



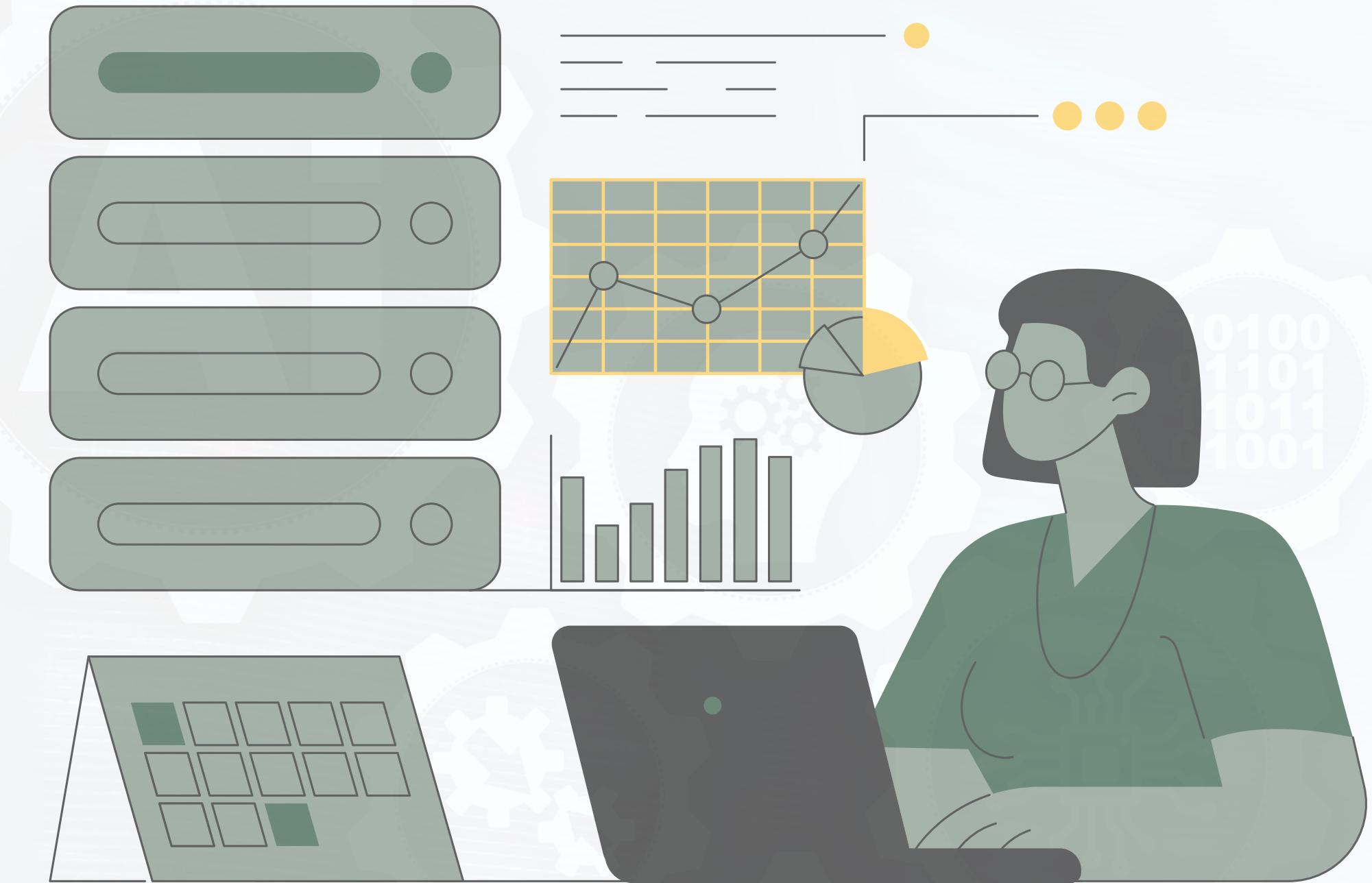
ពិភាក្សាលេខាគណ៍

SUNRISE INSTITUTE

DATA SCIENCE

EMBARKING ON A JOURNEY
INTO DATA SCIENCE

Instructor: YA MANON



You can have data without information but you
cannot have information without data.

-Daniel Keys Maran

COURSE STRUCTURE



This is a **project-based** course for students looking for a practical, hands-on, and highly engaging approach to learning, Data Science, Python essentials for data analysis.

Additional resources include:

-  **Source Code Link:** to serve as a helpful reference when you're offline or on the go
-  **Quizzes & Assignments** to test and reinforce key concepts, with step-by-step solutions
-  **Interactive demos** to keep you engaged and apply your skills throughout the course

SETTING EXPECTATIONS



Who this is for:

- Analysts or BI professionals looking to learn **Python for data analysis** and **Data Science**
- Students looking to learn the most popular open-source analytics tool (free)
- Anyone who wants to understand the core fundamentals of the Python language and syntax



Who this is NOT for:

- Experienced Python programmers or advanced users
- Students looking to learn Python for Software or Web Development
- Anyone who would rather copy and paste code or run packages without building the foundational skills



What is Data Science?

සිංහල සාකච්ඡා
SUNRISE INSTITUTE

What's Data Science?

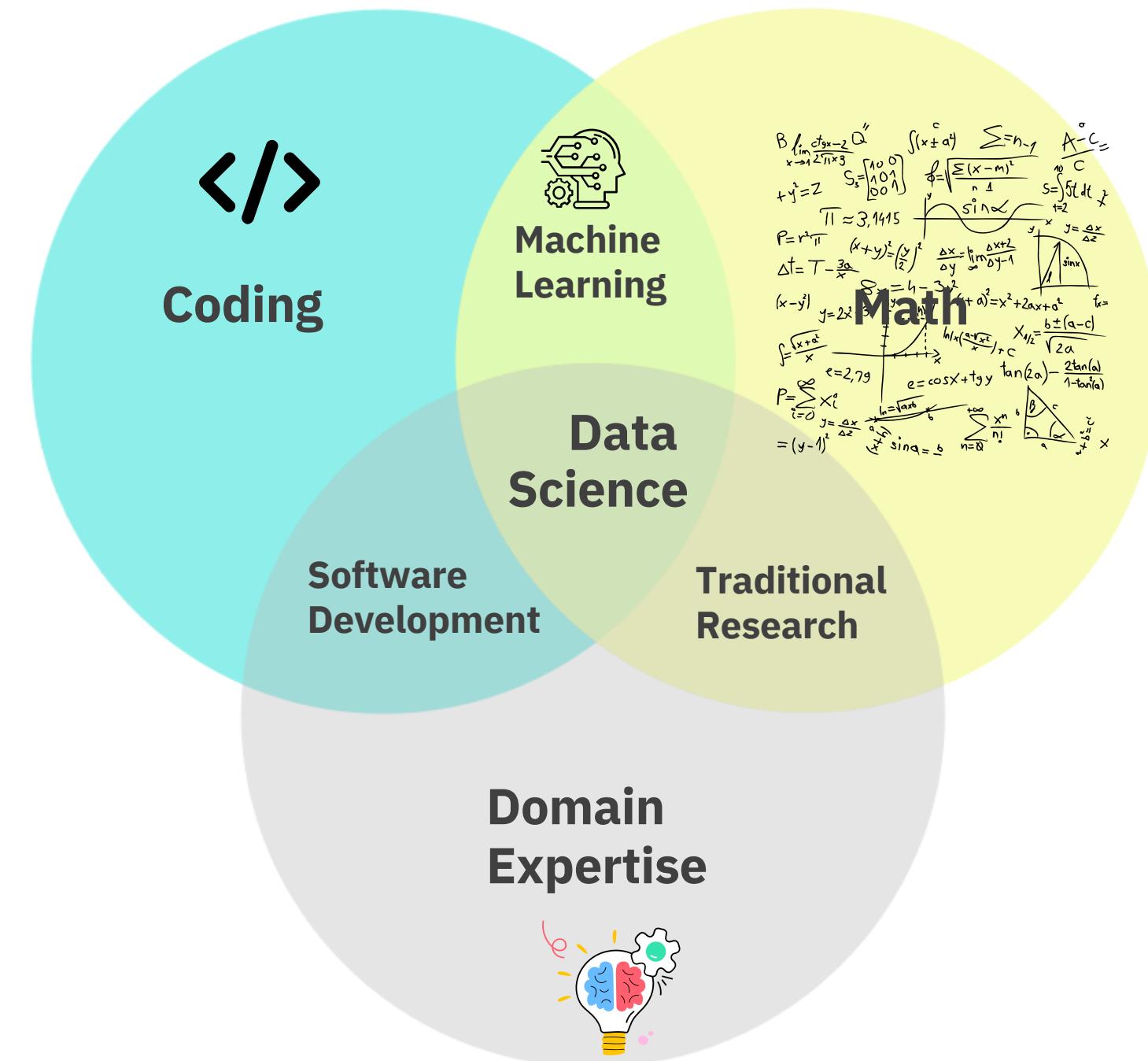
What is Data?

3202/9	1624	12040
3202/10	1624	12040
1420	220	12040
13242	627	470
355	51946	4703
462	4764981	4613433
51781	4311894	1321
54816	1845	13678
1729	5468778	190
680	655014	297
13	1682	51616
9	156453	1616
	1682	32027
	681	1808
31862	1808	15
1643	1616	
1643	1011	
1451	156984	5
846	808	

- **Data** information, experience, ...
- **Structures data:** Dimension (rows, columns), csv, excel , ..
 - **Example:** phone number, zip codes, identity number, gender,..
- **Unstructured data:** Everything else (no rows, columns)...
 - **Example:** emails, social media post, video, sensor, data...

Data Science

Data science is the study of data to extract meaningful insights for business.



The key is in applying these along with soft skills like:

- Communication
- Problem solving
- Curiosity & creativity
- Googling prowess



Data scientists & analysts approach problem solving in similar ways, but data scientists will often work with larger, more complex data sets and utilize advanced algorithms

Data Science

Data science is about *using data to make smart decisions.*



Real life Application

- Recommendation system (using machine learning , ai)
- Search engine (using machine learning, ai)





ත්‍රිජාත්‍යාල සෙවක ජෛව

SUNRISE INSTITUTE

INTRO PROGRAMMING

EMBARKING ON A JOURNEY
INTO DATA SCIENCE

Instructor: YA MANON

```
from sklearn.pipeline import Pipeline
from pathlib import Path
import os
import sys
import warnings
warnings.filterwarnings("ignore")
PACKAGE_ROOT = Path(os.path.abspath(os.path.dirname(__file__))).parent
sys.path.append(str(PACKAGE_ROOT))

from sklearn.linear_model import LogisticRegression

from building_prediction_model.config import config
import building_prediction_model.data_preprocessing.data_preprocessing as pp
import numpy as np
# print(config.DATASET)
classification_pipeline = Pipeline(
    steps = [
        ('Drop_columns', pp.DropColumns(variables_to_drop =['RowNumber','CustomerId','Surname'])),
        ('Encode_and_bind', pp.EncodeAndBind(encode='Gender', dummy='Geography')),
        ('Scale', pp.Scale(variables=
            ['CreditScore','Age','Tenure','Balance','NumOfProducts','EstimatedSalary'])),
        ('Model', LogisticRegression(random_state=12))])
```

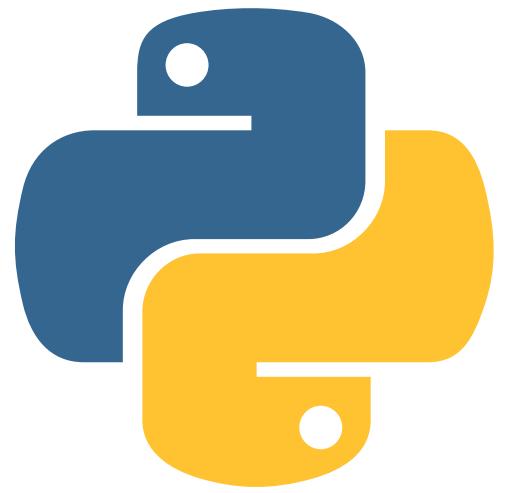
You can have data without information but you
cannot have information without data.

-Daniel Keys Maran



ເທິງ
SUNRISE
INSTITUTE

Python for Data Science





සිංහල සාකච්ඡා

SUNRISE INSTITUTE

WHY PYTHON?

MEET PYTHON

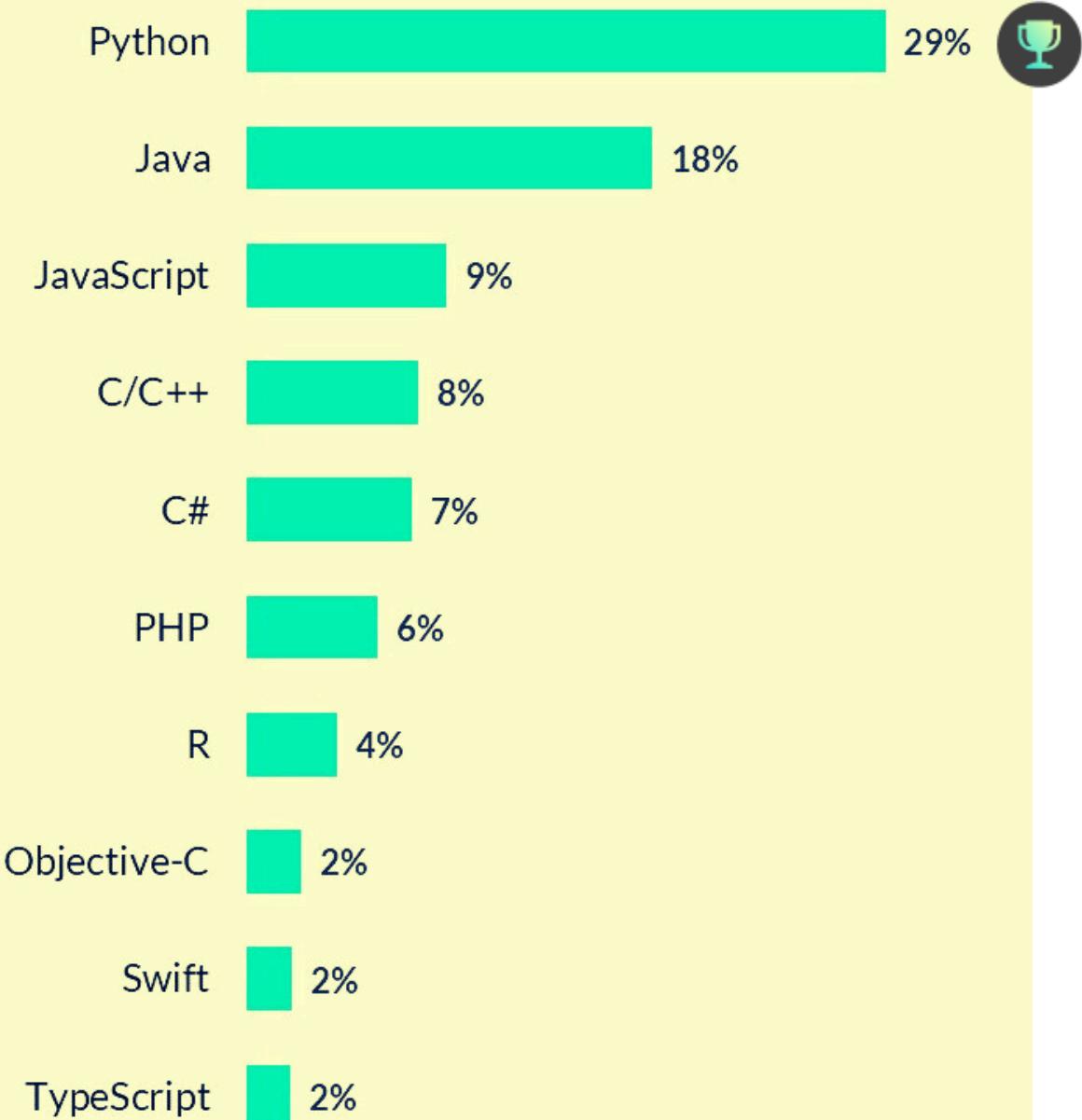


Python is a free, **open-source programming language** that is both powerful and easy to learn.

It has become one of the most popular languages in the world due to its accessibility, flexibility, ease of use, and wide range of applications, including:

- **Data Analytics**
- **Software Development**
- **Machine Learning**
- **Web Scraping**
- **Game Development And more!**

Popularity of Programming Language (PYPL):



Share of language tutorial searches on Google



COURSE OUTLINE

1

Why Python?

Introduce the Python analytics ecosystem and key reasons why it's the programming language of choice for many data analysts

2

Jupyter Notebook

Install Anaconda & create your first Jupyter Notebook, a user-friendly Python coding environment designed for data analysis and visualization

3

Data Types

Introduce native Python data types, common use cases, type conversion methods, and key concepts like iteration and mutability

4

Variables

Learn how to name and store values in memory using variables, as well as how to overwrite, delete and track them

5

Numerical Data

Learn how to work with numeric data, and use numeric functions to perform a range of arithmetic operations

6

String

Learn how to manipulate text via indexing and slicing, calculate string lengths, apply various string methods, and print f-strings to include variables



COURSE OUTLINE

7

Conditional Logic

Learn how to use IF statements and Boolean operators to establish conditional logic and control the flow of your programs

8

Sequence Data Types

Learn how to create, modify, and nest lists, tuples, and ranges, all of which allow you to store many values within a single variable

9

Loops

Understand the logic behind For and While loops and learn how to refine loop logic and handle common errors

10

Dictionaries and Sets

Address the limitations of working with lists and explore common scenarios for using dictionaries and sets in their place

11

Functions

Learn how to create custom functions in Python to boost productivity, and how to import external functions stored in modules or packages

12

Data Analytics with Pandas

Learn data manipulation and analysis, including one-dimensional data. Explore DataFrame and apply function for Data Analytics

COURSE OUTLINE

13

Intro Data Visualization

14

Matplotlib Basic

15

OO-ploting

16

Chart formatting

17

Chart types

18

Seaborn

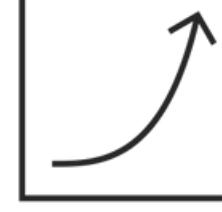
Understand how data visualization enhances for data visualization

Gain proficiency in using Matplotlib

Develop skills in interpreting and analyzing visualizations

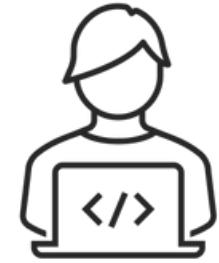
Matplotlib and Seaborn library and use it to build & customize several chart types, including line charts, bar charts, pie charts, scatterplots, and histograms

WHY PYTHON FOR ANALYTICS?



Scalability

Unlike most analytics tools or self-service Business Intelligence platforms, Python is **open source, free to use, and built for scale**



Versatility

With powerful libraries and frameworks, Python can add value at **every stage of the analytics workflow**, from data prep and analysis to machine learning and visualization



Community

Become part of a **large and active Python user community**, where you can share resources, get help, offer support, and connect with other users



Automation

Python can **automate complex tasks and workflows out of the box**, without complicated integrations or custom plug-ins



Demand

Python skills are **valuable and highly sought after**, and are becoming increasingly popular among analytics and Business Intelligence professionals



When it comes to data analytics, each tool has unique strengths and weaknesses; while Python shouldn't be the *only* tool in your stack, it can add tremendous value when combined with other tools like **Excel, SQL, Power BI & Tableau**

PYTHON ANALYTICS ECOSYSTEM

General Purpose Programming



Mastering **basePython** will give you a solid foundational understanding of the language, which is essential for using packages and libraries effectively

Data Manipulation & Analysis



Pandas helps us structure our data into dataset formats similar to that which you'd see in SQL or Excel. It also provides us with an arsenal of analytical functions that help us manipulate data and calculate the metrics we need to understand our data

Data Visualization



Matplotlib and **Seaborn** can create a wide array of visually appealing, static visualizations. **Plotly** can be used to create interactive visualizations and dynamic dashboards

Machine Learning



Scikit learn is among the most popular tools for building and testing machine learning models

Statsmodels provides a suite of tools for model building and statistical analysis

TensorFlow is the industry standard for developing deep learning models

DATA ROLES USING PYTHON

BI / DATA ANALYST

Data Analysts often use Python due to its cost effectiveness, or in collaboration with data science teams

Analysts are typically well-versed in base Python, Pandas, and at least one visualization library

→ **19%** of data analyst jobs require Python skills (**25%** in California)

DATA ENGINEER

Data Engineers commonly use Python to automate complex ETL processes or interact with APIs.

DBAs often use Pandas to manipulate data before storing it in a database or data warehouse.

→ **72%** of data engineer jobs require Python skills

DATA VIZ SPECIALIST

Data Visualization Specialists may use Python to design custom visuals that standard templates can't support.

They often utilize integrations with tools like Power BI or Tableau.

→ **33%** of data viz roles mention Python as a required or desired skill



DATA SCIENTIST

Data Scientists are most likely to use the 'full stack' of Python data tools. They leverage packages like Pandas, Scikit learn, Statsmodels & TensorFlow to build and deploy ML models.

→ **71%** of data science and machine learning jobs require Python skills