

Capstone 1 project proposal

*info in red is from grading rubric.

- What is the problem you want to solve?
 - Can you predict who how much medicare will reimburse (by percentage of total bill), based on the diagnosis while in the hospital?
- Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?
 - Client is the hospital/hospital system and/or insurance companies
 - Useful in a budgeting context because of increased resources required by hospital to treat diagnoses with less medicare coverage, allow for opening discussions with medicare (or lobbying at a government level) for improved reimbursement for those services with less coverage.
 - May also be useful for hospitals not currently treating these diagnoses who are looking into adding treatments/programs to their facilities
- 3. What data are you using? How will you acquire the data?
 - <https://data.cms.gov/inpatient-provider-lookup/view-data> (links to all data)
 - <https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-I-PPS-Provider/tcsp-6e99> (2017)
 - <https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-I-PPS-Provider/fm2n-hjj6> (2016)
 - <https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-I-PPS-Provider/w2du-it53> (2015)
 - <https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-I-PPS-Provider/9zmi-76w9> (2014)
 - Data sets are from the center for medicare/medicaid (CMS) for 2014-2017. They are downloadable as an excel or csv file.
- 4. Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution. This might include:
 - a. Is this a supervised or unsupervised problem? Supervised, everything has a label already
 - b. If supervised is it a classification or regression problem? Regression, looking for a value output, rather than a label output.
 - c. What variable is it you are trying to predict? The percentage of payment (out of total billed cost) that medicare provides for a given diagnosis.
 - d. What variables will you use as predictors? Past data from 2014-2017 on average bill for a given diagnosis and average medicare reimbursement
 - e. What will be your training data? I have the center for medicare data of all reimbursement for 2014-2017, broken down by hospital
 - In short:
Each year average = $\text{Sum}((\text{Average medicare reimbursement for a given diagnosis})/(\text{average bill for a given diagnosis})) / (\text{number of hospitals that discharge } > 10 \text{ people with that diagnosis})$

Plot difference in each year from 2014-2017 to see if there is a trend. If yes, attempt to predict future reimbursements.

- 5. What are your deliverables? Typically, this includes code, a paper, or a slide deck
 - At the end I expect to have code and a slide deck