DATA, DATA, EVERYWHERE

Module 01:

Q: Why Data is important in any industries?

A: Companies in many industries use data to improve processes, identifying opportunities and trends, launch new products and make decisions.

Q: What is Data?

A: Data is a collection of facts, including numbers, pictures, videos, words, observations and more. Data can be used in everyday life (e.g. fitness trackers, product reviews) and in business (e.g. learning about customers, improving processes, helping employees)

Dataset: A collection of data that can be manipulated or analyzed as one unit.

Q: What is Data Analytics?

A: Data Analytics is the science of data.

Q: What is Data Analysis?

A: Data analysis is the collection, transformation and organization of data in order to draw conclusions, make predictions and drive informed decision-making.

Q: Who is Data Analyst?

A: Someone who collects, transform and organize data in order to draw conclusions, make predictions and drive informed decision-making.

Data analysts finds data, analyzes it, and uses it to uncover trends, patterns, and relationships.

Data-driven decision-making is using facts to guide business strategy.

Benefits of data-driven decision-making include improved results, greater confidence in decisions, and more proactive approach to opportunities.

Google used data-driven decision-making to determine the value of managers in their organization

6 Phases of Data Analysis Process

1. **ASK**

   - Asked effective questions to define the problem.
   - Business Challenge/Objective/Question

2. **PERPARE**

- Identified what data they needed
- Data collecting and storing the information.
- Data generation, collection, storage, and data management

3. **PROCESS**

- Cleaning data and checking the information.
- Data cleaning/data integrity

4. **ANALYZE**

- To find patterns, relationships, and trends.
- Data exploration, visualization, and analysis

5. **SHARE**

- Shared the report with stakeholders
- Communicating and interpreting results

6. **ACT**

- Implement changes and take actions
- Putting your insights to work to solve the problem

Q: What are Data Ecosystems?

A: The various elements that interact with one another in order to produce, manage, store, organize, analyze and share data.

Data Scientists vs Data Analyst

Data Scientists create new questions using data, while analysts find answer to existing question by creating insights from the data sources.

Data + business knowledge = mystery solved

Module 02:

Analytical Skills:

Analytical skills are qualities and characteristics associated with solving problems using facts.

Five essential analytical skills for data-driven decision-making are below

5 Analytical Skills

i. **Curiosity**

- Wanting to learn something and seeking out new challenges and experiences.
- Helps to make predictions, research answers and draw conclusions

ii. Understanding Context
- How you group things into categories
- The condition in which something exist or happens
- Helps to make predictions, research answers and draw conclusions

iii. Having Technical Mindset
- The ability to break things down into smaller steps and work with them in an orderly and logical way
- Using facts to explore gut feelings

iv. Data Design
- How you organize information
- Helps to make data easy to access, understand, and make the most of.

v. Data Strategy
- The management of people, process, and tools used in data analysis.
- Incorporates people, processes, and tools to solve a problem

Analytical Thinking

Analytical Thinking involves identifying and defining a problem and then solving it by using data in an organized, step-by-step manner.

5 key aspects of Analytical Thinking

i. Visualization
- The graphical representation of information and helps data analysts understand and explain information more effectively

ii. Strategy
- Strategizing helps data analysts stay focused and on track and improves the quality and usefulness of the data collected

iii. Problem-orientation
- Keeping the problem top of mind throughout the entire project

iv. Correlation
- The relationship between two or more pieces of data, but correlation does not equal causation.

v. Big-picture and detail-oriented thinking

- Being able to see the big picture as well as the details, while detail-oriented thinking is about figuring out all of the aspects that will help execute a plan

Examples of questions data analysts ask: root cause of a problem, **Five Whys**

Gap analysis: A method for examining and evaluating the current state of a process in order to identify opportunities for improvement in the future

Module 03 & 04:

Data Life Cycle: Data has its own life cycle, which consist of 6 stages.

i.   Plan
- Deciding what kind of data is needed, how it will be managed, who will be responsible for it and the optimal outcomes.

ii.   Capture
- Collecting the data from a variety of sources and bringing it into the organization.

iii.   Manage
- Caring of the data, storing it, keeping it safe and secure, and taking actions to maintain it properly.

iv.   Analyze
- Using the data to solve problems, make decisions, and support business goals.

v.   Archive
- Storing data in a place where it is still available, but may not be used again.
- Keep relevant data stored for long-term and future reference.

vi.   Destroy
- Using secure data erasure software to delete the data from hard drives and shredding paper files.

Warning: Be careful not to mix up or confuse the six stages of the data life cycle with the six phases of the data analysis life cycle (Ask, Prepare, Process, Analyze, Share, and Act). They shouldn't be used or referred to interchangeably.

Data Analysis is the process of analyzing data, and is not a life cycle

Introduction to the tools data analyst use:

1. **Spreadsheets**
   - Digital worksheet to store, organize and sort data, organize and sort data
   - Formulas and functions to perform calculations and tasks.
   - Identify patterns and piece the data together in a way that works for each specific data project
   - Create excellent data visualizations, like graphs and charts.
   - ➢ **MS Excel or Google Sheet**
2. **Query Languages**
   - Computer programming language to retrieve and manipulate data from a database, most widely use SQL
   - Allow analysts to isolate specific information from a database(s).
   - Make it easier for you to learn and understand the requests made to databases.
   - Allow analysts to select, create, add, or download data from a database for analysis.
   - ➢ **SQL: (Structured Query Language)**

     Structure of basic query: SELECT, FROM and WHERE.

     - SELECT: is used to choose the columns from a table.
     - FROM: is used to choose the tables where the columns you want are located.
     - WHERE: is used to filter the data based on certain conditions.
3. **Visualization Tools**
   - Graphical representation of information
   - Tools like Tableau and Looker to create visuals that are easy to understand
   - Turn complex numbers into a story that people can understand
   - Help stakeholders come up with conclusions that lead to informed decisions and effective business strategies
   - ➢ **Tableau**
   - ➢ **Programming Language (R with RStudio) or Python**

Module 05:

Understanding Data and Fairness:

- Data analysts have a responsibility to make sure their analyses are fair.
- Fairness means ensuring that analysis does not create or reinforce bias.
- Conclusions based on data can be true and unfair.

Data Ethics:

- We think about what is the good and right way of using data?
- It's not just about minimizing harm but it's actually beneficial to people?
- Who is collecting the data, why are collecting it? How are they collecting it and for what purpose?
- Keeping data protected and private.