Page No.
Date

Unit 4 :- Data Analytics

Data analytics refers to qualitative & quantitative techniques & processes used to enhance productivity & business gain. Data analytics is a broader term that has analysis as a subhead & analytics is basically the concepts used to do analysis.

* Data analytics lifecycle

i) Discovery
ii) Data prep.
iii) Model planning
iv) Model building
v) Communicate result
vi) Operatinalize

~~Types~~ 4 General ways data analysts extract patterns

i) Data mining
ii) Text Analytics
iii) Data visualisation
iv) Business Intelligence

Page No.
Date

i) Data mining

Process of extracting usable data from large dataset. Data mining involves data collection, warehousing & compater processing.

Real life scenarion :- In healthcare industry. The algo. can evaluate behavioral patterns, knowledge building & decision making.

ii) Text Analytics

Drawing meaning out of written communication. Usually text analytics software relies on text mining & Natural language processing (NLP)

Real life scenario :- Autocorrect on your mobile devices.

iii) Data visualization

Data visualization presents a clear picture of what data actually means. Using bar graphs pie charts, tables & other visuals, data visualization makes the data easier for those making business decision to comprehend.

Real life scenario :- Excercise data, energy usess, weekly screen time charts.

Page No.
Date

iv) Business Intelligence

It leverages analytics tools to convert data to actionable insights. often paired with data visualization techniques,

Real life scenerion :- Use of BI to capitalize on customer trends & extend customized offers in real-time.

* Data Analytics tools

i) Artificial Intelligence → makes decision that can provide plausible likelihood in achieving goal.

ii) NoSQL Database → Delivers method for accumulation & retrieval of data.

iii) R programming → Assists Data scientistists in designing statistical software.

iv) Data Lakes

ETL → Extract, Transform & Load.

Page No.
Date

* Quality issues for data analysis

i) Duplicate Data → same data entered multiple times
ii) Incosistent formats → same type of data in incosistient format.
iii) Incomplete information → Crucial pieces of information are missing.
iv) Data inconsistency → Data in same field i.e. either in diffrent language or in diffrent units.
v) Inaccurate data → Potential mis-spellings exist or data is simply inaccurate.
vi) Invalid data → Data can't possibly correct based on simple rules & logic.
vii) Data imprecision → Lack of precision.

* where to address data quality issues

i) Address the issue in the source system.
ii) Fix during the ETL process.
iii) Fix at meta-data layer.

* Risks associated with poor data quality

i) missed opportunities.
ii) Lost revenue.
iii) Reduced operation efficiency & productivity.
iv) Customer dissatisfaction.
v) misanalysis

vi) Reputational damage
vii) Lack of compliance.
viii) Increased financial cost.

* Data Analysis task.

i) Defining the question
ii) collecting the data
iii) Cleaning the data
iv) Analysing the data.

* Types of data

i) First party data → structured & organised.
ii) second party data → structured but may not be that relevan
iii) Third party data → unstructured data

* Types of data analysis

i) Descriptive Analysis
ii) Diagnostic Analysis
iii) Predictive Analysis
iv) Prescriptive Analysis.
v) Cognitive Analysis.

Page No.
Date

i) Descriptive Analysis

Descriptive analysis identifies what has already happend. The descriptive analysis provides a response to the question "what happen" by presenting historical data ~~or~~ in the ~~form~~ of dashboards. while companies might not draw firm conclusion from any of these insights; ~~it~~ ~~we~~ summarizing & describing data will help them to determine how to proceed.

ii) Diagnostic Analysis

Diagnostic analysis focuses on why something has happend. Diagnostic analysis digs deeper into descriptive analysis to discover the root causes of outcomes. Creating comprehensive information is an important part of diagnostic analysis.

iii) Predictive

This form of analytics ~~p~~ makes prediction about future events based on prior data. The predictive analysis employs the information we've gathered to generate reasonable predictions about what will happen next.

iv) Prescriptive Analysis.

prescriptive analysis allows you to make recommendations for the future.

v) Cognitive Analysis.

Cognitive analytics system looks through the data in its knowledge base for answers to queries that make sense.

- Analytics having human like intelligence are known as cognitive analytics.
- AI & ML algorithms are frequently used in cognitive analytics.
- Unlike simple analytics, cognitive analytics can show patterns & correlations.

* Exploratory Data Analysis.

Exploratory data analysis (EDA) is used by data scientists to analyse and investigate data sets & summarize their main characteristics, often employing data visualisation methods. It makes easier for data scientists to discover patterns, spot anomalies, test a hypothesis or check assumptions. EDA is primarily used to see what data can reveal beyond formal modelling or hypothesis testing task & provides better understanding of dataset.

Page No.
Date

* Importance of EDA

It can help you identify obvious errors, as well as better understand patterns within the data, detect outliers or anomalous events, finding interesting relationship among the variables.
Data scientists can use EDA to ensure the results they produce are valid & applicable to any desired business outcomes & goals.

* EDA tools

i) Clustering & dimension reduction techniques.
ii) Univarient visualization.
iii) Bivarient visualization.
iv) multivarient visualization.
v) K-means clustering.
vi) Predictive models like linear regression.
vii) python.
viii) R programing language.

* Types of EDA.

i) Univarient Non-graphical → only one variable.
ii) Univarient graphical → Stem & leaf, Histogram, bar plot, Box plots.
iii) multivarient: ~~graphical~~ non-graphical → Two or more variables
iv) multivarient graphical → Bar plot or bar chart.