Assignment - 1

Q1. Define Data Analysis.

→ Data Analytics is defined as science of extracting meaningful, valuable ~~data~~ information ~~from~~ raw data.

Q2. Write Difference between data analysis & data analytics.

→

Data analysis	Data Analytics
1) Understanding past data.	Gaining insight + predicting future
2) It is descriptive type	It is descriptive + predictive + prescriptive
3) It has narrow scope	It has broader scope
4) Mainly used for reports & summaries	Mainly used for decisions, predictions.
5) One can find anonymous relations.	One cannot find anonymous relations.
6) Does not deal with inferential analysis	It supports inferential analysis.

Q3. Explain in detail types of data analytics.

→ ① Descriptive analytics - what happen?
② Diagnostic analytics - Why did it happen?
③ Predective analytics - What will it happen
④ Prescriptive analytics. - How can we make it happen?

① Descriptive analytics :- Here raw data is examine. or convert answers question. What happen? By analyzing Valuable information found from available data.
eg. Organisation record, Stakeholders & customer

② Diagnostic analytics :- To find root which causes issues. It try to gain deeper understanding of reason behind patterns of data found in past.
eg. Social media marketing campaign.

③ Predective analytics :- It predict future based on the available current & past data, Prediction based on historical data, build model & used them to forecast a future value.
eg. Demand of perticular package around holiday session.

④ Prescriptive analytics :- It not only expect what will happen? & when will happen? But also why it will happen
eg. health care industry.

Q4. Explain framework of data analysis.

→ ① Data Connection Layer :-
- This is first step where raw data is collected.
- Source can include websites, mobile apps sensors, social media, etc.
- The goal is to gather relevant & accurate data.
eg. collecting user clicks & purchases from an e-commerce website.

② Data management layer :-
Once data has been extracted data scientist must perform number of functions that are grouped under data management layer. The data may need to be normalized & stored in certain database architecture to improve data query & ~~access~~ by analytics layer.

③ Data Analysis Layer :-
In this layer, a data scientist uses number of engines to implement analytical functions depending on task at hand, a data scientist may use one or multiple engines to build on analytics application.

④ Presentation layer :-
It includes tools for building dashboard applications that display result of analytics engine.

Q5 What is AUC & ROC?

→ AUC :- AUC is insensitive to class imbalance if majority level of data are positive or negative.

ROC :- ROC curve is graphical plot that shows performance of binary classifier. curve is created by plotting true positive rate against false positive rate at various threshold.

Q6. Explain confusion matrix

→

		Predicted class: Positive	Predicted class: Negative
Actual class	Positive	True Positive (TP)	False Negative (FN) Type Error II
	Negative	False Positive (FP) Type I Error	True Negative (TN)

- Confusion matrix is techniques for summarising performance of classification algorithm.
- Table have 2 dimensional actual & predictable. Both dimension have TP, TN, FP & FN. Row represents actual class while column represent Predicted class

TP :- correctly predicted positives
TN :- correctly pregative negatives
FP :- Incorrectly predicted as positive
(Type I error)
FN :- Incorrectly predicted as negative
(Type II error)

Accuracy = (TP + TN) / Total
precision = TP / (TP + FP)
Recall = TP / (TP + FN)
F- score $\frac{2 \times (\text{precision} \times \text{Recall})}{(\text{precision} + \text{Recall})}$