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.**

- Wrong data in bus.latitude, bus.longitude, bus.phonenumber shows as -9999
- Wrong data in ins.score shows as -1

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]: name_violations_pivot = pd.merge(ins_named, ins2vio, left_on = 'iid', right_on = 'iid')
         name_violations_pivot = pd.merge(name_violations_pivot, vio, left_on = 'vid'\
                                                                     , right_on = 'vid')

         name_violations_pivot['count'] = 1
         name_violations_pivot= pd.pivot_table(name_violations_pivot, \
                                         values='count', \
                                         index='name', \
                                         columns=['description'], \
                                         aggfunc=np.sum, \
                                         fill_value=0)
         name_violations_pivot['Total'] = name_violations_pivot.sum(axis=1)
         name_violations_pivot = name_violations_pivot.sort_values(["Total"], ascending=False)

         name_violations_pivot

         # Business vialation piviot table sorted by total violation times, from most to least.
```

```
Out[78]: description                 Consumer advisory not provided for raw or undercooked foods  \
         name
         Peet's Coffee & Tea                                                     0
         … Omitting 246 lines …
         The Fillmore                                            0       1

         [5047 rows x 66 columns]
```