Notebook

February 16, 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 problem of these data sets is negative or not realistic numbers set of data. For example, in bus, they have -9999 latitdue adn longtidue, ins have negative scores.

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 [215]: right=pd.merge(ins[['score','bid']],bus[['bid','name','phone_number','latitude','longitude']]
          right=right[right['score']>0].groupby(['bid','name'], as_index=False).sum().sort_values(['sco
          right.head(90)
          num_phone=pd.Series(right['phone_number'].value_counts(),index=right['phone_number'].unique()
          right['phone']=~right['phone_number'].apply(str).str.len()==11
          count_t=(right['phone']==True).count()
          count_f=(right['phone']==False).count()
          right.head(60)


          #Form the data I collected, we can estblishe some intresting facts:
          #    1.The top scored buildings are school buidlings(first at least 90th rank by score is all
          #    2.High ranking building have all the latitude and longitude for most of the time
          #    3.Most of school building who doens't have correct phone number start with '-' and have
          #        (ex: -19998,-29997,-39996,-49995,-59994, etc)
          #    4.Out of 5724 building, only 5724 have the correct phone number
          #        (which mena half of the building have phoen while other half doesn't)
          #    5.Some of the buildign have the same latitude, phone number, and logntitude, which is fa
          #
          #

Out[215]:       bid                                 name   score  phone_number  \
          1151  5855              SOTA (SCHOOL OF THE ARTS)   778   113245565600
          1205  5953                 ST PETER'S PARISH SCHOOL   696         -69993
          … Omitting 116 lines …
          1807     226.348098   -734.505900   False
          1211     226.760184   -734.639004   False
          1206     226.259376   -734.688492   False
```