FOUNDATIONS OF DATA SCIENCE

Dr.Nagaraju Baydeti
Department of Computer Science and Engineering
National Institute of Technology Nagaland

Presentation Outline

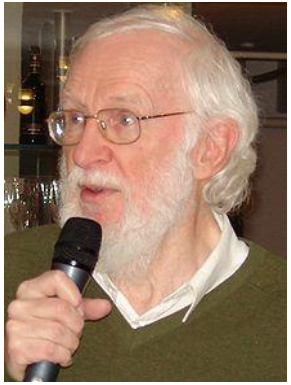
1. What is Data Science?
2. Where it started - The Origin of Data Science
3. Why Data Science?
4. The Data Science - Workflow
5. Life Cycle of Data Science
6. Data Science – Application Areas
7. Job Opportunities
8. Anatomy of Data Scientist
9. Wish to be a Data Scientist – Skillset Required
10. How to start? – Data Science using Python

1. What is Data Science?

- Study of data
- Process of deriving knowledge
- Gain insights from a huge and diverse set of data
- Organizing, processing and analyzing the data
- Works with Structured and Un-structured data
- Decision making
- An inter-discipline field

2. Origin of Data Science?

- John Wilder Tukey, an American mathematician.
- Developed Fast Fourier Transform algorithm and box plot.
- He introduced the word '***bit***' while working with John Von Neumann.
- In 1962, he described a field called '***data analysis***'.

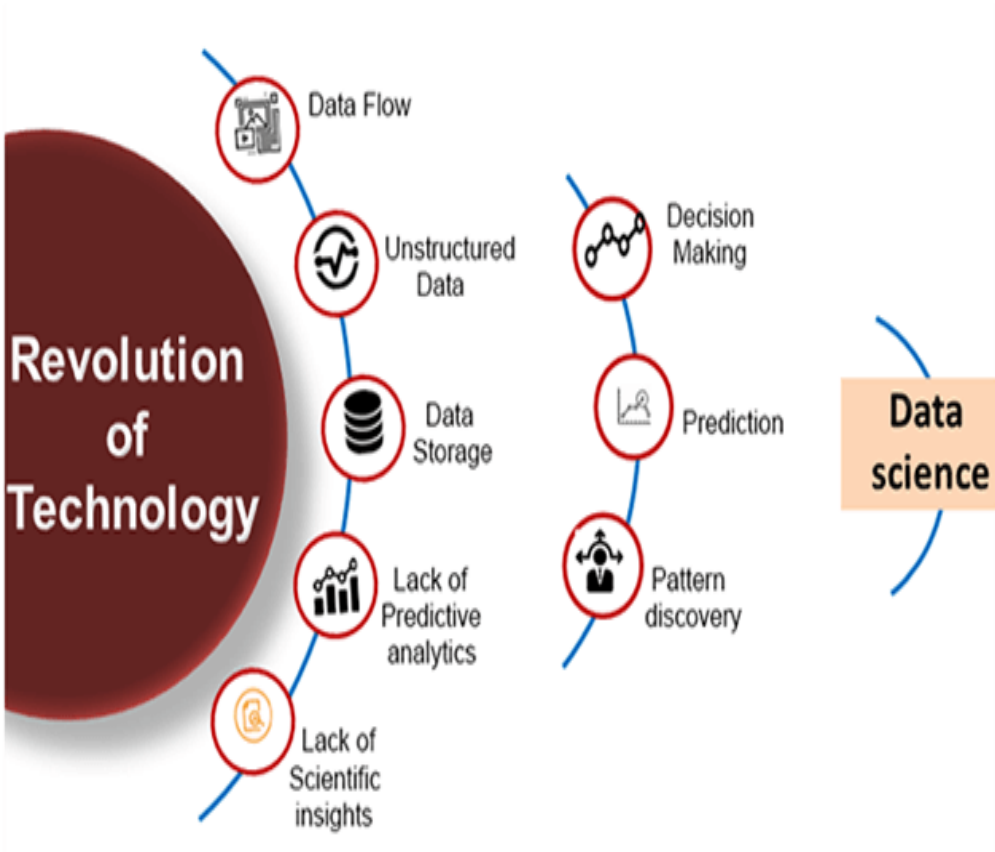


- Peter Naur, a Danish computer science pioneer & Turing award winner.
- Designed ALGOL 60 (Algorithmic Language 1960)
- Well known for the contributor of Backus-Naur form (BNF) notation used in describing the syntax for most programming languages.
- He disliked the term '***computer science***' and suggested it to be called as '***data science***'

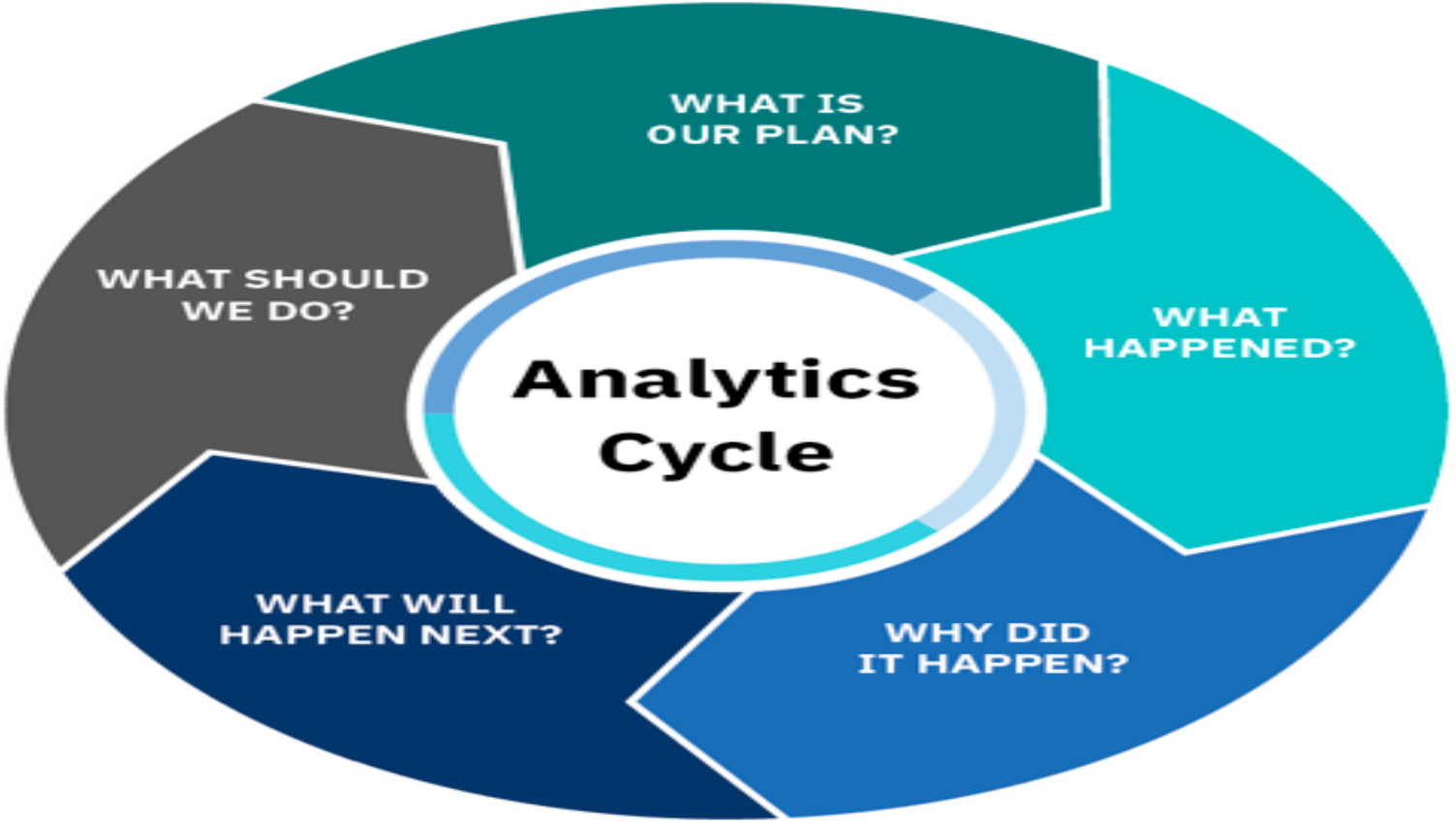
For more details about the history of Data Science: https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#6ab0108855cf

3. Why Data Science?

- Data – the new fuel for Industries
- Data-driven approach
- According to the U.S. Bureau of Labor Statistics
  - The number of roles for Data Scientists has grown by 650% since 2012
  - 11.5 million jobs will be created by 2026
- Data Scientist ranks top emerging jobs on LinkedIn

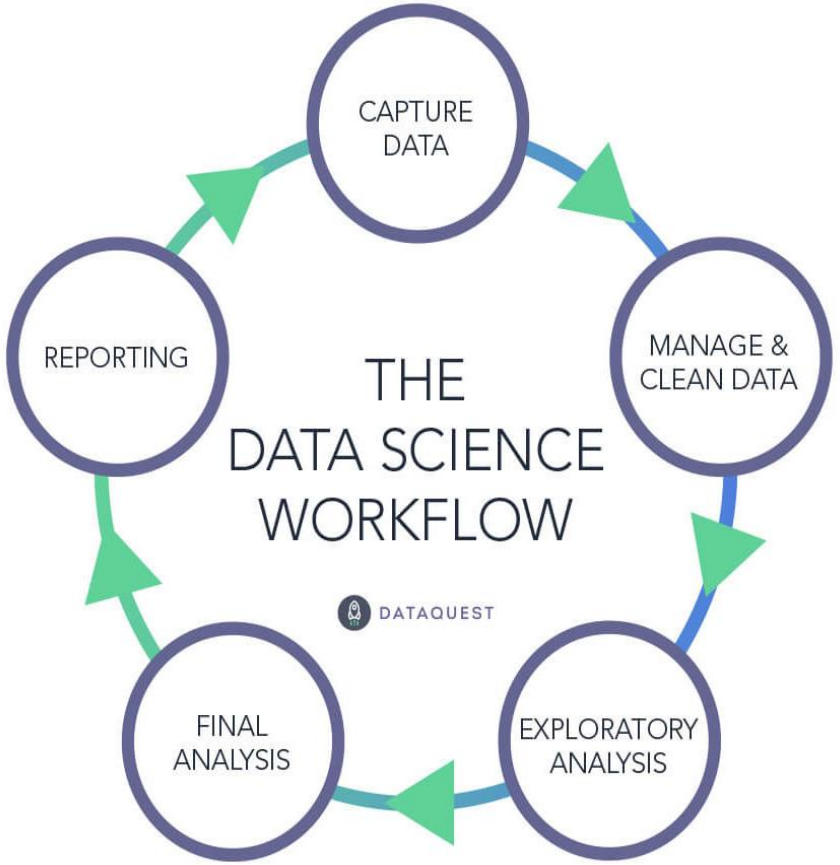


4. The Data Science - Workflow

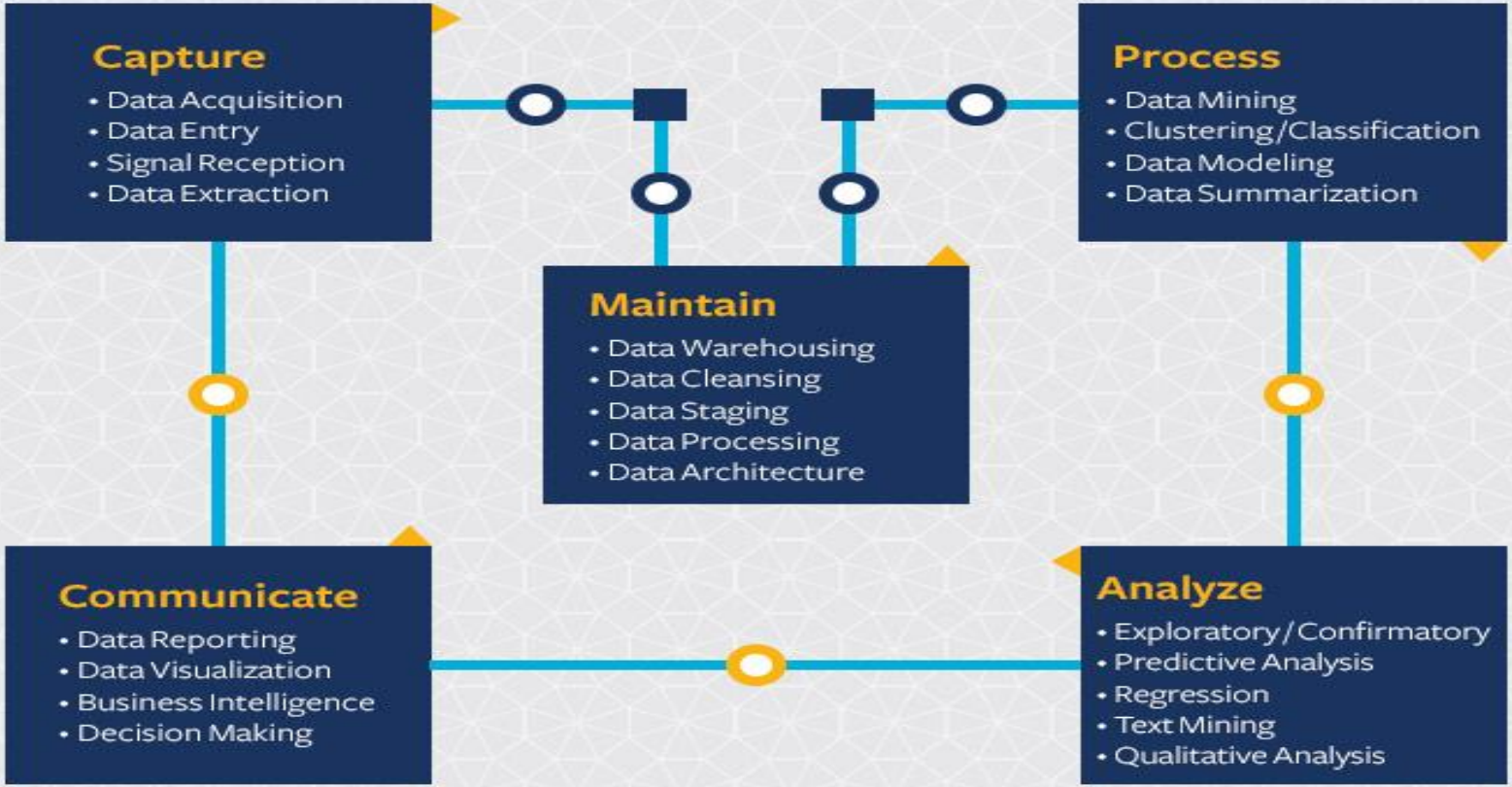


4. The Data Science – Workflow … continued

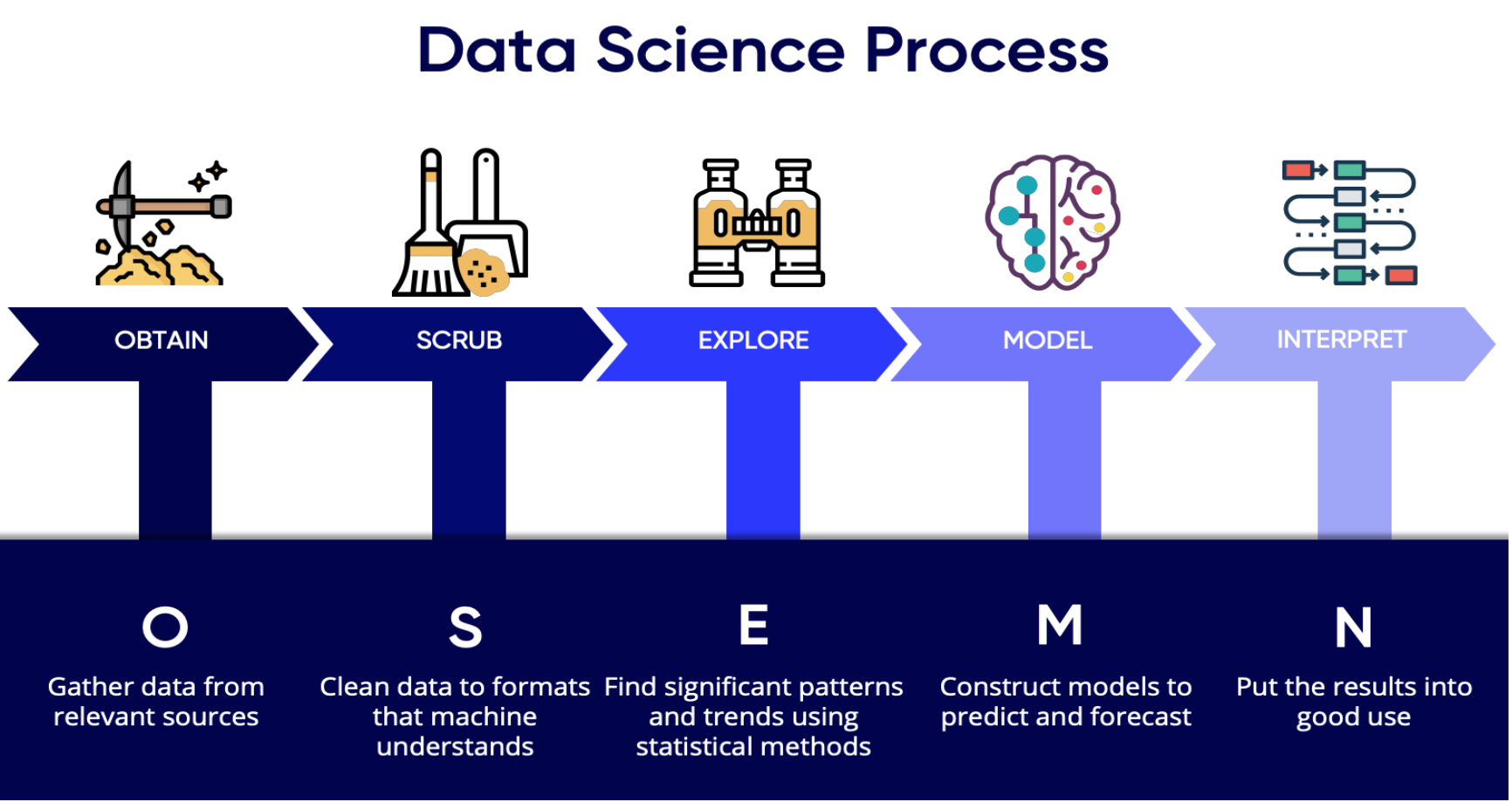
- Define a Problem
- Obtain the Data
- Scrubbing / Cleaning the Data
  - Missing Values
  - Data out of Range
  - Time Zone Differences
- Exploratory Data Analytics
  - Compute Descriptive Statistics to Extract Features and Test Significant Variables
- Data Modeling
  - Create an Efficient Method to Store Information
- Data Visualization



5. Life Cycle of Data Science



5. Life Cycle of Data Science … continued



6. Data Science – Application Areas

Source: https://data-flair.training/blogs/data-science-applications/

7. Job Opportunities

Various Job roles in Data Science domain:

- Data Analyst
- Machine Learning Expert
- Data Engineer
- **Data Scientist**
- Data Administrator
- Data Architect
- Business Analyst
- Business Intelligence Manager

8. Anatomy of Data Scientist

GLOBAL TECH COUNCIL

ANATOMY OF THE NEW ERA SCIENTIST
DATA SCIENTIST

AVERAGE SALARY

Data scientist get a handsome salary of ***$1,20,000*** on an average.

EDUCATION

- More than ***85%*** of data scientist have a masters degree.
- More than ***45%*** have a Ph.D.

COMMON FIELD OF STUDY

- Mathematics and Statistics ***(32%)***
- Computer Science ***(19%)***
- Engineering ***(16%).***

CAREER POSSIBILITIES

- World top most companies such as Apple, Amazon, Uber are employing data scientist.
- Data scientist get employed in sectors such as Finance, Marketing, Consulting, Healthcare, Pharmaceuticals, etc.

TECHNICAL SKILLS

Python Coding, SQL Database/Coding, Apache Spark, Machine Learning and AI, Statistic, Data Visualization.

BENEFITS

- Data scientist is termed as ""the sexiest job of the ***21st*** century" by Harward Business School.
- More than ***94%*** of graduates in the field of data sciences, successfully got employment since ***2011.***
- The data scientist is one of the best jobs in America according to the Glassdoor.

www.globaltechcouncil.org

9. Data Scientist – Skillset Required

DATA SCIENTIST
MUST-HAVE SKILLS

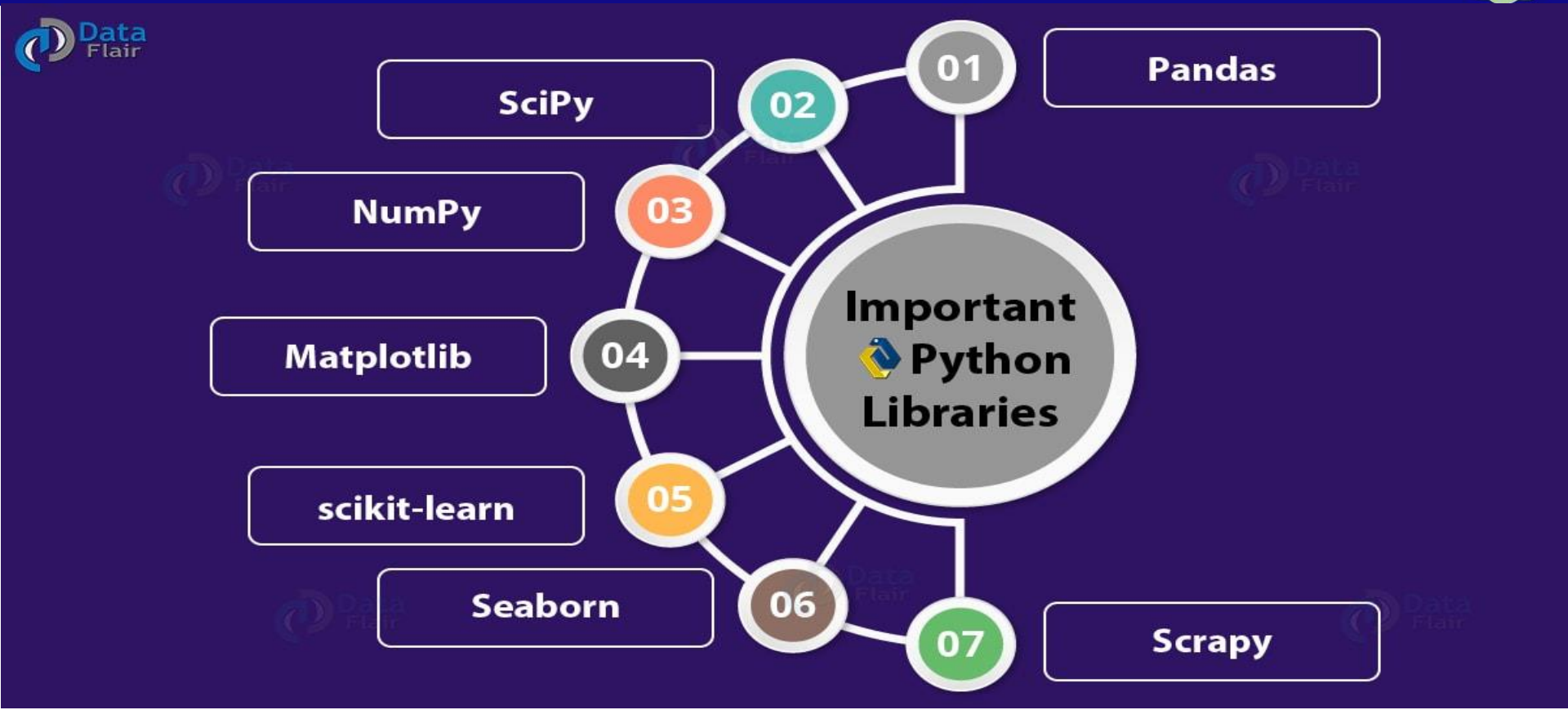
MATH & STATISTICS
- Machine Learning
- Statistical Modeling
- Exploratory Analysis
- Clustering
- Regression Analysis

DOMAIN KNOWLEDGE & SOFT SKILLS
- Inclination towards business operations
- Keen on working with data
- Problem solver
- Strategic, proactive, and cooperative
- Interested in hacking

PROGRAMMING & DATABASE
- Computer Science Fundamentals
- Database Management System
- Data Visualization
- Python
- Big Data

COMMUNICATION & VISUALIZATION
- Storytelling skills
- Convert data-based insights into decisions
- Collaborative with Sr. Management
- Knowledge of tools like Tableau
- Visual art design

10. Data Science using Python



References

1. https://datascience.berkeley.edu/about/what-is-data-science/
2. https://techterms.com/definition/data_science
3. https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#6ab0108855cf
4. http://sudeep.co/data-science/Understanding-the-Data-Science-Lifecycle/
5. https://towardsdatascience.com/5-steps-of-a-data-science-project-lifecycle-26c50372b492
6. https://www.javatpoint.com/data-science
7. https://www.mygreatlearning.com/blog/difference-data-science-machine-learning-ai/
8. https://www.dataquest.io/blog/what-is-data-science/
9. https://www.mygreatlearning.com/blog/different-data-science-jobs-roles-industry/
10. https://www.datacamp.com/community/blog/data-science-past-present-future
11. https://www.javatpoint.com/data-science
12. https://www.proschoolonline.com/blog/data-science-skills

Thank You
Question & Answers

Email: baydetinagaraju@nitnagaland.ac.in
baydetinagaraju@gmail.com

Contact: +919986502452