

UK Population Forecast (RNN)

Overview

This project's goal is to analyse and forecast the UK's population within the following: total population, urban population, rural population, male, and female population. The data has been extracted from the latest (retrieved 28/05/2020) World Bank statistics .csv for the UK.

This document provides a summary of the findings shown in the project's notebook, please see the notebook for all the graphs and methodologies.

The analysis and forecast have been successful as trends have been shown and interpreted, whilst the RNN forecast estimates a realistic simulation for the populations' futures.

This project lays a basic foundation for more features to be added and analysed, and could benefit from a deeper discussion with a domain expert with regards to adding to and refining the model. Note that the original .csv and the cleaned dataset includes much more data which could be used to improve the model.

Data source: <https://data.worldbank.org/country/united-kingdom?view=chart>

Data Analysis Changes

- Used Excel to remove many unnecessary rows from the original dataset, especially those with many nulls, this left a cleaned .csv with more pertinent information with manageable null amounts.

Population

Dissected the statistics dataframe to create a population dataframe with the following figures: Total population, Male population, Female population, Urban population, and Rural population.

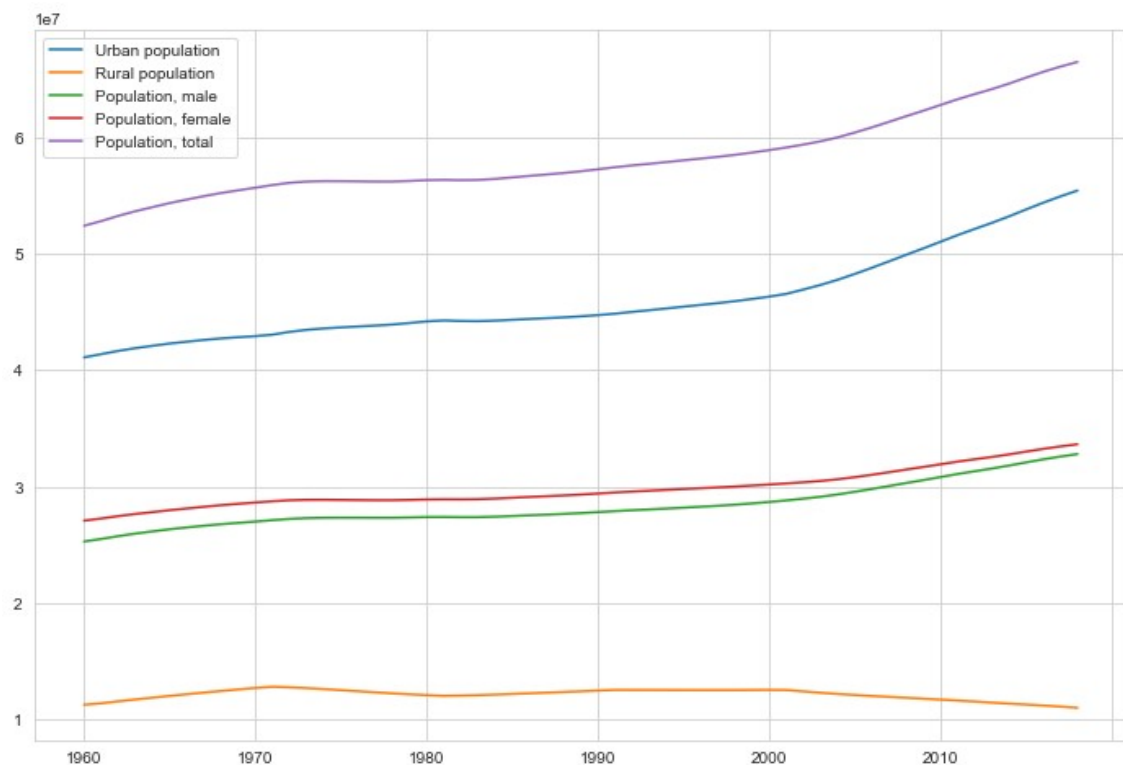
Populations Overview

The overall population is steadily increasing and the overall population increase trend seems to be following the population increase in urban areas specifically. Rural population is seeing a significant trend in declining population.

- This suggests that pressures on urban lifestyle will increase such as: housing prices, competition for jobs, amount of schools, strain on urban healthcare, etc.

- From 2000 to 2018, the urban population has increased by a large 19.7% and the steep trendline does not seem to be slowing any time soon.
- The opposite can be said for rural areas, where resources (educational, hospital, other government spend) will go to waste if they are not reallocated to urban areas accordingly.
- We can see that, particularly from 2000, rural population is sharply declining – a 12.2% drop in rural population from 2000 to 2018.

The male and female populations are increasing at around the same rate, and have been since the 1960.



RNN Model

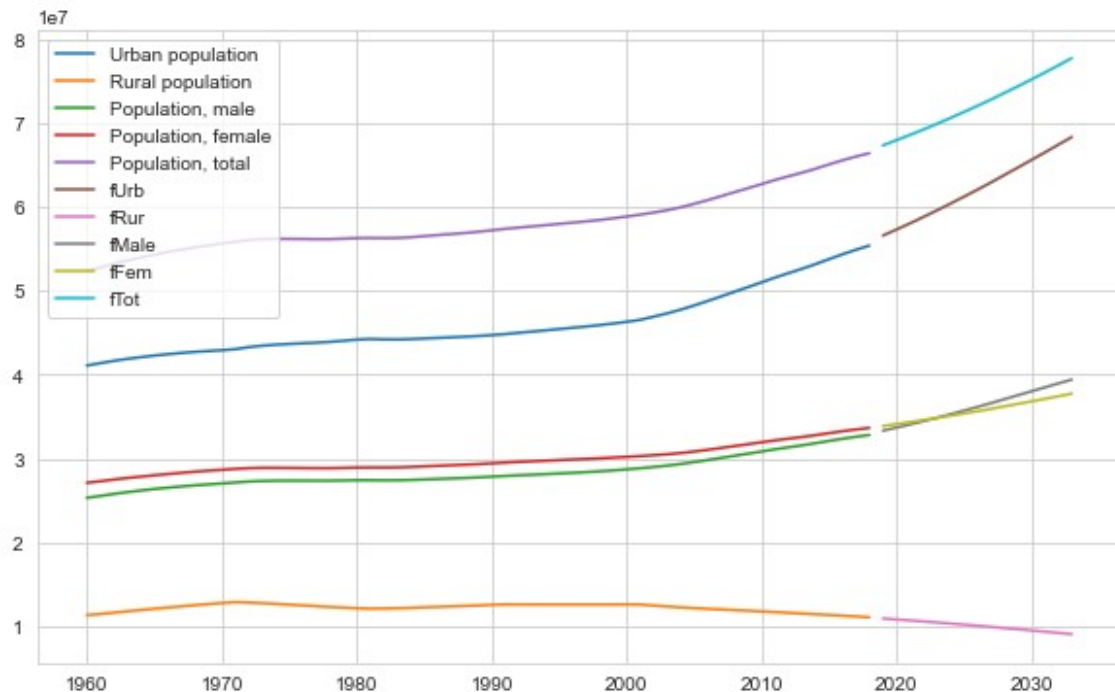
Target – Rural Population and Urban population

Using a Timeseries generator in conjunction with an LSTM layer and EarlyStopping to firstly predict and compare against a test set.

Initially tried using only the single Rural feature to predict the ‘future’ Rural population, but this failed miserably.

The fix was to use all 5 of the features and feed it into the model instead. This produced significantly better results, tweaking with the amount of neurons 50/100/20/12, etc., as well as the activation relu/tanh helped to get a better model before moving on to real forecasting (predicting the unknown future). Dropout layers were also tried.

Forecast



(slightly different from the notebook image)

The graph shows the potential forecast of population for the next 15 years. In particular, the forecasted rates of Rural, Urban and Total population seem to follow the natural trend well, whilst the Male and Female populations look less believable – most likely, like the past, Male and Female populations will move parallel to each other and not cross.

At these rates we'll see the Rural population fall to below 10 million by 2028, and the urban population reach almost reach a staggering 70 million people by 2033.

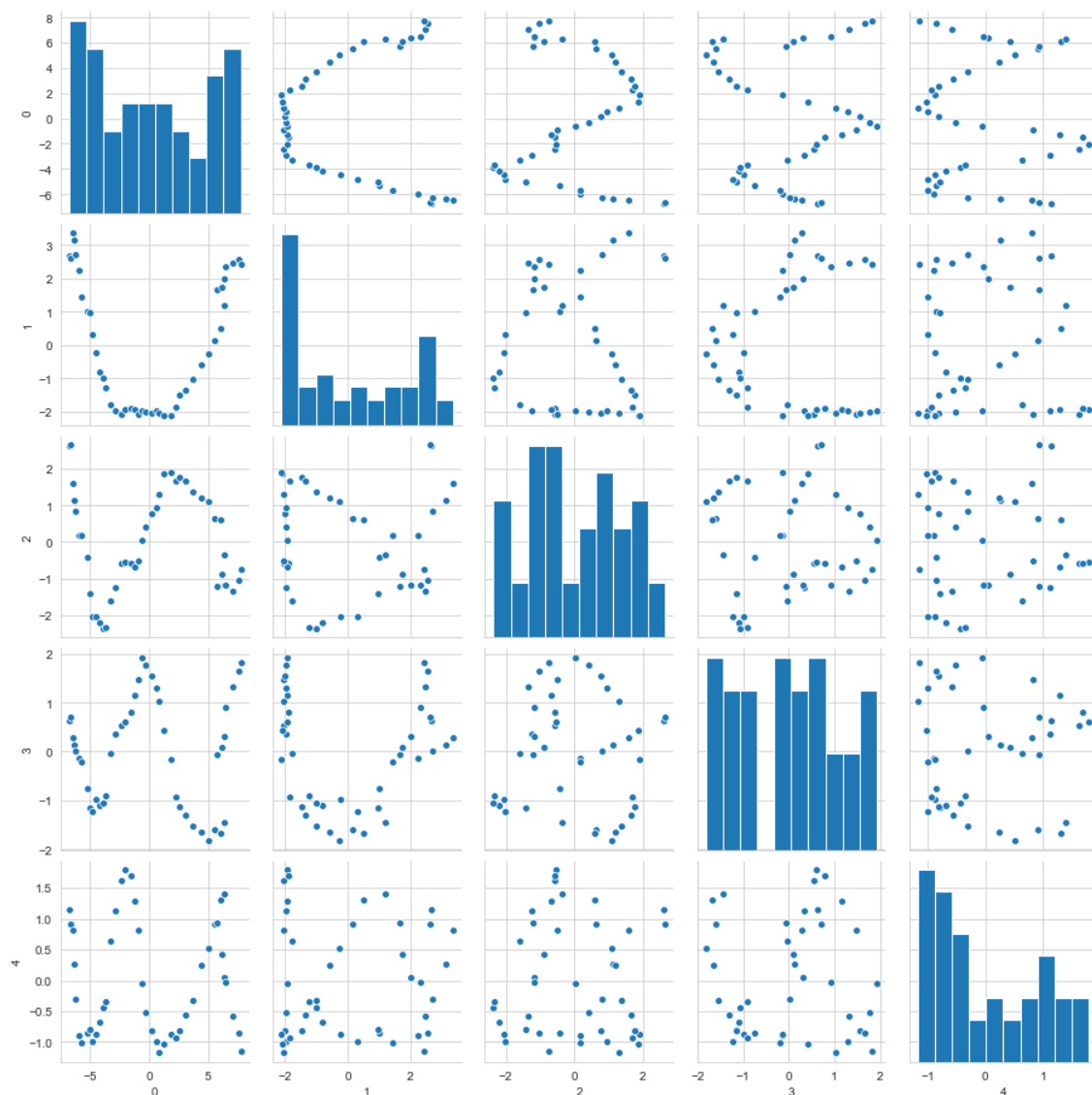
Insights

As in all forecasts these numbers are only a best guess and aren't completely reliable, although they are useful to plot potential trends to so that we are at least *prepared* for what the forecasts indicate.

We must also remember that populations do have soft upper limits and will eventually plateau due to more concrete constraints in multiple areas of society; housing, jobs, social security, other government spending, etc.

Principal Component Analysis (PCA)

To further investigate the entire dataset retrieved from the World Bank, I used PCA to reduced the set from 418 features to 5, after further cleaning that entire dataset.



The results produced an interesting grid, plotting the 5 condensed features against each other, showing quite a successful reduction in dimensions of the dataset – it managed to find cyclical patterns within the set and could be useful for further comparative analysis.

For example, taking a look at the comparisons of feature 0,3 and 0,4, we can see an almost inverse relationship occurring – a reflection of each feature 3 and 4, once variable 0 is introduced. 2 and 3 also show similarities when 0 is introduced, with them seeming to be positively correlated, but with inverse positive amplifications – when 2 shows a strong magnitude, 3 seems to show a weaker magnitude, and vice-versa.

In conclusion, it would be good to compare this with a reduction in other countries' dimensions to see if similar patterns exist, which may perhaps even lead to key features which explain differences in the comparative graphs and therefore in real life. Perhaps rich countries would be similar, where we could find features where poorer countries critically differ.