

# AI-Assisted Real Estate Data Modeling and SQL Analysis Report

## 1. Introduction

This report presents a database modeling and SQL analysis study based on field notes related to the real estate market. The study examines the impact of student demand on rental prices, sales prices, and payback (amortization) periods across different regions. Throughout the process, an AI-based tool (ChatGPT) was used strictly as a technical assistant.

## 2. Purpose and Scope of the Study

The main objectives of this study are:

- To transform qualitative field observations from the real estate market into a structured data model,
- To establish a conceptual data model using an ER (Entity–Relationship) diagram,
- To design a MySQL-compatible database schema,
- To generate synthetic data for testing purposes,
- To derive investment-oriented insights using analytical and complex SQL queries.

The study covers the regions of Cebeci, İncek, Öveçler, Cevizlidere, and Ön Cebeci.

## 3. Data Model and ER Diagram

Based on the field notes, the following core entities were identified:

- Region: Infrastructure level, transportation access, supply status, and student contribution
- Property Type: Housing characteristics, building age, and risk level
- Price Information: Rental price, sale price, and payback period
- Demand Structure: Ratios of students, civil servants, tradespeople, and single households
- Demographics: Student count, household count, and demographic ratios - Market Notes: Qualitative field observations

The relationships between these entities were modeled using an ER diagram and implemented in DBML format via dbdiagram.io.

## 4. Database Design (MySQL)

The conceptual data model was transformed into a physical database design compatible with MySQL. Primary key (PK) and foreign key (FK) relationships were defined to ensure data integrity. Boolean-type attributes were implemented using MySQL-compatible data types, and the InnoDB storage engine was used.

## **5. Synthetic Data Generation**

To enable testing and analysis, synthetic yet realistic data were generated. These data were created within plausible ranges derived from the original field notes, allowing meaningful query execution and result interpretation without using real market data.

## **6. Analytical SQL Queries and Findings**

### **6.1 Region-Based Investment Analysis**

SQL queries were developed to calculate average rental prices, sale prices, and payback periods for each region. The results indicate that regions with shorter payback periods are more attractive from an investment perspective. In this context, Ön Cebeci and Cebeci stand out.

### **6.2 Impact of Student Demand**

The analysis shows that regions with higher student ratios do not always exhibit higher rent per square meter. However, due to relatively lower sales prices, these regions tend to offer shorter payback periods. This suggests that student demand primarily affects investment return duration rather than directly increasing prices.

### **6.3 Risk and Return Relationship**

For properties classified as risky (e.g., older buildings or those with regulatory issues), sale prices tend to be lower while rental income remains at comparable levels. This highlights a clear high-risk, high-return dynamic.

## **7. Overall Evaluation and Conclusion**

The findings of this study indicate that:

- Student demand is a key determinant in real estate market dynamics,
- In regions with excess supply, students play a stabilizing role,
- Riskier properties often provide shorter payback periods for investors,
- Database-driven analysis is a powerful tool for systematic investment evaluation.

This study demonstrates that real estate markets can be systematically analyzed through data modeling and SQL-based analytical approaches.

## **8. AI Usage Disclosure**

In this project, ChatGPT was used as a supporting technical tool during the database modeling and SQL query development phases. The analytical approach, modeling decisions, and interpretation of results were conducted by the author.