Food Violations Report

Zachary Cripps
2810ICT Software Technologies
October 6, 2019

Abstract

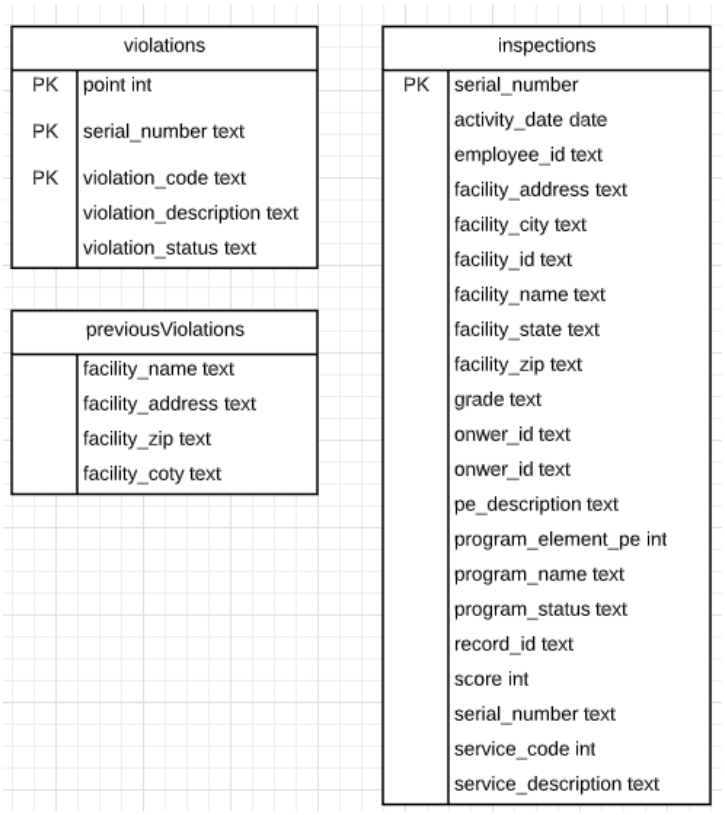
Between 2015 and 2017, an inspection of 191,371 businesses were conducted in California. It was found that there were 905,891 total food inspection and health violations. The highest contributes being large franchises that cater to hundreds of people per day. The highest violation codes that were issued to businesses were codes F044, F033 and F035 - all codes that represent failure in maintenance and cleanliness of facilities and tools for the restaurant. From these, fast food chains McDonalds and Burger King were compared with the number of violations per month. Comparison data shows that on average Burger King performed better with less violations. The spike in violations for both restaurants can also be assumed from being byproducts of public holidays or other events that occurred during the time period. However, the data is proven to be inaccurate as the number of stores owned by both facilities are too great in difference to make judgmental conclusions.

Introduction

Today, industries are competing in an increasingly data-driven world. The interrogation of raw data from an unusable form to produce meaningful information is critical for companies to hone their competitive edge and deliver critical business insights. This report analyses the 905,891 total food inspection and health violations that were issued following the inspection of 191,371 businesses in California over the period 2015 to 2017.

Database Structure

For this report, three tables (Inspections, Violations and Previous Violations) were created to be used for extrapolating data and representing data in a visual format. The tables were created from data presented in excel spread sheets that reported the inspections made and the violations linked to those inspections. Table inspections and violations schema were created as a mirror copy of the two excel spread sheets with keeping the same attributes and table names. Table Previous Violations was created based on the required information to be presented in Task 3. Presented below are the table schemas and the reasoning for attribute types.



All attribute types were given based on the information provided in the excel spread sheets. The type "text" was used for strings of words or characters while "ints" for numbers/integers with the exception of dates as they have a special data type called "date". It is worth noting that a primary key was added to the Violations Table to eliminate duplicate records when inserting values into the database.

Violation counts

Figure 1: The 35 highest violation codes issued between 2015 and 2017

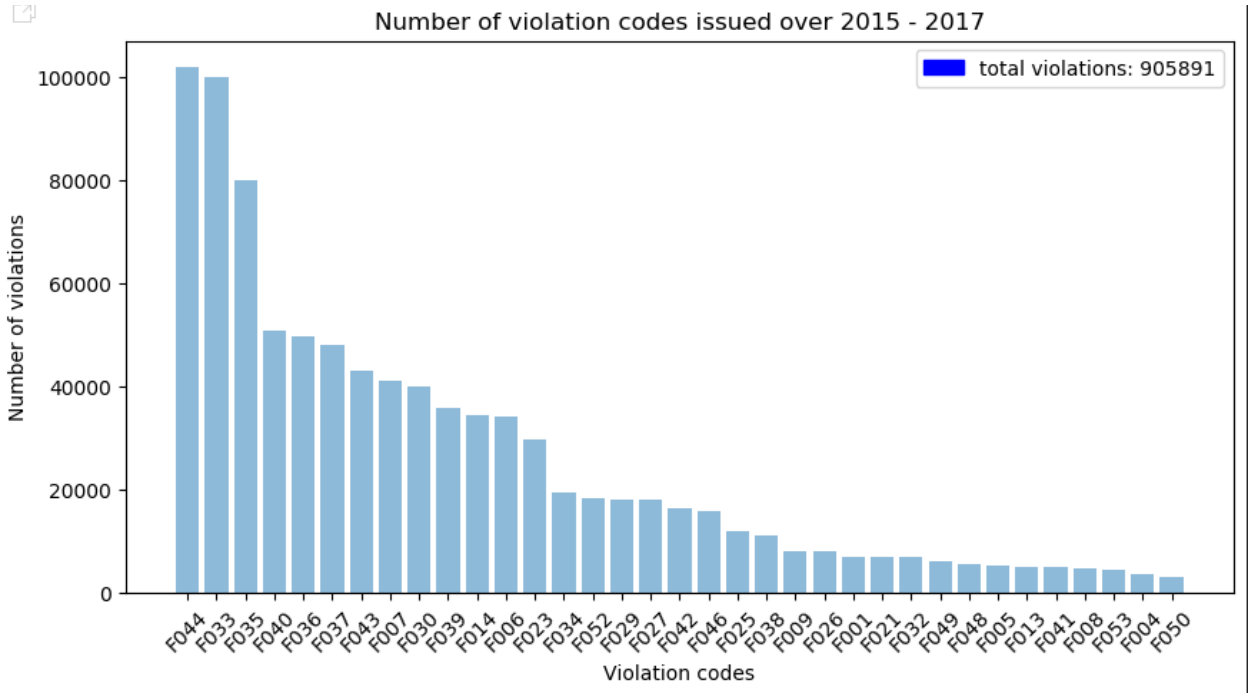


Figure 1 shows the 35 highest violation codes issued between 2015 and 2017 as a frequency histogram from highest violation code to lowest violation code. The source material used to create Figure 1 can be found in the Appendix section 2.

Between the years of 2015 and 2017 there had been a total of 905,891 food violations and safety issues across all business. Investigation into the excel spread sheet for 'violation types' shows that the code number F044, which is issued for restaurants that violate having not properly built and maintained facilities, was the highest grossing issue for restaurants ("# F044. Floors, walls and ceilings: properly built, maintained in good repair and clean"). The second and third highest counted violation codes were #F033 and #F035 ("#F033. Nonfood-contact surfaces clean and in good repair", "#F035. Equipment/Utensils - approved; installed; clean; good repair, capacity"). From these results it can be interpreted that the majority of restaurants fail in their inspections in the maintenance and cleanliness of facilities and tools. It would then be recommended to handle these issues and provide a workable solution before the next schedule inspection of the business.

The highest single store with the largest number of violations codes was "DODGER STADIUM" having 1,140 violations codes. With the fewest number of codes issued to smaller franchise stores such as "YUM YUM DONUTS" and "TRAVEL TRADERS" having a single count of 1 for each store under the franchise name. From this investigation only 1.59% of the restaurants that had been inspected received 1 to 10 violation codes. However, this information is not representative of all restaurants as there is a correlation between the high number of violations to the number of restaurants that are owned by a single franchise. Therefore, it can be assumed that the restaurants that are within the 1.59% range are single stores or have

fewer than 5 stores owned by the franchise. Figures used in this paragraph were extracted from excel_food.txt.

$$\frac{Number\ of\ stores\ with\ count(violations\ code) <= 10}{Total\ number\ of\ violation\ codes} \gg \frac{14478}{905891} \gg (0.0159)X100 \ \therefore = 1.59\%$$

Violations over time

Figure 2: Violations per month for the postcode with the highest, lowest and average total violations

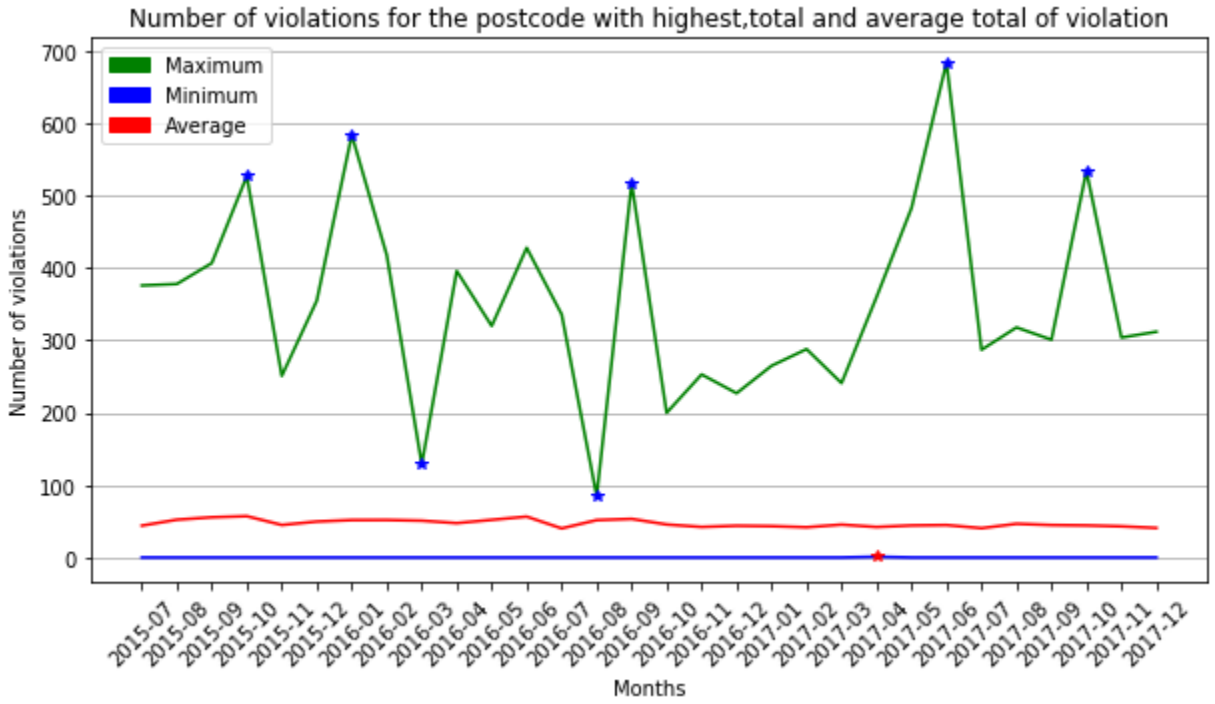


Figure 2 shows the line graph for the number of violations for the postcode with the highest and lowest number violations and number of violations for all California postcodes. Raw data of both postcodes and the average can be found in the Appendix section 3.

The results were calculated on a monthly basis. One of the lowest postcode selected was 91801-746 out of the total 525 restaurants that received 1 violation. This code area had a violation count of 1 which occurred in March of 2017 (marked by a red asterisk). The highest postcode selected was 97148 having a maximum of 684 violations in June 2017 and lowest count occurring in August 2016 with 85 total violations.

Reasons for the drastic changes in number of violations codes for postcode 97148 are assumed to rely on the fact that in that code area, 2,011 businesses exist which is an extremely large compared to the single business in post code area 91801-746. With 97148 containing 2,011 businesses, the overall number of violation counts would be extremely high but further assumptions can be made for certain data points. The sudden spikes of violations can be assumed to occur during public holidays and significant events such as

Black Friday in November and Christmas holidays which would result in increased activity for businesses (marked by blue asterisks). Months where there are low violations such as the sudden large drop in August in 2016 can be assumed as a trigger from the California wildfire which resulted in the evacuation of 80,000 civilians. When there were no large peaks in the number of violations for post code 97148, the average number approximated 200-351 recorded violations.

After evaluation of the average number of violations for each postcode in California, the recorded amount was shown to have a gradient of almost 0 with violation counts ranging between 40 and 50. These are not perfect results but represent a better indication of the quality of businesses compared to the highest postcode.

Figure 3: The average number of violations per month for all McDonalds and Burger Kings.

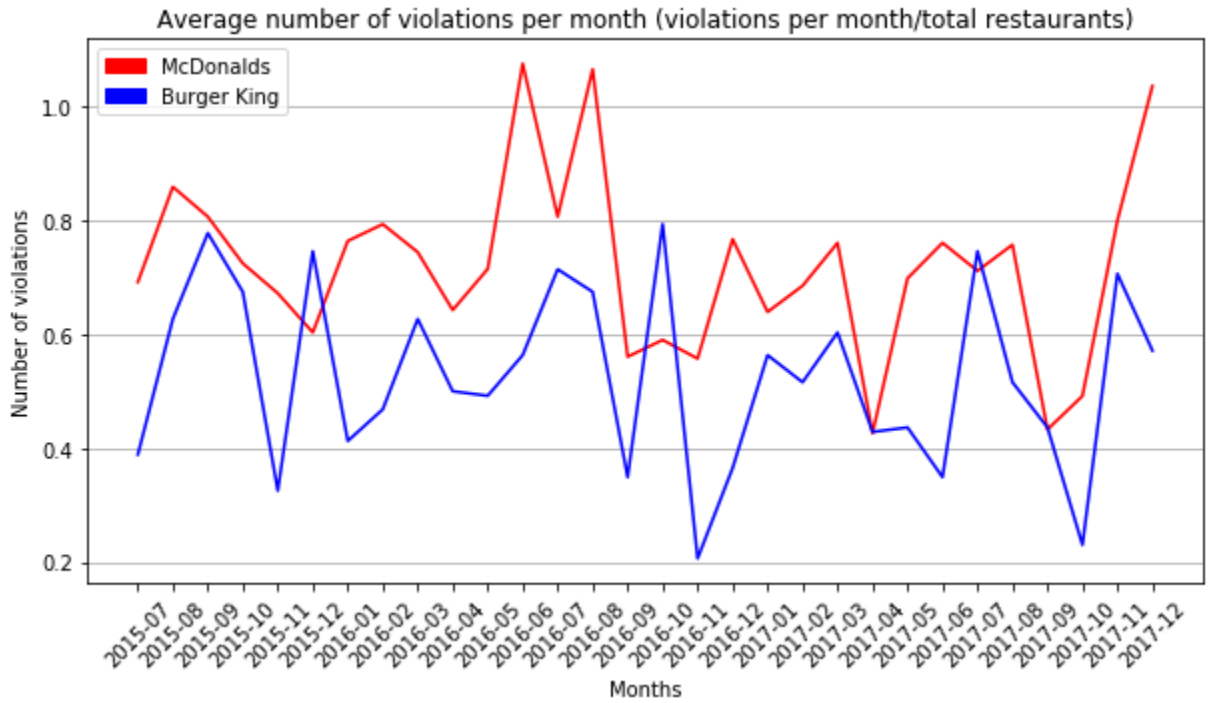


Figure 3 shows the average number of violations per month for all McDonalds compared with the average number of violations for all Burger Kings. The data for this graph was extracted by using python scripts from the provided data excel spread sheets (inspections.xlsx, vioaltions.xlsx).

The graph shows that on average Burger King performed better than McDonalds over the 2 years with all stores in the business receiving less violation codes. However, the results are only relative to the number of restaurants owned by each franchise and not a completely accurate representation of the violation codes received by each business. The McDonalds franchise owns 341 restaurants in California while Burger King operates only 38% of stores that McDonalds owns with 130 stores. When calculating the number of total violations per store over the two year period, the results show that Burger King receives twice the number of violations per store than does McDonalds (see Table 1). Therefore, the difference between both franchises and the number of violations received is not as significant as perceived from Figure 3 which shows that McDonalds has a larger number of violations with respect to Burger King. This is, however,

assuming that every restaurant owned by either franchise has received at least one violation. If the above assumption was false, then there could be some different results in Figure 3 but not any major outliers.

Table 1: Number of violations per month for McDonalds and Burger King stores

Average number of violations per month	
McDonalds	Burger King
0.691803	0.388889
0.859016	0.626984
0.806557	0.777778
0.72459	0.674603
0.672131	0.325397
0.603279	0.746032
0.763934	0.412698
0.793443	0.468254
0.744262	0.626984
0.642623	0.5
0.714754	0.492063
1.07541	0.563492
0.806557	0.714286
1.065574	0.674603
0.560656	0.349206
0.590164	0.793651
0.557377	0.206349
0.767213	0.365079
0.639344	0.563492
0.685246	0.515873
0.760656	0.603175
0.42623	0.428571
0.698361	0.436508
0.760656	0.349206
0.711475	0.746032
0.757377	0.515873
0.432787	0.436508
0.491803	0.230159
0.8	0.706349
1.036066	0.571429
Sum = 21.63934	Sum = 15.80952
Total number of stores = 341	Total number of stores = 130
Number of violations per store (Sum/ Total number of stores)	
=0.063458	=0.121612

Appendix

Section 1

```
create table violations(
point int,
serial_number text,
violation_code text,
violation_description text,
violation_status text,
PRIMARY KEY(point, serial_number, violation_code)
);
```

```
create table Previous_Violations(
facility_name text ,
facility_address text ,
facility_zip text ,
facility_city text
);
```

```
create table inspections(
activity_date date ,
employee_id text ,
facility_address text ,
facility_city text ,
facility_id text ,
facility_name text ,
facility_state text ,
facility_zip text ,
grade text,
owner_id text,
owner_name text,
pe_description text,
program_element_pe int,
program_name text,
program_status text,
record_id text ,
score int,
serial_number text ,
service_code int,
service_description text,
PRIMARY KEY(serial_number)
);
```

Section 2

Code	Description	Count
CL21	Public Health Permit/L	1
F001	# 01a. Demonstration (	6874
F002	# 02. Communicable di	38
F003	# 03. No discharge fror	36
F004	# 04. Proper eating, dri	3364
F005	# 05. Hands clean and p	5151
F006	# 06. Adequate handw	33952
F007	# 07. Proper hot and cc	41105
F008	# 08. Time as a public h	4706
F009	# 09. Proper cooling m	7993
F010	# 10. Proper cooking ti	51
F011	# 11. Proper reheating	854
F012	# 12. Returned and res	28
F013	# 13. Food in good con	5009
F014	# 14. Food contact surf	34443
F015	# 15. Food obtained fr	622
F016	# 16. Compliance with	1087
F017	# 17. Compliance with	44
F018	# 18. Compliance with	213
F019	# 19. Consumer adviso	255
F020	# 20. Licensed health c	1
F021	# 21b. Water available	6787
F022	# 22. Sewage and wast	1637
F023	# 23. No rodents, insec	29724
F024	# 24. Person in charge	983
F025	# 25. Personal cleanlin	11878
F026	# 26. Approved thawin	7798
F027	# 27. Food separated a	17866
F028	# 28. Fruits and vegeta	207
F029	# 29. Toxic substances	17986
F030	# 30. Food properly stc	39854
F031	# 31. Consumer self se	669
F032	# 32. Food properly lab	6734
F033	# 33. Nonfood-contact	100080
F034	# 34. Warewashing fac	19208
F035	# 35. Equipment/Uten:	80010
F036	# 36. Equipment, uten:	49742
F037	# 37. Adequate ventila	48042
F038	# 38. Thermometers pr	11060
F039	# 39. Wiping cloths: pr	35845
F040	# 40. Plumbing: Plumb	50868
F041	# 41. Garbage and refu	4742
F042	# 42. Toilet facilities: p	16153
F043	# 43. Premises; person	42943
F044	# 44. Floors, walls and	102009
F045	# 45. Sleeping quarters	437
F046	# 46. Signs posted; last	15735
F047	# 48. Plan Review requ	607
F048	# 47. Permits Available	5510
F049	# 50. Impoundment of	5926
F050	# 51. Permit Suspensic	2950
F051	# 49. Samples Collecte	40
F052	# 01b. Food safety cert	18359
F053	# 21a. Hot Water Avail	4218
F054	# 52. Multiple Major Cr	1214
F055	# 01a. Demonstration (	1515
F056	# 10. Proper cooking ti	12
F057	# 18. Compliance with	43
F058	# 19. Consumer adviso	185
MF07	# 07. Adequate handw	1
MF42	# 42. Garbage and refu	1
SF15	No Health Code Violat	1
SS33	Garbage / rubbish rece	1
W001	Proper hot and cold hc	9
W002	Food in good condition	1
W003	Food storage separate	6
W004	Food storage space	3
W005	Food elevated	11
W006	Food packaging protec	2
W008	Rodent	6
W009	Cockroaches	42
W011	Storage of materials 18	2
W012	Fly Breeding Material	3
W014	Fly Breeding	5
W016	Building rodent proof	3
W017	Hot and cold water ava	18
W018	Waste water or sewag	3
W019	Plumbing approved ar	15
W020	Wall(s) maintained cle	9
W021	Wall(s) maintained in	6
W022	Wall(s) constructed of	1
W023	Floor maintained clear	23
W024	Floor maintained in gc	5
W025	Ceiling maintained cle	2
W026	Ceiling maintained in	3
W027	Ceiling constructed of	5
W028	Toilet in good repair	3
W029	Toilet maintained clea	4
W030	Hand sink in good repa	1
W031	Hand sink maintained	4
W032	Toilet room floor / wal	3
W033	Toilet room floor / wal	2
W034	Toilet room with toilet	4
W035	Toilet room well venti	2
W036	Toilet room well lighte	2
W037	Toilet available	1
W038	Hand sink available	1
W039	Proper storage or use (	1
W040	Compliance with shell	2
W041	Premises maintained (	6
W042	Garbage / Rubbish rec	7
W043	Garbage / Rubbish rec	3
W044	Garbage / Rubbish rec	4
W045	No unapproved sleepi	1
W046	Live animals	1
W047	Thermometer: availab	7
W048	Permits Available	121
W049	Food from an approve	5
W050	Food properly labeled	14
W051	Walls, Floors, Ceilings:	32
W052	Equipment, Shelving, (	31
W053	Permit Suspension	35
WP13	# 13. Disease Transmis	1
WP15	# 15. Tobacco / Eating /	1
WP16	# 16. Hair Restraints / (	1
WP18	# 18. Personal Hygiene	1
	Total violations	**905891**

Section 3

Records structure = (date, number of violation codes)

Lowest postcode	Highest postcode	Average of all postcodes
('2015-07', 0)	('2015-07', 376)	('2015-07', 44.01364522417154)
('2015-08', 0)	('2015-08', 378)	('2015-08', 52.20499108734403)
('2015-09', 0)	('2015-09', 407)	('2015-09', 55.57192374350087)
('2015-10', 0)	('2015-10', 527)	('2015-10', 57.07116788321168)
('2015-11', 0)	('2015-11', 251)	('2015-11', 44.93613138686131)
('2015-12', 0)	('2015-12', 355)	('2015-12', 49.63884430176565)
('2016-01', 0)	('2016-01', 584)	('2016-01', 51.71107266435986)
('2016-02', 0)	('2016-02', 418)	('2016-02', 51.82101806239737)
('2016-03', 0)	('2016-03', 129)	('2016-03', 50.840873634945396)
('2016-04', 0)	('2016-04', 396)	('2016-04', 47.48264984227129)
('2016-05', 0)	('2016-05', 320)	('2016-05', 52.001508295625946)
('2016-06', 0)	('2016-06', 428)	('2016-06', 56.465014577259474)
('2016-07', 0)	('2016-07', 336)	('2016-07', 40.16894197952218)
('2016-08', 0)	('2016-08', 85)	('2016-08', 51.65932452276065)
('2016-09', 0)	('2016-09', 517)	('2016-09', 53.29761904761905)
('2016-10', 0)	('2016-10', 200)	('2016-10', 45.59577922077922)
('2016-11', 0)	('2016-11', 253)	('2016-11', 42.20552147239264)
('2016-12', 0)	('2016-12', 227)	('2016-12', 43.84307692307692)
('2017-01', 0)	('2017-01', 265)	('2017-01', 43.30104321907601)
('2017-02', 0)	('2017-02', 288)	('2017-02', 41.76967930029154)
('2017-03', 0)	('2017-03', 241)	('2017-03', 45.417015341701536)
('2017-04', 1)	('2017-04', 360)	('2017-04', 42.13384615384616)
('2017-05', 0)	('2017-05', 483)	('2017-05', 44.26571428571429)
('2017-06', 0)	('2017-06', 684)	('2017-06', 44.77065527065527)
('2017-07', 0)	('2017-07', 287)	('2017-07', 40.477037037037036)
('2017-08', 0)	('2017-08', 318)	('2017-08', 46.551169590643276)
('2017-09', 0)	('2017-09', 301)	('2017-09', 44.797734627831716)
('2017-10', 0)	('2017-10', 534)	('2017-10', 44.28082191780822)
('2017-11', 0)	('2017-11', 304)	('2017-11', 43.02388059701492)
('2017-12', 0)	('2017-12', 312)	('2017-12', 40.783987915407856)