IBM DATA ANALYTICS

Phase 1 : Learning Plan

Decoding Data : Insights & Impact Through Analytics 2025-26

I Introduction to Data analytics & tools

~17.5 hrs ⇒ 35 hrs

Data analytics for beginners

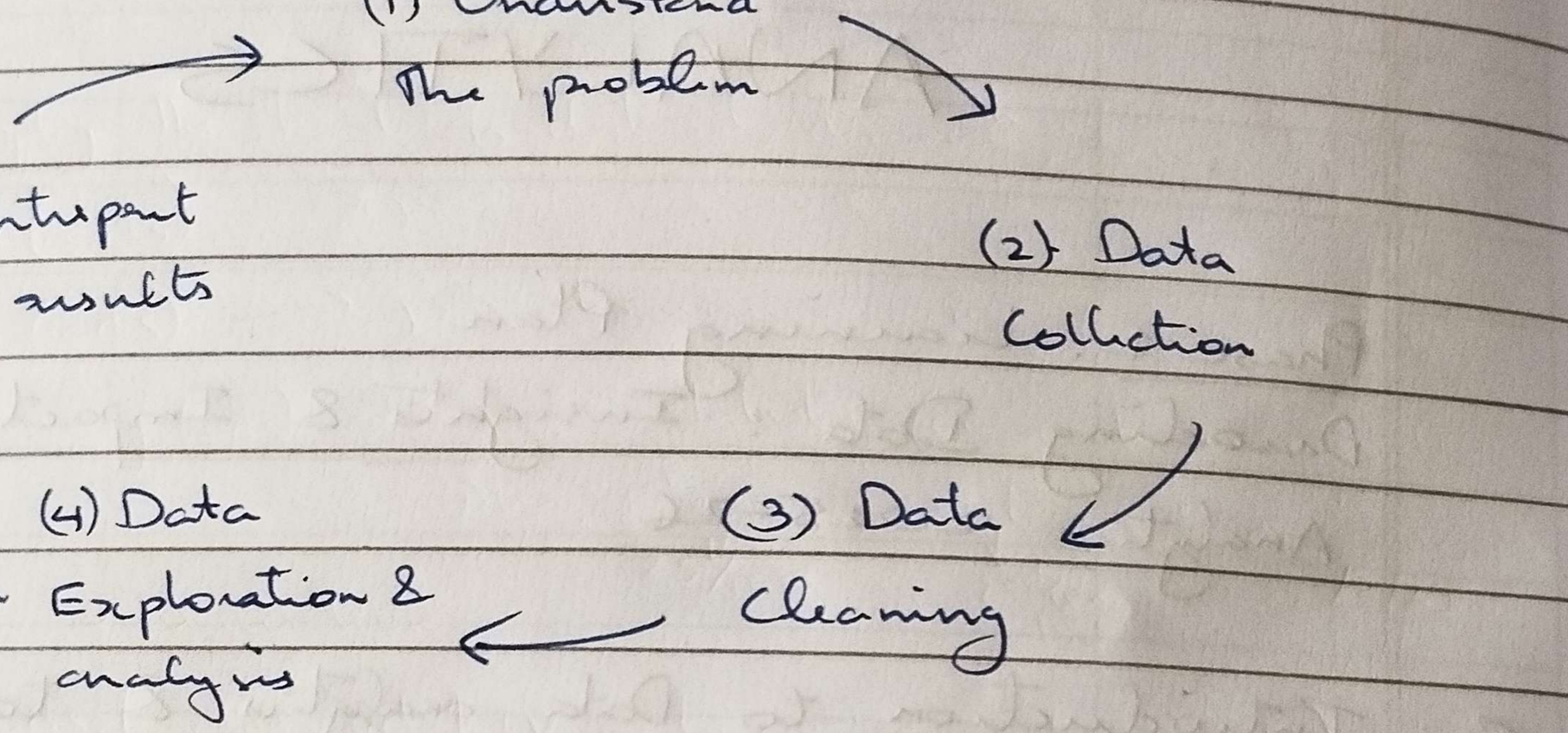
→ What is Data analytics? (DA)

- Process of exploring & analyzing large datasets to find hidden patterns, unseen trends, discover correlation & valuable insights.

- Ways to use data analytics (DA):

1. Improved decision making
2. Better Customer Service
3. Efficient Operations
4. Effective marketing

→ Steps involved in DA.



→ Tools for DA:

1. Python
2. R - an open source programming language used for numerical & statistical analysis
3. tableau - data visualisation & analysis tool
4. Power BI - business intelligence tool
5. QlikView - provides interactive analytics with in memory storage. to to analyse large datasets & use data discovery to support decision making.
6. Apache Spark - open source data analytics engine to process data in real time & carry out complex analytics using SQL & ML.
7. SAS - Statical analytics software, perform analytics visualise data, write SQL & build ML models

→ DA Applications

1. Retail
2. Healthcare
3. Manufacturing
4. Banking
5. Logistics

→ Case Study: Wall mart

→ Use case Demo: Linear Regression Analysis

- R :- open-source & free software
  - statistical analysis (data manipula" & mining)
  - data visualisation
  - Build models

* Download RStudio Then. new file → RScript
  Install all necessary packages
  install.packages("dplyr") # Deeplayr
  library(dplyr)
  install.packages("broom") # used to convert messy data to clean data / data cleaning
  library(broom)
  install.packages("caTools") # to built linear regress"
  library(caTools)
  install.packages("ggplot2") # data visualis"
  library(ggplot2)

* Run each line-by-line

Load the dataset

ads <- read.csv("../../.../..") # Path of csv file, remember to make '/'
variable

Display the head of dataset, dimension, summary

head(ads) # → gives first 6 rows

dim(ads) # → gives dimensions - no of rows & cols

summary(ads) # → gives statistics for each column
Min, 1st Quantile, Median, Mean, 3rd Qu, Max
25%/25 2nd Qu 50/50 75/25

Data visualisation

plot(ads$sales, ads$TV, type='p', col='red')
# ↳ (x axis value, y axis value) - shows how TV values change with sales.

pairs(ads) # → Shows multiple graph on how each column affects other. shown like a matrix of graphs
↓ upper triangular matrix

Correlation Analysis

install.packages("corrplot")

library(corrplot)

nums.cols <- sapply(ads, is.numeric) → checks if ads contains numbers

nums.cols # → display

Correlation only needs numeric data

cor.data <- cor(ads[, nums.cols] → filter out numeric columns

cor.data # → gives correlation o/p

1 means highly (close to 1), is more correlated.

corrplot(cor.data, method='color')

Simple linear regression
target ~ independent variable

model_simple <- lm(sales ~ TV, data = ads)

summary(model_simple) ↳ linear model

↳ This shows the Intercept Estimate: mean value of sales when budget is 0

tidy(model_simple) # Also shows the increase of col per units

Multiple linear regression

model_multi <- lm(sales ~ TV + newspaper + radio, data = ads)

summary(model_multi)

tidy(model_multi)

coef <- summary(model_multi)$coefficient

coef # → gives the coefficients of variables

how to train using CA tools library

set.seed(101) ↳ split to train & test set

sample <- sample.split(ads$TV, SplitRatio = 0.70)

train = subset(ads, sample = TRUE)

test = subset(ads, sample = FALSE) → have no value that are in train set

model <- lm(sales ~ ., train)

↳ all independent variables

summary(model)

res <- residuals(models) # → to get the residual value

res <- as.data.frame(res) # → convert to dataframe

sales.prediction ← predict(model, test)

sales.predictions

↳ combine predicted value with original data

results ← cbind(sales.prediction, test$sales)

results.

↳ organise to columns.

colnames(results) ← c('pred', 'real')

results ← as.data.frame(results)

results.

Data Fundamentals

Introduction to Data Concepts

5 Whys

- Ask 'why?' 5 times
- This is done by data scientists to analyse the data
- This can be repeated with a different set of questions
- Aim is to find the root cause

section from ← "Data science in our world"

Data → new information

Data analysis → process of collecting, cleaning & transforming data to obtain insights to make informed decisions

Date
Page

Use of data insights

- improve operations
- reduce cost
- understand customs
- increase profit
- find new innovation
- efficient improvement

Data → Structured Data (highly organized)

Eg: names, dates, geolocation, stock info ..

→ Unstructured Data (no predefined format)

Eg: texts, images, customer comments, emails ..

Database

- organized collection of structured data
- Relational database: collection of multiple data sets

- Data → Quantitative Data: numerical data.

  Eg: height, weight, length, temp ..

  → Qualitative Data: categorical data

  Eg: smell, location, color (characteristics) ..

Quantitative Data → Discrete: whole nos, cannot be divided.

Eg: 3, 76, 12; no. of people etc ..

→ Continuous: Bina limits. can be divided

Eg: 13.15, 56.351; height, weight ...

Qualitative Data
- Nominal : Labels
  Ej: gender, married, blonde ..
- Ordinal : Ordered
  Ej : likely, neutral, unlikely ; or grades A,B,C..

5 Vs - The elements of Big Data
- Volume → vasts amt of data generated
- Variety → structured, unstructured, semi - etc
- Velocity → speed of data generation & movement
- Veracity → quality of data, accuracy.
- Value → ability to get insights or value.

4 types of Data Analytics

1) Descriptive analytics - 'what is happening?'
- simplest & common
- uses raw data from multiple sources to create trends & patterns.
- uses historic & current data
- Ej: monthly revenue report, daily stock report..

2). Diagnostic analytics - 'why it happens?'
- takes the insights from descriptive to find causes of specific problems
- creates more connection & identifies patterns

of behaviour

3) Predictive analytics - 'What is likely to happen?'
- forecasting using historic data

4) Prescriptive analytics - 'what should happen?'
- combins all insights to form a course of action to take

Data Analysis Process

Collect → Clean → Analyze → Visualize

ETL (Extract, Transform & Load)
- data integrations
- combins data from various sources into a single consistent source & loads to data warehouse

Data Storytelling
- process of converting data analysis to a simple understandable story to influence business decisions
- combins data, visualisations & narrative

Visualization

Types of charts used:

- Pie chart
- Bar chart
- Column chart
- Line chart
- Scatter chart

} Quantitative

- Flow
- Structure
- Interrelationship

- Action plan
- Map

} Conceptual

Types of data comparison

- Relative proportion - pie chart, column
- Ranking - Bar
- Time - column, Line → series of points in time data
- Frequency - column, Line
- Correlation - Bar, Scatter → relationship blw 2 variables

Data Science in Our World

- Data analyst → collect & examine large datasets to identify trends, forecasts & visualization

- Data scientist → design & create new process for data modelling That use algorithm, prediction analysis & statistical analysis

3 Data Science methodologies

1) Cross-Industry Standard Process for Data Mining (CRISP-DM): contains 6 iterative non-sequential phases -
   1. Business Understanding
   2. Data Understanding
   3. Data Preparation
   4. Modeling
   5. Evaluation
   6. Deployment
   4.

2) Knowledge Discovery in Database (KDD):
   1. Selection
   2. Preprocessing
   3. Transformation
   4. Data Mining
   5. Interpretation / Evaluation

3) SEMMA
   1. Sample
   2. Explore
   3. Modify
   4. Model
   5. Access

Overview

(current business problem) (one-time proposed soln) (+future problem prediction)

1) Business Understanding → 2) Data exploration & preparation

~~current business~~

3) Data segmentation & transformation

4) Data visualisation & presentation → 5) Train models

6) Deploy models

Buisness Undistanding : ~~The buis~~

- The buisness problem is defined
- Buisness sponsor actively contributes
- After, The data scientist & team defines an analytical solution

Data Exploration & Prepration

- Data is identified & collected. They could be structured or unstructured.
- Data Exploration - finds pattens, relationship & insights from data.
- Data Prepration - Creates dataset & data cleaning

Data representation & transformation

- Descriptive statistics (minimum, mean, mode etc) is used to understand data.
- Exploratory visualisation is used further to understand
- Data transformation with machine learning.

Data visualisation & presentation

- Effective, Attractive, Impactive.

Train Data models

- Supervised learning: trains with labelled data to build general rules

- Unsupervised learning: trains on vast amounts of unlabelled data
- Reinforced learning: It is feedback based. It learns a good action & bad action from trial & error.

Deploy data models

- Integration of the ML model into business.
- for better prediction for future solution

Data Science Project Team

- Data Analyst
- Data Scientist
- Data Engineer

Overview of Data Tools & Languages

Open Source Software

- Software that is published public & anyone can use
- 4 key aspects: Use - anyone
  View - "
  Modify - "
  Share - "

- Community
- Contributin
- Committin - reviews & approves source code changes
- Code of Conduct - protects community
- Contribution Guidelin

Tools

→ Git & Github
→ ~~Micro~~soft Excel
→ Google Sheets
→ Python
→ SQL

→ IBM Watson Studio

- It is an IDE
- has built-in data refining tool
- other open source tools

→ Tableau :

- ~~inte~~ creates interactive visualisations

→ Matplotlib

- Python library
- visualisation in static, animated & interactive

→ R & RStudio
→ Apache Spark

→ Jupyth Notebook

⇒ Scala:

- scalabl language
- gnual purpose, robust, object-oriented & functional programming - programming language.

Selection of Tools to use

1) Determine The problem
   - Technique : 5 Why?, ask questions
   - Tools: Design Thinking

2) Collect & clean data
   - Techniques: Structured/unstructured data collection & cleaning
   - Tools: SQL, NoSQL, Excel, sheets

3) Analyze Data
   - Techniques: classification, regression, & clustering
   - Tools: Python, IBM Watson Studio, RStudio.

4) Validate analysis & model
   - Techniques: ML, Business model
   - Tools: Python, IBM Watson, RStudio

5) Visualize data
   - Techniques: storytelling, comm insights, presentation
   - Tools: Tableau, Matplotlib, Powerpoint

Clean, Refine & Visualize Data with IBM Watson Studio

* IBM Watson Studio
- collaborative DS & ML
- easy visualizations
- open source tools
- develop, train, deploy AI models

. Log into IBM cloud.
. ~~Select Catalog in Dashboard~~
(Shows a full guide on how to use Watson Studio for projects)

Your Future in Data: The Job Landscape

- There is an increasing demand for AI & Data Analytics

Scope
- Predictive analysis in IoT
- NLP in voice assistant (PA + AI)

DA in other fields

- Finance
- Healthcare
- Insurance
- Scientific servis
- IT
- Marketing
- Management of companies
- Manufacturing
- Public administration
- Education servis
- Arts & Entertainment
- Retail

Data Analyst responsibility

- Understand business direction & objective
- Explore meaning b/w the numbers
- Analyze comms of insights based on data
- Present easy-to-understand insights
- Contribute to decision making
- Determin significance of outliers.

Data Scientist responsibility

- Identify problems with greatest opportunity for business to solve
- Develop hypothesis - to research & find patterns
- Mine & analyze big data
- Use ML
- Build, train & deploy AI models
- Interpret data for actionable insights.
- Propose Soln/ strategies for business problem

Skills for Data Analysts

- Data Analysis
- Data Science
- Data Quality
- Comm. - problem solving
- teamwork & colab - research
- Tools: Excel, SQL, Python, Tableau, Power BI

Imp Skills for Data Scientist

- DA
- DS
- ML
- Predictive models
- AI
- Comm - problem solving
- teamwork & colab - research
- creating - presentation
- working
- Tools: SQL, Python, Tableau
  R, Apache Hadoop, Tensorflow, Scala, Scikit-le

Characteristics Companies look for

- curious
- confident about your position
- flexible & adaptable
- storytelling.