REPORT WRITING

Data Science For Developers

Prepared by

Samir Nath Arjyal

University Id:15373179

Prepared For

Mr.Sidhartha Neupane

B.S.c(Hons) Computing,Softwarica Collegeof IT and E-commerce ,Coventry University ST5014CEM Data Science for developers

26th Jan 2026

Table of Contents

Introduction

Background and Context

Data science is having an increasing role in aiding individuals in making well-informed decisions by really getting into very large, complex sets of information. This coursework covers an overview of each of the steps necessary in real life while working with data scientists, with an emphasis placed on how exactly we achieve each of these steps—not necessarily what they look like in an end presentation. Using R, you can access publicly accessible data sourced directly from UK government information systems in preparation for an understandable, structured, and reproducible approach with data analysis itself explained in each step of the way.

The project scenario involves property investors looking to invest in property in Cheshire or Cumberland in the UK. The investors' interests lie in the affordability of houses in the areas they choose the speed at which they can access internet connectivity for remote work from home, and the levels of crimes in the areas they wish to invest in. The UK government provides multiple datasets which cover various regions of Cheshire and Cumberland while a basic recommendation system identifies the most suitable locations.

Study Objectives

To attain this goal, the following questions are addressed by this study:

- What differences can be identified in property price between towns in Cheshire and Cumberland?
- What are the differences in broadband availability and quality across those areas?
- Is a correlation between crime rates and housing prices that we can, see?

- Which towns come with the package of affordable homes, strong broadband, and low crime scenarios?

Analytical Approach

This study uses descriptive statistics as well as simple exploratory techniques to compare the cost of houses, speed of internet connectivity, and crime rates in various towns. The towns utilized to implement grading systems to propose the investment locksmith service.. R is utilized as the programming language in processing data visualizations. It includes tidyverse packages.

Data Cleaning and Preprocessing

Data cleaning is one very important step needed to prepare this project, since every proper analysis requires good quality and homogenous data. Given that datasets come from UK government sources and the ONS Open Geography Portal, quite a few preparation steps had to be performed before any analysis could be done. Those steps ensured that the data were accurate and lined up consistently so it could be combined across the different sources.

House price data processing

House price data came from the UK Land Registry Price Paid dataset, which was used for gauging affordability in Cheshire and Cumberland. We restricted this dataset to only those transactions occurring within these two counties and cleaned it by removing any records showing missing or invalid price or postcode details. Duplicates were also removed to prevent duplicate counting in analysis. Since the data was rather structured and consistent, a little extra processing was necessary before continuing.

Broadband Speed data processing

For measuring how the towns in Cheshire and Cumberland connect to the Internet, we draw upon openly available broadband speeds published by the government of the UK. We filtered the data set to the areas we are looking at and removed records with missing or invalid data for the speed. Location identifiers were standardized so they can be merged with house price and crime datasets easily. We only kept the Average Download Speed and the geographic location fields for ease of integration and rapid comparison.
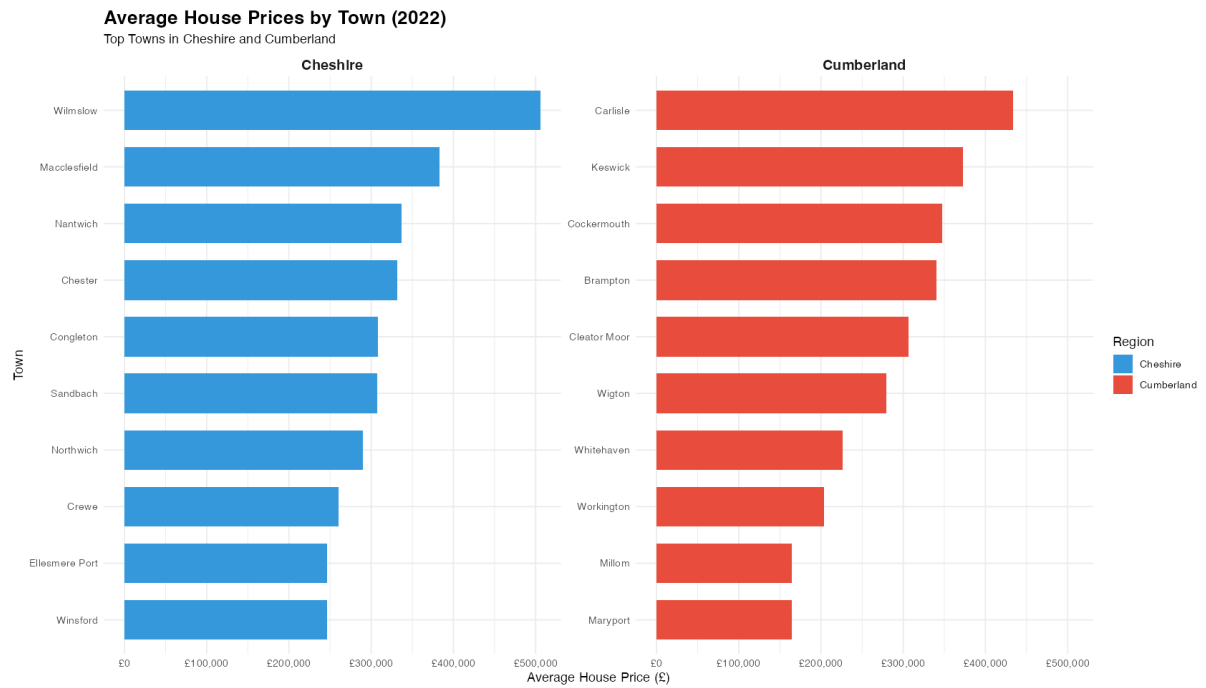
Crime data Processing

Crime data was provided by public UK government records to determine how safe each town in Cheshire and Cumberland is. We filtered the dataset to include only the crimes within the selected counties and the time of interest. Any rows with incomplete or missing location details had to be omitted, while the crime types were grouped in a way that we could get an overall, clear crime rate at the town level. The location fields had to be standardized so that they could match other datasets and integrated smoothly with housing and broadband data for further analysis.
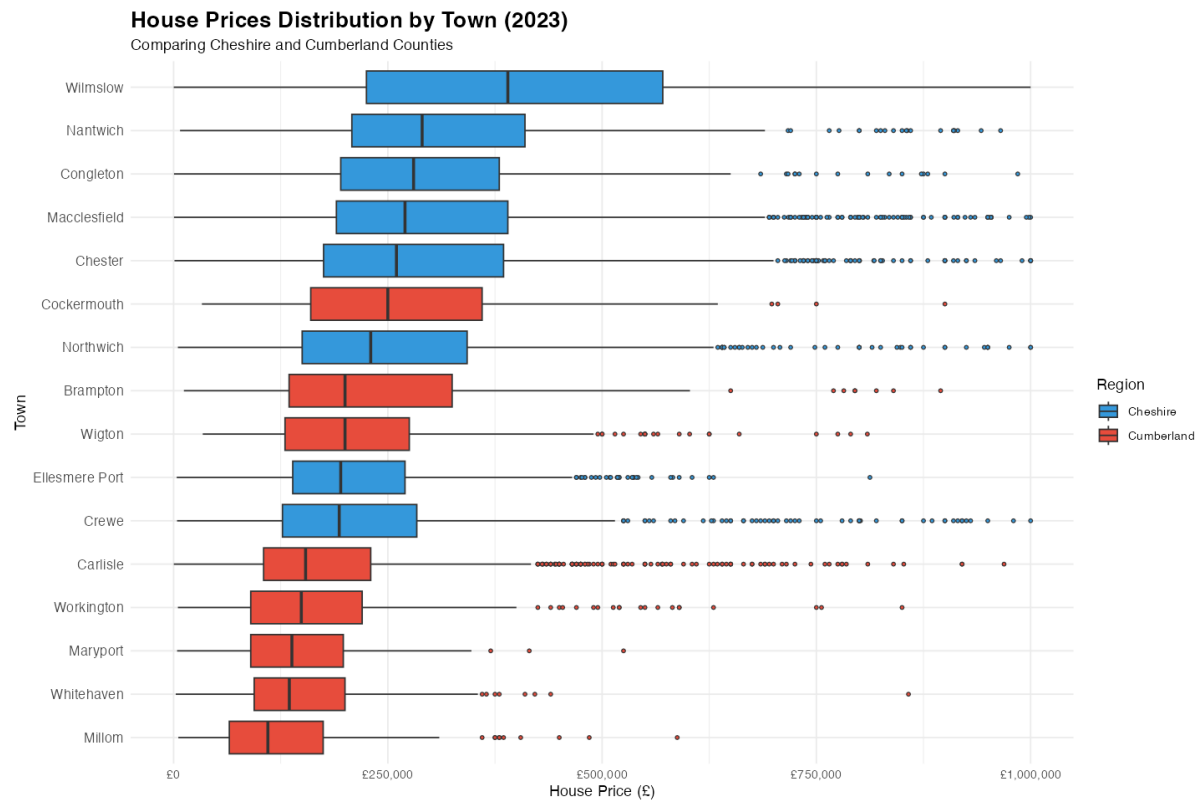
Exploratory Data Analysis (EDA)

The researchers conducted EDA to better comprehend how their combined dataset operated before they started their advanced analysis work. The research examined how residential property values and internet service speeds and criminal activity rates differed between various towns throughout Cheshire and Cumberland. Summary statistics together with basic visual displays enabled users to compare different regions while tracking overall patterns and discovering which areas excelled and which areas underperformed measurement categories. The EDA process helped identify nonstandard data points which confirmed that the dataset met requirements for both subsequent analysis and development of an effective recommendation engine.

House price data processing

The study analysed house prices from different towns in Cheshire and Cumberland to determine their affordability differences. The study calculated summary statistics through minimum and maximum and average price measurements and used visualizations to compare various towns. The process revealed that different property value areas which showed different value levels throughout the study area while identifying the towns which presented the most valuable investment opportunities. The process enabled researchers to discover areas with varying property values which displayed different value patterns based on their geographic location. The process enabled researchers to discover areas with varying property values which displayed different value patterns based on their geographic location.

Average House Prices by Town (2022)
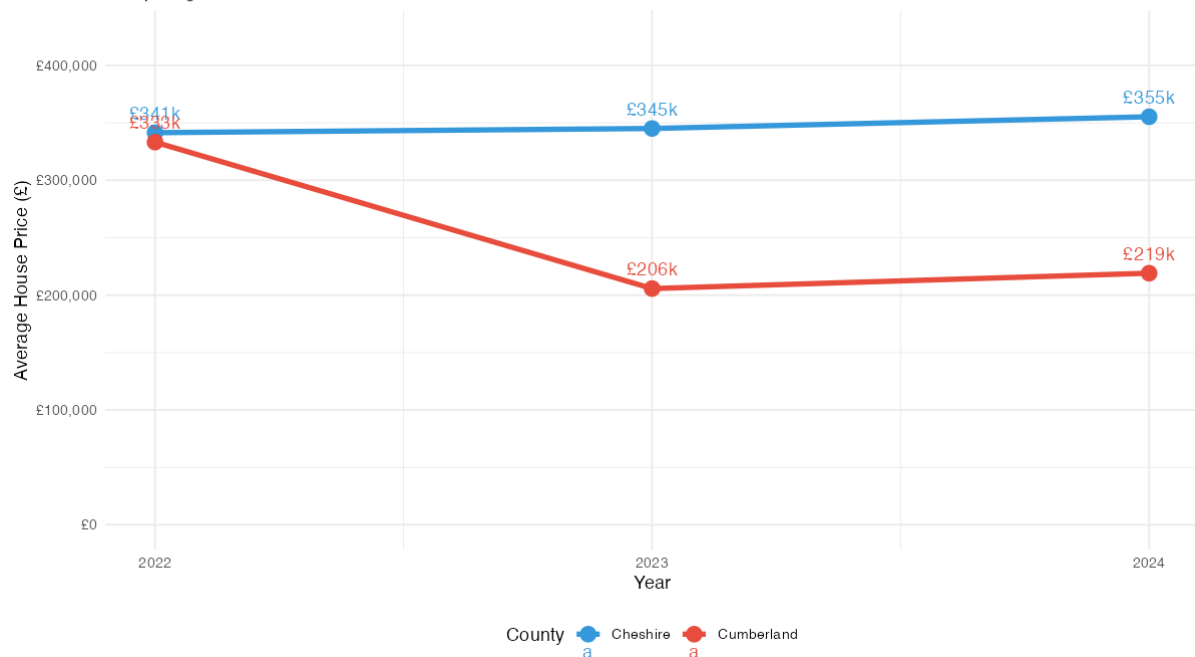Top Towns in Cheshire and Cumberland

The study evaluates the ten most expensive towns from every county. The town of Wilmslow in Cheshire blue color designation has higher property prices which reach approximately £500000. The Cumberland red color designation provides better value because Carlisle reaches a property value of approximately £500000 yet most towns offer lower prices. The research demonstrates that Cheshire displays higher overall housing market costs.

House Prices Distribution by Town (2023)
Comparing Cheshire and Cumberland Counties

The boxplots display the price distributions. Wilmslow has the highest prices because it displays multiple £1M+ price outliers. The Cumberland towns of Millom Whitehaven and Maryport all group together because their prices stay below £250k. The system demonstrates price distribution patterns by showing both average prices and their respective price variations in each town.
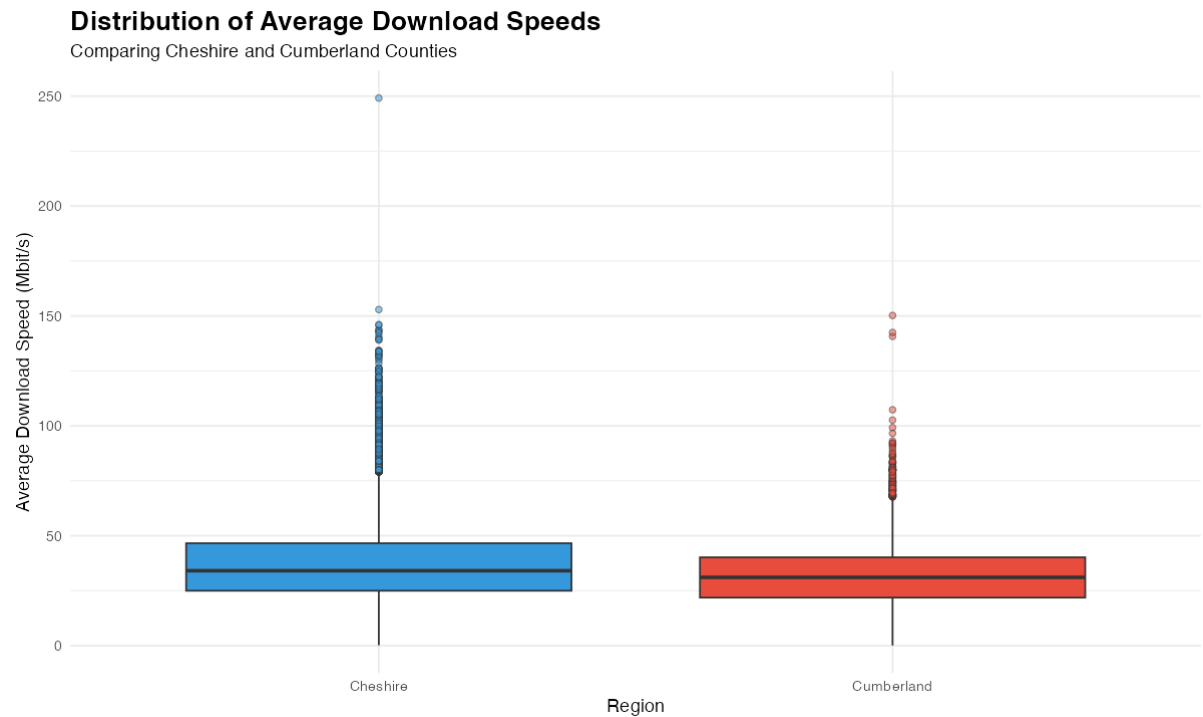
Figure 5: Line Graph Showing Yearly House Price by County (2022-2024)
Comparing Cheshire and Cumberland Counties

So,In Cheshire the value remains fixed which is between £340 and £355K. The major decrease is in Cumberland which is from £341K to £206K and rises till £219K. The data shows that Cheshire keeps its market conditions stable while Cumberland experiences a total market breakdown.
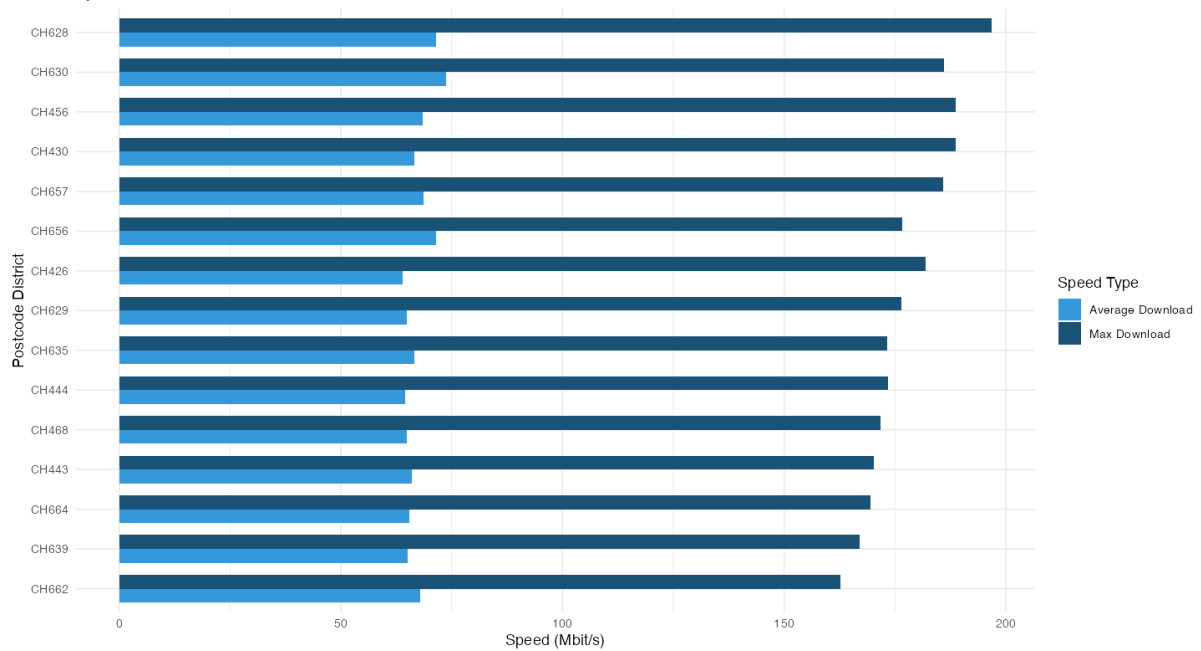
Broadband speed data processing

Researchers examined broadband speeds to assess which towns in Cheshire and Cumberland had better internet connectivity. Through visualizations the team calculated the average speed to show which areas had the access of internet faster also which areas had slower internet access

Distribution of Average Download Speeds

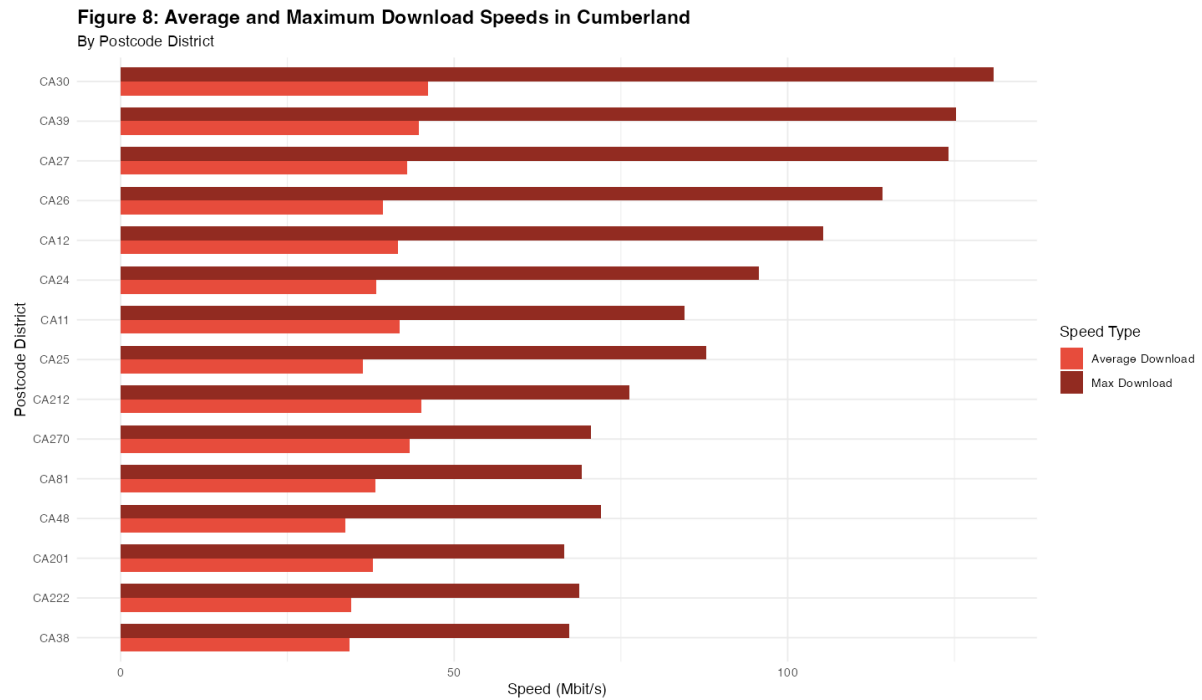Comparing Cheshire and Cumberland Counties

The two counties show equal median speeds which measure around 30 to 35 Mbit/s. Cheshire experiences higher speed variation because its extreme speed values reach above 250 Mbit/s. The distribution in Cumberland shows greater concentration because it has fewer extreme measurements. The broadband quality across all counties shows similar performance between the two regions.

Figure 7: Stacked Barchart of Average and Maximum Download Speed in Cheshire
By Postcode District

The fastest 15 postcode districts show their walking speeds from CH626 which records 200 Mbit/s as its highest speed. Light purple shows the average speed which falls between 50 Mbit/s and 70 Mbit/s while dark purple shows the maximum speed which ranges between 120 Mbit/s and 200 Mbit/s. The show demonstrates that across different districts there exists a significant difference between their normal operating speeds and their highest achievable speeds.

Figure 8: Average and Maximum Download Speeds in Cumberland
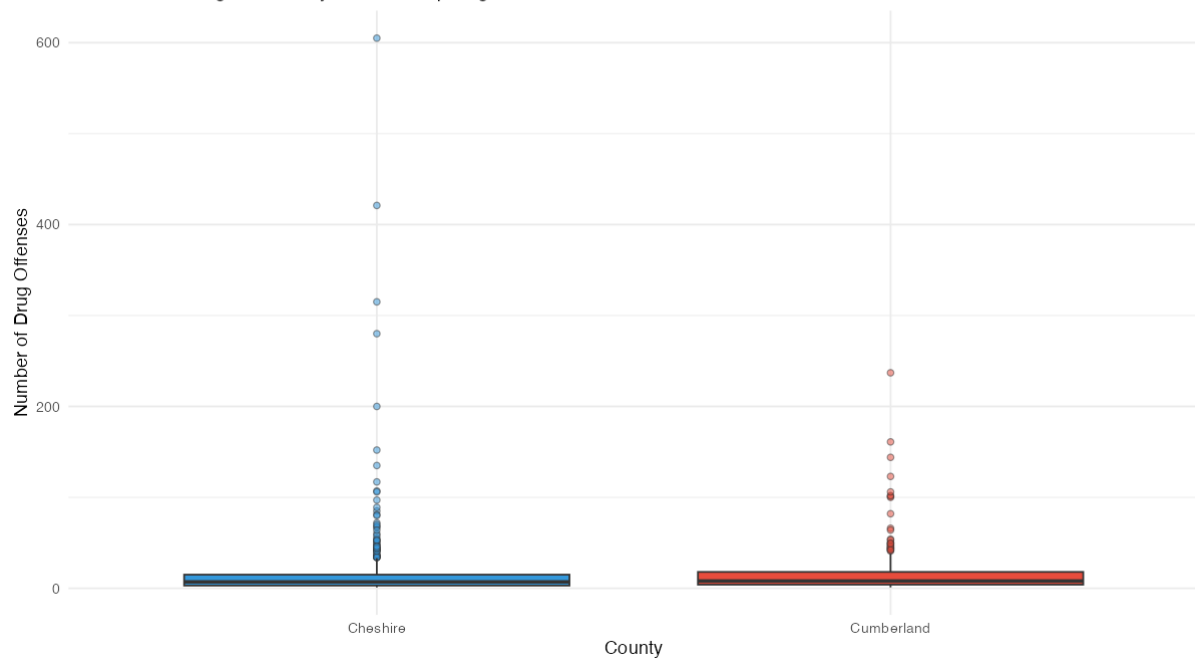By Postcode District

The CA30 district of the region provides its residents with the fastest internet speed which reaches 200 Mbit/s. The majority of districts in the area provide internet speeds which fall within the range of 40 to 70 Mbit/s and their maximum speeds extend from 80 to 180 Mbit/s. The top areas in Cumberland deliver broadband performance that matches the capabilities of Cheshire, which enables both counties to provide dependable high-speed internet service.

Crime data processing

Researchers examined crime data to determine safety levels in Cheshire and Cumberland towns. The researchers used total crime rate data to create visual displays which allowed them to compare crime levels between different towns and identify areas with higher and lower crime rates. The research indicated which towns provide a safer environment for potential property investors.
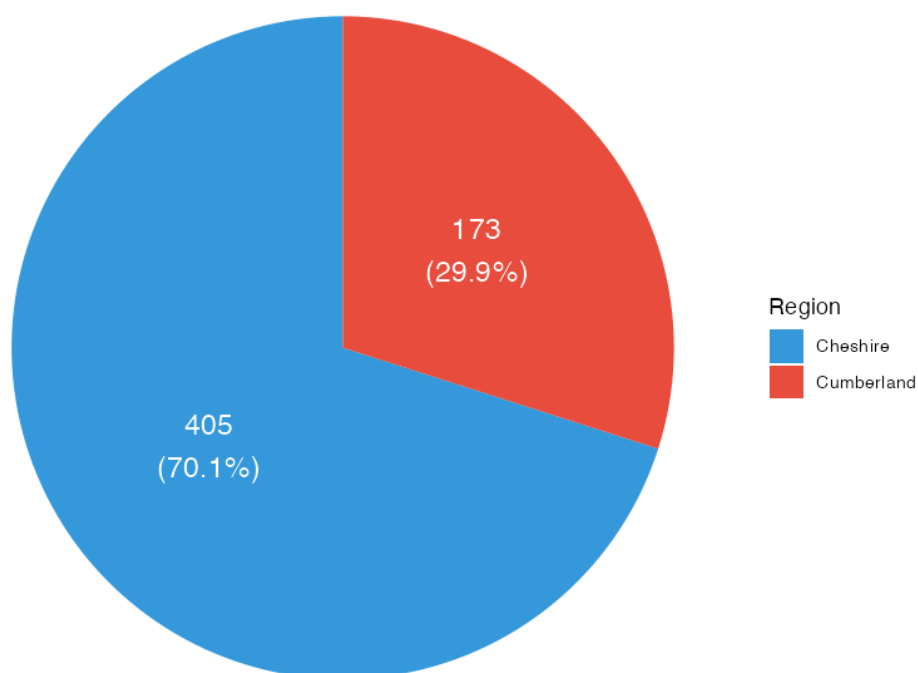
Figure 9: Drug Offence Rate by County

Distribution of Drug Offenses by LSOA - Comparing Cheshire and Cumberland

Cheshire (blue) has higher drug offense counts because its drug violations include outliers who committed more than 600 offenses but Cumberland (red) shows lower median values and fewer extreme cases. The analysis shows that drug-related problems in certain Cheshire districts have reached more serious levels than the Cumberland area which displays its criminal activities through high crime spots.
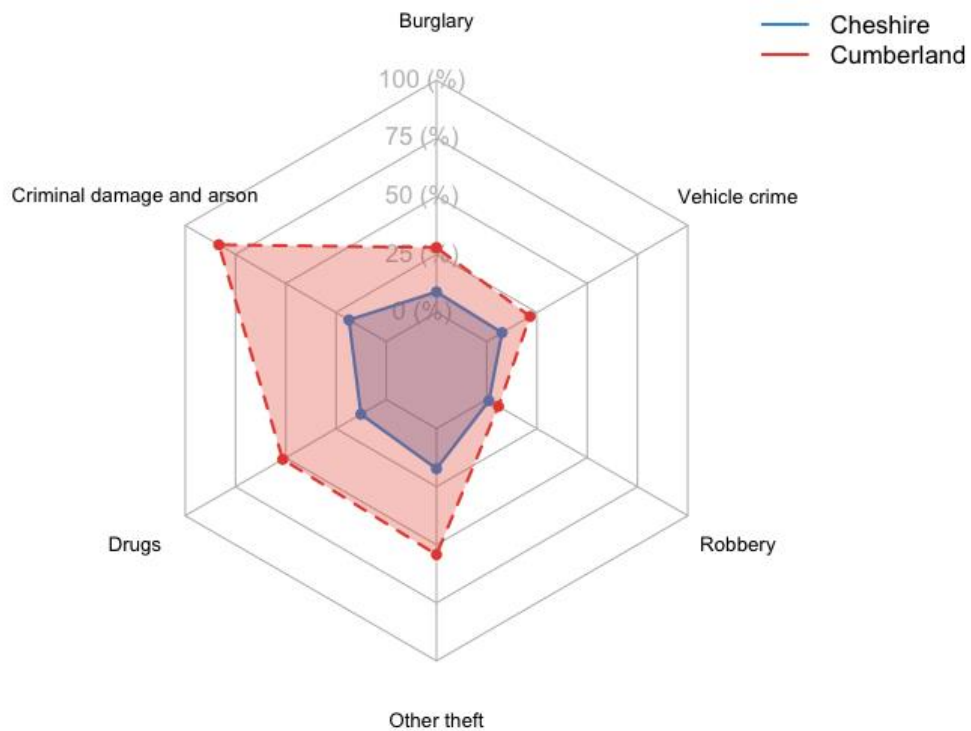
Robbery Incidents Distribution (2023)

Cheshire vs Cumberland

So,70.1% of the total units which amount to 405 units carried by Cheshire while 29.9% of the total units which amount to 173 units is carried by Cumberland The dataset shows larger distribution from Cheshire because of its larger population and greater number of towns which exist in its territory, while Cumberland presents smaller distribution. The visual display shows how the two regions differ in size, which helps viewers compare their house prices and broadband speeds and crime rates.

**Figure 10: Radar Chart of Vehicle Crime in December 2023
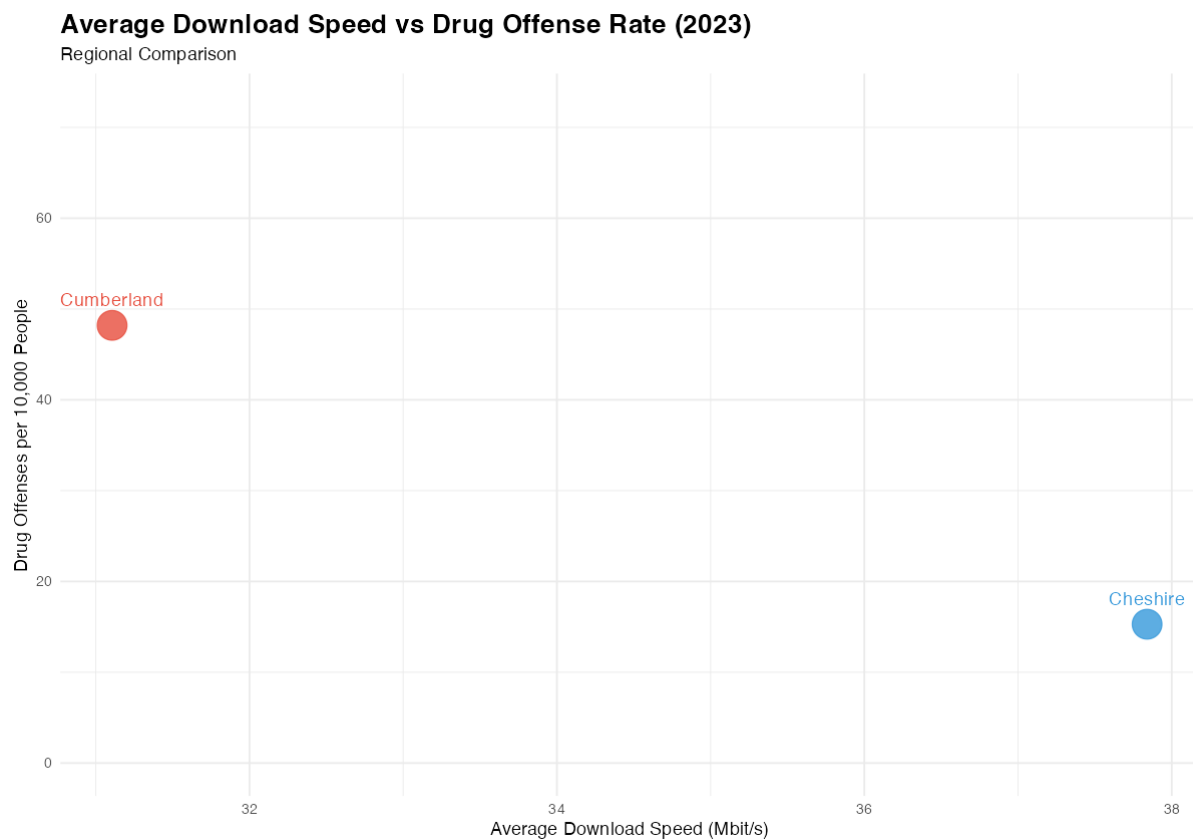(Crime Rate per 10,000 People)**

The study evaluates six different types of criminal activity. Cumberland shows superior performance in most test categories while its two specific areas of drug-related offenses and criminal damage and arson and other theft remain as its two lowest scoring test areas. Cheshire (blue/solid) displays lower crime rates across all categories. Cumberland encounters more serious crime problems than any other area.

Linear Modelling

The study employed linear regression to analyse how broadband speed together with crime rates impacts residential property prices throughout the towns of Cheshire and Cumberland. The researchers used house price as their main variable while they measured broadband and crime rates to establish their findings which they discovered during the exploratory data analysis. The researchers examined all variables for consistency while they treated outliers as

a critical factor to establish dependable research outcomes. The models demonstrated that towns with faster broadband connections experience higher property values while the higher crime rates lead to decreasing property values.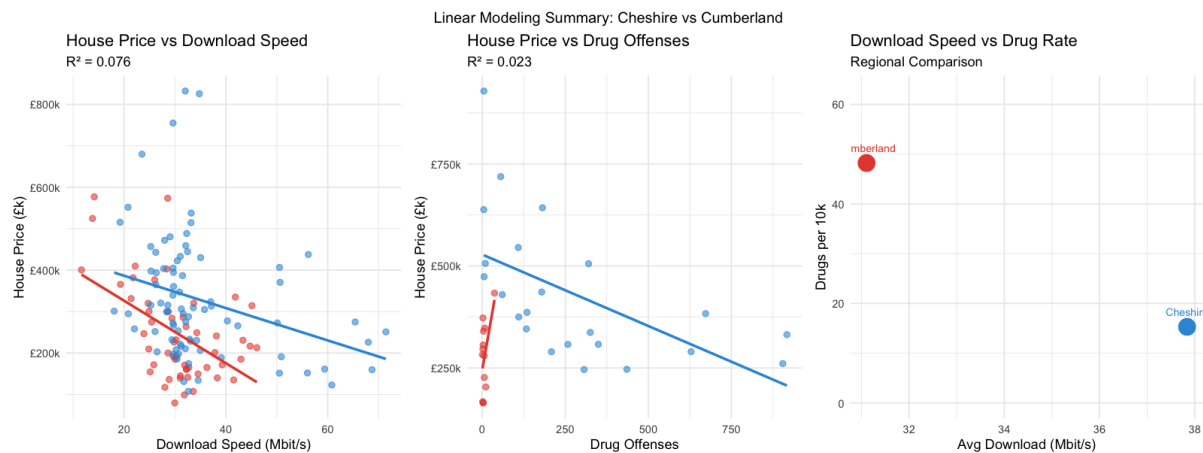 The research results established a statistical basis which served as the scoring system foundation for the recommendation framework that provides data-driven investment recommendations based on evidence.

Average Download Speed vs Drug Offense Rate (2023)
Regional Comparison

The bubble chart displays a comparison between two different measurements. Cumberland (red, top left) has ~31 Mbit/s download speed but higher drug offense rate (~50 per 10,000 people). Cheshire (blue, bottom right) has better broadband (38 Mbit/s) but significantly lower drug crime rate (15 per 10,000). The relationship between Cumberland and Cheshire shows that Cumberland has more criminal activity because its internet infrastructure is worse than Japanese connectivity.

House Price vs Download Speed
R² = 0.039

In the above Scatter plot it demonstrates the correlation is weak which has an R^2 value of 0. 039. The research found that higher broadband speeds have a minor connection to house prices which results in an insignificant relationship. The chart gets distorted by one extreme outlier which exceeds £4M from Cheshire. Most properties show a speed range between 20-50 Mbit/s and price their value under £1M. Hence, we can conclude that there is no correlation between the house price and the broadband speed as they fall under completely different categories.

Three correlation analyses are shown. The left panel shows that house prices decrease with higher broadband speeds except for the relationship which shows only a slight connection (R² = 0.076). The middle panel shows that house prices and drug offenses have almost no relationship (R² = 0.023) between them. The right panel demonstrates that Cumberland possesses slower internet speeds together with increased crime rates while Cheshire offers superior broadband services and experiences reduced drug-related incidents. The overall conclusion shows that broadband speed and crime rates do not effectively predict house prices because regional differences have a greater impact than these factors.

Recommendations Systems

Overview

The recommendation system was designed to provide a clear, data-based ranking of all towns in Cheshire and Cumberland through the three investment criteria which included house prices and broadband speed and crime rates. Each town received scores between 0 and 10 which measured its performance on each attribute with higher scores indicating better conditions through lower prices and faster broadband and lower crime rates. The scoring system established standardised measures which enabled direct comparison between different variables and the resulting total score was used to assess each town's performance. The

system displays complex data through a simple ranking system which enables investors to find locations that match their specific needs. The evaluation process included an extra confirmation factor which measured the distance to vital services because it was essential for assessing residents' standard of living.

Results

In the recommendation system the scores are generated for each town by evaluating the price of the house, speed of the internet, crime levels, and additionally the quality of life Towns with lower property prices, faster internet, and lower crime generally achieved higher individual scores, while those with higher prices or safety concerns scored lower. The research findings demonstrated distinct dissimilarities between the towns which showed persistent high performance across multiple evaluation criteria. The assessment process enabled researchers to evaluate each town's strengths and weaknesses through data analysis which helped investors assess three essential elements: property costs, internet access, and security. As per the study, all decision-making factors have the same bearing on outcomes, and towns that have multiple positive attributes end up being the best property investment options.

Reflection

Multiple factors are shown with the reccomendation system that need assessment to find best towns which will be properly invested by the investors. Towns receive their complete score through the combined impact of affordability and broadband speed and crime levels and overall quality of life. Some towns have affordable housing prices, yet the locals face two main problems which include slow internet connections and high crime rates whereas other towns provide safe environments with reliable and fast internet speed at higher costs. The

system combines these characteristics into a single score which shows towns that maintain identical performance levels across all assessment areas. The approach uses evidence-based data-driven decision-making to help investors understand trade-offs so they can make informed choices instead of using their basic assumptions. The reflection indicates that multi-dimensional evaluations offer a clearer viewpoint of the formidable investment and liveability in each one.

Overall Score Formula

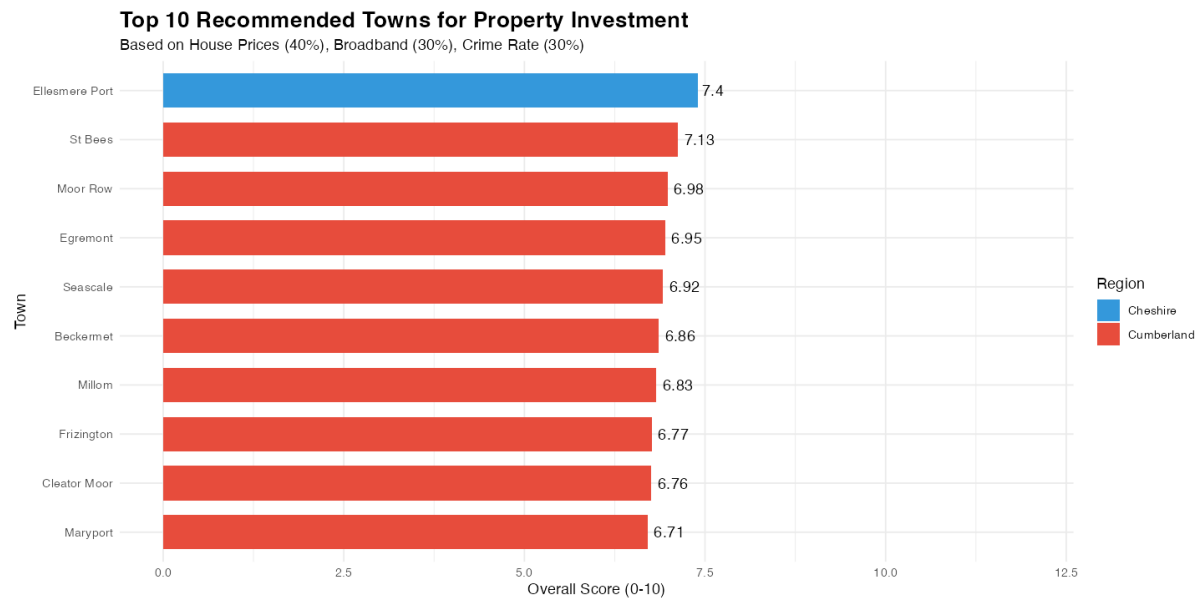Typically, the total score for each town was achieved using the weighted linear formula as shown below:

Overall Score = (House Price Score × 0.40)

    + (Broadband Score × 0.30)

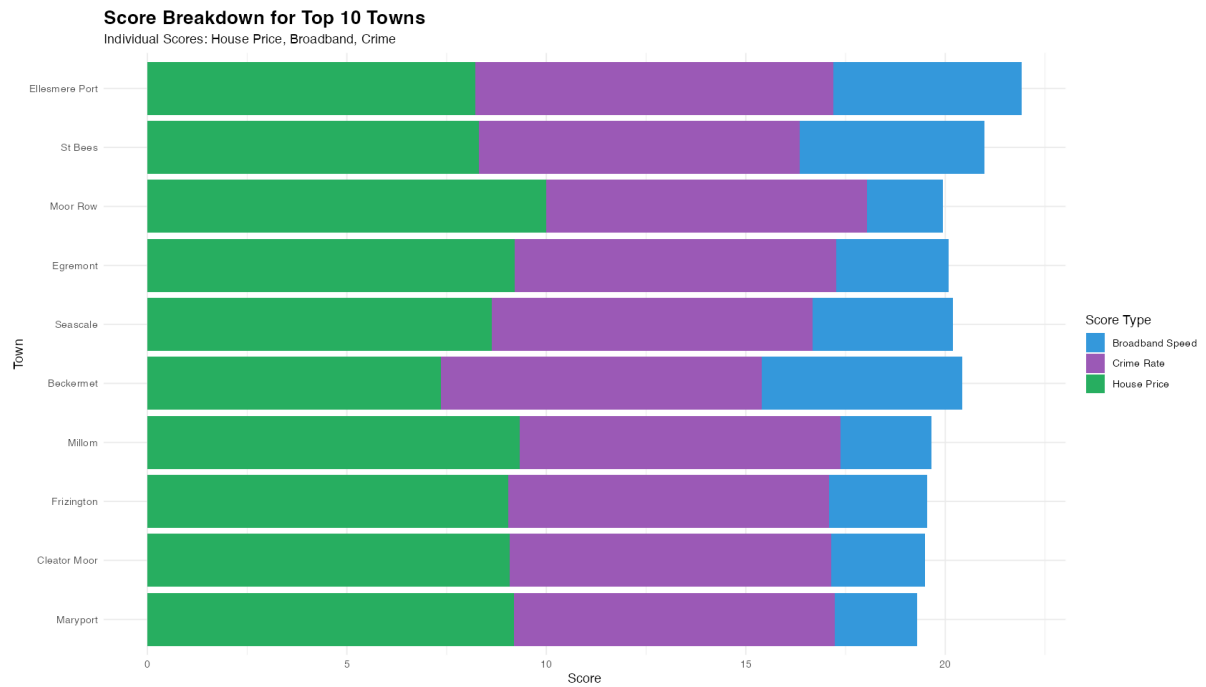   + (Crime Rate Score × 0.30)

County	House Price (40%)	Broadband (30%)	Crime (30%)	Overall Score
Cheshire	9.90	8.90	7.20	8.66
Cumberland	9.79	3.65	4.00	5.81

Top 10 Recommended Towns for Property Investment
Based on House Prices (40%), Broadband (30%), Crime Rate (30%)

In Cheshire, Ellesmere Port holds the first rank because it was able to get a good score 7.4 out of 10. Also, in Cumberland town St Bees has been able to score 7.13 out of 10. The assessments use house prices for 40% weight and broadband and crime each for 30% weight. The high affordability of Cumberland makes it the most affordable region.

Overall System Score

The total system score creates one score from its three components which include house prices and broadband speed and crime rate results. The highest weight is assigned to housing affordability by the System, with the qualities of broadband connectivity and safety considered equally important. The approach demonstrates that Cheshire achieves superior results because its residents possess better internet access while facing less criminal activity situations which help them to maintain their safety. In the other hand Cumberland performs worse because its residents experience both unreliable and slow internet access and higher rates of criminal offenses.

Score Breakdown for Top 10 Towns
Individual Scores: House Price, Broadband, Crime

Above graph represents the Ellesmere Port balanced score which represent house price for color green, purple for crime rates and broadband availability is shown with blue. In the Cumberland town the affordability with less crime and internet performance is indicated by green.

Legal and Ethical Issue

The project requires legal and ethical considerations to conduct data operations because they need to maintain responsible data handling and follow all regulatory requirements. Researchers obtained all datasets from publicly accessible UK government websites which distribute their data according to open data licenses. The data can be legally accessed and processed and shared according to legal requirements when proper attribution is provided. Official sources provide assurance that all personal data and sensitive information will either go through anonymization processes or be combined into aggregate data which decreases privacy risks.

For accurate methods, the analysis maintains ethical standards through its dedication to transparent. The documentation includes all data cleaning and processing procedures which enable others to replicate and verify the results. The researcher uses the research safety guard for protecting their work from false information because they use quantitative data which helps to rank the town based on affordable house, fast internet speed and low crime rate. The process of making interpretations and recommendations occurs with good intentions to help others make informed choices without any form of bias or discrimination. The project covers all the legal requirements while maintaining the ethical standards which guarantee the handling of the data through the entire process of data analysis.

Conclusion

The project achieved its goal through implementation of the data science lifecycle which enabled its researchers to conduct a data-based evaluation of property investment potential between towns in Cheshire and Cumberland. The study used house price data and broadband

speed data and crime rate data and multiple quality-of-life factors to create town rankings which showed which towns offered the best value through their combination of affordable housing and dependable internet service and secure locality. In this documentation the researcher documented all the steps for data cleaning also processed and analysed to make transparent research that helps recreate method and verify it.

The recommendations use quantitative evidence which eliminates any need for personal evaluation because it provides objective results that assist investors in making informed decisions. This project helps to maintain the ethical standards as well as legal because the team obtained all data sets from the government websites which is accessible for public and is managed in responsible way. The study shows how researchers can use structured data analysis to convert difficult datasets into practical results which maintain their ethical standards and legal compliance.

References

Price paid data. (2025, December 30). GOV.UK. https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads

House price statistics for small areas in England and Wales   - Office for National Statistics. (2022, September 14). https://www.ons.gov.uk/datasets/house-prices-local-authority

Price paid data - data.gov.uk. (n.d.). https://www.data.gov.uk/dataset/ec55e73e-61f4-4ac7-9702-27db85b127fc/price-paid-data

UK fixed-line broadband performance - data.gov.uk. (n.d.).
https://www.data.gov.uk/dataset/dfe843da-06ca-4680-9ba0-fbb27319e402/uk-fixed-line-broadband-performance

Find open data - data.gov.uk. (n.d.). https://www.data.gov.uk/

Statistical data sets. (2025, October 28). GOV.UK.
https://www.gov.uk/government/statistical-data-sets

UK Data Service. (n.d.). https://ukdataservice.ac.uk/find-data/browse/housing/

Appendix

Github: https://github.com/SssamiAryal/DS_CourseWork.git