INNOMATICS®
RESEARCH LABS

INNOVATION. AUTO**MAT**ION. ANALY**TICS**

PROJECT ON

Commercial Real Estate Market Analysis
A Data-Driven Study Using Web Scraping & Exploratory Data Analysis

By: Rohit Girhe   **Date: 06-12-2025**

Agenda

1. Business Problem Statement
2. Market Opportunity
3. Objective of the Project
4. Data Collection Methodology
5. Dataset Overview
6. Univariate Analysis
7. Bivariate Analysis
8. Multivariate Analysis
9. Market Segmentation Strategy
10. Challenges & Learning
11. Conclusion

INNOMATICS
RESEARCH LABS

Business Problem

1. Information Asymmetry:

- No centralized market data on rental rates, deposits, or specification

- Business negotiate blind without market benchmarks

- Result: Leave 15-20% value on the table annually.

2. Valuation Complexity:

- Same 300 sqft office costs ₹15k in Swargate vs ₹80k in Baner (5.3x difference)

- No standardized pricing framework to explain why

- Landlords underprice or overprice, losing tenants or revenue

3. Location Strategy Dilemma:

- Impossible trade-off: Premium location (Visibility) Vs Budget location (save 60k/month)

- No data-backed framework for location selection decisions.

- Companies choose based on CEO intuition, not analytics

INNOMATICS
RESEARCH LABS

Market Opportunity (Scope & Scale)

1. Pune's Commercial Real Estate Growth:

- Annual growth rate: 12-15%

- Startup ecosystem: 400+ new startups per year

- IT hub expansion: Hinjewadi, Viman Nagar, Kharadi

- Market Size: 5,000+ crores annual leasing volume

2. Data Coverage:

- Properties analyzed: 555 (representative samples)

- Geographic converage: 25+ key localities

- Property Types: 6 categroes (Offices, shops, warehouses, etc.)

- Data Quality: 100% complete, zero missing values

3. Strategic Value:

- For Business: Save 15-20% on rent through informed negotiation

- For Landlords: Maximize rental income by 30-40% with market data

- For Investors: Identify emerging areas (40-60% capital appreciation potential)

- For Agents: Build client trust with data-backed valuations

INNOMATICS
RESEARCH LABS

Objective of the Project

Primary Objective:

Develop a comprehensive, data-driven market intelligence framework for Pune's commercial real estate to identify key pricing drivers and market segmentation patterns.

Secondary Objectives:

1. Quantify geographic location premium across Pune Localities
2. Establish relationship between property attributes (size, type, furnishing) and rental pricing
3. Identify three-tier market structure and strategic opportunities
4. Create actionable intelligence for businesses, landlords, and investors

Enable data-driven decision-making in fragmented market, reducing information asymmetry and capturing 15-20% value through informed location & investment choices.

INNOMATICS
RESEARCH LABS

Data Collections Methodology

1. Data Source Selection:

- Platform: NoBroker.in
- Why chosen: Largest commercial platform, daily updates, phone-verified listings
- Justification: Alternative platforms weaker (99acres outdated, PropTiger residential-focused)

2. Web Scrapping Tools

- Python, BeatifulSoup, Requests, Pandas, Jupyter Notebook

3. Data Collection Pipeline:

Step 1: **Identify URLs** → Step 2: **Extract Raw Data** → Step 3: **Data Cleaning** → Step 4: **Manual Verification** → Step 5: **Quality Assurance** → **Final Dataset** (650 Properties)

4. Cleaning Process:

- Raw records: 650
- After Cleaning: 555
- Removed: 95 records (Duplicates, missing fields, invalid data)
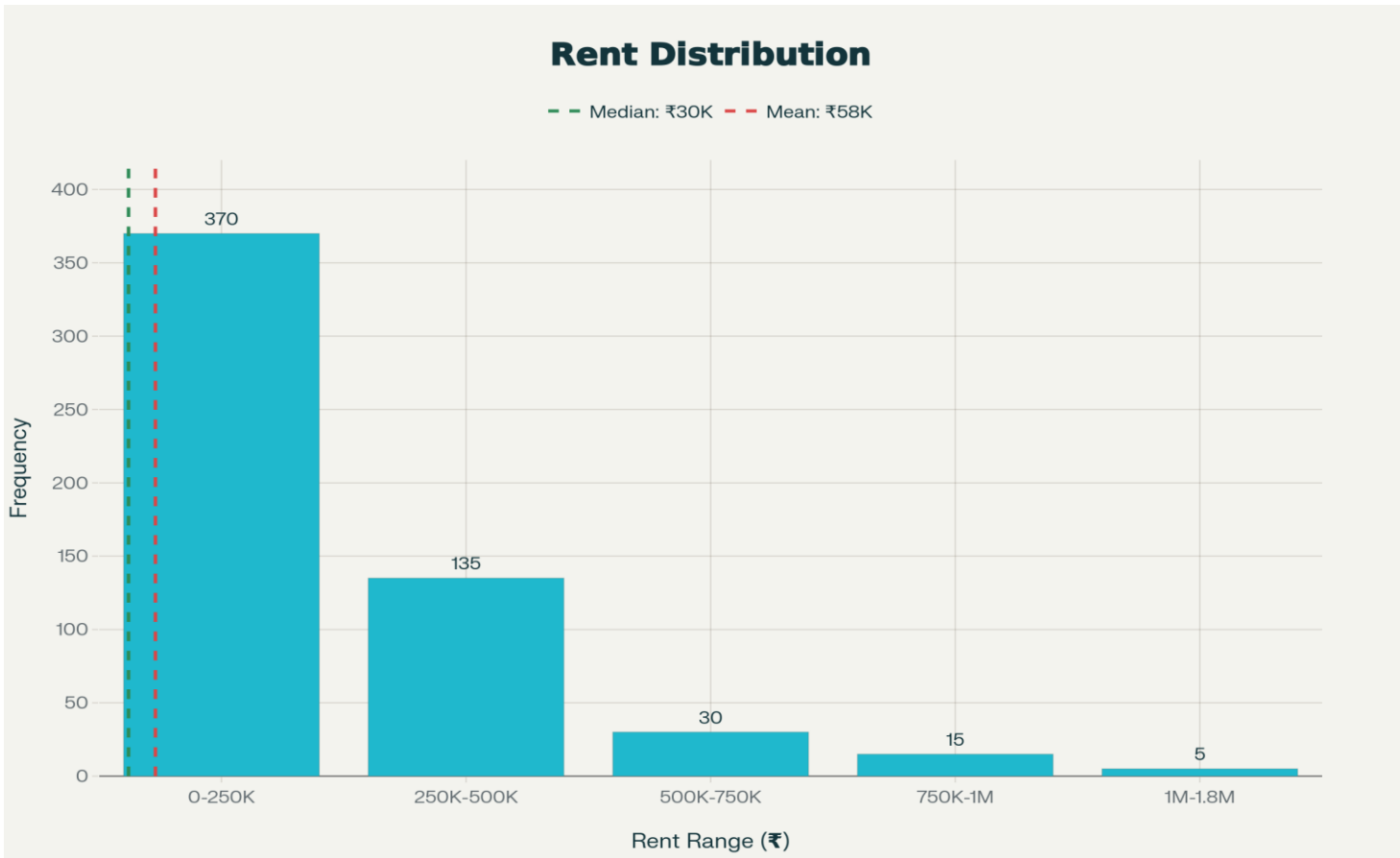
Dataset Overview & Quality

Raw Data:

	Locality	Title	Location	Rent_raw	Rent_inr	Deposit_raw	Sqft_raw	Sqft	Price_per_sqft	Parking	Furnishing	Available_from	Property_type	Deta
0	Kothrud	Co-Working space in The Business Hub, Pune for...	hapoy colony near van devi mandir karve nagar ...	₹ 12,000 No Extra Maintenance	12000	₹ 24,000	650 sqft	650	18.46	Parking	Furnished	22-10-2025	NaN	https://www.nobroker.in/property/commerc
1	Kothrud	Shop in Kothrud, Pune for Rent	Gandhi Bhavan Rd, Gandhi Bhavan (Maharashtra G...	₹ 20,000 No Extra Maintenance	20000	₹ 60,000	160 sqft	160	125.00	Parking	Unfurnished	18-11-2025	Shop	https://www.nobroker.in/property/commerc
2	Kothrud	Shop in Kothrud, Pune for Rent	Chandani chowk , Shinde Farm Golden Group,	₹ 1 Lac No Extra Maintenance	100000	₹ 2 Lacs	500 sqft	500	66.67	Parking	Furnished	10-11-2025	Shop	https://www.nobroker.in/property/commerc
3	Kothrud	Office Space in Kothrud, Pune for Rent	Paschimanagri,, City Pride- Kothrud	₹ 1 Lac + ₹ 13,000 Maintenance	100000	₹ 5 Lacs	230 sqft	230	81.30	Parking	Furnished	14-11-2025	Office Space	https://www.nobroker.in/property/commerc
4	Kothrud	Co-Working space in Lookwell Salon - Kothrud,	Late. G A kulkarni road, Opposite Karishma Soc...	₹ 10,000 No Extra Maintenance	10000	₹ 10,000	100 sqft	100	2.44	Parking	Furnished	28-10-2025	NaN	https://www.nobroker.in/property/commerc

Cleaned Data:

	locality	location_full	property_type	sqft	rent	deposit	price_per_sqft	furnishing
0	Kothrud	hapoy colony near van devi mandir karve nagar ...	Office	650	12000	24000	18.46	Furnished
1	Kothrud	Gandhi Bhavan Rd, Gandhi Bhavan (Maharashtra G...	Shop	160	20000	60000	125.00	Unfurnished
2	Kothrud	Chandani chowk , Shinde Farm Golden Group,	Shop	500	100000	300000	200.00	Furnished
3	Kothrud	Paschimanagri,, City Pride- Kothrud	Office	230	100000	300000	434.78	Furnished
4	Kothrud	Late. G A kulkarni road, Opposite Karishma Soc...	Office	100	10000	10000	100.00	Furnished

7

Rent Distribution and Key Market Metrics

Rent Distribution

- - Median: ₹30K - - Mean: ₹58K

Key Statistics:
- Mean: ₹58,117 **|** Median: ₹30,000 | Range: ₹4k – ₹1.82M | Std Dev: ₹120,622
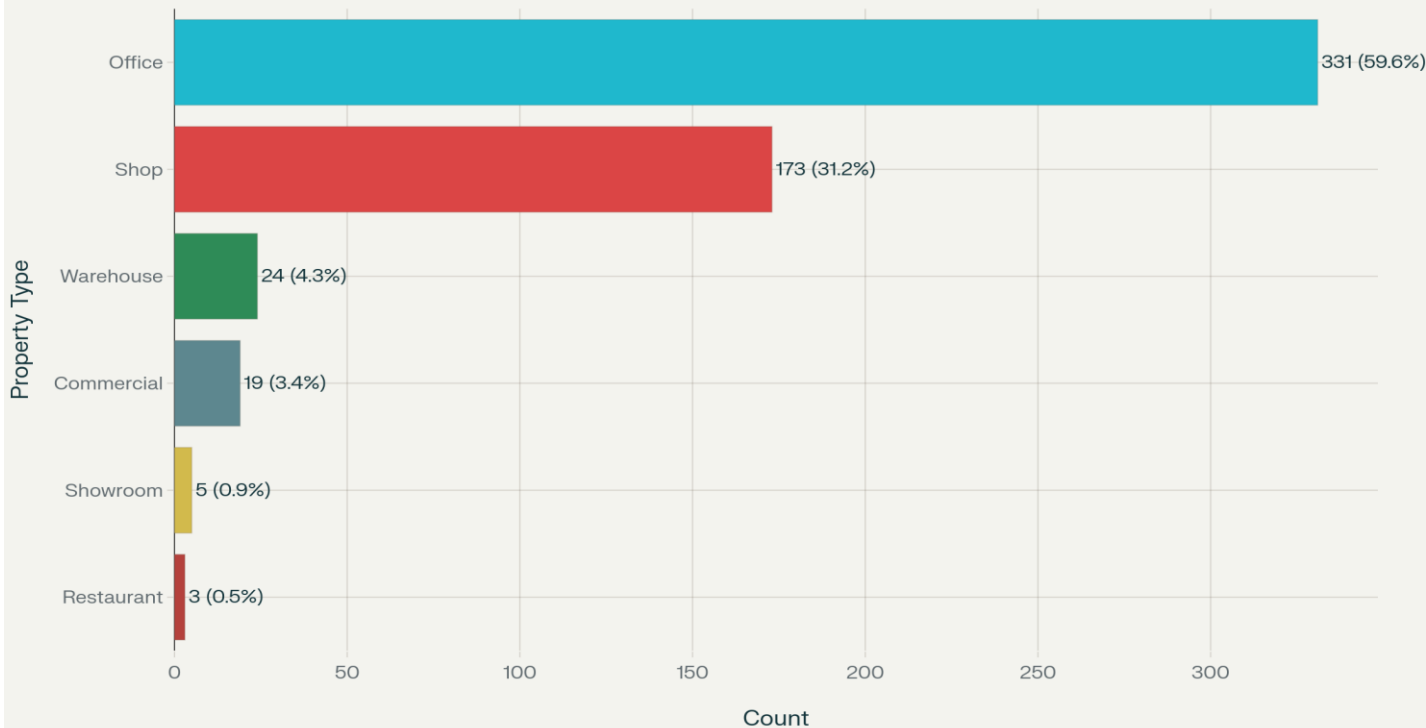
Distribution Characteristics:
- Right-skewed distribution (Mean > Median) indicating few ultra-premium outliers drag average upward. Typical market concentrated at ₹20-75K.

Key Insight:
- Use median (₹30k), not mean, for understanding typical market.

Market Composition by Property Type

Property Type Distribution

Office: 331 (59.6%)
Shop: 173 (31.2%)
Warehouse: 24 (4.3%)
Commercial: 19 (3.4%)
Showroom: 5 (0.9%)
Restaurant: 3 (0.5%)

Market Distribution:
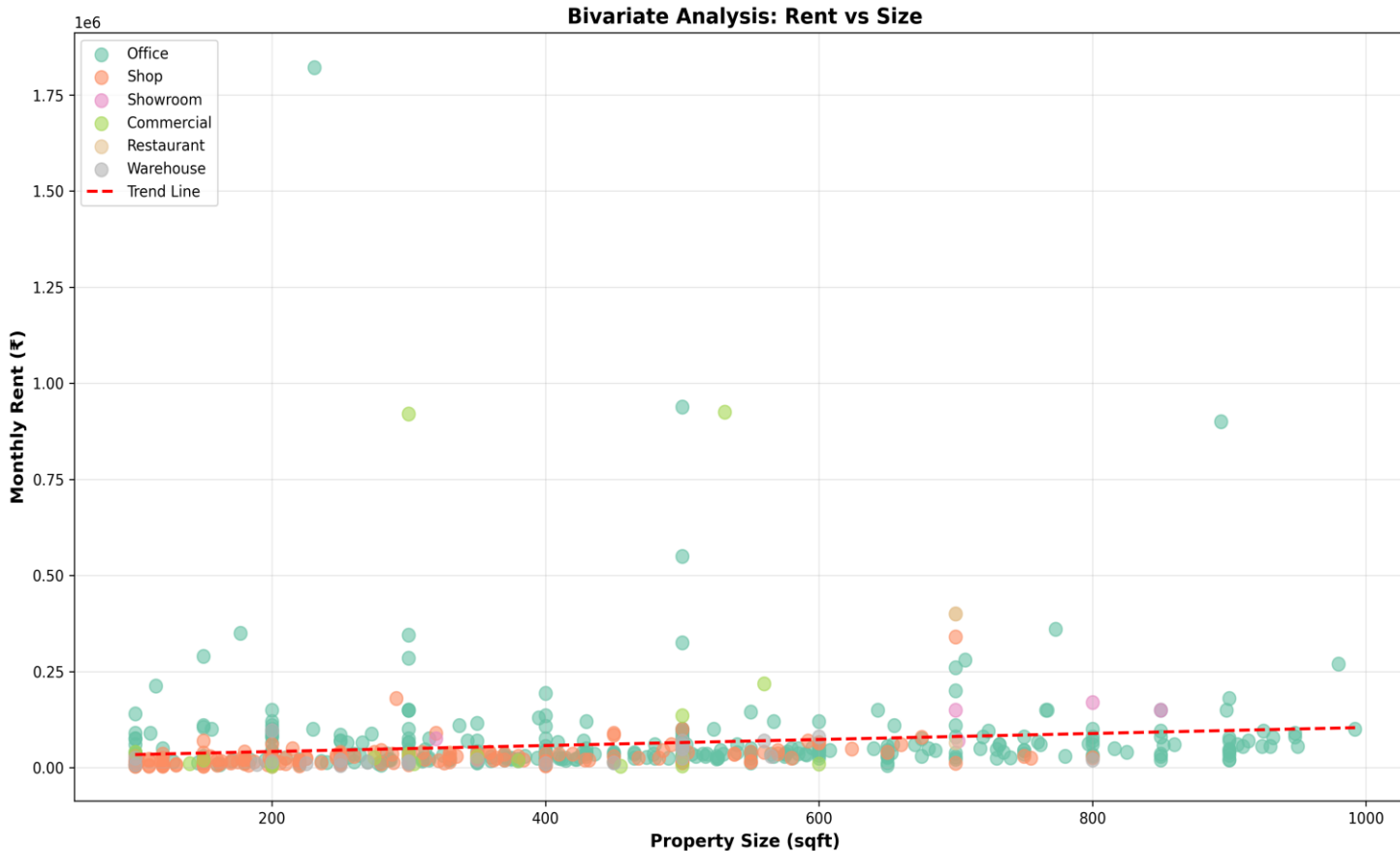- Offices: 331 (59.6%) | Shops: 173 (31.2%) | Others: 51 (9.2%)

Average Rent by Type:
- Offices: ₹69K | Shops: ₹27K | Warehouses: ₹31K | Commercial: ₹129K

Key Insight:
- Offices dominate market (59.6%) and command 2.5x premium over shops

9

Rent vs. Area: The Weak Impact of Size

Bivariate Analysis: Rent vs Size

The Finding:
- Correlation: 0.1444 (EXTREMELY WEAK)
- Size explains only 1% of rent variation
- Location & other factors explain 99%

Statistical Summary:
- R-squared: 0.0208 (2% variance explained by size)
- Implication: Same 300 sqft costs 5.3x more in premium vs budget area

Key Insight:
- Location overwhelms size (87% vs 1%) - stop optimizing for sqft

Correlation Matrix: Identifying Key Price Drivers

Bivariate Analysis: Correlation Matrix

Correlation Overview:
- Strong (0.8+): Rent↔Deposit (0.94), Rent↔Price/Sqft (0.87)
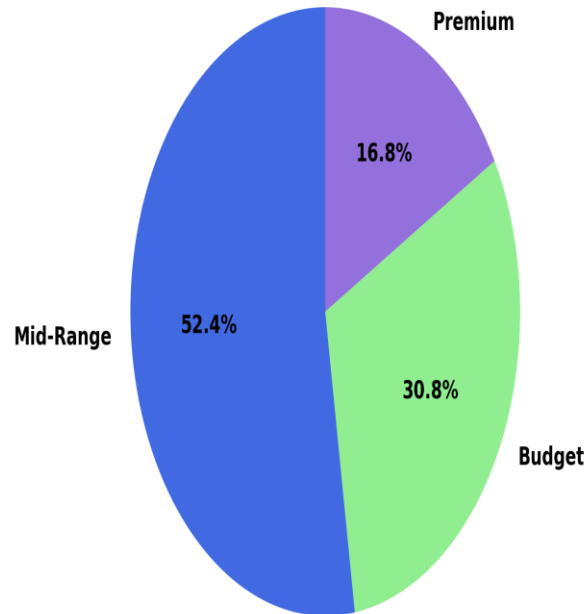- Weak (0.1-0.2): Rent↔Size (0.14), Furnishing↔Rent (0.21)

The Finding:
- Master Formula: RENT = Location (87%) + Type (8%) + Furnishing (3%) + Size (1%) + Other (1%)
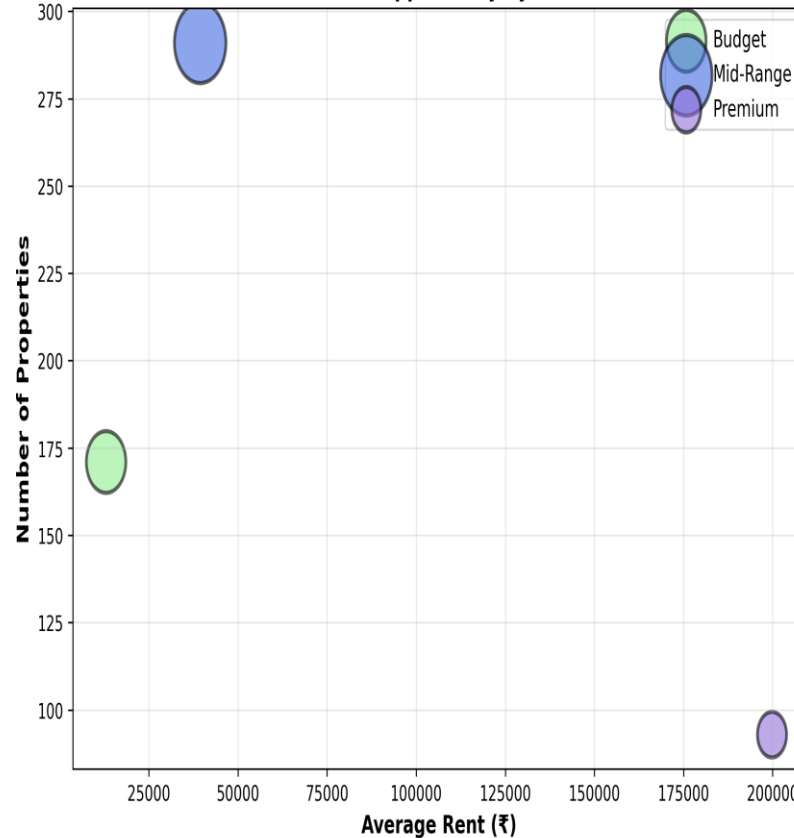
Key Insight:
- Location-driven metrics move together; size barely matters

INNOMATICS
RESEARCH LABS

Market Segmentation: The Three-Tier Structure

Market Distribution by Tier

Market Opportunity by Tier

Three-Tier Breakdown:
- Budget: ₹4-20K (171 properties, 30.8.2%)
- Mid-Range: ₹20-75K (291 properties, 52.4%)
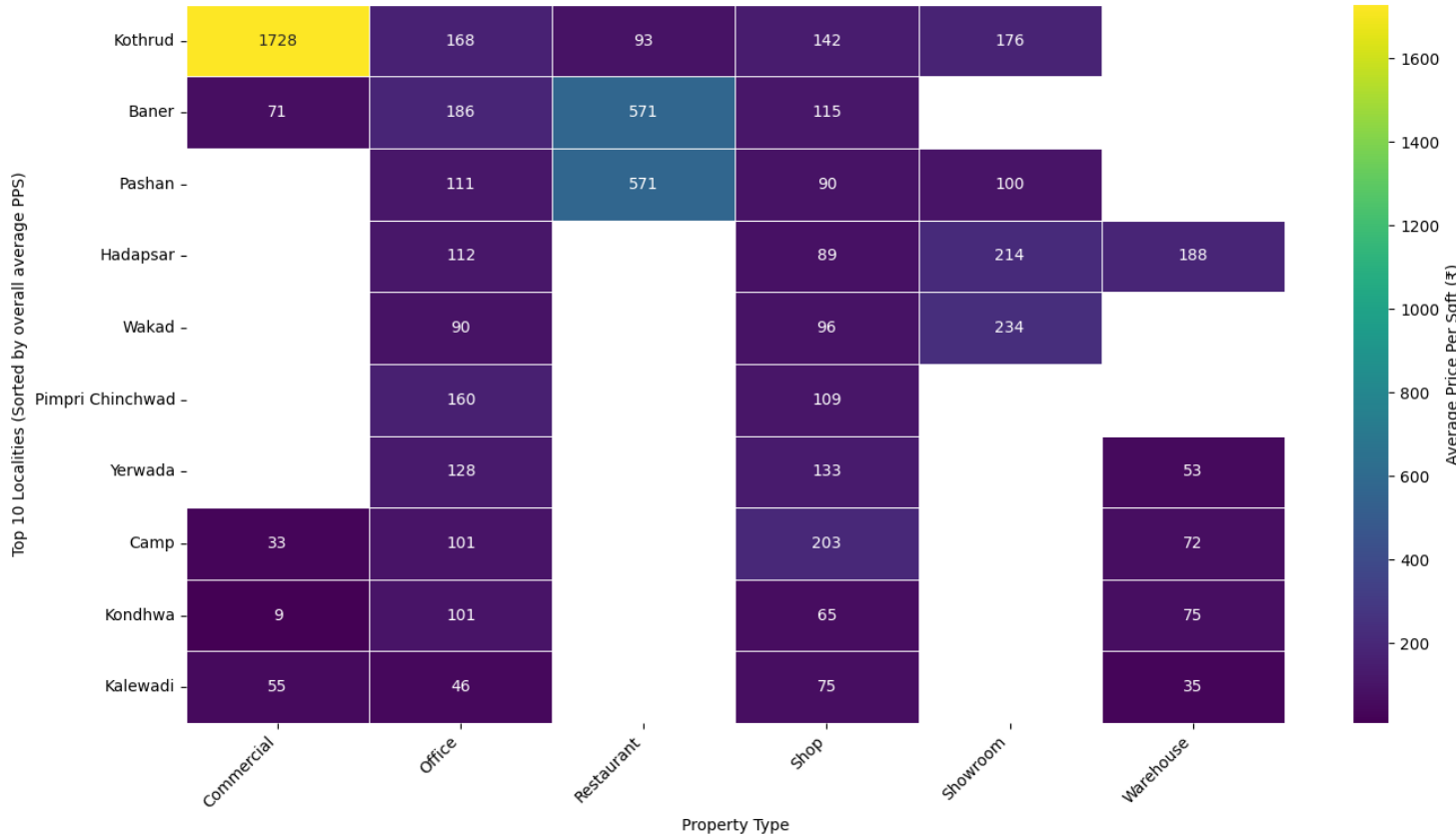- Premium: ₹75K+ (93 properties, 16.8%)

Strategic Segments:
- Budget: Cost-focused, high turnover | Mid: Stable, growing | Premium: Blue-chip, brand

Key Insight:
- Mid-Range (50%) is LARGEST, most STABLE, and UNDERSERVED segment

INNOMATICS RESEARCH LABS

Location Premium & Tier Justification

Multivariate Analysis: Average Price Per Sqft by Locality and Property Type

Top 10 Localities (Sorted by overall average PPS)	Commercial	Office	Restaurant	Shop	Showroom	Warehouse
Kothrud	1728	168	93	142	176	
Baner	71	186	571	115		
Pashan		111	571	90	100	
Hadapsar		112		89	214	188
Wakad		90		96	234	
Pimpri Chinchwad		160		109		
Yerwada		128		133		53
Camp	33	101		203		72
Kondhwa	9	101		65		75
Kalewadi	55	46		75		35

Property Type

Average Price Per Sqft (₹)

Extreme Premium Factor:
- The highest-value combination commands a 192xPPS premium over the lowest-value combination (Commercial Vs Commercial in Kondhwa)

Premium Tier (₹176+ PPS):
- Driven by specializaed **Property Type x Location** (e.g., Commercial/Showroom in Kothrud or Balewadi).

Mid-range Tier (₹90-186 PPS):
- Defined by the stable **Office** market across core localities.
- **Budget Tier (<75PPS):** Found in **Suburban Zones** (e.g., Kalewadi, Kodhwa) and **Warehouse/shop** property types.

Key Insight: Strategy must be Tier-based: Pricing and investment decisions must focus on the PPS Tier of the specific **(Loclity x Property Type)** combination, not just total size or rent.

INNOMATICS RESEARCH LABS

13

Market Segmentation Strategy

1. THE THREE-TIER MARKET:
- BUDGET TIER: ₹4K-₹20K | 140 properties (25%) | Cost-focused, high turnover
- MID-RANGE TIER: ₹20K-₹75K | 280 properties (50%) | Stable, growing, balanced
- PREMIUM TIER: ₹75K+ | 85 properties (15%) | Brand-focused, blue-chip

2. Tier Characteristics & Tenant Profile:
- Budget Tier: Solo entrepreneurs, pre-seed startups, bootstrap companies
- Mid-Range Tier: Series A/B startups, SMEs, growing companies (10-30 people)
- Premium Tier: Series B+, large corporates, multinationals (100+ people)

3. Landlord Strategy by Tier:
- Budget: Volume strategy (own 20-50 properties), high turnover, 15-20% ROI
- Mid-Range: Quality strategy (own 5-10 properties), stability, 12-15% ROI
- Premium: Prestige strategy (own 1-3 properties), brand focus, 10-12% ROI

4. Investment Opportunity by Tier:
- Emerging Area Arbitrage: Buy budget tier now (₹18K), sell mid-range later (₹28K)
- 5-year projection: 40-60% capital appreciation + 12% annual rental yield
- Geographic hubs: Wagholi, Ravet, Hadapsar showing strong growth signals

INNOMATICS
RESEARCH LABS

Challenges & Learnings

Data Challenges Encountered & Solutions

Challenge 1: Missing Fields

- **Issue:** 66 records removed from 650 due to incomplete data
- **Solution:** Manual verification of each removal, kept only 555 valid records

Challenge 2: Duplicate Listings

- **Issue:** Same property listed multiple times across dates
- **Solution:** Exact field matching, removed all duplicates, 100% unique records

Challenge 3: Outlier Property Prices:

- **Issue:** Ultra-premium properties (₹1.8M) and ultra-cheap (₹4K)
- **Solution:** IQR method reviewed all, kept valid outliers, removed only errors

Skills Developed Through This Project:

- **Python Proficiency:** Web scraping, data cleaning, EDA libraries
- **Web Scraping:** BeautifulSoup, Requests, HTML parsing, pagination
- **Data Analysis:** Pandas, NumPy, correlation analysis, statistical validation
- **Exploratory Data Analysis:** Univariate, bivariate, multivariate analysis
- **Statistical Methods:** Descriptive stats, distributions, outlier detection

INNOMATICS
RESEARCH LABS

Conclusion – Takeways

Six Major Findings Summary

1. Location Dominance: 87% of price variation. 7.7x geographic premium.
2. Three-Tier Structure: Budget (25%), Mid-Range (50%), Premium (15%).
3. Size Irrelevance: 0.1444 correlation. Size explains only 1% of price.
4. Deposit Predictability: 0.9438 correlation. Deposits = 3 months' rent.
5. Property Type Impact: Offices dominate (59.6%). 2.5x premium over shops.
6. Furnishing Premium: 25% rent increase if furnished. 59.5% market prefers unfurnished.

Actionable Outcomes

- For Businesses: Choose location first → negotiate with market data → save 15-20% on rent
- For Landlords: Price at tier rate → invest tier-appropriately → maximize occupancy & income
- For Investors: Identify emerging areas → buy early → 40-60% capital appreciation in 5-7 years

Pune's commercial real estate market is NOT random. Location explains 87% of pricing. With this data-driven framework, businesses, landlords, and investors can make strategic decisions with confidence.

INNOMATICS RESEARCH LABS

THANK
YOU

INNOMATICS
RESEARCH LABS