Predicting Hospital Utilization Rates as a Function of Socio-Demographic Factors

Background.

In this case you will work with data available through the Health Resources and Services Administration (HRSA). The broad topic we are interested in is access to healthcare and quality of care. We are also interested in disparities related to race and socioeconomic status. You can learn more about these issues through the links below:
https://archive.ahrq.gov/research/ruraldisp/ruraldispar.pdf
https://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/nhqrdr/nhqdr15/2015nhqdr.pdf

The Task.

For this competition your task is to create a model that predicts hospital utilization. Note that while there are some variables that directly measure aspects of utilization, in this case we are looking for a broader definition that incorporates several utilization aspects, given our objective to learn about access and quality of care. Note that simply using a single variable as your target will not suffice.

Specific tasks components include:

- Define and measure your target - there are several relevant variables in the file. Variables of interest include those touching directly on utilization but also variables such as readmission rates (think in line of "effective utilization") or access to care.
- Predictor variables - identify a relevant subset of fields that can be used to predict your target variable. Note that the file has approximately 3000 rows and 6000 records. You would have to reduce the number of fields (either subset or combine) in order to get a good model. Remember that we are particularly interested in disparities, as described in the introduction above. Explain how you selected the fields to use.
- Model building - choose and apply modeling technique(s) appropriate to the above task. Make sure to fully interpret the results, justify your model selection, and provide all relevant model quality and accuracy measures.

The Data.

For this competition you will work with the Area Health Resource File (AHRF), which is a data set comprised of data collected from more than 50 sources and contains more than 6,000 variables related to health care access at the county level. You can obtain the data directly from this website: http://ahrf.hrsa.gov, or you can use the R package available via github.

From R type:

```
> library(devtools)
> install_github("olyerickson/ahrf2016r")
> library(ahrf2016r)
> data(list="ahrf_county_train")
> str(get("ahrf_county_train"))
> data(list="ahrf_county_layout")
> str(get("ahrf_county_layout"))
> View(ahrf_county_train)
> View(ahrf_county_layout)
```

Via git:

```
$ git clone https://github.com/olyerickson/ahrf2016r.git
$ cd arhf2016/data
```

*Look for the *.rda files (direct load into R)*

```
$ cd arhf2016/metadata
```

Look for the data dictionary!

Deliverables and Judging.

Judging will take place on Sunday from 1-4pm. Announcement of winners will follow. There will be two judging components: (1) Model assessment, and (2) Presentation. For the presentation session you need to prepare a 10 minutes presentation of your solution. Make sure to include all required task components specified above. When preparing your presentation remember your audiences and make sure your slides present clear information with good flow. You will present off of your laptop so please make sure everything is set up and ready to go on your scheduled judging slot (see table below).

Judging will take into account the quality and accuracy of your model; the approach you took to defining the target variable and selecting the predictor variables; your model selection approach; and your presentation style and ability to answer questions.

Good Luck

Below is the judging schedule, the numbers represent **team numbers**. Note that each team has **two** separate judging slots. Please make sure you arrive to both **on time.**

	Room 1 (upstairs)	Room 2 (upstairs)	Room 3 (upstairs)	Room 4 (upstairs)
1:00pm	1	2	3	4
1:25pm	5	6	7	8
1:50pm	9	10	11	12
2:15pm	13	14	15	16
2:40pm	17	18	19	20
3:05pm	21	22	23	24
3:30pm	25	26	27	28
	Room 1 (downstairs)		Room 2 (downstairs)	
1:00pm	9, 10	2:40pm	1, 2	
1:10pm	11, 12	2:50pm	3, 4	
1:20pm	13, 14	3:00pm	5, 6	
1:30pm	15, 16	3:10pm	7, 8	
1:40pm	17, 18	3:20pm		
1:50pm	19, 20	3:30pm		
2:00pm	21, 22	3:40pm		
2:10pm	23, 24	3:50pm		
2:20pm	25, 26			
2:30pm	27, 28			
Cont'd on next column				