Interview Que and Ans

14 October 2022 14:46

DAY 1 14/10/2022

Q1) Difference between Data Mining and Data Profiling?

---> Data Mining -

1. Data mining is the process of finding relevant information which has not been found before.
2. It is the way in which raw data is turned into valuable information.

---> Data Profiling -

1. Data Profiling is usually done to assess a dataset for its Uniqueness, consistency and logic.
2. It cannot identify incorrect or inaccurate data values.

Q2) Define the term Data Wrangling in data analytics?

---> Data wrangling is the process of cleaning, structuring and enriching the raw data into a desired usable format for better decision making.

----> Different term of data wrangling.

Discover -- structure -- Clean -- Enrich -- Validate -- Analyse

Q3) Common problems that data analysts encounter during analysis?

---> 1. Handling duplicate and missing values.
2. Collecting the meaningful right data and right time.
3. Making the data secure and dealing with compliance issues.
4. Handling data purging and storage problems.

Q4) What are the various steps involved in any analytics project?

---> 1. Understand the problem.
2. Data collection.
3. Data cleaning.
4. Data exploration and analysis.
5. Interpret the result.

Q5) Technical tools have you used for analysis and presentation purpose?

----> 1. MySQL.
2. Power BI
3. Tableau
4. Excel
5. PowerPoint

Q6) What are the best practice for data cleaning?

--->

1. Make a data cleaning plan by understanding where the common error take place and keep communications open.
2. Identify and remove duplicates before working with the data. This will lead to an effective data analysis process.
3. Focus on the accuracy of the data. Maintain the value of data, provide mandatory constraints and set cross-fields validation.
4. Standardise the data at the point of entry so that it is less chaotic and you will be able to ensure that all information is standardised, leading to fewer error to entry.

Q7) How can you handle missing values in dataset ?

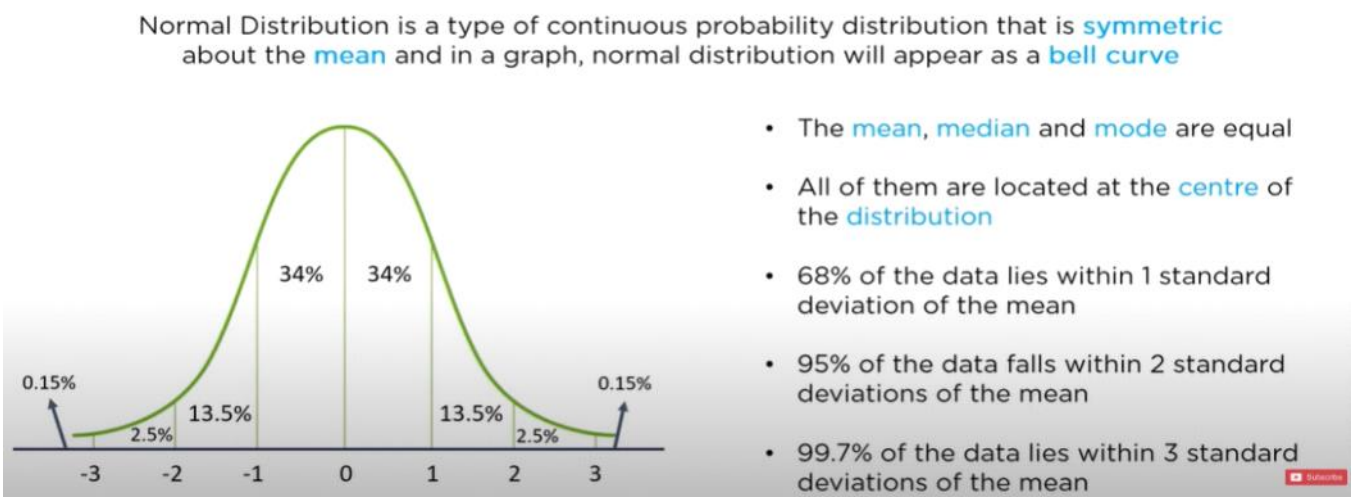
---->

1. Listwise deletion --> In listwise deletion method, an entire record is excluded from analysis if any single value is missing.
2. Average imputation --> Use the average value of response from the other participants to fill in the missing values.

3. Regression Substitution --> You can use multiple-regression analysis to estimate a missing value.
4. Multiple imputation --> It creates plausible values based on the correction for the missing data and then averages the simulated dataset by incorporating random errors in your predictions.

Q8) What do you understand by the term Normal Distribution?

---> Normal Distribution is a type of continuous probability distribution that is symmetric about the mean and in a
graph , normal distribution will appear as a bell curve.



Q10) How to joining different from blending in Tableau ?

---> Data joining -

1. Data joining can only done when the data comes from the same source.
2. E.g. Combination two table from the same datasets, or two or more worksheets from the same Excel file.
3. All the combined tables or sheets contains common set of dimensions and measures.

---> Data Blending -

1. Data Blending is used when the data is from 2 or more different sources.
2. E.g. Combining the oracle table with SQL Server, or combining Excel, or combining Excel sheet and oracle table.
3. In Data Blending, each data sources contains its own set of Dimensions and Measures.