



House Price Data, England & Wales, 1995 to 2019

Claire, Daniel, Mohanad, Sahas, Sahith, Steven

Summary

Our objective is to comprehensively understand housing market trends, especially in fluctuating economic conditions like the 2008 crash. This includes analyzing how price changes and demand shifts impact different regions in England and Wales.

Expectation

We want to enable real estate businesses and investors to make data-driven decisions, identify profitable opportunities, and understand market risks.

ASK 1

- **Dataset Description**
- **Why this dataset ?**
- **Dataset modification**
- **Business Questions**

Dataset Description

The "House Price Data for England & Wales (1995-2019)" dataset from Kaggle, sourced from the HM Land Registry, is an archive of residential property transactions in England and Wales. Here's an overview of this dataset:

- **Period Covered:** The dataset from 1995 to 2019 covers 24 years of data.
- **Volume of Data:** It comprises a massive collection of 25 million records.
- **Data Size:** The total size of the dataset is 4.5GB.
- **Last Update:** It was last updated four years ago.
- **Variables Included:** There are 16 different variables in the dataset.
- **Record Details:** Each record represents an individual property transaction.
- **Content of Records:** These records include critical information such as the transaction date, the price paid, and other relevant transaction details.
- **Utility of the Dataset:** This dataset is a valuable resource for analyzing the evolution of the housing market in England and Wales over the specified 24-year period.



Why this dataset ?

1

The dataset will help us understand the impact of fluctuating times in the real estate market, offering a meaning when the market is in economic uncertainty.

2

Identify areas with increasing or decreasing property values to guide us in making informed decisions about where to invest.

3

The dataset's tidy, clear descriptions and reasonable dimensions make it manageable and readily available for analysis.

4

The dataset is well-suited for dimensional modeling and data analysis and clearly describes house prices with the details of property type, location, year, and other information.

Dataset modification

Our examination of the dataset indicated that the period from 2006 to 2010 is particularly significant for studying the impact of the 2008 financial crisis on housing prices.

Concerns with the data

Data Quality:

- Addressing missing or inconsistent data through cleaning and preprocessing.

Privacy:

- Ensuring that the dataset does not contain personally identifiable information.

Scalability and Performance:

- Implementing efficient data processing techniques for handling a large volume of data.

ASK 2 – Data Wrangling

Cleaning and Removing

- Drive from the date attributes the following (year, month, week of the year, and quarter of the year) and Remove the Hour: Minute from time values because it is always zeros.
- Record status removed
- Remove unnecessary locality column with many missing values
- Removing the unique identifier
- Check the Null Values
- Remove Postal Codes, SAON, PAON, and Streets We removed the data since our analysis focuses on the county, district, and city levels to simplify the data and focus on higher-level geographic areas. Additionally, updating the table with postal code data required over 45 minutes, making it inefficient for our purposes.
- Final dataset: 4.5 million records and 14 attributes

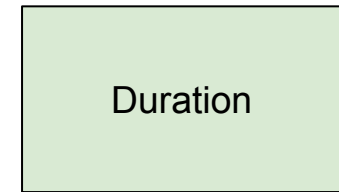
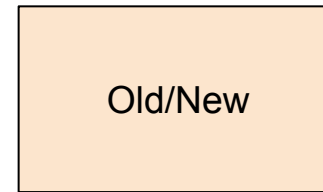
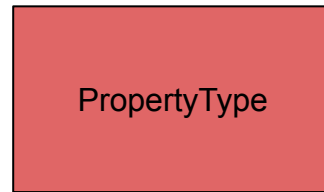
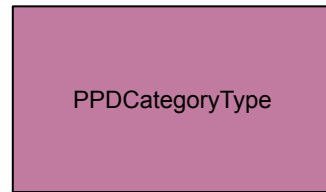
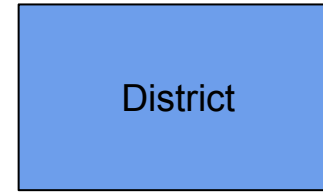
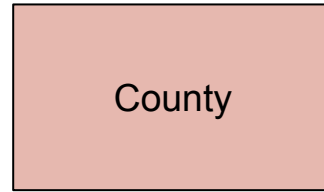
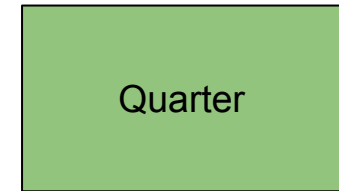
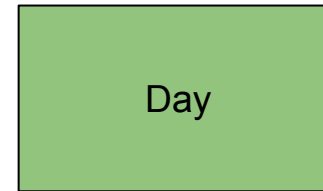
ASK 2 – Dimensional Modeling

Facts and Dimensions

Facts



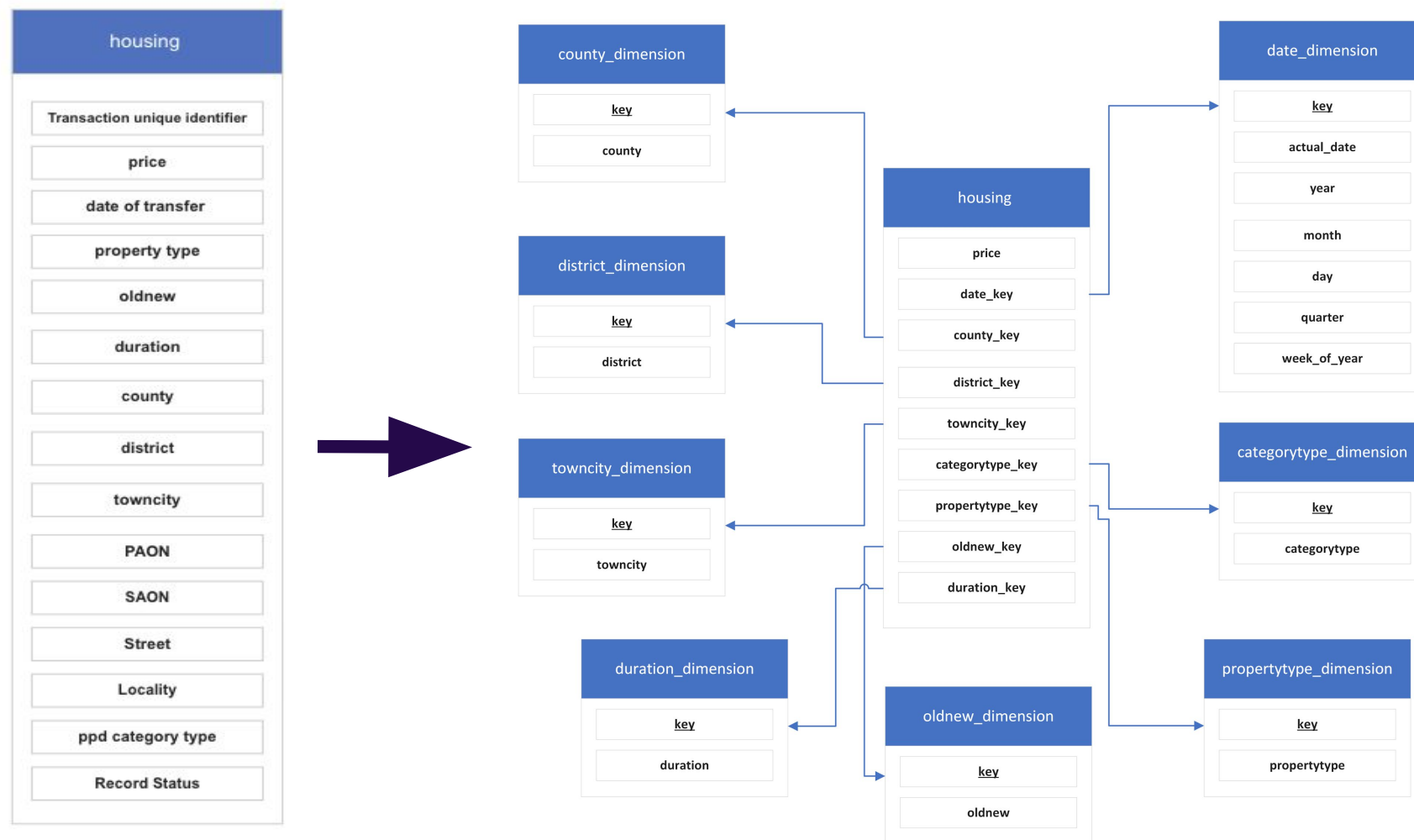
Dimensions



ASK 2 – Dimensional Modeling

For clarity, duration is categorical indicating whether a property sale is freehold, leasehold, etc.

We made the decision to use multiple dimension tables for location instead of a single dimension table for two main reasons. First, we decided it would be better to reduce redundancy by not repeating the town/city, district, or county data. Second, we want to do hierarchical analysis such as at the county and district levels.



ASK 3 - Data analysis

Business Questions

- Question 1 (description- answer - graphs - findings)
- Question 2 (description- answer - graphs - findings)
- Question 3 (description- answer - graphs - findings)

Business Questions - Question 1

Question 1

Give each county, district, and town city's total sales prices before and after 2008. Show answers by the quarter of the year, month ?

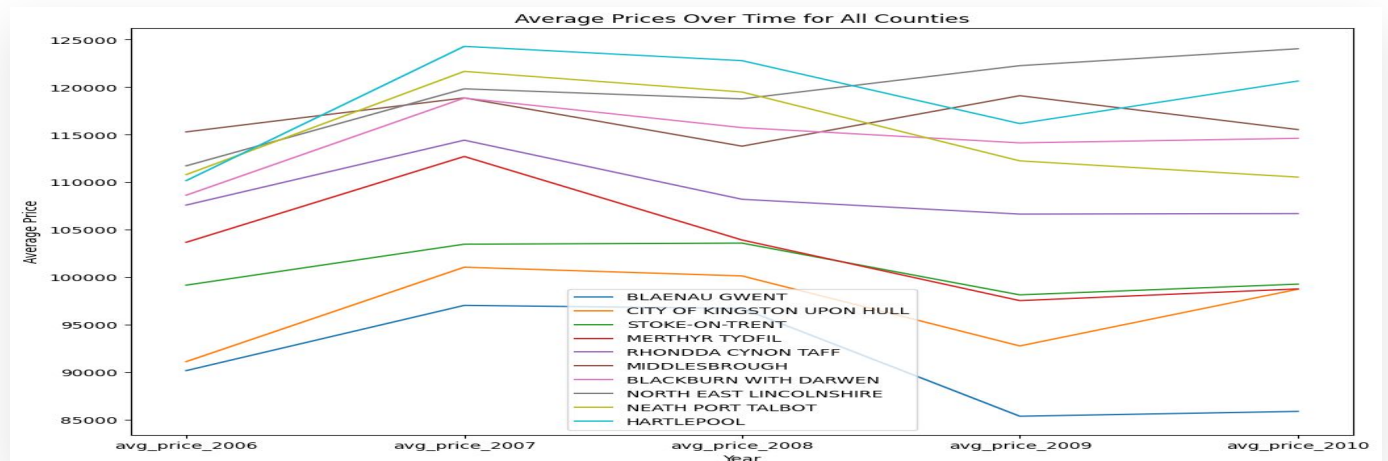
Description Q1

The goal of the question is to understand the impact of the 2008 events on real estate markets across various regions by examining sales prices, enabling a detailed comparison of market over time and across geographical areas, and providing essential information for decision-makers.

Answer 1

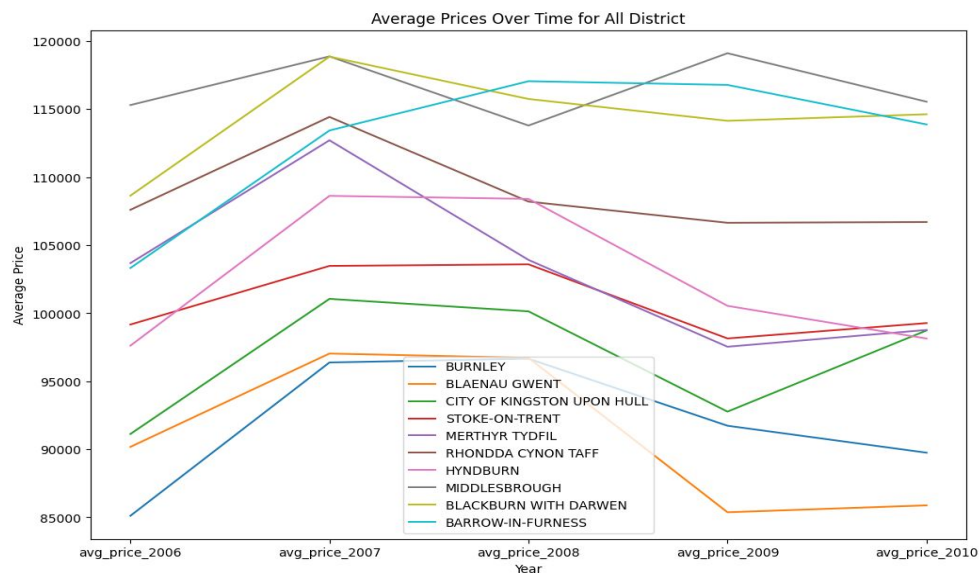
The result show of the top 5 counties with the lowest average sale prices (2006-2010) after 2008 to help us understand the actual time impact duration of 2008.

1. Blaenau Gwent
2. City of Kingston upon Hull
3. Stoke-on-Trent
4. Merthyr Tydfil
5. Rhondda Cynon Taff

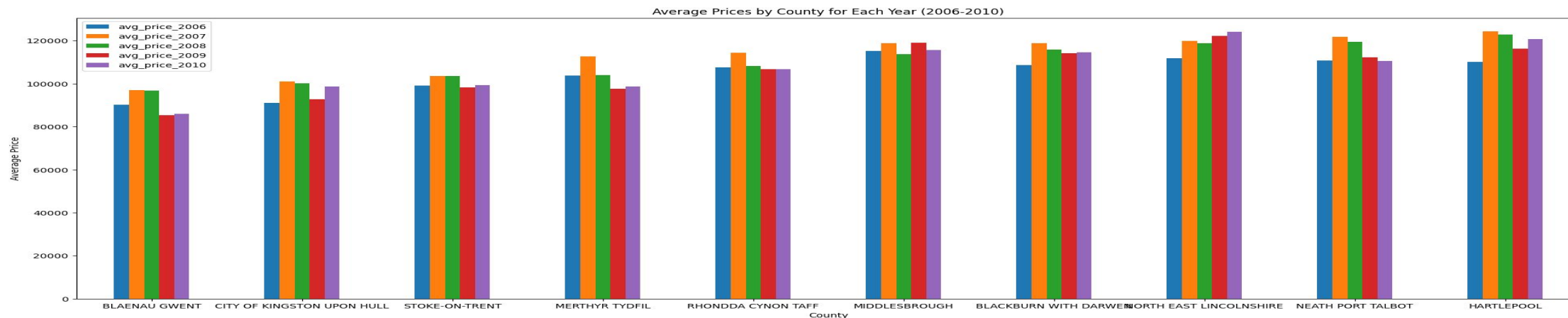
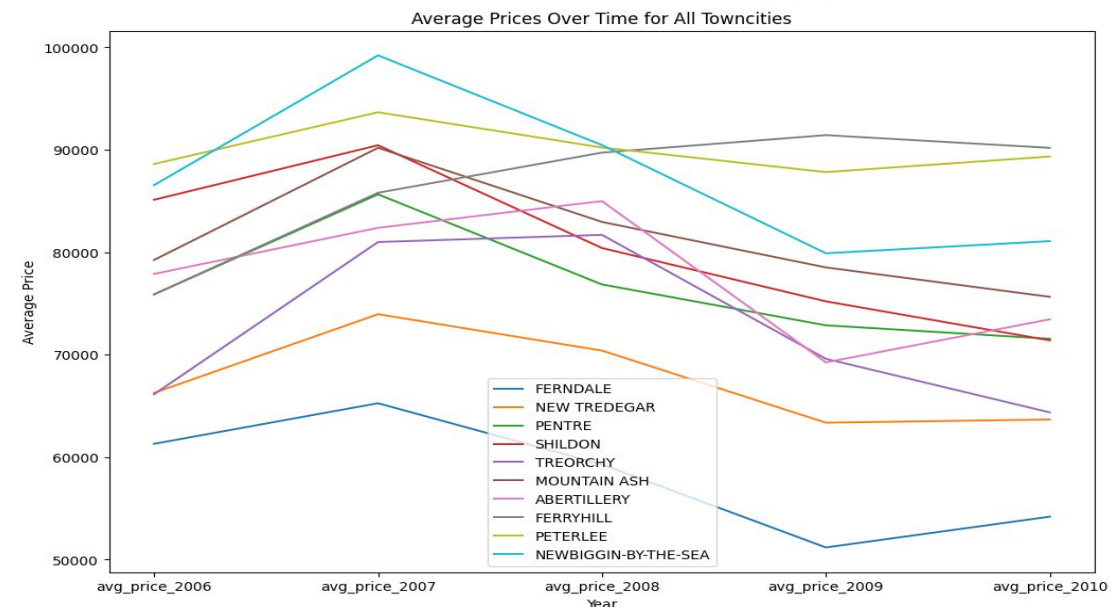


Answer 1 - Graphs

Average Prices by County for Each Year (2006-2010)



Average Prices Over Time for All Towncities (2006-2010)



Answer 1 - Findings

The analysis indicate that the last quarter of 2008 and the first quarter of 2009 significantly impacted the timing in the affected regions (country, district,town-city)

Some counties like the City of Kingston upon Hull and Stoke-on-Trent, which recovered quickly post-2009, demonstrate economic resilience.

Towns show a greater diversity in property price trends compared to counties and districts. This is likely due to the smaller scale of towns.

Some towns are located within the counties and districts. For example, Ferndale, Treorchy, Mountain Ash, and Abertillery are in Rhondda Cynon Taff. This geographic relationship means that economic and market trends in a county or district can impact the property prices in its towns.

Business Questions - Question 2

Question 2

What is the most impacted county, district, and town city? Are there any details dramatically impacted, such as (Property Type, Old/New, Duration, type of Price Paid transaction) What are your observations?

Description Q2

The goal of the question is to identify the most affected areas and property characteristics in real estate post-2008, focusing on specific regions and property details like type, old/new, and transaction conditions to gain insights into market changes.

Answer 2

After we identified the actual impact duration in question 1 (from the final quarter of 2008 to the first quarter of 2009), we have identified the most affected in and show the top 5 counties with net change per year:

	County	Average Price Before 2008 (£)	Average Price After 2008 (£)	Average Price Difference (£)	Percent Change (%)
1	Cheshire East	291,478.08	223,577.47	-67,900.61	-30.37
2	Cheshire West and Chester	240,822.56	196,770.46	-44,052.10	-22.39
3	County Durham	152,417.73	127,158.91	-25,258.82	-19.86
4	Central Bedfordshire	247,348.95	216,129.17	-31,219.78	-14.44
5	Newport	159,239.42	146,140.74	-13,098.68	-8.96



Answer 2

Most Impacted Districts

district	avn_bef ore 2008	avn_aft er 2008	avg_price _differenc e	Percent Change
SOMERSET WEST AND TAUNTON	309888. 89	163832 .50	-146056.39	-89.15
CHESHIRE EAST	291478. 08	223577 .47	-67900.61	-30.37
EAST SUFFOLK	256125. 00	208590 .00	-47535.00	-22.79
CHESHIRE WEST AND CHESTER	240425. 59	196770 .46	-43655.13	-22.19
WILTSHIRE	287213. 97	236578 .30	-50635.67	-21.40

Most Impacted County

county	avn_bef ore 2008	avn_aft er 2008	avg_price _differenc e	Percent Change
CHESHIRE EAST	291478. 08	223577 .47	-67900.61	-30.37
CHESHIRE WEST AND CHESTER	240822. 56	196770 .46	-44052.10	-22.39
COUNTY DURHAM	152417. 73	127158 .91	-25258.82	-19.86
CENTRAL BEDFORDSHIR E	247348. 95	216129 .17	-31219.78	-14.44
NEWPORT	159239. 42	146140 .74	-13098.68	-8.96

Most Impacted Town and city

towncity	avn_bef ore 2008	avn_aft er 2008	avg_price_ difference	Percent Change
NEWCAST LETON	175200.0 0	114000. 00	-61200.00	-53.68
ST NEOTS	301997.0 0	204754. 73	-97242.27	-47.49
ST ASAPH	246444.4 4	185796. 11	-60648.34	-32.64
ST HELENS	161945.8 8	122735. 55	-39210.33	-31.95
HINTON ST GEORGE	454071.4 3	349226. 19	-104845.24	-30.02

Answer 2 - Findings

We have identified the actual period of decrease from September 2008 to April 2009. The prices varied significantly across different months, reflecting the instability and regional differences in the housing market during this economic downturn.

Investigating the old and new houses average sale prices during the crash, some counties like Blaenau Gwent might not have decreased because of the economy, but investors started to buy old houses.

The city of Ravenglass experienced the most significant positive change with a substantial 45.25% increase in property value after the 2008 Financial Crisis. Which can be can valuable to both the government to implement policies and for investors.

Districts showed a significant shift in average home prices before and after the 2008 financial crash, reflecting the diverse impacts on local housing markets. With a trend of a major decrease starting December 2008 and following it into the first quarter of 2009.

Business Question– Question 3

Question 3

What are the highest sales prices per month for the town city of LONDON for all years?

Description Q3

The goal of the question is to provide insight into London real estate, helping to understand how property values and market change over time and identifying areas where property values are highest to help investors and real estate make decisions regarding investment or development in areas of London.

Answer 3

Year	Month	Sum of Prices	Number of Sales
2006	6	3,423,223,633	9495
2006	7	3,622,205,276	9837
2006	8	3,687,538,639	10192
2006	9	3,495,646,897	9468
2006	12	3,298,585,462	9021
2007	3	3,369,982,246	8751
2007	6	3,765,645,180	9201
2007	7	4,221,486,797	9648
2007	8	4,158,778,819	9877
2007	9	3,463,064,119	8219
2008	8	1,496,188,528	3569
2008	9	1,376,241,809	3052
2008	10	1,271,232,246	3092
2008	11	964,004,376	2364
2008	12	1,007,005,905	2607
2009	1	774,629,298	1867
2009	2	784,755,248	1930

Answer 3 - Findings

Based on our analysis we can conclude that housing pricing slowly increased from the middle of 2006 until they peaked mid 2007.

After July 2007 housing prices begin to decrease. It is not until August 2008 that we see the housing sum of prices decreases by 2 billion pounds.

Prices continue to drop hitting its lowest point January of 2009 at a sum of 774,629,298 pounds.

Our dataset does not contain the information of all months from January 2006-January 2009 but we can clearly see the pattern leading up to the 2008 housing crisis.