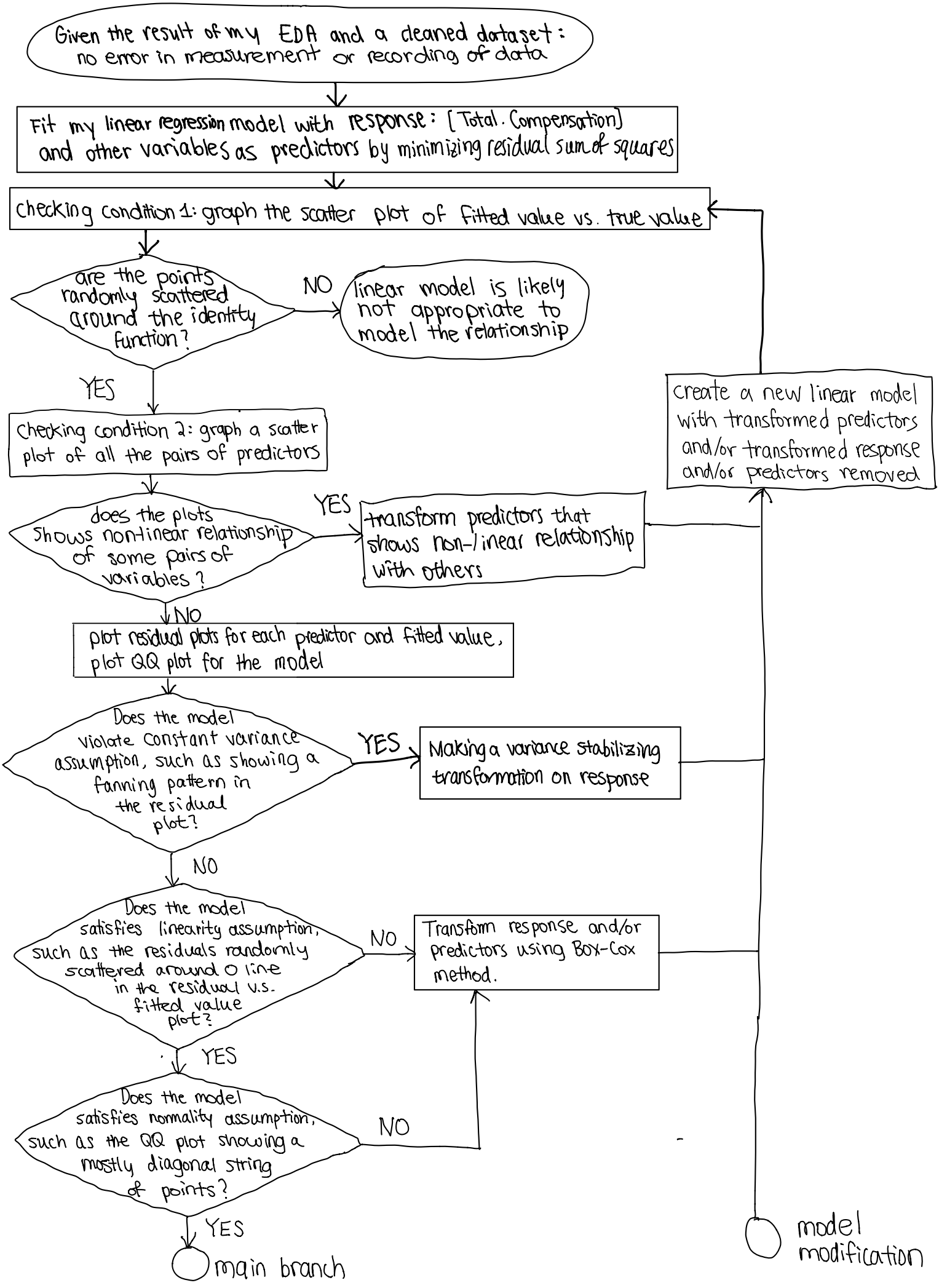
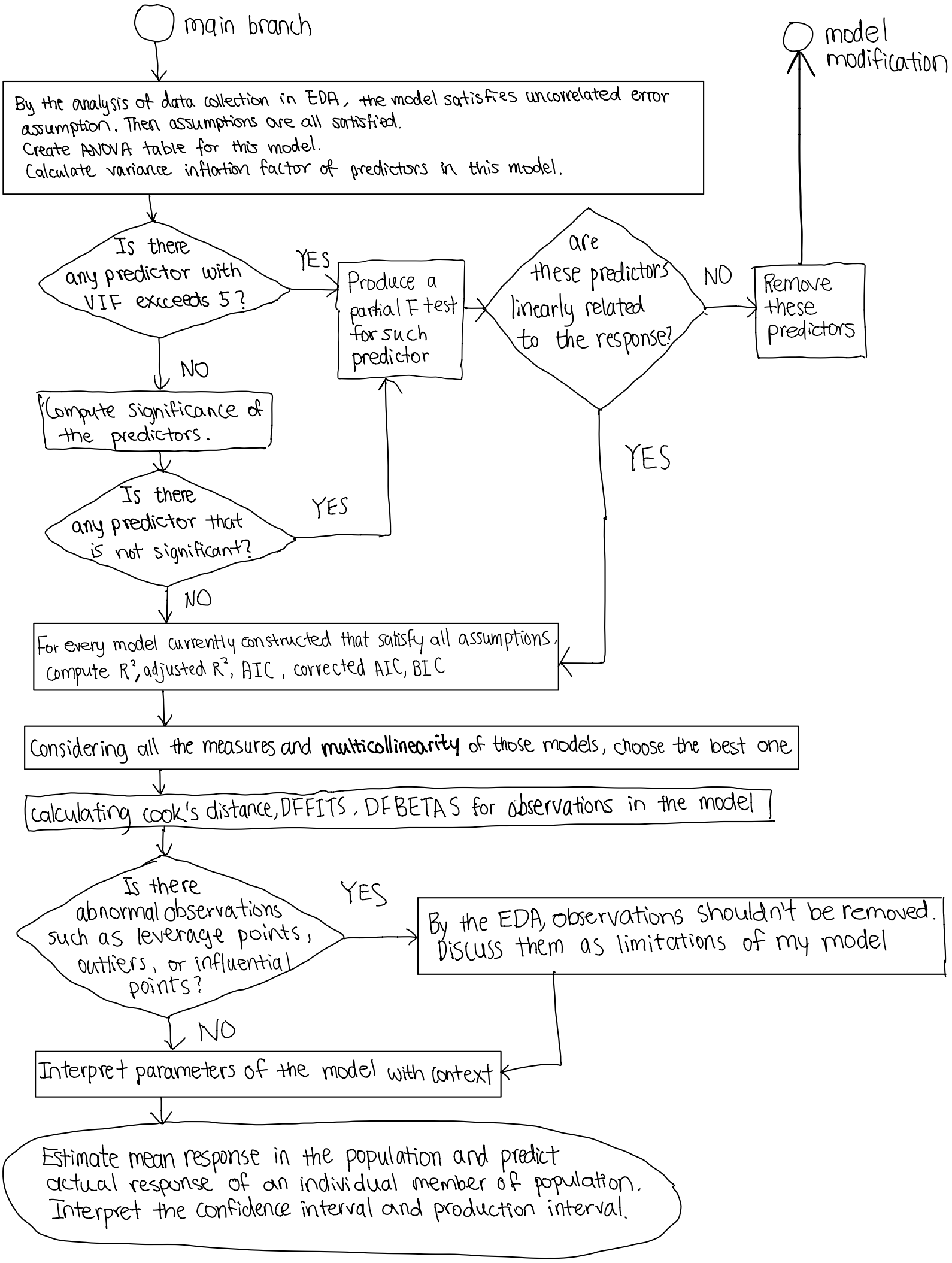

Given the result of my EDA and a cleaned dataset: no error in measurement or recording of data
Fit my linear regression model with response: [Total. Compensation] and other variables as predictors by minimizing residual sum of squares
checking condition 1: graph the scatter plot of fitted value vs. true value
are the points randomly scattered around the identity function?
NO
linear model is likely not appropriate to model the relationship
YES
Checking condition 2: graph a scatter plot of all the pairs of predictors
does the plots shows nonlinear relationship of some pairs of variables?
YES
transform predictors that shows non-linear relationship with others
NO
plot residual plots for each predictor and fitted value, plot QQ plot for the model
Does the model violate constant variance assumption, such as showing a fanning pattern in the residual plot?
YES
Making a variance stabilizing transformation on response
NO
Does the model satisfies linearity assumption, such as the residuals randomly scattered around 0 line in the residual v.s. fitted value plot?
NO
Transform response and/or predictors using Box-Cox method.
YES
Does the model satisfies normality assumption, such as the QQ plot showing a mostly diagonal string of points?
NO
YES
main branch
create a new linear model with transformed predictors and/or transformed response and/or predictors removed
model modification


main branch
model modification
By the analysis of data collection in EDA, the model satisfies uncorrelated error assumption. Then assumptions are all satisfied.
Create ANOVA table for this model.
Calculate variance inflation factor of predictors in this model.
Is there any predictor with VIF exceeds 5?
YES
Produce a partial F test for such predictor
are these predictors linearly related to the response?
NO
Remove these predictors
NO
Compute significance of the predictors.
Is there any predictor that is not significant?
YES
NO
YES
For every model currently constructed that satisfy all assumptions, compute R^2, adjusted R^2, AIC, corrected AIC, BIC
Considering all the measures and multicollinearity of those models, choose the best one
calculating cook's distance, DFFITS, DFBETAS for observations in the model
Is there abnormal observations such as leverage points, outliers, or influential points?
YES
By the EDA, observations shouldn't be removed. Discuss them as limitations of my model
NO
Interpret parameters of the model with context
Estimate mean response in the population and predict actual response of an individual member of population. Interpret the confidence interval and production interval.

M5:

Partial F Test

ANOVA F Test

M4: Inference Coeff. mean response, predictions

M3: assumptions

M2: fit model with

Statistical relationship

is there non