Input vector is defined by p features $X_1, X_2, ..., X_{13}$, output is a target feature Y.

1. For any feature X_i, estimate covariance and correlation between X_i and target feature Y (on a training set).
2. For any feature X_i, calculate absolute value of correlation $|\frac{\hat{Cov}(X_i,Y)}{\hat{\sigma}_{X_i}\hat{\sigma}_Y}|$.
3. **Question 1**: Based only on the latter correlation, can you list all variables that is relevant for prediction?
4. Calculate the correlation matrix $C = [correl(X_i, X_j)]_{1\leq i,j\leq 13}$, where $correl(X_i, X_j)$ is an estimator of the correlation between variables X_i and X_j.
5. **Question 2**: What conclusion can you make about the structure of your predictors after analysis of correlation matrix?
6. Write a Scikit-learn/Python code for least square estimation of weights β_i in the model (of course, using only training set):

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$$

7. Let us denote b_i a least square estimate of β_i. Obtain values of $b_i, i = \overline{0, p}$
8. Estimate variance $\hat{\sigma^2}$ of noise ϵ (on a training set) as MSE.
9. Let us denote t_i the t-value for a variable X_i, i.e. $t_i = \frac{b_i}{\hat{\sigma}\sqrt{h_{ii}}}$ where $[h_{ij}] = (X^T X)^{-1}$. Calculate t-value (on a training set) of every non-target variable.
10. Based on the value of t_i find all your variables that is relevant for prediction.
11. **Question 3**: List variables that can be discarded.
12. Calculate residuals and draw Q-Q plot.
13. **Question 4**: Is an error normally distributed, yes or no (based on Q-Q plot)?
14. Draw a plot "residuals against $\hat{Y}$".
15. **Question 5**: If your error is not normal, what would you assume about the real distribution of an error (based on the latter plot)?
16. Calculate R^2.
17. **Question 6**: Give the final verdict: did linear regression model solved the prediction problem or not?

18. Prepare a report (word or pdf file) where all those steps and their results are described in a clear way. I.e. with presentation of: all obtained values (in the form of tables/charts) and answers to questions. **Answers to questions 1-5 should be given with explanation, should be highlighted, your final grade will depend on it mainly.**

19. The length of report should not be longer than 4 pages.