



Toronto Data Workshop: How Liveable is Toronto?

Prospective Analytics

Sergio Zheng Zhou, Eric Zhu, Ashley Zhang, Muhammad Tsany

All pictures taken by Eric Zhu

About us - Prospective Analytics



Eric Zhu

Computer Science and Statistics
Deep Learning/Bayesian Modelling

[LinkedIn](#)



Sergio Zheng Zhou

Statistics and Economics
Business Analyst

[LinkedIn](#)



Ashley (Jing Yuan) Zhang

Statistics and Mathematics
Financial Modelling

[LinkedIn](#)

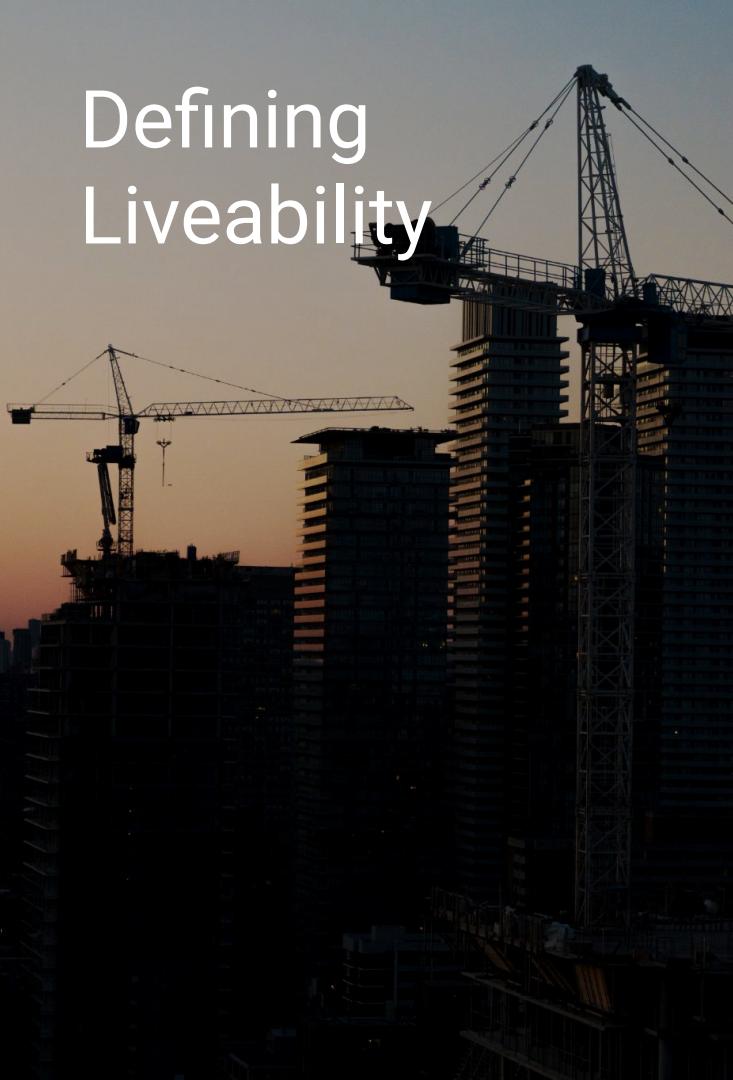


Muhammad Tsany

Statistics and Economics
Data Analytics

[LinkedIn](#)

Defining Liveability



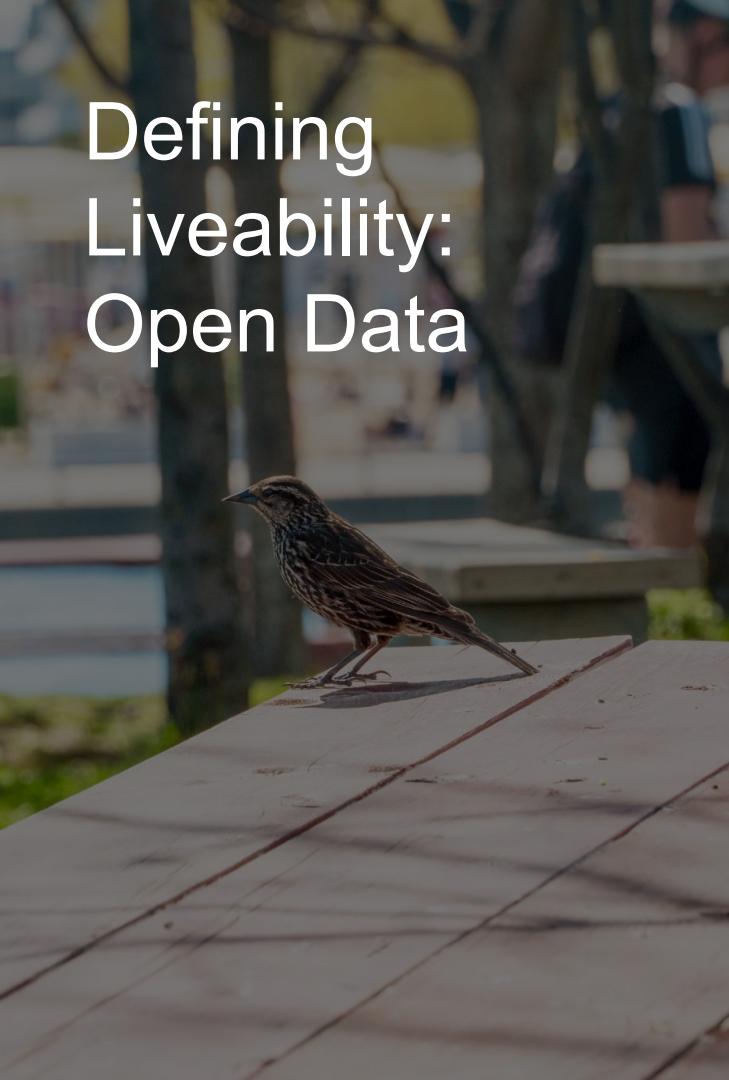
- Rating constructed based on “The Economist’s” *Global Liveability Index 2021*
- Neighbourhood liveability scores constructed with 5 weighted categories:
 - Stability (25%)
 - Healthcare (20%)
 - Education (10%)
 - Infrastructure (20%)
 - Culture and Environment (25%)

Reference:

https://pages.eiu.com/rs/753-RIO-438/images/global-liveability-index-2021-free-report.pdf?mkt_tok=NzUzLVJJUS00MzqAAAF9j-2A_8m7fB1aWnNABQKZ8QzBzVbqAs9mVmESiJFHj7htFqi7PKIjbMaV_7rgGNfvBuKEGs4Xn7DRxe5it5dGGJqofkPmLMUAOqwIYloYp4hUg



Methodology

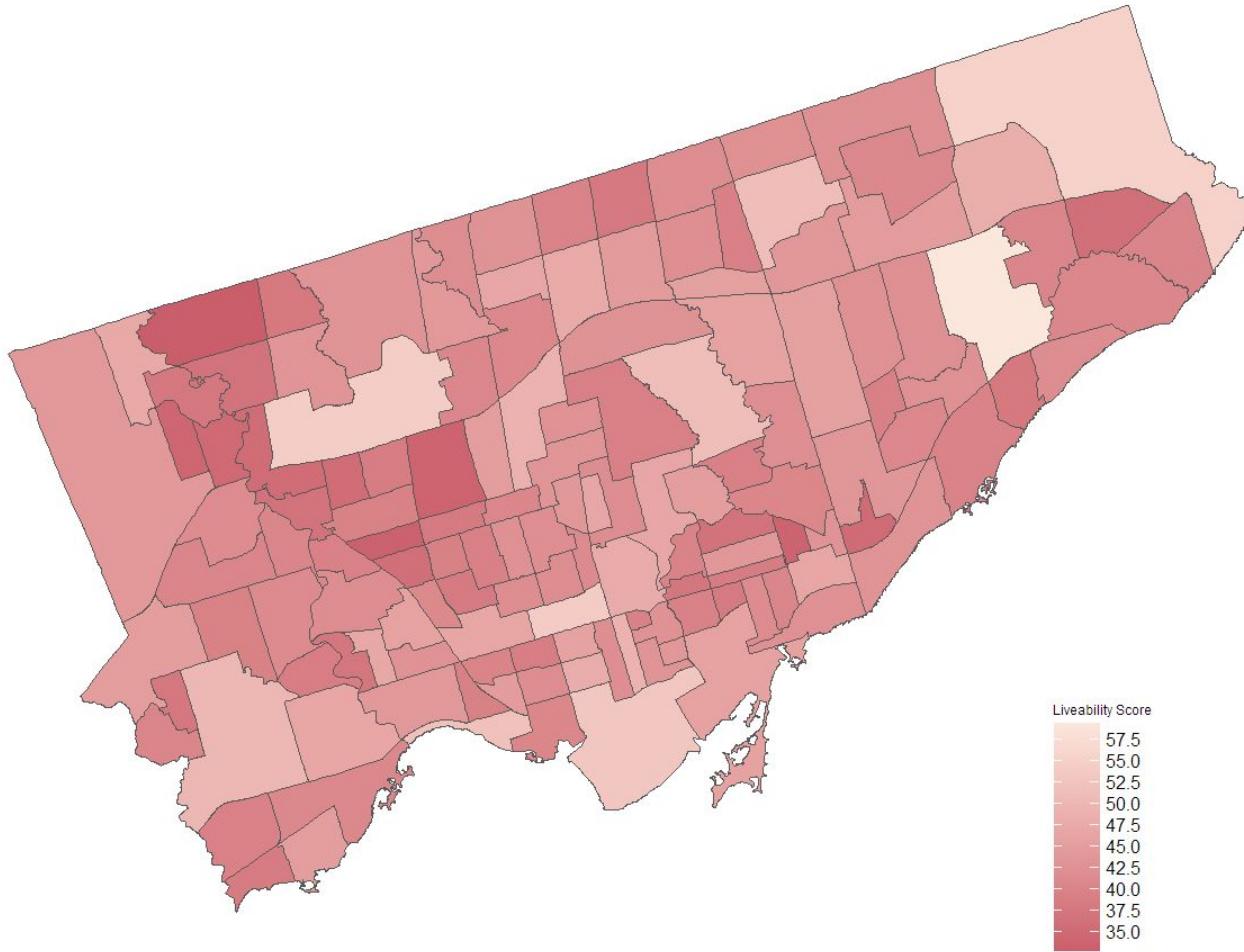


Defining Liveability: Open Data

- Stability (25%)
 - Petty and violent crime rates
- Healthcare (20%)
 - COVID-19 Vaccination and testing Sites
 - Mental health facilities
 - Healthcare providers
 - Premature mortality rates
- Education (10%)
 - Youth and adult education support centres
- Infrastructure (20%)
 - Water main breaks
 - Social housing
 - Walking score
- Culture & Environment (25%)
 - Youth recreation centres
 - Cultural spaces
 - Places of worship
 - Green space count

Liveability Scores Across Toronto

Neighbourhoods of Toronto, Canada



Clustering and Two Sample T-Test



- Clearly most neighbourhoods are darker, with some neighbourhoods having lighter colours
 - Cursory visual evidence of possible nesting structure
- Fit k-means model to differentiate
 - $k=2$
 - Examine scores in possible nesting structure
 - Cluster 1 - Higher liveability scores
 - Cluster 2 - Lower liveability scores
- Performed two sample t-test to gather evidence of approach appropriateness
 - $p\text{-value} < 2.2 \cdot 10^{-16}$
 - $T\text{-statistic} = 12.50101$
 - $df = 54.25$
 - Sample estimates
 - Mean of Cluster 1 = 47.86
 - Mean of Cluster 2 = 40.03

Defining Liveability: 2016 Statistics Canada Census

- Age
- Gender
- Income
- Education
- Ethnic Origin
- Language
- Employment

Evidence of Multicollinearity - VIF

- Used common rule of **VIF>10** as evidence of strong multicollinearity
- Most covariates had **extremely high** VIF values
 - Total Visible Minority: 262693.7
 - Employed: 125686.4
 - Working Age (25 - 54 years): 49817
- Strong evidence of multicollinearity hurts analysis
 - A “full model” using all available covariates isn’t meaningful
- Must choose covariates meaningfully
 - Arrived at analysing income levels + education levels
 - Meaningful for prospective immigrants

How does **income**
and **education** level
relate to liveability
scores across Toronto
and within higher &
lower liveability
neighbourhoods?





Linear Model and Linear Mixed Model Regression Analysis

Assumptions

- For every **linear** model constructed, the following assumptions are revised:
 - Linearity
 - Homoscedasticity
 - Independence
 - Normality
- For every **linear mixed model** constructed, the following assumptions are revised:
 - Continuous response variable
 - Independence
 - Normality
 - Homoscedasticity
 - Normality of random effect

Cluster 1: Income and Education

Estimates, 95% Confidence Interval, P-Values for Cluster 1 Income and Education Regression

	Estimate	2.5%	97.5%	pval
(Intercept)	42.2845394	39.6376080	44.9314708	0.0000
'Under \$25K'	0.0022862	-0.0001827	0.0047551	0.0683
'25K–49.9k'	0.0008594	-0.0026982	0.0044170	0.6257
'50K–89.9k'	-0.0003596	-0.0031972	0.0024779	0.7977
'\$90k and over'	0.0025931	0.0010437	0.0041426	0.0018
'No Degree'	-0.0004728	-0.0029265	0.0019810	0.6971
'Highschool Degree'	0.0013784	-0.0005656	0.0033224	0.1582
'Postsecondary Degree'	-0.0010654	-0.0021373	0.0000065	0.0513

Cluster 2: Income and Education

Estimates, 95% Confidence Interval, P-Values for Cluster 2 Income and Education Regression

	Estimate	2.5%	97.5%	pval
(Intercept)	35.9040482	34.5338855	37.2742109	0.0000
'Under \$25K'	-0.0000107	-0.0011084	0.0010870	0.9846
'25K–49.9k'	0.0027358	0.0002056	0.0052660	0.0344
'50K–89.9k'	-0.0033538	-0.0053046	-0.0014029	0.0010
'\$90k and over'	0.0014393	0.0001406	0.0027380	0.0302
'No Degree'	0.0001576	-0.0011253	0.0014405	0.8078
'Highschool Degree'	0.0000398	-0.0013306	0.0014102	0.9541
'Postsecondary Degree'	0.0004923	-0.0002615	0.0012460	0.1979

City of Toronto: Income and Education*

Estimates, 95% Confidence Interval, T-Values for Toronto Income and Education Regression

	Estimate	2.5%	97.5%	Tval
Intercept	38.9277559	1.3294764	12.1034959	13.1428
Under \$25K	0.0005405	2.0952872	2.6538764	1.0250
25K–49.9k	0.0023452	31.9255979	45.9772471	2.4787
50K–89.9k	-0.0025792	-0.0004735	0.0015546	-3.3525
\$90k and over	0.0018272	0.0005323	0.0041716	3.8714
No Degree	-0.0004103	-0.0040718	-0.0011116	-0.7109
Highschool Degree	0.0008995	0.0009234	0.0027388	1.7760
Postsecondary Degree	-0.0000606	-0.0015222	0.0006976	-0.2071

*Random intercept on liveability cluster

Conclusion

Overall Findings



- **\$90k and over** income bracket was significant in the positive direction in both clusters and at the city level.
- Contrarily, in the lower liveability scores cluster, the **\$50k-\$89.9k** bracket was significant in the **negative** direction and find **\$25k-\$49.9k** bracket has a **positive** association.
 - \$50-89.9k bracket may be a “mobility ceiling” for immigrants

Limitations and Next Steps



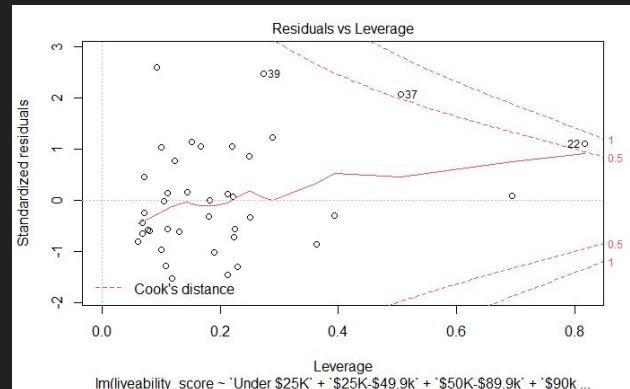
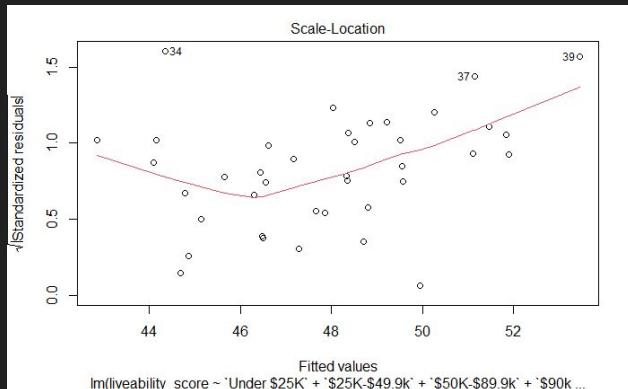
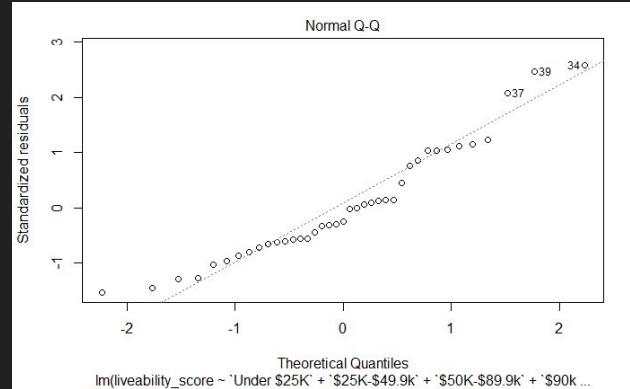
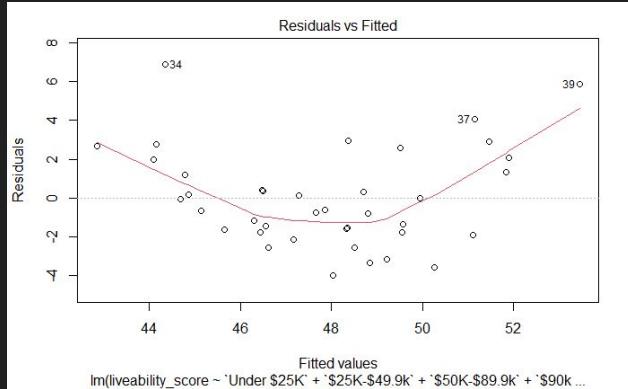
- Limitations
 - All datasets were the most recent and openly available from OpenDataToronto.
 - Confounders
- Next Steps
 - Examine individual neighbourhoods and condition education on other covariates.
 - Advise City of Toronto to look into possible gentrification of city and policy remedies.
 - Logistic modelling
 - Performing the same analysis on a country level.

The background of the image shows a wide, fast-moving river with white-capped waves. In the center, there's a small, rocky waterfall. A large, reddish-brown rock sits in the middle of the river. The banks are lined with dense green trees and bushes.

Appendix

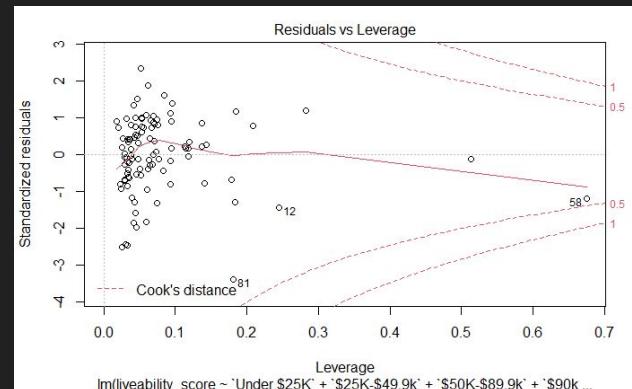
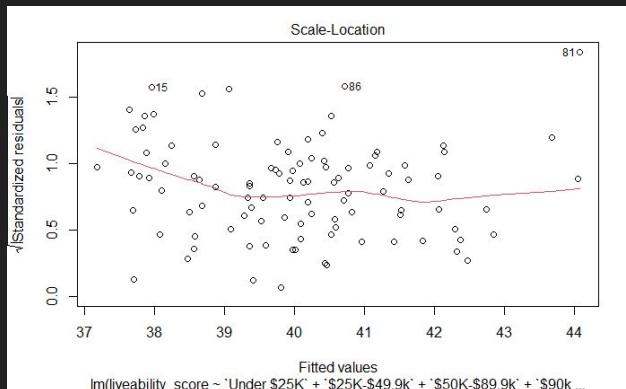
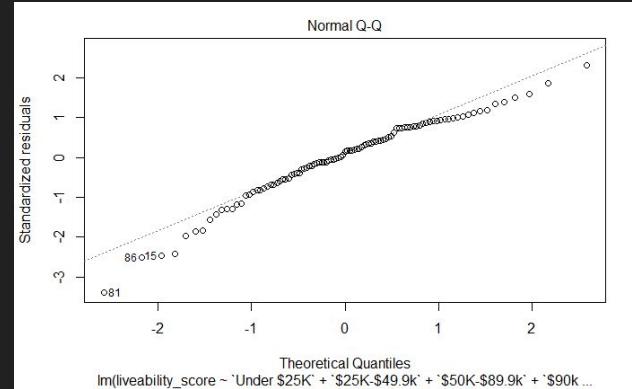
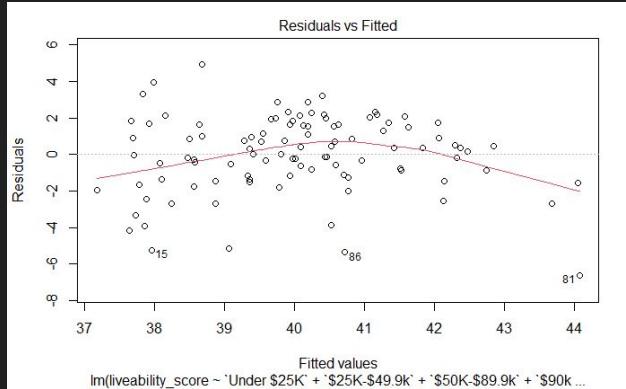
Linear Assumption plots

Cluster 1 (Higher Liveability): Income and Education



Linear Assumption plots

Cluster 2 (Lower Liveability): Income and Education



Linear Assumption plots

City of Toronto: Income and Education*

