Logical Data Model – UK Housing Transactions

1. Purpose and Scope

This document defines the **logical data model** for the UK housing transactions dataset. It translates the conceptual model into concrete analytical structures (facts and dimensions) while remaining technology-agnostic.

The logical model is designed for:

- Dimensional analytics (Kimball-style star schema)
- BI dashboards and KPI reporting
- Downstream advanced analytics and RAG-based querying

---

2. Identification of Logical Tables

Based on the conceptual model, the following logical tables are defined.

2.1 Fact Table

- fact_property_sales

2.2 Dimension Tables

- dim_date
- dim_region
- dim_property_type

Note: Postcode data is intentionally excluded from the analytical schema and is used only in upstream staging and mapping processes.

---

3. Fact Table Definition

3.1 fact_property_sales

Grain:

One row per completed property transaction

Logical Attributes

Column Name	Logical Data Type	Description
transaction_id	string	Unique transaction identifier
date_key	integer	Foreign key → dim_date.date_key

Column Name	Logical Data Type	Description
region_id	integer	Foreign key → dim_region.region_id
property_type_id	integer	Foreign key → dim_property_type.property_type_id
price	numeric	Sale price in GBP
old_new_flag	char(1)	O = Old, N = New
duration	char(1)	L = Leasehold, F = Freehold

Keys

- **Primary Key:** transaction_id
- **Foreign Keys:**
  - date_key
  - region_id
  - property_type_id

Measures and Additivity

Measure	Additivity
price	Fully additive
transaction_count	Derived (COUNT)

---

4. Dimension Table Definitions

4.1 dim_date

Purpose:
Supports time-series analysis, trends, and period-over-period comparisons (MoM, YoY).

Grain:

One row per calendar date

Column Name	Data Type	Description
date_key	integer	Surrogate key (YYYYMMDD)

Column Name	Data Type	Description
date	date	Actual calendar date
year	integer	Calendar year
quarter	integer	Quarter (1–4)
month	integer	Month number (1–12)
month_name	string	January, February, etc.
day_of_month	integer	Day of month (1–31)
day_of_week	integer	Day of week (1–7)
is_weekend	boolean	Weekend indicator

Primary Key: date_key

4.2 dim_property_type

Purpose:
Provides categorical classification of properties.

Grain:

One row per property type

Column Name	Data Type	Description
property_type_id	integer	Surrogate key
property_type_code	char(1)	F, D, S, T, O
property_type_desc	string	Flat, Detached, etc.

Primary Key: property_type_id

Cardinality:
Very low (approximately five rows).

4.3 dim_region

Purpose:
Enables geographic and regional analysis aligned with UK standards.

Grain:

One row per region

Column Name	Data Type	Description
region_id	integer	Surrogate key
region_name	string	London, North West, etc.
county	string	County name
district	string	District name

Primary Key: region_id

Design Note:

- Postcode is used upstream for region mapping
- Postcode is not stored in the analytical fact table

---

5. Logical Relationships

From Table / Column	To Table / Column	Cardinality
fact_property_sales.date_key	dim_date.date_key	Many → One
fact_property_sales.region_id	dim_region.region_id	Many → One
fact_property_sales.property_type_id	dim_property_type.property_type_id	Many → One

All dimensions are **conformed**, enabling reuse across facts.

---

6. Logical ER Diagram (Textual Representation)

dim_date

---

date_key (PK)

date

year

month

quarter

...

dim_property_type

---

property_type_id (PK)

property_type_code

property_type_desc

dim_region

---

region_id (PK)

region_name

county

district

fact_property_sales

---

transaction_id (PK)

date_key (FK)

region_id (FK)

property_type_id (FK)

price

old_new_flag

duration

This represents a clean, single-fact **Kimball-style star schema**.

---

7. Mapping: Raw CSV to Logical Model

Raw CSV Column	Logical Table	Logical Column
Transaction ID	fact_property_sales	transaction_id
Price	fact_property_sales	price
Date	dim_date	date
Property Type	dim_property_type	property_type_code
Old/New	fact_property_sales	old_new_flag
Duration	fact_property_sales	duration
Postcode	Staging only	Region mapping
District	dim_region	district
County	dim_region	county

Region names are derived using **ONS reference data**, not directly from the raw CSV.

---

8. Key Design Decisions

8.1 Use of Surrogate Keys

- Ensures stable joins
- Decouples analytics from source system volatility
- Aligns with industry-standard data warehousing practices

8.2 Exclusion of Postcode as a Dimension

- Extremely high cardinality
- Negative impact on query and dashboard performance
- Region-level analysis satisfies business requirements

8.3 Region-Based Geography

- UK housing analytics is region-centric
- Aligns with ONS and government statistical reporting
- Enables consistent benchmarking

---

9. Analytical Capabilities Enabled

This logical data model enables efficient analysis of:

- Average price by region and month
- Monthly transaction volumes
- Property type price trends
- Year-over-Year (YoY) and Month-over-Month (MoM) growth using dim_date

The model is suitable for **production-grade analytics**, BI dashboards, and AI-assisted querying.