February 4th, 2024

INF2178

Shion Guha

Benjamin Rabishaw

INF2178 Technical Assignment 1:

Exploratory Data Analysis

Housing is contemporarily, as it has been historically, a top priority for people and societies. Article 25 of the Universal Declaration of Human Rights states that all people have rights to "food, clothing, housing and medical care" (OHCHR). Meanwhile, however, housing availability and affordability are considered some of the top public problems facing Canada (Editorial Board, 2024). Root causes of the problem are variously identified as population growth (Al Mallees, 2023), policy shifts regarding the relationship between federal and provincial levels of government (Johnson, 2023), fundamental aspects of a capitalist socio-economic system (Durrah, 2021), or economic inflation arising out of the COVID-19 pandemic (Boisvert, 2022). Toronto specifically is considered a bell-weather municipality, sometimes associated with the decline of housing affordability in other municipalities and regions (Suhanic, 2024). As a result of this housing outlook, Toronto itself is estimated to be managing an unhoused population beyond 10,000 nightly (Jefford, 2023). This unhoused population is mainly involved with the city's network of shelters, though the city is often characterized as failing to keep up with needed shelter capacity (Jefford, 2023). So, to begin exploring Canada's broader, more chronic housing situation, we may begin at a smaller scale, by exploring Toronto's more acute shelter availability situation.

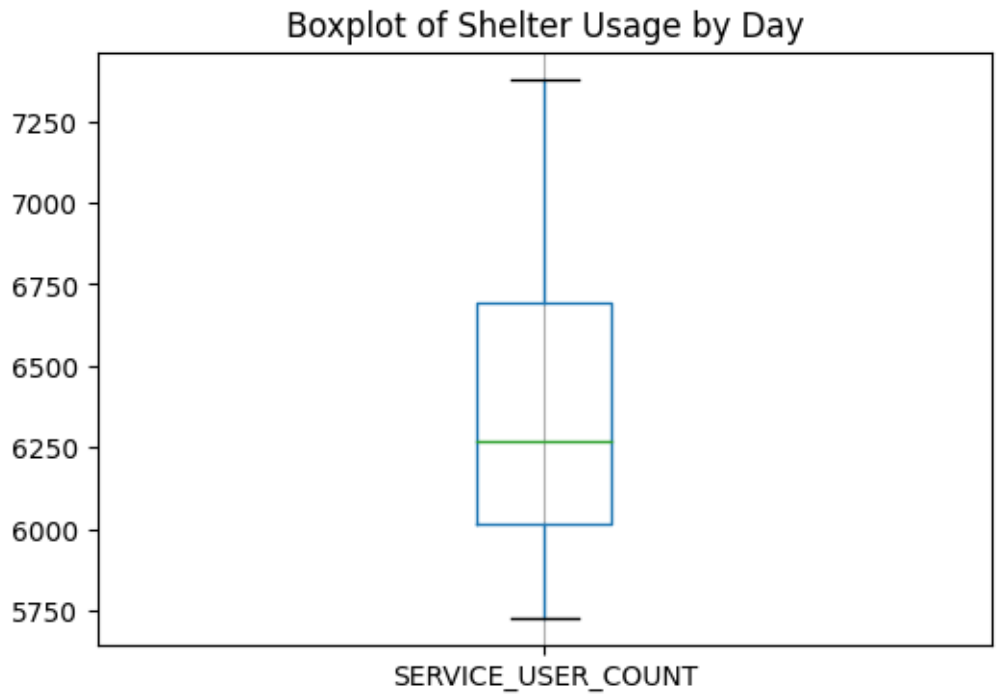
The data explored here is occupancy data for Toronto's shelters for each day of the year 2021. Data columns include: date; name and ID number of organization, name of program, sector (meaning basic demographic of people served, between families, men, women, and mixed); program model (emergency or transitional); service type; program area (meaning COVID-19 response, refugee response, and so on); user count, and capacity type (between single beds or rooms). The dataset has a length of 50,944, which contains a row for each day of the year for each shelter service available in the city.

In my exploratory data analysis of this dataset, I began with initial data cleaning in Python, and proceeded to pursue questions about the shape of the data. Univariate descriptive statistics of shelter usage via the SERVICE_USER_COUNT variable produced the following statistics:

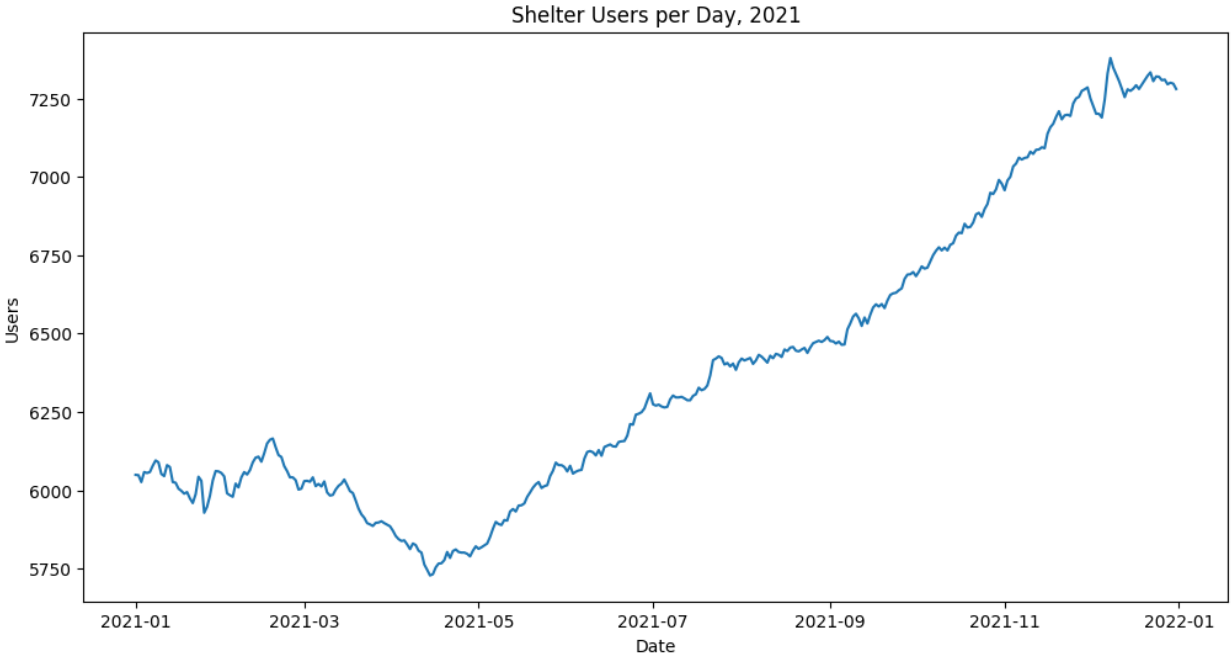
SERVICE_USER_COUNT	
count	50944.000000
mean	45.727171
std	53.326049
min	1.000000
25%	15.000000
50%	28.000000
75%	51.000000
max	339.000000

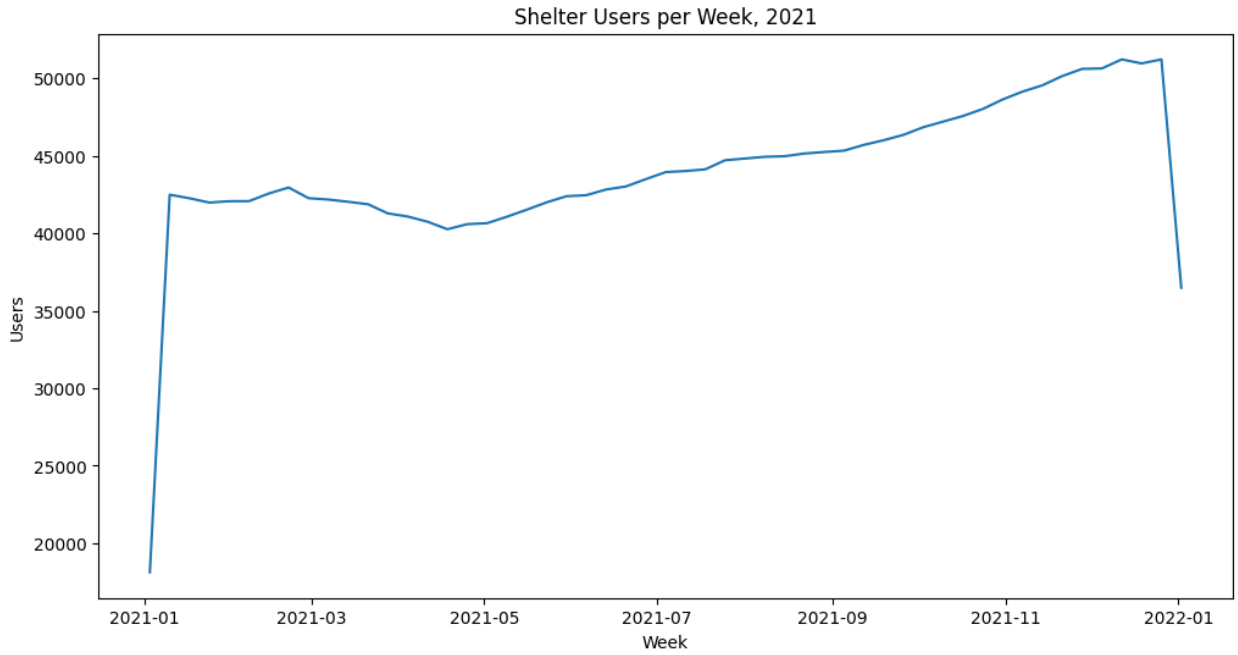
Of note with these initial descriptive statistics is that the minimum value is 1, rather than 0, indicating that there was never a scenario in 2021 where an open shelter location went unused, based on this data. The statistics here also suggest a possible skew to one aspect of the data: the 75th percentile is 51 users per program location per night, but the max is 339. This indicates that some program locations are significantly larger than others and used by significantly more people. To investigate descriptive statistics without this potential skew, by looking at the city's overall usage numbers rather than usage numbers fragmented between different locations of varying sizes, I resampled the data by day, and produced the following descriptive statistics and boxplot:

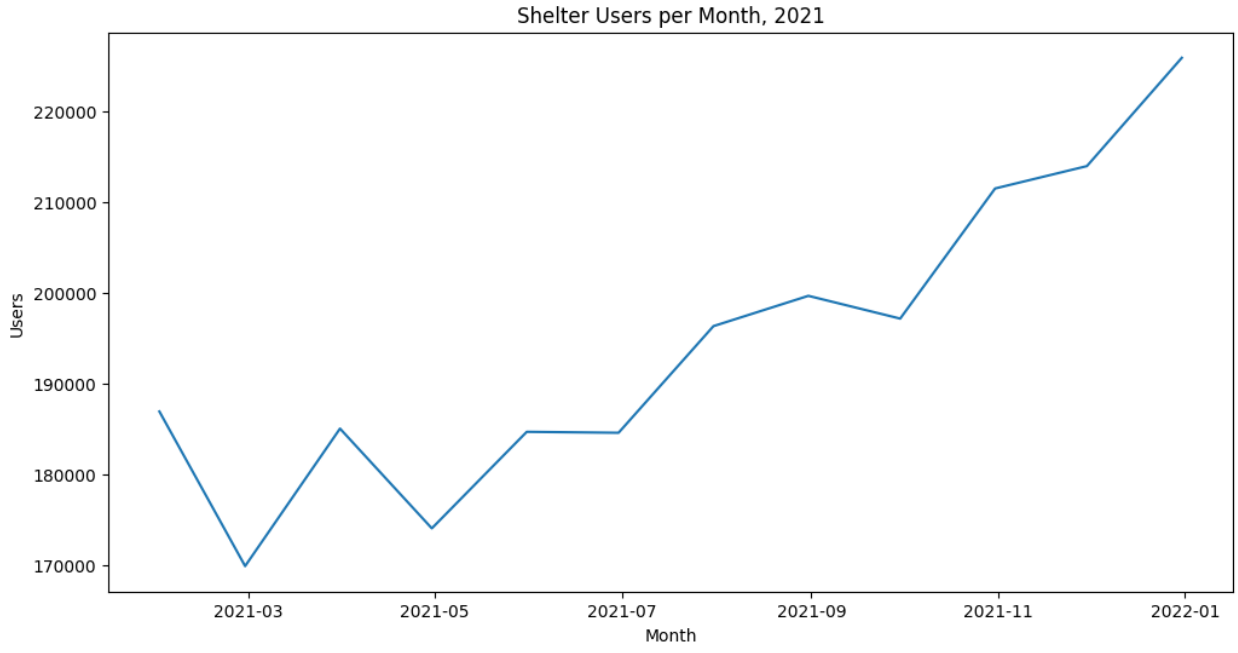
SERVICE_USER_COUNT	
count	365.000000
mean	6382.260274
std	469.984774
min	5728.000000
25%	6016.000000
50%	6270.000000
75%	6696.000000
max	7379.000000



These statistics paint a more even picture of shelter usage across the city: the IQR, meaning the difference between the 25th percentile and the 75th percentile, is only about 10% of the variable's mean. This suggests that variability in the middle 50 percentiles of the data is relatively narrow, meaning that daily usage numbers do not vary dramatically by day. To further explore this aspect of the data, I resampled by week and month, and produced the following figures:







With these, we notice a few initial potential trends and limitations to the data: first, there appears to be a marked increase in shelter usage as the year progresses, which is visible in all three figures but most noticeable with the monthly view. As well, there is an apparent dip in shelter usage during the Spring, between March and May. The week-based view, on the other hand, features clear decreases at the beginning of the year and the end of the year. Comparing the

weekly view to the daily view, we see that there were no drops in shelter usage at those moments, indicating that the shape of the weekly line graph is due to incomplete weeks of the calendar, rather than genuine decreases in shelter usage at the beginning and end of the year, as the year began partway through a week and ended partway through a week.

Proceeding from these, I formulated my first research question: are there statistically significant differences between overall shelter use over time? Guided by my visualizations, I used five separate one-sample t-tests to test whether there were significant differences between the mean of the data sampled by month and the individual months of April, May, November, and December. The results of all five one-sample t-tests are below:

```
t-statistic for March = -40.17437538019582
p-value for March = 1.2074728884529427e-27
t-statistic for April = -91.22779036119051
p-value for April = 3.2127949784583884e-37
t-statistic for May = -28.128668016946946
p-value for May = 4.0479896300111075e-23
t-statistic for November = 44.0767466506896
p-value for November = 3.9755292453323444e-28
t-statistic for December = 116.39415196869219
p-value for December = 2.111959529774661e-41
```

The t-statistics for March, April, and May are all negative, confirming that their means are below the overall monthly mean number of users. On the other hand, the t-statistics for November and December are positive, indicating increases relative to the overall monthly mean. Further, the p-values for each of these five t-tests are extremely small, indicating in each case that the difference is indeed statistically significant.

Of course, the early exploratory data analysis presented here is meager, with many more insights and discoveries possible within the rich dataset. Housing and unhoused shelter use, like any real-world issue, will demand varied analyses, of both quantitative and qualitative types, from many angles and employing many variables. What is presented here offers just the beginning of an exhaustive exploration into all the insights this data may hold, and all the clues it may provide about the experience of housing and the unhoused in Toronto, and even Canada more broadly.

References

Boisvert, N. (2022, February 2). *Inflation isn't the main factor driving Canada's sky-high housing costs, experts say | CBC news*. CBCnews. https://www.cbc.ca/news/politics/housing-inflation-conservatives-1.6335633

Darrah, D. (2021, July 9). *"Foreign Investors" aren't Causing Canada's Housing Crisis. Capitalism is.* Jacobin. https://jacobin.com/2021/09/canada-housing-crisis-foreign-investors-ban-ndp-tories

Jeffords, S. (2023, October 18). *Toronto eyes cheaper, long-term shelter model as hundreds turned away nightly | CBC News*. CBCnews. https://www.cbc.ca/news/canada/toronto/toronto-homeless-shelter-spaces-plan-1.6999159

Editorial Board (2024, February 2). *Globe Editorial: Canada is building a lot of housing. it's still not enough*. The Globe and Mail. https://www.theglobeandmail.com/opinion/editorials/article-canada-is-building-a-lot-of-housing-its-still-not-enough/

Johnson, D. (2023, September 2). *What's behind Canada's housing crisis? Decades of policy failures, says former deputy PM*. CTVNews. https://www.ctvnews.ca/business/what-s-behind-canada-s-housing-crisis-decades-of-policy-failures-says-former-deputy-pm-1.6544653

Suhanic, G. (2024, January 31). *Toronto housing crisis spreads price pain to nearby cities | sault this week.* Sault This Week. https://www.saultthisweek.com/news/housing-price-jumps-178-percent-in-this-city

United Nations. (1948). *OHCHR | Universal Declaration of Human Rights - English*. The Office of the High Commissioner for Human Rights. https://www.ohchr.org/en/human-rights/universal-declaration/translations/english