# UK Housing EDA

## Data Analyst: Mark Strathie

## Client/Sponsor: Google Certificate Final Project & UofG Data Journalism Honours Project

## Purpose:

*Write a brief description of why this project is happening below. Why is this project happening? What are the goals?*

The goal of this project is to answer 3 questions:

1. **How difficult is it to buy a home in the UK?**
2. **Why is this the case?**
3. **How do mortgages factor in?**

The target audience is the average UK population and there will be a focus on how mortgages affect the issue. This question is posed due to frequent news articles about a ‘UK housing crisis’ and in particular, a recent article featuring a quote from a NatWest chair who claimed it’s not “that difficult” to get on the property ladder. Is he right or wrong? I believe the majority of articles would benefit from focussed discussion of available data and incorporation of visualisations to make their case to the UK public, so they are better informed. This project will answer the question in this data-orientated manner to avoid sensationalism and present the raw facts. Main causes will be identified by this article so solutions can be targeted more effectively. Existing UK datasets will be used including, but not limited to, housing prices, housing availability, mortgages, inflation, and wages. I would like to analyse a time period of at least 1 generation (20-30 years). This project will be an exploratory data analysis (EDA) to attempt to identify any connections between these datasets by aggregating them and to specify the difficulty level of buying a home in the UK, and then propose why this is the case. The final deliverables will be a Jupyter notebook to demonstrate my skills and knowledge for the Google Data Analytics Professional Certificate, and a data journalism article showing some bivariate analyses and a report for my honours project at the University of Glasgow (UofG).

Main questions the project will answer constructed using the SMART framework:

**Specific**: Is there a connection between any of the datasets covered by my analysis? Housing price vs housing availability? Mortgage rate vs inflation? Difficulty vs wages? Etc.

**Measurable**: What percentage have housing prices risen over the last 30 years? How does this rate of change compare to inflation and rate of wage increase?

**Action-oriented**: If housing availability is increased will housing prices fall? If mortgage rates increase will housing prices fall?

**Relevant**: Based on the datasets, how can predictions about housing price be made given the conditions?

**Time-bound**: What is the date range for these datasets? Does the UK public need an understanding of this data for future housing decisions?

## Scope / Major Project Activities:

*What are the major parts of this project? List out the high-level steps, activities, or stages of the project, and give a brief description for each.*

|  |  |
| --- | --- |
| Activity | Description |
| Asking questions | Collect viewpoints and angles from various news articles to gain high-level understanding of issue and ideas of how to conduct EDA and present my final article. First article regarding NatWest chair was found online at The Guardian. |
| Prepare data | Download various datasets relevant to my project – housing prices, availability, mortgages, inflation, wages, etc. Look for open data on UK government websites and public data on Kaggle. Data will be stored on my PC and backed up to GitHub. Git will be used for version control. |
| Process data | Complete data cleaning and data validation checks to ensure integrity of data gathered. May need to re-visit collection phase. Use SQL and Python/Pandas. |
| Analyse data | Analyse data to identify correlations or other areas for further investigation. Aggregate data to spot trends. Draft visuals to help using Matplotlib/Seaborn. Perform bivariate analyses. |
| Draft article | Create report of findings on housing difficulty, reasons, and relation to mortgages. Draft into journalism article. Craft high quality visuals for report using Matplotlib/Seaborn and Tableau. |
| Produce final deliverables | Finished article, finished notebook, write project report. |

## This project does not include:

*Specify the things that this project isn’t responsible for doing (out of scope). For instance, “this project does not involve a summation of 2019 data analysis”*

* Using Excel, R, or a focus on any other software or Python libraries
* Solutions to the housing crisis. This is beyond the scope of my knowledge
* No focus housing data before 1945 or predictions of the future
* Extremely detailed analysis. This project is for a high-level overview of the housing market in the UK for a small data-journalism article aimed at the UK public.

## Deliverables:

*A specific list of things that your project will deliver.*

|  |  |
| --- | --- |
| Deliverable | Description/ Details |
| Jupyter notebook | Full notebook of EDA process using SQL, Python, Matplotlib, and Seaborn, as a minimum. Beginner level. Available on GitHub. |
| Tableau dashboard | At least 1 dashboard to practise using Tableau. This will be stored on my public Tableau profile, and I can share the link in my article and on my GitHub. |
| Data-journalism article | A finished new article discussing the difficulty of buying a house in the UK and getting a mortgage. Approx 1000 words. Complete with at least 3 visualisations. At least 1 of them from Tableau. Word document? Added to GitHub. |
| Final report | A final report detailing the whole EDA process to submit for my UofG honours project. Word document. Added to GitHub. |

## Schedule Overview / Major Milestones:

*The expected schedule for the project. This can be defined by milestones (e.g. “all data is cleaned and processed”), periods of time (“Week 1 / Week 2”), or other ways based on the needs of the project.*

|  |  |  |
| --- | --- | --- |
| Milestone | Expected Completion Date | Description/Details |
| *Scope of Work (SOW)* | *Week 1* | *Ask* ***SMART*** *questions and create a* ***SOW*** *for the project. Research articles to help narrow my topic. Start a* ***data journal*** *that will be used for bulk of my final report.* |
| *Data Collection and Review* | *Week 2* | *Collect initial datasets relating to topic and review them for suitability.* |
| *Learn data analysis processes and models* | *Week 3* | *Apply knowledge from courses to my project.* |
| *Data preparation 1* | *Week 4* | *Data cleaning, integration, and transformation. Multiple datasets to process. Start a* ***changelog****.* |
| *Data preparation 2* | *Week 5* | *Data cleaning, integration, and transformation. Multiple datasets to process.* |
| *Data Analysis and Insights* | *Week 6* | *Perform bulk of analysis and findings here in order to prepare presentation for next week. Then can fine tune from here. Identify connections between factors that influence housing prices and mortgage rates.* |
| *Project workshop* | *Week 7* | *A presentation of my project so far to my peer group for feedback.* |
| *Descriptive statistics* | *Week 8* | *Learned statistical methods to analyse and describe my data better.* |
| *Data visualisation* | *Week 9* | *Create high quality visuals in Python and Tableau to be used in my notebook and article.* |
| *Data journalism* | *Week 10* | *Focus on how to present all my findings in a journalistic article. Write approx. 1000 words.* |
| *Submissions* | *Week 11* | *Submit to UofG completed DOCX article and report. Upload my final notebook to my GitHub data analysis portfolio. Share my Tableau dashboard.* |

## \*Estimated date for completion:

*This is my “if all goes well and I have everything I need, this is when I’ll be done” date.*

Friday 22nd March 2024 – submission deadline