**UK Housing Market Analysis: Investigating Sales Trend Using HMLR Data (SQL & Tableau Project)**

**Project Goal**

“To provide data-driven insights into UK property sales trends, regional price variations, and buyer behaviour using HM Land Registry data to assist investors, financial institutions, and policymakers in decision making”.

In order to satisfy the project goal, the following business challenges will be explored:

1. How have UK house prices changed over time?
2. Which areas have seen the highest and lowest price growth?
3. How do property types impact pricing trends?
4. What is the impact of duration (Freehold Vs. Leasehold) on price?
5. How do new-build properties compare to older ones in terms of price and volume?
6. Are there any outliers or anomalies in property prices that require investigation?

**Business Context: Who Will Benefit?**

The following outlines the stakeholders who will benefit from this project, the insights they will gain from it, and how they can be applied in real-world scenarios:

* **Property Investors & Developers**: Aids in the identification of high-growth areas, regions with undervalued properties and market trends. Enables strategic investment decisions to be made into emerging locations with strong price appreciation, whilst avoiding declining markets and overvalued properties.
* **Banks and Mortgage Brokers**: Provides a data-driven risk assessment of property values and price trends, promoting the creation of competitive, adjusted mortgage policies that take the data into account.
* **Local Government & Urban Planners:** Insight gained into regional housing demand and transaction activity, which would facilitate the planning of new developments in high-demand areas e.g. housing projects & infrastructure investments.
* **Housing Policy Makers**: Understanding gained with regard to affordable housing trends and price inflation, which would allow for the more effective design of policies to improve affordability in struggling areas and regulate new developments.
* **Property Agents & Solicitors**: Insight gained into buyer preferences and regional transaction trends, lend support when advising clients on the best locations and times to buy and sell.
* **Home Buyers and Sellers:** Understanding provided with regard to affordable areas and price trends, which in turn aids buyers making buying or selling decisions, as they will be informed and data-driven based on market data.

**About HM Land Registry**

HM Land Registry (HMLR) is a UK government department responsible for registering the ownership of land and property in England and Wales. It maintains a central database of property ownership, mortgages, and rights affecting land. Key functions include:

* **Registering Land and Property**: Ensures ownership is officially recorded.
* **Providing Property Information**: Offers access to title deeds, ownership details, and property boundaries.
* **Guaranteeing Property Ownership**: Provides legal security and protection for property owners.
* **Supporting Property Transactions**: Facilitates buying, selling, and mortgaging of properties.
* **Resolving Land Disputes:** Helps clarify legal rights over land.

It plays a crucial role in property transactions, ensuring transparency and reducing fraud. Property investors or developers regularly interact with HMLR for title searches, ownership verification, or boundary issues.

**HMLR Price Paid Data Background**

Price Paid Data tracks property sales in England and Wales submitted to HM Land Registry for registration. Price Paid Data is based on the raw data released each year.

The amount of time it takes to register a sold property’s information with the HMLR can vary depending on various factors e.g. it might take longer to document a sold property’s information in Wales, than in England. The interval between sale and registration typically ranges between two weeks and two months, but occasionally, it has taken longer than two months. This means there might be a lag of up to two months with some data, therefore, November and December may be lighter than they should be.

Please note that the data excludes:

* sales that have not been lodged with HM Land Registry
* sales that were not for value
* transfers, conveyances, assignments or leases at a premium with nominal rent, which are:
  + ‘Right to buy’ sales at a discount
  + subject to an existing mortgage
  + to effect the sale of a share in a property, for example, a transfer between parties on divorce
  + by way of a gift
  + under a compulsory purchase order
  + under a court order
  + to Trustees appointed under Deed of appointment
* Vesting Deeds Transmissions or Assents of more than one property

The below table will make sense of what each column in the txt files means:

|  |  |
| --- | --- |
| Column Name | Explanation |
| Transaction Unique Identifier | A reference number which is generated automatically recording each published sale. The number is unique and will change each time a sale is recorded. |
| Price | Sale price stated on the transfer deed. |
| Date of Transfer | Date when the sale was completed, as stated on the transfer deed. |
| Postcode | This is the postcode used at the time of the original transaction. Note that postcodes can be reallocated and these changes are not reflected in the Price Paid Dataset. |
| Property Type | D = Detached, S = Semi-Detached, T = Terraced, F = Flats/Maisonettes, O = Other Note that: - we only record the above categories to describe property type, we do not separately identify bungalows - end-of-terrace properties are included in the Terraced category above - ‘Other’ is only valid where the transaction relates to a property type that is not covered by existing values, for example where a property comprises more than one large parcel of land |
| Old/New | Indicates the age of the property and applies to all price paid transactions, residential and non-residential. Y = a newly built property, N = an established residential building. |
| Duration | Relates to the tenure: F = Freehold, L= Leasehold etc. Note that HM Land Registry does not record leases of 7 years or less in the Price Paid Dataset. |
| PAON | Primary Addressable Object Name. Typically the house number or name. |
| SAON | Secondary Addressable Object Name. Where a property has been divided into separate units (for example, flats), the PAON (above) will identify the building and a SAON will be specified that identifies the separate unit/flat. |
| Street |  |
| Locality |  |
| Town/City |  |
| District |  |
| County |  |
| PPD Category Type | Indicates the type of Price Paid transaction. A = Standard Price Paid entry, includes single residential property sold for value. B = Additional Price Paid entry including transfers under a power of sale/repossessions, buy-to-lets (where they can be identified by a Mortgage), transfers to non-private individuals and sales where the property type is classed as ‘Other’.  Note that category B does not separately identify the transaction types stated. HM Land Registry has been collecting information on Category A transactions from January 1995. Category B transactions were identified from October 2013. |
| Record Status – Monthly File Only | Indicates additions, changes and deletions to the records.(see guide below). A = Addition C = Change D = Delete  Note that where a transaction changes category type due to misallocation (as above) it will be deleted from the original category type and added to the correct category with a new transaction unique identifier. |

**Dimensional Data Modelling: Schema Creation**

A star schema structure ensures efficient querying, reporting and visualisation by separating price paid data into a fact table with supporting dimension tables. This approach optimises performance for SQL queries and tableau visualisations. The concept of how the HMLR price paid dataset will be split into fact and dimension tables is as follows:

**Fact Table: FACT\_price\_paid**

This table stores all property sales data and links to various dimension tables for further analysis:

|  |  |
| --- | --- |
| Column Name | Description |
| transaction\_id (PK) | Same as ‘Transaction Unique Identifier’, just renamed. Unique identifier for each sale, as recorder in the dataset |
| price | Sale price stated on the transfer deed |
| date\_of\_transfer\_id (FK) | Foreign key linking to dim\_date. ID derived from ‘Date of Transfer’ column in original dataset. |
| location\_id (FK) | Foreign key linking to dim\_location. ID derived from various columns pertaining to location in the original dataset e.g. ‘postcode’, ‘county’ etc. |
| property\_type\_id (FK) | Foreign key linking to dim\_property\_type. ID derived from ‘Property Type’ column in original dataset. |
| old\_or\_new\_id (FK) | Foreign key linking to dim\_old\_new. ID derived from ‘Old/New’ column in original dataset. |
| duration\_id (FK) | Foreign key linking to dim\_duration. ID derived from ‘Duration’ column in original dataset. |
| pdd\_category\_type | Transaction category: Standard (A) or Special (B) |

**Dimension Table: DIM\_date**

This table allows time-based analysis such as yearly trends and seasonal patterns:

|  |  |
| --- | --- |
| Column Name | Description |
| date\_of\_transfer\_id (PK) | Unique identifier of each date |
| date\_of\_transfer | Actual sale date |
| year | Extracted year from date\_of\_transfer |
| month | Extracted month from date\_of\_transfer |
| quarter | Extracted quarter from date\_of\_transfer |

**Dimension Table: DIM\_location**

Stores location based insights for regional analysis:

|  |  |
| --- | --- |
| Column Name | Description |
| location\_id (PK) | Unique identifier for location |
| postcode | Property postcode at the time of the transaction. Same as column in the original dataset with the same name. |
| paon | Primary addressable object name (house number or name). Same as column in the original dataset with the same name. |
| saon | Secondary addressable object name (e.g., flat number). Same as column in the original dataset with the same name. |
| street | Street name. Same as column in the original dataset with the same name. |
| locality | Local Area Name. Same as column in the original dataset with the same name. |
| town\_city | Town or City where the property is located. Same as column in the original dataset with the same name. |
| District | Administrative district. Same as column in the original dataset with the same name. |
| county | County where the property is located. Same as column in the original dataset with the same name. |

**Dimension Table: DIM\_property\_type**

Stores Property Specific details

|  |  |
| --- | --- |
| Column Name | Description |
| property\_type\_id (PK) | Unique identifier for each property type. ID derived from ‘Property Type’ column in original dataset. |
| property\_type | Type of property: Detatched (D), Semi-Detached (S), Terraced (T), Flat (F), Other (O). Same as ‘Property Type’ column in the original dataset. |

**Dimension Table: DIM\_old\_or\_new**

Facilitates comparisons between new and old properties.

|  |  |
| --- | --- |
| Column Name | Description |
| old\_or\_new\_id (PK) | Unique identifier for property age. ID derived from ‘Old/New’ column in original dataset. |
| old\_or\_new | Property age: New (Y), Old (N). Same as ‘Old/New’ column in the original dataset. |

**Dimension Table: DIM\_duration**

Allows analysis of Freehold vs. Leasehold pricing differences.

|  |  |
| --- | --- |
| Column Name | Description |
| duration\_id (PK) | Unique identifier for each duration type. ID derived from ‘Duration’ column in original dataset. |
| duration | Duration type: Freehold (F), Leasehold (L). Same as ‘Duration’ column in the original dataset. |

**Star Schema Diagram**

A diagram of a type of type of text

AI-generated content may be incorrect.**Project Methodology: What Will Be Investigated?**

**Part 1: House Price Trends Over Time**

Understanding how house prices fluctuate over time helps investors and policymakers predict market trends and investment opportunities.

The following should be investigated:

* How have average house prices changed by year?
* Are prices showing seasonal trends (e.g. higher sales in summer vs winter)?

SQL Queries & Tableau Visuals:

* **SQL:** Compute the average sale price per year.
* **SQL:** Compute the average sale price per month.
* **SQL:** Compute the average sale price per season (quarter).
* **Tableau:** 
  + Line chart of house price trends over time.
  + Heatmap showing seasonal sales trends.

**Part 2: Regional Price Insights**

Understanding regional price variations helps investors, home buyers, and policymakers make informed decisions about where to invest, buy, or regulate housing markets.

The following should be investigated:

* What are the most expensive and most affordable counties / towns?
* How are property prices distributed geographically?
* Are there any emerging high-growth areas where prices are rising quickly?

SQL Queries & Tableau Visuals:

* **SQL:** Compute the average property price per county / district. Following this, rank the top 10 and bottom 10.
* **Tableau:** 
  + Bar chart plotting top 10 most expensive and affordable locations
  + Heatmap displaying price distributions across the UK

**Part 3: Property Type Comparison**

Different property types appreciate at different rates, affecting investment value and buyer affordability.

The following should be investigated:

* How do detached, semi-detached, terraced and flats compare in price?
* Which property type has the highest volume of transactions?

SQL Queries & Tableau Visuals:

* **SQL:** Calculate average price per property type (D, S, T, F).
* **Tableau:** 
  + Bar chart comparing average price by property type.
  + Stacked bar chart showing sales volume by property type.

**Part 4: Duration Impact on Pricing (Freehold Vs Leasehold)**

Understanding how duration impacts property value helps investors, banks, and home buyers assess long-term ownership costs and market demand.

The following should be investigated:

* Do Freehold properties (F) sell at a higher prices than Leasehold (L) properties?
* Are there regional differences in the proportion of Freehold Vs Leasehold sales?

SQL Queries & Tableau Visuals:

* **SQL:** Compare the average price of Freehold Vs Leasehold properties.
* **Tableau:** Boxplot comparing Freehold Vs Leasehold pricing

**Part 5: Transactions by Property Age (New Vs Old Properties)**

New-build properties may have different pricing trends and market appeal compared to older homes.

The following should be investigated:

* Do newly built properties (Y) sell at a premium compared to established properties (N)?
* How have new vs old property prices changed over time?

SQL Queries & Tableau Visuals:

* **SQL:** Compute the average price for new (Y) vs old (N) properties
* **Tableau:** 
  + Pie chart showing the proportion of new vs old property transactions
  + Time series chart showing price trends for new vs old properties

A screenshot of a white and black list

AI-generated content may be incorrect.