

Luxury Housing Sales Analysis

1. Pipeline Steps

Step 1: Data Cleaning & Feature Engineering (Python)

- Loaded 100k+ raw records into Jupyter Notebook.
- Cleaned inconsistent formats (`Ticket_Price_Cr`).
- Handled missing values (`Amenity_Score`, `Booking_Status`).
- Normalized categorical fields (`Builder`, `Micro_Market`).
- Engineered new features:
 - * Price_per_Sqft
 - * Quarter_Number
 - * Booking_Flag
- Exported cleaned dataset → CSV/SQL insertion.

The screenshot shows a Jupyter Notebook interface with two code cells. The first cell imports pandas and reads a CSV file named 'Luxury_Housing_Bangalore.csv'. The second cell displays the first 10 rows of the resulting DataFrame, which contains 18 columns. The columns are: Property_ID, Micro_Market, Project_Name, Developer_Name, Unit_Size_Sqft, Configuration, Ticket_Price_Cr, Transaction_Type, Buyer_Type, Purchase_Quarter, Connectivity_Score, Amenity_Score, Possession_Status, Sales_Channel, NRI_Buyer, Locality_Infra_Score, Avg_Traffic_Time_Min, and an unlabeled column with values 18, 106, 113, 106, 18, 22, 26, 44, 66, 58. The notebook interface includes a menu bar (File, Edit, Selection, View, Go, Run, Terminal, Help), a search bar, and a sidebar with icons for file explorer, search, and other tools. The status bar at the bottom shows 'Python 3.12.10' and '10/1000 rows x 18 columns'.

	Property_ID	Micro_Market	Project_Name	Developer_Name	Unit_Size_Sqft	Configuration	Ticket_Price_Cr	Transaction_Type	Buyer_Type	Purchase_Quarter	Connectivity_Score	Amenity_Score	Possession_Status	Sales_Channel	NRI_Buyer	Locality_Infra_Score	Avg_Traffic_Time_Min	
0	PRCP000001	Sejapur Road	Project_0	RMZ	4025.0	4bhk	12.75084603918798	Primary	NRI	2025-03-31	7.990091	5.462863	Launch	Broker	yes	9.212491	18	
1	PRCP000002	Indiranagar	Project_1	Puravankara	5760.0	3Bhk	16.29215187100594	Primary	Other	2024-06-30	4.839024	NaN	Under construction	NRI Desk	no	7.723898	106	
2	PRCP000003	Bannerghatta Road	Project_2	Tata Housing	7707.0	4bhk	10.517724412961911	Primary	HNI	2023-12-31	8.131315	8.669227	Ready to move	Direct	yes	6.905493	113	
3	PRCP000004	Bellary road	Project_3	Embassy	6192.0	3BHK	9.396367494232096	Primary	HNI	2024-03-31	7.501657	5.720246	Ready to move	Online	yes	6.100929	106	
4	PRCP000005	Koramangala	Project_4	SNH Raj	7147.0	4Bhk	15.345392444511946	Secondary	HNI	2024-12-31	4.525216	8.609649	Under construction	Broker	no	5.312310	18	
...
100999	PRCP004730	BELLARY ROAD	Project_229	Embassy	8546.0	5Bhk+	11.33081004147843	Secondary	CKO	2024-12-31	8.552797	6.221131	Under construction	NRI Desk	no	7.511827	22	
100996	PRCP009810	Bellary Road	Project_309	Brigade	3408.0	3bhk	10.829373158307602	Primary	CKO	2024-09-30	6.879269	9.783611	Ready to move	Direct	no	9.851849	26	
100997	PRCP065099	HENNUR ROAD	Project_98	RMZ	4091.0	4BHK	11.183303152058548	Primary	NRI	2023-12-31	6.753812	8.383013	Under construction	Broker	no	9.101604	44	
100998	PRCP083022	rajajinagar	Project_21	Embassy	7435.0	3BHK	10.914156376035923	Secondary	Other	2024-06-30	4.437787	5.756247	Under construction	Broker	yes	8.588551	66	
100999	PRCP023826	whitefield	Project_325	L&T Realty	3218.0	4bhk	₹13.27 Cr	Primary	Startup Founder	2023-12-31	7.700416	9.989290	Under construction	Direct	yes	9.917647	58	

10/1000 rows x 18 columns

The second code cell shows the following code:

```
df['Micro_Market'] = df['Micro_Market'].str.strip()
df
```

Step 2: Data Warehousing (SQL)

- Created SQL schema to store cleaned housing data.
- Inserted data using SQLAlchemy + MySQL
- Verified using queries:

```SQL

-- Row Count

```
SELECT COUNT(*) FROM house.house_data;
```

-- Booking Status Distribution

```
SELECT Booking_Status, COUNT(*) AS cnt
FROM house.house_data
GROUP BY Booking_Status
ORDER BY cnt DESC;
```

-- Average Ticket Price per Builder

```
SELECT Builder, AVG(Ticket_Price_Cr) AS avg_price
FROM house.house_data
GROUP BY Builder
ORDER BY avg_price DESC;
```
```

The screenshot displays the MySQL Workbench interface. The top menu bar includes File, Edit, View, Query, Database, Server, Tools, Scripting, and Help. The left sidebar shows the 'SCHEMAS' pane with a tree view of the database structure, including 'house' (Tables, Views, Stored Procedures, Functions) and 'sys'. The main editor window shows a SQL query: `SELECT * FROM house.house_data;`. Below the query editor, the 'Result Grid' displays the data from the 'house_data' table. The table has 13 columns: Property_ID, Micro_Market, Project_Name, Developer_Name, Unit_Size_Sqft, Configuration, Ticket_Price_Cr, Transaction_Type, Buyer_Type, Purchase_Quarter, Connectivity_Score, Amenity_Score, and Possession_Status. The data is sorted by Ticket_Price_Cr in descending order. The bottom pane shows the 'Output' tab with a message: 'Action Output'. The right sidebar contains a 'SQL Additions' pane with a message: 'Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.'

Property_ID	Micro_Market	Project_Name	Developer_Name	Unit_Size_Sqft	Configuration	Ticket_Price_Cr	Transaction_Type	Buyer_Type	Purchase_Quarter	Connectivity_Score	Amenity_Score	Possession_Status
PROP000001	Sarjapur Road	Project_0	RMZ	4025	4bhk	12.7508	Primary	NRI	2025-03-31	7.99000	7.99000	
PROP000002	Indiranagar	Project_1	Puravankara	5760	3bhk	16.2922	Primary	Other	2024-06-30	4.83902	8.13132	
PROP000003	Bannerghatta Road	Project_2	Tata Housing	7707	4bhk	10.5177	Primary	HNI	2023-12-31	8.13132	7.50166	
PROP000004	Bellary Road	Project_3	Embassy	6192	3bhk	9.39637	Primary	HNI	2024-03-31	7.50166	4.52522	
PROP000005	Koramangala	Project_4	SHN Raj	7147	4bhk	15.3454	Secondary	HNI	2024-12-31	4.52522	7.93607	
PROP000006	Yalahanka	Project_5	Brigade	4290	4bhk	9.82	Secondary	HNI	2023-09-30	7.93607	8.64312	
PROP000007	Bannerghatta Road	Project_6	Prestige	6776	3bhk+	10.8968	Secondary	Other	2025-03-31	8.64312	6.78131	
PROP000008	Whitefield	Project_7	Total Environment	5763	3bhk	9.35754	Secondary	NRI	2025-03-31	6.78131	7.28416	
PROP000009	Sarjapur Road	Project_8	Prestige	5815	3bhk	12.2994	Secondary	Other	2024-03-31	7.28416	9.5644	
PROP000010	3rd Stage	Project_9	Puravankara	8675	4bhk	9.19331	Primary	CIO	2023-09-30	9.5644	8.51803	
PROP000011	Rajajinagar	Project_10	Godrej	8872	4bhk	15.2218	Secondary	CIO	2024-06-30	8.51803	8.96303	
PROP000012	Sarjapur Road	Project_11	Brigade	8765	3bhk+	7.38479	Secondary	NRI	2024-03-31	8.96303	9.06152	
PROP000013	Koramangala	Project_12	SHN Raj	4606	3bhk+	11.0425	Primary	Startup Po...	2024-03-31	9.06152		

MySQL Workbench

Local instance MySQL80

File Edit View Query Database Server Tools Scripting Help

Navigator

house_data SQL File 5"

1 • SELECT COUNT(*) FROM house.house_data;

Limit to 1000 rows

Result Grid

COUNT(*)
161054

Administration Schemas

Information

Table: house_data

Columns:

Property_ID	varch
Micro_Market	varch
Project_Name	varch
Developer_Name	varch
Unit_Size_Sqft	float
Configuration	varch
Ticket_Price_Or	float
Transaction_Type	varch
Buyer_Type	varch
Purchase_Quarter	date
Connectivity_Score	float
Amenity_Score	float
Possession_Status	varch

Object Info Session

SQLAdditions

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

Read Only Context Help Snippets

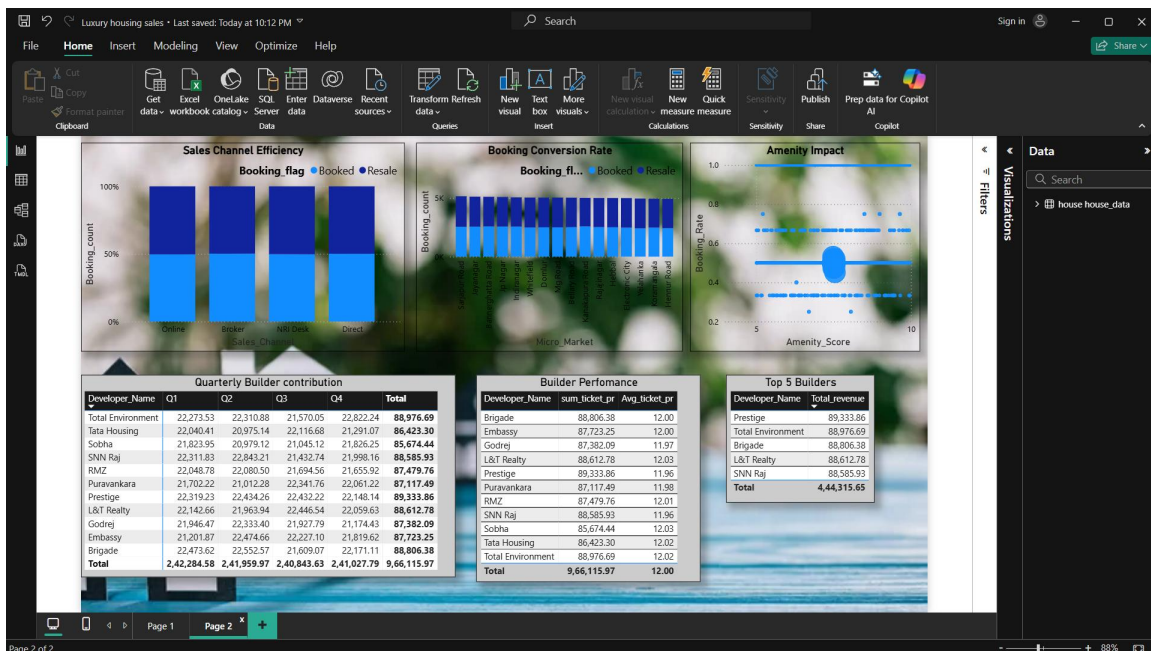
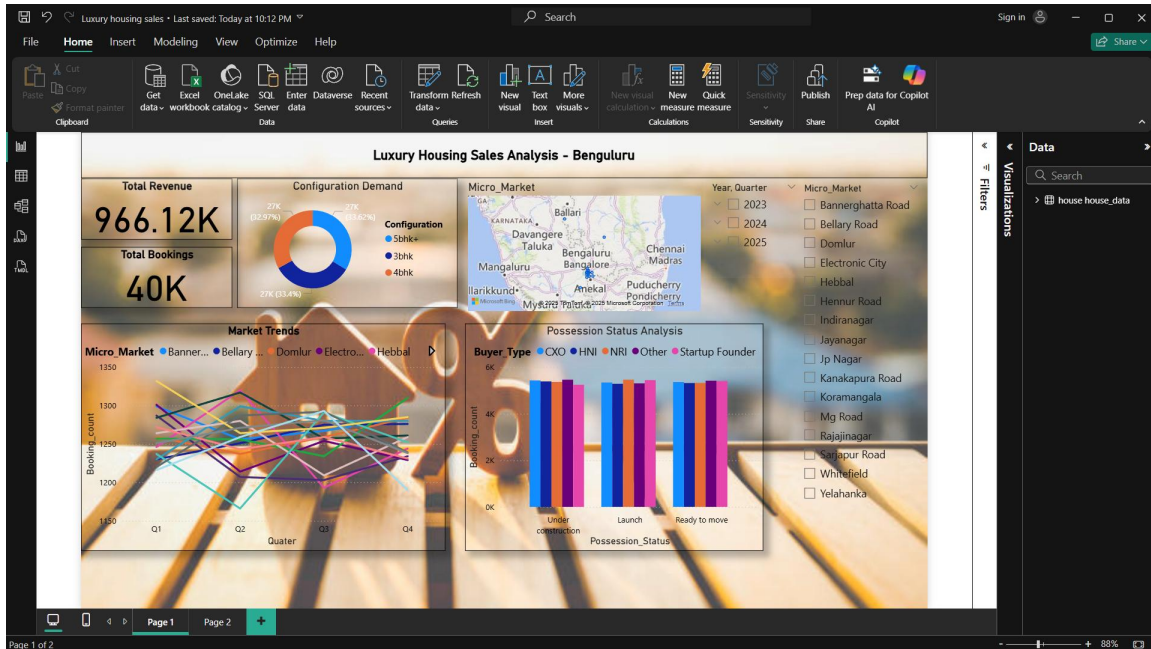
Output

Action Output

#	Time	Action	Message	Duration / Fetch
1	22:28:17	SELECT COUNT(*) FROM house.house_data LIMIT 0, 1000	1 row(s) returned	0.063 sec / 0.000 sec

Step 3: Visualization & Dashboard (Power BI)

- Connected Power BI to SQL database.
- Built relationships, DAX measures, KPIs.
- Created interactive visuals:
 - * Filters (Builder, Quarter, Micro-Market)
 - * Geo maps for project concentration
 - * Builder rankings & booking conversions
 - * Amenity impact analysis



2. Business Insights

Market Trends

- Quarterly booking trends vary strongly across micro-markets.
- Certain localities show consistent demand.

Builder Performance

- Top 5 builders dominate revenue & ticket size.
- Brand trust significantly impacts buyer decisions.

Amenity Impact

- Higher Amenity Scores → Higher Booking Conversions.
- Confirms that amenities drive luxury buyer demand.

Configuration Demand

- 3BHK dominates booking count.
- 4BHK contributes heavily to revenue.

Booking Conversion

- Central Bangalore → higher booking conversions.
- Peripheral areas lag behind.

Possession Status

- Ready-to-move projects have higher bookings.
- Buyers are risk-averse toward under-construction projects.

Geographical Insights

- Whitefield, Sarjapur Road, and North Bangalore are major housing clusters.

3. Tech Stack

- Python (pandas, numpy, sqlalchemy, matplotlib)
- SQL (MySQL)
- Power BI