

When is your next inspection?

Springboard Capstone Project

by Jonas Makonnen

Introduction

- ▶ Goal: Prediction of the number of days until next restaurant inspection
- ▶ Data: Open Data on restaurant inspections in NYC from 2011 to 2018

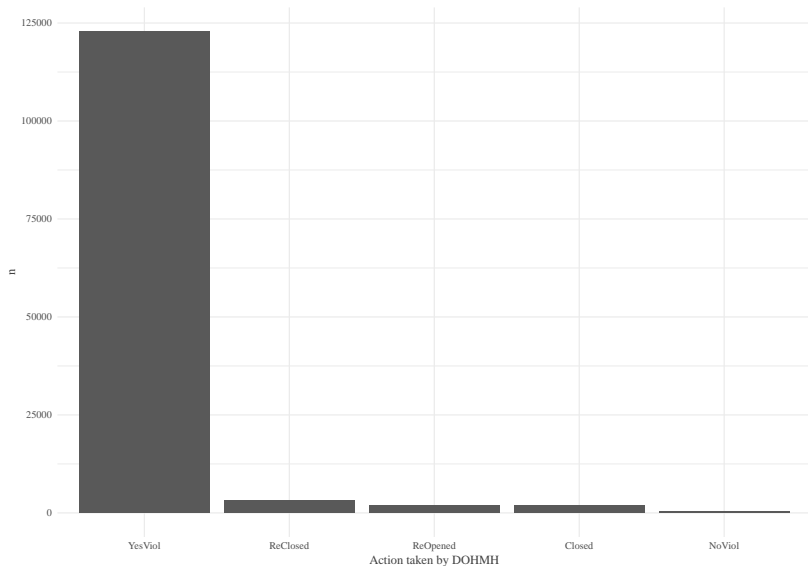
Data

- ▶ Restaurant inspections carried out by the Department of Mental Health and Hygiene (DOHMH) in New York City
- ▶ Information about inspection date, name of the restaurant, cuisine type, violations, grade, actions taken by the DOHMH etc.

##	[1]	"id"	"rest_name"
##	[3]	"boro"	"building"
##	[5]	"street"	"zipcode"
##	[7]	"phone"	"cuisine_descr"
##	[9]	"inspection_date"	"action"
##	[11]	"violation_code"	"violation_descr"
##	[13]	"critical_flag"	"score"
##	[15]	"grade"	"grade_date"
##	[17]	"record_date"	"violation_group"
##	[19]	"viol_vermin"	"viol_not_scored"
##	[21]	"viol_facility"	"viol_food_temperature"
##	[23]	"viol_hygiene"	"viol_food_protection"

Exploratory Data Analysis (Action taken)

Records of violations in the vast majority of cases.



Exploratory Data Analysis (Violation type) - 1

10F (general violation pertaining to non-food contact surfaces) is the most common violation type. followed by 08A (facility not vermin proof), 04L (Evidence of mice):

```
## # A tibble: 10 x 3
```

##	violation_code	number	rank
##	<chr>	<int>	<dbl>
##	1 10F	53154	1
##	2 08A	38706	2
##	3 04L	26768	3
##	4 06C	25521	4
##	5 06D	25215	5
##	6 02G	23866	6
##	7 10B	21819	7
##	8 02B	19023	8
##	9 04N	18882	9
##	10 04H	8174	10

Exploratory Data Analysis (Violation type) - 2

We can look at the frequencies while organizing codes into broader groups:

