1. Explain the linear regression algorithm in detail.
   - It captures linear relationship between independent variables and dependent variables.
   - It's used to build supervised continuous output variable model .
   - It requires labelled data to build model.
   - It's used for predicting impact of independent vars on target var as well as forecast target variable.
   - Types : Simple(1 independent) , multiple linear(more than 1 independent) regression.
   - It tries identify equation for best fit line using OLS(Gradient descent)

2. What are the assumptions of linear regression regarding residuals?
   - Residual errors are normally distributed around 0
   - Residual errors have homoscedesticity. Uniform residual error variance across range of independent vars

3. What is the coefficient of correlation and the coefficient of determination?
   - Coefficient of correlation is degree of relationship between 2 variables. (-1 to 1)
   - Coefficient of determination is R2 which explains overall how good model explains variations in target variable. Value is between 0 and 1

4. Explain the Anscombe's quartet in detail.
   - Anscombe's quartet comprises four data sets that have nearly identical simple descriptive statistics, yet have very different distributions and appear very different when graphed.
   - It gives significance of visualizing data to see distribution. Only seeing descriptive statistics may be misleading sometimes.

5. What is Pearson's R?
   - Pearson's correlation coefficient (r) is a measure of the strength of the association between the two variables. Value can be between -1(perfect negative), 0(No correlation), 1(perfect positive)

6. What is scaling? Why is scaling performed? What is the difference between normalised scaling and standardised scaling?

- Scaling means bring range of column values in specific range.
- Its important when interpretation of predictor variables needed as scaling helps being all variables in same range (0-1) and comparative analysis can be done on significance of each predictor variables.
- It also help gradient descent algorithm to run optimally while doing OLS
- Normalised scaling brings column values between 0 and 1 using MinMaxscalar
- Standardised scaling brings column mean at 0 and standard deviation at 1(Unit variance)

7. You might have observed that sometimes the value of VIF is infinite. Why does this happen?

- Perfect co-relation between variables such that 1 can be completely explain by others.

8. What is the Gauss-Markov theorem?

- The Gauss Markov theorem tells us that if a certain set of assumptions are met, the ordinary least squares estimate for regression coefficients gives you the best linear unbiased estimate possible.
- Its basis of Linear regression algorithm.
- There are five Gauss Markov assumptions:

    A. Linearity: the parameters we are estimating using the OLS method must be themselves linear.

    B. Random: our data must have been randomly sampled from the population.

    C. Non-Collinearity: the regressors being calculated aren't perfectly correlated with each other.

    D. Exogeneity: the regressors aren't correlated with the error term.

    E. Homoscedasticity: no matter what the values of our regressors might be, the error of the variance is constant.

9. Explain the gradient descent algorithm in detail.

- Gradient descent is brute force way of identifying least square error straight line for given dataset
- OLS internally used gradient descent algorithm
- It basically defines cost function and then takes on various values of B1,B2,.. (co-efficients) and intercept(B0) such that cost function is minimised to least value.

10. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.

- When quantiles of two datasets plotted against each other then we get qq plot
- It's used to compare distributions of 2 data sets.
- All point of quantiles lie on or close to straight line at an angle of 45 degree from x – axis. It indicates that two samples have similar distributions.
- The y – quantiles are lower than the x – quantiles. It indicates y values have a tendency to be lower than x values.
- The x – quantiles are lower than the y – quantiles. It indicates x values have a tendency to be lower than the y values.
- This helps in a scenario of linear regression when we have training and test data set received separately and then we can confirm using Q-Q plot that both the data sets are from populations with same distributions or not.