**Data science**

**DAY 1**

1. **What is Data?** Data refers to facts, statistics, or information collected for reference or analysis. It can be in various forms such as **numbers, text, images, or multimedia**. Data can be structured, semi-structured, or unstructured.
2. **What is Data Analytics?** Data analytics is the process of examining large data sets to uncover hidden patterns, correlations, trends, and insights. It involves applying statistical and mathematical techniques to interpret data and make informed business decisions.
3. **What is Data Science and its Methodology?** Data science is an interdisciplinary field that combines various techniques from mathematics, statistics, computer science, and domain knowledge to extract insights and knowledge from data. Its methodology typically involves:
   * **Problem Formulation**: Defining the problem and objectives of the analysis.
   * **Data Collection**: Gathering relevant data from various sources.
   * **Data Cleaning and Preprocessing**: Removing inconsistencies, errors, and irrelevant information from the data.
   * **Exploratory Data Analysis (EDA)**: Exploring and visualizing the data to understand its characteristics.
   * **Feature Engineering**: Selecting, transforming, and creating features to improve model performance.
   * **Model Building and Evaluation**: Developing predictive or descriptive models and assessing their performance.
   * **Deployment and Monitoring**: Implementing the solution in a real-world setting and continuously monitoring its performance.
4. **Data Science Jobs** Data science jobs encompass a wide range of roles, including:
   * Data Scientist
   * Data Analyst
   * Machine Learning Engineer
   * Data Engineer
   * Business Intelligence Analyst
   * Data Architect
   * Data Science Manager These roles involve tasks such as data analysis, machine learning model development, data engineering, and providing actionable insights to stakeholders.
5. **Data Science Tools and Technologies** Data scientists use a variety of tools and technologies to perform their tasks, including:
   * Programming Languages: Python, R, SQL
   * Data Visualization Tools: Matplotlib, Seaborn, Tableau
   * Machine Learning Libraries: TensorFlow, PyTorch, Scikit-learn
   * Big Data Technologies: Hadoop, Spark, Hive
   * Data Wrangling Tools: Pandas, dplyr, SQLAlchemy
   * Cloud Platforms: AWS, Azure, Google Cloud Platform
   * Version Control Systems: Git, GitHub, GitLab

https://www.emathzone.com/tutorials/basic-statistics.html