Project 2 plan

Time period	Doing	Expectation
11-17 ~ 11-24	Complete all categorical variable classification	Data cleaning Data dimension after cleaning
11~24-11~30	1.Predictors statistical understanding 2.Perform train/validation splits (random splits, LOOCV, k-fold) 3.Logistic regression and performance check	Complete statistical analysis Complete machine learning
12.1	Group presentation rehearsal – total 15 mins Written report (70% competition rate)	
12.1~12.4	Zoom meeting for final oral presentation	Oral presentation done
12.5	Submit assignment oral and written reports	Submission and confirmation

Suggestions:

- update data and conclusions if any
- I will update timely progress to Github and send notice to slack

Statistic analysis checking points (11-17~ 11-29)

- Data processing – mutate, cleaning and NA handling
- Variables classification
- Multicollinearity
- Model- predictors are all statistical significant?
- Residual plots review
- Train/Test splits, what is MSE for each case?
    - 50%50% train/test
    - LOOCV
    - K-fold validations
- Logistic regression?
    - ROC
    - AUC
    - What is cut-off value? What is true positive rate … ?
- Module 12 learning
- Data visualization