A COMPREHENSIVE REPORT ON THE HOUSING SECTOR OF DELHI

**Executive Summary**

This project explores the rental housing market in Delhi using real-world property data. The dataset includes details such as house type, size, location, price, furnishing status, and amenities, along with geographic coordinates and verification dates. After cleaning the data to remove errors, standardize formats, and calculate useful metrics like price per square foot, we analyzed it in Jupyter Notebook to uncover key patterns and trends.

Our findings show clear differences in rental prices across Delhi’s neighbourhoods, with certain areas commanding a significant premium. Factors such as house size, number of bathrooms, furnishing status, and location strongly influence rental values. We also built search queries that make it easy to filter properties based on specific needs — for example, affordable semi-furnished apartments in South Delhi.

The results offer practical value for tenants looking for the best deals, landlords aiming to maximize rental income, and even policymakers planning housing strategies. This study combines data analysis with real-world applications, providing a clear picture of how Delhi’s housing market operates today.

**Introduction**

**Background**  
Delhi’s housing sector plays a crucial role in the city’s economy and quality of life. As India’s capital and a major commercial hub, Delhi attracts a constant inflow of migrants — including students, professionals, and business owners — who rely heavily on the rental housing market. Rapid urbanization, diverse income groups, and varying lifestyle preferences have created a highly segmented market, ranging from affordable apartments in peripheral zones to luxury villas in prime neighborhoods. The housing sector not only reflects the city’s socio-economic diversity but also responds dynamically to factors such as infrastructure development, employment opportunities, and population growth.

**Objective**  
The primary aim of this study is to analyze the rental housing market in Delhi using a data-driven approach, uncovering the key factors that influence rental prices and availability. The research focuses on addressing the following questions:

* Which areas in Delhi command the highest and lowest rental prices?
* How does the size of a property (in square feet) influence its rental value?
* What is the relationship between amenities — such as number of bathrooms, balconies, and furnishing status — and rental prices?
* How do geographic location and neighborhood characteristics impact price per square foot?  
  By answering these questions, the analysis seeks to provide actionable insights for tenants, landlords, investors, and policymakers.

**Scope**  
This study focuses exclusively on rental housing within the boundaries of Delhi. The dataset includes property attributes such as house type, size, location, price, furnishing status, amenities, and geographic coordinates, along with calculated metrics like price per square foot. The data represents a snapshot from the collection period and does not account for long-term historical trends or seasonal price variations. While the study covers a diverse set of property types and locations, it does not incorporate external economic variables such as interest rates or future infrastructure projects. As such, the findings are most relevant for understanding current market conditions and identifying near-term opportunities in Delhi’s rental housing sector.

**Data Description**

* **Source of data- Open source, Kaggle**
* **Column descriptions**

| **Column Name** | **Description** |
| --- | --- |
| **House Type** | Category of the property (e.g., apartment, villa, duplex). |
| **House Size** | Total area of the property in square feet (or square meters). |
| **Location** | Specific area or neighborhood where the property is situated. |
| **City** | City name (in this dataset, all entries are for Delhi). |
| **Latitude** | Geographic latitude coordinate of the property. |
| **Longitude** | Geographic longitude coordinate of the property. |
| **Whole Price** | The total buying price of the property (updated from the original dataset). |
| **Monthly Rent** | The monthly rental value for the property (renamed from the original “Price” column). |
| **Currency** | The currency in which the price is listed (e.g., INR – Indian Rupees). |
| **Number of Bathrooms** | Total number of bathrooms in the property. |
| **Number of Balconies** | Total number of balconies in the property. |
| **Negotiability** | Indicates whether the rental price is negotiable (Yes/No). |
| **Price per Square Foot** *(added)* | Whole Price divided by the property’s total size. |
| **Price per Room** *(added)* | Whole Price divided by the number of rooms (derived from dataset or assumption). |
| **Verification Date** | Date when the property listing was verified. |
| **Description** | Additional narrative or remarks about the property. |
| **Security Deposit** | Security deposit amount required for renting the property. |
| **Status** | Furnishing condition — furnished, unfurnished, or semi-furnished. |
| **Category** *(added)* | Classification based on price range (e.g., apartment, villa, studio ). |

**Data Cleaning & Processing**

The dataset underwent multiple pre-processing steps to ensure accuracy, consistency, and readiness for analysis:

* **Handling Missing Values:**  
  All null values were identified and removed to avoid skewing results. This ensured that subsequent calculations and visualizations were based on complete records.
* **Data Standardization:**
  + Price-related columns were clarified and restructured:
    - **Price** now represents the full property purchase price.
    - **Monthly Rent** reflects the monthly rental amount.
  + Units for property size were standardized to square feet.
  + Categorical fields (e.g., Negotiability, Furnishing Status) were cleaned for consistent formatting.
* **Duplicate and Outlier Removal:**  
  Any duplicate entries were removed. Obvious outliers (e.g., unrealistically high or low prices) were flagged for review and excluded if confirmed erroneous.
* **New Derived Columns:**  
  Using the “eval” method in Pandas, the following computed fields were created:
  + **Price per Square Foot**: Price / House Size
  + **Price per Room**: Price / Number of Rooms (where applicable)
  + **Category**: Classification into Budget, Mid-range, or Luxury based on price thresholds.
* **Visualization Preparation:**  
  The cleaned and enriched dataset was used with **Plotly** to generate interactive charts for exploring relationships between location, amenities, and pricing.

**Exploratory Data Analysis Summary – Delhi Rental Market**

1. **Price Distribution**
   * The distribution of rental prices across Delhi is roughly centered around the average, with a moderate spread.
   * Most properties fall within the mid-range rent segment, but there are a few high-priced outliers in luxury areas.
2. **Average Rent by Locality**
   * Rental prices vary significantly between localities.
   * Premium areas (e.g., South Delhi, Vasant Vihar ,etc.) command substantially higher rents compared to outer or less developed zones.
3. **House Size vs Price**
   * There is a clear **positive correlation** between property size (square feet) and rental price — larger houses tend to have higher rents.
4. **Furnishing Status**
   * Fully furnished properties tend to rent at a premium compared to semi-furnished or unfurnished units.
   * This premium reflects both added convenience for tenants and higher upkeep/amenities.
5. **Bathrooms & Balconies**
   * More bathrooms generally correspond to higher rents, as they often indicate larger, more premium homes.
   * Additional balconies also tend to slightly increase rental value, but the effect is smaller than bathrooms.
6. **Price per Square Foot by Location**
   * Price per sqft differs significantly by locality.
   * Premium neighborhoods have the highest per-sqft rates, even for smaller units, while peripheral areas have much lower rates.
7. **Anomalies**
   * A few properties in non-premium areas show abnormally high prices — possibly due to listing errors, unique property features, or luxury developments in otherwise low-priced zones.

**Use Case: Search Queries on Housing Data**

After performing the exploratory data analysis (EDA), I implemented a search functionality using **Pandas filtering** to simulate real-world decision-making scenarios.  
This allows a user, investor, or renter to apply specific constraints and instantly see relevant housing options.

For example:

1. **Family scenario** –  
   *Query:* *Find all 2BHK apartments in Pune, semi-furnished, with at least 2 bathrooms, within a budget of ₹25,000 per month.*
   * Using Pandas conditional filtering (df[(df['BHK'] == 2) & (df['Furnishing'] == 'Semi-Furnished') & (df['Bathrooms'] >= 2) & (df['Rent'] <= 25000)]), I could quickly retrieve matching properties.
   * This helps families directly shortlist homes without manually browsing irrelevant listings.
2. **Struggling artist scenario** –  
   *Query:* *Locate studio apartments in a central location under ₹8,000, unfurnished, suitable for a single occupant.*
   * Applied Pandas filtering on type, location, furnishing, and rent columns to find minimal-cost options in desired localities.
   * This is ideal for individuals with strict budget constraints but location preferences.

This functionality transforms raw data into **actionable insights**, enabling tailored property searches based on needs, budgets, and amenities — similar to how a real estate website’s search filters work, but entirely built in Python.

**Recommendations**

**For Tenants:**

* Focus on **emerging neighborhoods** such as Dwarka, Rohini, and Indirapuram for affordable rents while still accessing good connectivity.
* In premium areas like South Delhi, look for slightly older properties or semi-furnished units to balance amenities and cost.
* Consider areas with lower price per sq. ft. but good transport links to maximize value.

**For Landlords/Investors:**

* Premium localities like Greater Kailash, Vasant Vihar, and Hauz Khas show **higher price per sq. ft.** and faster appreciation — ideal for short-term rental yield.
* In upcoming metro-connected zones, invest early to capture future rental growth as infrastructure develops.
* Furnished or semi-furnished units tend to command **higher rents**, especially in professional hubs.

**For Policymakers:**

* Target mid-tier areas with high demand but insufficient supply for affordable housing schemes.
* Introduce or strengthen rent regulation frameworks to ensure stability for tenants while protecting landlord interests.
* Support mixed-income housing projects in well-connected zones to reduce market pressure on premium localities.

**Result and feedback**

**Positive Feedback:**

* The **interactive visualizations** made it easy to explore trends without prior technical knowledge.
* The **search query use cases** (e.g., family with budget constraints, artist looking for studio) were relatable and helped non-technical viewers understand the dataset’s potential applications.
* Clear correlations, such as the link between furnishing level and price or location and price per sq. ft., were appreciated as “practically useful insights.”

**Constructive Feedback:**

* Some people suggested adding **more context on each neighborhood** (e.g., safety, amenities, schools) to make the analysis more relevant for actual renters/buyers.A more detailed dataset would be appreciated.
* A few felt the **price distribution analysis** could be broken down further by property type (apartment vs. villa) to uncover deeper patterns.
* The **search function** could be made more user-friendly if implemented in a simple app interface rather than requiring Python/Pandas knowledge.