Data Analysis

Data analysis involves converting raw data into different understandable forms to get clear insights of the underlying information there are different phases of data analysis and they are

1. Identify
2. Collect
3. Clean
4. Analyze
5. Intrepet

**Types Of Analysis**

**Descriptive Analysis:**

What’s going on the business( Generally done by the fresher data analysts to find the overall insights of the data

**Diagnostic Analysis:**

This deals with why problem i.e why there is a raise or fall etc.

**Predictive Analysis:**

Prediction of future based on the past or historical data also involves forecasting

**Prescriptive Analysis:**

This is combination of all the three previous analysis. This is the analysis which will decide the final summary of the analysis.generally done by the subject matter expert

**Tools For Data Analysis:**

* Excel Spreadsheets
* Can be useful by a feature called VBA ,can handle large data
* Amazon,Wipro uses excel for data analysis

**2.Rapid Miner**

* Has drag and drop
* Usage of any data
* Can create models
* Has graphs and other visualization tools
* Has over 15000 operations
* United Health Group,Paypal uses rapid Miner

**3.Talend**

* Open source
* Best for cloud and big data intrepetation
* Automation at its best
* Less development costs
* Integrates with many technologies
* Very faster
* Dominos,Capgemini use Talend

**4.KNIME**

* Open source
* ML models
* Best for pipelining
* Good GUI,drag and drop
* Less programming
* Automation and team based
* United Health Group,lego

**5.SAS**

* Automatic code generation
* Used by Microsoft
* Creates interactive dashboards
* Genpact,Accenture,IBM

**6.R and PYTHON**

* Useful for mainly research,
* Completely free and easy to use
* Parallel computation
* Fast updates
* Uber ,google,facebook

**7.Apache Spark**

* Specially designed for handling large databases
* Open source
* Access by different ways
* Simple programming
* More tools for ML and data visualization
* Access data quickly
* Executes veruy fast
* Java,scala,r,python useful
* Batch processing
* ML algorithms used
* Fault tolerant in data loss
* Netflix,Ebay

**8.Qlikview**

**Business analysis**

Self analysis

Leading business anlysis

In-memory storage technology

All the data stored in ram before analysis

Addition of annotations

Capgemini,cognizant,Accenture,Mercedes benz

9.PowerBI

Great dashboard

Collection of cloud based apps and services

Has excel inside it

Drag and drop

Capgemini,nestle

10.Tableu

Best in all the tools

Different visualization tools

Different ways to use it

Very fast

Dashboard

Queries turned to visualizations

Transparent filters

Deloitte,cisco,amazon,adobe