# PREDICTIVE ANALYSIS LAB PROJECT

## URBAN BUILT-UP AREA

|  |  |  |  |
| --- | --- | --- | --- |
| GROUP MEMEBRS NAME | SAP ID | ROLL NUMBER | BATCH |
| SURAJ KUMAR PANDEY | 500107301 | R2142220901 | 7 |
| VIVEK NATH | 500107662 | R2142220900 | 7 |

## SUBMITTED TO:

Dr. Achala Shakya

**Abstract/Introduction**

This theme deal with the predictive analysis of urban land/built up areas as this project targets to understand growth trends in population, land use changes and infrastructure growth into an process based environment for one of the most populous country i.e. India. Using four decades worth of data, we seek to gain insights in order to aid urban planning decisions.

Problem Statement

As urbanization continues to rise, there comes the need to examine population demographics and how it appears with land use along with infrastructure. And this project will solve the problem of predicting such trends to help policymakers with mitigation and effective urban planning.

**Objectives**

Determine the growth rate of population in India.

The only part that is still missing (albeit not in the strict sense) is exploring how regional population affects urban infrastructure

– Mapping trends in urban migration as well as land use transformations.

**Dataset Description**

This data set consists of variables related to villages, towns, households, population and area for each state in India. It contains features of both numerical and categorical types over 40 years.

**Methodology**

Data Preprocessing: Eliminating duplicates, managing the missing values

Feature Engineering: Encoding categorical variables and scaling numerical features.

Visualization → Plotting bar graphs, scatterplots and heatmaps to get a feel of the data

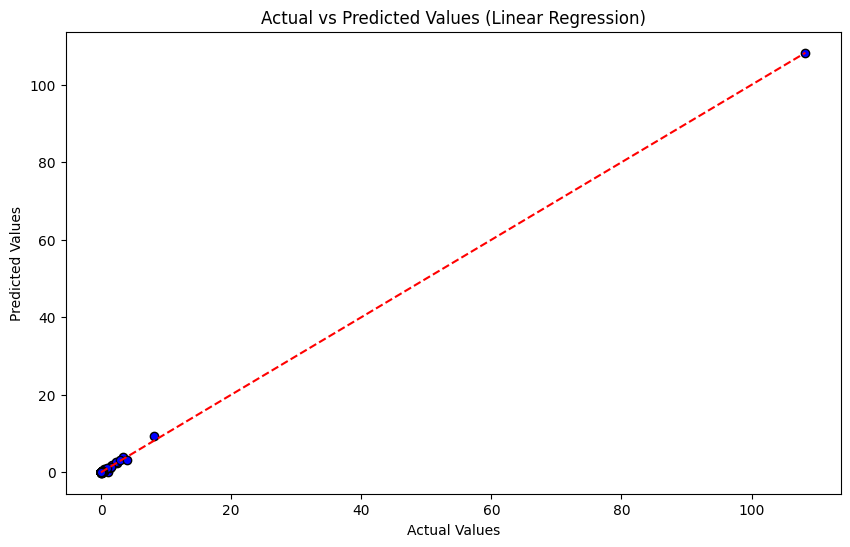
Modeling : Finding solution for predictions using machine learning algorithms like Linear Regression, Decision Trees etc.

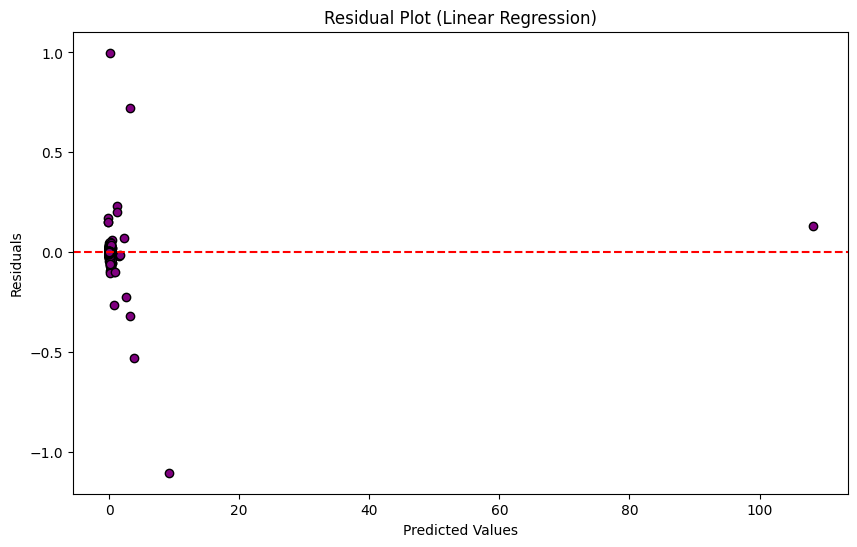
**Visualizations and Insights**

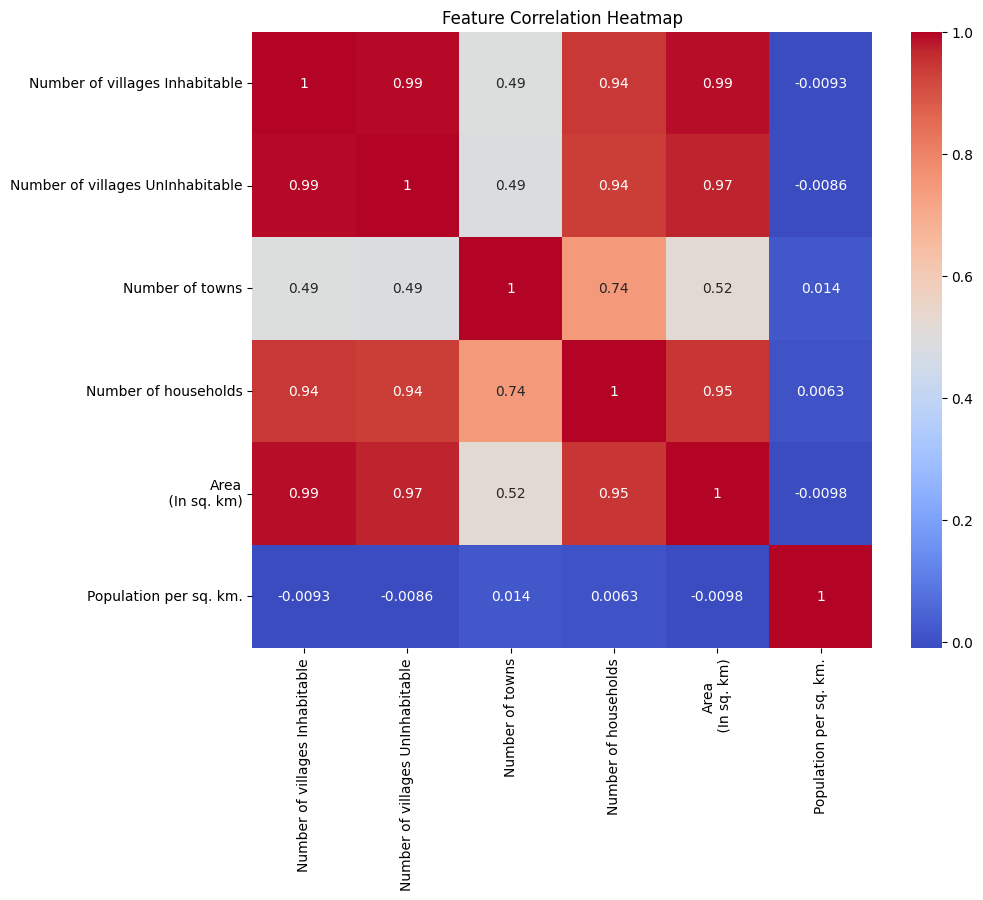
Some of the important visualizations are: — state wise population distribution, Population vs. Households relationship and Correlation Plot between numerical features. The visualizations enable extracting the valuable insights related to urban trends.

Performance of the Machine Learning models

Conducted Linear Regression to forecast population trends. We are shown R² Score and Mean Absolute Error (MAE) which indicate the accuracy and shortcomings of our model.

****

****

****

**Predictions**

Once the model was trained, it was used to predict for new input. i.e predicted population under defined urban parameters: [2.32449759e+08]

**Challenges Faced**

I went through several challenges like dealing with null values, processing the data as they were quite large and choosing the right features for the models.

**Future Scope**

More datasets, including live urban planning data

Finding and implementing advanced algorithms (deep learning is an option) to enhance predictions,

Extending the analysis to other geographical areas or situations.

**Conclusion**   
This research offers a thorough examination of urbanization trends, facilitating efficient urban planning decision-making. Important insights into population growth and changes in land use are provided by the predictive models and visualizations.  
Citations   
  
- Scikit-learn and Pandas are libraries for machine learning and data analysis.   
- Seaborn and Matplotlib for data visualization.   
Dataset: Historical information about Indian cities.