Airbnb – London Rental Predictors

Capstone Project

Benedict Medlam

Project Overview

- ## How to invest?

  Potential Airbnb Rental Property Investors.

- ## How to Improve?

  Current owners identifying areas for home improvements, to increase rental earning potential.

Current rental estimator is very basic, only considers:

- *Location*

- *Number of Bedrooms*

- *Entire place/private room*

No current tool to aid existing owners in determining how they can improve their properties market attractiveness.

Problem Statement

Using machine learning, how might we predict what property characteristics make for an attractive investment and beneficial home improvement, such that rental earning potential can be maximised?

Potential Impact:

- *Enable property investors to target their purchases based on the consumer demands.*

- *Allow current Airbnb landlords to gain most positive impact from home improvements.*

- *Improve the general Airbnb rental market environment, ultimately benefitting the renters through targeted property improvements and purchases.*

Dataset

Dataset found on Kaggle: *Airbnb Data: Listings Scraped Global top 10 cities.*

Currently focusing on the London Dataset.

Raw Data:

- *Raw data consists of 54 columns, with several columns redundant for the purpose of this project.*

    - Dropped unnecessary columns, left with 38 for the analysis.

- *Null values across entire dataset. Rows will null values in the 'Overall Rating' column have been dropped.*

- *Dropped all properties that aren't within a London postal district.*
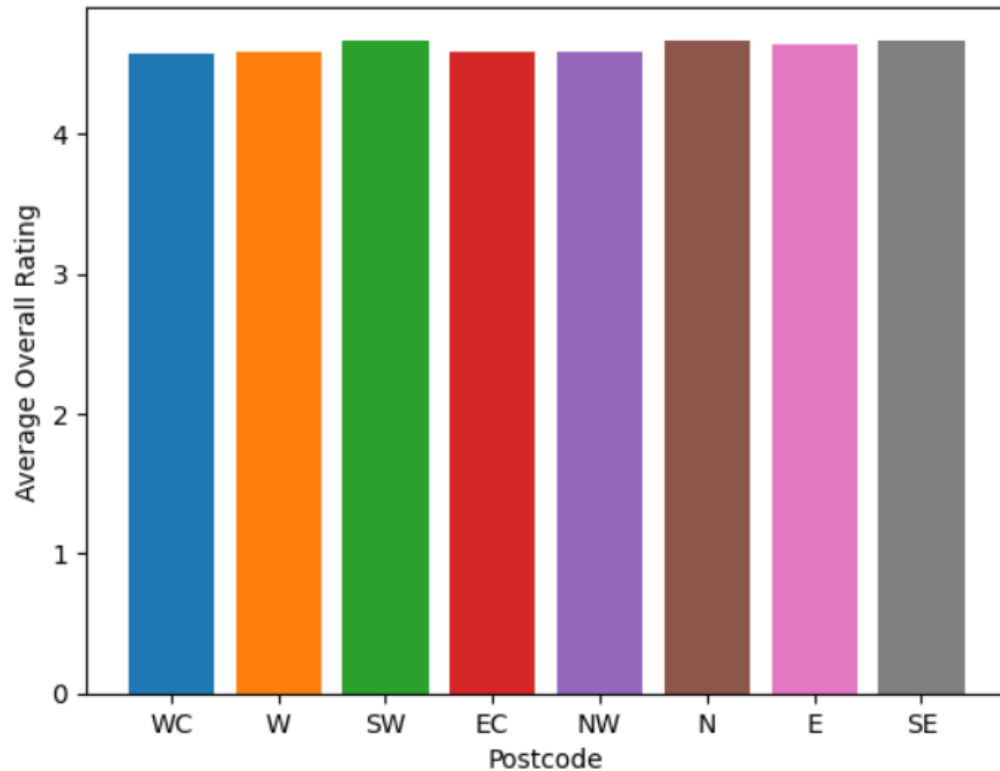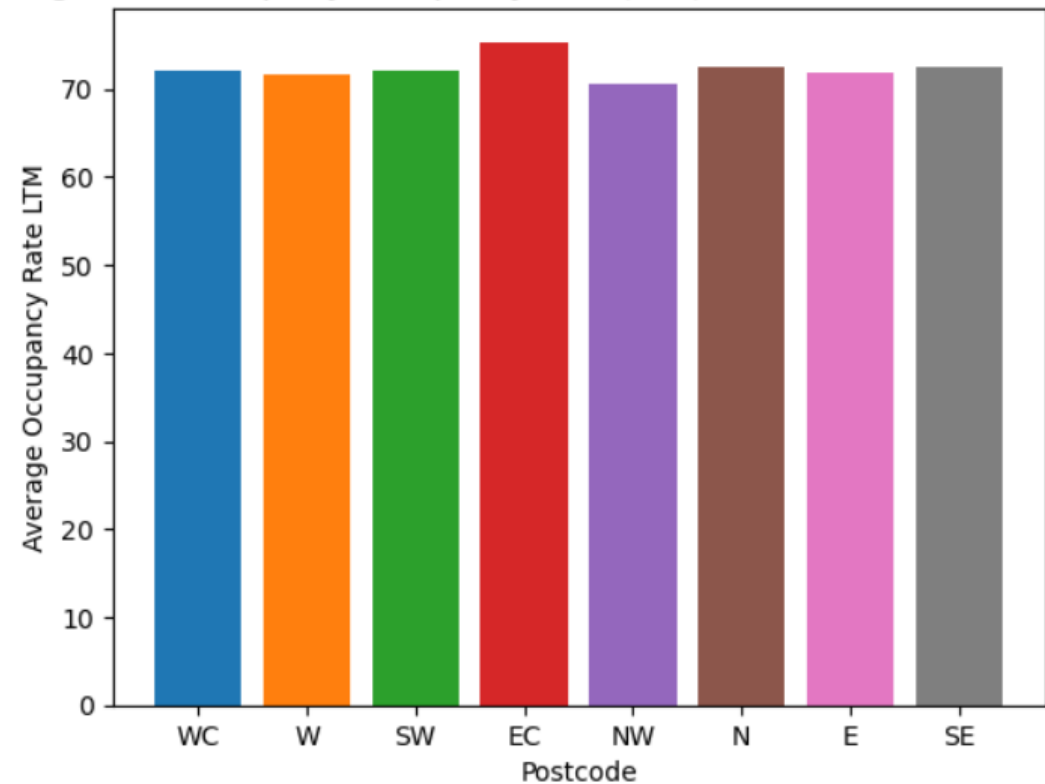
Initial EDA

Postcode Influence:
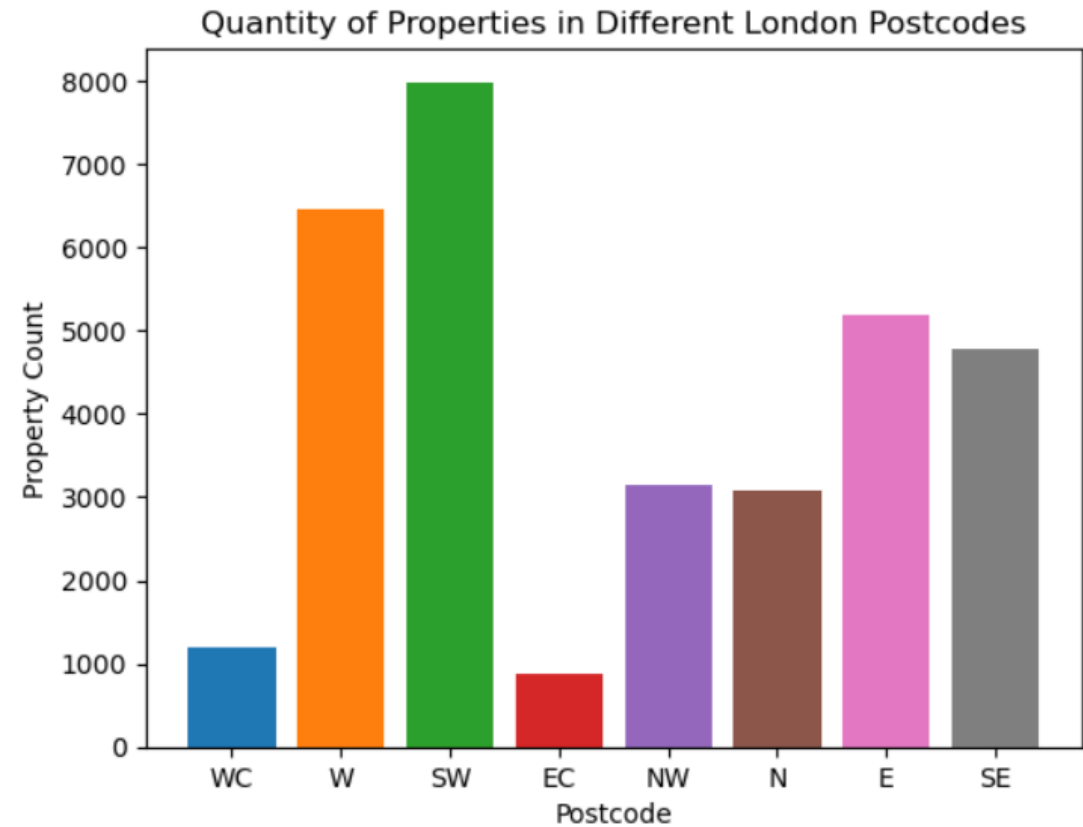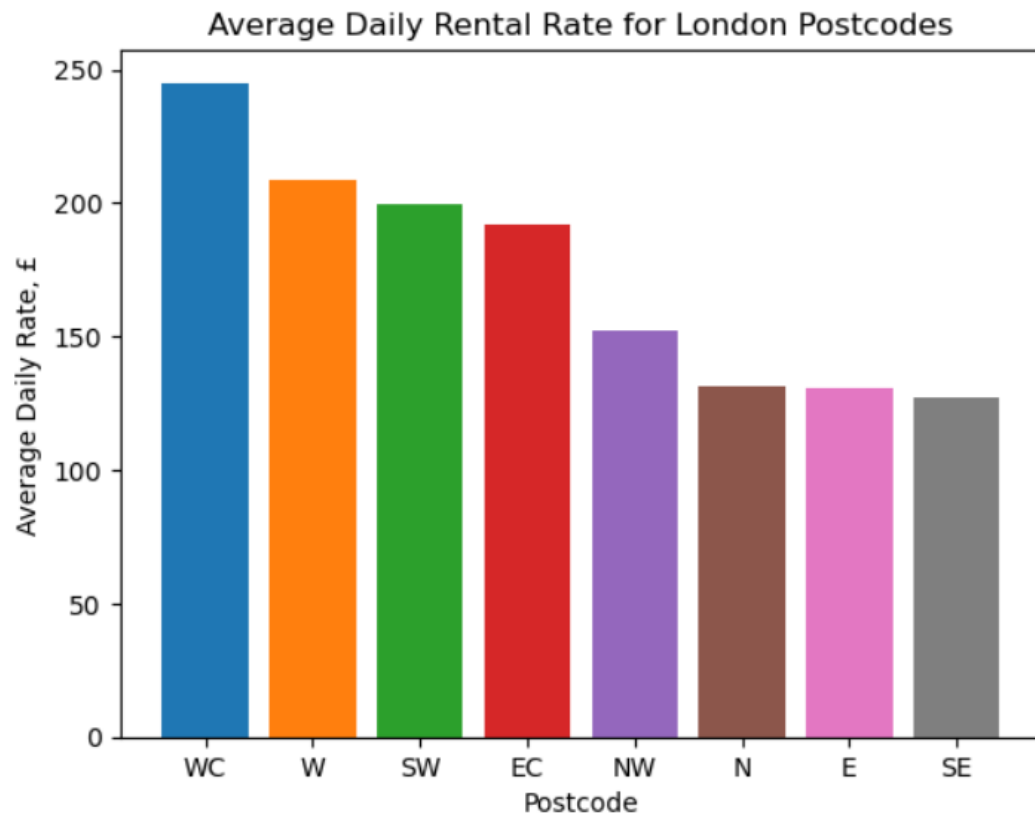
EDA Continued

Average Airbnb Property Rating for Different London Postcodes

Average Airbnb Property Occupancy Rate (LTM) for Different London Postcodes

- *SW, SE and N all tie at the top with an approximate average rating of 4.67.*

- *High average occupancy rates across all postcodes. W and SW are both above 70%.*

Next Steps:

- Additional EDA and Data Pre-Processing

More EDA, in-depth bivariate analysis and pre-processing: e.g. creating dummy variables for the different 'Property Types'.

- Baseline Modelling

Complete linear and logistic regression models to further analyse the relationship between variables.

- Machine Learning Models

Start training targeted machine learning models, using the refined and optimised data: e.g. potential use for NLP to determine which words within a property's description influence its success.