

PS4 Andy Fan Will Sigal

PS4: Due Sat Nov 2 at 5:00PM Central. Worth 100 points.

Style Points (10 pts)

Submission Steps (10 pts)

Download and explore the Provider of Services (POS) file (10 pts)

(This file records all providers certified to bill Medicare and Medicaid. Each provider has a unique CMS certification number, consistent across different years)

(The data dictionary is located at the bottom of the page as “Provider of Services File- Hospital & Non-Hospital Facilities Data Dictionary)

```
### SETUP
import pandas as pd
import altair as alt
import time
import os
import warnings
warnings.filterwarnings('ignore')
```

1. (Partner 1) This is a fairly large dataset and we won't be using most of the variables. Read through the rest of the problem set and look through the data dictionary to identify which variables you will need to complete the exercise, and use the tool on data.cms.gov into restrict to those variables (“Manage Columns”) before exporting (“Export”). Download this for 2016 and call it pos2016.csv. What are the variables you pulled?

required variables:

provider type code: PRVDR_CTGRY_CD

subtype code: PRVDR_CTGRY_SBTYP_CD

CMS certification number: PRVDR_NUM

Termination code: PGM_TRMNTN_CD

zip code: ZIP_CD

```
### import data
path1 = 'd:\\UChicago\\Classes\\2024Qfall\\Programming
↪ Python\\problem-set-4-andy-fan-will-sigal\\pos2016.csv'
df = pd.read_csv(path1, encoding="utf-8", on_bad_lines="skip")
```

2. (Partner 1) Import your pos2016.csv file. We want to focus on short-term hospitals. These are identified as facilities with provider type code 01 and subtype code 01. Subset your data to these facilities. How many hospitals are reported in this data? Does this number make sense? Cross-reference with other sources and cite the number you compared it to. If it differs, why do you think it could differ?

a.how many hospitals, does it make sense

b.cross ref. other sources, if it differs, why?

3. (Partner 1) Repeat the previous 3 steps with 2017Q4, 2018Q4, and 2019Q4 and then append them together. Plot the number of observations in your dataset by year.

4. (Partner 1) Each hospital is identified by its CMS certification number. Plot the number of unique hospitals in your dataset per year. Compare this to your plot in the previous step. What does this tell you about the structure of the data?

a.plot unique hospitals

b.compare to previous step, what does this tell about data structure

Identify hospital closures in POS file (15 pts) (*)

- 1.
- 2.
3.
 - a.
 - b.
 - c.

Download Census zip code shapefile (10 pt)

1. a.
 b.
- 2.

Calculate zip code's distance to the nearest hospital (20 pts) (*)

- 1.
- 2.
- 3.
4. a.
 b.
 c.
5. a.
 b.
 c.

Effects of closures on access in Texas (15 pts)

- 1.
- 2.
- 3.
- 4.

Reflecting on the exercise (10 pts)