Trends in Anti-Social Behaviour Crime Statistics in England and Wales

Helen Whitehead, November 2024

Executive summary

This report investigates Anti-Social Behaviour (ASB) crime statistics for England and Wales to identify trends and correlating factors, using Police crime data for October 2023 to September 2024.

It was observed that ASB levels are strongly seasonal, with more crimes reported in the summer months when days are longer.

The ASB statistics were analysed in small geographical areas, Lower Super Output Areas (LSOAs) which showed that the level of ASB varies enormously from one area to another, with LSOAs ranging from 0 to 1220 crimes in the year, with a median of 16.

A Support Vector Regression model was trained, and this demonstrated that the area statistics which most helpfully contribute towards an ASB count prediction are: the number of retail and entertainment sector workers (a proxy for town centre areas with pubs and shops); the proportion of housing which is owned (as opposed to rented or under shared ownership); and the proportion of residents who are aged 20 to 34. The model predicted ASB for unseen test data with better accuracy than a naive benchmark, with an R Squared of 0.38 compared to -0.00.

It is recommended that local and national policy makers take these factors into account when aiming to understand and reduce ASB, for example in housing planning decisions.

Introduction

Anti-Social Behaviour (ASB) refers to crimes where a person's behaviour is likely to cause "harassment, alarm or distress to persons not of the same household"[1] for example noise, verbal abuse, littering or graffiti. The impact on victims is known to include "negative mental health effects, avoidance behaviours and decreased economic productivity."[2] Given the significant implications of ASB, it is important to identify trends and contributing factors in order to inform local and national government policy to reduce ASB.

1 From the Antisocial Behaviour Act 2003 and Police Reform and Social Responsibility Act 2011 quoted by Met Police (https://www.met.police.uk/advice/advice-and-information/asb/asb/antisocial-behaviour/what-is-antisocial-behaviour/).

2 Home Office commissioned Ipsos survey (https://www.gov.uk/government/publications/impacts-of-anti-social-behaviour-on-individuals-and-communities/anti-social-behaviour-impacts-on-individuals-and-local-communities)

The police openly publish data on all crimes reported in England and Wales. The published files cover each month and each Police authority separately but by collating hundreds of files and aggregating data regarding almost a million ASB crimes, it becomes possible to identify trends in ASB which may not be apparent at a local level. While the data is heavily anonymised it none-the-less includes a location in the form of a Lower Super Output Area (LSOA), so it is possible to to analyse how ASB varies both from month to month and from one place to another. By combining the data with other datasets relating to LSOAs, such as deprivation measures and age demographics, we can identify what features of an area are associated with ASB, which may be useful for predicting ASB and intervening to reduce it.

This report set out to address:

- How did the level of ASB vary over the time period analysed?
- How much does the ASB level vary between LSOAs?
- Can we predict ASB using other factors?

Methods

Data Sources

Crime

- Used python's glob function to retrieve and then loop through 504 crime files from the UK Police[3] collating any crimes with the classification 'Anti-Social Behaviour' (one file for each police district in England and Wales for each month from October 2023 to September 2024).
- Of the 900,000 ASB crimes 1449 had no LSOA listed (0.16%). Given that this is such a small minority, and that any method of imputation would have been inaccurate (e.g. to fill using the most common LSOA for that police district), these rows were simply removed.
- Created a single DataFrame with an ASB count for each LSOA. Not every LSOA in England and Wales had an ASB crime in the year, so there were missing LSOAs, and these would later be filled with zero ASB counts on merging.

Census

- The Office of National Statistics (ONS) publish census data[4] broken down by LSOA (35,672 rows) as tidy csv files with no missing data, duplicates, or poor formatting.
- The most recent UK census took place in March 2021, so the pandemic would have affected some figures. With that in mind, the files which were identified as having the greatest potential for helping to predict ASB rates were: age, education, unemployment, central heating, and tenure.

3 https://data.police.uk/data/
4 https://www.nomisweb.co.uk/census/2021/data_finder

- In each case the file was read-in, omitting the header and footer. The LSOA name and code were split into separate features. In some cases some features were re-named, combined, or dropped.
- These area statistics for each LSOA relate to residents and households in the location of the crime, which may not be where the offender or victim reside.

Employees in Retail and Arts

- It may be that locations where people congregate, e.g. town centres with shops and pubs, have a different level of ASB compared to residential or rural areas. The proxy metric chosen for addressing this factor was the total number of employees in either of the industries 'Retail' or 'Arts, entertainment, recreation & other services' which is available via the Business Register and Employment Survey 2022 from the ONS[5].
- Data processing proceeded as with the ONS census data above.

Other supporting Data Sources

- Used a csv file from the ONS as the authoritative source of LSOA names / codes and their boundary definitions[6], plus a related shapefile for geocoding[7].
- Retrieved figures for UK daylight hours per month from World Data[8].

Merging and Scope

- Merged the data sources into a single DataFrame based on LSOA code. Filled LSOAs missing from the ASB crime count source with zero ASB as planned.
- One of the data sources (retail and entertainment workers) used older LSOA boundaries than the other sources (2011 vs 2021). The supporting data source above showed that 94% of the LSOAs had unchanged boundaries, so the analysis focussed on these unchanged LSOAs, and the LSOAs whose boundaries had changed were discarded.
- One of the 43 police forces didn't publish their statistics, so any LSOAs within the Greater Manchester Constabulary were also excluded.
- With these exclusions 32,011 of the 35,672 LSOAs in England and Wales were represented, which equates to 90%. Figure 2.1 shows these excluded LSOAs as the non-coloured polygons.

Exploratory statistics

- Used matplotlib and seaborn libraries for exploratory statistics. In particular used seaborn's regplot to loop through the features visualising relationships, first as linear, then when one of the features (education) showed no linear relationship adding the argument "order = 2" to investigate whether it benefited from adding a second-degree polynomial. Scatter plots used a random sample of the data, since it can be hard to observe trends with 32,000 data points.

---

5 https://www.nomisweb.co.uk/datasets/newbres6pub

6 https://www.data.gov.uk/dataset/b82d9b7d-7626-4aa1-9973-c6b19836c26b/lsoa-2011-to-lsoa-2021-to-local-authority-district-2022-best-fit-lookup-for-ew

7 https://www.data.gov.uk/dataset/c481f2d3-91fc-4767-ae10-2efdf6d58996/lower-layer-super-output-areas-lsoas

8 https://www.worlddata.info/europe/united-kingdom/sunset.php

- Used the co-efficient and intercept from a trained Linear Regression model from sklearn to define a linear equation to express the relationship between daylight hours and ASB levels.
- Exported the DataFrame as a csv to use as a data source in Tableau, alongside the supporting shapefile, to create a map of the ASB crimes by LSOA at Figure 2.1.

Analytical Methods

- Removed features with collinearity (i.e. % population old / % population young, and % homes rented / % homes owned) because there is no benefit from representing the same information in two ways, and the model may have behaved erratically if these had been included.
- Outliers were not removed. These were determined to be a real part of the data, not just noise, so it was more responsible to leave them in, as they will assist with understanding real extremes in ASB.
- Split off 25% of the data to be excluded from training and hyper-tuning, and to remain unseen for testing purposes.
- Created a benchmark for comparison purposes, using a Dummy Regressor from sklearn which made predictions based purely on mean ASB, without using any other features.
- Scaled features using StandardScaler from sklearn, to ensure that features with bigger values didn’t dominate over features with smaller values.
- Used Support Vector Regression (SVR) from sklearn as the chosen model. Having looked at the distribution of the data and found it to be leptokurtic (more outliers) and skewed, and having found some non-linearity in one of the features, the model needed to be able to handle all these properties, which SVR can. Unfortunately due to having a large dataset this took a long time to tune, train and explore. Other methods were investigated, but due to the data being skewed and having a lot of outliers these performed poorly.
- For hyperparameter tuning, looped through:
  - values of C to determine how many misclassifications to allow. The optimal value was 100 which is a relatively soft margin decision boundary, allowing for outliers.
  - types of kernels due to potential non-linearity in the data, especially the education feature. Performance was best with the Radial Basis Function (rbf), which allows as many auxiliary input variables as required.
  - values of gamma to determine how sensitive the model should be to individual data points. The optimal value was 0.01.

  On each loop, five fold cross-validation was used in order to minimise over-fitting and give the best chance of strong performance on the unseen test dataset.
- Trained the SVR model with the hyper-parameters identified, and then compared the R squared score for the training set and the test set (to check for over-fitting), and compared to the benchmark (to evaluate performance).
- Looked at feature importance using sklearn’s permutation importance function, which randomly shuffles the values of the features, and uses the resulting degradation of the model’s performance to indicate how important each feature is to the success of the model.
- Plotted the predictions for the test dataset against actuals and against residuals to visualise how the model performed.

Results

This section will lay out results in terms of the key questions: how did ASB vary throughout the year; how much does the ASB level vary between LSOAs; and can we predict ASB using other factors?

How did ASB vary throughout the year?

The total number of reported ASB crimes across England and Wales varies considerably from month to month, with the highest month's figure (July 2024 at 90,237) being 60% higher than the lowest month's figure (December 2023 at 56,266).

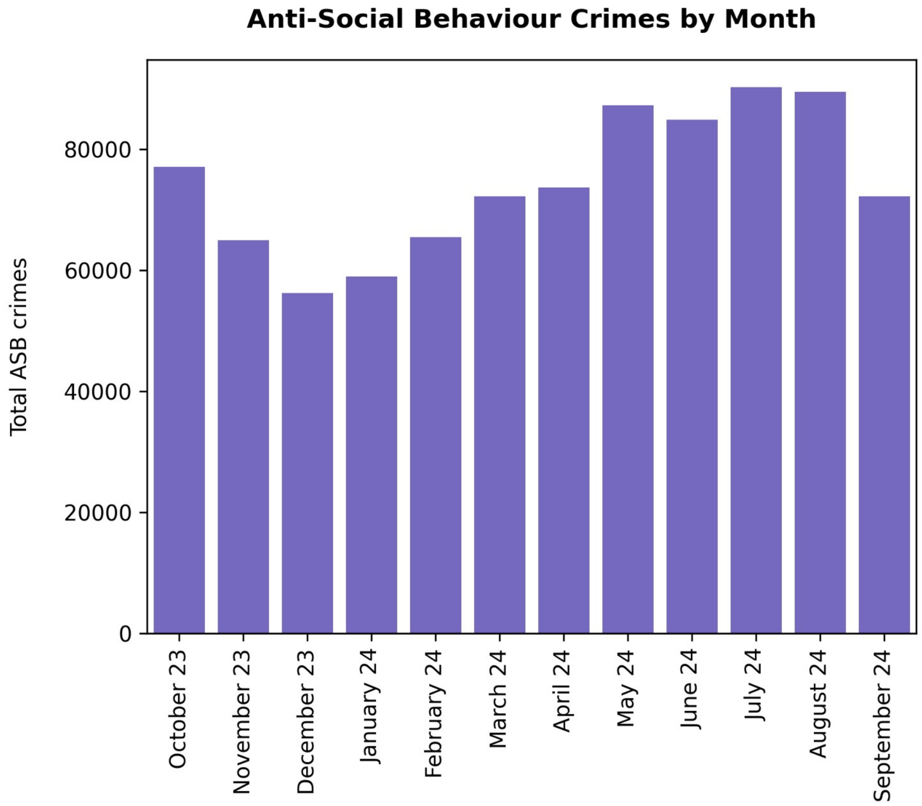

Fig 1.1: Anti Social Behaviour crimes are total reported for England and Wales, excluding Greater Manchester Police district. The level varies considerably through the year, peaking in Summer and dipping in Winter.

Whilst the monthly ASB figures vary considerably, this does not reflect that ASB is erratic or unpredictable. Figure 1.1 suggests a seasonal element, where perhaps longer daylight hours or warmer temperatures keep people outside for longer, and result in more problematic behaviour.

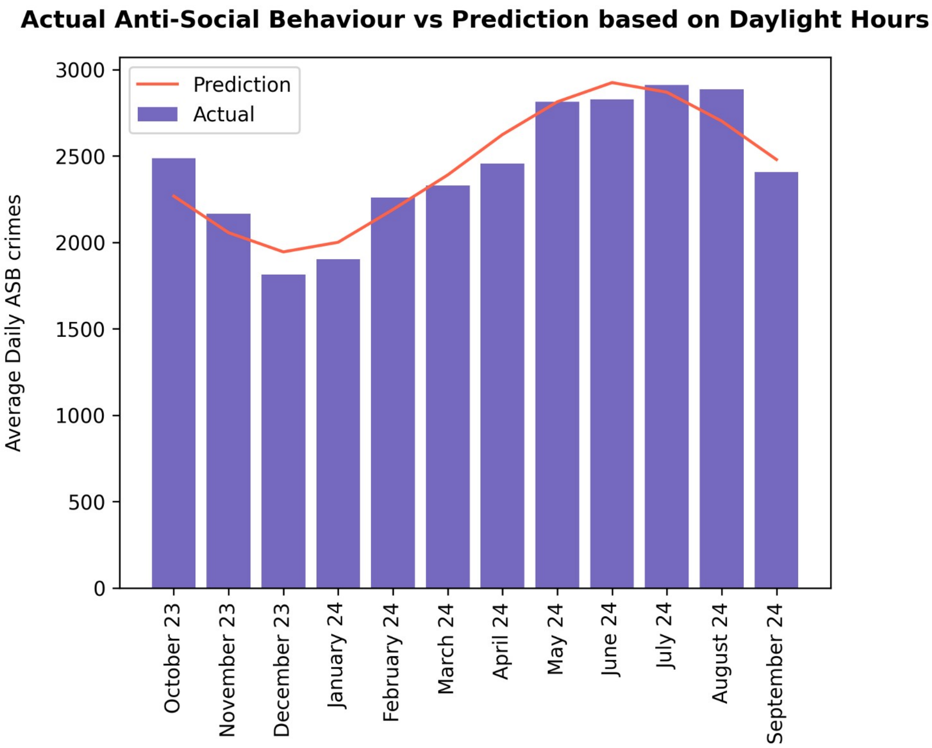

Fig 1.2: The actual ASB figures are closely predicted to 88.5% success by the linear equation:
Average Daily ASB = (Average Daily Light Hours in Month * 111.283) + 1065.835

In Figure 1.2 the effect of the number of days in the month is removed by looking at a daily average for each month rather than a monthly total. There is also a simple prediction for ASB shown using a linear equation:

Average Daily Anti Social Behaviour in any given month =
(Average Daily Light Hours in Month * 111.283) + 1065.835

This linear regression model is very over-fitted and would not have much predictive power outside of this dataset, but none-the-less it predicts the ASB crimes in 23/24 with 88.5% accuracy, which demonstrates that most of the monthly variation in ASB is accounted for by the number of daylight hours. Perhaps this is because people are more likely to be outdoors and engaging in problematic behaviours in the daylight, or to have windows open leading to more noise complaints, or because the odour of litter can be more offensive in warmer temperatures.

How much does the ASB level vary from LSOA to LSOA?

There are 32,011 LSOAs in scope for this analysis, and their ASB counts for October 2023 to September 2024 range from 0 to 1220. These LSOAs are plotted on a map below, Figure 2.1, where the 10% of LSOAs which were omitted from the analysis are shown as the non-coloured polygons. Note that LSOAs are designed to have an average population of 1,500 residents, so they vary considerably in physical size, with densely populated

LSOAs being small, and sparsely populated LSOAs being large. It appears that the rural parts of England and Wales may be more likely to be shaded in blue, that is to have lower levels of ASB.

Lower Super Output Areas in England and Wales by Anti Social Behaviour crime count

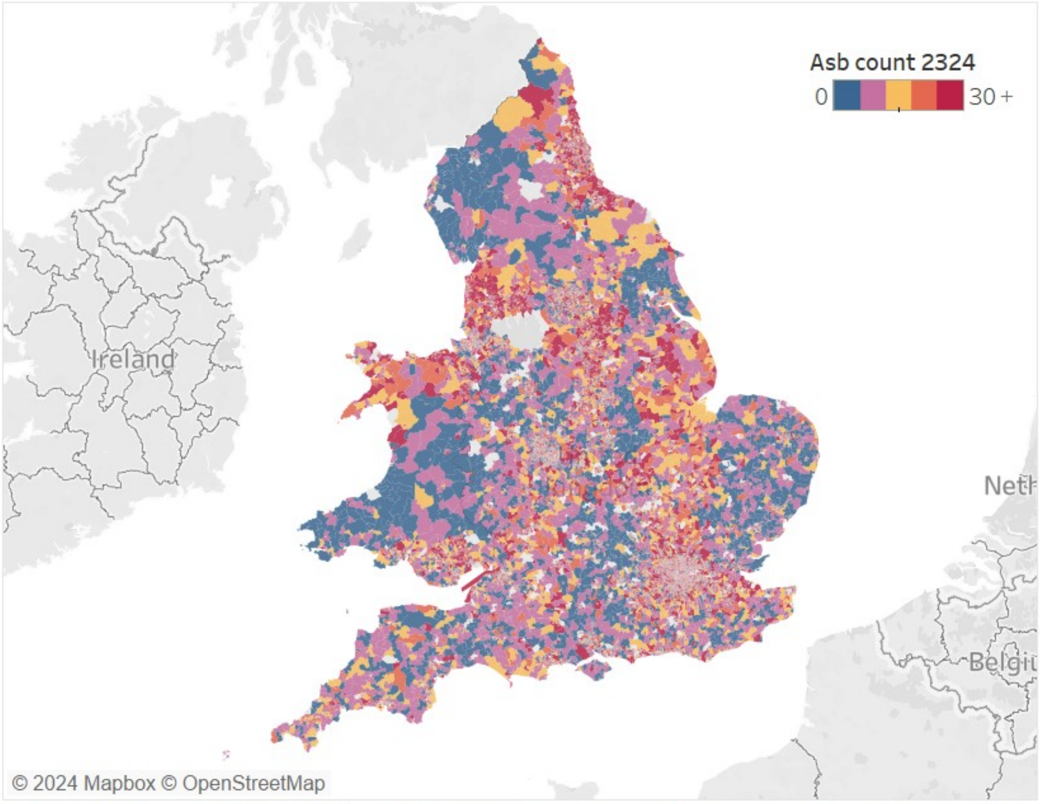

Fig 2.1: Map shows LSOAs shaded according to the number of Anti-Social Behaviour crimes reported in the period October 2023 to September 2024, from 0 to 30+. Areas with missing data, particularly Greater Manchester, are ommited (grey). The blue shades tend to indicate that rural areas such as Wales and East Anglia may belower in ASB.

The histogram below, Figure 2.2, shows that the mean number of ASB crimes 23/24 per LSOA is 25.5. While an LSOA can only have up to 25.5 crimes under the mean (no negative values possible) they can have 1000 or more crimes over the mean, so the distribution is logically right-skewed rather than normal.

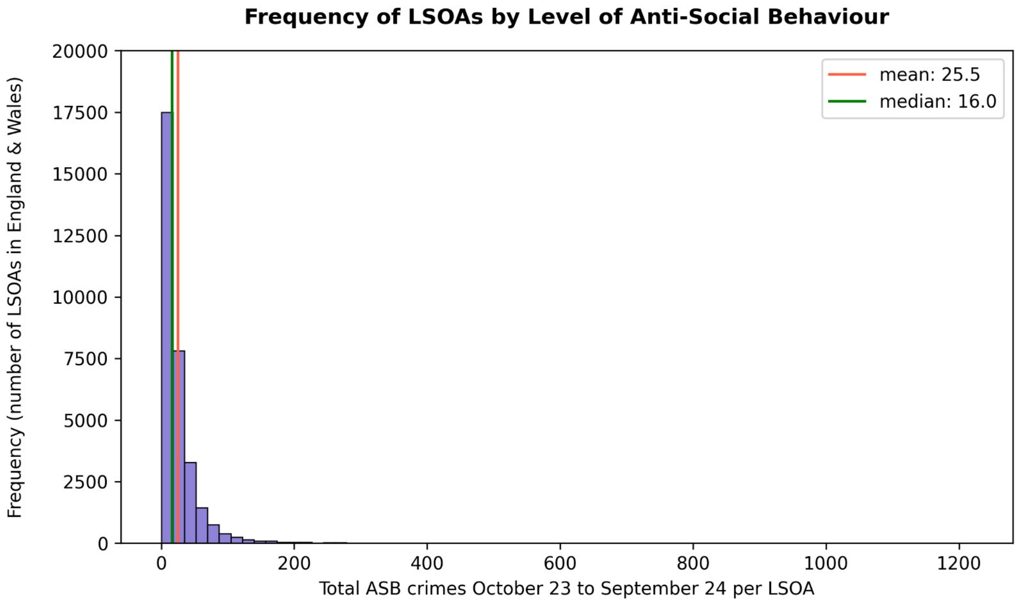

Fig 2.2: The number of ASB crimes varies a great deal. The distribution is heavily skewed to the right, with the maximum rate being 1220 compared to an average of 25.5

In Figure 2.2 we can see that the mean is being pulled upwards due to the distribution being skewed to the right. In statistical terms there is a kurtosis of 122.6 (where 3.0 would be a normal distribution and >3 is leptokurtic: more outliers) and a skewness of 7.7 (where 0.0 would be normal and >0 is right-tailed). With so many outliers skewing the data it's more useful to look at the median and the spread rather than rely too heavily on the mean.

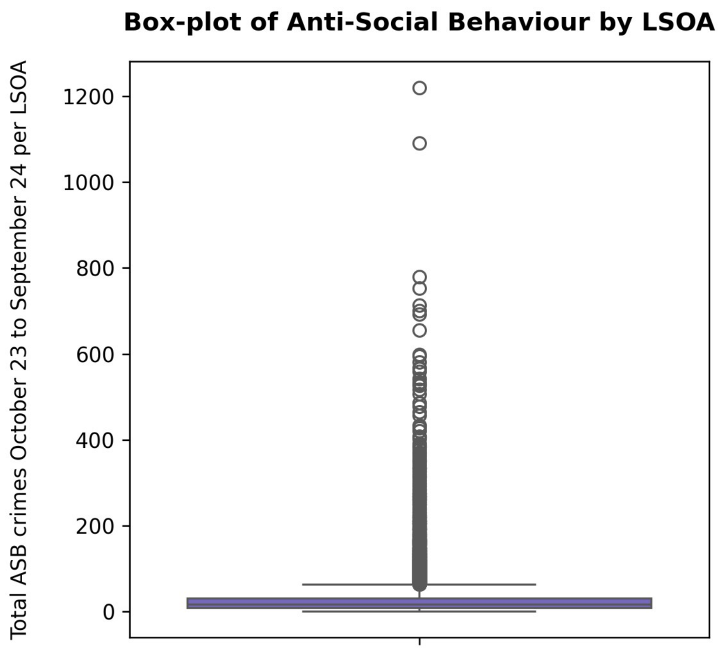

Fig 2.3: Whilst any LSOAs with over 63 ASB crimes could be termed outliers, there are a great many of these which seem to be a natural part of the distribution. There are no corresponding negative outliers due to the logic of the measure.

As the box-plot at Figure 2.3 shows, there are a high number of outliers in the data. As a rule of thumb any value above 63 could be seen as an outlier due to being above the third quartile by more than 1.5 times the Inter Quartile Range.

By looking in detail at some of the LSOAs with highest ASB counts, it appears that the high ASB is a natural part of how the data is distributed, rather than being incorrect or anomalous. Whilst the top LSOAs have have exceptionally high ASB of 40 to 50 times the average, their other features are within expected ranges, and the ASB crimes are distributed across the year (which implies that the crimes are unrelated rather than being clustered into a single month as they would be if they were part of a single incident, or a data error in one of the monthly files). Since these high figures are a real part of the data, they were not removed as outliers, as they can tell us about areas with exceptionally high ASB. The LSOAs with the highest ASB tend to be areas with a lot of retail and entertainment workers, i.e. town centres, and there is some repetition in the cities with three of the top ten being in Newcastle Upon Tyne and three in London. It looks like these areas are genuine hot-spots for ASB.

Can we predict ASB using other factors?

The ASB count for each LSOA was compared to eight other area statistics to investigate any factors which may be associated with ASB:

- Proportion of homes where the highest qualified person is qualified to level 2 or below (i.e. relatively low education)
- Proportion of residents unemployed
- Proportion of homes without central heating (as an indicator of deprivation)
- Proportion of homes owned
- Proportion of homes rented (either socially or privately)
- Number of retail or entertainment sector workers (used as a proxy for town centres with pubs and shops)
- Proportion of residents aged 20 to 34
- Proportion of residents aged 50 plus

On plotting each of the eight selected features against ASB counts in Figure 3.1, the features all appear to have potential relationships with ASB, although these are fairly weak in some cases. In most instances these relationships are linear, but in the case of low education, this feature benefited from adding a second-degree polynomial (i.e. it has a curved trend).

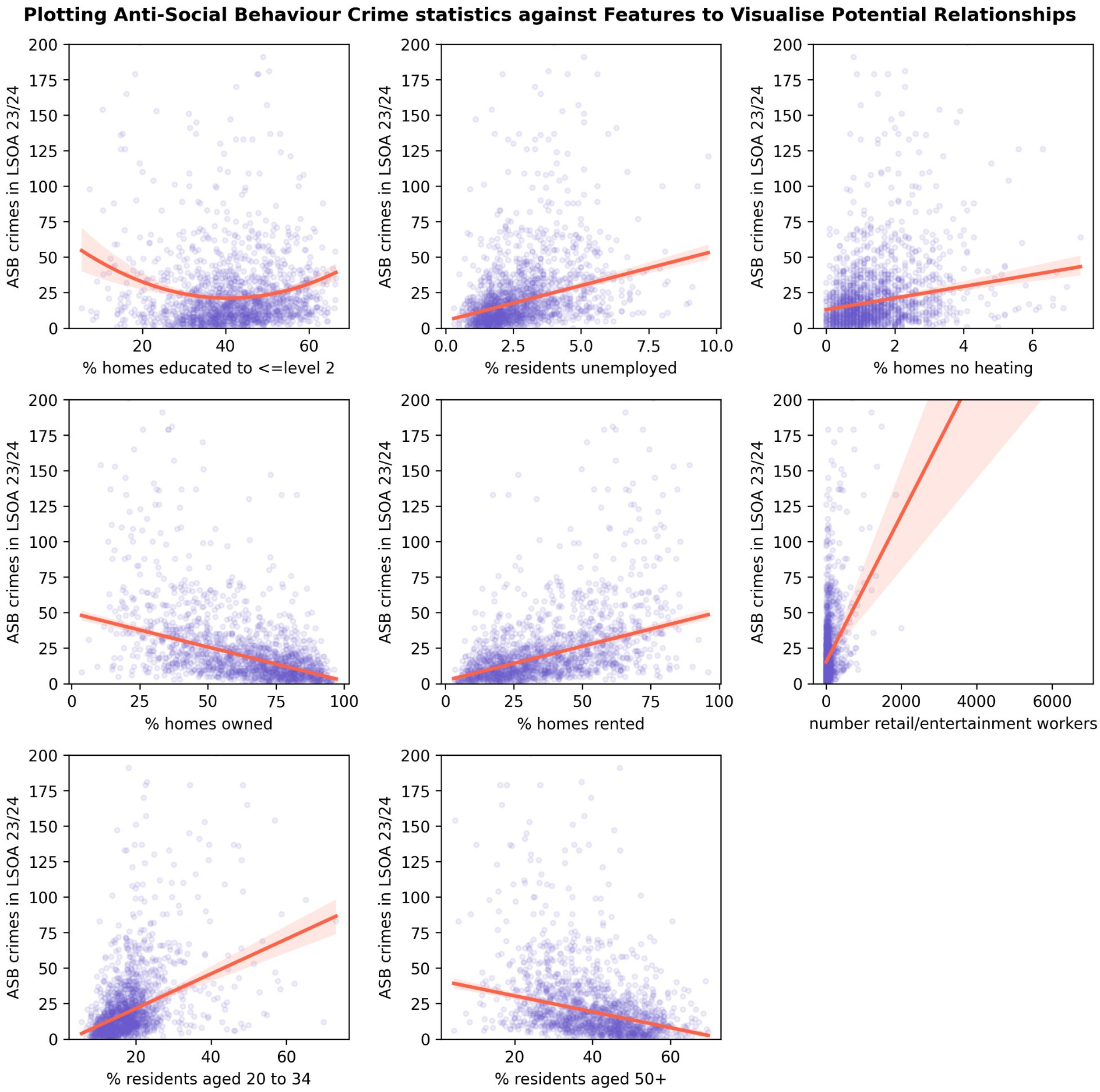

Fig 3.1: Anti-Social Behaviour Statistics are plotted against all the features. Positive linear relationships are shown for retail/entertainment workers, for deprivation indicators (unemployment, homes with no heating, and rented homes), and for young populations. Negative linear relationships are shown for home ownership and older populations. Education may have a quadratic relationship.

As Figure 3.1 shows, features which are positively correlated with ASB are number of retail and entertainment workers; proportion of residents aged 20 to 34; and deprivation measures (proportion of rented housing, proportion of homes without heating, and proportion of residents unemployed).

Features which are negatively correlated with ASB are home ownership and proportion of residents over 50.

The proportion of households where the highest qualified person is only educated to Level 2 or below may have a quadratic relationship with ASB.

A Support Vector Regression (SVR) model was trained in order to predict ASB levels using six of the eight features (omitting ‘% residents aged 50+’ and ‘% homes rented’, due to collinearity with other features).

Using R squared as an accuracy measure, the trained model scored 0.38 for the unseen test data, which indicates that the features selected are clearly related to ASB, but there is variance that we’re not accounting for in the model. None-the-less, the performance far exceeded the naive benchmark of -0.00 which is the baseline accuracy you’d expect by simply predicting the mean, so the model has some value in explaining and predicting ASB.

The R squared score for the training set was 0.42, so the fact that the score on the unseen test data was only slightly lower suggests that the model was not excessively over-fitted and could be expected to perform at a similar level of accuracy for any other unseen data. This model could therefore be used to estimate ASB in areas where crime statistics are not known, but other population figures are known – for example the Greater Manchester LSOAs for which police data was missing. As well as having predictive power, the model also informs us about how the input features are related to ASB.

Figure 3.2 demonstrates which input variables were most critical to the model using permutation importance scores. The feature which scores the highest is the number of retail or entertainment sector workers, which is used as a proxy for town centres, and indeed it is intuitive that ASB is likely to be higher in areas where people congregate. The next highest ranked feature is home ownership, where high home ownership is associated with low ASB. Perhaps home ownership brings a level of accountability for one’s behaviour, a sense of pride in one’s neighbourhood, or a willingness to foster good relationships which may not be present in areas with more rented accommodation. The next most important feature is the proportion of young adults, which may be related to noise such as parties. The remaining measures are typical measures of social deprivation, which are less important to the model.

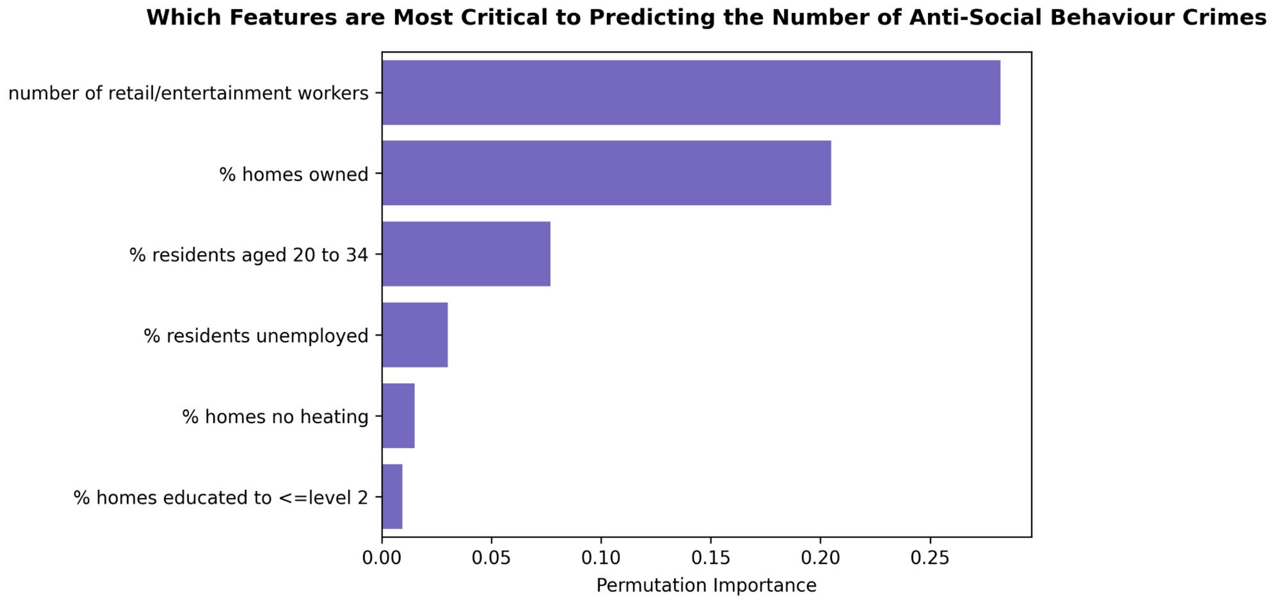

Fig 3.2: Permutation importances for each feature indicate that the features which are most critical for model performance are number of retail/entertainment workers, % of homes owned, and % of residents aged 20 to 34

Figure 3.3 plots the ASB predictions for the LSOAs in the unseen test data against the residuals (how far each prediction differs from the actual ASB figure). There are points above and below the line, demonstrating that the predictions of the model are sometimes under-estimates and sometimes over-estimates. The predictions are more accurate where ASB is lower, so the model is heteroscedastic, because the residuals are not consistent across the whole range.

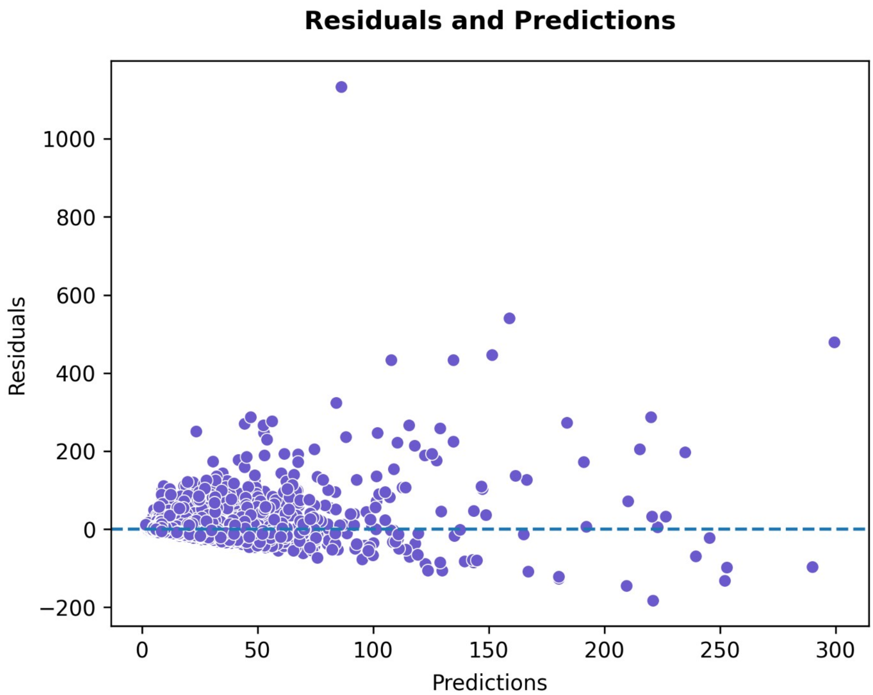

Fig 3.3: Plotting residuals against predictions demonstrates that whilst the model is not consistantly under or over-estimating (points are distributed boths sides of the horizontal line), it does experience some heteroscedasticity (the residuals are more spread at higher predictions) and a big outlying data point.

The SVR model could perform better, and perhaps there are some relevant area statistics which ought to have been in scope. In spite of the performance not being perfect, it is well above the benchmark, so the following features are clearly useful in predicting ASB: places where people congregate, with low home ownership and high proportion of young adults.

Conclusion

In this study, ASB crime statistics were analysed and a Support Vector Regression model was fitted. It was determined that:

- ASB across England and Wales in October 2023 to September 2024 had a seasonal pattern, with levels being higher in the summer months when days were longer.
- ASB varied a great deal between different LSOAs, with a median of 16 crimes per LSOA in the year, but a range of 0 to 1220. Geographically, the higher and lower areas were spread across the country, with some rural areas appearing to have lower ASB. The very highest counts tended to be in town centres, especially in Newcastle and London.
- The characteristics of an area which are most helpful in predicting ASB are number of retail and entertainment workers (used as a proxy for town centres areas with pubs and shops) which is a positive correlation; proportion of homes owned (rather than rented) which is a negative correlation; and proportion of residents aged 20 to 34 which is a positive correlation.

In understanding that these factors are related to ASB, this could be useful for setting policy. Policing could be increased in longer days and in town centre locations. Housing policy could aim to distribute young people’s housing (e.g. student accommodation and houses of multiple occupancy) across wider areas.

Future study could further analyse the relationship between daylight hours and ASB. If daylight hours are strongly correlated with ASB as was found in this study, then it follows that areas in the north of England would have more seasonality in their ASB levels, and southern areas would have less seasonality, because northern areas have more variation in the number sunlight hours across the year.