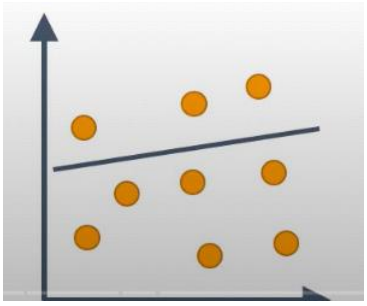
Interview Que and Ans

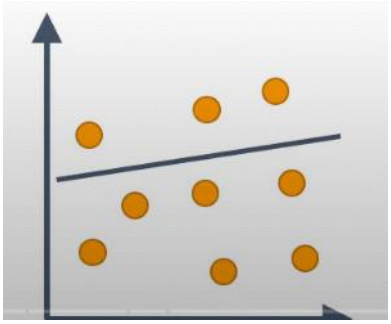
17 October 2022 16:50

Q11) How is Overfitting different from underfitting?

--> **Overfitting** -

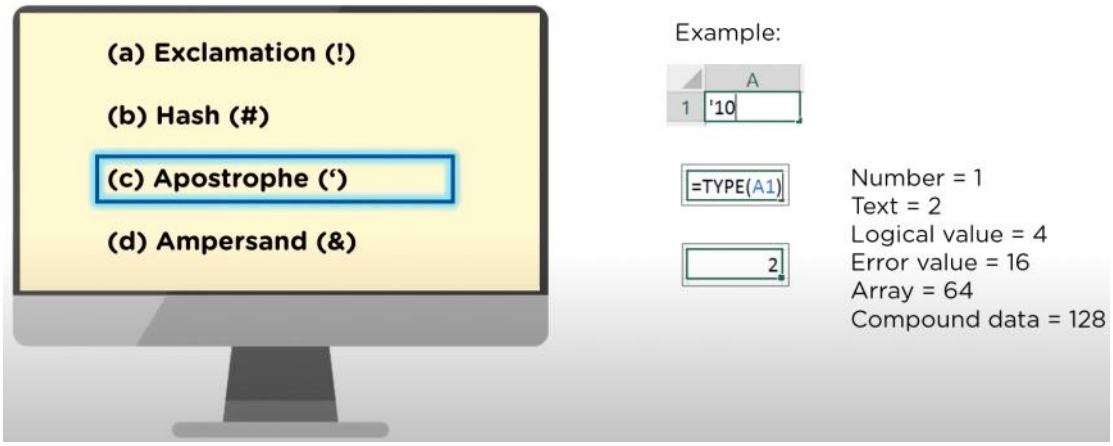
1. Model trains the data too well using the training set.
2. The performance drop significantly over the test set.
3. Happens when the model learns the noise and random fluctuations in the training dataset in details.
4. 

-->**Underfitting** -

1. The model neither trains the data well nor cam generalize to new data.
2. Perform poorly both on train and the test set.
3. Happens when there is less data to build an accurate model and also when we try to build a linear model with a non-linear.
4. 

Q12) In MS EXCEL, a numeric value can be treated as a text value if it proceeds with.

-->



Q13) How do you subset or filter data in SQL?

--> To subset or filter in SQL, we can use **WHERE** and **HAVING** clause.

Q14) Difference between WHERE and HAVING clause in SQL ?

--> WHERE clause -

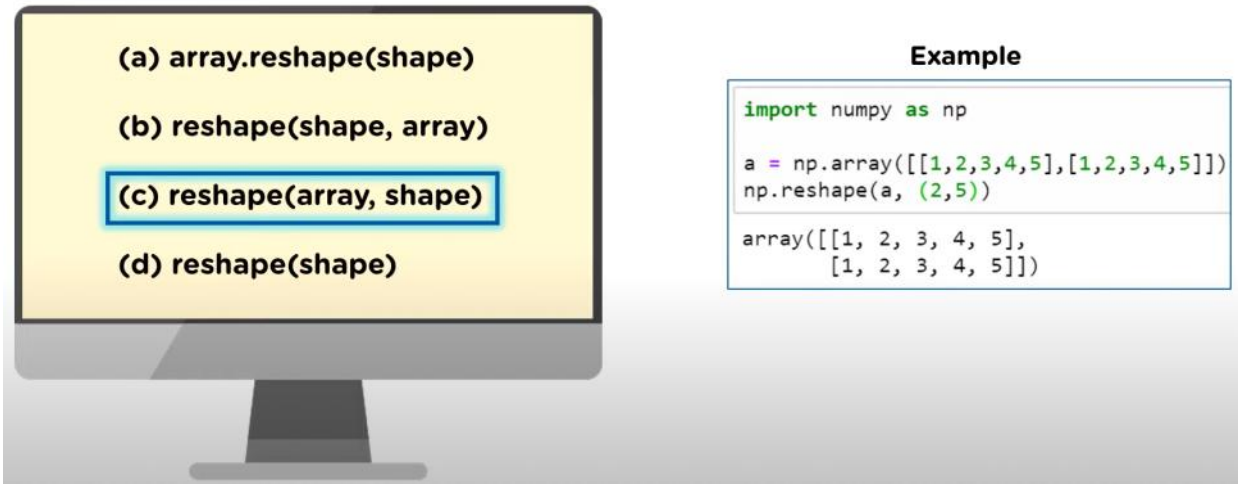
1. WHERE clause works on row data
2. In where clause, the filter occurs before any grouping are made.
3. Aggregate function cannot be used.
4. Syntax : Select **** from ****where***

--> HAVING clause -

1. HAVING clause works on aggregates data.
2. HAVING is used to filter values from a group.
3. Aggregate function can be used.
4. Syntax : select***from***where***group by***having***order by***

Q15) Correct syntax for reshape() function in NumPy?

--->



Q16) What is the criteria to say whether a developed data model is good or not ?

---->

1. A Good model should be insightful and self-explanatory.
2. The model developed should be able to easily consumed by the client for actionable and profitable results.
3. A good model should easily adapt to changes according to business requirements.
4. If the data gets updated, the model should be able to scale according to the new data.

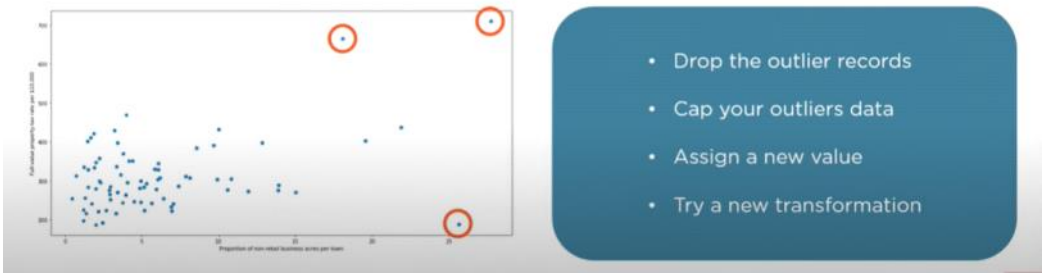
Q17) What is the significance of exploratory data analysis ?

--> Exploratory data analysis is an important steps in any data analysis process.

1. Exploratory data analysis (EDA) helps to understand the data better.
2. It helps obtain confidence in your data to a point where you're ready to engage a machine learning algorithm.
3. It allows you to refine your selection of further variables that will be used later for model building.
4. You can discover hidden trends and insights from the data.

Q18) How do you treat outliers in a dataset ?

--> An outlier is a data point that is distant from other similar points. They may be due to variability in the measurement or may indicate experimental errors .



Q19) Explain descriptive, predictive, and prescriptive analytics.

--->

Descriptive -

1. Provide insights into the past to answer 'What has happened'.
2. Uses data aggregation and data mining techniques.
3. E.g. An ice cream company can analyse how much ice cream was sold, which flavour were sold, and whether more or less ice cream was sold than day before.

Predictive -

1. Understand the future to answer ' what could happen'.
2. Uses statistical models and forecasting techniques.
3. E.g. Predict the sale of ice creams during summer, spring and rainy days.

Prescriptive -

1. Suggest various courses of action to answer ' what should you do'.
2. Uses optimization and simulation algorithms to advice possible outcomes.
3. E.g. Lower prices to increase sales of ice creams, produce more/less quantities of a certain flavour of ice cream.

Q20) What are the different types of sampling techniques used by data analysts?

--> Sampling is a statistical method to select a subset of data from an entire dataset(population) to estimate the characteristics of whole population.

1. Simple random sampling.
2. Systematic sampling
3. Cluster sampling
4. Stratified Sampling
5. Judgmental or Purposive Sampling.