# Statistics for Data Science – Group 2 Proposal

*Submitted 2020-07-13*

*Approval Received 2020-07-14*

Members (alphabetical order):

* Walid Gherwi
* Bryan Mallinson
* Steven McAvoy
* Vernon Naidoo
* Mahshad Najafi Ragheb
* Mohit Ramnani

Our proposed topic is the occupancy and capacity of homeless shelters in Toronto, and what factors impact those measures.

Our core data set can be found here: <https://open.toronto.ca/dataset/daily-shelter-occupancy/>

The preliminary analysis will be to clean and examine the data and its patterns, and apply transformations if required. Then, we’ll look at factors to help explain trends/cycles in the data.

**Initial questions:**

1. Does geography make a difference to the type of occupancy (i.e. single person vs. family)?
2. How does the data trend overall?
3. Are there annual, monthly, daily cycles?
4. Do increases in occupancy precede increases in capacity, or do increases in capacity precede increases in occupancy?

Other factors we plan to investigate for relationships are:

|  |  |
| --- | --- |
| **Factor** | **Confidence in Data Availability** |
| Population (absolute, density) | Confident |
| Economic measures (income, unemployment, GDP, interest rates, home sales, poverty indices) | Confident |
| Shelter funding | Not Confident |
| Weather (temperature, precipitation, warnings) | Confident |
| Violent Crime | Confident |

Our focus will be to make a hypothesis about a correlation in the dataset and test the hypothesis using a statistical inference technique.

**Possible Hypotheses:**

1. Areas with higher unemployment and lower income will have a higher level of shelter usage.
2. Areas with higher violent crime will have a higher level of shelter usage.
3. Periods with extreme cold or precipitation will have a higher level of shelter usage.
4. Changes in availability of affordable housing over time can be correlated with similar changes in homelessness/shelter use.
5. Changes in measurable economic factors over time - such as employment, GDP per capita (\* Toronto population), interest rates, home sales and poverty indices - can be correlated with measurable levels of homelessness, such as shelter occupancy.