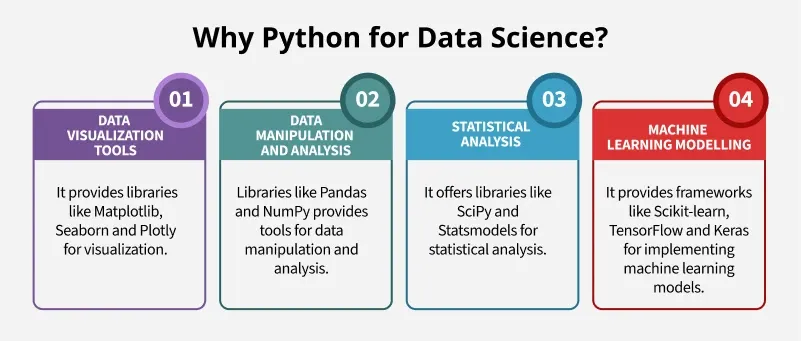
**Data science notes**

**Learn Data Science Tutorial With Python**

Data Science has become one of the fastest-growing fields in recent years, helping organizations to make informed decisions, solve problems and understand human behavior. As the volume of data grows so does the demand for skilled data scientists. The most common languages used for data science are Python and R.

In this ***Data Science with Python***tutorial will guide you through the fundamentals of both data science and Python programming.



**Python For Data Science Benefits**

In summary, Python is a popular language for data science because it is easy to learn, has a large and active community, offers powerful libraries for data analysis and visualization, and has excellent machine-learning libraries.

In terms of application areas, Data scientists prefer Python for the following modules:

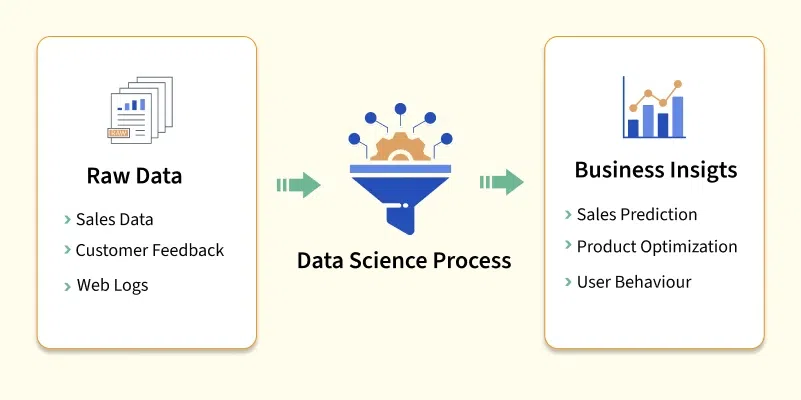
1. [Data Analysis](https://www.geeksforgeeks.org/data-analysis-with-python/)
2. [Data Visualizations](https://www.geeksforgeeks.org/python-data-visualization-tutorial/)
3. [Machine Learning](https://www.geeksforgeeks.org/machine-learning/)
4. [Deep Learning](https://www.geeksforgeeks.org/introduction-deep-learning/)
5. [Image processing](https://www.geeksforgeeks.org/image-processing-in-python-scaling-rotating-shifting-and-edge-detection/)
6. [Computer Vision](https://www.geeksforgeeks.org/computer-vision-introduction/)
7. [Natural Language Processing (NLP)](https://www.geeksforgeeks.org/introduction-to-natural-language-processing/)

**What is Data Science?**

Data science is the study of data that helps us derive useful insight for business decision making. Data Science is all about using tools, techniques, and creativity to uncover insights hidden within data. It combines math, computer science, and domain expertise to tackle real-world challenges in a variety of fields.

Data Science processes the raw data and solve business problems and even make prediction about the future trend or requirement. For example, from the huge raw data of a company, data science can help answer following question:

* What do customer want?
* How can we improve our services?
* What will the upcoming trend in sales?
* How much stock they need for upcoming festival.



In short, data science empowers the industries to make smarter, faster, and more informed decisions. In order to find patterns and achieve such insights, expertise in relevant domain is required. With expertise in Healthcare, a data scientists can predict patient risks and suggest personalized treatments.

**Data science involves these key steps:**

* **Data Collection:** Gathering raw data from various sources, such as databases, sensors, or user interactions.
* **Data Cleaning:**Ensuring the data is accurate, complete, and ready for analysis.
* **Data Analysis:**Applying statistical and computational methods to identify patterns, trends, or relationships.
* **Data Visualization:** Creating charts, graphs, and dashboards to present findings clearly.
* **Decision-Making:**Using insights to inform strategies, create solutions, or predict outcomes.

**Real Life Example of Data Science**

There are lot of examples you can observe around yourself, where data science is being used. For Example – Social Media, Medical, Preparing strategy for Cricket or FIFA by analyzing past matches. Here are some more real life examples:

**Social Media Recommendation:**

Have you ever wondered why you always get Instagram Reels aligned towards your interest? These platforms uses data-science to Analyze your past interest/data (Like, Comments, watch etc) and create personalized recommendation to serve content that matches your interests.

**Early Diagnosis of Disease:**

Data Science can predicts the risk of conditions like diabetes or heart disease, by analyzing a patient’s medical records and lifestyle habits. This allows doctors to act early and improve lives. In Future, it can help doctors detect diseases before symptoms even start to appear. For example, predicting a Tumor or Cancer at a very early stage. Data Science uses medical history and Image-data for such prediction.

**E-commerce recommendation and Demand Forecast:**

E-commerce platforms like Amazon or Flipkart use data science to enhance the shopping experience. By analyzing your browsing history, purchase behavior, and search patterns, they recommend products based on your preferences. It can also help in predicting demand for products by studying past sales trends, seasonal patterns etc.

**Applications of Data Science**

Data science has a wide range of applications across various industries, by transforming how they operate and deliver results. Here are some examples:

* Data science is used to analyze patient data, predict diseases, develop personalized treatments, and optimize hospital operations.
* It helps detect fraudulent transactions, manage risks, and provide personalized financial advice.
* Businesses use data science to understand customer behavior, recommend products, optimize inventory, and improve supply chains.
* Data science powers innovations like search engines, virtual assistants, and recommendation systems.
* It enables route optimization, traffic management, and predictive maintenance for vehicles.
* Data science helps in designing personalized learning experiences, tracking student performance, and improving administrative efficiency.
* Streaming platforms and content creators use data science to recommend shows, analyze viewer preferences, and optimize content delivery.
* Companies leverage data science to segment audiences, predict campaign outcomes, and personalize advertisements.

**Industry where data science is used**

Data science is transforming every industry by unlocking the power of data. Here are some key sectors where data science plays a vital role:

* **Healthcare:** Data science improves patient outcomes by using predictive analytics to detect diseases early, creating personalized treatment plans and optimizing hospital operations for efficiency.
* **Finance:** Data science helps detect fraudulent activities, assess and manage financial risks, and provide tailored financial solutions to customers.
* **Retail:** Data science enhances customer experiences by delivering targeted marketing campaigns, optimizing inventory management, and forecasting sales trends accurately.
* **Technology:** Data science powers cutting-edge AI applications such as voice assistants, intelligent search engines, and smart home devices.
* **Transportation:** Data science optimizes travel routes, manages vehicle fleets effectively, and enhances traffic management systems for smoother journeys.
* **Manufacturing:** Data science predicts potential equipment failures, streamlines supply chain processes, and improves production efficiency through data-driven decisions.
* **Energy:** Data science forecasts energy demand, optimizes energy consumption, and facilitates the integration of renewable energy resources.
* **Agriculture:** Data science drives precision farming practices by monitoring crop health, managing resources efficiently, and boosting agricultural yields.

**Important Data Science Skills**

Data Scientists need a mix of technical and soft skills to excel in this domain. To start with data science, it’s important to learn the basics like Mathematics and Basic programming skills. Here are some essential skills for a successful career in data science:

* **Programming:** Proficiency in programming languages like [Python](https://www.geeksforgeeks.org/python-programming-language-tutorial/), [R](https://www.geeksforgeeks.org/r-programming-for-data-science/), or SQL is crucial for analyzing and processing data effectively.
* **Statistics and Mathematics:** A strong foundation in statistics and linear algebra helps in understanding data patterns and building predictive models.
* [**Machine Learning**](https://www.geeksforgeeks.org/machine-learning/)**:** Knowledge of machine learning algorithms and frameworks is key to creating intelligent data-driven solutions.
* [**Data Visualization**](https://www.geeksforgeeks.org/data-visualization-with-python/)**:** The ability to present data insights through tools like Tableau, Power BI, or [Matplotlib](https://www.geeksforgeeks.org/matplotlib-tutorial/) ensures findings are clear and actionable.
* **Data Wrangling:** Skills in cleaning, transforming, and preparing raw data for analysis are vital for maintaining data quality.
* **Big Data Tools:** Familiarity with tools like Hadoop, Spark, or cloud platforms helps in handling large datasets efficiently.
* **Critical Thinking:** Analytical skills to interpret data and solve problems creatively are essential for uncovering actionable insights.
* **Communication:** The ability to explain complex data findings in simple terms to stakeholders is a valuable asset.

*Python and R language are widely used for data science. To learn data science effectively, we have curated step-wise guide for both:*

[***Data Science with Python***](https://www.geeksforgeeks.org/data-science-with-python-tutorial/)

[***Data Science With R***](https://www.geeksforgeeks.org/data-science-tutorial-with-r/)

**How to Become a Data Scientist?**

Data Science is a high demand career and opportunity in multiple growing industries. Let’s discuss some key steps to becoming a successful data scientists:

* **Learn Programming Skills:**Master essential programming languages like Python and R.
* **Build a Strong Foundation First:** Study statistics, mathematics, and data structures.
* **Start Machine Learning:** Learn algorithms, models, and frameworks for building AI solutions.
* **Data Visualization Skills:** Use tools like Tableau or Power BI to present insights effectively.
* **Gain Practical Experience along with Learning:**Work on projects, internships, or competitions to apply your knowledge.
* **NLP and Deep Learning:** These are very important, after you finish above areas.
* **Learn Big Data Tools:** Get familiar with Hadoop, Spark, and cloud computing platforms.
* **Stay Updated with Trends:** Follow the latest trends and advancements in the field of data science.
* **Network and Collaborate:** Join data science communities, attend meetups, and connect with professionals.

*Refer this – For detailed* [*Roadmap to become a Data Scientists*](https://www.geeksforgeeks.org/data-scientist-roadmap/)

**Jobs and Career in Data Science**

Here are some of the key data science job roles:

**1. Data Scientist**

**Responsibilities:** Analyzing large datasets, developing machine learning models, interpreting results, and providing insights to inform business decisions.

**Skills:** Proficiency in programming languages like Python or R, expertise in statistics and machine learning algorithms, data visualization skills, and domain knowledge in the relevant industry.

**2. Data Analyst**

**Responsibilities:** Collecting, cleaning, and analyzing data to identify trends, patterns, and insights. Often involves creating reports and dashboards to communicate findings to stakeholders.

**Skills:** Strong proficiency in SQL for data querying, experience with data visualization tools like Tableau or Power BI, basic statistical knowledge, and familiarity with Excel or Google Sheets.

**3. Machine Learning Engineer**

**Responsibilities:** Building and deploying machine learning models at scale, optimizing model performance, and integrating them into production systems.

**Skills:** Proficiency in programming languages like Python or Java, experience with machine learning frameworks like TensorFlow or PyTorch, knowledge of cloud platforms like AWS or Azure, and software engineering skills for developing scalable solutions.

**4. Data Engineer**

**Responsibilities:** Designing and building data pipelines to collect, transform, and store large volumes of data. Ensuring data quality, reliability, and scalability.

**Skills:** Expertise in database systems like SQL and NoSQL, proficiency in programming languages like Python or Java, experience with big data technologies like Hadoop or Spark, and knowledge of data warehousing concepts.

**5. Business Intelligence (BI) Analyst**

**Responsibilities:** Gathering requirements from business stakeholders, designing and developing BI reports and dashboards, and providing data-driven insights to support strategic decision-making.

**Skills:** Proficiency in BI tools like Tableau, Power BI, or Looker, strong SQL skills for data querying, understanding of data visualization principles, and ability to translate business needs into technical solutions.

**6. Data Architect**

**Responsibilities:** Designing the overall structure of data systems, including databases, data lakes, and data warehouses. Defining data models, schemas, and data governance policies.

**Skills:** Deep understanding of database technologies and architectures, experience with data modeling tools like ERWin or Visio, knowledge of data integration techniques, and familiarity with data security and compliance regulations

**Python Libraries for Data Analysis**

* [Numpy Tutorial](https://www.geeksforgeeks.org/numpy-tutorial/)
* [Pandas Tutorial](https://www.geeksforgeeks.org/pandas-tutorial/)
* Scipy Tutorial
* Statsmodel Tutorial
* SweetViz

**Python Libraries for Data Visualization:**

* [Matplotlib Tutorial](https://www.geeksforgeeks.org/matplotlib-tutorial/)
* [Seaborn Tutorial](https://www.geeksforgeeks.org/python-seaborn-tutorial/)
* [Plotly Tutorial](https://www.geeksforgeeks.org/python-plotly-tutorial/)

**Python Libraries for Image Processing:**

* [OpenCV Tutorial](https://www.geeksforgeeks.org/opencv-python-tutorial/)

**Machine Learning:**

* Sckit-Learn Tutorial
* [Deep Learning Tutorial:](https://www.geeksforgeeks.org/deep-learning-tutorial/)
  + Tensorflow Tutorial
  + PyTorch Tutorial

**Python Libraries for Data Science**

To gain expertise in data science, you need to have a strong foundation in the following libraries:

* [Pandas for Data Manipulation](https://www.geeksforgeeks.org/pandas-tutorial/)
* [NumPy for Numerical Computing](https://www.geeksforgeeks.org/python-numpy/)
* [Matplotlib for Data Visualization](https://www.geeksforgeeks.org/data-visualization-using-matplotlib/)
* [Seaborn for Data Visualization](https://www.geeksforgeeks.org/data-visualization-with-python-seaborn/)
* [Scikit-learn for Machine Learning](https://www.geeksforgeeks.org/learning-model-building-scikit-learn-python-machine-learning-library/)

**Data Loading**

Data loading means importing raw data from various sources and storing it in one place for further analysis.

* [Loading a CSV File into a DataFrame](https://www.geeksforgeeks.org/python-read-csv-using-pandas-read_csv/)
* [Loading Data from an Excel File](https://www.geeksforgeeks.org/how-to-import-an-excel-file-into-python-using-pandas/)
* [Loading Data from JSON File](https://www.geeksforgeeks.org/how-to-read-json-files-with-pandas/)
* [Loading Data from SQL Databases](https://www.geeksforgeeks.org/read-sql-database-table-into-a-pandas-dataframe-using-sqlalchemy/)
* [Web Scraping using BeautifulSoup to Scrape Data](https://www.geeksforgeeks.org/implementing-web-scraping-python-beautiful-soup/)
* [Loading Data from MongoDB into DataFrame](https://www.geeksforgeeks.org/convert-pymongo-cursor-to-dataframe/)

**Data Preprocessing**

Data preprocessing involves cleaning and transforming raw data into a usable format for accurate and reliable analysis.

* [What is Data Preprocessing?](https://www.geeksforgeeks.org/data-preprocessing-machine-learning-python/)
* [Working with Missing Data using Pandas](https://www.geeksforgeeks.org/working-with-missing-data-in-pandas/)
* [Removing Duplicates using drop\_duplicates()](https://www.geeksforgeeks.org/python-pandas-dataframe-drop_duplicates/)
* [Scaling and Normalization of Data](https://www.geeksforgeeks.org/ml-feature-scaling-part-2/)
* [Aggregating and Grouping Data](https://www.geeksforgeeks.org/pandas-groupby-summarising-aggregating-and-grouping-data-in-python/)
* [Feature Selection using Sklearn](https://www.geeksforgeeks.org/feature-selection-in-python-with-scikit-learn/)
* [Handling Categorical Data using Label Encoding](https://www.geeksforgeeks.org/ml-label-encoding-of-datasets-in-python/)
* [Handling Categorical Data using One-Hot Encoding](https://www.geeksforgeeks.org/ml-one-hot-encoding/)
* [Detecting outlier using Z score](https://www.geeksforgeeks.org/z-score-for-outlier-detection-python/)
* [Detecting outlier using Interquartile Range](https://www.geeksforgeeks.org/interquartile-range-to-detect-outliers-in-data/)
* [Handling Imbalanced Data](https://www.geeksforgeeks.org/handling-imbalanced-data-for-classification/)
* [Efficient Preprocessing for Large Datasets](https://www.geeksforgeeks.org/handling-large-datasets-in-pandas/)

**Data Analysis**

Data analysis is the process of inspecting data to discover meaningful insights and trends to make informed decision.

* [What is Data Processing?](https://www.geeksforgeeks.org/introduction-to-data-processing/)
* [Exploratory Data Analysis in Python](https://www.geeksforgeeks.org/exploratory-data-analysis-in-python/)
* [Univariate and Multivariate Analysis](https://www.geeksforgeeks.org/what-is-univariate-bivariate-multivariate-analysis-in-data-visualisation/)
* [Calculating Correlation](https://www.geeksforgeeks.org/python-pandas-dataframe-corr/)
* [Hypothesis testing using Python](https://www.geeksforgeeks.org/understanding-hypothesis-testing/)
* [One-sample t-test using Python](https://www.geeksforgeeks.org/how-to-conduct-a-one-sample-t-test-in-python/)
* [Two Sample t-test using Python](https://www.geeksforgeeks.org/how-to-conduct-a-two-sample-t-test-in-python/)
* [ANOVA (Analysis of Variance) in Python](https://www.geeksforgeeks.org/how-to-perform-a-one-way-anova-in-python/)
* [Mann-Whitney U Test in Python](https://www.geeksforgeeks.org/mann-and-whitney-u-test/)
* [Z-test in Python](https://www.geeksforgeeks.org/z-test/)
* [Chi-Square Test](https://www.geeksforgeeks.org/ml-chi-square-test-for-feature-selection/)
* [PCA with Python](https://www.geeksforgeeks.org/principal-component-analysis-with-python/)
* [Shapiro-Wilk Test in Python](https://www.geeksforgeeks.org/how-to-perform-a-shapiro-wilk-test-in-python/)
* [Wilcoxon Signed-Rank Test in Python](https://www.geeksforgeeks.org/how-to-conduct-a-wilcoxon-signed-rank-test-in-python/)

**Data Visualization**

Data visualization uses graphical representations such as charts and graphs to understand and interpret complex data.

**Data Visualization using Matplotlib**

* [Line Chart](https://www.geeksforgeeks.org/line-chart-in-matplotlib-python/)
* [Bar Plot](https://www.geeksforgeeks.org/bar-plot-in-matplotlib/)
* [Histogram](https://www.geeksforgeeks.org/matplotlib-pyplot-hist-in-python/)
* [Heatmap](https://www.geeksforgeeks.org/how-to-draw-2d-heatmap-using-matplotlib-in-python/)
* [Box Plot](https://www.geeksforgeeks.org/box-plot-in-python-using-matplotlib/)
* [Scatter Plot](https://www.geeksforgeeks.org/matplotlib-pyplot-scatter-in-python/)
* [Pie Chart](https://www.geeksforgeeks.org/plot-a-pie-chart-in-python-using-matplotlib/)
* [3D Plot](https://www.geeksforgeeks.org/three-dimensional-plotting-in-python-using-matplotlib/)

**Data Visualization using Seaborn**

* [Pair Plot](https://www.geeksforgeeks.org/data-visualization-with-pairplot-seaborn-and-pandas/)
* [Count Plot](https://www.geeksforgeeks.org/countplot-using-seaborn-in-python/)
* [Violin Plot](https://www.geeksforgeeks.org/violinplot-using-seaborn-in-python/)
* [Strip Plot](https://www.geeksforgeeks.org/stripplot-using-seaborn-in-python/)
* [KDE Plot (Kernel Density Estimate)](https://www.geeksforgeeks.org/kde-plot-visualization-with-pandas-and-seaborn/)
* [Joint Plot](https://www.geeksforgeeks.org/how-to-set-title-on-seaborn-jointplot-python/)
* [Reg Plot](https://www.geeksforgeeks.org/python-seaborn-regplot-method/)

**Interactive Visualization**

* [Scatter Plot](https://www.geeksforgeeks.org/scatter-plot-using-plotly-in-python/)
* [Bar Chart](https://www.geeksforgeeks.org/bar-chart-using-plotly-in-python/)
* [Line Chart](https://www.geeksforgeeks.org/line-chart-using-plotly-in-python/)
* [Animated Data Visualization](https://www.geeksforgeeks.org/animated-data-visualization-using-plotly-express/)
* [Choropleth Maps using](https://www.geeksforgeeks.org/choropleth-maps-using-plotly-in-python/)
* [Interactive Visualization using Bokeh](https://www.geeksforgeeks.org/python-bokeh-tutorial-interactive-data-visualization-with-bokeh/)
* [Visualizing Geospatial Data using Folium](https://www.geeksforgeeks.org/visualizing-geospatial-data-using-folium-in-python/)

**Machine Learning**

Machine learning focuses on developing algorithms that helps computers to learn from data and make predictions or decisions without explicit programming.

* [Machine Learning Tutorial](https://www.geeksforgeeks.org/machine-learning/)
* [Deep Learning Tutorial](https://www.geeksforgeeks.org/deep-learning-tutorial/)

**Data Science step1**

Introduction to data and its types

Continuous data and discrete data

Nominal and ordinal data

Data Processing: -

Reading data, handling missing data, categorical data, splitting data in training and testing, Normalize data

Web Scrapping:

What is Web Scrapping, Web Scrapping processing, search elements in different ways, how to use developers tools in browser

Exploratory data analysis: -

EDA of pima Indian diabetes dataset, visualize it,

Data Transformation and scaling data: -

Rescal data , standardize data, normalize data and binarize data

For predicting categorical output which machine learning types of algorithm used

Top of Form

In which type of machine learning technique output label has been given

* **supervised learning**
* **Unsupervised learning**

Bottom of Form

Top of Form

Predicting whether image contains cat dog or human face.

This type of problem can be solved with which type of machine learning system

* **Regression**
* **classification**

Bottom of Form

Top of Form

Predicting probability of given email is spam.

This type of problem can be solved with which type of machine learning system

* **Regression**
* **Classification**
* **Clustering**
* **Reinforcement learning**

Bottom of Form

Which system will you preferred to solve given problem below.

1. Predicting Cat or Dog

2. Find number is odd or even

3. Find shortest route between two node

4. Predict tomorrow stock price.

TS - Traditional system

ML - Machine Learning

Top of Form

Which type of machine learning I did not discuss.

* **Supervised learning**
* **Reinforcement learning**
* **Classification**
* **Unsupervised learning**
* **Regression**

Bottom of Form

<https://www.udemy.com/course/machine-learning-course-with-python/learn/lecture/16722374?start=0#overview>

**Statistics & probability**

