

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.**

For the bus csv file, some of the potential problems that can be seen is the there isn't a proper latitude, and longitude for some of the restraunts as well as a phone number to reach them. This could be because that there is no current information given for them just as of yet. To deal with this missing information, -9999 was used to fill in this missing information. In the ins.csv file we can see some places have have a score of -1, which seems to be bad data as this information is missing. In ins2vio.csv file, in the iid column, there are diffenent lengths for the numbers, which signifies there could be missing or bad data that was given/ written down incorrectly.

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 [203]: All_data

```
Out[203]:
```

	iid	vid	date	score	\
0	97975_20190725	103124	07/25/2019 12:00:00 AM	96	
1	85986_20190815	103124	08/15/2019 12:00:00 AM	83	
... 0mitting 59 lines ...					
40209	Improperly displayed mobile food permit or sig...			Moderate Risk	

[40210 rows x 20 columns]