A PROJECT REPORT

Project Title	Luxury Housing Sales Analysis – Bengaluru

Submitted by

SAGHEER AHMED

1. Project Overview

The goal of this project is to build an **end-to-end real estate analytics solution** for Bengaluru's luxury housing market. The workflow includes:

- **Data Cleaning & Feature Engineering** using Python (Pandas & NumPy).
- **Data Storage** in a SQL database for structured queries.
- **Interactive Dashboards** in Power BI for business insights and visualization.

The dataset comprises **100,000+ records** containing information about projects, builders, micro-markets, unit sizes, ticket prices, buyer types, and more.

---

2. Folder Structure

```
└── Luxury Housing Sales Analysis Bengaluru/
    ├── LUXURY HOUSING SALES.pbix
    ├── requirements.txt
    ├── src/
    │   ├── data_cleaning.py
    │   ├── geodata.py
    ├── Docs/
    │   └── Document.docx
    │   └── Luxury Housing Sales Analysis Bengaluru.pptx
    ├── asserts/dashboard.png,map.png, table.png
    ├── .git/
    ├── Dataset/
    │   ├── raw/
    │   │   └── Luxury_Housing_Bangalore.csv
    │   ├── coordinates/
    │   │   └── bangalore_places_latlong.csv
    │   ├── cleaned/
    │   │   └── Luxury_Housing_Bangalore_Cleaned_data.csv
```

---

2. Data Pipeline

Step 1: Raw Data

Source: https://drive.google.com/file/d/1nLi3uMk3MPj5TnBZ1hO5_30_-MGVq_Vm/view

Dataset/raw/Luxury_Housing_Bangalore.csv

Key columns:

Column Name	Data Type	Description
Property_ID	object	Unique identifier for each property.
Micro_Market	object	Local area or micro-market in Bengaluru (e.g., Whitefield).
Project_Name	object	Name of the residential project.
Developer_Name	object	Builder/developer company name.
Unit_Size_Sqft	float64	Built-up area of the unit in square feet.
Configuration	object	Unit type (e.g., 4 BHK, 5BHK+).
Ticket_Price_Cr	float64	Sale price in crore INR.
Transaction_Type	object	Primary or resale transaction.
Buyer_Type	object	End-user or investor classification.
Purchase_Quarter	datetime64[ns]	Date/quarter of transaction.
Connectivity_Score	float64	Index rating of transport connectivity.
Amenity_Score	float64	Composite score of project amenities.
Possession_Status	object	Ready-to-move, under construction, etc.
Sales_Channel	object	Channel of sale (direct, broker, VR desk).
NRI_Buyer	object	“Yes” if buyer is a Non-Resident Indian.
Locality_Infra_Score	float64	Infrastructure rating of the locality.
Avg_Traffic_Time_Min	int64	Average commute time to hubs (minutes).
Buyer_Comments	object	Free-text comments/feedback.
Price_per_Sqft	float64	Derived metric = (Ticket_Price_Cr*1e7)/Unit_Size_Sqft.

Column Name	Data Type	Description
Quarter_Number	int32	Quarter (1–4) extracted from Purchase_Quarter.
Booking_Flag	int64	1 = booked (primary sale), 0 = others.
Place	object (16 non-null)	Micro-market/place name used for geocoding.
Latitude	float64 (16 non-null)	Latitude of the micro-market.
Longitude	float64 (16 non-null)	Longitude of the micro-market.

Step 2: Python Data Cleaning & Feature Engineering

Data Cleaning Tasks

- Checked and handled missing values in Ticket_Price_Cr, Unit_Size_Sqft, Amenity_Score, and Buyer_Comments.
- Normalized text fields (Micro_Market, Developer_Name, Configuration).
- Converted Purchase_Quarter to datetime format.

Feature Engineering

- Price_per_Sqft = (Ticket_Price_Cr * 1e7) / Unit_Size_Sqft
- Quarter_Number extracted from Purchase_Quarter.
- Booking_Flag = 1 if Transaction_Type = Primary, else 0.

Output: Cleaned CSV → Dataset/cleaned/Luxury_Housing_Bangalore_Cleaned_data.csv

---

Step 3: SQL Database

Database: MySQL (luxurydb)

Table: luxury_housing_bangalore

Schema: Stores all cleaned columns including derived features.

Python-SQL Integration:

- Used SQLAlchemy and pymysql to insert cleaned data into MySQL.

```
engine = create_engine("mysql+pymysql://root:12345@localhost:3306/luxurydb")

df.to_sql("luxury_housing_bangalore", con=engine, if_exists="replace", index=False)
```

- Validation queries:

```
SELECT COUNT(*) FROM luxury_housing_bangalore;

SELECT Booking_Flag, COUNT(*) FROM luxury_housing_bangalore GROUP BY Booking_Flag;

SELECT Developer_Name, AVG(Ticket_Price_Cr) AS AvgPrice FROM luxury_housing_bangalore
GROUP BY Developer_Name;
```

Geospatial Data

- Micro Market coordinates fetched via geopy,stored in bangalore_places_latlong.csv for map visualizations.

Step 4: SQL schema

Column Name	Data Type	Description
Property_ID	VARCHAR(50)	Primary Key. Unique identifier for each property
Micro_Market	VARCHAR(255)	Sub-location or neighborhood within Bangalore
Project_Name	VARCHAR(255)	Name of the real estate project
Developer_Name	VARCHAR(255)	Name of the developer or builder
Unit_Size_Sqft	FLOAT	Size of the housing unit in square feet
Configuration	VARCHAR(50)	Unit configuration (e.g., 3BHK, 4BHK)
Ticket_Price_Cr	FLOAT	Price of the unit in Crores of INR

Column Name	Data Type	Description
Transaction_Type	VARCHAR(50)	Type of transaction (e.g., Primary, Resale)
Buyer_Type	VARCHAR(50)	Type of buyer (e.g., Individual, Investor)
Purchase_Quarter	DATE	Date representing the quarter of purchase (e.g., 2023-07-01)
Connectivity_Score	FLOAT	Score indicating location connectivity
Amenity_Score	FLOAT	Score indicating available amenities
Possession_Status	VARCHAR(50)	Status like Ready-to-move, Under Construction
Sales_Channel	VARCHAR(50)	Sales medium (e.g., Direct, Broker)
NRI_Buyer	VARCHAR(50)	Flag for NRI buyer (Yes/No or True/False as string)
Locality_Infra_Score	FLOAT	Score for locality infrastructure quality
Avg_Traffic_Time_Min	FLOAT	Average commute time in minutes
Buyer_Comments	TEXT	Free-text comments from the buyer
Price_per_Sqft	FLOAT	Price per square foot (calculated or provided)
Quarter_Number	INT	Numerical representation of the quarter (e.g., Q1 = 1)
Booking_Flag	INT	Indicates if the unit was booked (1/0)

Step 5: Power BI Dashboard

Connection: Live DirectQuery to MySQL (127.0.0.1:3306)

Pages & Features:

1. Home Page – Executive summary with KPIs:

- Total sales price, Avg house price, Min/Max Sqft, Total units sold

Visuals:

- Market Trends (Line Chart by Quarter & Micro-Market)
- Builder Performance (Stacked Bar Chart)
- Booking Conversion (Stacked Column)

- Configuration Demand (Donut Chart)
- Amenity Impact (Scatter Chart)
- Sales Channel Efficiency (Stacked Column)
- Possession Status Analysis (Clustered Column)
- Top 5 Builders (Multi-row Card using Builder_Revenue_Rank)

2. Map Page – Geographical insights:

- Maps Bengaluru micro-markets with bubble size indicating sales volume
- Highlights high-demand zones like MG Road, Koramangala, and Whitefield

3. Table Page – Quarterly Builder Contribution:

- Matrix table showing revenue by builder across quarters
- Highlights top-performing builders per quarter

4. Drill Through Builder Detail Page – Transaction-level analysis:

- Filter by specific Developer (e.g., Brigade)
- KPIs: Total revenue, Avg house price
- Detailed table: Project_Name, Property_ID, Ticket_Price_Cr, Unit_Size_Sqft, Booking_Flag

DAX Measures:

Total_Revenue = SUM('luxury_housing_bangalore'[Ticket_Price_Cr])

Booking_Success_Rate = DIVIDE(SUM('luxury_housing_bangalore'[Booking_Flag]), COUNTROWS('luxury_housing_bangalore'),0)

Builder_Revenue_Rank = RANKX(ALL('luxury_housing_bangalore'[Developer_Name]), [Total_Revenue], , DESC)

Slicers & Filters:

- Configuration, Micro-Market, Builder Name, Quarter
- Reset all filters button for convenience

---

3. Business Insights

Micro-Markets: Sarjapur Road, Jayanagar, and Bannerghatta Road dominate luxury sales.

Builders: Prestige, Total Environment and L&T Realty lead in revenue and booking success.

Configurations: 5BHK+ units are the most in-demand.

Amenities: Higher Amenity_Score strongly correlates with higher booking success.

Sales Channels: Broker and Online perform better for high-ticket sales.

Quarterly Trends: Q4 consistently shows peak sales volume.

4. Summary

This project replicates a real-world enterprise-level data pipeline:

Raw CSV → Python ETL (cleaning + features)

SQL database → Data storage, validation, and querying

Power BI → Live analytics, KPIs, geospatial insights, and interactive dashboards

The final output provides actionable insights for **luxury housing developers, sales teams, and urban market analysts** in Bengaluru.

---

5. Required packages:

Data Ingestion & Cleaning

- pandas - core library for loading CSVs, data cleaning, manipulation.
- datetime - parsing and formatting Purchase_Quarter dates.

Data Quality & Feature Engineering

- Geopy - geocoding micro-market names to latitude/longitude.

Database Integration

- SQLAlchemy - ORM/connection engine for sending DataFrame to CSV to MySQL.
- pymysql - MySQL DB driver used by SQLAlchemy.
