World's Most Livable Cities: Factors Affecting Quality of Life

Ayliah Fani
Department of Computer Science
City, University of London
London, United Kingdom
ayliah.fani@city.ac.uk

***Abstract*—The significance of predictors of quality of life and liveability of a city, along with the relationship of liveability with a country's human development index were investigated. The Human Development Index indicated a moderate linear relationship with a country's liveability, while predictor importance could not be represented in a valuable manner. More robust nonlinear modelling is required to draw advantageous insight for both the importance of predictors and the influence of a country's development.**

Keywords— human development, quality of life, liveability, liveable cities

I. Introduction

Today's world is becoming increasingly urbanised, with more than half of the global population living in urban centres, as compared to 30% in 1950. In 2018, the United Nations Department of Economic and Social Affairs stated that 68% of the world population is expected to be living in urban centres by 2050 [1]. With the rapid influx of people moving to cities, the desire and necessity of creating "liveable" cities and increased Human Development have become ever more apparent.

Liveability can be defined as a combination of physical and social characteristics that are experienced in a city- such as economic opportunities, welfare, health, safety, travel connectivity, and recreation, that drive the appeal to live, work, and invest in a city [2][3]. As such, liveability can be regarded as "issues relating to the overall quality of life and wellbeing" within a city [2].

Thirty years ago the United Nations Development Programme (UNDP) developed a measure of a country's progress in terms other than a purely economic method, Gross Domestic Product (GDP). The UNDP began ranking the world's countries by their human development: "by whether people in each country have the freedom and opportunity to live the lives they value"[4]. UNDP now releases an annual Human Development Report assigning each country a Human Development Index: "A composite index measuring average achievement in three basic dimensions of human development—a long and healthy life, knowledge and a decent standard of living" [5].

It is through these rankings and insights that the factors affecting the quality of life in a city can be evaluated and derived.

II. Analytical Questions and Data

A. Data

Data on the liveability of cities was acquired from Kaggle's "City Quality of Life" dataset based on data available through Teleport.org's API. Teleport is an online database comprised of information on cities with emphasis on liveability data, informing a user on ideal cities to live based on their criteria and preferences. This dataset rates a city on a scale of 1 to 10 for each of the following categories: housing, cost of living, start-ups, venture capital, travel connectivity, commute, business freedom, safety, healthcare, education, environmental quality, economy, taxation, internet access, leisure and culture, tolerance, and outdoors.

The second dataset ranking the HDI of world's countries was acquired from the United Nations Development Programme Human Development Data Centre.

B. Analytical Questions

With increased globalisation and urbanisation, moving to a new city has become increasingly accessible. In turn, creating a new focus on which city has the greatest quality of life to offer, not only in a national landscape but the world. It is expected that the HDI of a country and it's city's average liveability will be correlated. The following questions will aim to be answered:

- What are the most important factors for predicting liveability?
- Does a city's liveability ranking relate to its nation's Human Development Index?

By investigating these questions insights can be drawn as to what makes a city desirable to live in, allowing people to connect with urban centres, and governments the tools to build them.

III. Analysis

From these questions, the main goal is to use the data to determine the connection between economic and environmental predictors (housing, cost of living, start-ups, venture capital, travel connectivity, commute, business freedom, safety, healthcare, education, environmental quality, economy, taxation, internet access, leisure and culture, tolerance, and outdoors), and how they influence the quality of life a city has to offer. Additionally, liveability indicators of a city will be compared with the HDI of its country. This will be pursued with the methodology outlined.

A. Data Preparation

The *City Quality of Life* data set was received with mislabelled column names and an index that was mistaken as a column. Additionally, American cities' country index were labelled as the state in which it belonged to instead of the proper index of "United States of America". This was remedied by removing the redundant columns, renaming the remaining necessary columns, cleaning, checking for missing values, reshaping, and correcting the mislabelled American cities.

The *Human Development Index* dataset was received with the HDI for each country from 1990 to 2019 and several redundant columns. Only the most recent HDI of each country was of interest so HDI rankings from 1990-2018 were removed. Countries in this dataset had their official names, whereas the *City Quality of Life* dataset had countries informal names, therefore *HDI* countries were renamed. Finally, countries in the *HDI* dataset that were not represented in the *City* dataset were removed.

B. Data Derivation

Several new features were derived from the *City Quality of Life* Dataset, specifically a *City [Liveability] Score* and an *Average City [Liveability] Score* of a given country. The *city score* was derived by summing the score (based on a scale of 1 to 10) that a city had received for each respective category, dividing by the total number of points possible to score, followed by multiplying by 100. This gave a city a composite score out of 100.

The *Average City Score* represents the liveability of a country. It was derived by averaging the awarded points of each liveability predictor as well as the city score of the represented cities of a country. This yielded individual country's average city liveability indicators, and *average city score*.

These derived features were used to create two data frames: (1) *City Liveability*, detailing the points awarded to a city for each liveability indicator and composite city score, and (2) *Country Liveability and Development*, which is comprised of the average city liveability, and categorical points of which it is derived from, along with the country's respective Human Development Index.

C. Construction of Models

Two models were built to investigate the relationship between HDI and a country's liveability, and the liveability factors that define a city. First, a linear regression model was built to predict a country's *Average City Score* based on the country's *Human Development Index.*

Second, an OLS multiple regression model was built to predict a city's *liveability score* based on the predictors housing, cost of living, start-ups, venture capital, travel connectivity, commute, business freedom, safety, healthcare, education, environmental quality, economy, taxation, internet access, leisure and culture, tolerance, and outdoors.

D. Validation of Results

The linear regression model was validated by calculating the residuals and their distribution. The residuals are distributed around zero with a mean of zero, meaning that assumptions were not violated. The standard deviation of the model.

The OLS multiple regression model was validated by looking at the distribution of residuals between the actual city score and the predicted.

IV. Findings, Reflections, and Future Work

A. Country Liveability and Development

The world's *Average City Liveability Score* for a country was found to be normally distributed around 55 points (Fig.1) and skewed to the left. Although there are outliers towards the far left, it was chosen to keep them in the dataset since they could still provide valuable insights to the connection between HDI and Liveability.

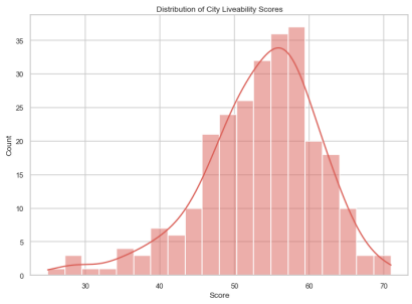

Fig. 1. Distribution of Liveability Scores

This distribution of Liveability scores was also investigated by continent (Fig. 2). The median score does differ between continents, with some continents having outliers. Again, these were not removed from the dataset since a low score could be related to other factors not perceived by the violin plot.

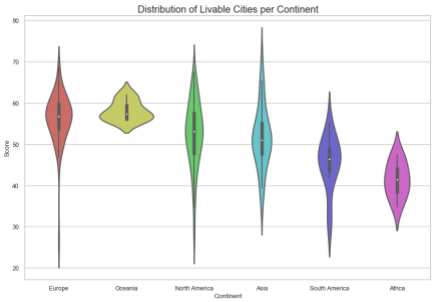

Fig. 2. Distribution of Liveability Scores by Continent

The Average City Liveability Score was plotted against its Human Development Index, and grouped by continent (Fig. 3) before fitting a linear regression model (Fig. 4). From Fig.3 an assumption can be drawn that there is a moderate positive linear relationship. This was confirmed from the linear regression model (Fig. 4a).

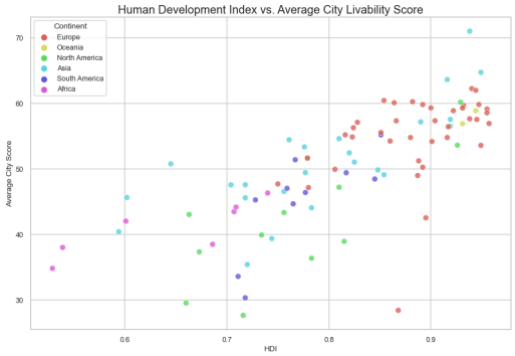

Fig. 3. Country's Human Development Index versus Average City Liveability

The linear relationship between a country's average city liveability score and that of its human development index was represented by the equation:

$$y = 63.7x - 1.75 \quad (1)$$

The R^2 value indicated a poor fit with a value of 0.55, indicating only a moderate positive relationship between the variables and partial explanation of variability of the Average City Score around its mean. There are limitations to the R2 value so residuals were also studied (Fig. 4b). The residuals are normally distributed with a mean of zero, meaning that assumptions were not violated. The standard deviation was found to be 6 points, which is not considering the average city score and the number of points available.

Fig. 4a. Average National City Liveability versus Human Development Index and Fig. 4b. Residuals of Predicted versus Actual Average City Score

B. *Predictors of Liveability*

To find the relationship between the liveability indicators and a given city's liveability score first a correlation plot was created to draw insights from (Fig.5). The largest predictor of city liveability was seen to be education, healthcare, and business freedom. Education from teleport's description of the framework is based on the quality of higher education institutes, student performance, and student wellbeing. Business freedom is comprised of performance in categories of freedom from corruption, labour restrictions, and time to open a business. Healthcare is evaluated on expenditure, quality, and life expectancy.

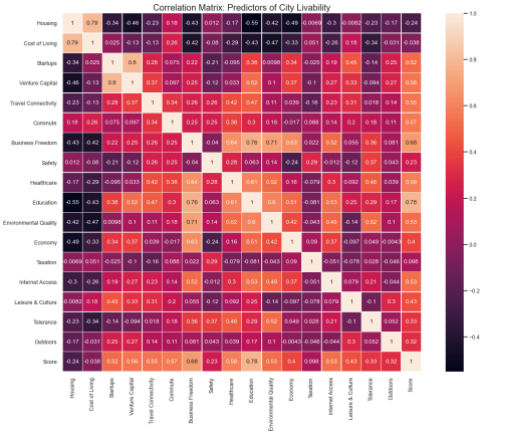

Fig. 5. Correlation Matrix of City Liveability Predictors

Since liveability is measured by several predictors, a multiple regression model was built (Fig.6). From the OLS multiple regression results the R^2 is seen to be 1, which is suspicious and indicates an overfitted model. Due to this, the model needs to be further validated.

OLS Regression Results

Dep. Variable:	Score	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	2.399e+30
Date:	Sun, 20 Dec 2020	Prob (F-statistic):	0.00
Time:	13:25:12	Log-Likelihood:	8024.5
No. Observations:	266	AIC:	-1.601e+04
Df Residuals:	248	BIC:	-1.595e+04
Df Model:	17		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0	1.41e-14	0	1.000	-2.78e-14	2.78e-14
Housing	0.5882	1.23e-15	4.8e+14	0.000	0.588	0.588
Cost of Living	0.5882	1.07e-15	5.5e+14	0.000	0.588	0.588
Startups	0.5882	1.1e-15	5.33e+14	0.000	0.588	0.588
Venture Capital	0.5882	9.48e-16	6.2e+14	0.000	0.588	0.588
Travel Connectivity	0.5882	8.35e-16	7.04e+14	0.000	0.588	0.588
Commute	0.5882	1.05e-15	5.6e+14	0.000	0.588	0.588
Business Freedom	0.5882	1.35e-15	4.35e+14	0.000	0.588	0.588
Safety	0.5882	9.85e-16	5.97e+14	0.000	0.588	0.588
Healthcare	0.5882	1.48e-15	3.97e+14	0.000	0.588	0.588
Education	0.5882	1.23e-15	4.77e+14	0.000	0.588	0.588
Environmental Quality	0.5882	9.66e-16	6.09e+14	0.000	0.588	0.588
Economy	0.5882	1.32e-15	4.46e+14	0.000	0.588	0.588
Taxation	0.5882	8.59e-16	6.85e+14	0.000	0.588	0.588
Internet Access	0.5882	8.29e-16	7.09e+14	0.000	0.588	0.588
Leisure & Culture	0.5882	8.2e-16	7.18e+14	0.000	0.588	0.588
Tolerance	0.5882	9.69e-16	6.07e+14	0.000	0.588	0.588
Outdoors	0.5882	8.57e-16	6.86e+14	0.000	0.588	0.588

Omnibus:	5.615	Durbin-Watson:	0.570
Prob(Omnibus):	0.060	Jarque-Bera (JB):	5.783
Skew:	-0.248	Prob(JB):	0.0555
Kurtosis:	3.525	Cond. No.	266.

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Fig. 6. OLS Multiple Regression Results

The actual liveability versus predicted was observed (Fig.7a), where it can be seen that there is quite a large difference between observed and predicted results. This is further seen in figure 7b describing the residuals. The residuals are normally distributed but the mean is not centred around zero. Therefore, this model gravitates to overestimate the city score. It is likely that the relationship is not linear, or that removing extreme outliers is required.

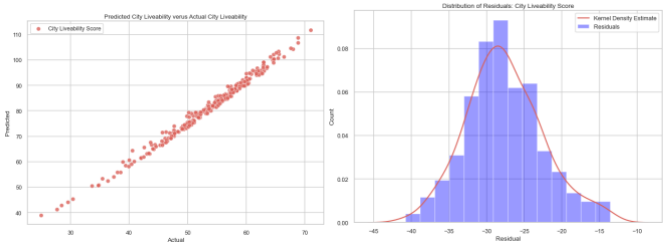

Fig. 7a. Average City Liveability vs Predicted and Fig. 7b. Distribution of City Liveability Score Residuals.

C. *Future Work*

The analysis completed during this project has provided a good foundation for future work. Models built and investigated did not represent the data as much as would be preferred, nevertheless, this has given guiding insights towards further steps. The multiple regression model underfits the data and could not gain valuable traction to predict liveability. Alternatively, while the HDI of a country did give a positive linear relationship with a country's liveability, it only gave a partial explanation to the observed scores. This is representative that more robust nonlinear models are required to make headway in regards to determining the weight of predictors and their influence on city liveability, and the relationship between HDI and Liveability witnessed across a country.

REFERENCES

[1] United Nations, Department of Economic and Social Affairs, and Population Division, *World urbanization prospects: the 2018 revision*. 2019.

[2] T. K. Giap, W. W. Thye, and G. Aw, 'A new approach to measuring the liveability of cities: the Global Liveable Cities Index', *World Rev. Sci. Technol. Sustain. Dev.*, vol. 11, no. 2, p. 176, 2014, doi: 10.1504/WRSTSD.2014.065677.

[3] M. Z. Gough, 'Reconciling Livability and Sustainability: Conceptual and Practical Implications for Planning', *J. Plan. Educ. Res.*, vol. 35, no. 2, pp. 145–160, Jun. 2015, doi: 10.1177/0739456X15570320.

[4] 'Human Development Reports', *United Nations Development Programme*. http://hdr.undp.org/en/2020-report (accessed Dec. 18, 2020).

[5] 'Human Development Report 2020, The Next Frontier: Human Development and the Anthropocene', United Nations Development Programme, New York, 2020. Accessed: Dec. 18, 2020. [Online]. Available: http://hdr.undp.org/sites/default/files/hdr2020.pdf.

Section	**Word Count**
Abstract	150
Introduction	260/300
Analytical questions and data	243/300
Analysis	568/1000
Findings, reflections and further work	594/600