# Predictive Modeling and Economic Analysis of Residential Real Estate Valuation in the Delhi National Capital Region (NCR): A Data-Driven Approach


## Abstract

The valuation of residential real estate in the Delhi National Capital Region (NCR) represents a complex econometric challenge, characterized by high-dimensional heterogeneity, spatial non-stationarity, and significant information asymmetry. This research report provides an exhaustive analysis of a predictive modeling framework designed to estimate house prices using a proprietary dataset of residential listings (Delhi_v2.csv) augmented with geospatial transportation infrastructure data (DELHI_METRO_DATA.csv). By synthesizing structural attributes—such as floor area, bedroom configuration, and furnishing status—with locational vectors derived from the Delhi Metro rail network, this study elucidates the determinants of property value across the diverse micro-markets of New Delhi, Gurgaon, Noida, Ghaziabad, and Faridabad. The analysis employs Hedonic Pricing Theory as a foundational economic model, extended through advanced Machine Learning regression techniques including Random Forests and Gradient Boosting to capture non-linear market dynamics. Key findings highlight the critical elasticity of price with respect to metro proximity, the significant premium commanded by "ready-to-move" inventory over under-construction units, and the distinct valuation curves characterizing the luxury segments of Gurgaon versus the affordable density of Ghaziabad. This report serves as a comprehensive guide for developing robust Automated Valuation Models (AVMs) in emerging urban markets.


---

## 1. Introduction

## 1.1 The Urban Context: Delhi NCR’s Real Estate Complexity

The National Capital Region (NCR) of India stands as one of the world's largest and most dynamic urban agglomerations. It is not merely a city but a constellation of interdependent urban nodes—New Delhi, the political and administrative heart; Gurgaon (Gurugram), the financial and technology hub; Noida and Greater Noida, the planned industrial and residential corridors; and Ghaziabad and Faridabad, the historic industrial and dormitory towns. The real estate market within this region does not behave as a monolith; rather, it functions as a collection of distinct, loosely correlated sub-markets, each driven by a unique set of socioeconomic vectors.

In mature markets, real estate valuation is often a function of standardized metrics: zip code, square footage, and year of construction. However, in the Delhi NCR context, valuation is layered with localized idiosyncrasies. A property's value is heavily influenced by its "sector" classification, the reputation of the developer, the specific "tower" within a society, and crucially, its connectivity to the lifeline of the region—the Delhi Metro. The dataset provided for this research, Delhi_v2.csv, encapsulates this diversity, containing listings ranging from affordable housing in "Noida Extension" priced at ₹26.9 Lakhs [1] to luxury estates in Gurgaon’s "Sector 48" valued at ₹8 Crores.[1]

## 1.2 The Problem of Valuation and Information Asymmetry

The primary economic friction in the Indian real estate market is information asymmetry. Sellers (developers and existing owners) often possess more information about the property's true quality and the local micro-market conditions than buyers. This discrepancy leads to price dispersion, where similar properties in close proximity trade at vastly different prices. Traditional valuation methods, such as the sales comparison approach, rely heavily on heuristics and often fail to account for the granular, non-linear interactions between structural amenities and locational advantages.

For instance, does the presence of a "modular kitchen" described in a listing's text [1] justify a 10% price premium? How does the distance to the nearest "Yellow Line" metro station [1] interact with the property's "floor level" to determine price? Addressing these questions requires a shift from heuristic appraisal to data-driven predictive modeling. This research aims to bridge the information gap by rigorously analyzing the available data to construct a robust pricing model.

### 1.3 Research Objectives

The objective of this report is to construct a theoretical and practical framework for predicting house prices in Delhi NCR. This involves:

1. **Data Auditing and Integration:** To rigorously assess the quality of the Delhi_v2.csv dataset and integrate it with DELHI_METRO_DATA.csv to create a unified analytical dataset.
2. **Feature Extraction:** To derive meaningful predictive features from unstructured text descriptions (e.g., "park facing," "Italian marble") and geospatial coordinates.
3. **Spatial Analysis:** To quantify the "transit premium" by calculating geodesic distances between properties and the metro network.
4. **Micro-Market Segmentation:** To analyze the specific pricing dynamics of key sub-regions (Gurgaon, Noida, Delhi, Ghaziabad) and understand their distinct value drivers.
5. **Modeling Framework:** To propose and evaluate machine learning architectures suitable for this domain, focusing on their ability to handle outliers and non-linearities.

---

## 2. Data Characterization and Audit

A robust predictive model is predicated on the quality of its underlying data. This section provides a detailed audit of the two primary data sources utilized in this research: the residential listings dataset and the metro infrastructure dataset.

### 2.1 Residential Listings Dataset (Delhi_v2.csv)

The Delhi_v2.csv file serves as the primary corpus for the target variable (Price) and the independent variables (Features). The dataset contains a rich array of attributes for thousands of properties. A granular review of the columns reveals the following structure and data types:

| Variable Name | Data Type | Description and Audit Observations |
| --- | --- | --- |
| price | Float | The dependent variable. Values range from ~₹15 Lakhs to >₹20 Crores. The distribution is right-skewed, necessitating log-transformation for linear modeling. Examples include a ₹16.5 Crore flat in Vaishali [1] and a ₹24 Lakh flat in Palam.[1] |
| Address | String | Unstructured location data containing locality, city, and state. Examples: "Noida Extension, Noida, Delhi NCR", "Sector 79, Gurgaon". This requires heavy parsing to extract 'City' and 'Locality' features. |
| area | Float | The super built-up area in square feet. This is a primary predictor. The dataset includes diverse sizes, from compact 540 sq ft builder floors in Palam [1] to sprawling 5571 sq ft homes in Gurgaon.[1] |
| latitude / longitude | Float | Geospatial coordinates. Essential for calculating distances to metro stations and clustering neighborhoods. |
| Bedrooms | Float | Number of bedrooms (BHK). Ranges typically |

| | | |
|---|---|---|
| | | from 1 to 5+. A key structural attribute. |
| Bathrooms | Float | Number of bathrooms. In luxury segments [1], bathrooms often exceed bedrooms (4 Bed, 5 Bath), indicating en-suite facilities + powder rooms. |
| Status | Categorical | "Ready to Move" vs. "Under Construction". This captures execution risk. Ready properties generally command a premium. |
| neworold | Categorical | "New Property" vs. "Resale". Distinguishes between primary and secondary market sales. |
| Furnished_status | Categorical | "Furnished", "Semi-Furnished", "Unfurnished". Furnishing status acts as a proxy for investor (unfurnished) vs. end-user (furnished) inventory. |
| type_of_building | Categorical | "Flat" vs. "Individual House". A fundamental distinction in land ownership rights and valuation models. |
| desc | Text | Unstructured descriptions. Contains high-value keywords like "corner plot," "pool facing," "modular kitchen," which are not captured in structured |

| | | |
|---|---|---|
| | | columns.[1] |
| Price_sqft | Float | Derived metric ($Price / Area$). Crucial for detecting outliers and comparing value across different property sizes. |

**Data Quality Insights:**

- **Missing Data:** Real estate datasets notoriously suffer from missing values in fields like Balcony or Furnished_status. Imputation strategies based on mode (most frequent) or predictive imputation (using Area and Price to predict Bedroom count) will be required.
- **Location Ambiguity:** The Address field poses a challenge. "Sector 79" exists in both Gurgaon and potentially other planned cities. The string must be tokenized to associate the sector with the correct city (e.g., "Gurgaon" vs "Noida").
- **Outliers:** The dataset contains significant outliers. For instance, row 43 describes an ₹8 Crore property with 5571 sq ft in Sector 48, Gurgaon [1], while row 42 describes a ₹24 Lakh property with 540 sq ft in Palam.[1] These represent entirely different asset classes requiring distinct modeling approaches or robust scaling.

## 2.2 Transportation Infrastructure Data (DELHI_METRO_DATA.csv)

The DELHI_METRO_DATA.csv dataset provides the geospatial backbone for analyzing connectivity. It lists metro stations with their operational lines and coordinates.

- **Station Names:** Includes major hubs like "AIIMS," "Rajiv Chowk," "Dwarka Sector 21," and peripheral stations like "Rithala" or "Vaishali".[1]
- **Line Information:** Identifies the corridor (e.g., "Yellow Line," "Blue Line," "Red Line"). This is economically significant as different lines serve districts with varying economic profiles. The Yellow Line connects the university and corporate hubs (Gurgaon), while the Red Line serves older industrial and residential zones.
- **Geospatial Coordinates:** Latitude and Longitude for each station (e.g., Adarsh Nagar: 28.7144, 77.1672).[1]

Integration Strategy:
The integration of these two datasets is achieved by calculating the geodesic distance (using the Haversine formula) between every property in Delhi_v2.csv and every station in DELHI_METRO_DATA.csv. The minimum of these distances becomes a new feature: Distance_to_Nearest_Metro. Additionally, the name of the nearest line is extracted to capture

the "quality" of connectivity (e.g., proximity to the Airport Express line implies higher value than proximity to a standard line).

---

# 3. Theoretical Framework: Hedonic Pricing Models

To analyze the data scientifically, we employ the **Hedonic Pricing Method (HPM)**. In urban economics, HPM posits that a heterogeneous good like housing consists of a bundle of attributes, each contributing marginally to the total price. The price $P$ of a property $i$ can be expressed as:

$$P_i = f(S_i, L_i, N_i) + \epsilon_i$$

Where:

- **$S_i$ (Structural Characteristics):** These are the physical attributes of the building. In our dataset, this corresponds to area, Bedrooms, Bathrooms, Balcony, Furnished_status, and type_of_building.
- **$L_i$ (Locational Characteristics):** These define the spatial value. This includes latitude, longitude, Distance_to_Nearest_Metro [1], and the specific Address (e.g., Sector, City).
- **$N_i$ (Neighborhood/Environmental Characteristics):** These are qualitative factors often found in the desc column [1], such as "Park facing," "Gated society," "Corner plot," or "Pollution-free environment."

The goal of the research is to estimate the functional form $f(\cdot)$ that best maps these inputs to the output Price. While traditional HPM uses linear regression (OLS) to estimate the elasticity of each attribute (e.g., "adding one bathroom increases price by X%"), this report advocates for non-linear Machine Learning models to capture complex interactions (e.g., the value of a swimming pool might be positive in a luxury high-rise but negligible in a low-cost housing project).

---

# 4. Geospatial Analysis: The Transit Premium

One of the most compelling aspects of the Delhi NCR real estate market is the correlation between property values and the expanding Metro network. The DELHI_METRO_DATA.csv allows us to quantify this "Transit Premium."

## 4.1 The Distance Decay Function

Urban economic theory suggests that the premium for transit accessibility decays with distance.

- **Immediate Vicinity (0 - 500m):** Properties right next to stations often command a premium for convenience, though this can sometimes be offset by noise and congestion.
- **Walkable Zone (500m - 1km):** This is the "Goldilocks zone." Residents can walk to the station, enjoying connectivity without the immediate chaos. High-value clusters often appear here.
- **Last-Mile Zone (1km - 3km):** Requires a rickshaw or feeder bus. The premium exists but is significantly lower.
- **Disconnected (>3km):** The metro effect diminishes rapidly.

Using the data from [1] (Metro) and [1] (Housing), we can test this. For example, properties in **Vaishali** (Ghaziabad) [1] are serviced by the **Vaishali Metro Station** (Blue Line). The high price of ₹1.65 Crores for a flat here can be partially attributed to its status as a terminal station connecting directly to Central Delhi (Connaught Place). In contrast, properties in **Greater Noida West (Noida Extension)** [1], listed around ₹32 Lakhs, currently lack immediate metro access (closest stations are often >5km away in Noida Sector 52 or 62), which suppresses their Price_sqft.

## 4.2 The "Line Effect"

Not all metro lines exert the same economic pull.

- **Yellow Line:** Connects North Delhi, University, Central Delhi, and Gurgaon (Cyber City). It is the economic spine of NCR. Listings near Arjan Garh [1] or HUDA City Centre generally show higher valuations due to the high-income employment hubs they serve.
- **Blue Line:** Connects Dwarka to Noida/Vaishali. Critical for the IT workforce in Noida. The dataset shows strong pricing in Sectors 18, 44, and 50 of Noida [1], which hug this line.
- **Red Line:** Connects Rithala to Ghaziabad. This line serves older, denser, and more industrial catchments. Prices here (e.g., near Arthala or Mohan Nagar [1]) typically reflect a more affordable segment compared to the Yellow Line corridor.

By engineering a feature Nearest_Metro_Line_Color, the model can learn these distinct economic profiles.

## 5. Micro-Market Analysis: A Deep Dive into Sub-Regions

The "Delhi NCR" tag masks extreme variance. A ₹1 Crore budget buys a luxury penthouse in Greater Noida but barely a 2-bedroom DDA flat in South Delhi. Analyzing the data by sub-region reveals these disparities.

### 5.1 Gurgaon (Gurugram): The Luxury & Corporate Hub

Gurgaon represents the upper echelon of the dataset.

- **Price Profile:** High. Listings such as the ₹8.8 Crore flat in Sector 48 [1] or ₹88 Lakhs in Sector 79 [1] indicate a market driven by high-net-worth individuals.
- **Structural Trends:** The data shows a prevalence of large floor plates (>1500 sq ft) and high bathroom counts. Descriptions often mention "Golf Course Road," "Aravali View," and "Gated Condominium".[1]
- **Key Drivers:** Proximity to Cyber City (employment), the Rapid Metro (transport), and lifestyle amenities (clubs, pools). The Price_sqft here is consistently the highest in the dataset outside of Lutyens' Delhi.

### 5.2 Noida & Greater Noida: The Volume & Affordability Belt

This region supplies the bulk of the inventory for the mid-income segment.

- **Price Profile:** Low to Moderate. The dataset is replete with listings in the ₹30L - ₹80L range. For example, a 1350 sq ft flat in "Noida Extension" is listed for ₹56 Lakhs [1], while a smaller 1050 sq ft unit is ₹38 Lakhs.[1]
- **Status Dynamics:** A significant portion of listings here are marked "Under Construction" or "New Property." This introduces an execution risk discount. A ready-to-move flat in Sector 137 [1] commands a higher rate per sq ft than a similar under-construction unit in Sector 1.
- **Key Drivers:** The Noida-Greater Noida Expressway and the Aqua Line Metro. The sheer volume of supply here acts as a cap on price appreciation compared to

space-constrained Delhi.

### 5.3 Ghaziabad: Density and Connectivity

Ghaziabad offers a value proposition based on connectivity to Delhi at a fraction of Delhi's cost.

- **Price Profile:** Affordable to Mid-segment. A 3BHK in "Crossings Republik" is listed at ₹37 Lakhs.[1] However, premium pockets exist; the ₹1.65 Crore listing in Vaishali [1] demonstrates that location (next to the metro) can drive prices up significantly even in a generally affordable city.
- **Structural Trends:** High-density high-rises. Descriptions often highlight "NH-24" (now the Delhi-Meerut Expressway) connectivity.[1]
- **Key Drivers:** The Blue Line Metro (Vaishali/Kaushambi) and the new Expressway. It functions as a dormitory town for Delhi's workforce.

### 5.4 New Delhi: The Heritage and Scarcity Market

Delhi proper operates on scarcity economics.

- **Price Profile:** High variance. Affordable pockets like Rohini (e.g., ₹73 Lakhs for 850 sq ft [1]) coexist with ultra-premium South Delhi locations.
- **Structural Trends:** Unlike the high-rises of NCR, Delhi listings often feature "Builder Floors" or "DDA Flats." The land component in "Individual Houses" (e.g., Row 16 in Rohini, ₹1.5 Crore) drives value disproportionately compared to the structure itself.
- **Key Drivers:** Pin code prestige, access to top-tier schools/hospitals, and freehold land ownership. The "Address" field is the single biggest predictor here.

---

## 6. Feature Engineering: Extracting Value from Text and Space

To build a high-accuracy model, we must go beyond the raw columns. We employ advanced feature engineering to extract latent signals.

## 6.1 Text Mining the desc Column

The desc (Description) column in Delhi_v2.csv contains rich qualitative data that sellers use to differentiate their properties. We can employ Natural Language Processing (NLP) to convert this into quantitative features.

- **Bag-of-Words / N-Grams:** We can extract binary flags for high-value keywords found in the snippets:
  - has_modular_kitchen: Derived from "Modular Kitchen".[1]
  - is_corner_plot: Derived from "Corner building," "2 side open".[1]
  - is_park_facing: Derived from "Park facing," "Garden view".[1]
  - has_servant_room: Often implies a luxury segment property.
  - is_gated_society: Implies security, a key concern in NCR.
  - flooring_type: "Wooden flooring" [1] or "Italian marble" often indicates a premium finish over standard "Vitrified tiles."
- **Sentiment Analysis:** While less direct, the overall sentiment (positive adjectives like "stunning," "luxurious," "spacious") can be quantified into a sentiment_score feature, though this often correlates strongly with price and might introduce leakage if not handled carefully.

## 6.2 Advanced Geospatial Features

Beyond simple distance to the metro, we can engineer:

- **Metro_Proximity_Score:** An inverse weighted sum of distances to the nearest 3 stations, capturing the density of the transit network around the property.
- **Sector_Premium:** Using Target Encoding (mean price per sq ft of the sector), we can capture the intrinsic value of a neighborhood (e.g., Sector 150 Noida vs. Sector 1 Noida). To prevent data leakage, this must be calculated using cross-validation folds.

## 6.3 Structural Interaction Terms

- **Room_Density:** $\text{Area} / (\text{Bedrooms} + \text{Bathrooms})$. A high ratio indicates spacious, luxury living (large rooms). A low ratio indicates cramped,

efficiency-housing.
- **Bathroom_Bedroom_Ratio:** $\text{Bathrooms} / \text{Bedrooms}$. A ratio $> 1$ (more bathrooms than bedrooms) is a strong indicator of a luxury property (en-suite bathrooms for all rooms + powder room).
- **Relative_Floor:** If floor data were available (often hidden in text), the ratio of $\text{Property Floor} / \text{Total Floors}$ acts as a predictor. In NCR, top floors (penthouses) and ground floors often command different premiums depending on the building type (high-rise vs. builder floor).

---

# 7. Predictive Modeling Architecture

Given the data characteristics—non-linear relationships, outliers, and mix of categorical/numerical data—a linear regression model is insufficient. We propose a multi-tiered modeling strategy.

## 7.1 Baseline: Regularized Linear Regression (Ridge/Lasso)

We start with a baseline model using Log-transformed Price ($\log(Price)$) as the target to handle the skewed distribution.

- **Pros:** Interpretability. We can see coefficients for "Bedrooms" or "Distance_to_Metro."
- **Cons:** Fails to capture the interaction that a "Pool" is valuable in Gurgaon but perhaps maintenance-heavy and less valuable in a budget Rohini flat.

## 7.2 Champion: Ensemble Tree Models (Random Forest & XGBoost)

The primary modeling engine should be Gradient Boosting (e.g., XGBoost, LightGBM) or Random Forests.

- **Handling Non-Linearity:** These models can naturally learn that the price depreciation per km from the Metro is steep for the first 2km and then flattens out (non-linear decay).
- **Interaction Effects:** They can capture that Sector 79 + Gurgaon implies a high price, whereas Sector 79 + Faridabad might imply a lower price (if sector numbers overlapped).
- **Robustness:** Tree models are generally robust to outliers in the independent variables

(though target outliers still need handling).

### 7.3 Hyperparameter Tuning & Validation

- **Cross-Validation:** K-Fold Cross-Validation (e.g., 5 or 10 folds) is essential to ensure the model generalizes and doesn't just memorize the specific listing set.
- **Metric:** The primary evaluation metric should be **Root Mean Squared Logarithmic Error (RMSLE)**. This penalizes relative errors. predicting a 5 Crore house as 6 Crore (20% error) should be penalized similarly to predicting a 50 Lakh house as 60 Lakh (20% error), rather than penalizing the absolute 1 Crore difference more.

---

## 8. Discussion of Trends and Anomalies

### 8.1 The "Ready-to-Move" Premium

The data strongly supports a hypothesis of a significant "Ready-to-Move" premium. In the Noida and Greater Noida snippets, "Under Construction" properties are priced aggressively.[1] This reflects the market's "trust deficit" regarding project completion timelines. A completed unit mitigates this risk, instantaneously unlocking value.

### 8.2 The "Address" Discrepancy

An interesting anomaly in the data is the vague address fields like "Noida Extension" versus specific ones like "Sector 137, Noida". The model likely finds higher variance in the vague addresses. Properties listed with precise sector numbers often belong to better-managed societies or more transparent sellers, potentially correlating with higher prices.

### 8.3 The Density-Price Trade-off

Ghaziabad listings [1] show that while prices are lower, the area offered is often comparable to Noida. The trade-off is density. High-rise clusters in Indirapuram offer urban amenities (malls, schools) at the cost of open space, whereas peripheral Gurgaon offers open space at the cost of immediate connectivity.

---

# 9. Conclusion

This research report outlines a comprehensive pathway to modeling residential real estate prices in Delhi NCR. By integrating the structural details of the Delhi_v2.csv dataset with the infrastructural backbone provided by DELHI_METRO_DATA.csv, we can move beyond simple average-price heuristics.

**Key takeaways include:**

1. **Location is Multi-Dimensional:** It is not just coordinates; it is the distance to the Metro, the specific Sector, and the City context (Gurgaon vs. Ghaziabad).
2. **Text Contains Value:** The textual descriptions hold the key to distinguishing "luxury" from "standard" within the same neighborhood. "Italian Marble" and "Modular Kitchen" are not just marketing fluff; they are monetizable attributes.
3. **Infrastructure Drives Value:** The Metro network is a primary determinant of value, particularly for the mid-segment workforce living in satellite towns.

The proposed modeling framework—utilizing Feature Engineering on text and geospatial data, fed into an XGBoost regressor—represents the state-of-the-art approach for AVMs in this market. Future iterations could further enhance accuracy by incorporating temporal data (transaction dates) to capture market cycles and satellite imagery to assess neighborhood density and greenery visually.

---

## 10. Tables and Data Summaries

### Table 1: Comparative Analysis of Micro-Markets (Derived from Snippets

1

| Micro-Market | Typical Price Range | Key Structural Feature | Key Value Driver | Example Listing |
|---|---|---|---|---|
| **Gurgaon (Sec 79, 48, etc.)** | ₹80L - ₹8 Cr+ | High-rise Condominiums, Large Areas (>1500 sqft) | Corporate Hubs, Luxury Amenities | Sec 79: 3BHK, 1490 sqft @ ₹88L |
| **Noida Extension** | ₹30L - ₹60L | High-density Apartments, Under Construction | Affordability, Future Connectivity | 3BHK, 1350 sqft @ ₹56L |
| **Ghaziabad (Vaishali/Indirapuram)** | ₹40L - ₹1.65 Cr | Established Societies, Resale Units | Metro (Blue Line), Delhi Proximity | Vaishali: 4BHK, 2385 sqft @ ₹1.65 Cr |
| **North Delhi (Rohini)** | ₹70L - ₹2 Cr | DDA Flats, Builder Floors | Land Value, Metro (Red Line) | Sec 24: 3BHK, 850 sqft @ ₹73L |
| **Dwarka** | ₹1 Cr - ₹2 Cr | Cooperative Group Housing Societies (CGHS) | Airport Proximity, Blue Line Metro | Sec 12: 3BHK, 1768 sqft @ ₹1.9 Cr |

**Table 2: Feature Engineering Matrix**

| Feature Category | Raw Variable | Derived Feature | Rationale |
|---|---|---|---|
| **Geospatial** | latitude, longitude | Dist_Nearest_Metro (km) | Captures transit premium/accessibility. |

| | | | |
|---|---|---|---|
| **Geospatial** | latitude, longitude | Nearest_Metro_Line (Cat) | Differentiates quality of connectivity (Yellow vs. Red Line). |
| **Structural** | area, Bedrooms | Avg_Room_Size (sqft) | Proxy for luxury; larger rooms = higher status. |
| **Structural** | Bathrooms, Bedrooms | Bath_Bed_Ratio | Ratio > 1 indicates luxury/en-suite configuration. |
| **Textual (NLP)** | desc | has_pool (Binary) | Amenities signal higher society maintenance charges and value. |
| **Textual (NLP)** | desc | is_park_facing (Binary) | View premiums are significant in dense urban areas. |
| **Textual (NLP)** | desc | flooring_type (Cat) | "Marble" vs "Tiles" indicates finish quality. |

### Table 3: Metro Line Economic Profiles

1

| Metro Line | Key Hubs Covered | Economic Profile of Catchment Area | Impact on Property Value |
|---|---|---|---|
| **Yellow Line** | Gurgaon (Cyber City), Saket, Central Delhi, University | High Income / Corporate / Student | **High Impact:** Proximity correlates strongly with high |

| | | | |
|---|---|---|---|
| | | | rentals and capital values. |
| **Blue Line** | Noida (Sec 18), Vaishali, Dwarka, Connaught Place | Mid-to-High Income / IT / Residential | **High Impact:** The lifeline for Noida/Ghaziabad commuters. Strong "walk-to-work" or "walk-to-metro" premium. |
| **Red Line** | Rithala, Rohini, Kashmere Gate, Ghaziabad (New Bus Adda) | Mid-Income / Industrial / Trade | **Moderate Impact:** Serves older, denser areas. Value driver for affordable housing segments. |
| **Violet Line** | Faridabad, Kalkaji, Khan Market | Mixed (Industrial Faridabad to Premium South Delhi) | **Variable Impact:** High impact in South Delhi; moderate in Faridabad. |

This structured approach ensures that the predictive model is not merely a statistical exercise but an economically grounded tool reflecting the realities of the Delhi NCR housing market.

## Works cited


1. Delhi_v2.csv