Q1: Explain the linear regression algorithm in detail.

Linear algorithm tries to find a linear relationship between variables. In a single linear regression we try to find association between 2 variables. One is predictor and the other predicted value. In a multiple linear regression we try to find relationship (coefficients) between multiple predictors and the predicted variable (like car price). Following are the steps for performing linear regression:

- Understand the business problem and objective. What value is to be predicted and what are the predictor variables. Developing good business understanding will help.
- Differentiate between categorical and numeric variables.
- **Visualize**: Find correlation between numeric variables. Identify the ones that have very high correlations.
- Visualize categorical variables using boxplots and observe how value of predicted variable changes with each categorical value
- Convert categorical values to dummy variables or 0 and 1 (in case of only 2 values). Drop the first one so that variable with n values will have n-1 dummy variables.

Linear Regression Phase

- Split the data into training and test datasets
- Rescale the features – use minmax technique (standardize is another technique). MinMax only for non-dummy and the ones not having 0 or 1 values only. For these not required.
- Divide train set into X and Y sets (predictor and predicting variables)
- Add a constant to X_train
- Build the model
    - Build incrementally starting with 1 variable and then adding more
    - Build reverse by starting with all and reduce
    - Build Automated way (using RFE) start with 10-15 variables
    - Build mixed way. Start with RFE then using VIF / p-value to adjust until you get perfect model
- Inspect the summary (R square, p value and VIFs or Use RFE method)
- Adjust the variables in the model until you get satisfactory results
- Residual Analysis (observe errors are normal distribution around mean 0)
- Make Predictions

Q2: What are the assumptions of linear regression regarding residuals?

- There is linear relationship between X and Y (predictors and predicted)
- Error terms are normally distributed
- Error terms are indepdent of each other
- Error terms have constant variance (homoscedasticity)

Q3: What is the coefficient of correlation and the coefficient of determination?

- Coefficient of correlation is Beta1. It indicates a unit change in X will bring Beta1 change in Y
- Coefficient of determination is R-square. A value of R square closer to 1 indicates how well the variance in predicted variable is explained by the model

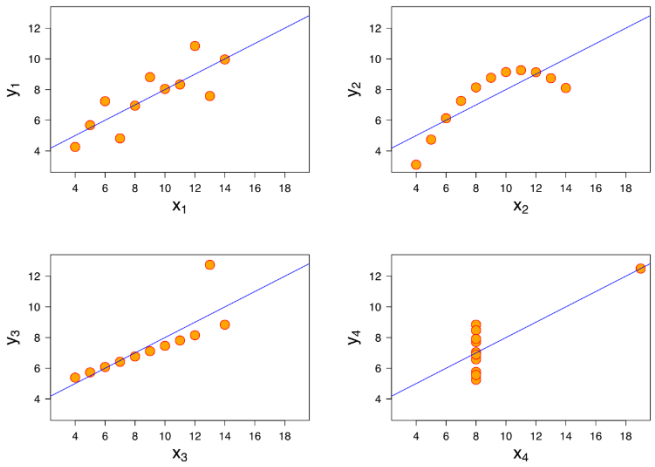
Q3: Explain the Anscombe's quartet in detail.

This quartet emphasizes the importance of visualization in Data Analysis. Looking at the data reveals a lot of the structure and a clear picture of the dataset. There are 4 datasets for which summary statistics like mean, sum, standard deviation and correlation coefficients are same. However, when these datasets and visualized they reveal completely different patterns. Anscombe's quartet explain's importance of visualization in data science.

	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

The summary statistics show that the means and the variances were identical for x and y across the groups :

1. Mean of x is 9 and mean of y is 7.50 for each dataset.
2. Similarly, the variance of x is 11 and variance of y is 4.13 for each dataset
3. The correlation coefficient (how strong a relationship is between two variables) between x and y is 0.816 for each dataset



- Dataset I appears to have clean and well-fitting linear models.
- Dataset II is not distributed normally.
- In Dataset III the distribution is linear, but the calculated regression is thrown off by an outlier.
- Dataset IV shows that one outlier is enough to produce a high correlation coefficient.

Q4: What is Pearson's R?

Pearson's R is a measure of the linear correlation between two variables X and Y. Its similar to R-square what we used in this course.

Q5: What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?

Scaling is to bring all variables at same level. This make comparison of coefficients comparable and easy. Since all variables are on same scale coefficients can indicate which variable has higher coefficient vs low.

Narmalized scaling is done by x – mean(x) / sd (x) . It makes mean 0 and standard deviation 1.

MinMax scaling is done by x – min(x)/max(x) – min(x). It bring all values between 0 and 1.

MinMax scaling takes care of outliers better.

Q6: You might have observed that sometimes the value of VIF is infinite. Why does this happen?

VIF = 1 / 1-R-square. So VIF will become infinite when R-square is 1. R-square is 1 when TSS = RSS. Which means poor fit.

Q7: What is the Gauss-Markov theorem?

In statistics, the Gauss–Markov theorem states that in a linear regression model in which the errors are uncorrelated, have equal variances and expectation value of zero, the best linear unbiased estimator (BLUE) of the coefficients is given by the ordinary least squares (OLS) estimator, provided it exists. Here "best" means giving the lowest variance of the estimate, as compared to other unbiased, linear estimators. The errors do not need to be normal, nor do they need to be independent and identically distributed (only uncorrelated with mean zero

and homoscedastic with finite variance). The requirement that the estimator be unbiased cannot be dropped, since biased estimators exist with lower variance.

Q8: Explain the gradient descent algorithm in detail.
Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model. Parameters refer to coefficients in Linear Regression and weights in neural networks.

Q9: What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.
Q-Q plot is plot between quantiles (quantiles represent number of elements under certain percentile values. Like 50% percentile = 0 means 50% of elements are below 0 value.). Q-Q plot being straight line indicate that data came from some theoretical distribution such as a normal or exponential.

Q10: Provide R Square Value at the end.

.889