CHAPTER 01

INTRODUCTION

The real estate sector is one of the most data-intensive industries, involving complex information related to property prices, locations, construction stages, buyer preferences, and sales performance. In metropolitan cities like Bengaluru, luxury housing projects have grown rapidly due to increased urbanization, IT sector expansion, and rising disposable income. However, understanding market behavior and identifying key business drivers remain challenging without systematic data analysis.

This project aims to analyze luxury housing sales data in Bengaluru using a complete data analytics and business intelligence pipeline. The analysis focuses on booking trends, builder performance, amenity influence, geographical distribution, and sales efficiency. The final outcome is an interactive Power BI dashboard that enables stakeholders to make informed, data-driven decisions.

The real estate industry generates large volumes of data related to property pricing, buyer behavior, project status, and sales performance. Analyzing this data helps builders, investors, and stakeholders make informed business decisions.

This project focuses on analyzing luxury housing projects in Bengaluru using a complete data analytics pipeline. The dataset was cleaned and prepared, loaded into a database, and visualized using Power BI to generate meaningful business insights. The dashboard provides insights into booking trends, builder performance, amenity impact, geographical concentration, and top-performing developers. The real estate sector plays a crucial role in the economic growth of metropolitan cities in India. In recent years, Bangalore has emerged as one of the fastest-growing luxury housing markets, driven by rapid urbanization, expansion of the IT and startup ecosystem, rising disposable incomes, and increased demand for premium residential living. Luxury housing projects are characterized by high-value properties, advanced amenities, prime locations, and superior construction quality.

With the growing scale of real estate data, traditional analysis methods are no longer sufficient to derive meaningful insights. Data-driven decision-making has become essential for builders, investors, and policymakers to understand market behavior, buyer preferences, and sales performance.

CHAPTER 02

PROBLEM STATEMENT

Real estate developers and investors often struggle to answer critical business questions such as:

- Which builders are performing best in terms of revenue and booking success?
- Which micro-markets show high or low booking conversion?
- Does the availability of amenities influence buyer decisions?
- How are luxury housing projects geographically distributed across Bengaluru?
- Which housing configurations are most preferred by buyers?

The lack of integrated analytics tools makes it difficult to derive these insights from raw datasets. Therefore, this project addresses the need for a structured, end-to-end analytics solution that transforms raw housing data into meaningful business intelligence.

The goal of this project is to build an end-to-end analytics solution that transforms raw real estate data into actionable insights using modern business intelligence tools. Despite the availability of large volumes of real estate data, many stakeholders still rely on intuition-based decisions. Factors such as market trends, builder performance, booking conversion rates, amenity impact, and sales channel efficiency are often not analyzed in an integrated manner. There is a need for an interactive, analytical dashboard that provides a comprehensive view of market trends, performance metrics, and conversion insights to support informed decision-making.

The luxury housing market in Bangalore is highly competitive, with multiple builders operating across various micro-markets and quarters. However, stakeholders face challenges such as:

- Lack of visibility into quarter-wise booking trends
- Difficulty in identifying high-performing builders and locations
- Limited understanding of buyer behavior and configuration demand
- Inefficient evaluation of sales channels and amenities

CHAPTER 03

OBJECTIVES

The primary objective of this project is to design and implement an end-to-end data analytics and business intelligence solution for analyzing luxury housing sales in Bengaluru. The project aims to convert raw, unstructured real estate data into meaningful insights that support strategic decision-making for builders, investors, and stakeholders.

The detailed objectives are as follows:

3.1 Data Cleaning and Preparation

To clean, preprocess, and standardize the raw housing dataset by handling missing values, correcting data types, and removing inconsistencies to ensure data accuracy and reliability for analysis.

3.2 Feature Engineering for Business Analysis

To create derived attributes such as Booking Status and key performance indicators (KPIs) that are not directly available in the raw dataset but are essential for analyzing booking conversion, sales performance, and market trends.

3.3 Analysis of Booking Trends

To analyze booking patterns across different quarters, micro-markets, and buyer types in order to understand seasonal demand, regional performance, and customer behavior in the luxury housing segment.

3.4 Builder Performance Evaluation

To evaluate and compare real estate developers based on total revenue, booking volume, and average ticket size, thereby identifying top-performing builders and understanding competitive positioning in the market.

3.5 Micro-Market Comparison

To study and compare the performance of various Bengaluru micro-markets in terms of booking conversion rates, project concentration, and pricing trends, helping identify high-growth and underperforming locations.

3.6 Amenity Impact Assessment

To assess the relationship between amenity score and booking success in order to determine whether enhanced amenities influence buyer decisions in luxury housing projects.

3.7 Configuration Demand Analysis

To identify buyer preferences by analyzing demand across different housing configurations such as 2BHK, 3BHK, and 4BHK, enabling insights into product mix optimization.

3.8 Sales Channel Effectiveness

To evaluate the effectiveness of different sales channels by comparing their booking success rates, thereby identifying the most efficient channels for luxury housing sales.

3.9 Geographical Distribution Analysis

To visualize and analyze the geographical concentration of luxury housing projects across Bengaluru using map-based visualizations, helping stakeholders understand spatial market dynamics.

3.10 Dashboard Development and Visualization

To design and develop an interactive Power BI dashboard that consolidates all analytical insights into an easy-to-use visual interface with filters, slicers, and KPIs for real-time business analysis.

3.11 Decision Support and Business Value

To provide a data-driven decision support system that enables stakeholders to make informed strategic, operational, and investment decisions in the luxury housing market.

CHAPTER 04

SCOPE OF THE PROJECT

This project focuses exclusively on luxury housing projects within Bengaluru city. The analysis is limited to the attributes available in the dataset, such as pricing, possession status, amenities, and buyer type. External factors like interest rates, government policies, and macroeconomic indicators are not included. The project is analytical and descriptive in nature, with optional scope for predictive modeling in the future.

The scope of this project defines the boundaries, coverage, and limitations of the analysis carried out on luxury housing sales data in Bengaluru. The project is focused on providing descriptive and diagnostic analytics through data visualization and business intelligence techniques.

4.1 Geographical Scope

The analysis is restricted to luxury housing projects located within Bengaluru city. Micro-markets such as Whitefield, Sarjapur Road, Hebbal, and other prominent locations are included. No other cities or regions are considered in this study.

4.2 Data Scope

The project uses a single consolidated dataset containing information related to luxury housing properties. The analysis is limited to the columns available in the dataset, including pricing, configuration, possession status, amenities, buyer type, sales channel, and purchase quarter.

External datasets such as interest rates, economic indicators, government policies, or demographic data are not included.

4.3 Analytical Scope

The project focuses on:

- Booking trend analysis across time and locations
- Builder performance evaluation
- Booking conversion analysis by micro-market
- Amenity impact assessment

- Configuration demand analysis
- Sales channel efficiency
- Geographical concentration analysis

The analytics performed are descriptive (what happened) and diagnostic (why it happened), not predictive.

4.4 Tool and Technology Scope

The scope includes the use of:

- Power BI for data visualization and dashboard development
- Python and SQL for data cleaning and validation
- Excel/CSV for data handling

Advanced machine learning models, AI-based forecasting, and real-time streaming analytics are outside the current scope.

4.5 Visualization Scope

The project includes the creation of:

- Bar charts, line charts, donut charts, stacked charts
- Scatter plots for correlation analysis
- Map visualizations using micro-market data
- KPI cards and tables for performance tracking

Highly complex custom visuals and third-party Power BI extensions are not used.

4.6 User Scope

The dashboard is designed for:

- Real estate developers
- Sales and marketing teams
- Investors and analysts
- Management stakeholders

The project does not include role-based access control or multi-user authentication.

4.7 Time Scope

The analysis is limited to the time period represented in the dataset. Real-time or continuously updating data sources are not part of this project.

The scope of this project is confined to the analysis of luxury housing projects in Bangalore using structured real estate data to derive meaningful business insights. The study focuses on examining quarter-wise market trends, micro-market performance, builder-wise revenue contribution, booking conversion rates, and buyer preferences across different housing configurations. It also includes the evaluation of amenity impact on booking success, possession status influence on buyer behavior, and the effectiveness of various sales channels in driving successful bookings. The project leverages Business Intelligence techniques using Power BI for data modeling, visualization, and interactive dashboard development, enabling stakeholders to explore data dynamically. The analysis is limited to the dataset provided and does not include affordable or mid-segment housing, future market predictions, or external economic factors. The primary objective of this scope is to support data-driven decision-making for builders, investors, and management by presenting clear, accurate, and visually intuitive insights into the luxury real estate market.

CHAPTER 05

DATASET DESCRIPTION

The dataset contains 100,000+ records representing luxury housing properties in Bengaluru.

Key Attributes:

- Property_ID: Unique identifier for each property
- Project_Name: Name of the housing project
- Developer_Name: Builder or real estate developer
- Micro_Market: Sub-location within Bengaluru
- Ticket_Price_Cr: Property price in crores
- Configuration: Property type (2BHK, 3BHK, 4BHK, etc.)
- Possession_Status: Stage of construction or readiness
- Amenity_Score: Numeric score representing available amenities
- Purchase_Quarter: Quarter in which the transaction occurred
- Sales_Channel: Mode of sales (Direct, Broker, Online, etc.)
- Buyer_Type: Investor or End User

The dataset is designed to simulate a real-world enterprise housing dataset.

DATASET :- LINK

CHAPTER 06

METHODOLOGY

Data Cleaning & Preprocessing

Data cleaning was performed to ensure accuracy, consistency, and usability of the dataset.

Cleaning Steps Performed:

- Verified data types for numerical and categorical fields
- Ensured pricing values were in consistent numeric format
- Checked for missing or null values
- Removed redundant or invalid entries where necessary
- Standardized categorical values such as possession status

Derived Feature Creation:

Since booking status was not explicitly available, a new column Booking_Status was derived using domain logic:

- Ready to Move → Booked
- Under Construction → Booked
- Launch → Not Booked

This derived feature enabled booking conversion analysis across different dimensions

Tools & Technologies Used

- Power BI – Interactive dashboard creation
- SQL – Data storage, validation, and querying
- Python (Pandas, NumPy) – Data preprocessing and EDA
- Microsoft Excel – Data inspection and formatting
- GitHub – Project version control and documentation

- The methodology of this project follows a systematic and structured data analytics workflow, ensuring accuracy, reliability, and meaningful insight generation. The project adopts an end-to-end approach, starting from raw data acquisition and ending with interactive dashboard development and insight interpretation.

6.1 Data Collection

- The dataset used in this project consists of structured luxury housing sales data for Bengaluru. The data was provided in CSV format and contains information related to property pricing, builder details, possession status, amenities, buyer type, and sales channels.
- The dataset represents a simulated real-world enterprise dataset and was chosen to reflect realistic business scenarios in the luxury real estate domain.

6.2 Data Understanding and Initial Exploration

Before cleaning, the dataset was explored to:

- Understand the structure and meaning of each column
- Identify numerical and categorical attributes
- Detect missing values, duplicates, and anomalies
- Assess data distribution and consistency

Exploratory checks were performed using Python and Excel to ensure familiarity with the data and identify potential issues.

6.3 Data Cleaning and Preprocessing

Data cleaning was performed to improve data quality and analytical accuracy.

The following preprocessing steps were applied:

- Verified and corrected data types for numerical and categorical fields
- Ensured uniform formatting of categorical values such as possession status
- Checked for missing or null values and validated their impact
- Removed inconsistencies and ensured logical data integrity

These steps ensured that the dataset was reliable and suitable for analytical modeling and visualization.

6.4 Feature Engineering

Since certain analytical attributes were not directly available in the dataset, feature engineering was performed to create meaningful derived fields.

A new column Booking_Status was created using domain logic based on possession status:

- Properties marked as Ready to Move or Under Construction were classified as Booked
- Properties marked as Launch were classified as Not Booked

Additionally, key analytical measures such as total revenue and booking counts were created using DAX in Power BI.

6.5 Data Modeling

The cleaned dataset was imported into Power BI, where data modeling was performed to ensure correct aggregation and analysis.

Key activities included:

- Validating table structure and relationships
- Ensuring proper cardinality where applicable
- Setting appropriate data types and summarization behavior
- Verifying measures and calculated columns

Proper data modeling ensured accurate calculations across all visuals.

6.6 Measure Creation and KPI Development

Business-relevant metrics were created using DAX (Data Analysis Expressions) to support analysis.

Key measures included:

- Total Revenue
- Booked Property Count
- Not Booked Property Count
- Booking Conversion Rate
- Average Ticket Size

These KPIs were used across multiple visuals to maintain consistency and accuracy.

6.7 Visualization Design

Interactive dashboards were created in Power BI using appropriate visual types selected based on business questions.

Visualization principles followed:

- Clear mapping of fields to axes and legends
- Minimal clutter and readable layouts
- Use of filters and slicers for interactivity
- Consistent color and labeling standards

Visuals included bar charts, line charts, stacked charts, donut charts, scatter plots, maps, tables, and KPI cards.

6.8 Insight Generation and Interpretation

Each visualization was carefully analyzed to identify:

- Trends and patterns
- Performance comparisons
- Correlations between variables
- Outliers and anomalies

Insights were documented in business-friendly language to ensure interpretability for non-technical stakeholders.

6.9 Validation and Accuracy Checks

To ensure correctness:

- KPI values were cross-verified using table visuals
- Aggregations were validated against raw data
- Visual filters and slicers were tested
- Edge cases were reviewed for logical consistency

This validation ensured trustworthy insights.

6.10 Dashboard Finalization

The final Power BI dashboard was refined by:

- Aligning visuals according to business flow
- Adding titles, tooltips, and labels
- Ensuring smooth user interaction
- Optimizing layout for presentation and evaluation

The completed dashboard serves as a decision-support tool for luxury housing market analysis.

CHAPTER 07

Power BI Visualizations & Analysis

1. Market Trends – Quarter-wise Booking Analysis

Description

This analysis focuses on understanding quarter-by-quarter changes in luxury housing bookings across different micro-markets in Bangalore. By tracking booking trends over time, we can identify seasonal patterns, growth phases, and declining markets.

Objective

- To analyze how booking volumes vary across quarters
- To identify high-growth and low-performance micro-markets
- To understand market momentum in luxury housing segments

Insights

- Certain micro-markets show consistent growth, indicating strong demand
- Some markets experience seasonal spikes during festive or financial quarters
- Declining trends highlight areas requiring pricing or marketing intervention

Business Value

This helps developers and investors plan launches, optimize pricing strategies, and allocate resources efficiently based on demand cycles.

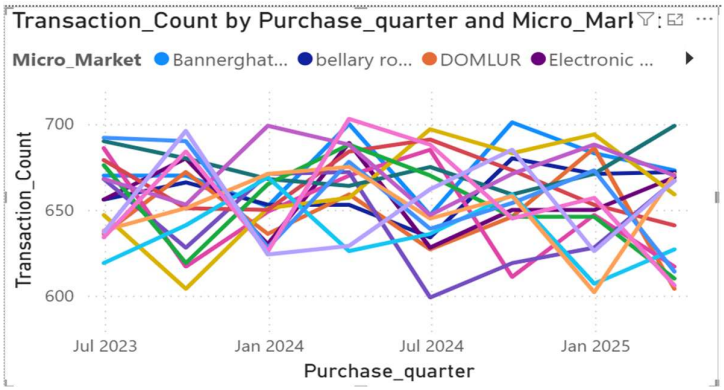


Fig 7.1 Market Trends

2. Builder Performance – Revenue and Ticket Size

Description

This section evaluates builder performance by comparing total ticket sales revenue and average ticket size. It highlights both volume-driven and premium-focused builders.

Objective

- To identify top-revenue-generating builders
- To compare builders based on average property value
- To rank builders on financial performance

Insights

- Some builders generate high revenue through volume sales
- Others focus on fewer but high-value premium projects
- Builders with balanced revenue and ticket size show stable market positioning

Business Value

Helps buyers, investors, and stakeholders assess builder credibility, pricing strategy, and market dominance.

Developer_Name	Total_Ticket_Sales	Avg_Ticket_Size
Total Environment	97,096.23	12.62
Tata Housing	94,258.76	12.52
Sobha	94,368.59	12.63
SNN Raj	96,344.34	12.43
RMZ	95,297.92	12.54
Puravankara	95,050.58	12.48
Prestige	97,375.80	12.56
L&T Realty	97,481.99	12.66
Godrej	96,754.88	12.60
Embassy	94,990.06	12.47
Brigade	95,548.15	12.45
Total	**10,54,567.30**	**12.54**

Fig 7.2 Builder Performance

3. Amenity Impact – Correlation Analysis

Description

This analysis studies the relationship between amenity score and booking conversion rate to understand whether better amenities drive higher sales success.

Objective

- To examine the impact of amenities on booking decisions
- To identify correlation between project quality and sales performance

Insights

- Projects with higher amenity scores generally show better conversion rates
- Larger projects with more amenities attract higher buyer engagement
- Some outliers indicate pricing or location influence beyond amenities

Business Value

Guides builders on amenity investment decisions and helps prioritize features that improve conversion.

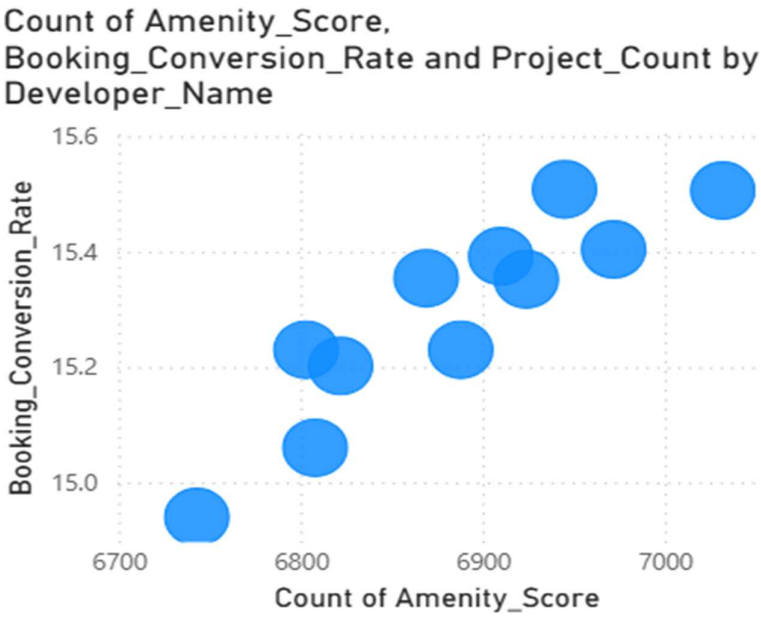


Fig 7.3 Amenity Impact

4. Booking Conversion – Micro-Market Comparison

Description

This visualization compares booking conversion rates across micro-markets, showing successful vs unsuccessful bookings.

Objective

- To identify high-conversion and low-conversion micro-markets
- To assess buyer readiness and market maturity

Insights

- Certain micro-markets consistently outperform others
- Low conversion markets may suffer from affordability or oversupply
- Balanced markets show healthy buyer interest

Business Value

Supports location-based strategy planning for sales, marketing, and new project launches.

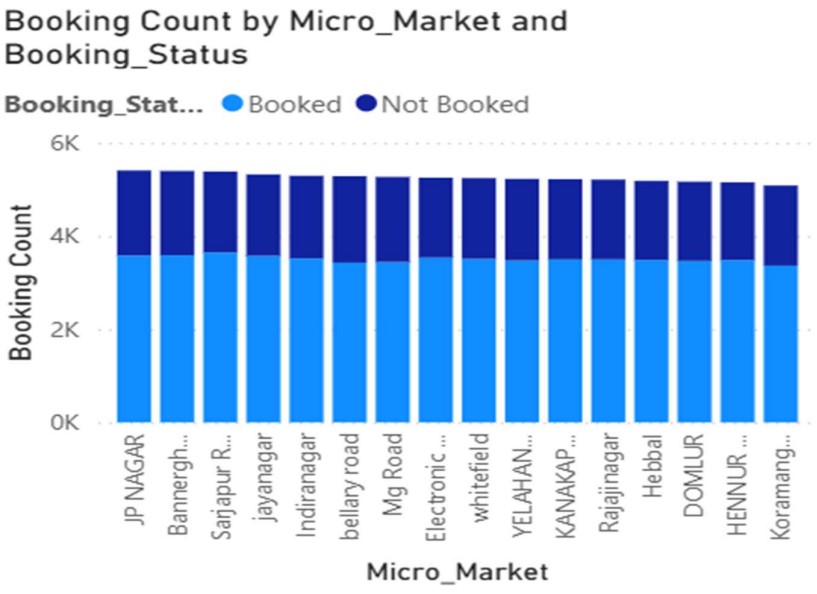


Fig 7.4 Booking Conversion

5. Configuration Demand – Buyer Preferences

Description

This section analyzes demand across housing configurations such as 2BHK, 3BHK, and 4BHK based on booking count.

Objective

- To identify the most preferred luxury housing configuration
- To understand buyer space requirements

Insights

- 3BHK units dominate luxury housing demand
- Larger configurations appeal to premium buyers and investors
- Smaller configurations show limited demand in luxury segments

Business Value

Helps builders design inventory mix aligned with buyer preferences.

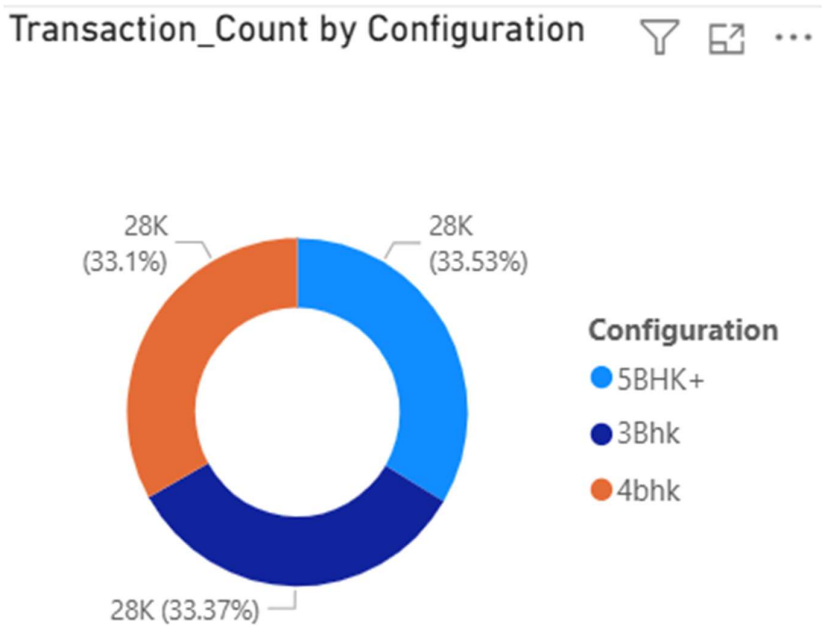


Fig 7.5 Configuration Demand

6. Sales Channel Efficiency

Description

This analysis evaluates the effectiveness of different sales channels in converting leads into successful bookings.

Objective

- To identify the most successful sales channels
- To compare conversion efficiency across channels

Insights

- Direct sales and referrals show higher conversion rates
- Online platforms generate volume but lower conversion
- Channel mix optimization improves overall performance

Business Value

Enables builders to optimize marketing spend and focus on high-ROI channels.

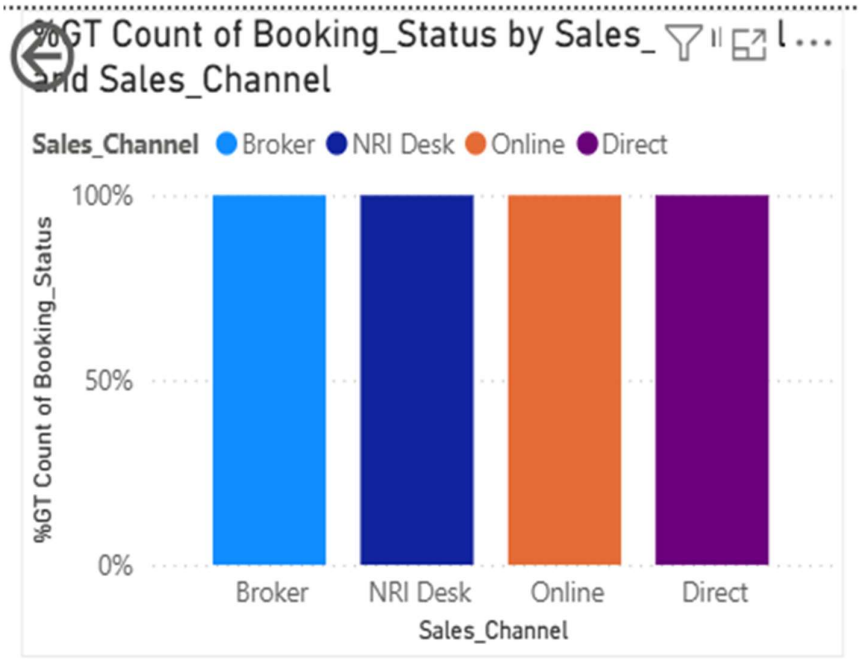


Fig 7.6 Sales Channel Efficiency

7. Quarterly Builder Contribution

Description

This matrix analysis shows builder-wise revenue contribution per quarter, highlighting seasonal dominance and consistency.

Objective

- To identify builders dominating specific quarters
- To track consistency in revenue generation

Insights

- Some builders perform strongly during specific quarters
- Consistent performers indicate strong brand presence
- Seasonal dips reflect launch timing strategies

Business Value

Useful for competitive benchmarking and strategic planning.

Developer_Name	2023	2024	2025	Total
Brigade	35,366.04	48,173.94	12,008.18	**95,548.15**
Embassy	35,935.13	47,926.78	11,128.15	**94,990.06**
Godrej	37,163.64	47,916.86	11,674.38	**96,754.88**
L&T Realty	35,825.89	49,896.97	11,759.13	**97,481.99**
Prestige	36,299.69	48,947.26	12,128.84	**97,375.80**
Puravankara	36,601.17	47,118.04	11,331.37	**95,050.58**
RMZ	36,262.09	47,125.59	11,910.24	**95,297.92**
SNN Raj	36,583.92	47,744.43	12,015.99	**96,344.34**
Sobha	35,013.78	47,773.84	11,580.96	**94,368.59**
Tata Housing	33,896.22	48,379.56	11,982.99	**94,258.76**
Total Environment	36,562.65	49,006.83	11,526.76	**97,096.23**
Total	**3,95,510.22**	**5,30,010.09**	**1,29,046.99**	**10,54,567.30**

Fig 7.7 Quarterly Builder Contribution

8. Possession Status Analysis

Description

This section studies how possession status (Ready / Under Construction) influences booking outcomes across different buyer types.

Objective

- To understand buyer preference based on possession stage
- To analyze booking behavior across buyer segments

Insights

- End-users prefer ready-to-move properties
- Investors show interest in under-construction projects
- Possession status significantly affects booking success

Business Value

Guides builders in project phasing and marketing communication.

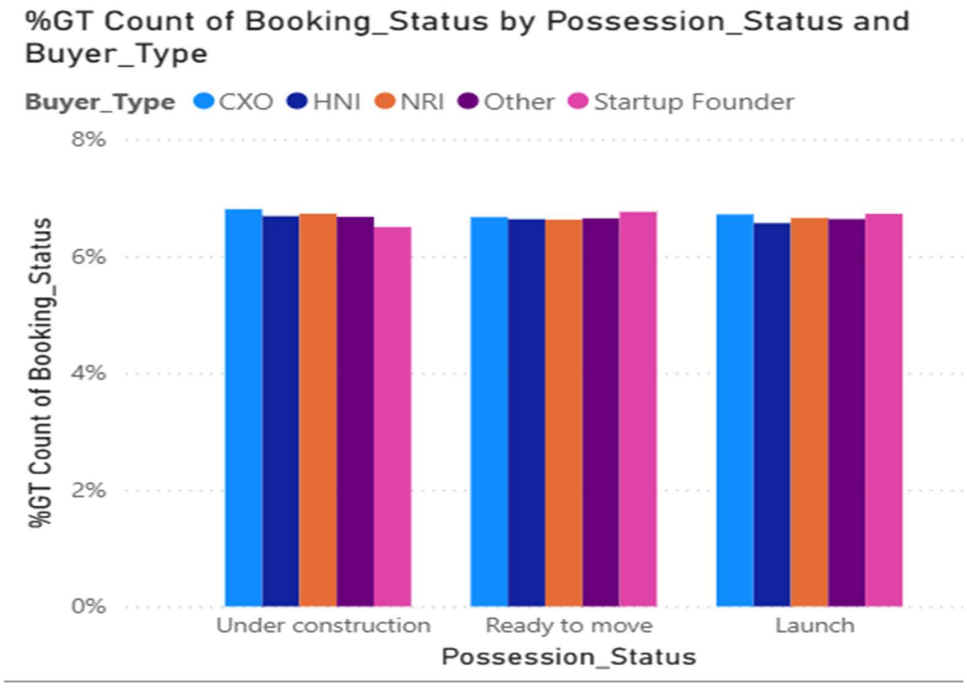


Fig 7.8 Possession Status Analysis

9. Geographical Insights – Bangalore Luxury Market

Description

This visualization maps the concentration of luxury housing projects across Bangalore micro-markets.

Objective

- To identify geographic hotspots for luxury housing
- To understand regional development trends

Insights

- North and East Bangalore show high project concentration
- Emerging micro-markets indicate future growth potential
- Central zones remain premium but supply-constrained

Business Value

Supports location-based investment and expansion decisions.

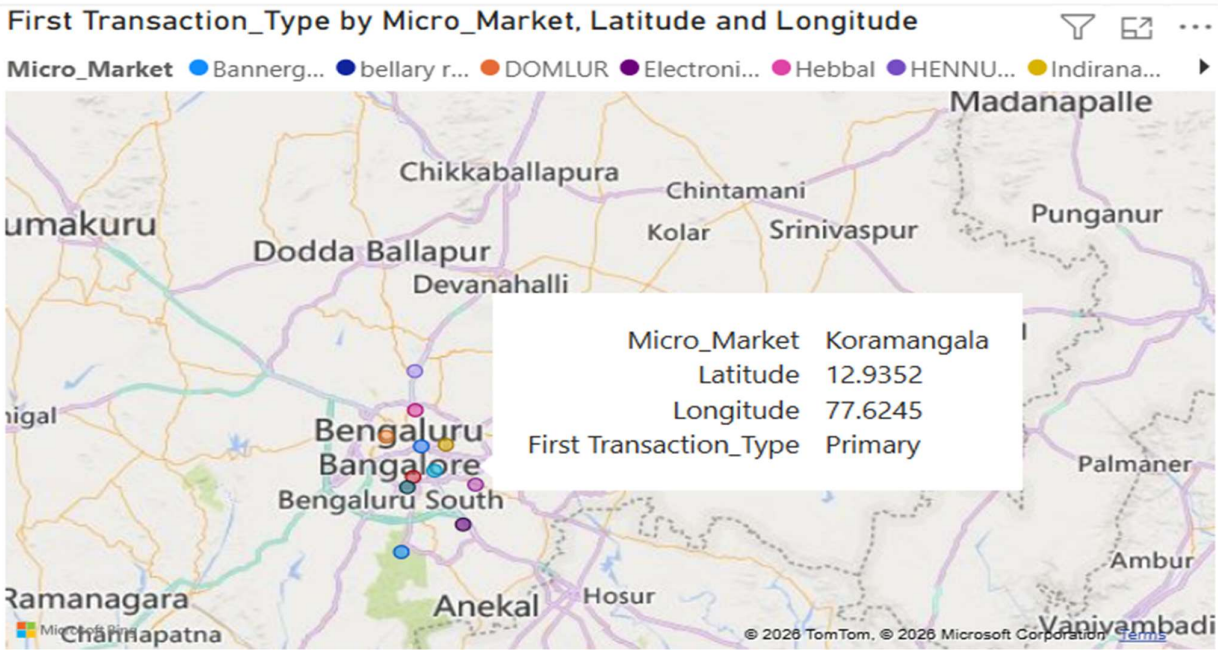


Fig 7.9 Geographical Insights

10. Top Performers – Builder Excellence

Description

This KPI-based analysis identifies the top 5 builders based on revenue and booking success rate.

Objective

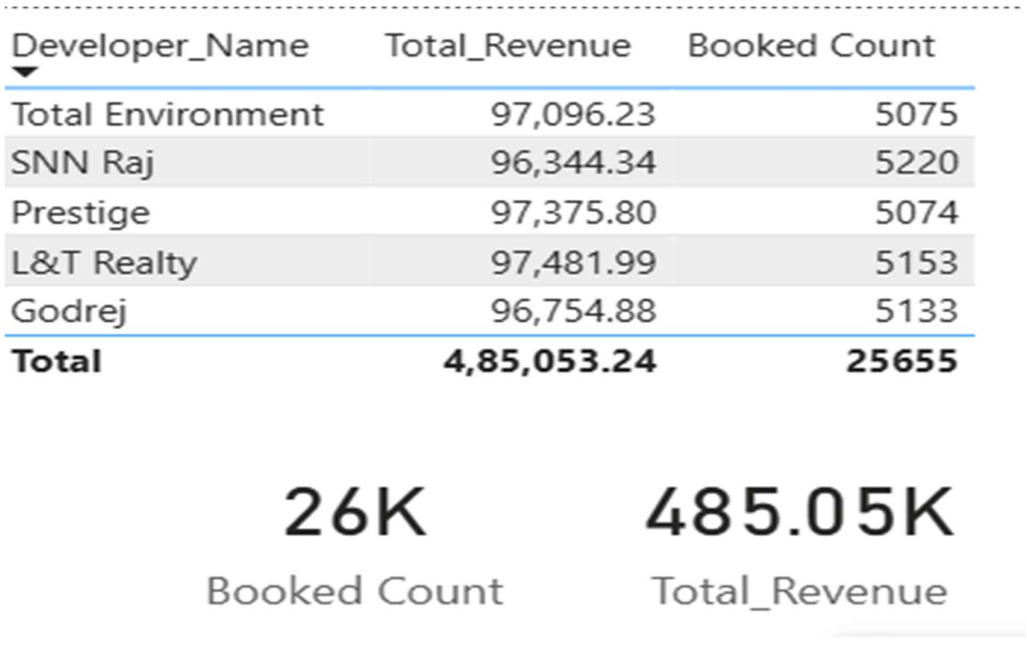
- To recognize market leaders
- To compare performance on revenue and conversion efficiency

Insights

- Top builders maintain both high revenue and strong conversion
- Strong branding and trust drive booking success
- Drill-through enables deeper project-level analysis

Business Value

Provides a quick executive summary for decision-makers and investors.



Developer_Name	Total_Revenue	Booked Count
Total Environment	97,096.23	5075
SNN Raj	96,344.34	5220
Prestige	97,375.80	5074
L&T Realty	97,481.99	5153
Godrej	96,754.88	5133
Total	**4,85,053.24**	**25655**

Fig 7.10 Top Performers – Builder Excellence

CONCLUSION

In conclusion, this project successfully demonstrates how data analytics and business intelligence tools can be effectively used to analyze and interpret the performance of the luxury housing market in Bangalore. Through interactive Power BI dashboards, the study provides clear insights into quarter-wise market trends, micro-market performance, builder-wise revenue contribution, booking conversion rates, buyer preferences, and the impact of amenities and sales channels on booking success. The analysis highlights key market leaders, high-demand configurations, and geographically strong zones, enabling stakeholders to identify growth opportunities and performance gaps. By transforming raw real estate data into meaningful visual insights, the project supports informed, data-driven decision-making for builders, investors, and management teams. Overall, this study emphasizes the importance of analytical approaches in the real estate sector and showcases the practical application of business intelligence in solving real-world market problems.