Problem stat:

1. First it needs to check supplied data source (DB, csv etc)
2. Size
3. Problem type (Any data Science project ends in one of the categories namely, Regression, Classification, Clustering, Anomaly detection and Reinforcement learning).
4. There are different type of regression categories: Linear, Logistic, polynomial, stepwise, Ridge, Lasso, Elastic net,
5. Difference correlation and regression: **Correlation quantifies the degree to which two variables are related**. **Correlation** does not fit a line **though data points**. Correlation is calculated using correlation coefficient (value range from -1 to +1).
6. Cost function used in Linear regression is Mean Squared error.
7. Cost function used in Logistic regression (David Cox, 1958, B‘ham UK) is **cross entropy** or **log loss**.
8. Classification model measuring matrix are **AIC** (Alkaline information criteria – smaller the value is better), **NULL deviation** (TSS) and **Residual** **deviation** (RSS), (smaller is he difference then better), Confusion matrix, ROC-AUU, F1score(Harmonic mean of precision & recall).
9. ROC-AUC CURVE: **True positive rate (Sensitivity)** Vs **False positive rate (1-Specificity)**.
10. Difference between Multinomial and Polynomial logistic regression: Multinomial means more than two class of target classification whereas Polynomial is higher order predictors like x\*\*2, x\*\*3 etc.