# Applied Data Science with Python

# Essentials of Python Programming

# Learning Objectives

By the end of this lesson, you will be able to:

- Examine why Python is a suitable program for data analytics
- Set up a Jupyter notebook
- Explore Python functions and the different types and sequences in Python
- Apply Python strings, discuss Lambda, and list comprehension in Python
- Recognize the different packages available in Python for data science, like NumPy, Pandas, SciPy, and statsmodels

# Business Scenario

ABC is a leading telecom provider that provides a wide range of services, including prepaid, postpaid, broadband, DTH, payment banks, and business solutions.

The company is currently struggling with a customer churn issue, which is a measure of the number of customers who discontinue using a product or service. Customer churn is a significant problem in the telecom or banking industries, and it is a vital metric for determining service quality, target demographics, and understanding why a customer might leave.

To address the problem, the organization has decided to utilize Python packages like Pandas and Seaborn to create informative graphs and analyze the characteristics of lost customers. They will also study the types of customers who remained and used certain products to gain insights.

# Lab Walkthrough

A lab walkthrough of Jupyter lab is given in the word document called Lab Guide.  It can be downloaded from the Reference Material section.

ASSISTED PRACTICE

# Assisted Practices

Let’s understand the topics below using Jupyter Notebooks.

- 3.3_Setting Up Jupyter Notebook
- 3.4_Python Functions
- 3.5_Python Data Types and Sequences
- 3.6_Python Strings
- 3.7_Reading and Writing CSV Files in Python

**Note**: Please download the pdf files for each topic mentioned above from the Reference Material section.

ASSISTED PRACTICE

# Date and Time in Python

# Discussion: Date and Time in Python

Duration: 10 minutes

- What is the date and time in Python?
- What are the six primary classes in the datetime module?

# Date and Time in Python

Import **datetime** module to work with date and time in Python.

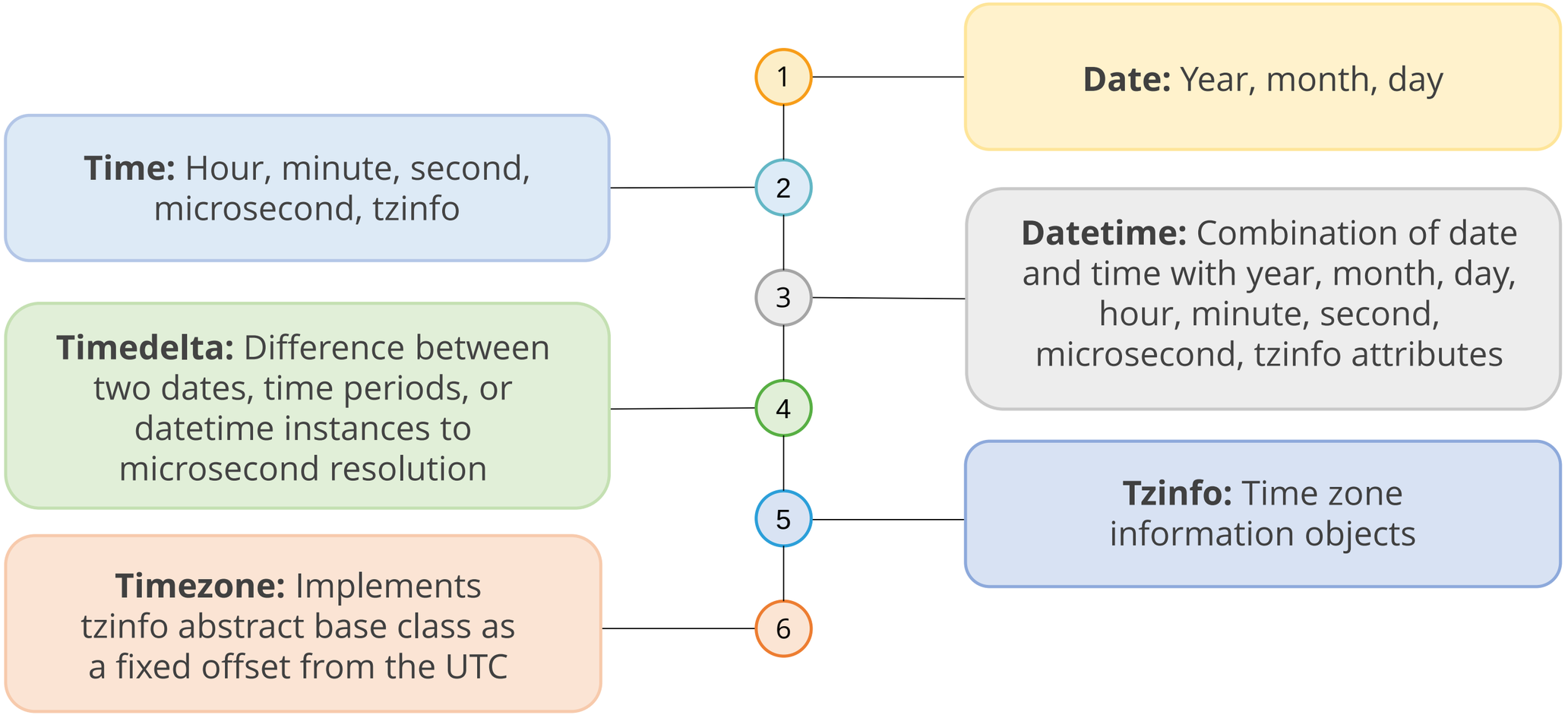
It is a standard built-in library in Python

It offers classes for manipulating date and time.

The functions in the classes can be used to perform operations on date and time.

# Date and Time in Python

There are six main classes in **datetime** module:



# Date and Time in Python

The Python code implementation of the **datetime** module can be seen below:

**Example 1:** Passing date as an argument

```
from datetime import date
date_argument = date(2021,12,24)
print("Example of passing date as an argument",
date_argument)
```

**Output:**

```
[11]: from datetime import date
      date_argument = date(2021,12,24)
      print('Example of passing date as an argument', date_argument)

      Example of passing date as an argument 2021-12-24
```

# Date and Time in Python

**Example 2:** Using the datetime module for getting today's date

```
from datetime import date
today_date = date.today()
print('Today date is', today_date)
```

**Output:**

```
Today date is 2023-05-02
```

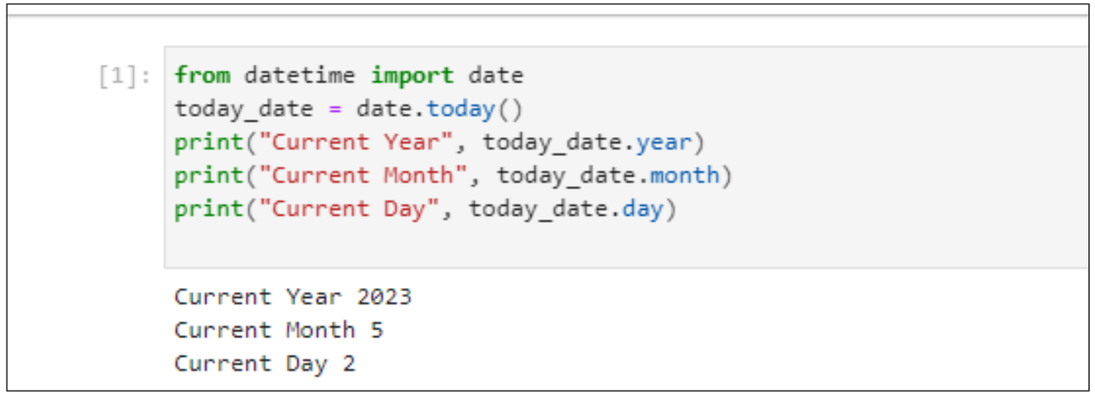
**YYYY-MM-DD** will be replaced by the actual date when the code is run.
The date format is year-month-day.

# Date and Time in Python

**Example 3:** Using the datetime module for getting the current year, month, and date

```
from datetime import date
today_date = date.today()
print("Current Year", today_date.year)
print("Current Month", today_date.month)
print("Current Day", today_date.day)
```

**Output:**




**YYYY, MM,** and **DD** will be replaced with the current year, month, and day, respectively, when the code is run.

Similarly, other classes from the **datetime** module can be used to perform basic operations on dates and time.

# Discussion: Date and Time in Python

Duration: 10 minutes

- What is the date and time in Python?

**Answer:** The date and time in Python refer to a standard built-in library that provides classes for manipulating and working with dates and times.

- What are the six primary classes in the datetime module?

**Answer:** The datetime module consists of six main classes: Time, Timedelta, Timezone, Date, Datetime, and Tzinfo.

# Assisted Practices

Let’s understand the topics below using Jupyter Notebooks

- 3.9_Python Objects and Map Function
- 3.10_Lambda and List Comprehension in Python

**Note**: Please download the pdf files for each topics mentioned above from the Reference Material section.

ASSISTED PRACTICE

# Python for Data Science

# Discussion: Python for Data Science

Duration: 15 minutes

- Why should Python be used for data science?
- What are the various Python packages for data science?

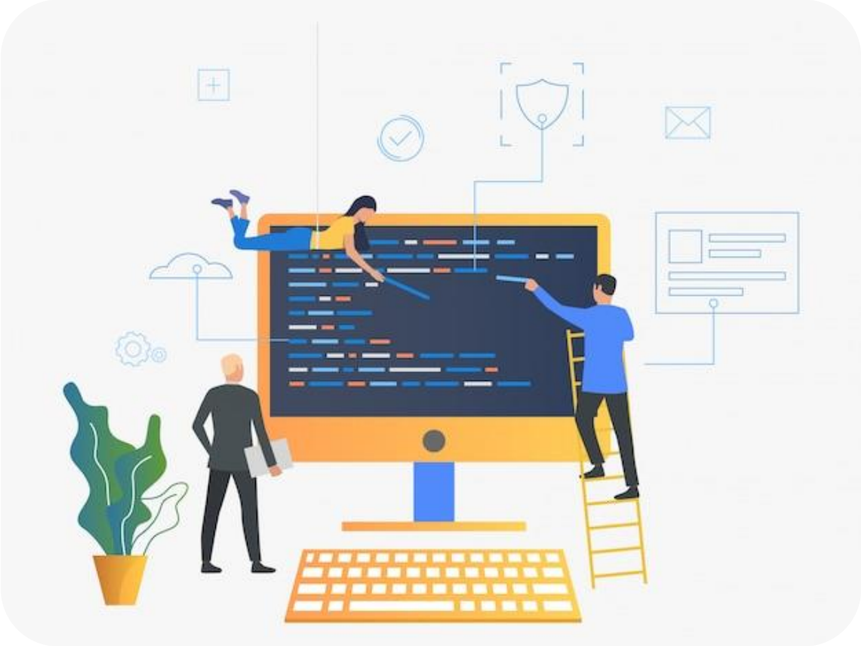
# Why Use Python for Data Science?

Python is the preferred programming language for data science projects across industries.


It has multiple open-source packages like NumPy and Pandas for data cleaning, exploration, and visualization.

# Why Use Python for Data Science?

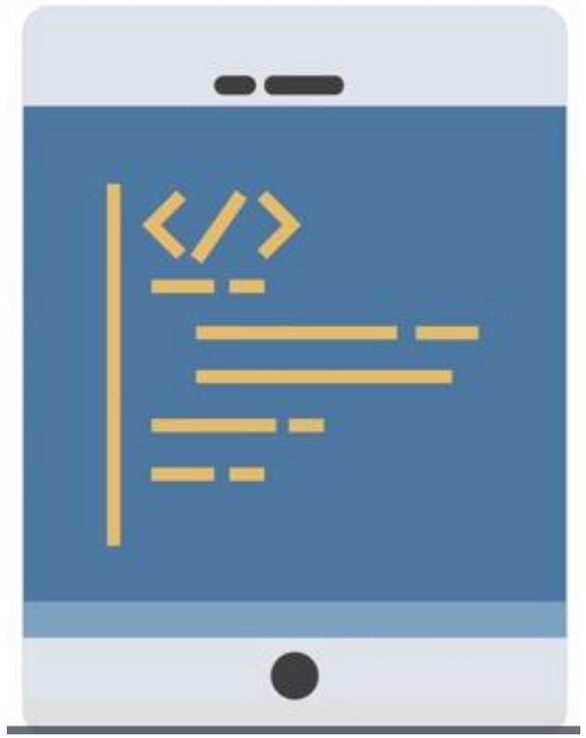
**Advantage 1:** Open-source, interpreted, high-level language that’s great for object-oriented programming.



One of the core components of Data Science is mathematics and statistics. Python is very effective when dealing with quantitative analysis.

# Why Use Python for Data Science?

**Advantage 2:** Ease of use and simple syntax



Suitable for users who are new to programming languages

Provides quick prototyping

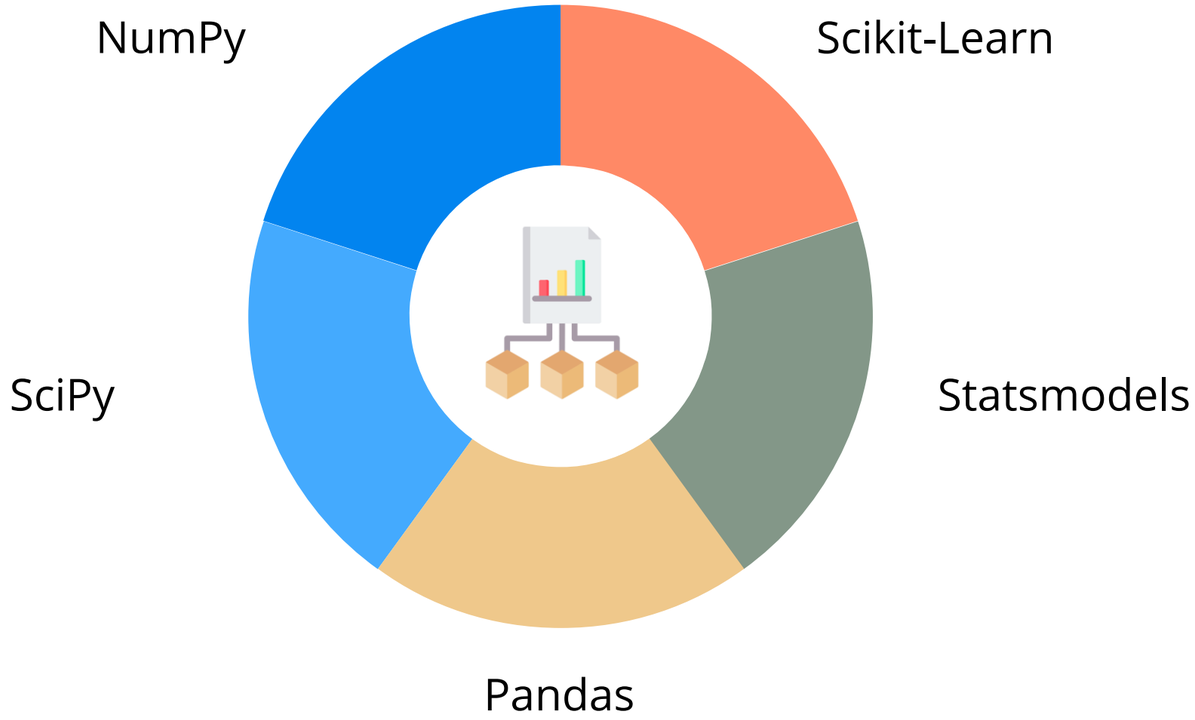
Has a shorter learning curve

# Why Use Python for Data Science?

**Advantage 3:** Scalability when compared to R

# Why Use Python for Data Science?

**Advantage 4:** Availability of a wide variety of data science libraries and packages



# Why Use Python for Data Science?

**Advantage 5:** Compatibility with all major operating systems

Windows

Mac


Unix


Linux

# Why Use Python for Data Science?

**Advantage 6:** A vast number of online user communities are creating new data science libraries daily.

A simple Google search can solve any programming challenge with Python.

# Why Use Python for Data Science?

**Advantage 7:** Powerful visualization libraries

Matplotlib, Seaborn, Pandas plotting, and ggplot can convey the insights gained from data science algorithms using charts, graphics, and other interactive visualization tools.

# Python Packages for Data Science

# Python Packages for Data Science

Python gives access to many libraries to perform quantitative and visualization tasks.

The most popular Python packages are:


Let us examine them in depth.

# NumPy

NumPy (Numerical Python) is an open-source library, predominantly used when working with arrays.

- It enables most of the operations required in linear algebra.
- It uses arrays instead of typical Python lists, which makes it more computationally efficient.
- It is used with SciPy and Matplotlib and has replaced Matlab as the industry standard for technical and engineering calculation.

# Pandas

Pandas is an open-source library built on top of NumPy and is used for data manipulation.


The word Pandas is derived from panel data, a term from econometrics.

It can be used with NumPy to analyze and manipulate data.

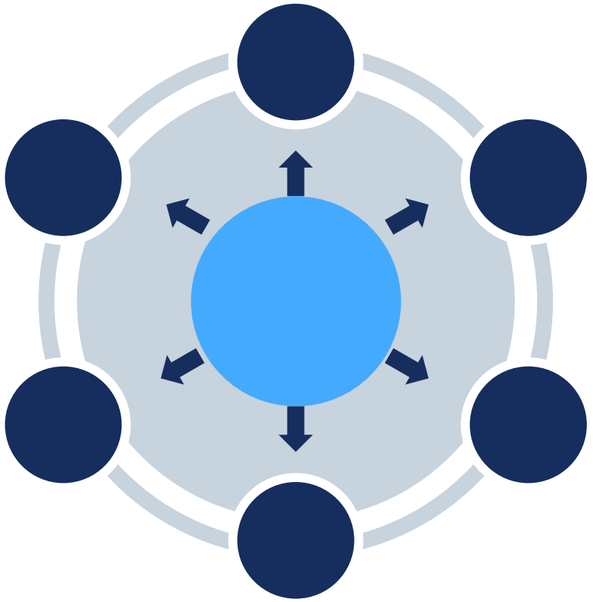
It allows working with tabular data, time series data, and matrix data.

# Pandas

Some of the areas that Pandas is best suited for are:

Importing data from CSV, JSON files, etc.

Handling time series data



Handling missing values in the dataset

Exporting results as CSV, JSON files, etc.

Data wrangling and manipulation

Merging multiple datasets

# SciPy

SciPy (Scientific Python) is an open-source library built on top of NumPy and is used for implementing scientific formulas.


It is tailored for scientific and engineering applications.

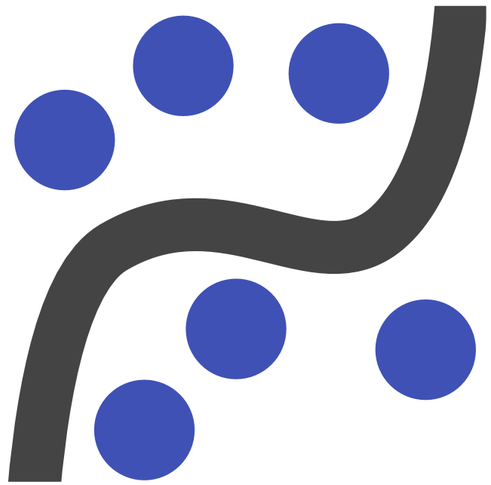
# SciPy

It has submodules for computationally intensive areas like:


Its fundamental data structure employs multidimensional arrays that are supported by the NumPy library.

# Statsmodels

Is an important statistical analysis library that:



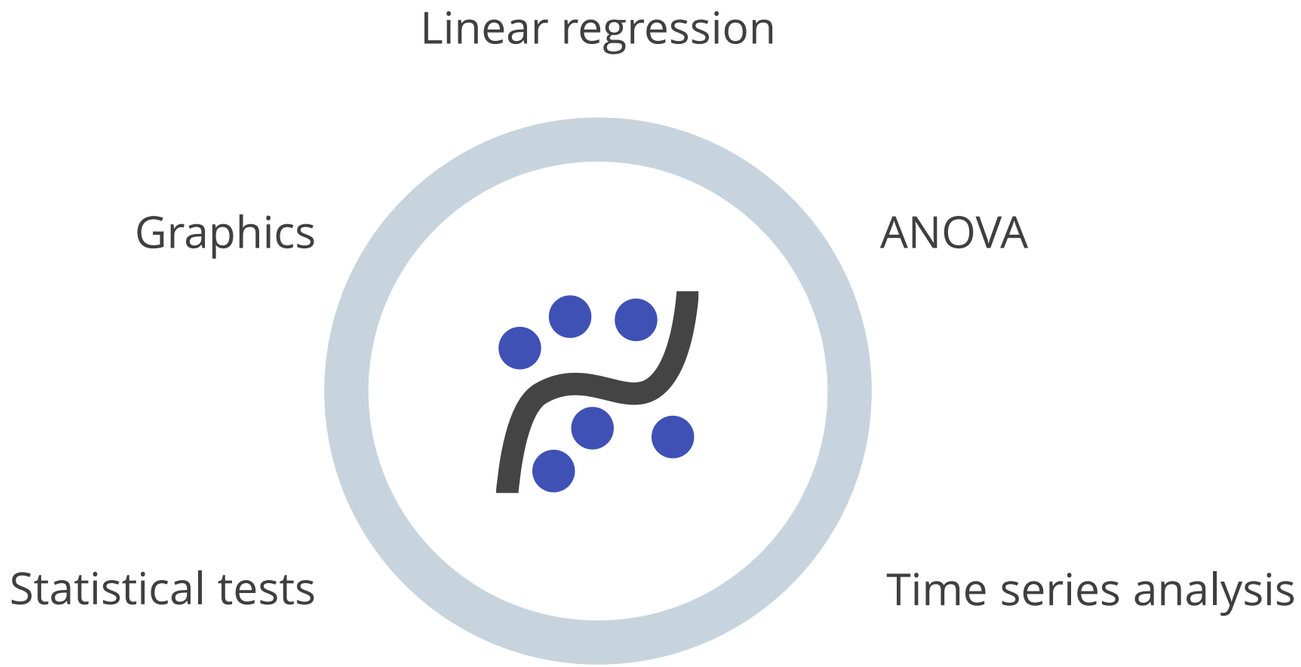
Allows the estimation of statistical models and performs statistical tests

Covers descriptive statistics, statistical tests, plotting functions, etc.

Is capable of handling deep statistical research projects

# Statsmodels

Some of the important features provided by statsmodels are:



# Discussion: Python for Data Science

Duration: 15 minutes

- Why should Python be used for data science?

**Answer:** Python is an open-source, interpreted, and high-level language well-suited for object-oriented programming, making it a valuable choice for data science tasks.

- What are the various Python packages for Data Science?

**Answer:** There are several Python packages commonly used in data science, including NumPy, Statsmodels, Pandas, and SciPy.

# Assisted Practices

Let’s understand the topics below using Jupyter Notebooks.

- 3.13_Introduction to Statsmodels API: Part A
- 3.14_Introduction to Statsmodels API: Part B

**Note**: Please download the pdf files for each topics mentioned above from the Reference Material section.

ASSISTED PRACTICE

# Key Takeaways

- Python is an open-source, interpreted, high-level language that's great for object-oriented programming.
- The **datetime** module is needed to work with date and time in Python.
- Python gives access to many libraries to perform quantitative and visualization tasks.
- NumPy works with arrays instead of traditional Python lists and is computationally efficient.
- Pandas is an open-source library built on top of NumPy.

# Key Takeaways

- SciPy uses a multidimensional array, provided by the NumPy library, as a basic data structure.
- Statsmodels is a crucial Python library that allows the estimation of statistical models and performs statistical tests.


# Knowledge Check

**Knowledge Check**

**1**

**Which of the following works with arrays instead of traditional Python lists?**

A. NumPy

B. Pandas

C. SciPy

D. Statsmodel

**Knowledge Check**

**1**

**Which of the following works with arrays instead of traditional Python lists?**

A. NumPy

B. Pandas

C. SciPy

D. Statsmodel

The correct answer is **A**

**NumPy works with arrays instead of traditional Python lists.**

**Knowledge Check 2**

**Which of the following uses a multidimensional array provided by the NumPy library as a basic data structure?**

A. Matplotlib

B. Pandas

C. SciPy

D. Statsmodel

**Knowledge Check 2**

**Which of the following uses a multidimensional array provided by the NumPy library as a basic data structure?**

A. Matplotlib

B. Pandas

C. SciPy

D. Statsmodel

The correct answer is **C**

**SciPy uses a multidimensional array provided by the NumPy library as a basic data structure.**

**Knowledge Check**

**3**

**Which Python library is used for data wrangling and manipulation?**

A. NumPy

B. Pandas

C. SciPy

D. Statsmodel

**Knowledge Check**

**3**

**Which Python library is used for data wrangling and manipulation?**

A. NumPy

B. Pandas

C. SciPy

D. Statsmodel

The correct answer is **B**

**Pandas is used for data wrangling and manipulation.**

# Thank You