difference betn data science & big data

① It is field of scientific analysis of data in order to solve the complex problem & necessary cleaning & preparing the data.	① Big data is storing & processing large volume of structured & unstructured data that can not be possible with traditional application
② It is used in bio-tech, energy, gaming & insurance	② used in retail, education, healthcare & social-media
③ Goals:- data classification, anomaly, prediction, scoring & ranking	③ Goal:- To provide better customer service, revenue identification & Marketing

Applications of data Science:- ① healthcare ② Gaming ③ Image recognition ④ Logistics ⑤ predict future market trends ⑥ Recommendation sys ⑦ Streamline manufacturing

hgilp rs

hpgirls

data-Explosion:- Business model transformation

BGNA

Globalization

personalization of services

New Sources of data

5 v's of big data :-

① Volume, ② Velocity ③ Variety ④ Value ⑤ Veracity

Volume:- Machine data, App logs, clickstreams log, Emails, Contracts, geographical info sys & geo-spatial data

Velocity:- Speed of creation data

Variety:- nature of data structured & unstructured data.

① Sensor data ② Mob networks ③ Amazon, facebook, Yahoo ④ Social Media

5 V's of Data :-

① Volume :- i] Larger than conventional relational database
ii] Consist of terabyte or petabyte of data.

② Velocity :- i] refers to speed of Generat° of data.
ii] How fast data is generated & processed to meet the demands.
iii] Determines real potential in data.

③ Variety :- i] Hetrogeneous source
ii] Nature of data - structure & unstructure

④ Value :- i] Business value derrived from data.
ii] Objective :- Generate some sort of value for the company doing an analysis.

⑤ Veracity :- True Data relates to assurance of data quality, intgrity, credibility & accuracy.

Business Intelligent

1. Business Intelligence tends to provide reports, dashboard & queries on business question for the current period & in the past
2. BI make it easy to answer the question like quarter-to-date rev, progress towards quarterly targets, & show prodt performance
3. Understand history
4. Power BI

Data Science

1. DS is tends to use disaggregated data & looking forward to extract info in such a way to get the future predictions
2. data science tends to be more exploratory in nature & may use scenario optimization to deal with more open-ended question.
3. definition.
4. Application - HP GIRLS

CC

1. it provide recourses on demand
2. It refers to internet services from SaaS, PaaS to IaaS
3. Cloud is used to store data & info on remote servers
4. low maintainance
5. on demand soln
6. eg:- GCP & Salesforce

Big data

1. handle huge volume of data
2. It can be structured, unstructured
3. It deals with volume of data & info
4. big data is highly scalable, robust & cost-effective
5. eg:- Cloudera, Apache.

data science life cycle :- BDE DVDC PDMFE

BDE DVDC PDMFE

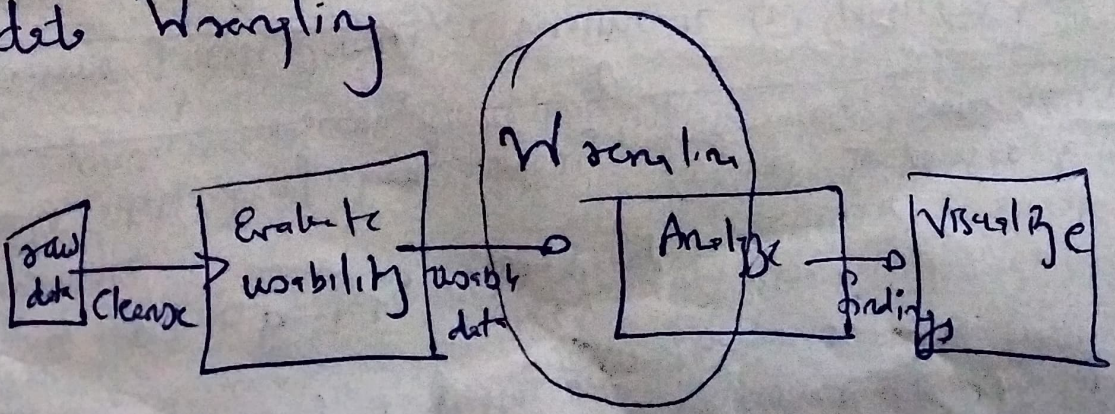
Business understanding
data Exploration
Data Visualization
Data cleaning
Predictive Modeling
data Mining
Feeds Eng.

Parameter	Structured data	Unstructured data
Representation	Stored in rows & column format.	Unstructured data do not follow specified format
Meta data	Syntax	Services
Storage	database management system	Unmanaged file structure
Standard	SQL, ADO.net, ODBC	Open XML, SMTO, SMS
Tool for integration	ETL	Manual data integr
Used by ORG	Low vol of ^{r}	High vol of ^{r}

Information	data
① It is processed data ② Info is specific ③ Info depends on data ④ Info is output of computer ⑤ refers to knowledge of interpretation of data	① Raw data ② data is not specific ③ doesn't depend on info ④ data is input to computer ⑤ data refers to facts, measurements etc.

Qualitative data	Quantitative data
data which cannot be measured Type:- Nominal data & Ordinal data Makes use of appearance, color textures & other qualities These are indescribable terms eg:- Jean is bad	It can be measured Types:- Interval data & Ratio data Measures quantities such as length, size, amount, price & even duration These are numerical terms Eg:- Jean is of 7 people

data Wrangling



Data Wrangling :-

① Defination :-
Data wrangling is the process of cleaning, organizing and transforming raw data into desired format for better decision making in less time.

② Six STEP Data Wrangling Process :-
1] Discovering
2] Structuring
3] Cleaning
4] Enriching
5] Validating
6] Publishing

③ Benifits :- 1] Enables users to make timely & consumed decisions
2] Process large volume of data easily
3] Merges several datasets into one for analysis.