

# Python Ecosystem for Data Scientist

---

WEEK-1

# Outline

---

Demystify “Data Scientist”

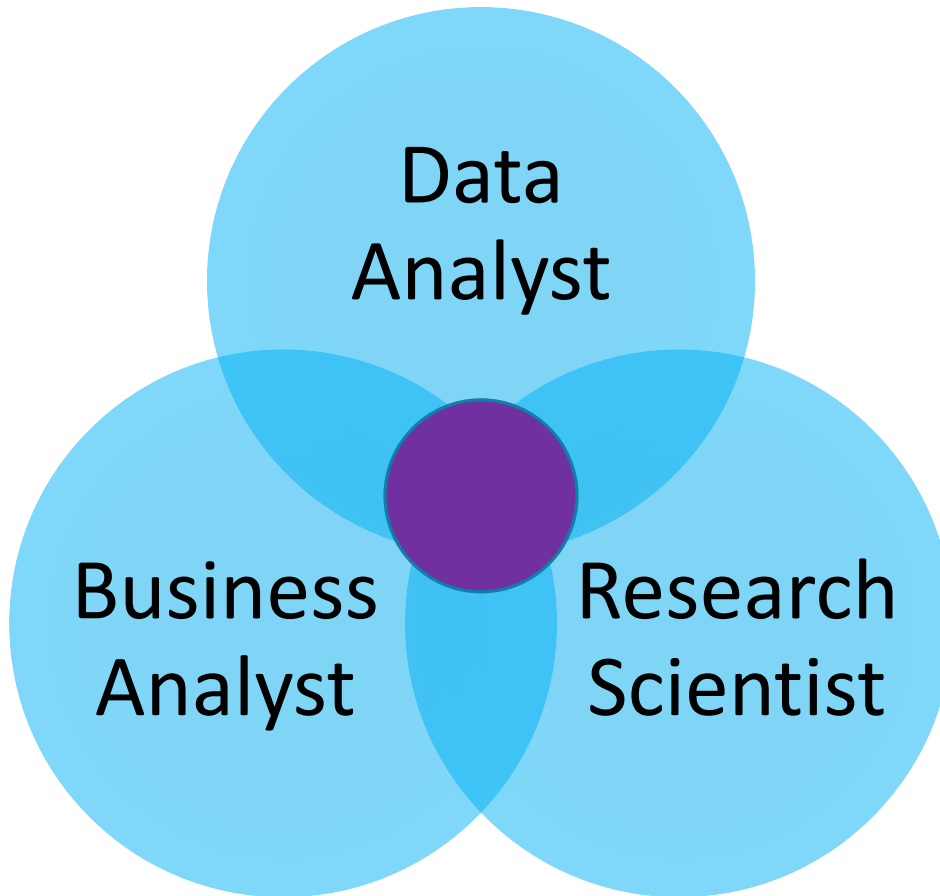
Python key packages

Pandas – data analytics

Fast visualization in Pandas

# What is Data Scientist?

---



Analytical skills

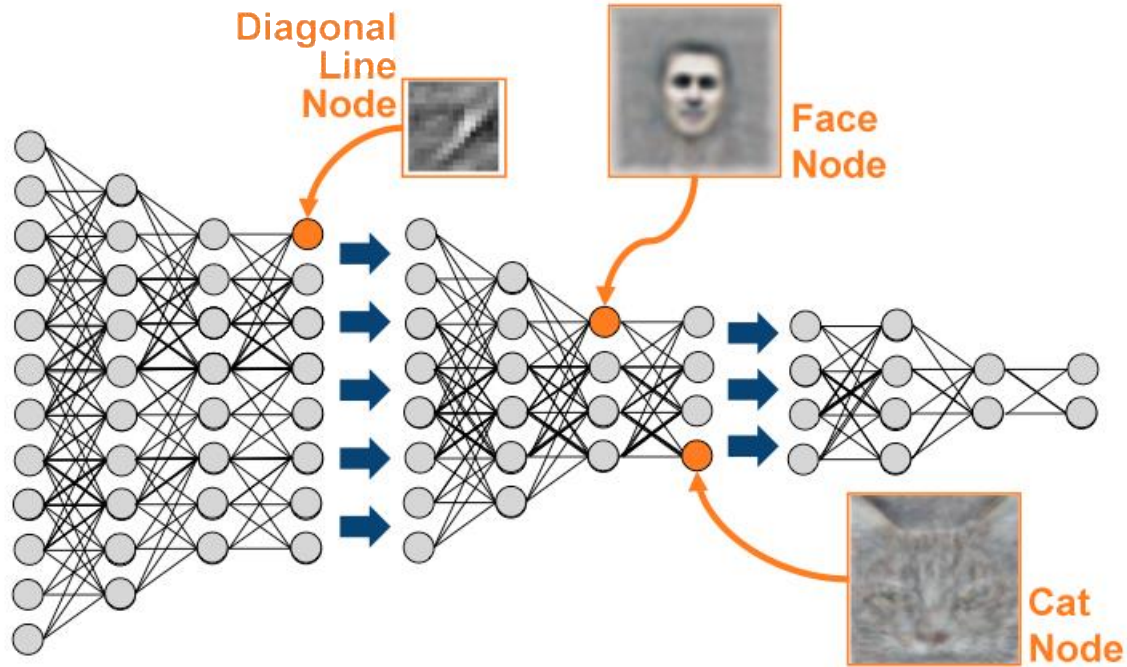
Design algorithm

Business insight

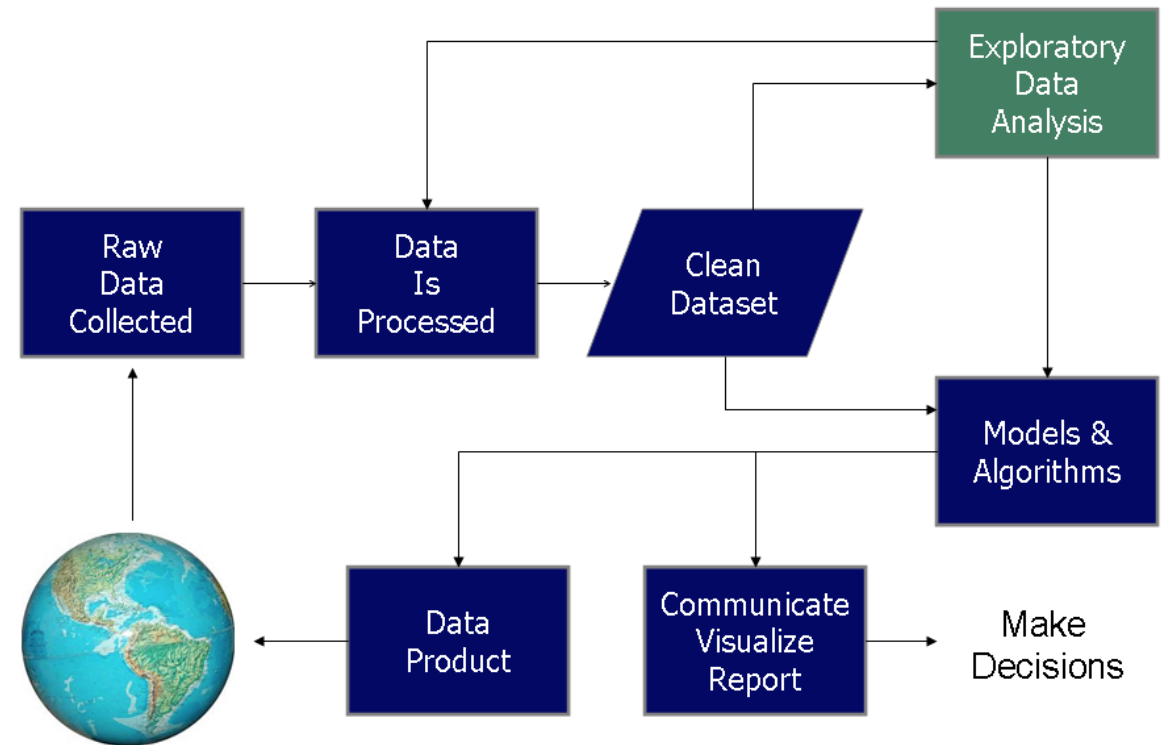
# Life as a Data Scientist

“Reality”

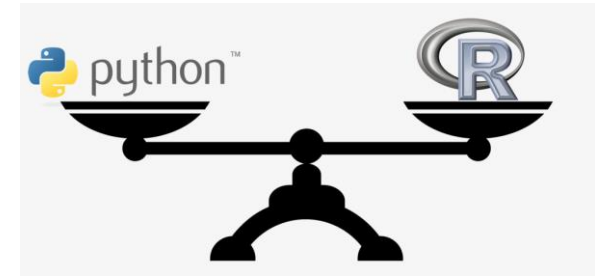
“Ideal”



Data Science Process



# Why Python?



General purpose programming language

Purpose	
R focuses on better, user friendly data analysis, statistics and graphical models.	Python emphasizes productivity and code readability.
Used By?	
R has been used primarily in academics and research. However, R is rapidly expanding into the enterprise market.  <i>"The closer you are to statistics, research and data science, the more you might prefer R."</i>	Python is used by programmers that want to delve into data analysis or apply statistical techniques, and by developers that turn to data science.  <i>"The closer you are to working in an engineering environment, the more you might prefer Python."</i>

# Pandas: Panel Data System

---

Rich data structures and functions to work with structured data fast, easy and expressive

Build on top of NumPy

Ideal tool for: munging/cleaning/analyzing/modeling data

**PANel DAta S**

