

Conceptual Data Model – UK Housing Transactions

1. Objective and Context

This document describes the **conceptual data model** for a UK housing transactions database derived from CSV source files. The model is designed to support analytical reporting, dashboarding, and downstream analytics, including a Retrieval-Augmented Generation (RAG) model layered on top of curated data.

The modelling approach follows dimensional modelling principles, with a **star schema** centred on property sale transactions.

2. Business Process Overview

The core business process is property sales. Each sale occurs at a point in time, at a specific location, and involves a specific property type. I model this using a star schema with a central fact table at the transaction grain, surrounded by time, location, and property dimensions to support regional, temporal, and categorical analysis.

Core Business Process

The core business process being modelled is **property sales**.

Each property sale:

- Occurs at a specific point in time
- Takes place at a defined geographical location
- Involves a specific property classification
- Has an associated monetary value

To support analytical queries across geography, time, and property characteristics, the model uses a **central fact table at the transaction grain**, surrounded by descriptive dimensions.

3. Key Business Questions Supported

The conceptual model is designed to answer the following analytical questions:

- How do house prices vary by region?
- How does the housing market change over time?
- How do prices differ by property type?
- What are monthly transaction volumes and year-over-year (YoY) growth trends?

4. Conceptual Entities (Nouns)

4.1 Property Sale (Core Event)

A **Property Sale** represents a completed transaction in which a property is sold for a price on a specific date.

This is the **central event** around which the model is designed.

4.2 Property

A **Property** represents the classification and characteristics of the housing unit involved in the sale.

Conceptually includes:

- Property type
 - New build indicator
 - Tenure type
-

4.3 Location

A **Location** represents the geographical context of a property.

Conceptually includes:

- Postcode
 - District
 - County
 - Region (ONS-defined)
-

4.4 Time

Time represents when the property transaction occurred.

Conceptually includes:

- Day
- Month
- Quarter
- Year

5. Fact and Dimension Definitions

5.1 Fact Entity: Property Sale

Represents:

A single completed property transaction.

Characteristics:

- Occurs at a specific point in time
- Has a monetary value
- Occurs at one location
- Refers to one property classification

Measures (Metrics):

- Sale price
 - Transaction count (implicit, derived by row count)
-

5.2 Dimension Entity: Property

Represents:

The descriptive classification of the property sold.

Attributes:

- Property type (e.g., Flat, Semi-Detached, Detached)
- New build indicator (New / Old)
- Tenure (Leasehold / Freehold)

Why it is a dimension:

- Descriptive rather than transactional
 - Low cardinality
 - Used for grouping, filtering, and segmentation
-

5.3 Dimension Entity: Location

Represents:

The geographical hierarchy associated with a property sale.

Attributes:

- Postcode
- District
- County
- Region

Why it is a dimension:

- Enables spatial and regional analysis
 - Supports hierarchical roll-ups (postcode → region)
-

5.4 Dimension Entity: Time

Represents:

The calendar context of a property sale.

Attributes:

- Date
- Month
- Quarter
- Year

Why it is a dimension:

- Enables trend analysis
 - Supports Month-over-Month (MoM) and Year-over-Year (YoY) comparisons
 - Shared across multiple facts
-

6. Entity Relationships

6.1 Property Sale → Time

- Each property sale occurs on exactly one date
- Each date can have many property sales

Relationship:

- Many Property Sales → One Time
-

6.2 Property Sale → Property

- Each sale involves one property classification
- Each property classification appears in many sales

Relationship:

- Many Property Sales → One Property
-

6.3 Property Sale → Location

- Each sale occurs in one geographical location
- Each location can have many sales

Relationship:

- Many Property Sales → One Location
-

7. Conceptual Star Schema (Business View)

Time

|

|

Property — Property Sale — Location

This star schema structure enables efficient analytical queries across time, geography, and property attributes.

8. Grain Definition

Grain Statement:

One row represents one completed property transaction.

All facts and dimensions in the model strictly conform to this grain.

Grain Alignment Examples

- Sale price = price of the individual transaction
- Date = transaction date
- Region = derived via the Location dimension
- Property type = descriptive attribute, not a measure

Maintaining a consistent grain is critical for correctness, aggregation, and analytical integrity.

9. Summary

This conceptual data model establishes a clear, business-aligned foundation for:

- Dimensional modelling (star schema)
- Analytical dashboards and KPIs
- Time-series and regional analysis
- Advanced analytics and RAG-based insights

Subsequent logical and physical models will refine this structure into implementable schemas and optimized storage designs.