Notebook

February 19, 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.**

When I see the vio table, it seems there isn't enough space to show the data. If there are datas with similar and long description, we may not distinguish those unless we know other key like id.

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 [78]:
```
#YOUR CODE HERE
ins_zip = pd.merge(ins, bus['postal_code'], left_on = ins['bid'], right_on = bus['bid']);
s_n = ins_zip.loc[:, ['postal_code', 'type', 'score']];
s_n = s_n[s_n['score'] != -1];
s1 = pd.value_counts(s_n['postal_code']);
s1 = s1.to_frame(name = 'val');
lowest_postal = s_n.groupby(['postal_code'], as_index = False).median().sort_values('score', as
s1['postal'] = s1.index;
low_m = lowest_postal.merge(s1['val'], left_on = lowest_postal['postal_code'], right_on = s1['
low_m
#YOUR EXPLANATION HERE (in a comment)
#I wanted to get score data according to postal code, and to get good data, I also counted the
#I expected there would be a big gab as locations, but there isn't that big gab.
#The interesting thing is the food trucks (with postal code -9999) have a good score.
```

Out[78]:

	key_0	postal_code	median score	val
0	CA	CA	78	1
1	92672	92672	84	2
… Omitting 48 lines …				
51	94301	94301	100	7
52	95132	95132	100	1
53	94013	94013	100	6