1st capstone idea: Regression model for predicting hospital readmission within 30 days after discharge

1. Information about the data set:
2. It consists of 10,000 records of faked patients with 14 fields (columns).
3. The data set was imported in Tableau for predictive analyses. Add hoc A-B test was used to examine and visualize the effect of given variables on the outcome. The outcome is a binary outcome and it is presented by the last field (Readmitted).
4. The 1st A-B test was for the effect of gender on readmission. <https://public.tableau.com/profile/omar.al.naimi#!/vizhome/HospitalReadmissionProject1/Gender>
5. The 2nd A-B test was the effect of hospital location on readmission. <https://public.tableau.com/profile/omar.al.naimi#!/vizhome/HospitalReadmissionProject2/HospitalLocation>
6. The 3rd A-B test was the effect of having health insurance on readmission. <https://public.tableau.com/profile/omar.al.naimi#!/vizhome/HospitalReadmissionProject3/HealthInsurance>
7. The 4th A-B test was the effect of at least one checkup after discharge on readmission. <https://public.tableau.com/profile/omar.al.naimi#!/vizhome/HospitalReadmissionProject4/RegularCheckup>
8. The 5th A-B test was the effect of the number of previous admissions on readmission. <https://public.tableau.com/profile/omar.al.naimi#!/vizhome/HospitalReadmissionProject5/PreviousAdmission>
9. The 6th A-B test was the effect of the period of previous admission (in days) on readmission. <https://public.tableau.com/profile/omar.al.naimi#!/vizhome/HospitalReadmissionProject6/AdmissionPeriodday>
10. The 7th A-B test was the effect of socio-economic status on readmission. <https://public.tableau.com/profile/omar.al.naimi#!/vizhome/HospitalReadmissionProject7/SES>
11. The 8th A-B test was the effect of age on readmission. <https://public.tableau.com/profile/omar.al.naimi#!/vizhome/HospitalReadmissionProject8/Age>
12. The last A-B test used the last digit of the patient insurance number to evaluate the data set as a sample. <https://public.tableau.com/profile/omar.al.naimi#!/vizhome/HospitalReadmissionProject9/EvaluatingDataSet>
13. Statistics: the visualizations those showed difference between the compared parameters have been checked by chi-squared test to acquire statistical significance (if there is a significance). The website <http://www.evanmiller.org/ab-testing/chi-squared.html> and <http://vassarstats.net/> were used to conduct chi-squared test. The results were the following:
14. The 1st A-B test showed that females have more readmission than males (p < 0.001).
15. The 2nd A-B test showed that there are difference between hospital locations (<.0001).
16. The 3rd A-B test showed that health insurance has no effect on readmission (p = 0.48).
17. The 4th A-B test showed that regular checkup has reduced the readmission (p < 0.001).
18. The 5th A-B test showed that 1 previous admission has effect on readmission more than 2 previous admissions (p < 0.001).
19. The 6th A-B test was the effect of the period of previous admission (in days) on readmission.
20. The 7th A-B test showed that socio-economic status has von effect on readmission (0.6714).
21. The 8th A-B test showed that age group has effect on readmission (<.0001).

The project will continue to regression model development in next time

2nd capstone idea: regression model for predicting medical claim frauds.

3rd capstone idea: regression model for predicting people at risk of being obese.