📄 COVER PAGE

Nagpur Real Estate Analytics & Forecasting Dashboard
Project Documentation

Submitted by: Ritik Kumar

Submission Date: 26-02-2026

📄 TABLE OF CONTENTS

1. Introduction

This project presents an end-to-end real estate analytics pipeline for Nagpur. It includes web scraping, data cleaning, exploratory analysis, price forecasting using Facebook Prophet, and deployment of an interactive Streamlit dashboard.

The system enables locality-wise price comparison, trend visualization, and future price forecasting to support investment insights.

2. Objectives

- Scrape real estate listings from MagicBricks
- Clean and standardize raw property data
- Perform locality-wise price analysis
- Forecast future price per square foot
- Build a multi-page Streamlit dashboard
- Enable dataset download for further analysis

3. Data Collection

Data was scraped from MagicBricks using:

- requests for HTTP requests

- BeautifulSoup for HTML parsing

Collected fields:

- Locality

- Property type

- Price

- Area (sqft)

- Price per sqft

- Scrape date

Scraping was conducted at a controlled rate for academic use.

4. Scraping Challenges (MagicBricks)

- Dynamic content loading limited static HTML access

- Frequent HTML structure changes

- HTTP 403 blocking on repeated requests

- Pagination required manual handling

- Inconsistent price and area formats

Mitigation techniques included custom headers, time delays, and post-scraping normalization.

5. Data Preprocessing

Data cleaning steps:

- Removed duplicates and missing values

- Converted price strings (₹, Lac, Crore) to numeric

- Standardized area units to sqft

- Created avg_price_per_sqft feature

- Normalized locality names

- Removed outliers

Generated datasets:

- nagpur_real_estate_raw.csv

- nagpur_real_estate_cleaned.csv

- locality_stats.csv

- forecast_summary.csv

.

6. Exploratory Data Analysis

EDA included:

- Locality-wise average and median price analysis

- Price distribution visualization

- Identification of premium and affordable areas

7. Forecasting Methodology

Since the dataset is cross-sectional, a simulated time series was created per locality.

Facebook Prophet was used to:

- Model trend components

- Generate future forecasts

- Compute confidence intervals

- Estimate growth percentage

8. Streamlit Dashboard

Page 1 — Dashboard

- Locality selection

- Metric cards (avg price/sqft, median price, listings)

- Price distribution histogram

- Top 5 expensive and affordable localities

Page 2 — Trend & Forecast

- Historical trend visualization

- Prophet forecast with confidence interval

- Forecast summary table

Page 3 — Compare Localities

- Multi-select locality comparison

- Bar chart of average price per sqft

- Locality statistics table

Page 4 — Download Data

- Download cleaned dataset

- Download forecast summary

9. Tools & Technologies
- Python
- Pandas, NumPy
- Plotly
- Facebook Prophet
- Streamlit
- BeautifulSoup

10. Key Insights
- Premium localities have consistently higher price per sqft
- Emerging areas show higher forecast growth potential
- Affordable regions offer better entry-level investment opportunities

11. Limitations
- Forecast is based on simulated time series
- Scraped data may not represent the entire market
- Dynamic content limited feature extraction

13. Conclusion

This project demonstrates a complete real estate analytics workflow, combining web scraping, preprocessing, forecasting, and interactive visualization. It provides actionable locality-level insights for price comparison and trend analysis in Nagpur.