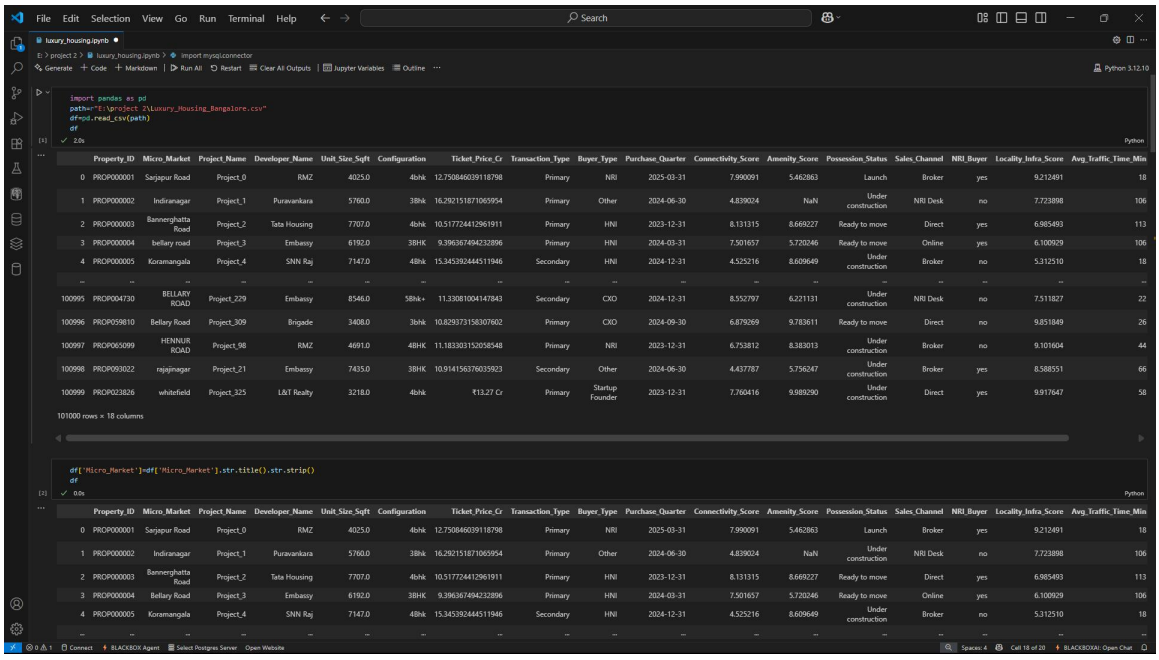
Luxury Housing Sales Analysis

1. Pipeline Steps

Step 1: Data Cleaning & Feature Engineering (Python)

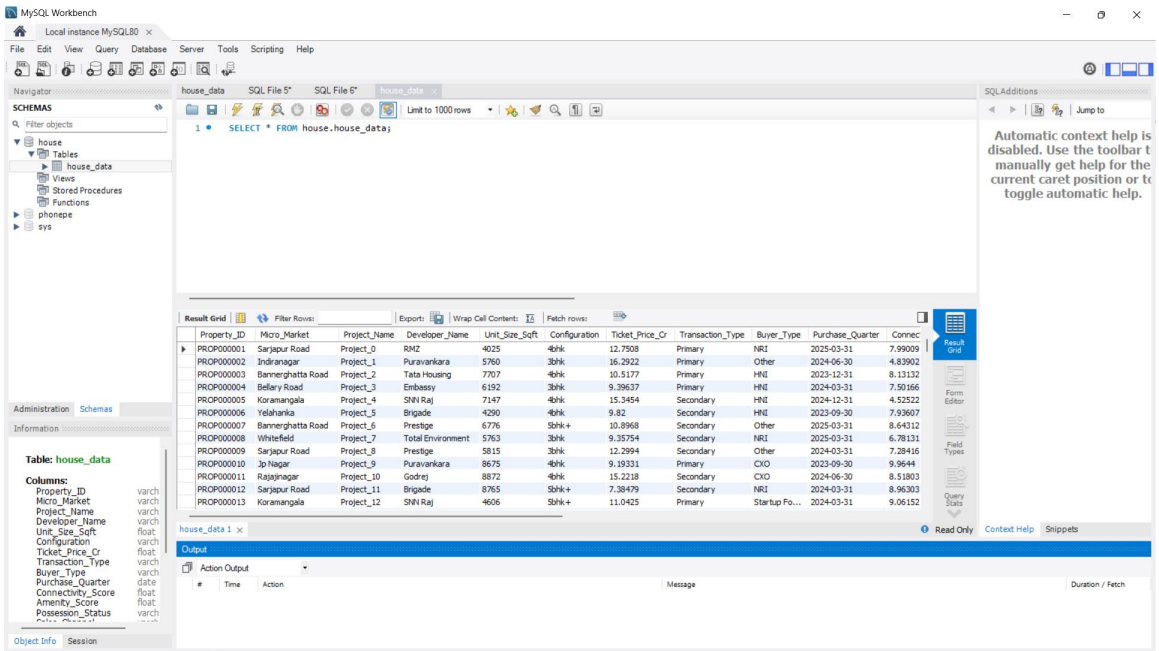
- Loaded 100k+ raw records into Jupyter Notebook.
- Cleaned inconsistent formats (`Ticket_Price_Cr`).
- Handled missing values (`Amenity_Score`, `Booking_Status`).
- Normalized categorical fields (`Builder`, `Micro_Market`).
- Engineered new features:
  * Price_per_Sqft
  * Quarter_Number
  * Booking_Flag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

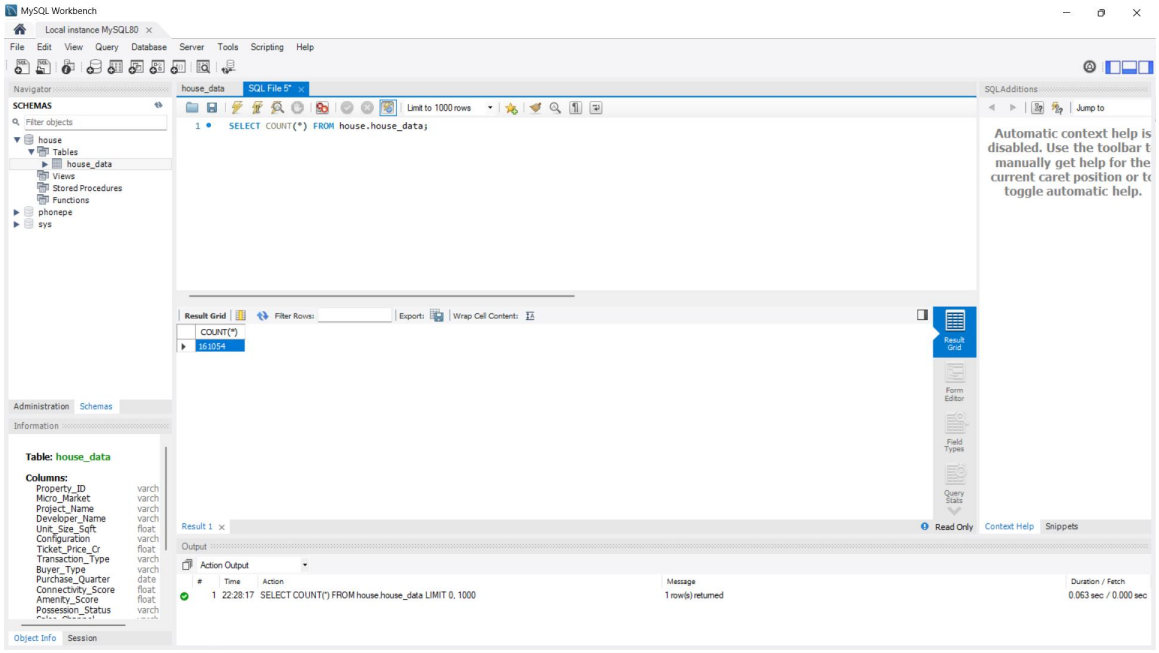


Step 2: Data Warehousing (SQL)

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

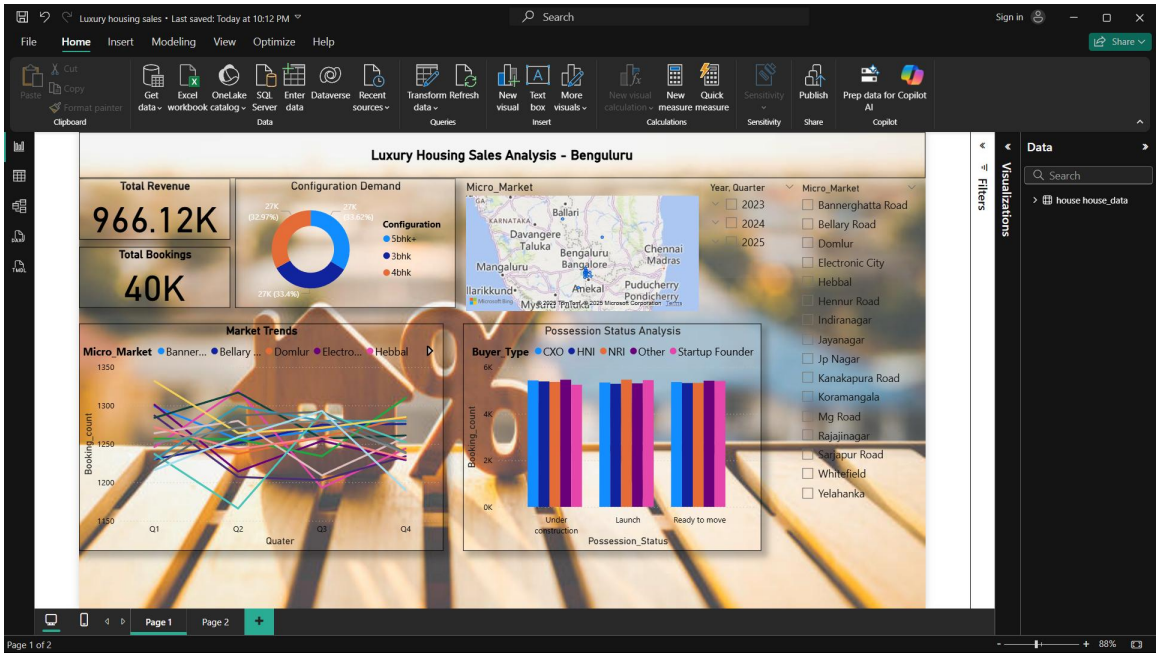
```SQL
-- Row Count
SELECT COUNT(*) FROM house.house_data;

-- 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;
```

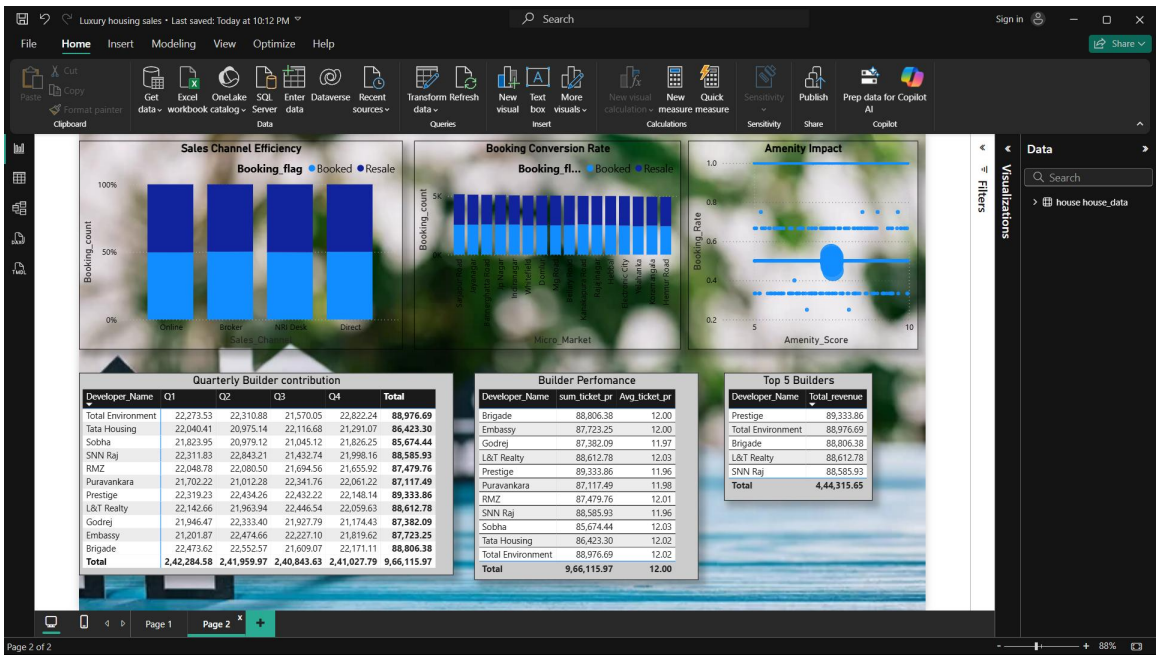



MySQL Workbench
Local instance MySQL80
File Edit View Query Database Server Tools Scripting Help
Navigator
SCHEMAS
Filter objects
house
Tables
house_data
Views
Stored Procedures
Functions
phonepe
sys
Administration Schemas
Information
Table: house_data
Columns:
Property_ID varch
Micro_Market varch
Project_Name varch
Developer_Name varch
Unit_Size_Sqft float
Configuration varch
Ticket_Price_Cr float
Transaction_Type varch
Buyer_Type varch
Purchase_Quarter date
Connectivity_Score float
Amenity_Score float
Possession_Status varch
Object Info Session
house_data
SQL File 5*
Limit to 1000 rows
1 SELECT COUNT(*) FROM house.house_data;
Result Grid
Filter Rows:
Export:
Wrap Cell Content:
COUNT(*)
161054
Result Grid
Form Editor
Field Types
Query Stats
Result 1
Read Only
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar t manually get help for the current caret position or tc toggle automatic help.
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