

# Notebook

February 17, 2020



Use the `head` command on your four files again. This time, describe at least one potential problem with the data you see. Consider issues with missing values and bad data.

Please write your answer in the markdown cell below. You may create new cells below your answer to run code, but **please never add cells between a question cell and the answer cell below it.**

One potential issue with the data is bad data regarding the latitude and longitude. `-9999.000000` is definitely not the correct value, and trying to map out these locations will not produce the correct results. The first phone number is also not correct.



---

## 1 6: Compute Something Interesting

Play with the data and try to compute something interesting about the data. Please try to use at least one of groupby, pivot, or merge (or all of the above).

Please show your work in the cell below and describe in words what you found in the same cell. This question will be graded leniently but good solutions may be used to create future homework problems.

**Please have both your code and your explanation in the same one cell below. Any work in any other cell will not be graded.**

In [225]: *#YOUR CODE HERE*

```
helper_merge = ins.merge(ins2vio[['iid', 'vid']])
merge_insvio = helper_merge.merge(vio[['vid', 'description']])
merge_insvio = merge_insvio[merge_insvio['score'] > 0]
merge_insvio.sort_values('score', ascending = False).head(20)
```

*#YOUR EXPLANATION HERE (in a comment)*

*# Something interesting I noticed from this data is that restaurants that earn a perfect score have low risk descriptions that they are written up for. However, there are no restaurants that have a perfect score in all. All of the perfect scores are also given in recent years (2018 - 2019), and all of the perfect scores are in recent years. This is most definitely because of the massive increase in recent years.*

Out[225]:

	iid	date	score	type \
9350	95224_20190719	07/19/2019 12:00:00 AM	100	Routine - Unscheduled
8005	5562_20180605	06/05/2018 12:00:00 AM	100	Routine - Unscheduled
... Omitting 58 lines ...				
9816	Improper storage of equipment utensils or linens			
9911	Improper storage of equipment utensils or linens			
35452	Inadequate warewashing facilities or equipment			