

# Assignment 1

## Assignment 1

Esther Shalini Rajasekaran

25 January 2021

### Abstract

First sentence. Second sentence. Third sentence. Fourth sentence.

## 1 Introduction

Toronto has the largest shelter system in Canada. It supports 75 shelter/respite sites sheltering approximately 7,000 people every night. 11 locations are operated by the City and 64 programs are operated by community non-profit agencies. The City has implemented a tiered approach to COVID-19 where the primary goal is to save lives and make sure the health care system does not become over. (*Housing and Homelessness Research and Reports* 2020) The dataset chosen for analysis is a listing of all active shelters serving the City of Toronto from Jan - Dec, 2020 and is published by Shelter, Support and Housing Administration. The dataset provides information on the name of the shelter, program name, sector served, addresses of the shelter, total capacity and total occupancy as of 4.00 A.M daily. Due to privacy and confidentiality the dataset does not include details regarding violence against women shelters. The total capacity also does not reflect the physical distancing measures that were implemented due to COVID-19. (*Daily Shelter Occupancy* 2021)

## 2 Data

The dataset consists of the following variables: file, OCCUPANCY\_DATE, ORGANIZATION\_NAME, SHELTER\_NAME, SHELTER\_ADDRESS, SHELTER\_CITY, SHELTER\_PROVINCE, SHELTER\_POSTAL\_CODE, FACILITY\_NAME, PROGRAM\_NAME, SECTOR, OCCUPANCY, CAPACITY and has a total 41061 rows. It provides details on occupancy, vacancy, and capacity data for the shelter and overnight services system.

1. The OCCUPANCY was chosen as the **independent variable** for our analysis and is measured on a continuous scale. It includes a count of all individuals and, where capacity is measured at the room level, the number of rooms occupied. (*Daily Shelter and Overnight Service Usage* 2021)
2. The following variables were chosen as **dependent variables** for the analysis.
  - SECTOR - This variable is measured categorically and it captures the various sectors who occupy the shelters and has the following values - Men, Women, Families, Co-ed & Youth.
3. The following variables were chosen as **control variables** for the analysis.
  - CAPACITY - This variable is measured on a continuous scale and is measured in rooms for family programs and hotel and interim housing COVID-19 response programs.
  - OCCUPANCY\_DATE - This is the date for which the occupancy was calculated for each shelter. It is captured in MM/DD/YYYY format and is measured ordinally.

Lets have a look at the data. Given below is a snapshot of the data

Table 1: Snapshot of the data

file	OCCUPANCY_DATE	ORGANIZATION_NAME	SHELTER_NAME	SHELTER_CAPACITY
1	01/01/2020	COSTI Immigrant Services	COSTI Reception Centre	100 Lippin
1	01/01/2020	COSTI Immigrant Services	COSTI Reception Centre	100 Lippin
1	01/01/2020	COSTI Immigrant Services	COSTI Reception Centre	100 Lippin
1	01/01/2020	COSTI Immigrant Services	COSTI Reception Centre	100 Lippin
1	01/01/2020	COSTI Immigrant Services	COSTI Reception Centre	100 Lippin
1	01/01/2020	Christie Ossington Neighbourhood Centre	Christie Ossington Men's Hostel	973 Lansdo

Table 2: Total Occupancy Vs Capacity based on Occupancy Dater

month	TOTALOCCUPANCY	TOTALCAPACITY	OCCUPACYCAPACITYRATIO
2020-01-01	209510	220887	0.9484940
2020-02-01	195677	207260	0.9441137
2020-03-01	204855	221286	0.9257477
2020-04-01	177962	213393	0.8339636
2020-05-01	151625	211296	0.7175952
2020-06-01	126144	191975	0.6570856
2020-07-01	125861	191462	0.6573680
2020-08-01	123682	175435	0.7050019
2020-09-01	115034	165432	0.6953552
2020-10-01	111280	168892	0.6588826
2020-11-01	102865	148175	0.6942129
2020-12-01	102148	150073	0.6806554

### 3 Data Cleaning and Preparation

The columns name are cleaned so it is easier to read .

The column name file refers to the year from which the dataset was pulled . In this case since all the data was pulled from 2020 that field can be dropped.

The date field appears to be in character hence changing it to a date field using mm/dd/yyyy format

The categorically variable - Sector is recoded to the following +Men = 1 +Women = 2 +Co-ed = 3 +Youth = 4 +Families = 5

### 4 Descriptive Analysis

To gain an understanding of the profile of the dataset, descriptive statistical techniques were used to explore the data. Frequency tables and proportion distributions were built to give us an overview of how the variables were related. Now lets look at the monthly occupancy.The table 2 below shows the monthly total occupancy and capacity and it clearly shows that there is a decrease in the percentage of monthly occupancy from Jan to December 2020.This might also be due to the fact that more people prefer to stay at their homes than moving to shelter homes due to COVID.

The total capacity for various months was calculated to identify if there are any particular months where the occupancy was high.Based on the Figure?? there is a decrease in the monthly occupancy rate from Jan to December.

Now lets look at the total occupancy Vs Capacity for various sectors.The figure?? below shows that more families tend to move to shelter homes than other sectors.

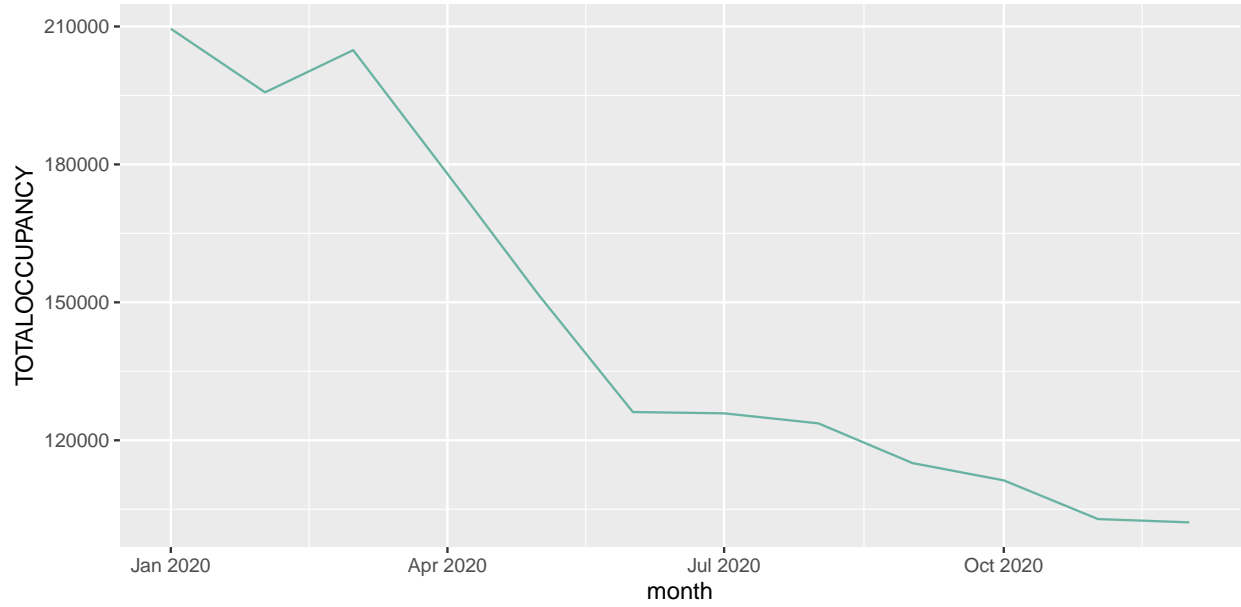


Figure 1: Monthly total Occupancy

Table 3: Total Occupancy Vs Capacity based on Sector

sector	TOTALOCCUPANCY	TOTALCAPACITY	DIFFERENCE
Co-ed	243400	304093	60693
Families	693144	966576	273432
Men	473166	589245	116079
Women	201774	252533	50759
Youth	135159	153119	17960

Table 4: Total Occupancy Vs Capacity based on Sector

sector	TOTALOCCUPANCY	TOTALCAPACITY	DIFFERENCE
Co-ed	243400	304093	60693
Families	693144	966576	273432
Men	473166	589245	116079
Women	201774	252533	50759
Youth	135159	153119	17960

The table 4 below shows that families exceed the total capacity allocated than other sectors. One reason could be become these might be large families who would occupy more be rooms than the normal family size. The details on family size is not mentioned in the dataset.

## 5 Model

### 5.1 Ordinary Least Squares (OLS) Regression

In this study due to multiple parameters affecting the total occupancy of a shelter home, to minimize the errors, regression analysis with least squares method is used. OLS reflects the relationship between X and y variables following the simple formula:  $y = b_0 + b_1X_1 + b_2X_2 + E$

Where  $+b_0$  — y — intercept  $+b_1, b_2$  — slope  $+X, X_1, X_2$  — predictor  $+y$  — Target variable

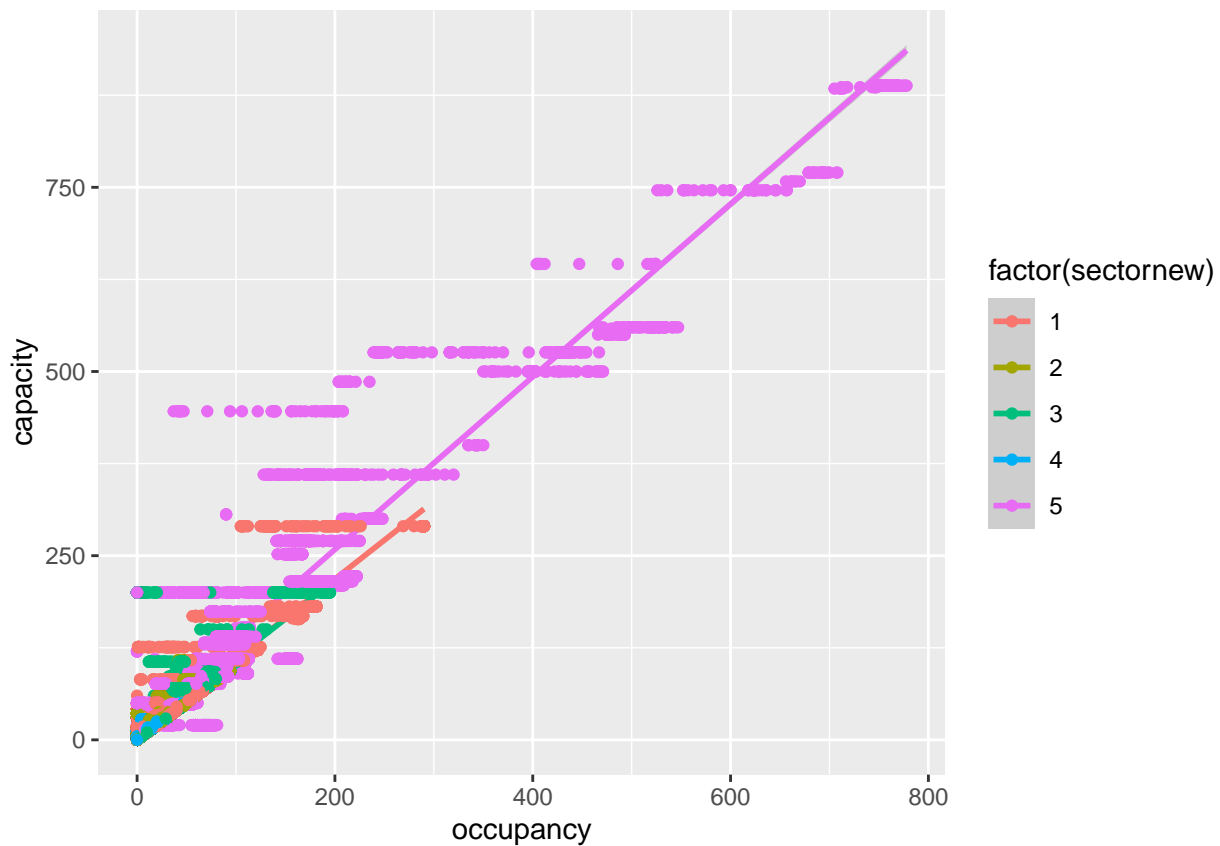
OLS is an estimator in which the values of  $b_1$  and  $b_0$  (from the above equation) are chosen in such a way as to minimize the sum of the squares of the differences between the observed dependent variable and predicted dependent variable. The analysis suggests that  $+R\text{-Square} = 88.32\%$  of the variance in total occupancy of each shelter home can be explained by various Sectors and total capacity  $+Intercept =$  The average total occupancy is 2.385 when it is Men and total capacity is 0  $+The$  coefficient for women is -1.031 and significantly different from 0, suggesting that the total occupancy for women are on an average 1.031 lower than men holding other variables constant  $+The$  coefficient for Co-ed is -0.07 suggesting that the total occupancy for co-ed are on an average 0.07 lower than men holding other variables constant. The model states that total occupancy and total capacity are not statistically significant.  $+The$  coefficient for Youth is 0.92 and significantly different from 0, suggesting that the total occupancy for Youth are on an average 0.92 higher than men holding other variables constant  $+The$  coefficient for Families is -7.43 and significantly different from 0, suggesting that the total occupancy for Families are on an average 7.43 lower than men holding other variables constant Based on the Figure??

```
##
## Call:
## lm(formula = occupancy ~ sectornew + capacity, data = datashelter2020)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -292.719   -3.356    0.680    6.838   135.088
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.385101   0.210935  11.307 < 2e-16 ***
## sectornew2   -1.031362   0.311991  -3.306 0.000948 ***
## sectornew3   -0.075352   0.343619  -0.219 0.826426
## sectornew4    0.925670   0.353269   2.620 0.008789 **
## sectornew5   -7.430636   0.381382 -19.483 < 2e-16 ***
```

```
## capacity      0.750594    0.001521 493.386 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.67 on 41055 degrees of freedom
## Multiple R-squared:  0.8832, Adjusted R-squared:  0.8832
## F-statistic: 6.208e+04 on 5 and 41055 DF,  p-value: < 2.2e-16
```

graph

```
## 'geom_smooth()' using formula 'y ~ x'
```



*Daily Shelter and Overnight Service Usage.* 2021. Shelter, Support; Housing Administration. <https://www.toronto.ca/city-government/data-research-maps/research-reports/housing-and-homelessness-research-and-reports/shelter-census/>.

*Daily Shelter Occupancy.* 2021. Shelter, Support; Housing Administration. <https://open.toronto.ca/dataset/daily-shelter-occupancy/>.

*Housing and Homelessness Research and Reports.* 2020. Shelter, Support; Housing Administration. <https://www.toronto.ca/city-government/data-research-maps/research-reports/housing-and-homelessness-research-and-reports/>.