[](https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwikxez5-9vTAhVD0oMKHVxNCA8QjRwIBw&url=https://en.wikipedia.org/wiki/File:NYU_logo.png&psig=AFQjCNEKMjDIM3BjN_ueuJYTL3okEaMgBw&ust=1494184160353163)

**P.A.I.S.A**

**Property and Investment Statistical Analysis**

**Project Report**

May 11, 2016

Course: Cloud Computing (CS-GY 9223)

Faculty: Professor Sambit Sahu

Semester: Spring 2017

Live Demo Link:

<http://ec2-54-218-102-247.us-west-2.compute.amazonaws.com:8000>

YouTube Link:

[**https://www.youtube.com/watch?v=DlCUJY5mOfA&t=8s**](https://www.youtube.com/watch?v=DlCUJY5mOfA&t=8s)

Team Members:

Aishwarya Ajay Das (aad532)

Chinmay Sumant (cs4681)

Harsh Trivedi (hat280)

Suril Shah (sus214)

**Table of Content**

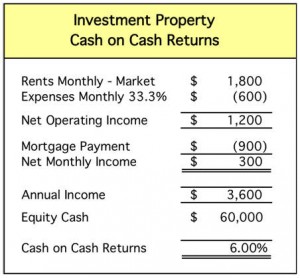
|  |  |
| --- | --- |
| Topic | Page No. |
| Abstract | 3 |
| Motivation | 3 |
| Introduction | 4 |
| Architecture | 4 |
| Technology Stack | 4 |
| Flow Diagram | 5 |
| Implementation I | 5 |
| Implementation II | 8 |
| Testing our Analytics | 11 |
| Future Scope | 13 |
| Conclusion | 13 |

Abstract:

Property Investment is a financial decision which needs deep insight into market values, property rates and real estate know-how. This decision should be data driven so that it helps the individual buying the property to achieve maximum return on investment.

Motivation:

“Cash on Cash return” is the most important measurement. So, while the price is important, one’s actual cash equity investment is the vital issue. So, for every dollar invested what is our percentage yield return on our equity cash investment. Cash Deposits offer 1.5%, Bonds 4.5%, stocks 7.5% and real estate is generally high risk, so we want fairly high returns to compensate for the risk.

[](http://www.zillow.com/blog/files/2011/08/cash-on-cash.jpg)

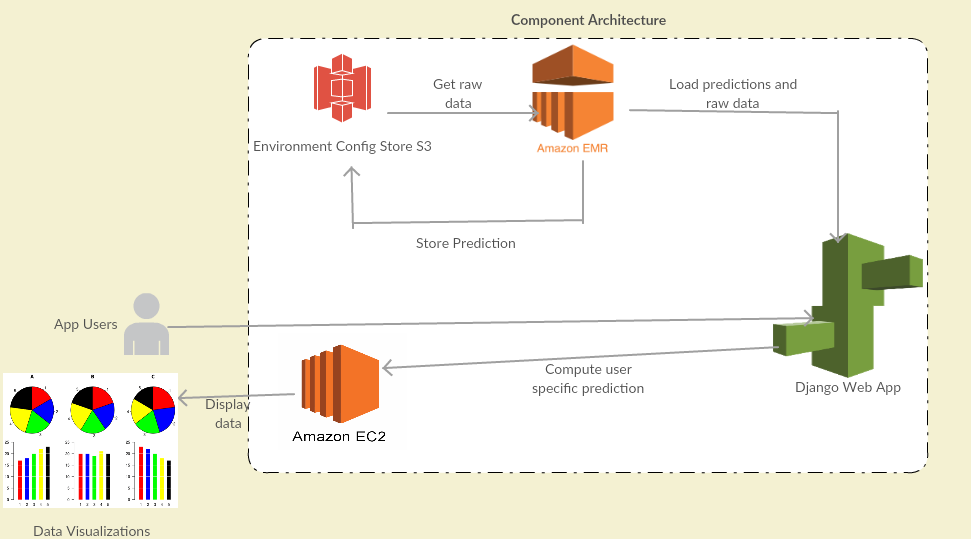
If one is buying a $200,000 investment property they probably put down 25% or $50,000 plus another 5% or $10,000 for closing costs, loan fees and rehab costs. So, the mortgage is $150,000 and a buyer’s cash equity is $60,000 from the start. Again: The property price of $200,000 is important too, but how much cash equity one invests is much more important.

Using a conservative estimate, depending on the local market, that property might generate $1,800 per month in rent and have 33.3% operating expenses ($600) leaving net operating income of $1,200. Then subtracting the monthly mortgage payment of $900 leaves $300 of monthly cash flow or $3,600 per year.

Introduction:

We present an application which incorporates large amount of property data, data processing, machine learning algorithms, web front end to display results and cloud technology. The application provides probable locations where the person should invest in real estate considering its market value in near future. The prediction of probable location takes into account the budget constraint of the individual. All the predictions are done using a proprietary algorithm. The list of counties will be shown on google maps. Multiple datasets are taken into consideration and various attributes like property values, rent etc. will be analyzed to predict highest property appreciation areas and top current rent prospect.

Architecture:



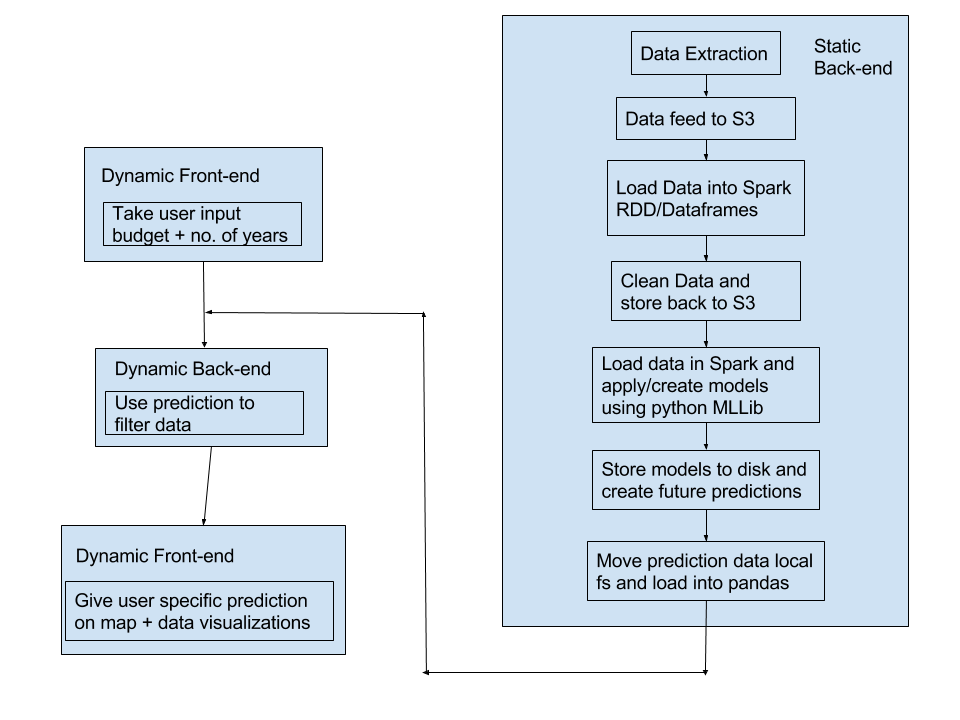
Technology Stack:

1. Spark on top of Hadoop Cluster
2. Python
3. Pandas
4. Django
5. JavaScript
6. MLlib
7. HTML
8. CSS

Cloud Components:

1. AWS Elastic MapReduce
2. Amazon S3 storage
3. AWS EC2

Flow Diagram

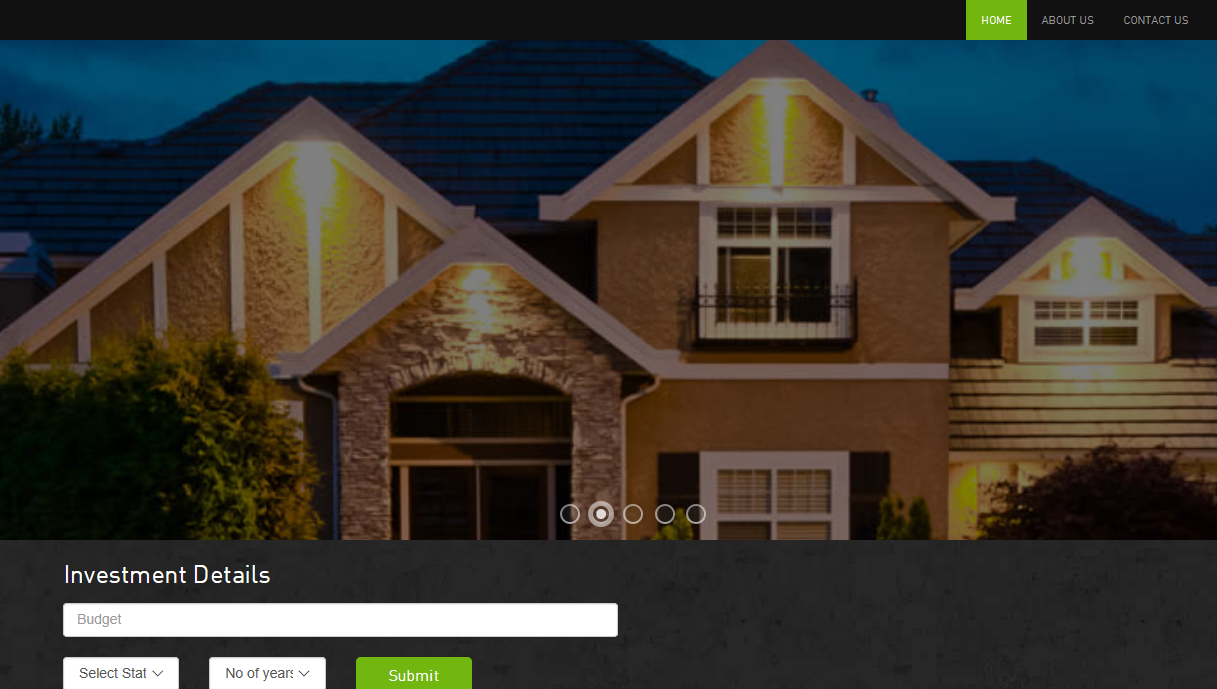


Implementation I (up till the Hackathon):

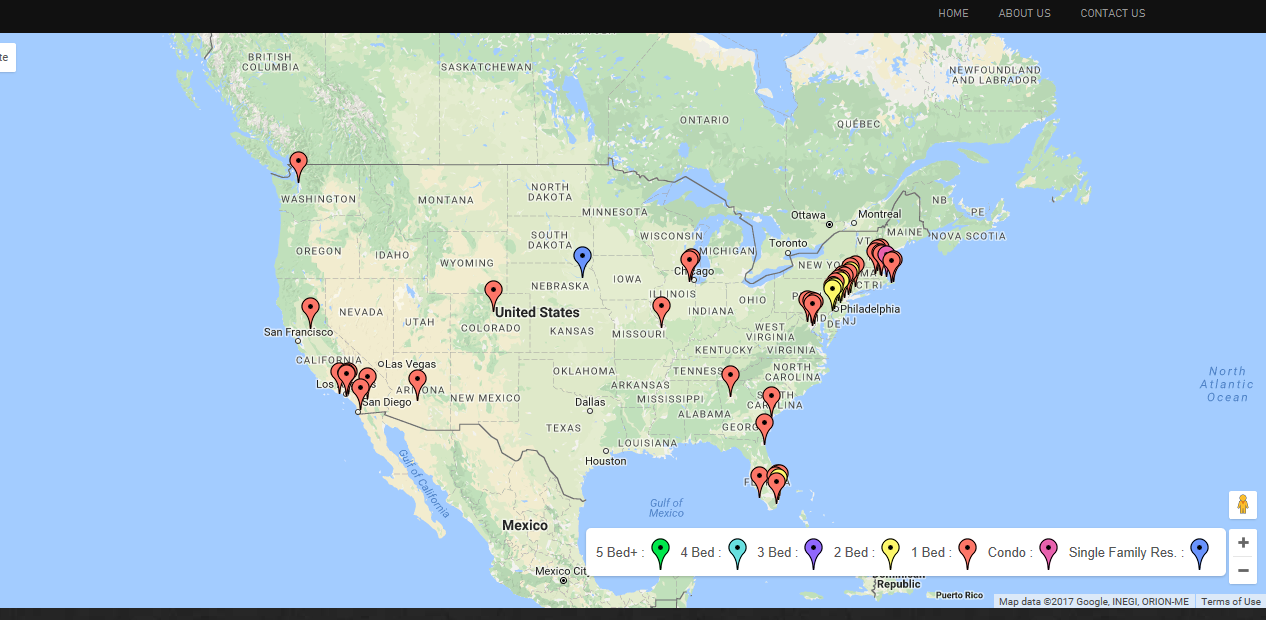
1. Designed Static and Dynamic workflow
2. Designed Project Architecture and Finalized AWS Components
3. Finalized Technology Stack
4. Coded Front End template

The Front End includes the following pages:

1. Home Screen



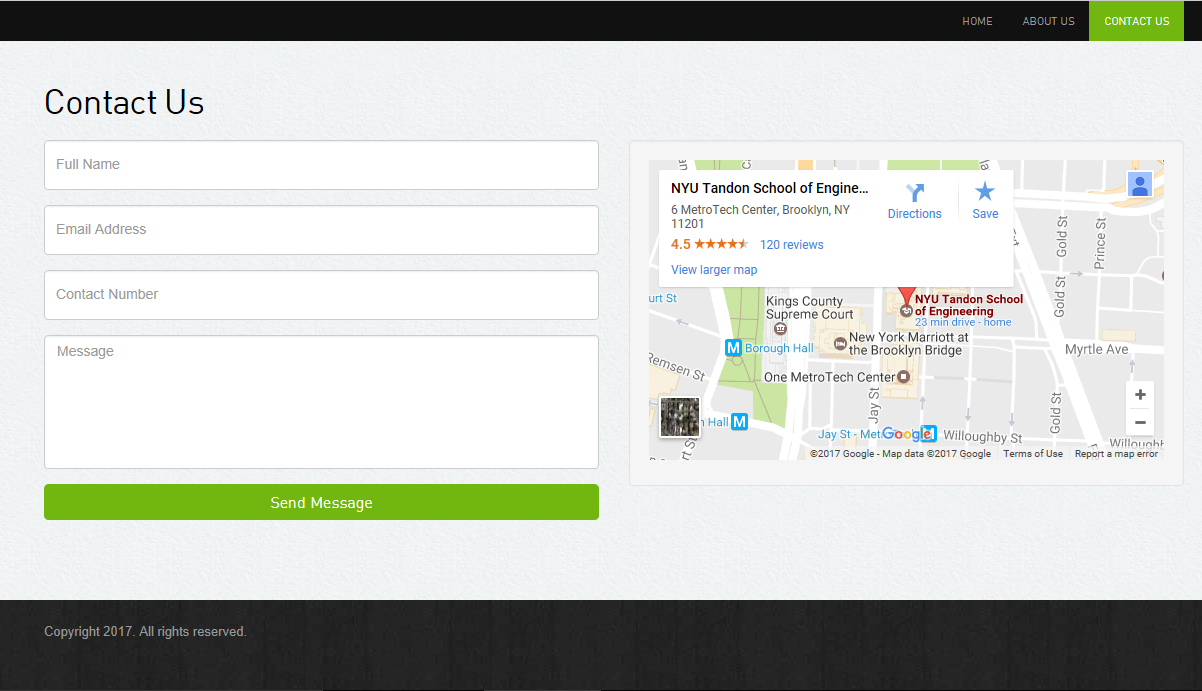
1. Result Page without the graph



1. About Us

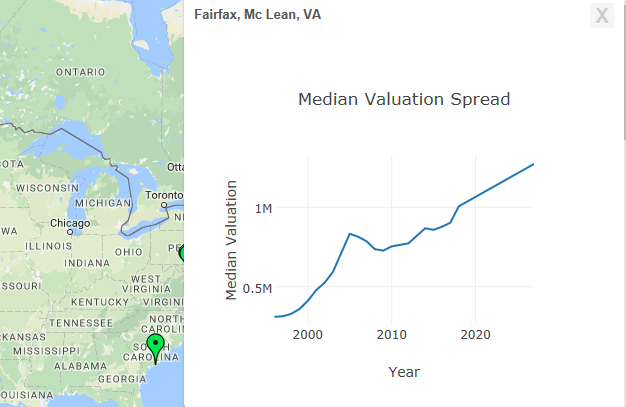


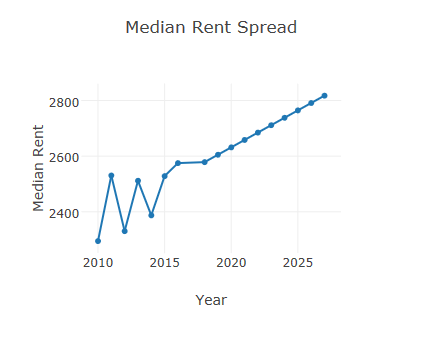
1. Contact Us



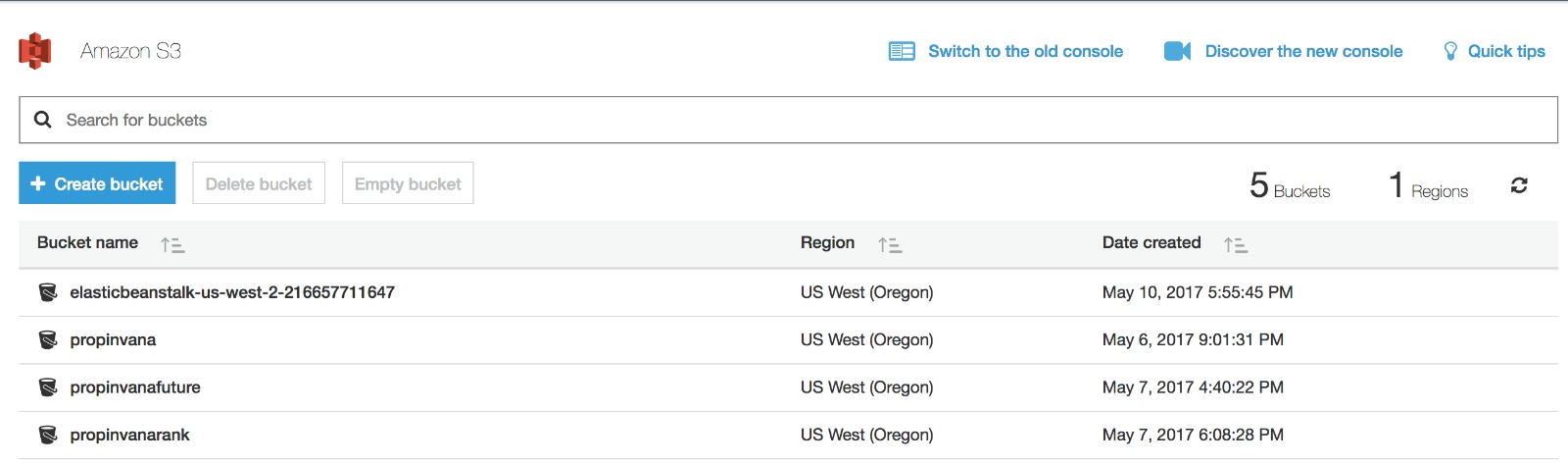
Implementation II (after the Hackathon):

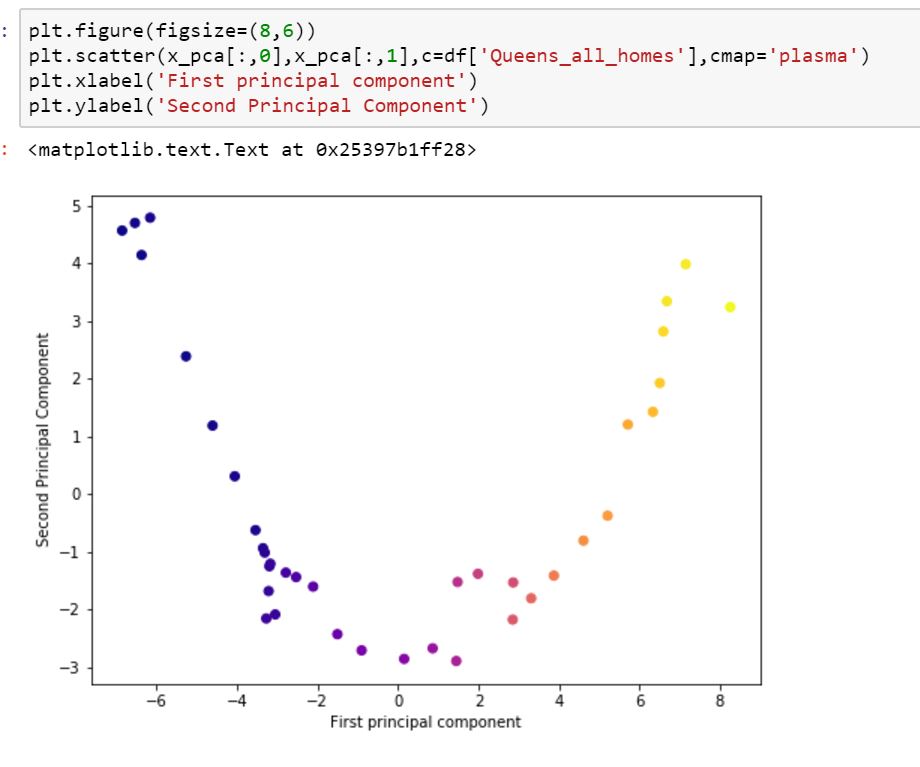
1. Implemented the Graph on the Result page using Plotly API

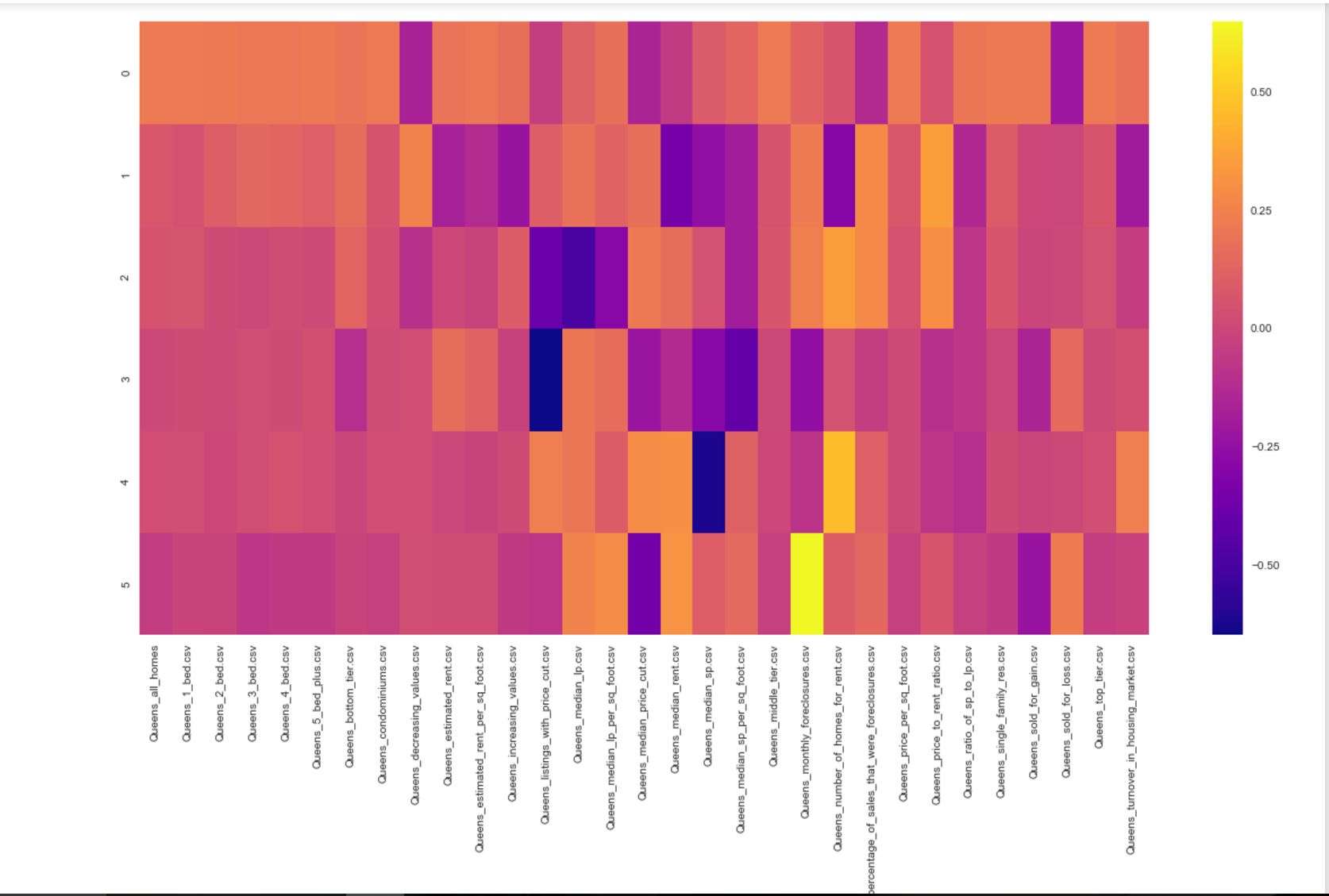




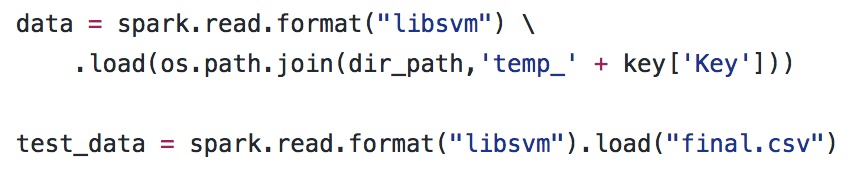
1. Aggregate all data on Amazon S3.



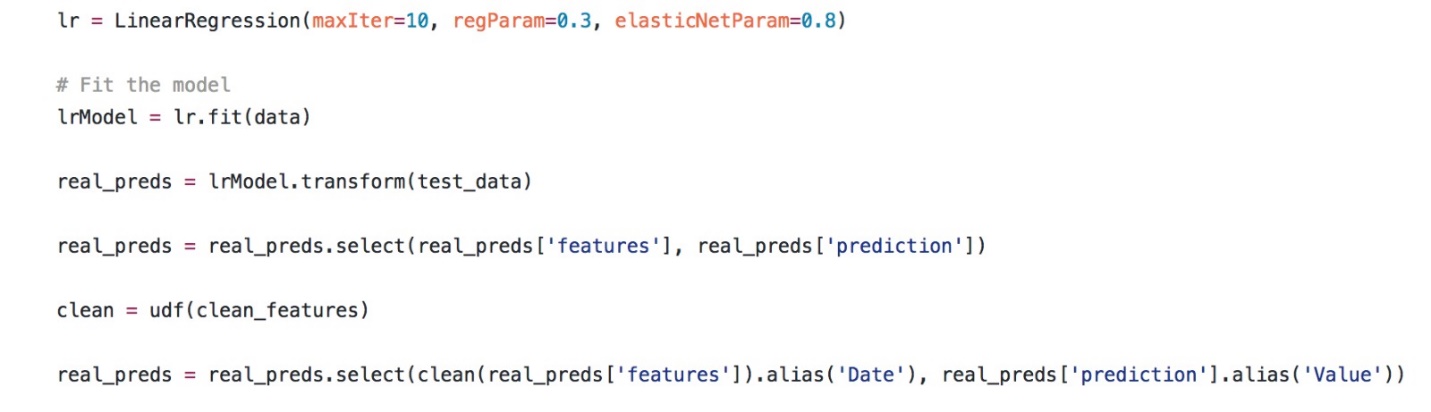
1. Implement Principal Component Analysis on features and decide on which features to use in the model.



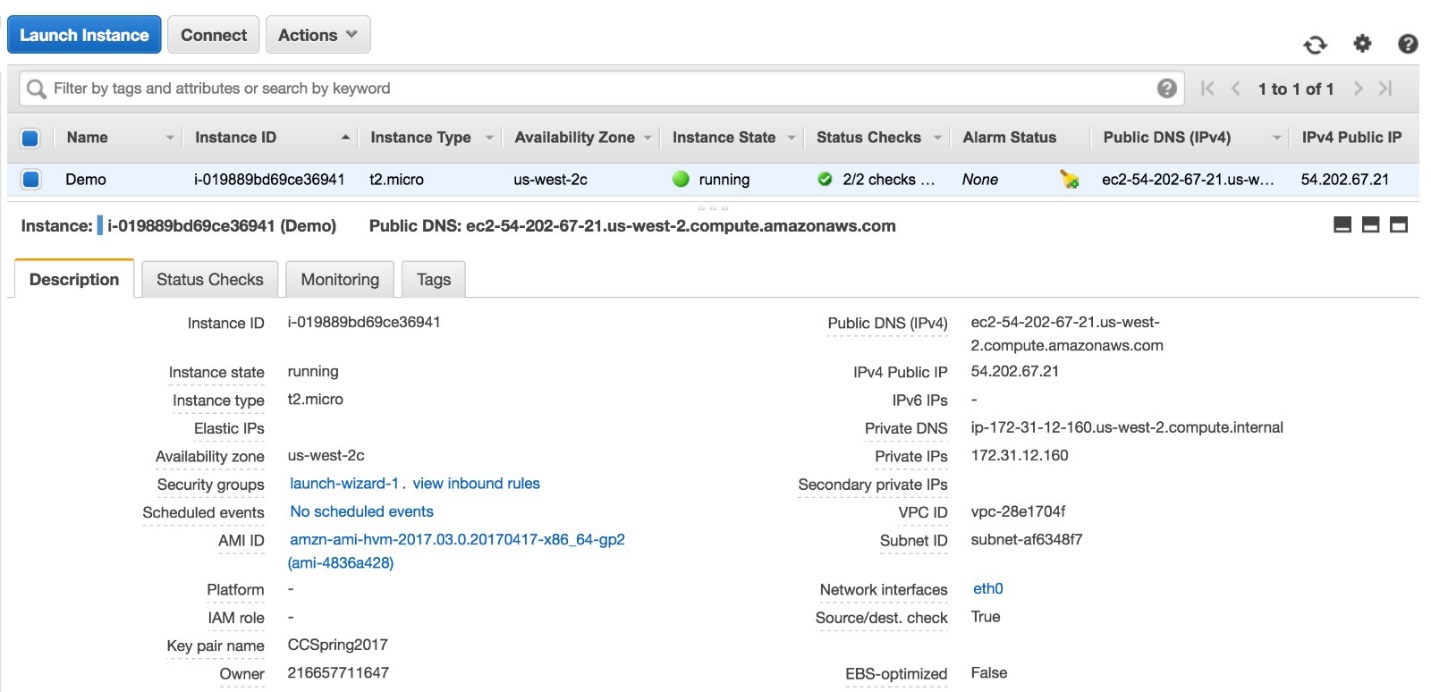
1. Write Spark Analytics scripts to implement the above points.



1. Shortlist Machine Learning Algorithm that will be used for prediction.



1. Integrate code into AWS component workflow.



1. Integrate front end workflow with back-end search using Django Framework.

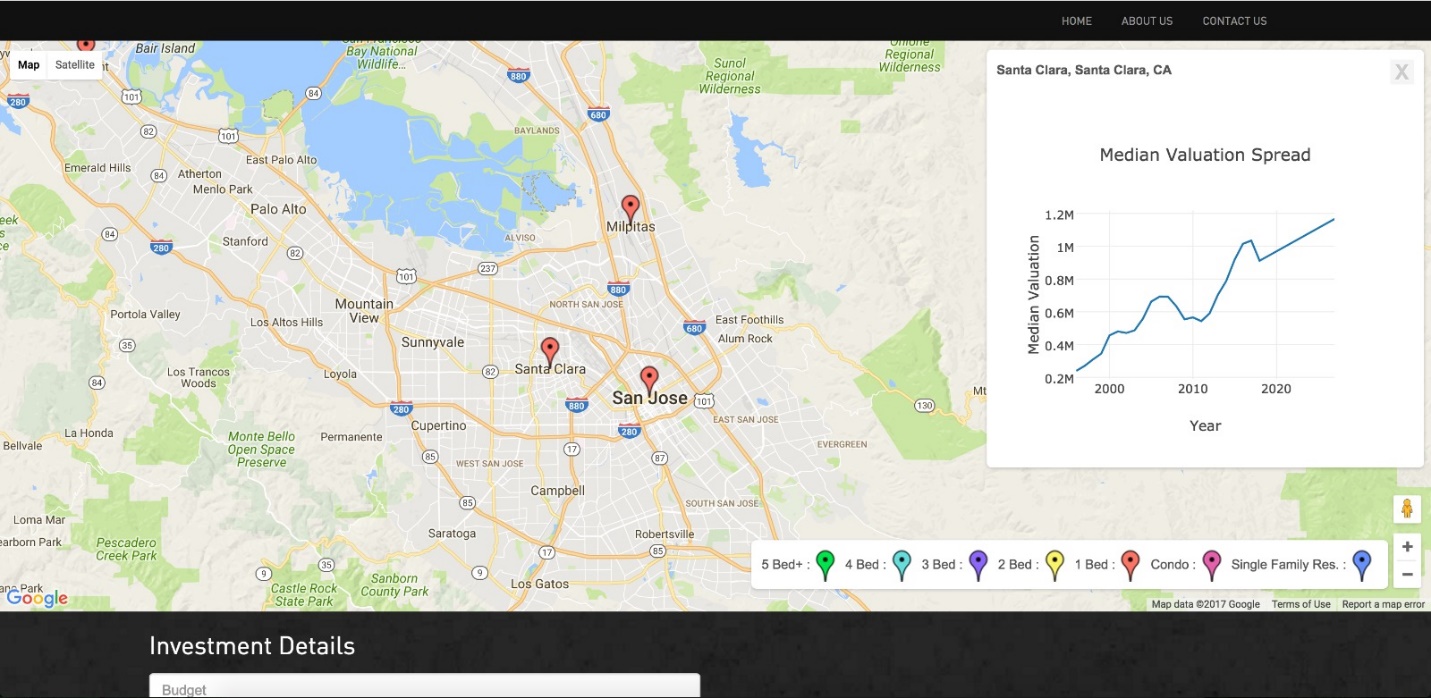
Testing our Analytics:

We took random values and saw the predictions that our model made.

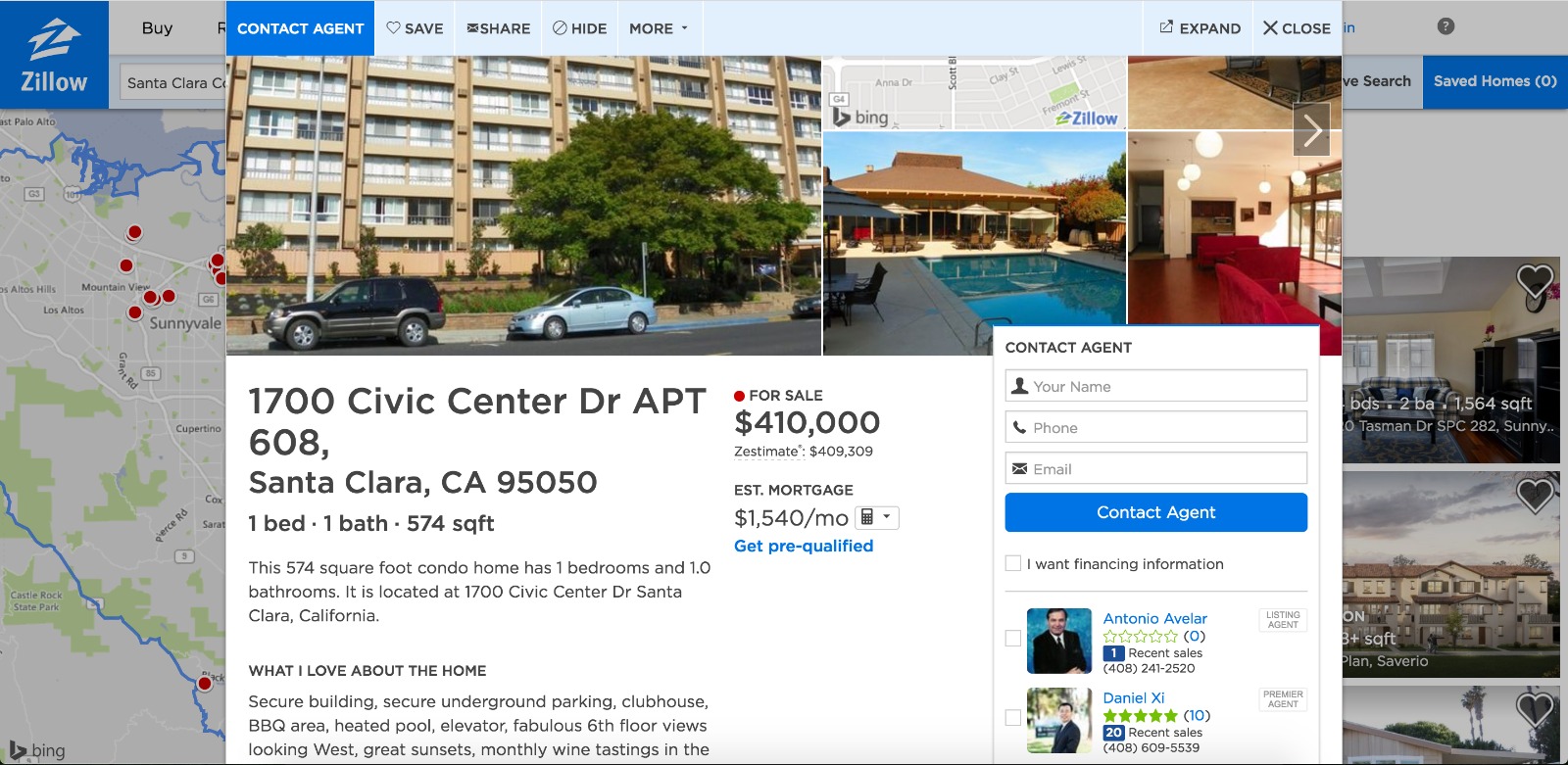
1. One such test case is discussed below:
   1. We inputted the following details

|  |  |
| --- | --- |
| Budget | $ 500,000 |
| State | All |
| No. of Years | 5 |

* 1. The output screen we got was



* 1. To test the if our prediction is correct we looked randomly and picked one location
     1. We picked Santa Clara
     2. To test the correctness, we went to Zillow to find whether any 1 BHK property was available within our budget and we found the following listings.



C:\Users\acer\AppData\Local\Microsoft\Windows\INetCache\Content.Word\56AC23B1-18F3-4F15-9405-3DB48E7852C5.jpg

* 1. This proved our data is right.
  2. Now we checked the ROI in our backend and compared it other places to see if our prediction was right. It proved to be correct.

Future Scope:

1. Scale this to Worldwide level.
2. Create a Mobile application for the same.
3. Integrate with third party vendors to aid the user to show him the properties available.
4. Gather more data to improve on our predictions.
5. Use of Hedonic Regression to instead of Linear Regression

Conclusion:

PAISA helps user to take a data driven approach while investing their money and reduce their risk. We thank the Professor and the Teaching Assistants for giving us this opportunity for creating a working prototype of a part of a real-world problem and guiding us at multiple stages during the project.