Abdullah Kise

Çözüm Mimarı
abdullahkise@hotmail.com

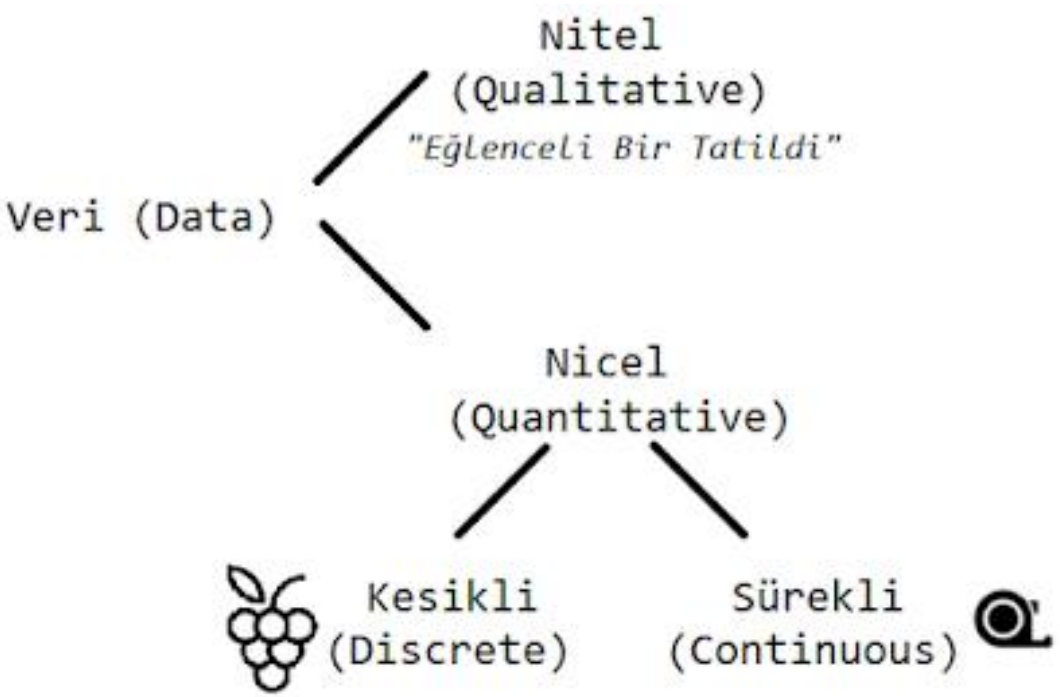
Microsoft Certified TRAINER

MVP Microsoft Most Valuable Professional

Python ile Veri Analizi

Bizi Takip Edin

Data ...



Yığın Veri
(Batch Data)

Akan Veri
(Stream Data)

★Structured (table) ★Semi-structured (csv, xml, json, yaml ...) ★Unstructured (image, text, video, sounds ...)

Ad	Renk	Fiyat
Elma	Kırmızı	10
Ayva	Sarı	5

```
{
  "id": 1,
  "name": "Foo",
  "price": 123,
  "tags": [
    "Bar",
    "Eek"
  ],
  "stock": {
    "warehouse": 300,
    "retail": 20
  }
}
```

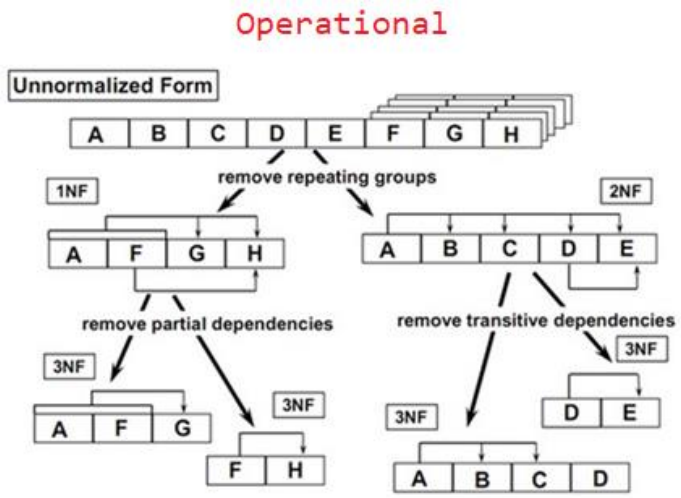
JSON

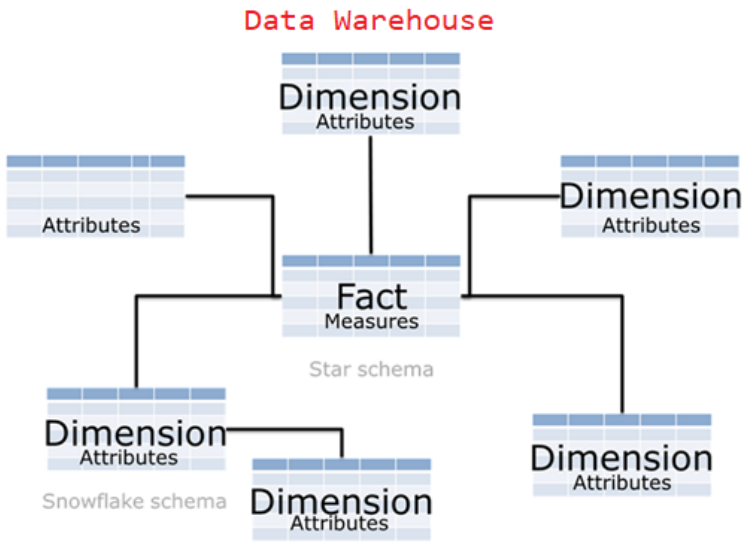
```
<?xml version="1.0" encoding="UTF-8"?>
<note>
  <to> Tove</to>
 <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this weekend!</body>
</note>
```

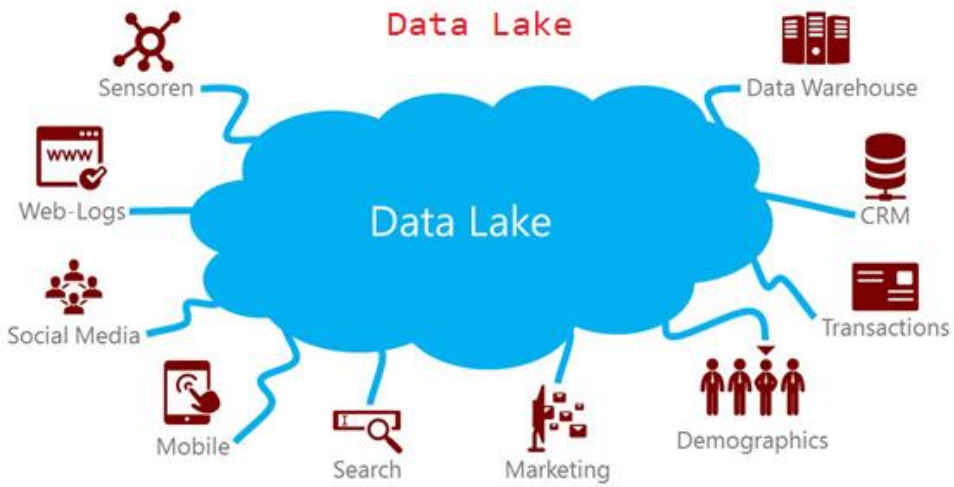
XML

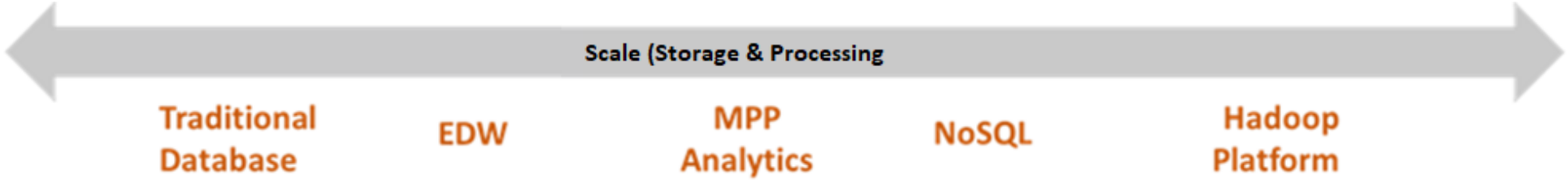
Data Management Evolution

+Lakehouse









Data Analysis Process

Capture

Integrate

Organize

Analyze

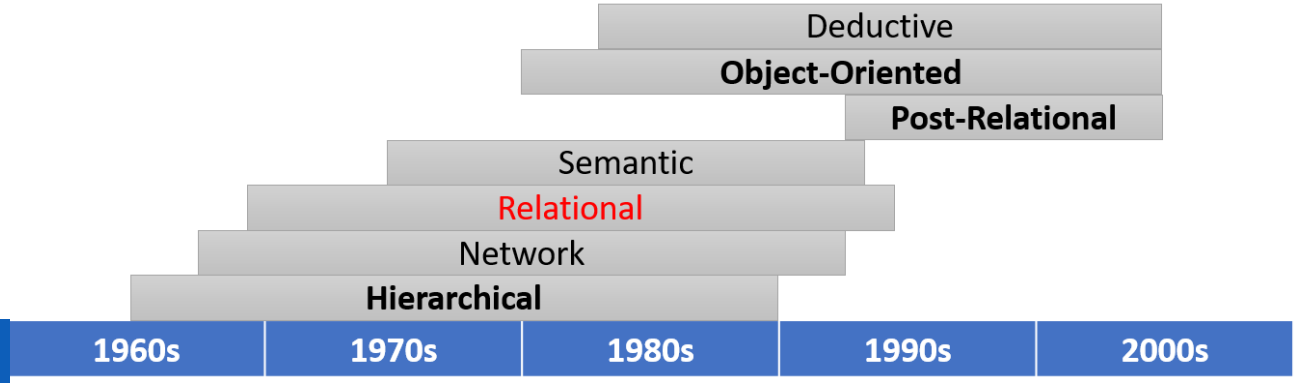
Act

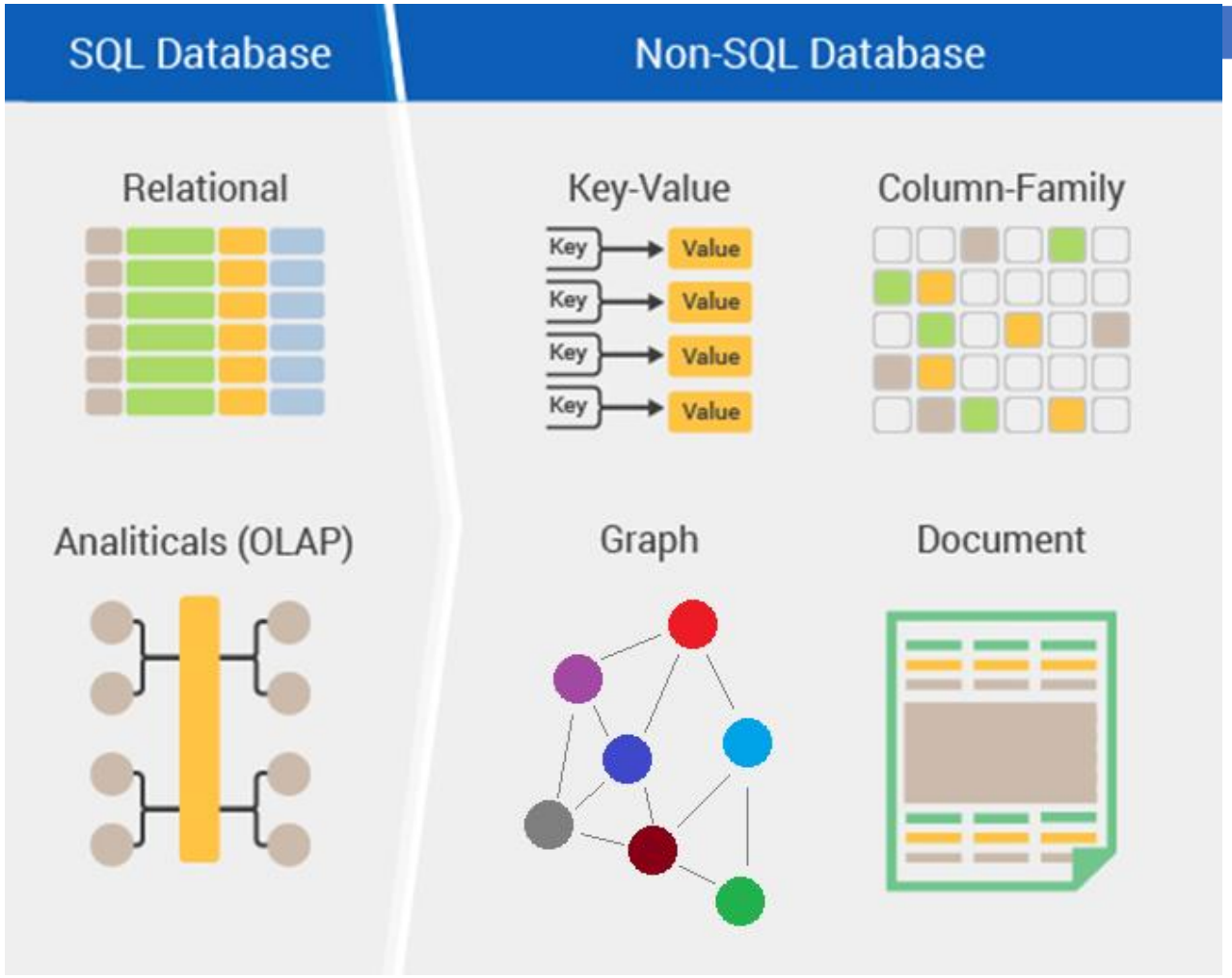
Large and Diverse

Too Fast

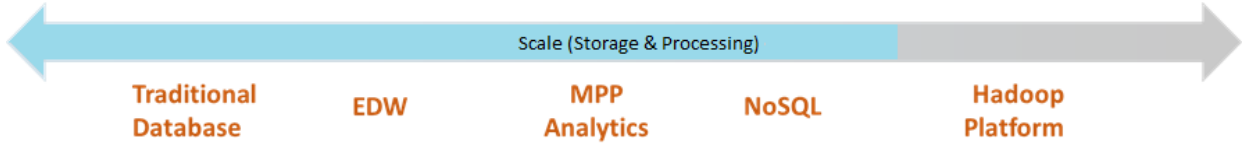
Very Mysterious

Large and Diverse





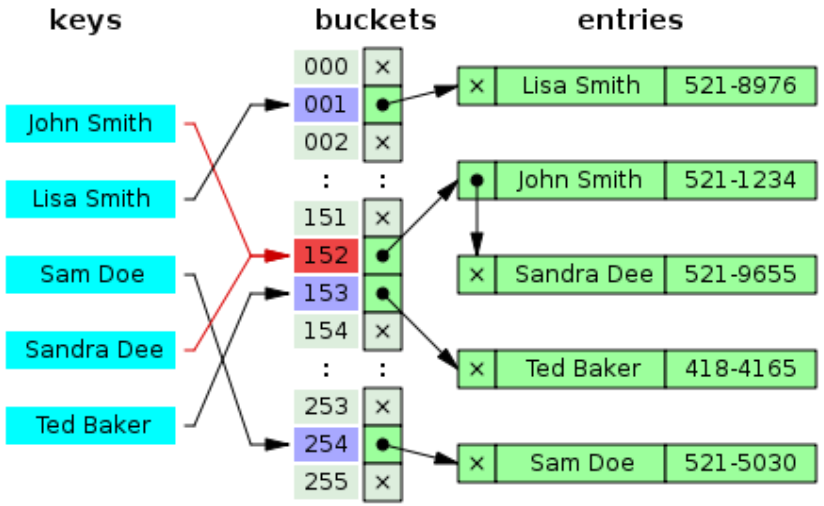
- **RDBMS**
  - ✓ MSSQL, ORACLE, MYSQL ...
- **NoSQL**
  - ✓ Cosmos DB, Riak, Redis, Mongo DB ...
- **NewSQL**
  - ✓ Clustrix, VoltDB, MemSQL ...
- **Storage**
  - ✓ On-prem, Azure Storages, Amazon S3 …



NoSQL Types

Key-Value Store
(***Azure Tables***, *Riak, Redis*)

User session data,
Preferences,
Shopping cart data



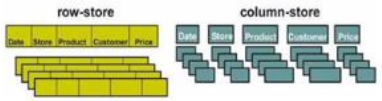
Column-based Store
(***Azure HDInsight Hbase***, *Cassandra, Google BigTable*)

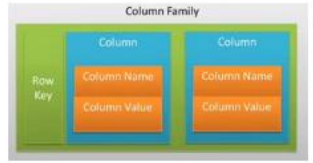
Content management,
blogging,
heavy write request (log)

```
{
  Fuar : {
      Adres: {
          Sehir: istanbul
          postakodu:34033
          },
      Detay: {
          Kapasite: 100
          Kalite: 5
          }
}
```

Key
Column Families
Columns

Sehir	Postakodu	Kapasite	Kalite
İstanbul	34033	100	5





Document-based Store
(***Azure DocumentDB***, *Mongo DB*)

Content management,
Blogging,
E-commerce

```
receipt:     Oz-Ware Purchase Invoice
date:        2012-08-06
customer:
    given:   Dorothy
    family:  Gale

items:
    - part_no:   A4786
      descrip:   Water Bucket (Filled)
      price:     1.47
      quantity:  4

    - part_no:   E1628
      descrip:   High Heeled "Ruby" Slippers
      size:      8
      price:     100.27
      quantity:  1
```
YAML

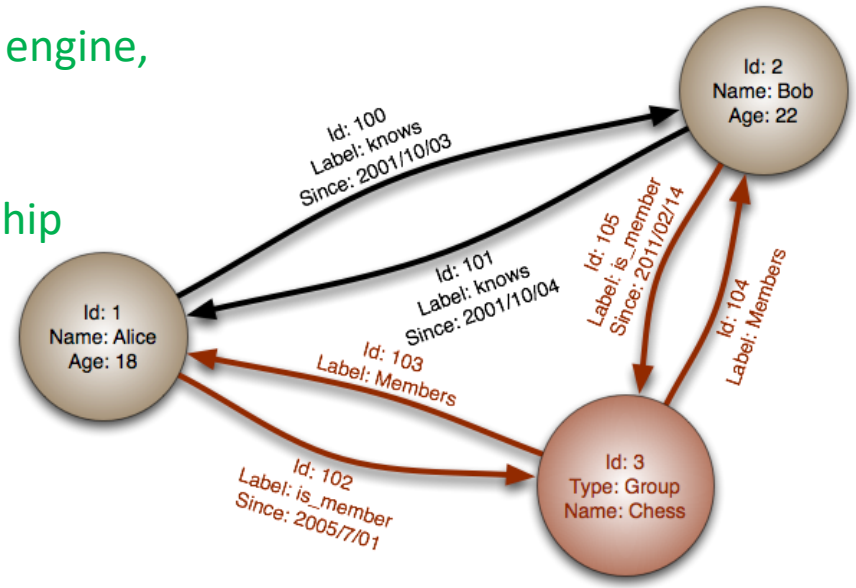
```
{
  "id": 1,
  "name": "Foo",
  "price": 123,
  "tags": [
    "Bar",
    "Eek"
  ],
  "stock": {
    "warehouse": 300,
    "retail": 20
  }
}
```
JSON

```
<?xml version="1.0" encoding="UTF-8"?>
<note>
  <to> Tove</to>
 <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this weekend!</body>
</note>
```
XML

```
Json: {"hello": "world"}

Bson:
  \x16\x00\x00\x00                  // total document size
  \x02                              // 0x02 = type String
  hello\x00                         // field name
  \x06\x00\x00\x00world\x00         // field value
  \x00                              // 0x00 = type EOO ('end of object')
```
BSON

Graph-based Store
(Neo4J,InfoGrid)

Social networks,
Recommendation engine,
Spatial data,
Shortest path,
n degree relationship



Too Fast

Streaming Data

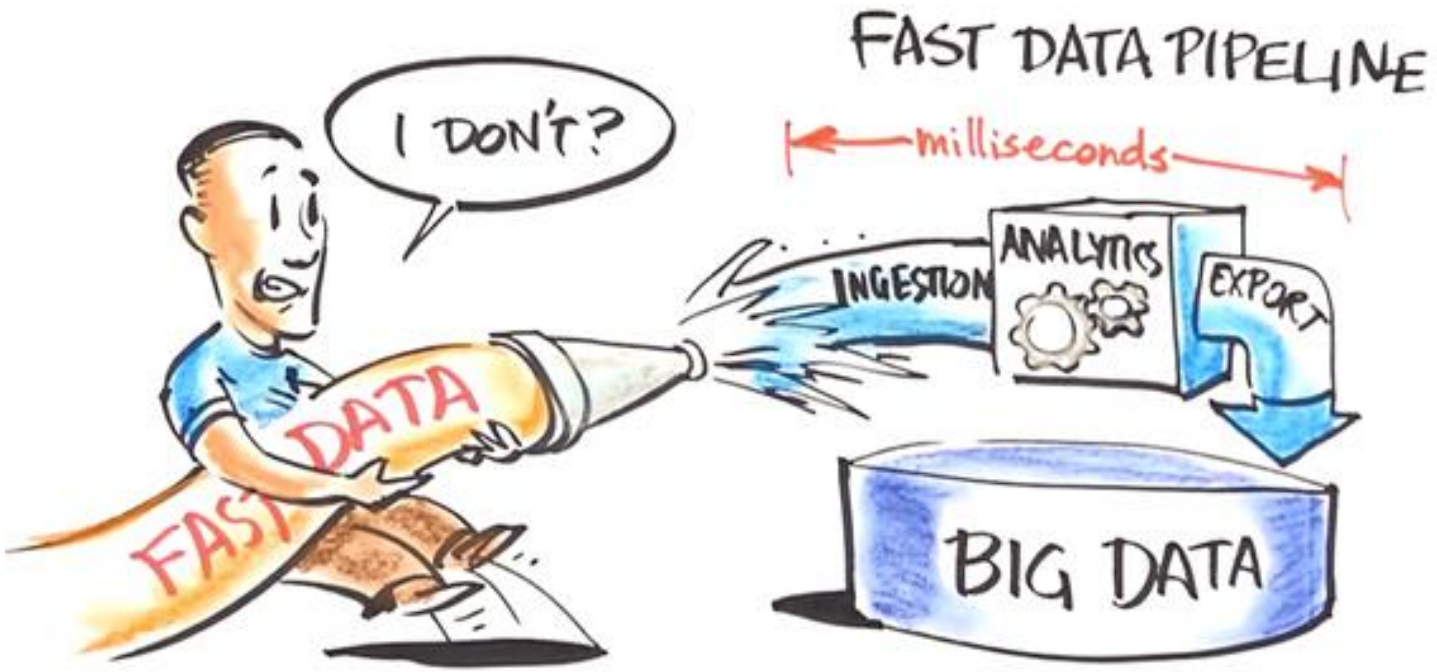
Sensor. Machine Data

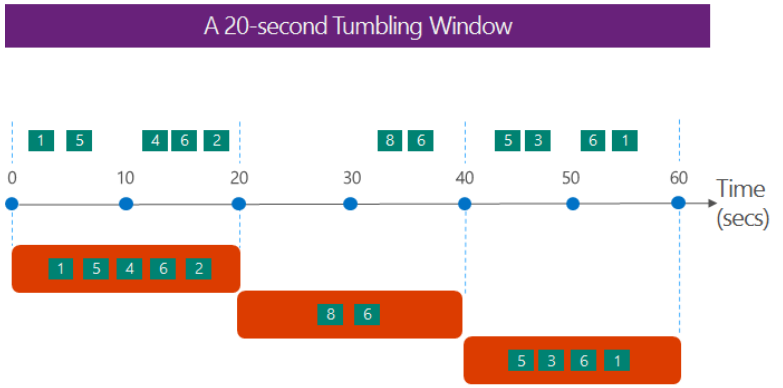
Server logs

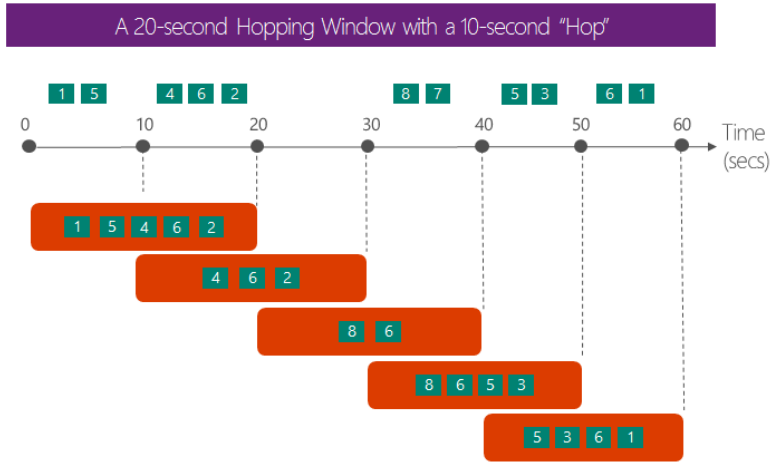
Sentiment, Web Data

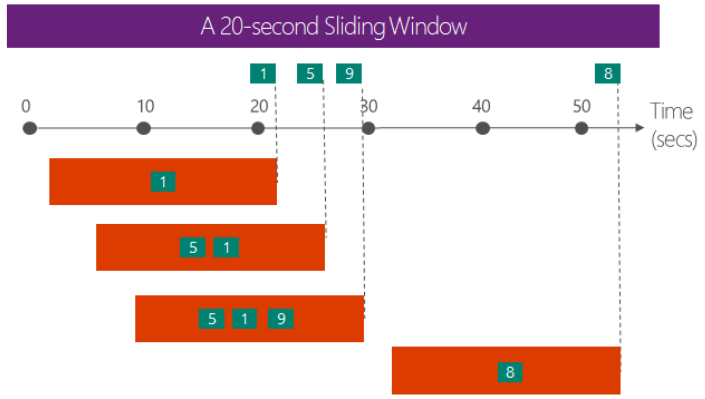
Clickstream

Geolocation





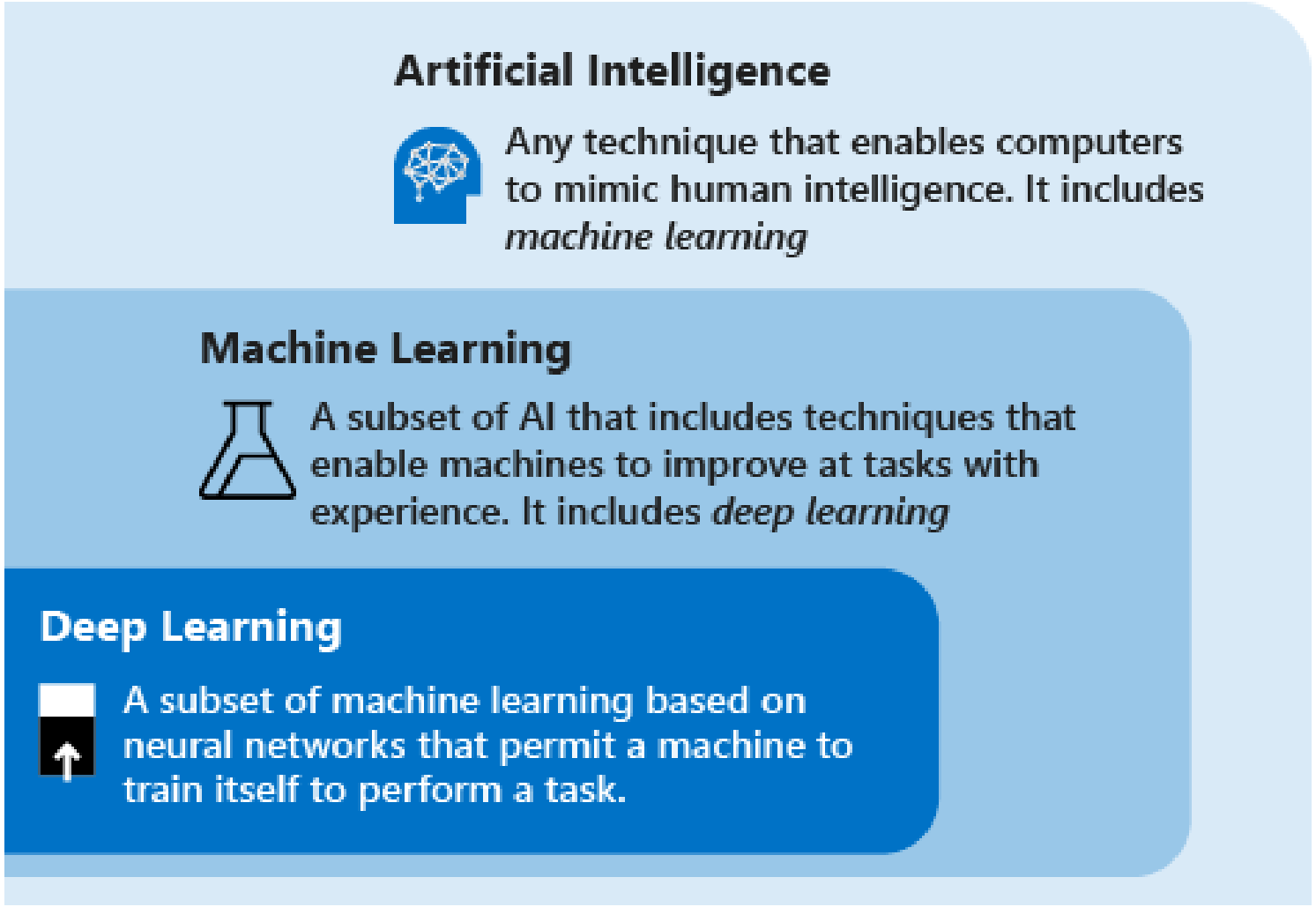




Very Mysterious

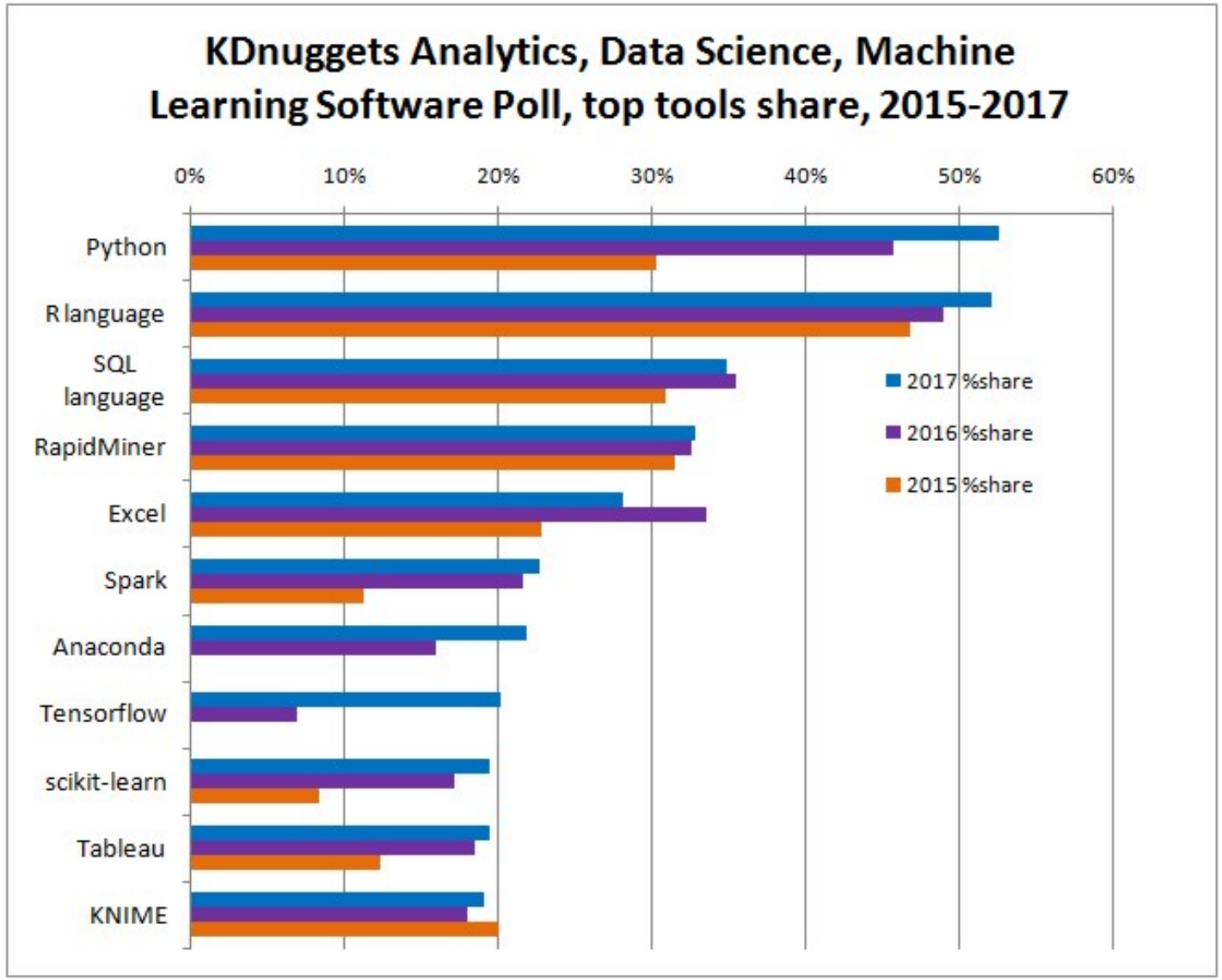
Traditional Programming
Rules + Data -> Answers

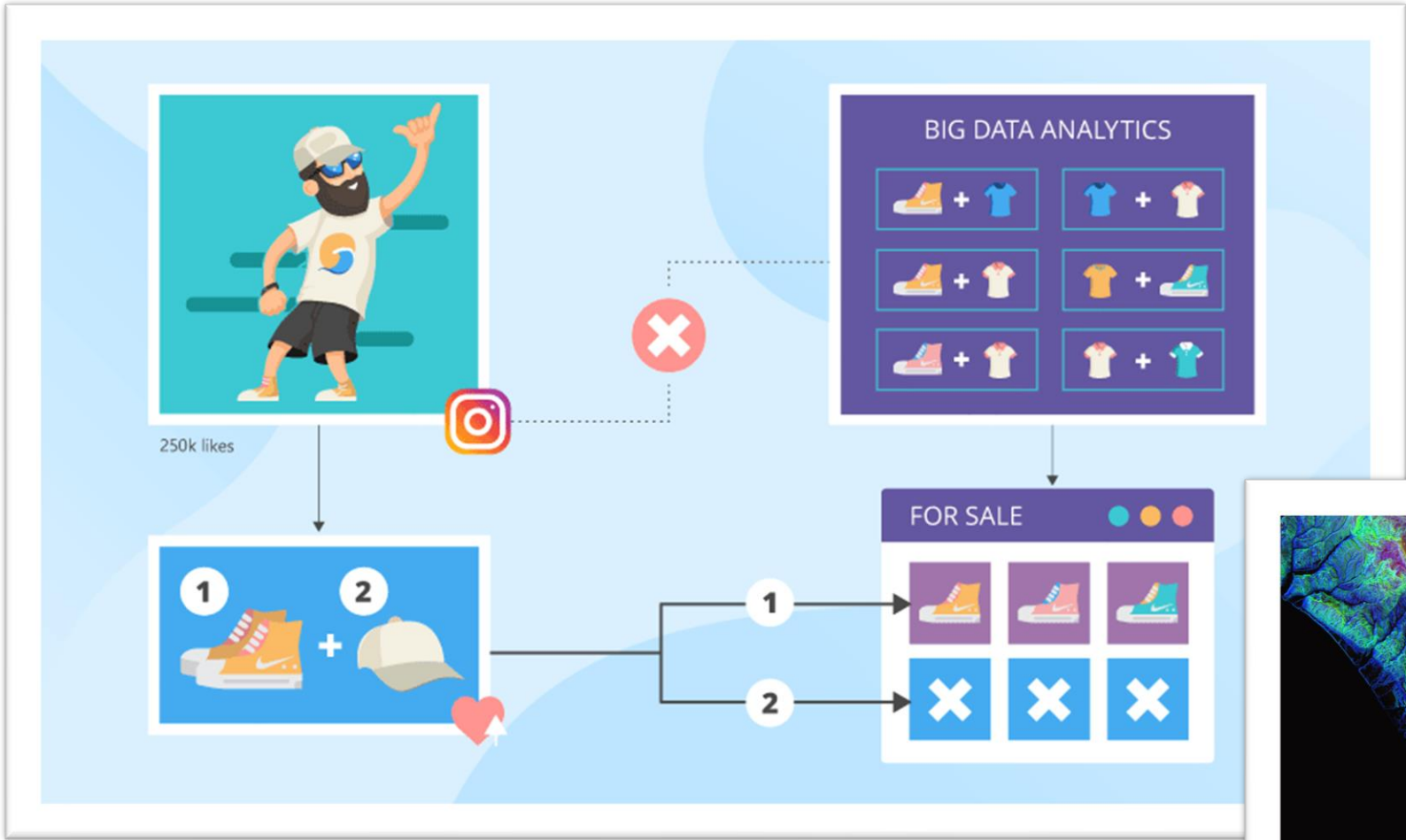
Machine Learning (Training Phrase)
Answers + Data -> Rules

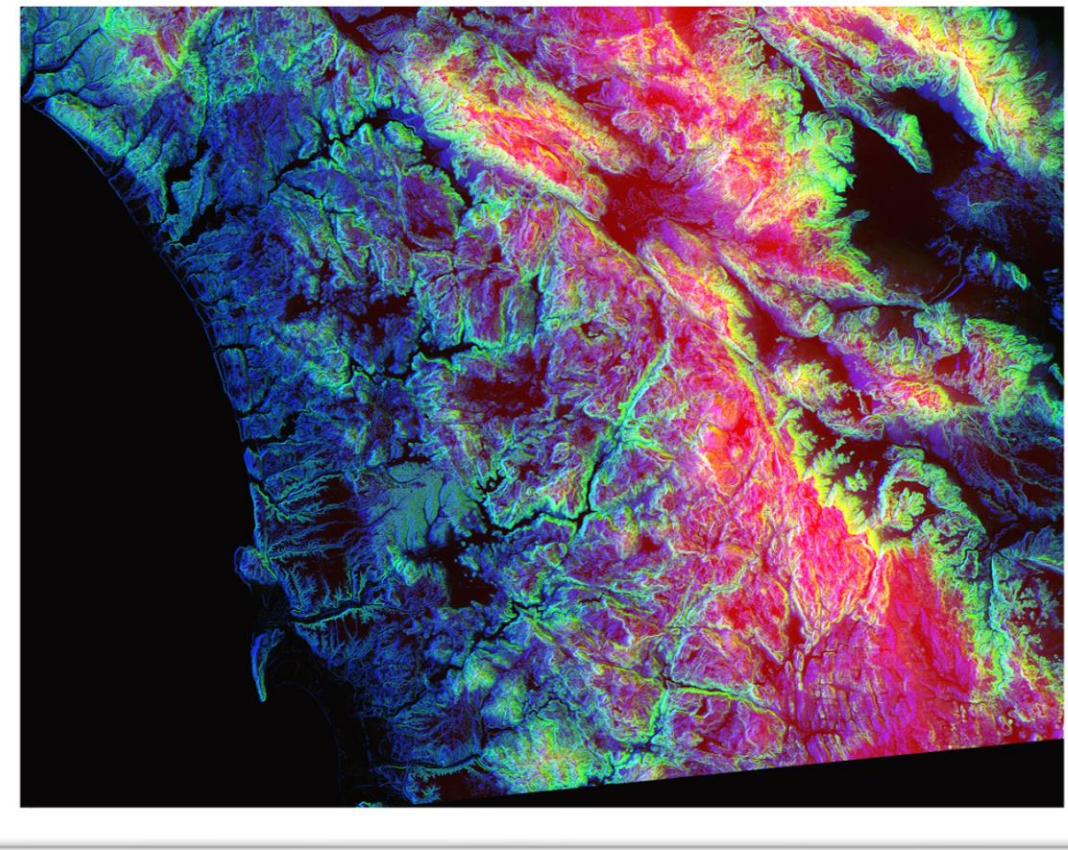


- Sentiment analysis (Duygu Analizi)
- Security analytics
- Social CRM/network analysis (Ağ Analizi)
- Churn mitigation (Abone Kaybı Azaltma)
- Brand monitoring (Marka,Ürün İzleme)
- Cross and Upsell (Çapraz Satışlar)
- Loyalty and promotion analysis (Sadakat ve Tanıtım)
- Web application optimization
- Operations analytics (İşlemler Analizi)
- Marketing campaign optimization (Pazarlama Kampanya)
- Brand management (Marka Yönetimi – Hedef Markette)
- Enterprise search (Kurumsal – Kurum İçi Arama)
- Social media analytics (Sosyal M. Analizi)
- Pricing optimization (Fiyatlandırma Opt.)
- Internal risk assessment (Risk Değerlendirmesi)
- Customer behavior analysis
- Revenue assurance (Gelir Güvencesi yat/gel/gid)
- Logistics optimization (Lojistik Optimazyonu)
- Clickstream analysis
- Influencer analysis (Etkileyici Analizi)
- IT infrastructure analysis (Altyapı Analizi)
- Legal discovery (Hukuki Keşif)
- Equipment monitoring (Ekipman İzleme)

Top DS Language together in-database




BIG DATA ANALYTICS
250k likes
FOR SALE
1
2
1
2



Çalışma ortamını hazırlama

- Python dünyasına giriş
- Python Engine kurulumu
- Anaconda kurulumu
- Jupyter Notebook ile çalışma
- Visual Studio Code ile çalışma
- Help kullanımı
- Paket kavramı ve paket indirme
- Interactive mode veya Script dosyaları ile çalışma

What is Python?

- ✓ Created in the **early 1990s**
- ✓ **Automating** repetitive tasks and writing **web apps** to building **machine learning** models and implementing **neural networks**
- ✓ **Researchers**, **mathematicians**, and **data scientists** in particular like Python because of its rich and easy-to **understand syntax** and the wide range of **open-source packages** available *(Packages are commonly used, **shared code** libraries that are freely available for anyone to use.)*
- ✓ Applications written in Python can run on almost any computer, including those running **Windows, macOS**, and popular distributions of **Linux**.
- ✓ The ecosystem contains a rich set of **development tools** for writing, debugging, and publishing Python applications

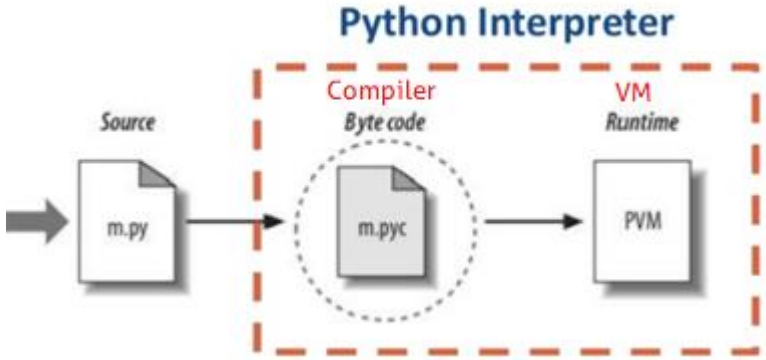
Execute Python Code

1-Interactive mode (REPL - Read-Eval Print Loop)

```
C:\WINDOWS\system32\cmd.exe - py
C:\Users\Kise>py    py -3
Python 3.8.2 (tags/v3.8.2:7b3ab59, Feb 25 2020, 22:45:29) [MSC v.1916 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> print("Merhaba Dünyalı")
Merhaba Dünyalı
>>>
```

2-Script mode

```
PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL                                    1: cmd

(base) C:\Users\Kise\Desktop\Python L\C_IPFDS-Python ile Veri Analizine Giriş\M2>python merhaba.py
merhaba dünyalı
```



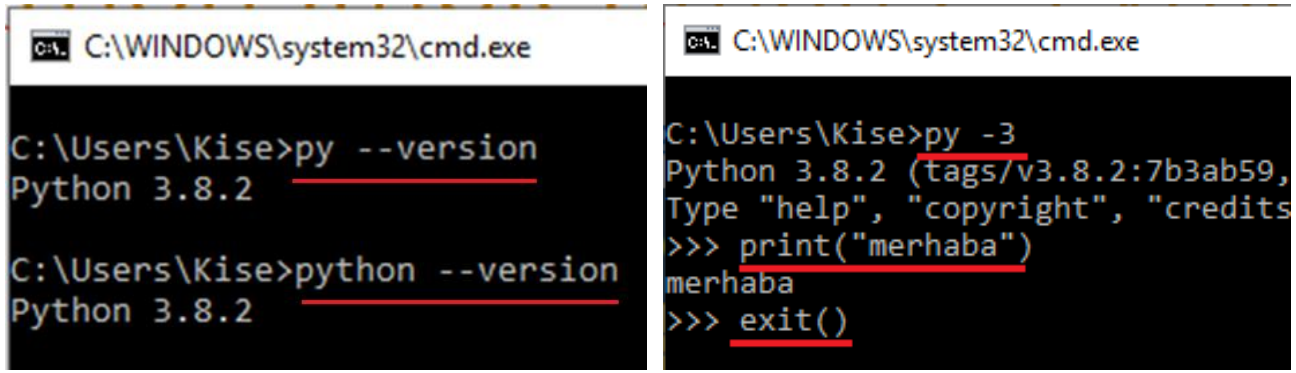
Python Implementations

- **CPython** (most popular)
  - https://www.python.org/
- **Anaconda** (a Python distribution for scientific tasks)
  - https://www.anaconda.com/
- **Jupyter Notebooks** (web-based interactive programming environment)
  - https://jupyter.org/
- **PythonAnywhere** (freemium hosted Python installation which lets you run Python in the browser, e.g. for tutorials, showcases, etc.)
  - https://www.pythonanywhere.com/
- **IronPython** (built on the .NET runtime)
- **Jython** (Python running on the Java Virtual Machine)
- **PyPy** (A fast python implementation with a JIT compiler, is a replacement for CPython)
- **MicroPython** (Python running on micro controllers)

https://www.python.org/download/alternatives/

Python 2, Python 3
Anaconda, Jupyter, VSCode

Check Python Installation



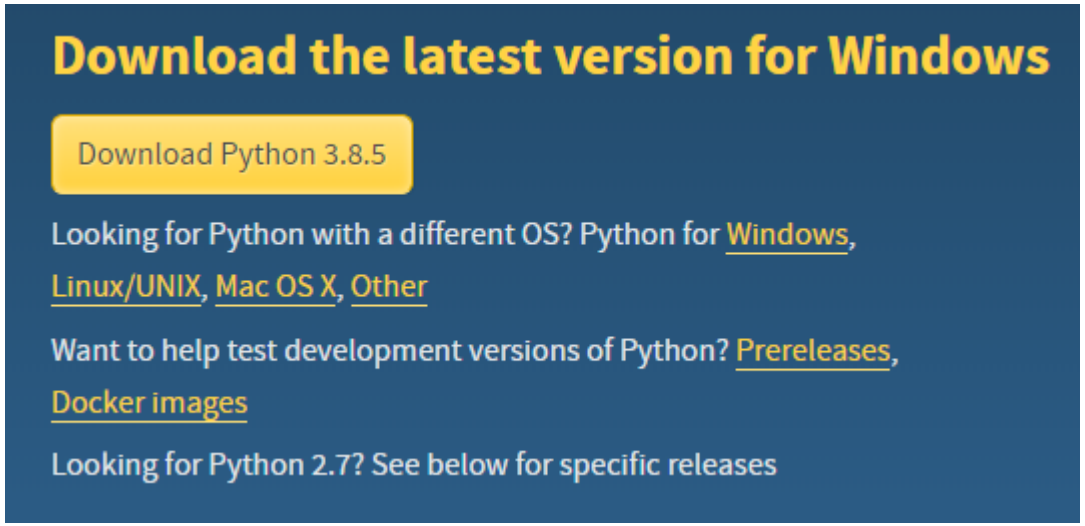
Some distributions of Linux have **Python 2** pre-installed on them, you'll need to take extra steps to install **Python 3**
`python3 --version` or `python --version`

Install Distribution

Anaconda https://www.anaconda.com/products/individual

Install Python Engine

https://www.python.org/downloads/



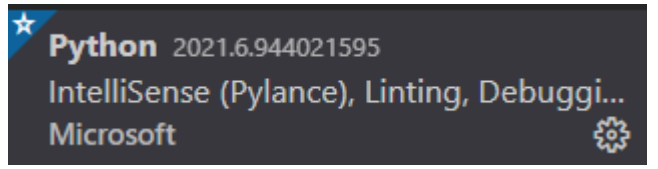
Install IDEs

https://jupyter.org/install

```
pip install jupyterlab
```

```
pip install notebook
```

https://code.visualstudio.com/Download



Exercise

(version, # ‘’, print, help, exit)

```
C:\WINDOWS\system32\cmd.exe

C:\Users\Kise>py --version
Python 3.8.2

C:\Users\Kise>python
Python 3.8.2 (tags/v3.8.2:7b3ab59, Feb 25 2020, 22:45:29) [MSC v.1916 32 bit
Type "help", "copyright", "credits" or "license" for more information.
>>> print("Merhaba Dünyalı")
Merhaba Dünyalı
>>> # Bu bir yorum satırıdır. Kod olarak değerlendirmez.
>>> print("Kodları birden fazla satırda yazmak için \
... sola yatık çizgi kullanabilirsin")
Kodları birden fazla satırda yazmak için sola yatık çizgi kullanabilirsin
>>> exit()

C:\Users\Kise>
```

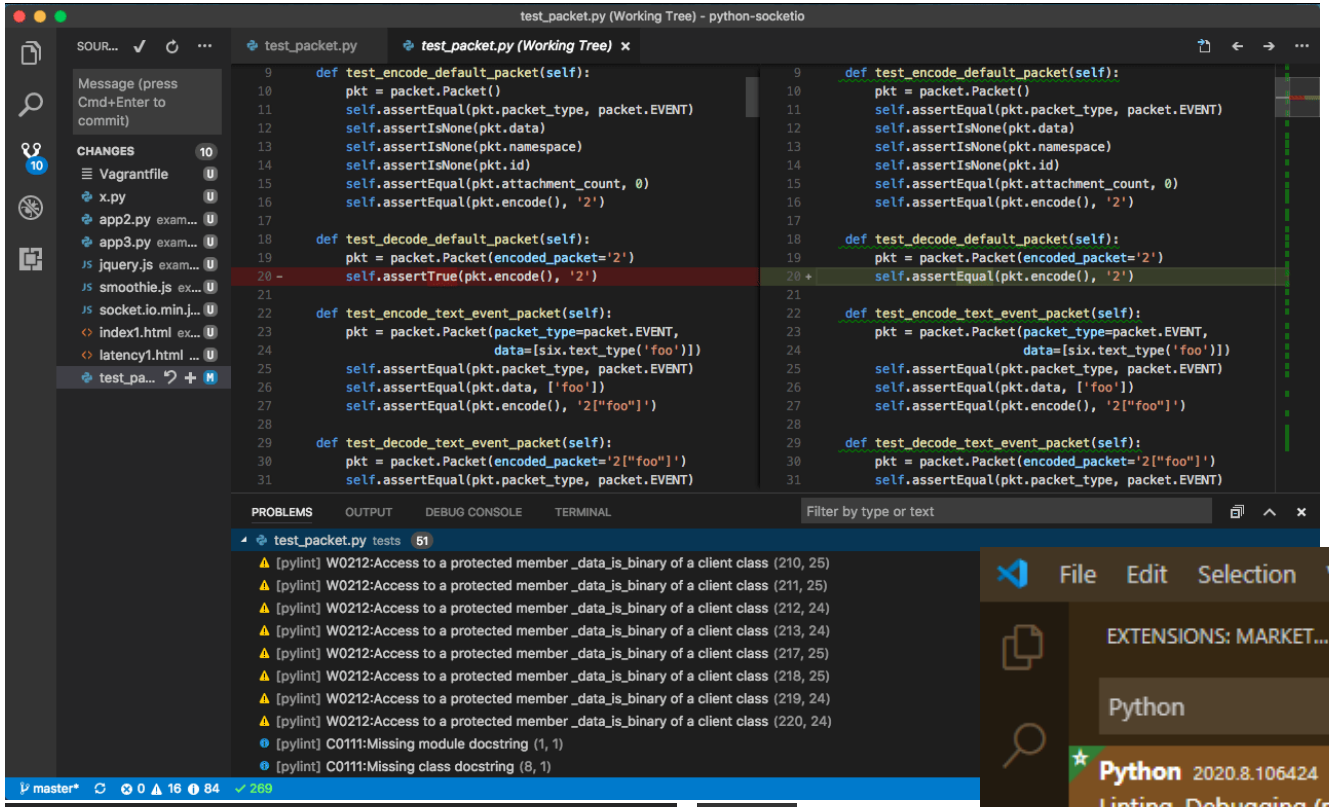

```
>>> print("merhaba","dünyalı",sep=">")
merhaba>dünyalı
```

Enter to display line-by-line

Space to go page-by-page.

```
C:\WINDOWS\system32\cmd.exe - python

C:\Users\Kise>python
Python 3.8.2 (tags/v3.8.2:7b3ab59, Feb 25 2020, 22:45:29) [MSC v.1916 32 bit
Type "help", "copyright", "credits" or "license" for more information.
>>> help(print)
Help on built-in function print in module builtins:

print(...)
    print(value, ..., sep=' ', end='\n', file=sys.stdout, flush=False)

    Prints the values to a stream, or to sys.stdout by default.
    Optional keyword arguments:
    file:  a file-like object (stream); defaults to the current sys.stdout.
    sep:   string inserted between values, default a space.
    end:   string appended after the last value, default a newline.
    flush: whether to forcibly flush the stream.

>>> help()

Welcome to Python 3.8's help utility!

If this is your first time using Python, you should definitely check out
the tutorial on the Internet at https://docs.python.org/3.8/tutorial/.

Enter the name of any module, keyword, or topic to get help on writing
Python programs and using Python modules.  To quit this help utility and
return to the interpreter, just type "quit".

To get a list of available modules, keywords, symbols, or topics, type
"modules", "keywords", "symbols", or "topics".  Each module also comes
with a one-line summary of what it does; to list the modules whose name
or summary contain a given string such as "spam", type "modules spam".

help> print
Help on built-in function print in module builtins:

print(...)
    print(value, ..., sep=' ', end='\n', file=sys.stdout, flush=False)

    Prints the values to a stream, or to sys.stdout by default.
    Optional keyword arguments:
    file:  a file-like object (stream); defaults to the current sys.stdout.
    sep:   string inserted between values, default a space.
    end:   string appended after the last value, default a newline.
    flush: whether to forcibly flush the stream.

help> q          Helpten Çıkış
```

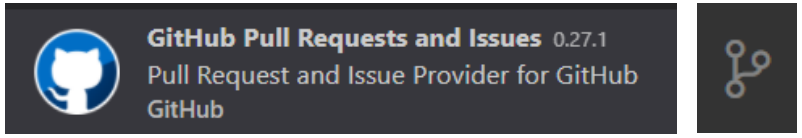
Visual Studio Code & Python Extension & Pylint

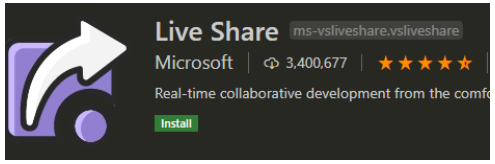


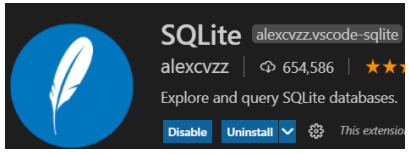
Pylint (optional) is one of the more popular Python linters.

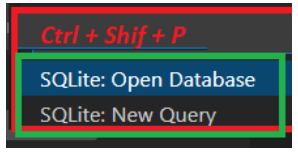
A linter is a useful tool that can check your code for possible bugs and coding style. If you install a linter, Visual Studio Code will **check your code each time** you save your code file.

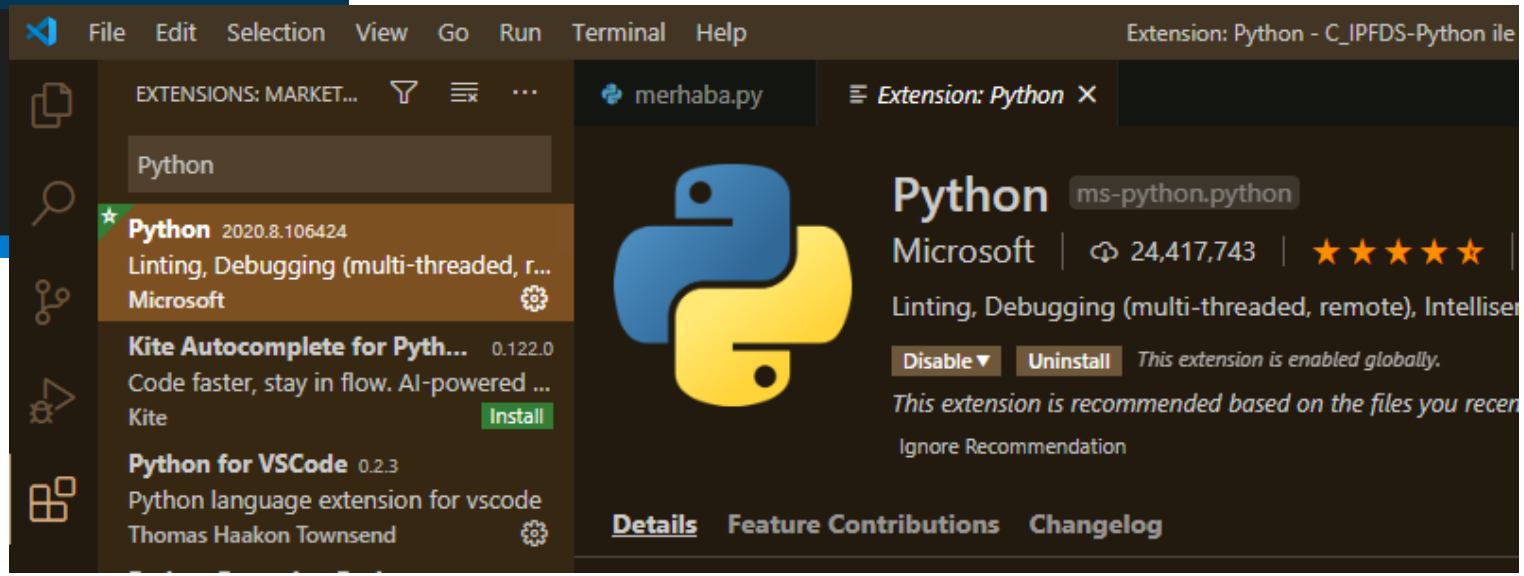
To start installation, you'll create a file with a **.py file** extension, which will trigger a **pop-up window** with an install button.



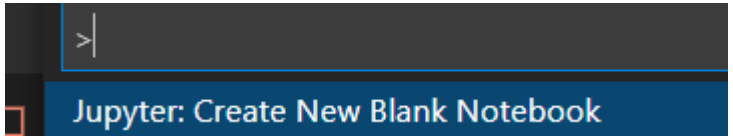








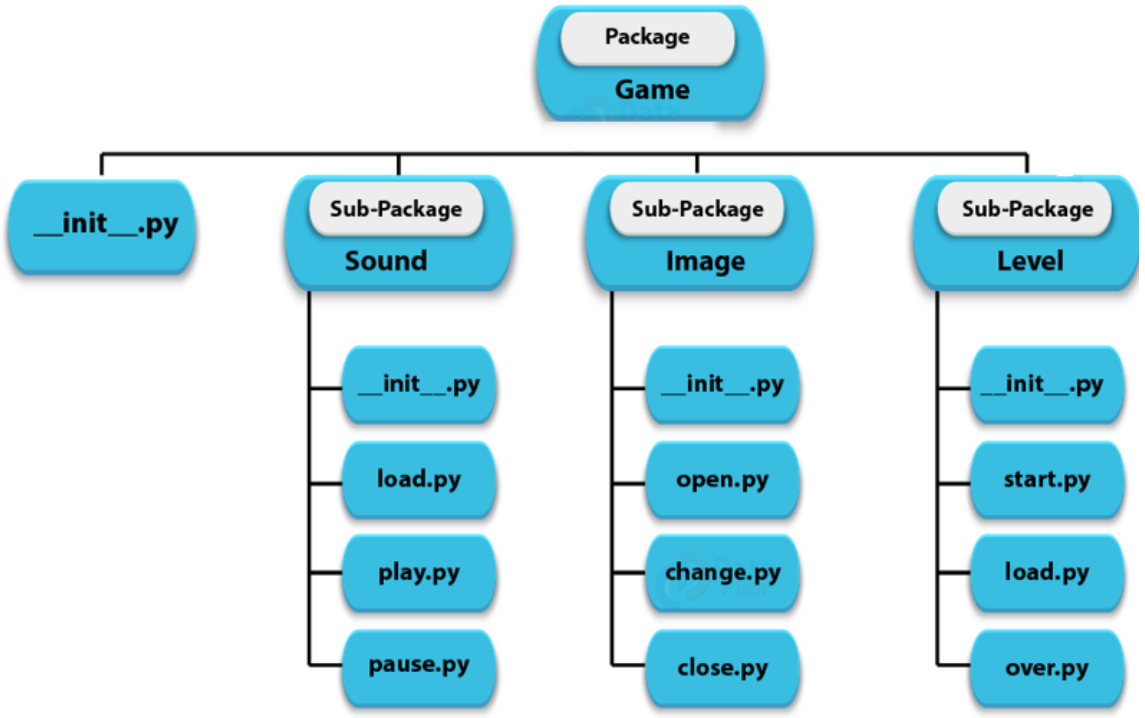
https://code.visualstudio.com/Download



Python Paket Yapısı

- Package, Module
- Virtual Environment

Package, Module Intro




```
import Game #  paket
import Game as gm # paket gm isminde

import Game.Image # Image modulu
from Game import Image # Image modulu
from Game import Image as gim # Image modulu gim isminde

#__init__.py içinde __all__=["",""] şeklinde tanımlananlar
from MyPackage import *
```

__init__.py içeriği yüklenir. Bu dosya içerisinden diğer modüller çağırılabilir.

Python packages: https://pypi.org/
Python Standart Library: https://docs.python.org/3/library/

- **pip list, pip show paketAdi**
- **pip upgrade (Terminal):**

```
python.exe -m pip install --upgrade pip
```

- **Paket yükleme (Terminal):**

```
py -m pip install emoji
      pip install Game
```

- **Paket silmek (Terminal):**

```
      pip uninstall emoji
```

- **Paket listesini dosyadan almak**

```
pip install -r requirements.txt
```

requirements.txt

```
numpy
scipy
Pillow
cython
matplotlib
scikit-image
tensorflow==1.14.0
keras==2.0.8
opencv-python
h5py==2.10.0
imgaug
IPython[all]
```

- **Paketleri çalışma ortamına alma örnekleri**

```
import matplotlib.pyplot as plt
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn import datasets
```

```
import sys, pprint
pprint.pprint(sys.path)
```

```
['c:\\Users\\Kise\\Desktop\\Python L\\C_IPFDS-Python ile Veri Analizine '
 'Giriş\\M2',
 'C:\\Users\\Kise\\Anaconda3\\lib',
 'C:\\Users\\Kise\\Anaconda3',
```

```
[14]
import sys
print(sys.platform)
```

```
win32
```

```
import math
math.pi
```

```
3.141592653589793
```

```
import os
os.mkdir("YeniKlasor")
```

✓ Paketler varsayılan olarak genel kapsamda yüklenir. (**pip list**)
✓ İstenirse **Virtual Environments** kullanılarak belirlenen paketler ayrı bağımsız bir sanal ortama yüklenir ve yönetilir. **Paket = Klasör**

Virtual Environment

- By Default, packages are installed globally
- Virtual environments can be used to contain and manage package collections.

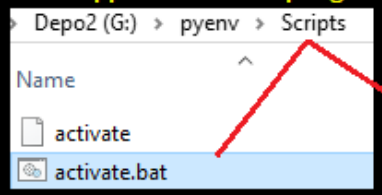
Package = Folder

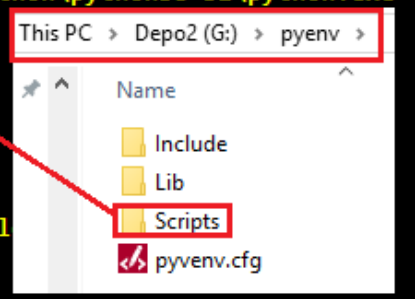
```
# Install virtual environment
pip install virtualenv

# Windows systems
python -m venv <folder_name>

# OSX/Linux (bash)
virtualenv <folder_name>
```

```
# Windows systems
# cmd.exe
<folder_name>\Scripts\Activate.bat
# Powershell
<folder_name>\Scripts\Activate.ps1
# bash shell
. ./<folder_name>/Scripts/activate

# OSX/Linux (bash)
<folder_name>/bin/activate
```

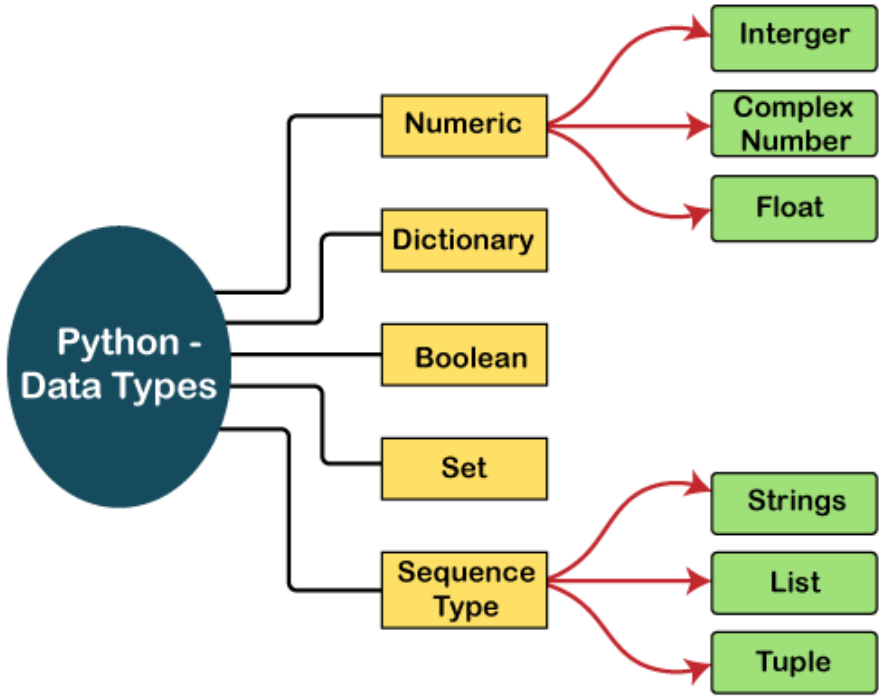
C:\WINDOWS\system32\cmd.exe

```
C:\Users\Kise>pip install virtualenv
Collecting virtualenv
  Downloading virtualenv-20.2.1-py2.py3-none-any.whl (4.9 MB)
     |████████████████████████████████| 4.9 MB 544 kB/s
Collecting appdirs<2,>=1.4.3
  Downloading appdirs-1.4.4-py2.py3-none-any.whl (9.6 kB)
Collecting filelock<4,>=3.0.0
  Downloading filelock-3.0.12-py3-none-any.whl (7.6 kB)
Requirement already satisfied: six<2,>=1.9.0 in c:\users\kise\appdata\roaming\python\python38\site-packages
Collecting distlib<1,>=0.3.1
  Downloading distlib-0.3.1-py2.py3-none-any.whl (335 kB)
     |████████████████████████████████| 335 kB 284 kB/s
Installing collected packages: appdirs, filelock, distlib, virtualenv
Successfully installed appdirs-1.4.4 distlib-0.3.1 filelock-3.0.12 virtualenv-20.2.1
WARNING: You are using pip version 20.2.4; however, version 20.3.1 is available.
You should consider upgrading via the 'c:\users\kise\appdata\local\programs\python\python38-32\python.exe

C:\Users\Kise>python -m venv "G:\pyenv"

C:\Users\Kise>G:\pyenv\scripts\Activate.bat

(pyenv) C:\Users\Kise>pip install emoji
Collecting emoji
  Using cached emoji-0.6.0.tar.gz (51 kB)
Using legacy 'setup.py install' for emoji, since package 'wheel' is not install
Installing collected packages: emoji
    Running setup.py install for emoji ... done
Successfully installed emoji-0.6.0
WARNING: You are using pip version 20.2.1; however, version 20.3.1 is available.
You should consider upgrading via the 'g:\pyenv\scripts\python.exe -m pip install --upgrade pip' command.

(pyenv) C:\Users\Kise>

(pyenv) C:\Users\Kise>pip list
Package    Version
---------- -------
colorama   0.4.4
emoji      0.6.0
pip        20.2.1
setuptools 49.2.1

(pyenv) C:\Users\Kise>G:\pyenv\scripts\deactivate.bat
C:\Users\Kise>
```





Standart Data Types



Name	Type	Description
Integers	int	Whole numbers, such as: 3 300 200
Floating point	float	Numbers with a decimal point: 2.3 4.6 100.0
Strings	str	Ordered sequence of characters: "hello" 'Sammy' "2000" "楽しい"
Lists	list	Ordered sequence of objects: [10,"hello",200.3]
Dictionaries	dict	Unordered Key:Value pairs: {"mykey" : "value" , "name" : "Frankie"}
Tuples	tup	Ordered immutable sequence of objects: (10,"hello",200.3)
Sets	set	Unordered collection of unique objects: {"a","b"}
Booleans	bool	Logical value indicating **True** or **False**

Function	Description
int(x [,base])	Converts x to an integer. base specifies the base if x is a string.
long(x [,base])	Converts x to a long integer. base specifies the base if x is a string.
float(x)	Converts x to a floating-point number.
complex(real [,imag])	Creates a complex number.
str(x)	Converts object x to a string representation.
repr(x)	Converts object x to an expression string.
eval(str)	Evaluates a string and returns an object.
tuple(s)	Converts s to a tuple.

Class	Description	Immutable?
bool	Boolean value	✓
int	integer (arbitrary magnitude)	✓
float	floating-point number	✓
list	mutable sequence of objects	
tuple	immutable sequence of objects	✓
str	character string	✓
set	unordered set of distinct objects	
frozenset	immutable form of set class	✓
dict	associative mapping (aka dictionary)	

Operators

Operator	Name (Comparison)	Example
==	Equal	x == y
!=	Not equal	x != y
>	Greater than	x > y
<	Less than	x < y
>=	Greater than or equal to	x >= y
<=	Less than or equal to	x <= y

Operator	Name (Arithmetic)	Example
+	Addition	x + y
-	Subtraction	x - y
*	Multiplication	x * y
/	Division	x / y
%	Modulus	x % y
**	Exponentiation Power	x ** y
//	Floor division	x // y

Operator	Description (Logical)	Example
and	Returns True if both statements are true	x < 5 and x < 10
or	Returns True if one of the statements is true	x < 5 or x < 4
not	Reverse the result, returns False if the result is true	not(x < 5 and x < 10)

Operator	Example (Assignment)	Same As
=	x = 5	x = 5
+=	x += 3	x = x + 3
-=	x -= 3	x = x - 3
*=	x *= 3	x = x * 3
/=	x /= 3	x = x / 3
%=	x %= 3	x = x % 3
//=	x //= 3	x = x // 3
**=	x **= 3	x = x ** 3

Operator	Description (Identity)	Example
is	Returns True if both variables are the same object	x is y
is not	Returns True if both variables are not the same object	

Operator	Description (Membership)	Example
in	Returns True if a sequence with the specified value is present in the object	x in y
not in	Returns True if a sequence with the specified value is not present in the object	