com/dhruvkanakia/ADS/blob/master/ADS\_Assignment2/ADS\_Assignment2/rawEda.ipynb](url)

Below attached are few graphs which shows the plots we have tried to explore the data.

![download](<https://user-images.githubusercontent.com/10628795/28014055-3cbf2ad4-6539-11e7-8f2a-75087bd2b5fc.png>)

The above plot shows the correlations between each field with each other. This can be used to fill the missing values.

![download 1](<https://user-images.githubusercontent.com/10628795/28014061-3f364770-6539-11e7-8aa8-3fc0bd960210.png>)

This plot shows how the room count is increasing over the years. Not much can be concluded. But, we see considerable rise between 1950 and 1990 which suggests that maybe during those period the number of apartments having more rooms were built in large numbers.

![download 2](https://user-images.githubusercontent.com/10628795/28014063-40830938-6539-11e7-975c-18d5b406e9c8.png)

This link is to find out if there is any relation between square feet and room count. There are outliers but we can see with increase in square feet the number of rooms are also increasing.

![download 3](https://user-images.githubusercontent.com/10628795/28014068-44bb37e6-6539-11e7-84a5-f1cb32de8f56.png)

After finding out there obviously are relations between columns, we wanted to check how are the missing values affecting the plots. The above plot clearly shows missing values being plotted at 0 which is affecting the pattern.

In the next section we will clean this file and upload it on S3.. Below is the link for it……

[https://github.com/dhruvkanakia/ADS/blob/master/ADS\_Assignment2/ADS\_Assignment2/Ingestion.ipynb](url)

Using various functions we first find out the shape, mean, median, count, standard deviation to see how is the data distributed. Based on the results we clean the data.

The clean data csv file has been uploaded on S3.

Dataingestion.py

Cleaning and storing data on S3

RawDataEDA

Uploading the Data into IBM Cloudant

Getting the ten nearby places using the cloudant data and algorithm

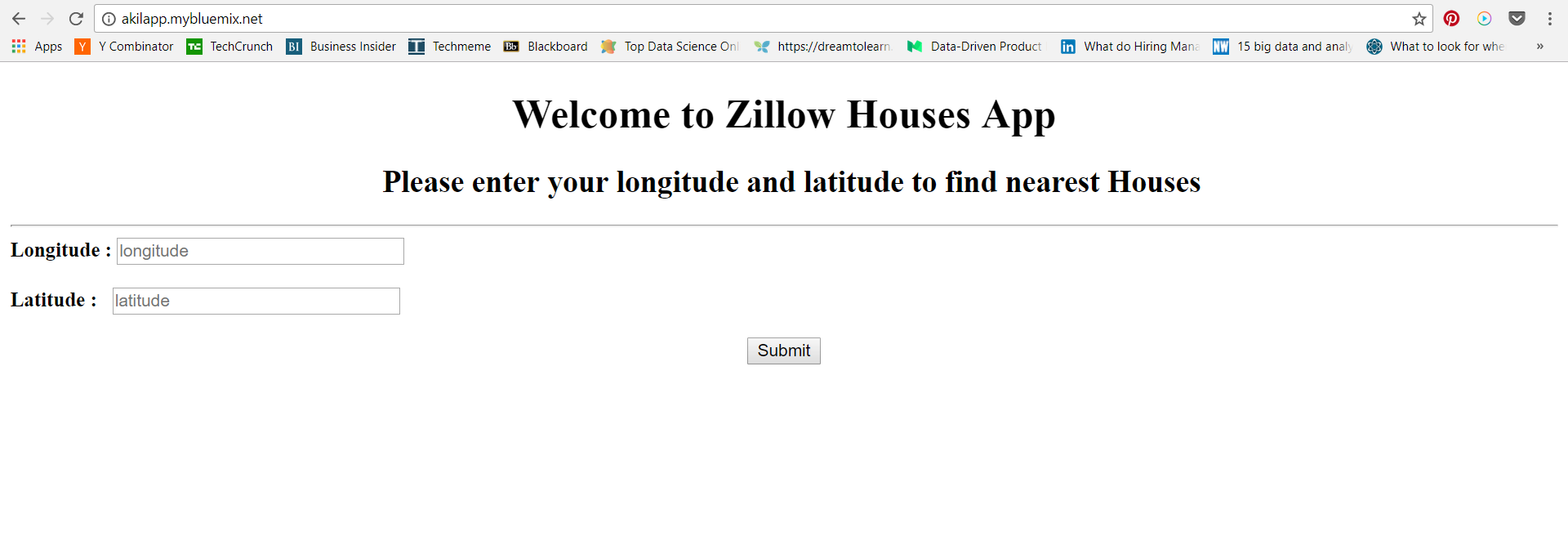
The Working of the REST API using Python Flask is inspired from :

<https://regebro.wordpress.com/2010/12/13/python-implementing-rich-comparison-the-correct-way/>

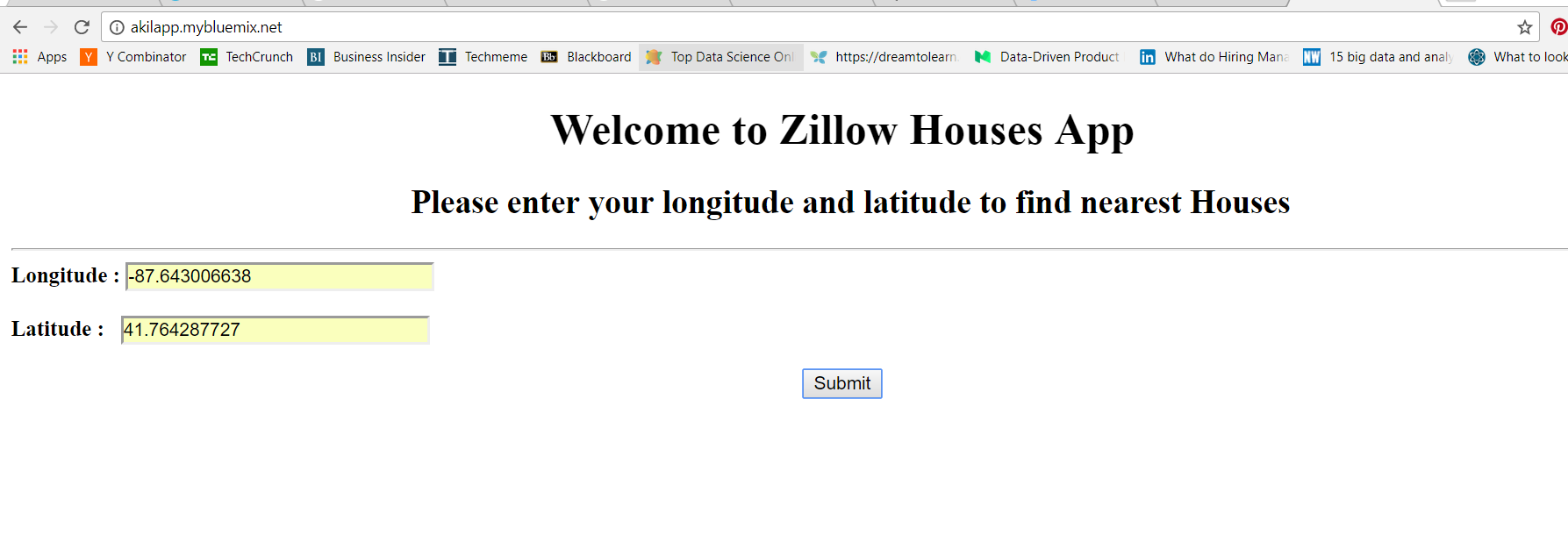
<https://jeffknupp.com/blog/2014/06/18/improve-your-python-python-classes-and-object-oriented-programming/>

<https://stackoverflow.com/questions/4913349/haversine-formula-in-python-bearing-and-distance-between-two-gps-points>

The REST API hosted in the IBM Bluemix Cloud looks like



After Submitting the results



Search Results page looks like

