|  |  |
| --- | --- |
| **Project Title** | **Luxury Housing Sales Analysis – Bengaluru** |
| **Skills take away From This Project** | **Data Cleaning using Python (Pandas, NumPy)**  **Data Preprocessing and Transformation**  **SQL Data Warehousing and Querying**  **Building Interactive Dashboards using Power BI**  **Data Visualization and Storytelling**  **Exploratory Data Analysis (EDA)**  **Business Insight Generation**  **Real Estate Market Analytics** |
| **Domain** | **Real Estate**  **Business Intelligence**  **Data Analytics**  **Urban Market Research** |

**Problem Statement:**

Build a complete real estate analytics solution using Python for advanced data cleaning, load the refined dataset into a SQL database, and use Power BI to directly connect to SQL and build a dashboard. The goal is to replicate a real-world enterprise-level data pipeline and analysis environment using a complex housing dataset with 1,00,000+ records.

**Project Deliverables:**

1. Python Script/Notebook:

* Data cleaning
* Feature creation
* Data loaded into SQL

1. SQL Scripts:

* Table schema
* Insertion validation
* Basic aggregation queries

1. Power BI Dashboard:

* Live connection to SQL
* KPI visuals and filters

1. Final Documentation:

* Pipeline steps
* Business Insights
* SQL query and Power BI screenshots

**Business Use Cases:**

* **Market Intelligence**: Identify high-performing localities, builder-wise trends, and configuration demand shifts.
* **Sales Optimization**: Use booking and inquiry data to uncover drop-off patterns.
* **Buyer Persona Building**: Use Buyer Type and Comment sentiment to group and understand buyer personas.
* **Competitive Pricing**: Analyze pricing strategies across builders and market segments.
* **Amenity Score & Conversion**: Determine the correlation between amenities and booking rates.
* **Quarterly Trend Tracking**: Track real estate patterns across fiscal quarters to aid investment decisions.

**Approach:**

**🐍 Step 1: Python — Data Cleaning & Feature Engineering**

* Load the raw .csv file
* Clean inconsistent formats (e.g., Ticket\_Price\_Cr)
* Handle nulls in Amenity\_Score, Booking\_Status, etc.
* Normalize text fields (Builder, Micro\_Market)
* Derive columns like Price\_per\_Sqft, Quarter\_Number, Booking\_Flag
* Output: Cleaned CSV or DataFrame ready for DB insertion

**🧠 Step 2: SQL — Load Clean Data into RDBMS**

* Create appropriate SQL table schema
* Use Python (via SQLAlchemy or pymysql) to insert data into MySQL/PostgreSQL
* Run initial SQL validation queries:
  + SELECT COUNT(\*)
  + GROUP BY booking status
  + AVG ticket price per builder

**📊 Step 3: Power BI — Visualize via Direct SQL Connection**

* Connect Power BI to SQL DB
* Build relationships, DAX calculations
* Create interactive dashboards:  
  + Filter by Builder, Quarter, Market
  + Map visuals, booking conversion KPIs
  + Text box insights from Buyer Comments

**Technical Tags:**

Python, SQL, Power BI, ETL, Data Pipeline, Data Cleaning, Dashboard, Real Estate, Luxury Housing, Micro-Market Analytics, SQLAlchemy, Business Intelligence, Feature Engineering

**Results:**

* Clean, loaded SQL table with normalized data
* Interactive Power BI dashboard connected to SQL live
* Analytical insights useful to real estate firms
* Hands-on with 3 critical tools in the data analytics workflow