Midterm Project Proposal Group 16

- Group members: Jieqi Tu (jt3098), Jiayi Shen (js5354)
- The tentative project title: House sale price prediction in King County
- The anticipated data source:
    - Data link: https://www.kaggle.com/harlfoxem/housesalesprediction
    - Description: This dataset contains house sale prices for King County, which includes Seattle. It includes homes sold between May 2014 and May 2015. There are 16 feature variables, excluding ID and response. Predictors include number of bedrooms or bathrooms, number of floors, and the year the house was built, etc..
- Response and predictors:
    - Response: House sale price
    - Predictors:
        - bedrooms (Number of Bedrooms/House)
        - bathrooms (Number of bathrooms/House)
        - sqft_living (square footage of the home)
        - sqft_lot (square footage of the lot)
        - floors (Total floors (levels) in house)
        - waterfront (House which has a view to a waterfront)
        - view (Has been viewed)
        - condition (How good the condition is (Overall))
        - grade (overall grade given to the housing unit, based on King County grading system)
        - sqft_above (square footage of house apart from basement)
        - sqft_basementsquare (footage of the basement)
        - yr_built (Built Year)

        - yr_renovated (Year when house was renovated)
        - sqft_living15 (Living room area in 2015(implies-- some renovations) This might or might not have affected the lotsize area)
        - sqft_lot15lot (Size area in 2015(implies-- some renovations))

- The planned analysis:
    - Data visualization: to check the correlation between each predictor and response and the distribution of predictors.
    - Fit linear, ridge, lasso, PCR, PLS, polynomial, smoothing spline, GAM and MARS models. Use cross-validation to obtain MSE from different models. Compare these models with the obtained test MSEs.
    - Make some graphs and plots (e.g. boxplot, bar plot, etc.) to show the distribution of MSEs, and the fit of models, etc.