A Statistical Analysis of Toronto Residents' Engagement and Their City Impression.*

Exploring the Correlation Between Engagement and Perception.

Nikhil Iyer

January 25, 2024

Cities go through great lengths to make sure their residents are content, and Toronto is no different. Year after year Toronto is listed as one of the best cities in the world, so how do the people actually living in the city perceive it? Through statistical analysis of survey's conducted by Toronto, the top-level finding was that the overall sentiment did not change, in fact it slightly decreased as resident engagement increased.

1 Introduction

There has been an exponential growth in the number of people living in urban areas than rural, and for the first time, more people live in urban than rural areas (United Nations Department of Economic and Social Affairs 2020). As a result, residents have a huge impact on the future of the city, and the decisions it makes. Cities are constantly being transformed by the decisions of democratically elected officials and the people (Natalie Bicknell Argerious 2020).

Generally speaking, residents control the direction of their city. Developments, Costs, Facilities, and more are done based on the demographic of the city, and what the residents want. To obtain this data, cities have a variety of collection methods such as surveys, and annual census'. Due to the amount of data greatly increasing in recent years, statistical analysis has never been more important, and analyzing resident data offers insights to guide the city in choosing it's next steps.

The remainder of this paper analyzes and discusses the results that were formed from survey data collected by the city of Toronto ("City of Toronto Open Data" 2009). The following sections will show the statistical methodologies used in this analysis, which all adhere to the best

*Code and data are available at: https://github.com/Niyer02/Toronto-Public-Engagement-Analysis data science practices. The sections can be broken down as follows: Section 1 - Introduction, Section 2 - Data, Section 3 - Results, Section 4 - Discussion, Section 5 - Appendix.

This analysis was performed in R(R Core Team 2022), using tidyverse(Wickham et al. 2019), tibble(Müller and Wickham 2023), dplyr(Wickham et al. 2023), ggplot2(Wickham 2016), viridis(Garnier et al. 2023), knitr(Xie 2023), and KableExtra(Zhu et al. 2024). The desired outcome is that the reader fully understands the methods used, and how the end result was derived.

2 Data

The raw data was collected from Open Data Toronto ("City of Toronto Open Data" 2009), and was cleaned in R (R Core Team 2022) using the tidyverse (Wickham et al. 2019) and dplyr (Wickham et al. 2023) packages. The raw data consisted of 2 types or variables. The first being engagement questions, with binary response values `(NA, 1)`. The second being perception questions with responses ranging from (Strongly agree - Strongly disagree). The cleaned data can be seen in Table 4.

Feature engineering is the process of converting raw data (Or in our case, cleaned data) into use able features. We performed feature engineering on the cleaned data set to end up with our desired features, which can be seen in Table 1:

Table 1: Analysis Data Example

engagement_prop	sentiment_average
0.25	0.45
0.55	0.50
0.10	0.20
0.65	-0.40
0.25	-1.20
0.25	0.15

2.1 Variables

The first variable is `engagement_prop` (Engagement Proportion). There were a total of 21 engagement questions, each associated with a distinct type of engagement (Survey, Consultation, etc.), this variable is the mean of all those questions. It is important to note that all engagement questions were structured such that `1` is positive (Participant engaged with the city via this method) and `0` is negative (Participant did not engage with the city via this method). Therefore, higher values of this variable can be interpreted as higher engagement with the city by the participant, while lower values can be interpreted as lower engagement with the city by the participant.

The second variable is `sentiment_average` (Sentiment Average). There were a total of 22 perception questions. Each question was structured such that `2` is a strong positive sentiment, `-2` is a strong negative sentiment, `0` is neutral, and `1` and `-1` represent a lesser positive and negative sentiment respectively and this variable is the mean of all those questions. Thus, values in the range of `[2, 0)` can be interpreted as positive sentiment, and values in the range of `(0, -2]` can be interpreted as negative sentiment, with the magnitude being represented in the numeric scale.

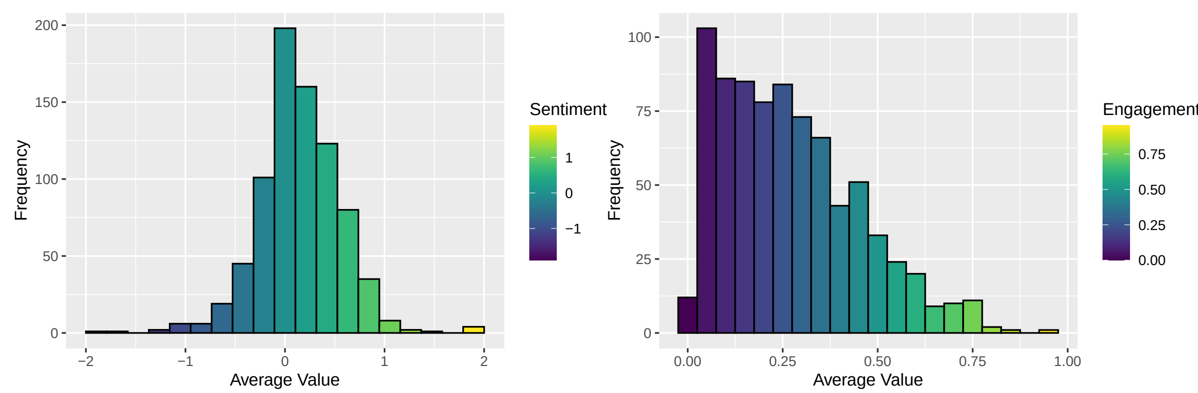
Table 2: Summary statistics for `sentiment_average`

Variable	n	Mean	SD	Median	Min	Max	Skew	Kurtosis
sentiment_average	792	0.150305	0.4159707	0.15	-2	2	-0.0347103	2.910171

Table 3: Summary statistics for `engagement_prop`

Variable	n	Mean	SD	Median	Min	Max	Skew	Kurtosis
sentiment_average	792	0.2707702	0.1802542	0.25	0	0.95	0.7326325	0.0255641

From Table 2 we can see that the mean sentiment is slightly positive, and has a standard deviation (SD) of 0.41. The skew of the graph is very slightly towards a negative sentiment. The distribution has a kurtosis of 2.9 which implies it has a high peak, and little tails. From Table 3 we can see that the mean engagement is 0.27, which is an average of 5.6 distinct methods of engagement with the city. The distribution also has a standard deviation (SD) of 0.18 The skew and kurtosis are not worth mentioning, as the distribution is modeled on binary data. The distributions of the variables can be seen in Figure 1. See Figure 1a for `sentiment_average` and Figure 1b for `engagement_prop`.



(a) Distribution of `sentiment_average`

(b) Distribution of `engagement_prop`

Figure 1: Distributions for variables of interest

3 Results

The scatter plot of `engagement_prop` and `sentiment_average` can be seen in Figure 2. The expected behavior of the trend line is a positive growth as engagement increases, however as seen in Figure 2, the trend line stays flat, and even decreases near the end. This shows a very low correlation between `engagement_prop` and `sentiment_average`. The Pearsons Correlation value is -0.007 which confirms our findings, more can be seen in the appendix. In this context, a low correlation between the two variables means that as a participant engages more and gives more data to their city, their overall sentiment towards the city remains the same, or slightly decreases. The inference drawn from this analysis is that Toronto may not prioritize its residents' opinions as much as it should, which leads to the residents' sentiment towards Toronto slightly decreasing.

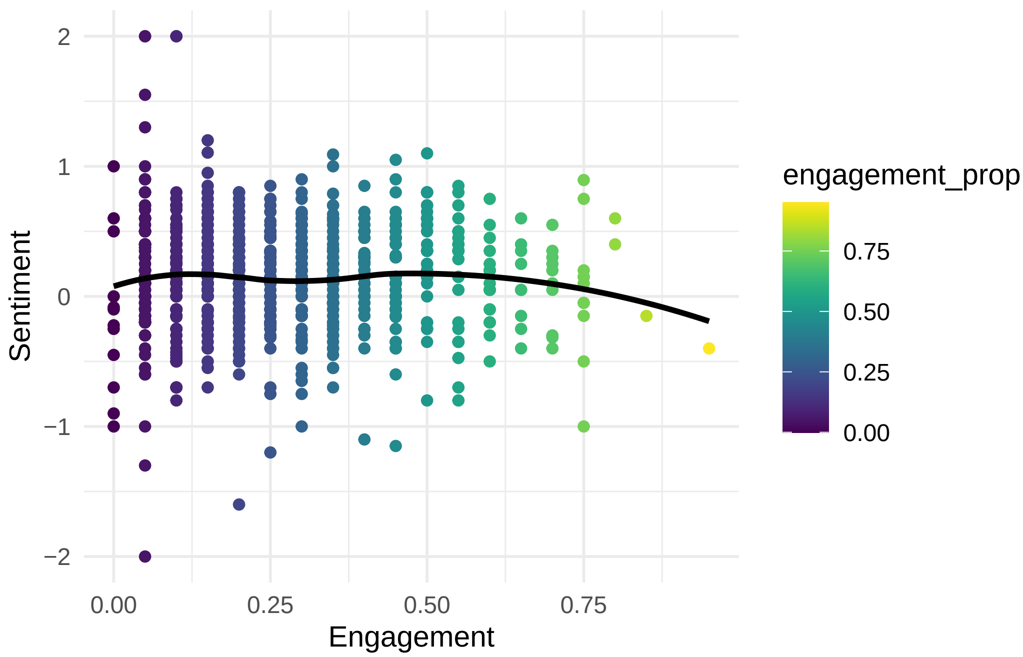


Figure 2: Scatterplot of `engagement_prop` and `sentiment_average`

4 Discussion

4.1 Discussion of the results

From the results section, we conclude that Toronto may not prioritize its residents' opinions as much as it should, which leads to the residents' sentiment towards Toronto slightly decreasing. This conclusion is somewhat worrying when looking at how a city should perform and grow. As Toronto is a democracy, the growth should reflect the populations overall decision. However the findings in this paper show otherwise. The findings instead infer that Toronto does not use its residents' opinions as much as expected, as the trend line stayed quite flat. This does not mean that Toronto is making inherently bad decisions, as that would reflect in the trend greatly decreasing, but rather means Toronto is only doing an adequate job at applying its residents' opinions and/or advice.

On average, residents that engage with Toronto more should have a positive sentiment, as their ideas and data should be reflected in some way, whether that be by changes made throughout the city, reactions from city officials, reactions from the general public, or another way. Evidence of this can be seen in neighborhoods, where residents empower each other, aid local governments, and therefore improve the outcomes (Hussey 2024). This theory should hold when applied to a larger group like a city, but falls short in Toronto. Residents' should feel empowered, and community engagement should "serve as a catalyst for changing policies, programs, and practices"(Schiavo 2021). As residents' engage more with their community and city, they "are able to gain experience in organizing, identifying resources, and developing strategies to improve the community's well being an achieve their goals"(Mary L. Ohmer and Adams 2022).

If Toronto does not change its ways, and continues to somewhat disregard its residents' there may be a major change in government officials as residents look to elect a leader that they feel represents and listens to them.

4.2 Weaknesses and next steps

Weaknesses in this analysis can be seen in the data itself, and the correlation between the two variables. The data was collected from a survey, and records residents' perception of Toronto. While this data is important, and accurate, it is also subjective. This raises concerns of how accurate the data can be when looking at the overall sentiment. While it is true that community engagement should boost overall sentiment, the data collected did not have enough coverage and variety of questions to make a hard conclusion, which is why the results show a flat and static trend line around the neutral category. Additionally, many respondents chose not to answer some questions, and while most `NA` rows were cleaned, not all of them could be. The presence of missing data points may have led to some noise in the statistics and graphs.

Next steps would be to collect more complete and comprehensive data regarding this subject. This would allow for a more in-depth analysis and more powerful statistical tests. This would also allow for better inferences, which cities could use to make more-informed decisions.

Appendix

A Cleaned Data

Table 4: Base Cleaned Data

Respondent	Engagement1	Engagement2	Sentiment1	Sentiment2
29	1	0	Agree	Agree
30	1	1	Agree	Agree
31	0	0	Neither agree nor disagree	Agree
34	1	1	Agree	Agree
35	0	0	Strongly agree	Neither agree nor disagree
36	0	1	Neither agree nor disagree	Neither agree nor disagree

B Data After Applying Quantitative Mapping

Table 5: Base Quantitative Data

Respondent	Engagement1	Engagement2	Sentiment1	Sentiment2
29	1	0	1	1
30	1	1	1	1
31	0	0	0	1
34	1	1	1	1
35	0	0	2	0
36	0	1	0	0

C Pearsons Correlation Test on Variables of Interest

```
    Pearson's product-moment correlation

data:  quantitative_data$engagement_prop and quantitative_data$sentiment_average
t = -0.20006, df = 790, p-value = 0.8415
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.07674307  0.06257699
sample estimates:
         cor
-0.007117582
```

References

"City of Toronto Open Data." 2009. https://open.toronto.ca/dataset/public-engagement-review-survey/.

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, et al. 2023. *viridis(Lite) - Colorblind-Friendly Color Maps for r*. https://doi.org/10.5281/zenodo.4679423.

Hussey, Sally. 2024. *Why Is Community Engagement Important?* https://granicus.com/blog/why-is-community-engagement-important/.

Mary L. Ohmer, Michele Mohr Carney, Amy N. Mendenhall, and Deborah Adams. 2022. "Community Engagement: Evolution, Challenges and Opportunities for Change." *Journal of Community Practice* 30 (4): 351–58. https://doi.org/10.1080/10705422.2022.2144061.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*.

Natalie Bicknell Argerious. 2020. "Democracy and Cities." https://www.theurbanist.org/2020/11/03/democracy-and-cities/.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Schiavo, Renata. 2021. "What Is True Community Engagement and Why It Matters (Now More Than Ever)." *Journal of Communication in Healthcare* 14 (2): 91–92. https://doi.org/10.1080/17538068.2021.1935569.

United Nations Department of Economic and Social Affairs. 2020. "Urbanization: Expanding Opportunities, but Deeper Divides." https://www.un.org/development/desa/en/news/social/urbanization-expanding-opportunities-but-deeper-divides.html.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao, Will Beasley Thomas Travison and Timothy Tsai and, Yihui Xie, Rob Shepherd GuangChuang Yu and Stéphane Laurent and, Yoni Sidi, Brian Salzer, George Gui, et al. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra .