

Google Data Analytics Certificate

Introducing Data Analytics

↓
A collection of Facts:

- * Numbers, pics, videos,
- words, measurement observation

Data Analysis.

Data Analytics

→ The collection, transformation & organization

of data in order to draw conclusion

make predictions, and drive informed

decision-making

process of examining data to draw conclusion about that information

ASK question to define both the issue to be solved and

what would equal a successful result

Prepare by building a timeline and collecting data, store the data



Process by cleaning it so that it is complete, correct, relevant,

free of errors & outliers

Analyze data to find patterns, relationship and trends

Share the findings and recommendations



Act on the result and focus on improving key areas

Why Businesses needs to control data?

- * Improve processes
- * Identify opportunities & trends
- * Launch new products
- * Serve customers
- * make thoughtful decision

draw conclusion

Make prediction

Take decision

Transforming data into insights

Data Science

Machine Learning

* to automate or

to make many-many

decision under
uncertainty

Statistics

* Make few

important ashiv, dec, medium *

decision under
uncertainty

Analytics

* not so much A

Data Ecosystem

↳ A group of elements that interact with one another

↳ Various elements that interact with one another in order to produce,
manage, store, organize, analyze and share data

↳ include software, hardware and people

Cloud ↳ A place to keep data online, rather than a computer hard drive

Cloud not ↳ Virtual location

Diff b/w Data Scientist and Data Analyst

↓ Similarities to diff ↓

Creating new ways of modeling

and understanding the unknown

by using raw data

find new question using data

find answers to existing question

by creating insights from data

Data driven decision making

* Using facts to guide business strategy

Data Analyst Skills

Analytical skills → Qualities and characteristics associated with

 solving problems using facts.

1. Curiosity

2. Understanding context → The condition in which smtg exist or happens

3. having a technical mindset → The ability to break things down into smaller steps

or pieces and work with them in orderly and logical way

4. Data design → How we organize information

5. Data strategy → Management of people, processes and tool used in

Thursday 14.10.2021 Thinking About Analytical Thinking

14.10.2021



Identifying and defining a problem and then

Solving it by using data in an organized step-by-step manner.

5 Aspects of how to think analytically

1. Visualization → The graphical representation of information

→ ex: graphs, maps, or other design elements

→ can help data analysts understand and explain information more effectively.

2. Strategy → Strategic mindset is key to stay focused and on track.

→ help data analyst see what they want to achieve with the data and how they can get there

3. Problem orientation → to identify, describe and solve problem

4. Correlation → ASK a lot of questions

→ is like a relationship

→ Correlation does not equal causation (just because 2 pieces of data both are trending in same direction, that doesn't necessarily mean they are related)

5. Big picture and Detail oriented thinking → zoom out and see possibilities and opportunities

Exploring core Analytical Skills

- i) Analytical Thinking ii) creative thinking iii) critical thinking

When have problem the first question data analyst wanna find out is

1) What is the root cause?

→ 5 "why". The fifth and final answer will give insights

2) Where are the gaps in our process?

→ Most people use gap analysis

→ Gap analysis lets us examine and evaluate how a process works currently in order to get where we want to be in future

→ Use of gap analysis: 1) To improve a product

→ To become more efficient

→ General approach : 1) understand where we are now compared to where we want to be
2) can identify the gaps between current and future state
and determine how to bridge the gaps.

3) What we did not consider before?

→ What informations not procedures might be missing from a process

Using Data to Drive successful outcomes

Data driven decision making → using facts to guide business strategy

With Data: 1) can gain valuable insight

→ verify theories or assumptions

→ better understand opportunities and challenges

→ support in objective

→ help make a plan

Data Phases and Tools

Life cycle of Data

1. Plan : * What kinda data is need

- * How the data is managed throughout the life cycle
- * Who will be responsible for it
- * What are the optimal outcomes

2. Capture : * Data is collected from variety of sources

3. Manage : * How we care for our data

- * How and where it's stored
- * The tools used to keep it safe and secure
- * Actions taken to maintain it properly

4. Analyze : * Data used to solve problems, make great decisions, support business goals

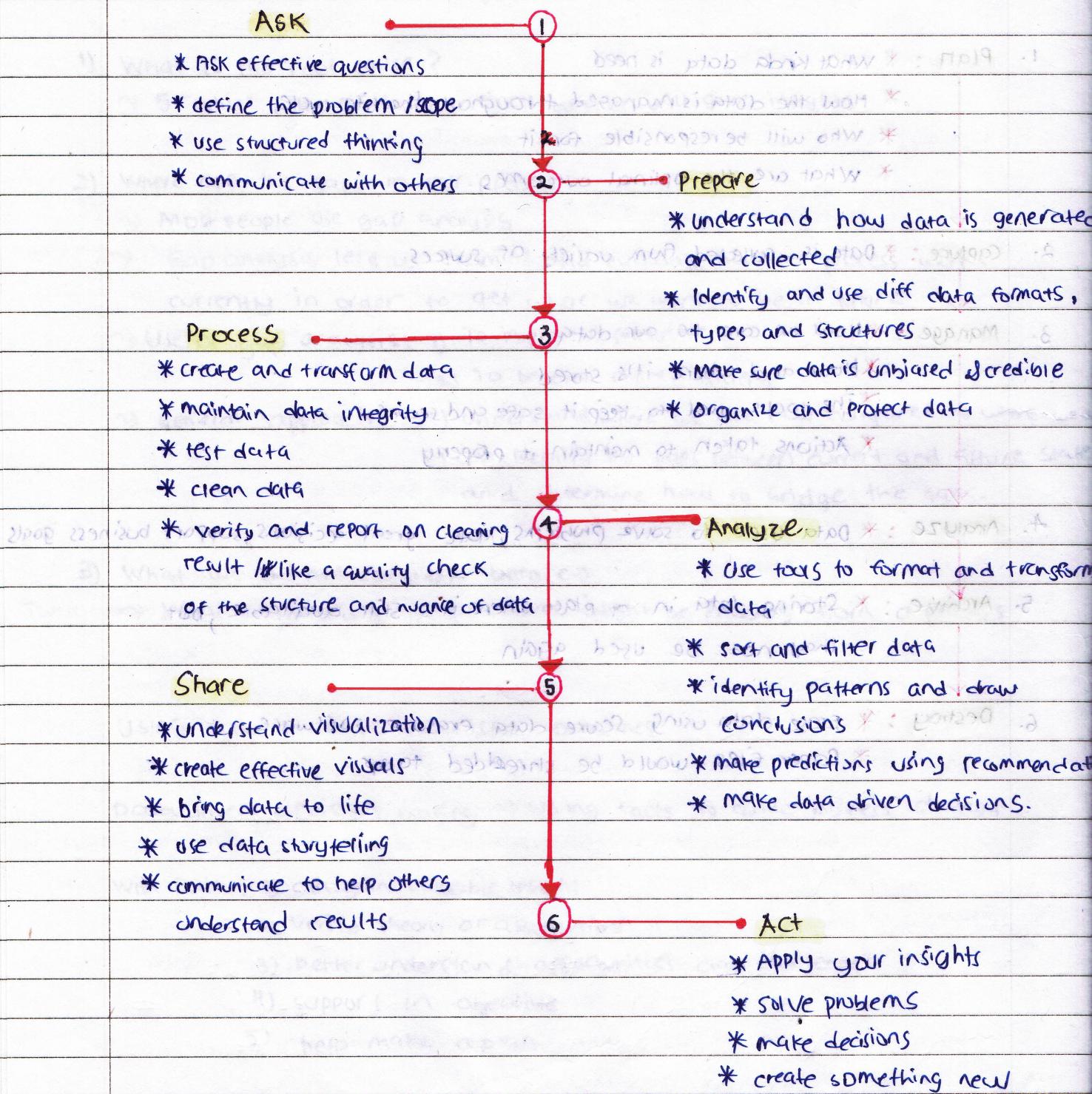
5. Archive : * Storing data in a place where it's still available, but may not be used again

6. Destroy : * Erase data using secure data erasure software

* Paper files would be shredded to prevent reuse

still at work until now *

6 Phases of Data Analysis



The data analysis toolbox

Exploring data analyst tools

1. Spreadsheets → digital worksheet, stores, organizes and sorts data

i) Microsoft Excel ii) Google Sheets

→ have formulas and functions

2. Query Language → Computer programming language that allows to retrieve and manipulate data from a database.

i) SQL - Structured query language

3. Data visualization → i) Tableau ii) Looker

Spreadsheets

	A	B	C	D
1	Cell A1			
2				Row 2
3				

column

* column labels are called as 'attribute'.

cell = A1

row

SQL ↗ For large datasets

- Store
- Organize → Data
- Analyze

Examples of database that use SQL:

- i) Oracle
- ii) Microsoft SQL Server
- iii) MySQL
- iv) PostgreSQL

Basic structure of a SQL query

SELECT

[choose the column you want] #2

FROM

[from the appropriate table] #1

WHERE

[a certain condition is met] #3

Suggested order to write SQL queries
Start Big(data table) and go small
(specific conditions)

Postgres SQL

Postgres=# \l → To have a look on our database currently present in our server

postgres=# Create Database studentdb; → To create database name studentdb

postgres=# \c studentdb → studentdb=# → From postgres to switch to studentdb database

\s studentdb =# drop database demo; → To delete a database

studentdb=# Create Table students(name text, address text, age int, number int);

studentdb=# Insert into students (name, address, age, number) values ('John', 'LA', '34', '011345');

studentdb=# SELECT * from student;

studentdb=# SELECT * from student WHERE age = 45;

Machine Learning

Evaluating performance for classification

$$1. \text{ Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

→ Accuracy useful when target classes are balanced

$$2. \text{ Recall} = \frac{\text{Number of True Positive}}{\text{Number of True Positive + Number of False Negative}}$$

$$\hat{L}(R-A) = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} (y_i - \hat{y}_i)^2$$

→ Ability of a model to find all the relevant cases within a dataset

$$3. \text{ Precision} = \frac{\text{Number of True Positive}}{\text{Number of True Positive + Number of False Positive}}$$

→ Ability of a model to identify only the relevant point

$$4. \text{ F1-score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

→ harmonic mean of precision and recall

→ optimal blend of precision and recall

$$\hat{L}(R-A) = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} (y_i - \hat{y}_i)^2$$