**1. Business Context**

**1.1 Problem Statement**

Raw real estate data contains inconsistencies, unstructured text, currency formatting issues, and duplicate or missing information. Without a well-defined transformation process, the data cannot be effectively used for **property valuation, market trends, and buyer analytics**.

**1.2 Medallion Architecture Overview**

- **Bronze Layer (Raw Data):** Stores unprocessed property listing data from multiple sources (online portals, real estate agencies).

- **Silver Layer (Cleansed & Enriched Data):** Applies data cleaning, standardization, and structuring to improve data quality.

- **Gold Layer (Business-Optimized Data):** Aggregates, transforms, and structures the data for business intelligence, pricing analysis, and location-based insights.

---

**2. Business Requirements for Gold Layer**

**2.1 High-Level Requirements**

- **Standardized & Business-Optimized Property Data** for accurate listings and pricing analysis.
- **Aggregated & Structured Features** to enhance market trend insights.
- **Performance-Optimized Data** for dashboards and real estate analytics.
- **Consistent & Accurate Pricing Metrics** to support home buyers and investors.

**2.2 Feature Engineering for Gold Layer**

| Feature Name | Description | Business Value | Transformation Logic |
|---|---|---|---|
| **Standardized Property Name** | Ensures clean property listing names. | Improves search and comparison. | Extract project names and locations separately. |
| **Cleaned Square Feet** | Converts square_feet into a numeric value. | Enables accurate property size filtering. | Remove "sqft" and convert to integer. |

| Transaction Type Standardization | Standardizes transaction types (New, Resale). | Enables trend analysis based on property type. | Extract "New Property" or "Resale", remove project names. |
|---|---|---|---|
| Numeric Price per Sqft | Converts price_per_sqft into a numeric format. | Enables price trend analysis. | Remove "₹", "per sqft" and convert to float. |
| Standardized Total Price | Converts price into numeric value in INR. | Enables accurate price comparisons. | Convert "₹45 Lac" to 4,500,000 INR and "₹1.5 Crore" to 15,000,000 INR. |
| Floor Number & Total Floors | Extracts numeric values from floor column. | Enables better search filters. | Extract "5" from "5 out of 10" and "10" as total floors. |
| Standardized Furnishing | Standardizes furnishing levels. | Improves listing filters for buyers. | Map values to (Unfurnished, Semi-Furnished, Fully Furnished). |
| Standardized Facing Direction | Ensures consistency in property orientation. | Enables direction-based property preferences. | Standardize values (North, South, East, West). |
| Location Standardization | Extracts structured location data from property_name. | Enables location-based analysis and trends. | Extract locality, city (e.g., "Althan, Surat"). |
| Price per Sqft Category | Categorizes properties based on price per square foot. | Helps segment properties into budget, mid-range, and luxury. | Assign categories: Budget (<₹3,000/sqft), Mid-Range (₹3,000-₹6,000/sqft), Luxury (>₹6,000/sqft). |

## 3. Gold Layer Data Model (Optional)

| Column Name | Data Type | Description |
|---|---|---|
| property_id | STRING | Unique identifier for the property. |

| property_name | STRING | Standardized property listing name. |
|---|---|---|
| location | STRING | Extracted locality and city. |
| square_feet | INTEGER | Numeric value of property size. |
| area_type | STRING | Carpet Area or Super Area. |
| transaction_type | STRING | Standardized transaction type (New, Resale). |
| status | STRING | Standardized property status (Ready to Move, Under Construction). |
| floor_number | INTEGER | Extracted floor number. |
| total_floors | INTEGER | Total floors in the building. |
| furnishing | STRING | Standardized furnishing type. |
| facing | STRING | Standardized direction (North, South, East, West). |
| price_per_sqft | FLOAT | Numeric price per square foot. |
| price | FLOAT | Standardized total price in INR. |
| **Engineered Features (Gold Layer Enhancements)** | | |
| price_per_sqft_category | STRING | Categorized property pricing (Budget, Mid-Range, Luxury). |

## 4. Business Benefits of the Gold Layer

- **Consistent & High-Quality Property Listings** – Ensures accurate, structured property details.
- **Enhanced Pricing & Market Insights** – Enables better trend analysis and price forecasting.
- **Improved Buyer & Investor Decision-Making** – Clean and reliable data helps in property selection.
- **Optimized for Search, Filtering & Reporting** – Enables efficient queries for dashboards and comparisons.

- **Better Sales & Marketing Strategies** – Helps real estate businesses segment properties based on pricing, location, and size.