



# *****Road Map*****

1. Programming Language
2. Data Structures and Algorithms 🡪 Problem solving skills.
   1. Array 🡪 Easy to Moderate level 🡪 LeetCode
   2. Strings 🡪 Easy to Moderate level 🡪 LeetCode
3. Mathematics 🡪 Linear Algebra, Calculus, Probability and Random Variables, Matrix Theory, Statistics
4. Database 🡪 SQL
5. Data Science Life cycle
   1. Data Acquisition
   2. Data Pipeline
   3. EDA (Exploratory Data Analysis)
      1. Null values
      2. Outliers
      3. Mean, median, mode.
      4. Categorical handling etc.…...
   4. Data Modeling 🡪 Applying..
      1. ML algorithms depending upon scenario.
         1. Supervised
         2. Unsupervised
      2. Deep Learning Algorithms (Tensor flow, Pytorch, etc.)
      3. NLP
   5. Model Evaluation 🡪 Confusion Matrix, Accuracy etc.
   6. Model Deployment
6. Data Science Projects
7. Prepare for Interview

Table Of Contents

[Road Map 1-2](#_Toc145182410)

[Chapter 1 Data 1-9](#_Toc145182411)

[A diagram of information and data

Description automatically generated 1-9](#_Toc145182412)

[Data? 1-9](#_Toc145182413)

[Categories of Data 1-10](#_Toc145182414)

[Quantitative data: 1-11](#_Toc145182415)

[Qualitative data: 1-11](#_Toc145182416)

[Data Examples 1-11](#_Toc145182417)

[Types of Data 1-11](#_Toc145182418)

[Chapter 2 Python 2-15](#_Toc145182419)

[Chapter 3 DSA 3-16](#_Toc145182420)

[ 3-16](#_Toc145182421)

[Algorithm Analysis 3-16](#_Toc145182422)

[Recursion 3-16](#_Toc145182423)

[Array Based Sequences 3-16](#_Toc145182424)

[Stacks, Queues and Deques 3-16](#_Toc145182425)

[Linked Lists 3-16](#_Toc145182426)

[Trees 3-16](#_Toc145182427)

[Priority Queues 3-16](#_Toc145182428)

[Maps, Hash Tables and Skip Lists 3-16](#_Toc145182429)

[Search 3-16](#_Toc145182430)

[Sorting and Selection 3-16](#_Toc145182431)

[Text Processing 3-16](#_Toc145182432)

[Graph Algorithms 3-16](#_Toc145182433)

[Memory Management and Binary Trees 3-16](#_Toc145182434)

[Chapter 4 Linear Algebra 4-18](#_Toc145182435)

[Chapter 5 Calculus 5-19](#_Toc145182436)

[Chapter 6 Statistics 6-20](#_Toc145182437)

[Chapter 7 Probability 7-22](#_Toc145182438)

[Chapter 8 Data Manipulation and Analysis 8-23](#_Toc145182439)

[Chapter 9 Machine Learning 9-24](#_Toc145182440)

[Chapter 10 Deep Learning 10-25](#_Toc145182441)

[Chapter 11 Natural Language Processing (NLP) 11-26](#_Toc145182442)

[Chapter 12 AI 12-27](#_Toc145182443)

[OOPS 12-28](#_Toc145182444)

[What is OOP? 12-28](#_Toc145182445)

[Why OOP? 12-28](#_Toc145182446)

[When OOP? 12-28](#_Toc145182447)

[How OOP? 12-28](#_Toc145182448)

[Where OOP? 12-28](#_Toc145182449)

[Goals, Principles, Design Pattern 12-28](#_Toc145182450)

[Chapter 13 Data 13-29](#_Toc145182451)

[Chapter 14 Machine Learning 14-30](#_Toc145182452)

Learning data science and AI is an exciting journey that requires a combination of theoretical understanding and hands-on practice. Here's a roadmap to help you get started:

\*\*1. **Foundations:**

* **Mathematics:** Brush up on linear algebra, calculus, and probability/statistics. These are essential for understanding the algorithms and concepts in data science and AI.

**2. Programming:**

* **Python:** Learn Python, a widely-used programming language in data science and AI. Familiarize yourself with libraries like NumPy, Pandas, and Matplotlib for data manipulation, analysis, and visualization.

**3. Data Manipulation and Analysis:**

* **Pandas:** Learn how to clean, manipulate, and analyze data using the Pandas library.
* **Data Visualization:** Learn Matplotlib and Seaborn for creating meaningful visualizations to understand your data.

**4. Machine Learning:**

* **Introduction to Machine Learning:** Understand the basics of supervised, unsupervised, and reinforcement learning.
* **Scikit-Learn:** Explore the Scikit-Learn library for implementing various machine learning algorithms.
* **Model Evaluation:** Learn about metrics, cross-validation, and hyperparameter tuning.
* **Feature Engineering:** Understand how to preprocess and engineer features for better model performance.

**5. Deep Learning:**

* **Neural Networks:** Learn the basics of artificial neural networks and their components.
* **Deep Learning Frameworks:** Choose a deep learning framework like TensorFlow or PyTorch and learn to build and train neural networks.
* **Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs):** Dive into advanced neural network architectures for image and sequence data.

**6. Data Preprocessing and Feature Engineering:**

* **Feature Scaling, Normalization:** Understand techniques to scale and normalize data for better model performance.
* **Handling Missing Data:** Learn strategies to handle missing values in datasets.
* **One-Hot Encoding, Label Encoding:** Explore methods to transform categorical data into numerical formats.

**7. Advanced Topics:**

* **Natural Language Processing (NLP):** Learn how to process and analyze text data using techniques like tokenization, sentiment analysis, and text generation.
* **Reinforcement Learning:** Dive into the field of reinforcement learning, which deals with training agents to make sequential decisions.
* **Time Series Analysis:** Understand techniques for analyzing and predicting time-dependent data.

**8. Projects:**

* Work on real-world projects to apply your knowledge and gain practical experience.
* Start with small projects and gradually work your way up to more complex ones.

**9. Online Resources:**

* Utilize online platforms like Coursera, edX, Udacity, and Khan Academy for structured courses on data science and AI.
* Follow blogs, YouTube channels, and forums to stay updated on the latest trends and techniques.

**10. Kaggle:**

* Join Kaggle, a platform for data science competitions. Participate in competitions to solve real-world problems and learn from others.

**11. Books:**

* Consider reading foundational books like "Python for Data Analysis" by Wes McKinney, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron, and "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

Remember that learning data science and AI is an ongoing process. The field is constantly evolving, so staying curious and open to new ideas is crucial. Practical experience and building a strong portfolio of projects will be invaluable as you progress in your journey.

youtube channels to learn datascience and ai

ChatGPT

There are numerous YouTube channels that offer high-quality tutorials and educational content on data science and AI. Here are some popular YouTube channels that you might find valuable for learning:

1. **Sentdex** - Provides tutorials on various topics including Python programming, machine learning, deep learning, and data analysis. Link: [Sentdex YouTube Channel](https://www.youtube.com/user/sentdex)
2. **3Blue1Brown** - Offers visually engaging videos that explain complex mathematical and AI concepts, making them easier to understand. Link: [3Blue1Brown YouTube Channel](https://www.youtube.com/c/3blue1brown)
3. **StatQuest with Josh Starmer** - Breaks down statistical concepts and machine learning algorithms using easy-to-understand animations. Link: [StatQuest YouTube Channel](https://www.youtube.com/user/joshstarmer)
4. **Krish Naik** - Covers a wide range of topics in data science, machine learning, and AI through tutorials and practical demonstrations. Link: [Krish Naik YouTube Channel](https://www.youtube.com/user/krishnaik06)
5. **Tech With Tim** - Offers tutorials on Python programming, data science, machine learning, and game development. Link: [Tech With Tim YouTube Channel](https://www.youtube.com/c/TechWithTim)
6. **Codebasics** - Provides tutorials on Python, machine learning, and data science, with a focus on practical examples. Link: [Codebasics YouTube Channel](https://www.youtube.com/c/codebasics)
7. **Data School** - Offers tutorials on Python, Pandas, machine learning, and data visualization. Link: [Data School YouTube Channel](https://www.youtube.com/user/dataschool)
8. **Siraj Raval** - Covers a variety of AI and machine learning topics, often discussing cutting-edge developments and concepts. Link: [Siraj Raval YouTube Channel](https://www.youtube.com/c/SirajRaval)
9. **AI Engineering** - Focuses on topics related to artificial intelligence, machine learning, and data science, with in-depth tutorials and explanations. Link: [AI Engineering YouTube Channel](https://www.youtube.com/c/AIEngineering)
10. **DataCamp** - Provides tutorials, courses, and webinars on data science and machine learning topics. Link: [DataCamp YouTube Channel](https://www.youtube.com/c/datacamp)

Top of Form

# *****Data*****

# A diagram of information and data Description automatically generated

***Data***

https://www.mathsisfun.com/data/index.html

Data comes in various forms such as numbers, words, etc. that describes discrete facts about an objective reality.

Example: BuyMore, consumer electronic goods, 12%, South India, 80%, North India

***Information:***

It comes from providing context to data. It is also the description of data. Information helps to understand patterns between factual data and give it a meaning. Information helps us to answers questions like who, when, what, where etc.

**Example:** BuyMore is a retail chain that has had done  12%  sales of consumer electronic goods in south India 80% sales in north India.

***Knowledge:***

It is synthesis of data and information leads us to answer the how question and take business decisions.

**Example:** BuyMore management got a knowledge that sales of consumer electronic goods is poor in south India compared to north India  and hence management of BuyMore retail store initiates some strategy to promote the products.

## ***Data?***

Data is defined as a collection of individual facts or statistics.

(While “datum” is technically the singular form of “data,” it’s not commonly used in everyday language.)

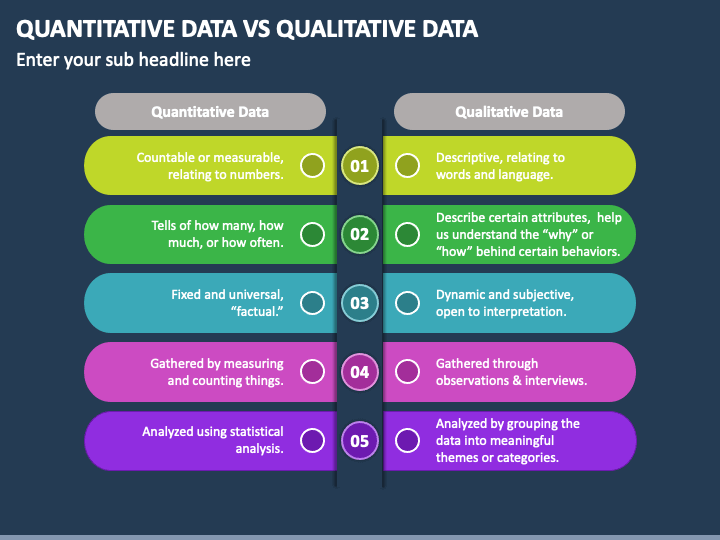
Data can be in the form of text, figures, images, numbers, graphs, or symbols. Data gathered by observations, measurements, research, or analysis.

Data is a raw form of knowledge and simple. It may even seem useless until it is analyzed, organized, and interpreted.

### ***Categories of Data***

Data can be broadly categorized as

* **Quantitative data** is provided in numerical form that can answer questions like “How many”, “How much”, “How often”. For example, the weight, volume, or cost of an item.
* **Qualitative data** is descriptive, but non-numerical, like the name, sex, or eye color of a person.



### *Quantitative data:*

**Quantitative data** can be Discrete or Continuous:

* **Discrete data** can only take certain values (like whole numbers)

Example: the number of students in a class

* **Continuous data** can take any value (within a range)

Example: A person's height: could be any value (within the range of human heights), not just certain fixed heights.

Time in a race: you could even measure it to fractions of a second,

simply: **Discrete data** is counted; **Continuous data** is measured.

Data are easily amenable (responsive)

### *Qualitative data:*

Qualitative data, also known as the categorical data, describes the data that fits into the categories. Qualitative data are not numerical. Categorical measures are defined in terms of natural language specifications, but not in terms of numbers. One of the examples is a grouped data. More precisely, categorical data could be derived from qualitative data analysis that are countable, or from quantitative data analysis grouped within given intervals.

### ***Data Examples***

* The number of visitors to a website in one month
* Inventory levels in a warehouse on a specific date
* Individual satisfaction scores on a customer service survey
* The price of a competitors’ product.Top of Form

### ***Types of Data***

* **Structured Data:**

Data that is organized in a tabular format, often found in databases. Each row corresponds to a record, and each column corresponds to a field.

* **Unstructured Data:**

Data that does not have a predefined structure, such as text documents, images, and videos.

* **Semi-Structured Data:**

Data that is not strictly organized like structured data but still has some level of structure, such as JSON or XML files.

Data can be categorized into various types based on its nature, structure, and characteristics. The main types of data include:

**\*\*1. Numerical Data:**

* Numerical data consists of numbers and is used for quantitative measurements.
* Two subtypes:
  + **Continuous Data**: Data that can take any value within a specific range. For example, height, weight, temperature.
  + **Discrete Data**: Data that can only take specific, distinct values. For example, the number of cars in a parking lot, the count of items sold.

**2. Categorical Data:**

* Categorical data represents categories or labels and is used for qualitative analysis.
* Two subtypes:
  + Nominal Data: Categories without any inherent order or ranking. For example, colors, types of animals.
  + Ordinal Data: Categories with a meaningful order or ranking. For example, educational levels (high school, bachelor's, master's).

**3. Text Data:**

* Text data consists of unstructured textual content, such as documents, articles, emails, and social media posts.
* Often used in natural language processing (NLP) tasks like sentiment analysis and text classification.

**4. Time Series Data:**

* Time series data is collected over successive time intervals. It's used for analyzing trends, patterns, and seasonality.
* Common in fields like finance, economics, and environmental monitoring.

**5. Image Data:**

* Image data represents visual information in the form of pixels.
* Used in computer vision tasks like image recognition, object detection, and image generation.

**6. Audio Data:**

* Audio data represents sound waves and is used in applications like speech recognition and music analysis.

**7. Geospatial Data:**

* Geospatial data includes information related to geographical locations, coordinates, and maps.
* Used in geographic information systems (GIS) and location-based services.

**8. Binary Data:**

* Binary data consists of only two possible values, often represented as 0s and 1s.
* Used in computer systems, cryptography, and digital communication.

**9. Mixed Data:**

* Mixed data involves a combination of different data types within a dataset. For instance, a dataset about customers might include numerical attributes (age), categorical attributes (gender), and text attributes (address).

Understanding the types of data is crucial for determining appropriate analysis methods, data preprocessing techniques, and modeling approaches. Different types of data require different methods for handling, visualizing, and extracting insights.

**Importance of Data**

Data serves as the foundation for making informed decisions, deriving insights, identifying trends, and developing models in various fields such as business, science, healthcare, and more.

**Data Collection**

The process of gathering data from various sources, which can include surveys, sensors, websites, databases, and more.

**Attributes or Features**

These are the individual characteristics or variables that make up a dataset. For example, in a dataset about customers, attributes could include name, age, and location.

SQL is a powerful language used for managing and manipulating relational databases. Below is a basic SQL course outline to get you started:

**1. Introduction to Databases:**

* Understand what databases are and why they are used.
* Learn about the differences between relational and non-relational databases.

**2. Basics of SQL:**

* Learn about SQL syntax, statements, and clauses.
* Explore SELECT statements to retrieve data from a database.

**3. Filtering and Sorting Data:**

* Use WHERE clause to filter data based on conditions.
* Learn to use ORDER BY clause to sort query results.

**4. Working with Functions:**

* Understand various SQL functions like COUNT, SUM, AVG, MAX, MIN, etc.
* Explore how to perform calculations on data.

**5. Joins and Relationships:**

* Learn about different types of joins (INNER, LEFT, RIGHT, FULL) to combine data from multiple tables.
* Understand primary keys and foreign keys for establishing relationships between tables.

**6. Grouping and Aggregating Data:**

* Use GROUP BY clause to group data based on specific columns.
* Learn about HAVING clause to filter grouped results.

**7. Subqueries and Derived Tables:**

* Explore subqueries to perform operations within queries.
* Understand derived tables (also known as subquery factoring or Common Table Expressions).

**8. Data Modification:**

* Learn how to INSERT, UPDATE, and DELETE data from tables.
* Understand the importance of data integrity and transaction management.

**9. Advanced SQL Topics:**

* Study concepts like indexes, views, stored procedures, and triggers.
* Explore techniques for optimizing SQL queries.

**10. Practical Projects:** - Apply SQL concepts to real-world scenarios by working on projects that involve data manipulation, analysis, and reporting.

# *****Python*****

**Aghsdnhg**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**Sd**

**D**

**Df**

**F**

**F**

**F**

**F**

**F**

**S**

**S**

# *****DSA*****

## 

<https://www.programiz.com/dsa>

https://github.com/manishbisht/Competitive-Programming/tree/master/Algorithms/Array

## ***Algorithm Analysis***

## ***Recursion***

## ***Array Based Sequences***

## ***Stacks, Queues and Deques***

## ***Linked Lists***

## ***Trees***

## ***Priority Queues***

## ***Maps, Hash Tables and Skip Lists***

## ***Search***

## ***Sorting and Selection***

## ***Text Processing***

## ***Graph Algorithms***

## ***Memory Management and Binary Trees***

**Aghsdnhg**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**Sd**

**D**

**Dg**

**G**

**G**

**G**

**G**

**G**

**Gg**

**G**

# *****Linear Algebra*****

**Aghsdnhg**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**Sd**

**D**

**Dg**

**G**

**G**

**G**

**G**

**G**

**Gg**

**G**

# *****Calculus*****

**Aghsdnhg**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**Sd**

**D**

**Dg**

**G**

**G**

**G**

**G**

**G**

**Gg**

**Gf**

# *****Statistics*****

Road Map

1. Descriptive type (part) 🡪 Summarizing, Analyzing, Visualizing
   1. Measure of Central Tendency
      1. Mean
      2. Median
      3. Mode
   2. Measure of Dispersion
      1. Range
      2. IQR
      3. Standard Deviation
      4. Variance
   3. Outliers
      1. Box Plot
      2. Histogram, etc.
   4. Distribution
      1. Normal /Gaussian Distribution Theorem
      2. Central Limit Theorem (Real time Examples also)
      3. Skewed/ Non skewed distribution.
      4. Poisson distributions
      5. Bernoulli/Binomial distributions
      6. Other Distributions
2. Inferential type
   1. Hypothesis Testing
      1. Null
      2. Alternative
      3. Null vs Alternative
   2. Errors
      1. Type 1 (Alpha)
      2. Type 2 (Beta)
   3. P-value
   4. Confidence Interval
3. Correlation Coefficient
4. Covariation
5. Z-Score (Standard normal Distribution)
6. Z-test
7. T-test
8. Anova test
9. Log normal Distribution.
10. Univariant, bivariant , multi variant

**Aghsdnhg**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**Sd**

**D**

**Dg**

**G**

**G**

**G**

**G**

**G**

**Gg**

**G**

# *****Probability*****

**Aghsdnhg**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**Sd**

**D**

**Dg**

**G**

**G**

**G**

**G**

**G**

**Gg**

**G**

# *****Data Manipulation and Analysis*****

**Aghsdnhg**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**Sd**

**D**

**Dg**

**G**

**G**

**G**

**G**

**G**

**Gg**

**G**

# *****Machine Learning*****

**Aghsdnhg**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**Sd**

**D**

**Dg**

**G**

**G**

**G**

**G**

**G**

**Gg**

**G**

# *****Deep Learning*****

**Aghsdnhg**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**Sd**

**D**

**Dg**

**G**

**G**

**G**

**G**

**G**

**Gg**

**G**

# *****Natural Language Processing (NLP)*****

**Aghsdnhg**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**A**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**S**

**Sd**

**D**

**Dg**

**G**

**G**

**G**

**G**

**G**

# *****AI*****

Road Map

1. Descriptive type (part) 🡪 Summarizing, Analyzing, Visualizing
   1. Measure of Central Tendency
      1. Mean
      2. Median
      3. Mode
   2. Measure of Dispersion
      1. Range
      2. IQR
      3. Standard Deviation
      4. Variance
   3. Outliers
      1. Box Plot
      2. Histogram, etc.
   4. Distribution
      1. Normal /Gaussian Distribution Theorem
      2. Central Limit Theorem (Real time Examples also)
      3. Skewed/ Non skewed distribution.
      4. Poisson distributions
      5. Bernoulli/Binomial distributions
      6. Other Distributions
2. Inferential type
   1. Hypothesis Testing
      1. Null
      2. Alternative
      3. Null vs Alternative
   2. Errors
      1. Type 1 (Alpha)
      2. Type 2 (Beta)
   3. P-value
   4. Confidence Interval
3. Correlation Coefficient
4. Covariation
5. Z-Score (Standard normal Distribution)
6. Z-test
7. T-test
8. Anova test
9. Log normal Distribution.
10. Univariant, bivariant , multi variant

***Road Map***

## ***OOPS***

<https://python-programming.quantecon.org/intro.html>

### ***What is OOP?***

Object-oriented Programming (OOPs) is a programming pattern that based on objects, which contain data in the form of fields and classes in programming.

### ***Why OOP?***

### ***When OOP?***

### ***How OOP?***

### ***Where OOP?***

### ***Goals, Principles, Design Pattern***

# *****Data*****

# *Machine Learning*