

Dinesafe explore percentage of pass*

My subtitle if needed

Siyu Li

January 25, 2024

From the information from Open Data Toronto about Dinesafe, the dataset shows the establishment type, establishment status, inspection date, inspection issue and other factors based on Ontario Food Premises Regulation (493/17). The author uses this data to analyze the number of restaurants' status as pass, conditional pass and closed in 2022 and 2023 and compare the percentage of passes between two years. Finally, author find the percentage of pass decrease every year but the difference only 0.003% which can be ignored and the percentage of establishments have problem is over 50% which means half of establishments have problems. This paper helps to get a more general idea of how Toronto's food and drink establishments' perform, which can give readers a clear idea to explore more about food safety in the city.

1 Table of Contents

- [Introduction](#)
- [Data](#)
- [Reference](#)

2 Introduction((section-1?))

Health become a common topic with the development of living standards nowadays. Food safety is one of the most important parts of health. Food safety is the science of handling, preparing and storing food to reduce the risk of foodborne illnesses. Food contaminated with harmful bacteria, viruses, parasites or chemical substances can cause many illnesses(Public Health Ontario).

*Code and data are available at: https://github.com/SiyuLii/Dinesafe_Analysis.git

This paper explores the number of establishments that satisfy the Ontario Food Premises Regulation (493/17) which is enacted by the Ontario government and how it is distributed. This data is highly related to public health and the tourism industry. By comparing the number of passed establishments, readers can easily know if the number is increased, decreased or kept constant. The number of pass establishments corresponds to the sanitary condition of the city. If the condition for establishments violates the enactment, it may cause foodborne diseases. Foodborne diseases impede socioeconomic development by straining healthcare systems and harming national economies, tourism and trade(World Health Organization). As an indicator, it is necessary to have the data on establishment status. In another way, this data reflects living standard of citizens ,which estimates the ability of the government. In this paper, author will explore the how the number of passed and conditional pass restaurant changes in years and how the number of different severity changed.

The reminder of this paper is construed with data section Section 3 and reference section. Section 3 The Data section discusses origin of data, how the data is distributed by years and establishment status. Then, compare the establishment status by years. This lets the difference and change can be observed directly.

3 Data

The data is from OpenDataToronto([opendatatoronto?](https://opendatatoronto.org/)) by city of Toronto. Only one dataset is used. The original dataset include seventeen variables which is id of the establishments, address, number of inspection per year and other relative variables. However, some of those variables are not related to author’s hypothesis. Thus, only three variables are selected for the analysis. Those three variables are establishment status, inspection date and severity. With those variables’ data, it express how the establishments perform and number of establishments pass the inspection. This can directly

Author analysis data in R(R Core Team 2022) by using package ‘tidyver()’(Wickham et al. 2019) and ‘knitr()’(Xie 2014) to generate the table. Also,‘summarise()’(Wickham et al. 2023) used to get number of pass, conditional pass, and different severity level. This form to create Table(Table 1).

Table 1: Number of Establishment Get Pass, Conditional Status in Different Year

Status	Severity	Count
2022	Conditional Pass	64
2022	Pass	27260
2023	Conditional Pass	177
2023	Pass	46617
2024	Conditional Pass	150
2024	Pass	2538

Status	Severity	Count
NA	Pass	415

Next, by using similar package to get table for severity. In this table(Table 2), it shows five level of severity of inspection. By the table, it obviously show that for number of state is empty(refer to establishment have no problem during inspection) is the most. Establishments have minor state be the second large, which means those establishments have small problem but it is not serious.

Table 2: Number of Establishment Have Five Different Severity Level From 2022 to 2024

Inspection Year	Severity	Count
2022	C - Crucial	867
2022	M - Minor	9609
2022	NA - Not Applicable	1089
2022	S - Significant	5051
2022	NA	10708
2023	C - Crucial	1251
2023	M - Minor	17635
2023	NA - Not Applicable	2161
2023	S - Significant	8917
2023	NA	16830
2024	C - Crucial	73
2024	M - Minor	933
2024	NA - Not Applicable	103
2024	S - Significant	497
2024	NA	1082
NA	NA	415

By the first table(Table 1), it shows there is increase in number of establishments in 2023. It implies more establishments open in 2023 compare to 2022. In this table, it include data in 2024. It is obvious the number of establishments is smaller than 2022 and 2023.

```
#| label: fig-establishment_status
#| fig-cap: Number of Pass And Conditional Pass From 2022 to 2024
#| echo: false
#| warning: false
#| message: false

# this code from Professor Rohan's note
```

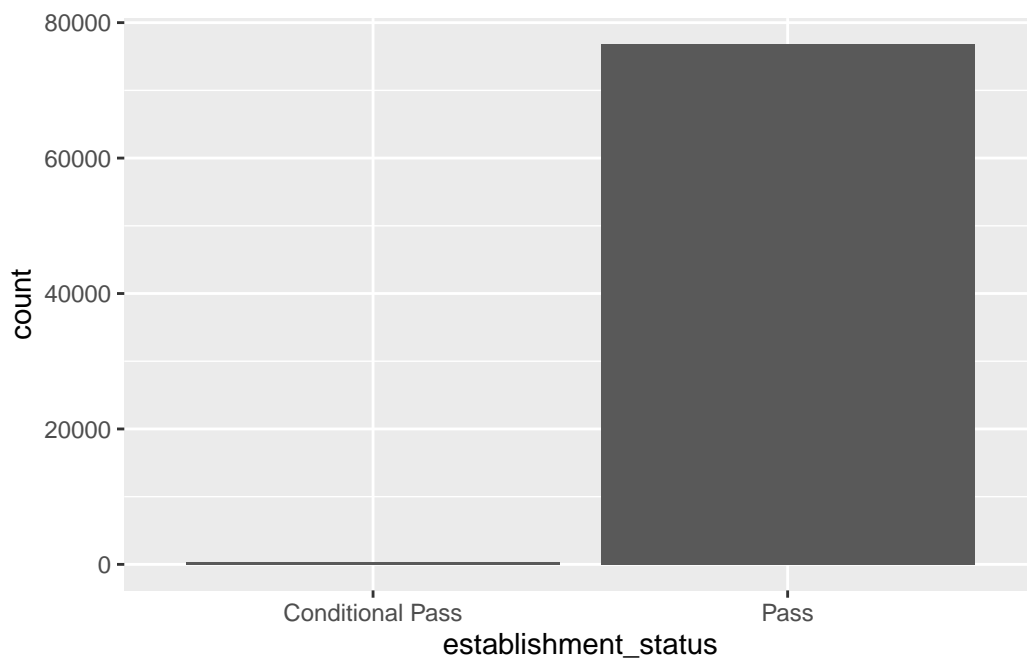
```
# https://tellingstorieswithdata.com/03-workflow.html
```

```
data("dinesafe_program_clean", package = "AER")
```

```
Warning in data("dinesafe_program_clean", package = "AER"): data set  
'dinesafe_program_clean' not found
```

```
dinesafe_program_clean |>  
  ggplot(aes(x = establishment_status)) +  
  geom_histogram(stat = "count")
```

```
Warning in geom_histogram(stat = "count"): Ignoring unknown parameters:  
`binwidth`, `bins`, and `pad`
```



By the figure(Figure 1)above, it shows the distribution of different level. From it, it express that most of establishments in Toronto have no problems during inspection and the number of significant is large. This means there are certain number of establishments exist.

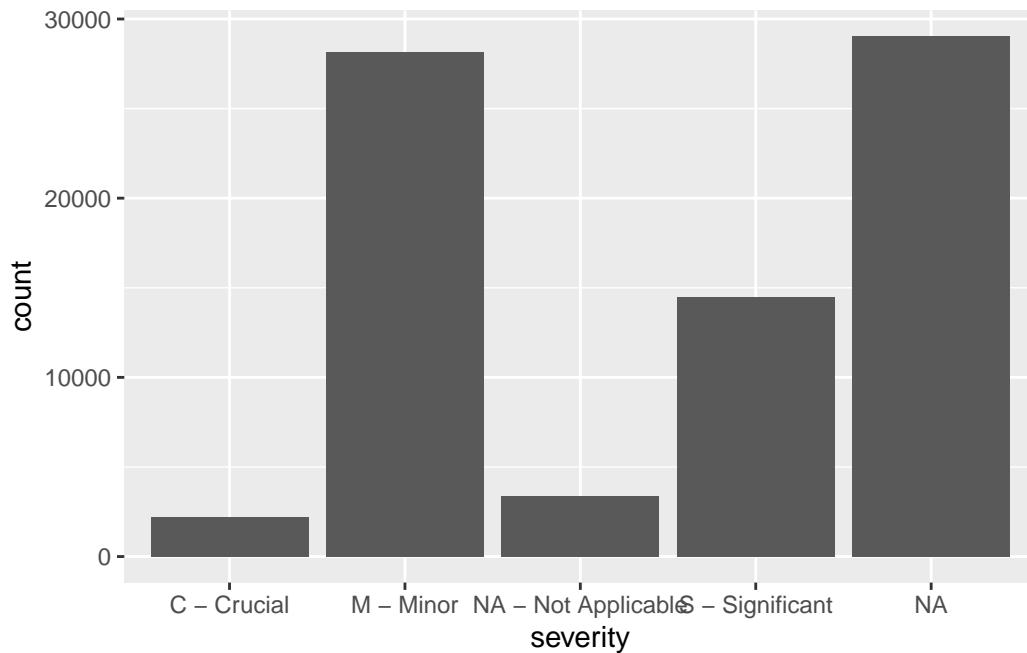


Figure 1: Number of Five Different Severity Level From 2022 to 2024

References

- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, 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. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.