**Property and Investment Statistical Analysis**

**Aishwarya Ajay Das, Chinmay Sumant, Harsh Trivedi, Suril Shah**

**Project Link: -** [**http://ec2-54-187-64-142.us-west-2.compute.amazonaws.com:8000/**](http://ec2-54-187-64-142.us-west-2.compute.amazonaws.com:8000/)

|  |  |  |
| --- | --- | --- |
| **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.  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** | **Work Flow**  **https://lh3.googleusercontent.com/b6vvbl1oTTTYY-1HkBLu65RtBjk068w3VEagwIQmSO8cu53DI3zM-Qvn5Y12sBo6iFSTel1AAfhkNtRX94jJfTv2mYeHMDLJSenZD1aj00LGXuejn9i51Zrq6JIUO_Oj2pZDW04l** |
| **Technologies Used**  **Web Framework:**  Spark on top of Hadoop Cluster, Python, Pandas, Django, JavaScript, MLlib, HTML, CSS  **Cloud Technologies:**  AWS Elastic MapReduce, AWS Elastic Beanstalk, Amazon S3 storage, AWS EC2 | **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 |
| **Algorithm**   1. The Data is obtained using the Quandl API. 2. Feature Selection and Principal Component Analysis is performed to find relative importance of the attributes that affect the rent and resale value. 3. Now the values of these attributes are linearly regressed to get future values. 4. Depending upon the user inputs, the sum of the resale value at that time and the rent obtained until that time is calculated and top properties are suggested. | **Conclusion**  PAISA helps user to take a data driven approach while investing their money and reduce their risk. |
| **Dataset**  The Quandl API for housing data is used in our project. Quandl offers 30+ real estate market indicators for over 35,000 geographical areas across the United States. |