WHICH CITY IN UK IS THE MOST LIVEABLE?

**ABSTRACT**

In order to find the most liveable city in the UK, we first needed to define a model with all the factors that the majority of people find important. After doing various researches and conducting our own survey, we came up with twelve factors and their weights.

Then, we found data for these factors. We managed to get about 80 UK cities as the number of observations. After that, we run linear regression and tested our factors for their statistical significance and eliminated the ones that were not significant at that level.

We combined the estimated coefficients from the regression with our survey weights and got the final weight by which we judge how much each factor is important for determining the liveability of a UK city. And finally we come up with the city that has the maximum combined weight of all factors considered.

**INTRODUCTION**

Moving to a new city is always challenging! UK has always been a preferred choice for a wider population in all respects, may it be for higher studies, or work, or relocating in general. We, as a group of students who have moved to Southampton recently, came up with an analysis which helps people to decide which is the most liveable city in the United Kingdom. This project aims to find the most preferred city based on chosen factors which most people find important when moving to a new city.

**BODY**

DATA COLLECTION

For various factors that determine the liveability quotient of a city, related datasets are required. Various government websites provide data for these factors with respect to the cities in UK. The data has been collected from the following sources:

1. UK Government website
2. Kaggle.com
3. National Health Survey

…and some more.

These datasets provided the data for various cities in UK. But to understand the weights of each factors affecting the liveability, as well as to ensure that the factors collected are realistic, we relied on the following two sources additionally:

1. The Happiness Ranking dataset: This is the dataset from the official government website, which provides the happiness levels of people bases on various factors. The data explains what factors determine people’s happiness, and to what level. For example, *Weather* could be a factor which influences people’s happiness. So, the happiness level would be high for this factor in particular. Therefore, the factors containing the maximum happiness level were chosen.
2. Survey: A survey was conducted on a sample size of about 120 people. They were asked to rank each factor from 1 to 10 (10 being the most important), based on their preferences when moving to a new city.

DATA CLEANSING

For the datasets collected from various government websites, consistency was a major concern. To do this, we compared all the datasets to confirm if data was available for all the cities, in all the factors. For the cities which had no data across all the factors, we eliminated them. As a result, the final count of cities ready for processing was narrowed down to 82.

For the datasets obtained after performing general formatting/cleaning, normalization process was carried out. The normalization techniques used were Z-score and Min-Max. In some cases, the cities had no numerical data. For example, if we consider Medical facilities as one of the important factors, the dataset to work upon would be ‘The Number of Hospitals each city has’. To normalize this data, the Z-score method was used, as there are no min-max scenarios in this case. Similarly, other datasets were normalized too, based on the type of data available. The Min-Max normalization method was used to normalize the weights from the survey. Further analysis was done on the normalized data.

DATA MANAGEMENT

The final datasets had different formats like a table-like structure, or a JSON file. Therefore, it was essential to have them all stored in the database for consistent storage. The storage is done on MongoDB, in the form of collections.

PROCESSING / METHODOLOGY

For the analysis, we used

1. For selecting the most important factors to work upon, the Feature selection technique called Lasso was used. The final factors are:
2. We double-checked it using F and T tests in R. Description:
3. Linear regression was used to get the estimates and standard errors, as well as, R squared of the whole model. For this, all the datasets were merged. We regressed happiness rating (the dependent variable) on all of the 12 factors (the independent variables). The result was estimated coefficients and standard errors for each of the 12 factors. Certain limitations have to be noted here: namely, some of the assumptions of linear regression might have been violated. More preciously, the zero conditional mean assumption that says that the unobserved factors in the error term should not be dependent on your independent variables. However, we are positive to suspect that there might be some factors which we could not have included in the model but which explain the happiness rating and are dependent of our independent variables. For example, there were no data for crime rates. However, crime rate can explain our dependent variable to some extent too. In summary, all this means that the estimated coefficient we got, might be and probably are slightly biased. We also violated the random sample assumption and we might have measurement errors in the reported data we got from the government websites.
4. For merging the factors and their weights, the Baye’s Decision rule was used.

APPLICATION

After performing the analysis, a web application is built, where individual users can interact with the application, and based on the results entered by the user, like the age, gender, the most preferable city would be recommended. This recommendation is a result of the analysis carried out by our model.

Another piece of recommendation could also come up from the survey results calculated before. For example, if a user enters a certain age, gender, location, etc., the application could get the data from the survey reports, and would recommend a city that was chosen by people of similar gender/age/location and others.

RESULT

As mentioned under the Application section, the result obtained will be two different recommendations, one based on our analysis, and the other one based on similar parameters from the survey conducted before.

LIMITATIONS

There are a few limitations of this model:

1. Most of the datasets are concentrated during the years 2015 and 2016.
2. There were insufficient entries for all factors across all cities. Therefore the list had to be narrowed down to 82 cities only. Further calculations are carried out on these cities.
3. In order to get unbiased estimated coefficient when running linear regression, 4 assumptions have to be satisfied. Zero conditional mean assumption was most likely violated due to not having all the factor data available. For instance, we wanted to collect data for crime rates of each of the city of interest, however, there are no data for this and so the crime rate hidden, unobserved in the error term making our estimates biased. The assumption of random sampling was violated too and finally we suspect that some of the data we downloaded might have measurement errors, for instance the population data.
4. For the survey, the sample size is small, and as a result, a bias has to be included in further calculations.
5. The survey that was conducted for calculating the importance of factors, was majorly concentrated on the people of age group 20-30. As for other age groups, there were insufficient responses and hence the application does not have the survey responses for these age groups. The same is true for location constraints. Most of the responses have been collected from Asia and Europe. Therefore, for other regions, the survey is limited. Thus we can say that the survey is not equally distributed.

REFERENCES

Data sources:

Add the website where you guys found the happiness rating data

Add the website source for education 2017

<https://data.gov.uk/dataset/noise-exposure-data-england>

<https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/townsandcitiesanalysis>

<http://www.centreforcities.org/data-tool/#graph=map&city=show-all>

<https://www.gov.uk/government/statistics/travel-time-measures-for-the-strategic-road-network-and-local-a-roads-july-2016-to-june-2017>

<https://data.gov.uk/dataset/historic-monthly-meteorological-station-data>

<https://www.theguardian.com/news/datablog/2011/may/19/train-stations-listed-rail>

<https://data.gov.uk/dataset/european-quality-of-life-survey>

<https://www.kaggle.com/getthedata/open-pubs>

<https://www.kaggle.com/PromptCloudHQ/londonbased-restaurants-reviews-on-tripadvisor>

<https://www.timeshighereducation.com/student/best-universities/best-universities-uk>

<https://data.gov.uk/dataset/historic-monthly-meteorological-station-data>

<https://data.gov.uk/dataset/european-quality-of-life-survey>

<https://data.gov.uk/dataset/hospitals_>

<https://www.getthedata.com/open-pubs>

<http://www.bbc.co.uk/news/uk-41203240>

<https://www.gov.uk/government/statistics/travel-time-measures-for-the-strategic-road-network-and-local-a-roads-july-2016-to-june-2017>

SOURCES:

We have 13 factors:

1. Flat/House prices

2. Population

3. Road Traffic

4. GVA per worker

5. Unemployment rate

6. Noise level at night

7. Total Jobs available

8. Weather

9. Number of schools

10. Entertainment

11. Quality of higher education

12. Number of hospitals

13. Connectivity to other cities

We have chosen these factors based on these articles/blog posts: ·

* + https://www.domain.com.au/news/what-makes-a-suburb-liveable-the-16-factors-that-make-or-break-a-neighbourhood-20160730-gqhdkw/ : This blog post talks about factors: 1, 3, 13, 9, 10·
  + https://www.ytravelblog.com/makes-place-livable/ : This blog post talks about factors: 8, 10, 3, 1, 5, 7 ·
  + http://theconversation.com/how-do-we-create-liveable-cities-first-we-must-work-out-the-key-ingredients-50898 : 3, 11, 10 ·
  + https://livability.com/best-places/ranking-criteria : 4, 9, 12, 1
  + http://www.city-data.com/forum/urban-planning/1694005-what-makes-city-attractive-young-people.html : 7, 10,
  + https://www.huffingtonpost.com/localeur/what-factors-make-a-city-urban-planning\_b\_5511883.html: 2, 4, 13 ·
  + http://www.agta.asn.au/conf2015/presentations/Chaffer\_L.pdf : this article talks about factors: 8, 6, 13, 10, 12, 9
  + extra source for 55db noise level and why we have chosen this level over others: <http://www.euro.who.int/en/health-topics/environment-and-health/noise/policy/who-night-noise-guidelines-for-europe>