

The most liveable city in the UK

Team Data Diggers:
Yifang Zhou
Richa Ranjan
Anton Okhotnikov
Shihuai Wang
Daniel Pelnar

CONTENTS

01 Aim

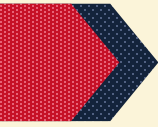
02 Implementation

03 Analysis

04 Result

05 Application

06 Limitation



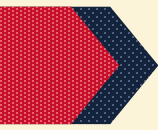
AIM

01

Finding the important/relevant factors that influence liveability of cities (citizens' happiness level)

02

Designing an interactive application to predict the most liveable city based on an individuals preferences of factors



IMPLEMENTATION

4

Data Collection

- 16 datasets in total : (weather, traffic, entertainment, infrastructure)
- UK Government data, Online Survey conducted, etc.
- Scrapy open source data

Data Cleansing

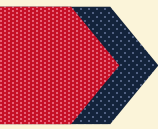
- Missing data handling- replace with average or zero
- Techniques Treating illegal characters, missing values by Python

Data Normalization

- Z-score method

Data Management

- MongoDB



ANALYSIS

5

01

Selecting Factors that determine the City Rank

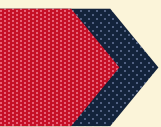
- LASSO or Random Forest
- What do we need? Importance ranking of factors
- Why not LASSO? Linear/Ranking?/No selection
- Advantage of Random Forest: Non-linear/Ranking

02

Correlation between two factors

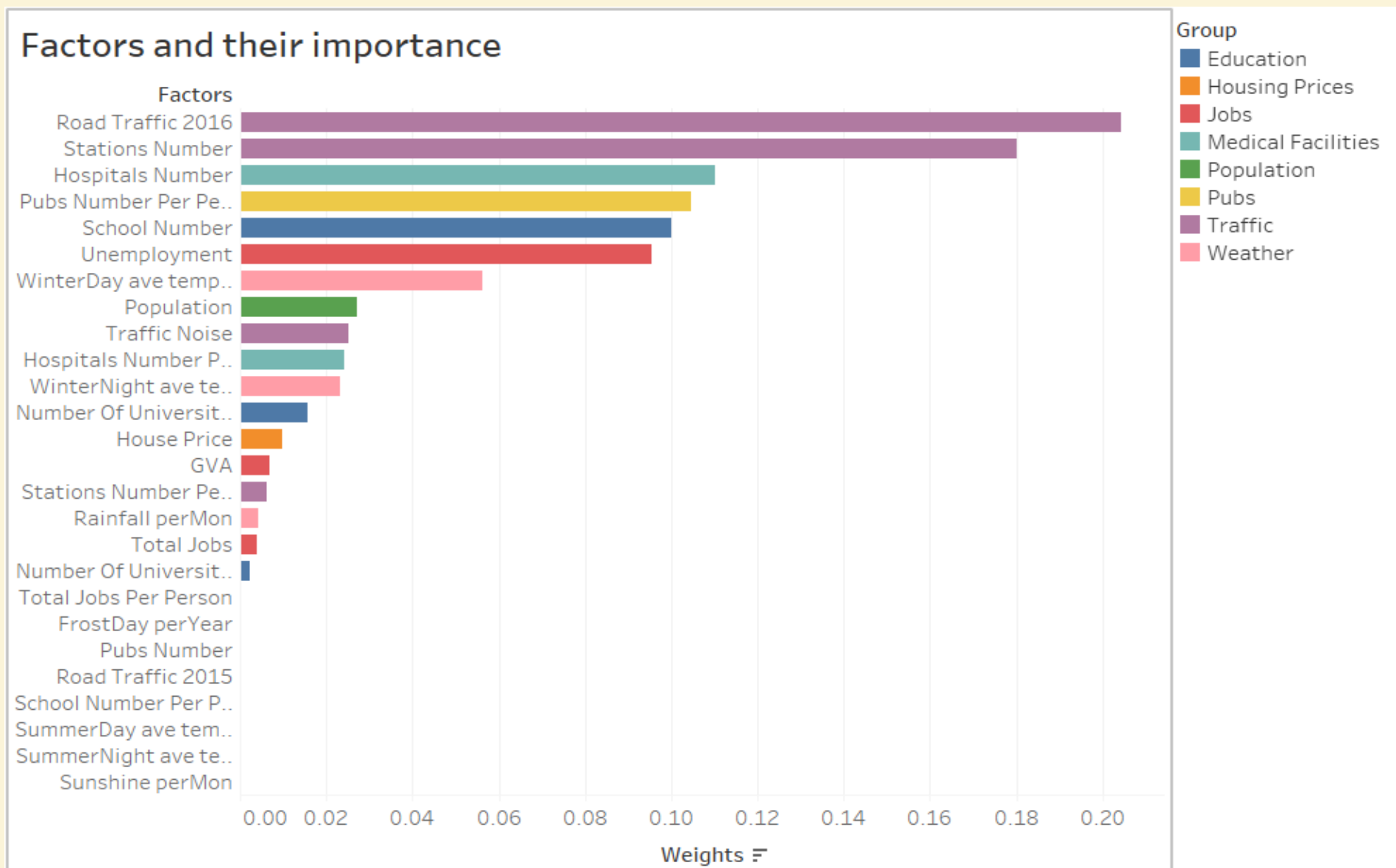
- Pearson' s Correlation Coefficient
- E.g. strong correlation with population

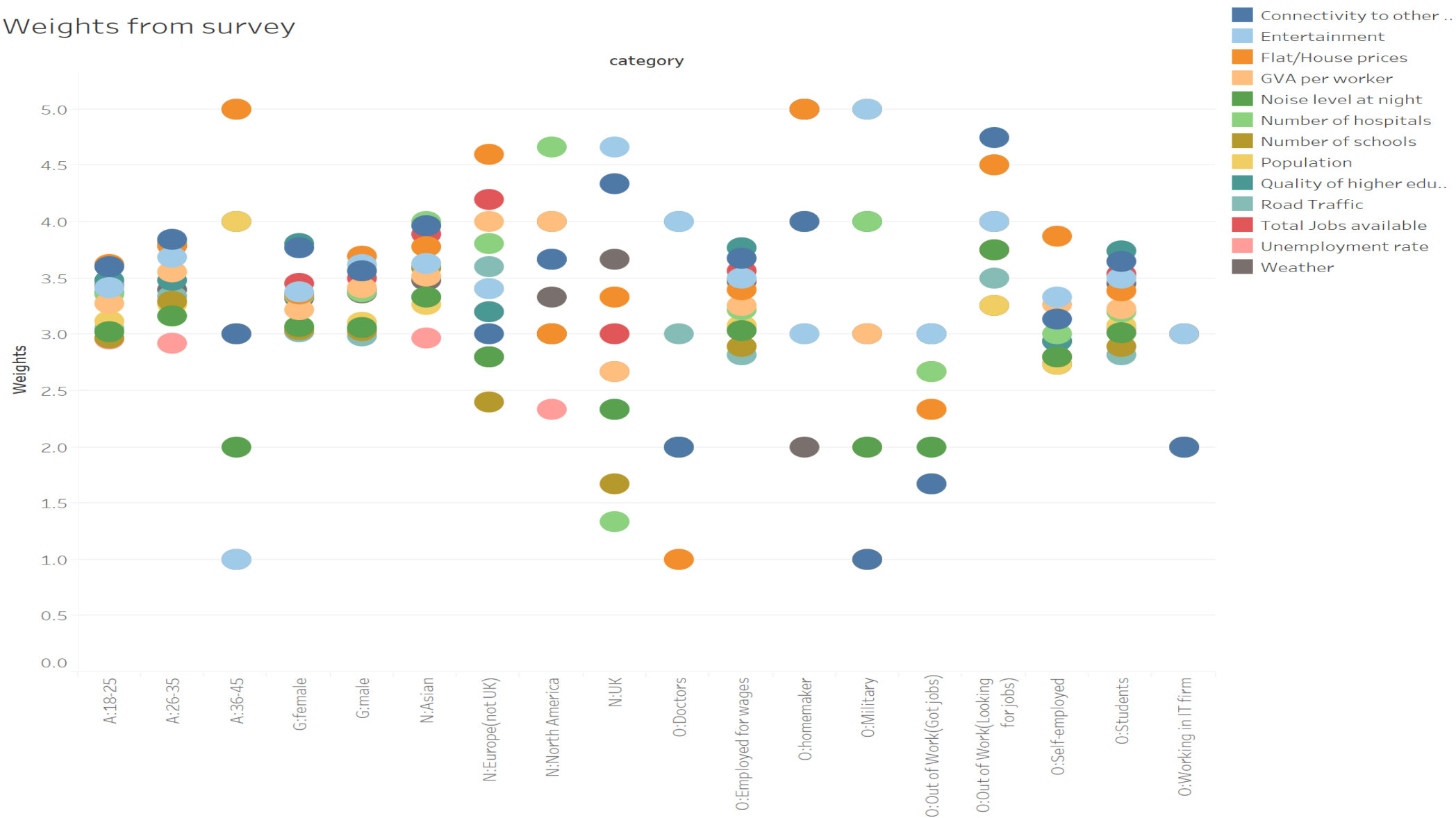
Factors	Correlation Coefficient
School	0.810281
Hospital	0.726583
Stations	0.686357
University	0.662404

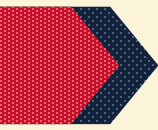


RESULTS (from the Model)

6







APPLICATION

8

Web application

- HTML, CSS and Python

User input

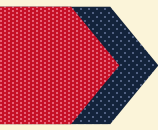
- The user enters age, gender, job, location and other factors and the results are displayed based on similar parameters obtained from the survey.
- User's factor ranking is also taken as an input

Algorithm

$$(Score_{city} = \sum_i^k r_i \cdot w1_i \cdot V_i)$$

Output

- Prediction of top-3 most liveable cities for person based on his/her preferences.



CONCLUSION

01

Survey Result: Spread of the weights for different factors is almost same across all categories

02

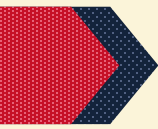
Model Result: Road Traffic (2016) was the most important factor

03

Comparison: The Weights for the Factors are similar across our model and survey.

04

Application Result: Calculated dynamically, based on inputs by user.



LIMITATIONS

10

01

Limited datasets:
data for some factors are
available only for 2015 and
2016

02

Small data sample:
the number of cities considered
is 84 and the survey responses
are also a small sample size.

03

Application: Weightage of
the Factors is not
quantifiable

04

No specific rules for
defining the weights



Questions?

Thank you for listening!

Team Data Diggers:

Yifang Zhou

Richa Ranjan

Anton Okhotnikov

Shihuai Wang

Daniel Pelnar