

Business intelligence

Machine Learning

DATA SCIENCE

DataTech

Database

Python/R

Data Mining

Data Mining

Python/R

DataTech

Business intelligence

# دوره جامع تکنولوژی داده (DataTech)

برنامه نویسی در علم داده با Python

Python/R

Statistics

DATA SCIENCE

Database

Statistics

## محمد امینی



- دکتری تخصصی مهندسی فناوری اطلاعات (تجارت الکترونیک)  
دانشگاه علم و صنعت ایران
- پژوهشگر ارشد مرکز مطالعات و تحقیقات فناپ
- مدرس رشته مهندسی فناوری اطلاعات
- متخصص و مدرس علم داده



[m.amini84@gmail.com](mailto:m.amini84@gmail.com)

پل های ارتباطی



**LinkedIn:** [www.linkedin.com/in/mohammad-amini-iust](https://www.linkedin.com/in/mohammad-amini-iust)



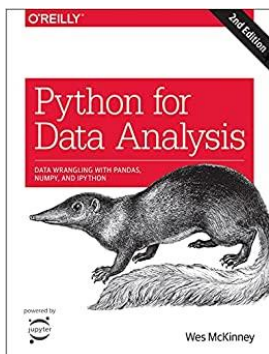
**GitHub:** <https://github.com/Farhad9660/Python-for-Data-Science>



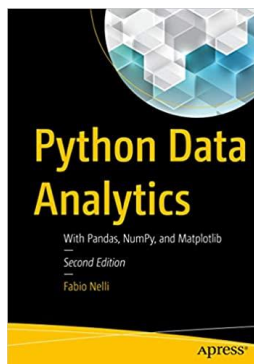
وزارت علوم، تحقیقات و فناوری  
دانشگاه خاتم  
Khatam University



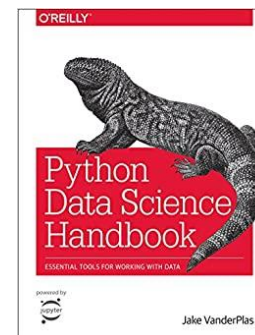
## منابع انگلیسی



**Python for Data Analysis:  
Data Wrangling with Pandas,  
NumPy, and IPython 2nd  
Edition**



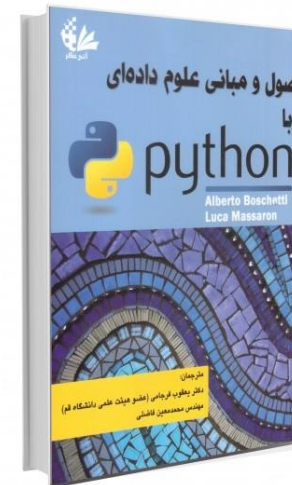
**Python Data Analytics: With  
Pandas, NumPy, and  
Matplotlib 2nd Edition**



**Python Data Science  
Handbook: Essential  
Tools for Working with  
Data**

## منابع فارسی

اصول و مبانی علوم داده‌ای با Python  
انتشارات آتی نگر



## وب سایت‌های مفید

<https://www.kaggle.com/>

<https://machinelearningmastery.com/>

<https://www.kdnuggets.com/>

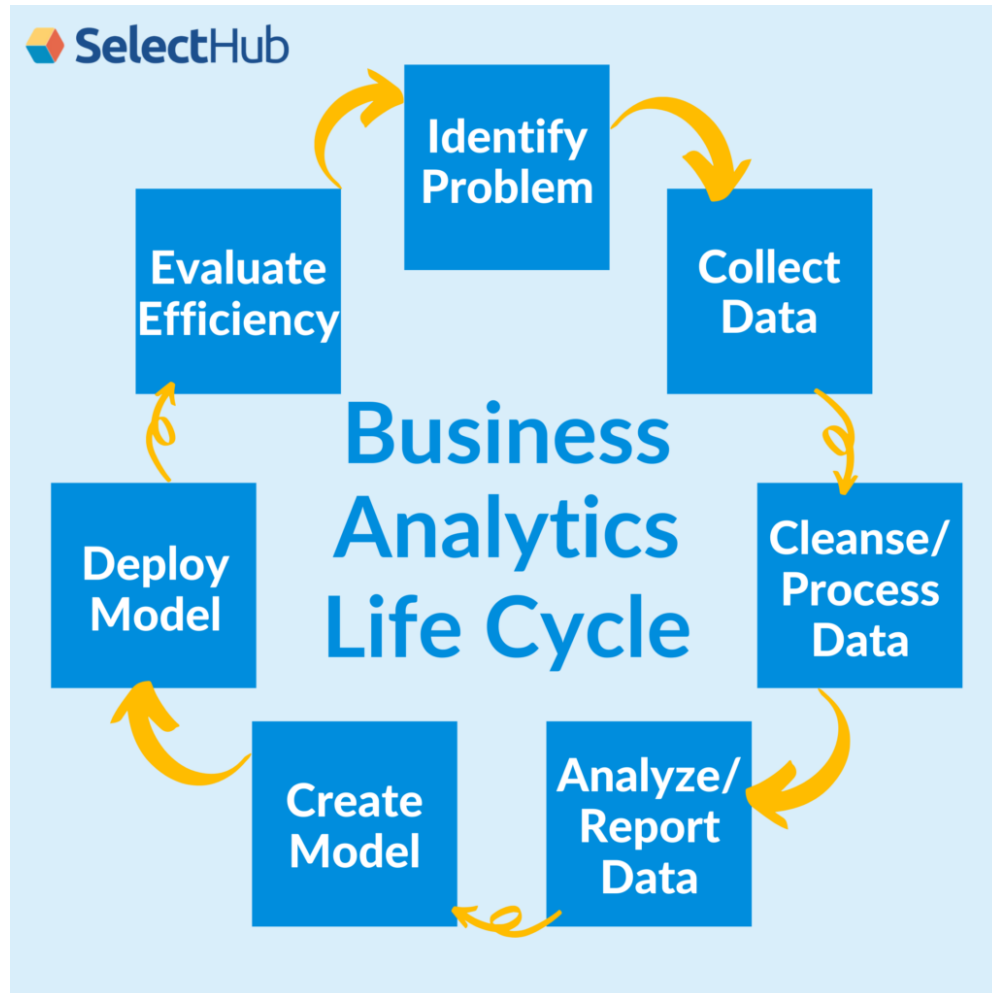
<https://towardsdatascience.com/>



وزارت علوم، تحقیقات و فناوری  
دانشگاه خاتم  
Khatam University







فرایند تحلیل داده‌ها در حل  
مسائل کسب و کار

# مقدمه



مسیر دانشمند داده  
Data Science



وزارت علوم، تحقیقات و فناوری  
دانشگاه خاتم  
Khatam University



# مهارت‌های علم داده با پایتون



 pandas



NumPy

matplotlib  
seaborn



وزارت علوم، تحقیقات و فناوری  
دانشگاه خاتم  
Khatam University





## مباحث این دوره

- آشنایی با زبان پایتون و محیط برنامه نویسی ۴ ساعت
- کنترل جریان، کار با انواع داده‌ای ۳ ساعت
- توابع، عبارات لامبدا، مدیریت استثنائات، شی گرای ۴ ساعت
- پردازش داده ها با NumPy ۴ ساعت
- تحلیل داده‌ها با Pandas ۵ ساعت
- مصورسازی داده‌ها در پایتون (matplotlib, Seaborn, Pandas) ۵ ساعت
- کار با پروژه های عملی ۳ ساعت



# راه اندازی محیط برنامه نویسی

## نصب و راه اندازی بسته Anaconda



وزارت علوم، تحقیقات و فناوری  
دانشگاه خاتم  
Khatam University

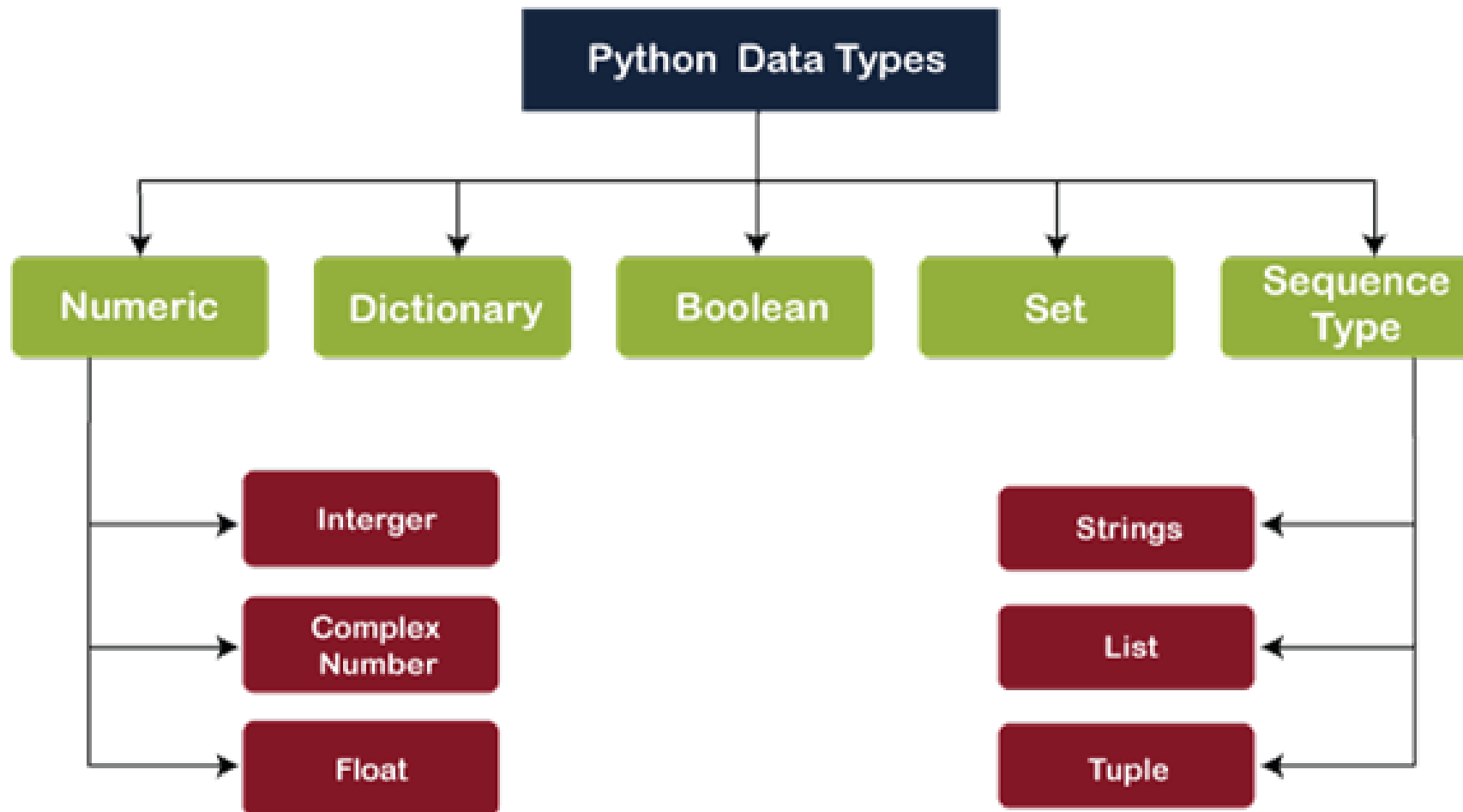


## آشنایی با محیط Jupyter Notebook



وزارت علوم، تحقیقات و فناوری  
دانشگاه خاتم  
Khatam University





## انواع داده‌ای

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

### Rules for Python variables:

- A variable name must start with a letter or the underscore character .
- A variable name cannot start with a number .
- A variable name can only contain alpha-numeric characters and underscores (A-z, 0-9, and \_ )
- Variable names are case-sensitive (age, Age and AGE are three different variables)
- The reserved words(keywords) cannot be used naming the variable. .



❑ The reserved words(keywords) cannot be used as variable names

متغیرها در پایتون

## Python Keywords

False	def	if	raise
None	del	import	return
True	elif	in	try
and	else	is	while
as	except	lambda	with
assert	finally	nonlocal	yield
break	for	not	
class	from	or	
continue	global	pass	

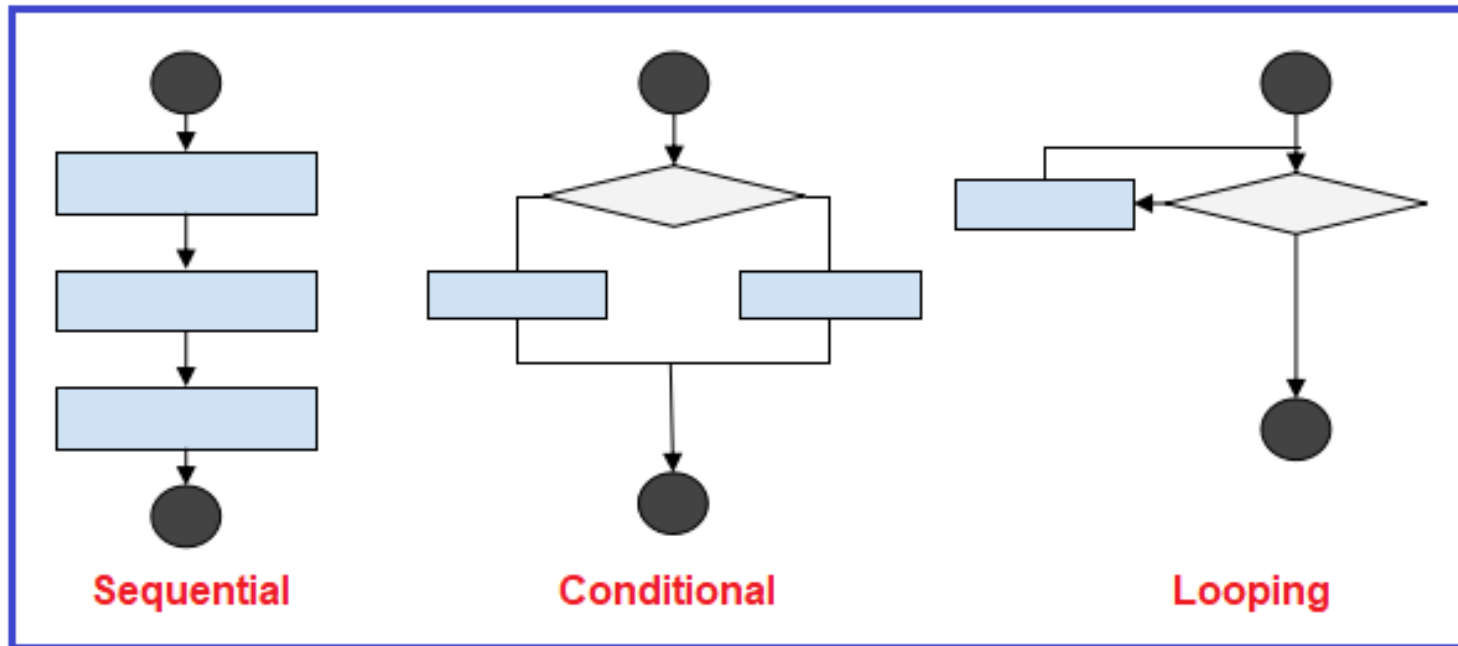


وزارت علوم، تحقیقات و فناوری  
دانشگاه خاتم  
Khatam University



## کنترل جریان

- ساختار شرطی : if, elif, else
- ساختار حلقه ای : while, for



## رشته ها

- تعریف، عملیات (indexing, slicing, reverse indexing)، متد len()
- متد print و قالب دهی خروجی
- محاسبات و اعمال بر روی رشته ها: متدهای رشته، خصوصیات رشته و ..
- (lower, upper, strip, split, find, islower, isupper, istitle, )



## لیست‌ها

- تعریف
- عملیات (indexing, slicing, reverse indexing)
- متدها (حذف، اضافه، تغییر) و پردازش
- پردازش لیستها با ساختار حلقه‌ای
- اعمال بر روی رشته‌ها به شکل list comprehension



## نوع داده ای تاپل

- تعریف
- عملیات و متدها



## نوع داده ای مجموعه

- تعریف
- عملیات و متدها
- پردازش (حذف ، اضافه، تغییر)





## نوع داده ای دیکشنری

- تعریف
- عملیات و متدها
- پردازش (حذف ، اضافه ، تغییر)



## توابع و عبارات لامبدا

- توابع: تعریف، استفاده
- توابع گمنام، عبارات لامبدا



# مقدمات برنامه نویسی

## مدیریت استثنائات



وزارت علوم، تحقیقات و فناوری  
دانشگاه خاتم  
Khatam University



# پردازش داده ها با NumPy



# NumPy



وزارت علوم، تحقیقات و فناوری

دانشگاه خاتم

Khatam University



FANAP

# پردازش داده ها با NumPy



NumPy

- بسته بنیادی پایتون برای انجام محاسبات بر روی ساختارهای داده ای آرایه و ماتریس
- دارای توابع برای انجام محاسبات سریع جبر خطی

- mathematical,
- logical,
- shape manipulation,
- Sorting
- selecting, I/O,
- discrete Fourier transforms,
- basic linear algebra, basic statistical operations,
- random simulation

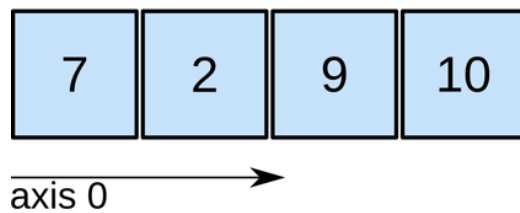


وزارت علوم، تحقیقات و فناوری  
دانشگاه خاتم  
Khatam University



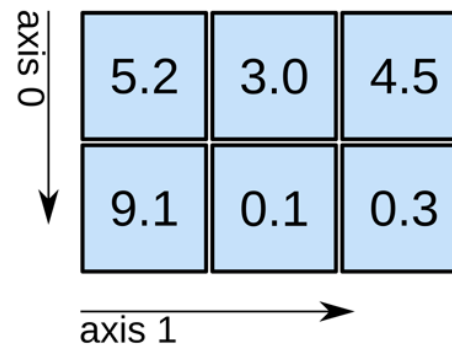
# پردازش داده ها با NumPy

1D array



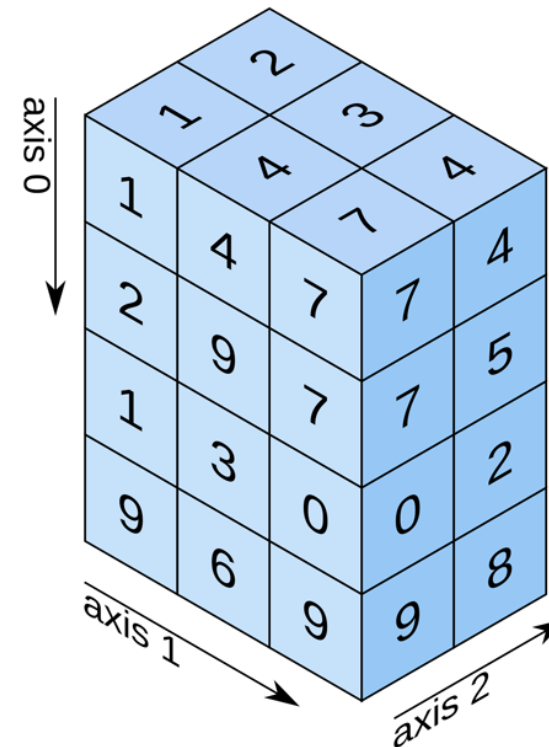
shape: (4,)

2D array



shape: (2, 3)

3D array



shape: (4, 3, 2)





# پردازش و تحلیل داده ها با Pandas



وزارت علوم، تحقیقات و فناوری  
دانشگاه خاتم  
Khatam University



- پانداز یک کتابخانه متن باز در پایتون برای تحلیل تخصصی و کارآمد داده‌هاست
- پانداز بر اساس کتابخانه نامپای نوشته شده است
- در حال حاضر پانداز، کتابخانه اصلی و پذیرفته شده برای تحلیل داده‌ها توسط پایتون است

# پردازش و تحلیل داده ها با Pandas



## • ویژگی‌های pandas

- کارایی و سرعت بالا در پردازش داده‌های ماتریسی و جدولی
- توانایی پردازش داده‌ها از منابع مختلف و با فرمت‌های مختلف
- توانایی انجام عملیات مختلف پردازشی بر روی مجموعه داده‌ها

merging, groupBy, re-ordering, and re-shaping. subsetting, slicing, filtering, •

- انجام پیش پردازش و پاکسازی داده ها
- توانایی مدلسازی، توصیف و تحلیل آماری داده‌ها
- توانایی یکپارچه سازی با دیگر کتابخانه های پایتون مانند scikit-learn, statsmodels, SciPy
- قابلیت‌های مصورسازی داخلی

# پردازش و تحلیل داده ها با Pandas

## • ساختارهای داده‌ای در pandas

**Series**

1
2
3
4

**dataframe**

1	2	3
4	5	6
7	8	9
10	11	12



وزارت علوم، تحقیقات و فناوری  
دانشگاه خاتم  
Khatam University



# پردازش و تحلیل داده ها با Pandas

## • ساختارهای داده‌ای در pandas

index	A	3
	B	-5
	C	7
	D	4

Series

A **one-dimensional** labelled array  
capable of holding any data type

	columns			
	Country	Capital	Population	
index {	1	Switzerland	Bern	8585405
	2	Australia	Canberra	24976386
	3	Canada	Ottawa	37157431

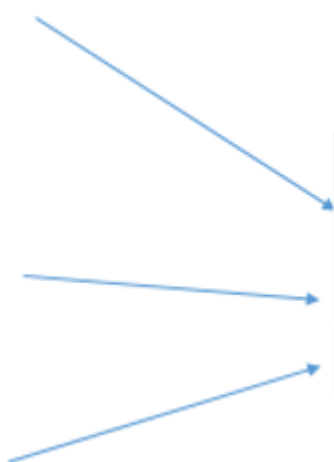
DataFrame

A **two-dimensional** labelled data  
structure capable with columns of  
potentially **different types**

## گروه بندی سطرها: Group By

گروه بندی سطر یک دیتافری بر اساس مقادیر یک ستون و اعمال توابع جمعی بر روی هر گروه

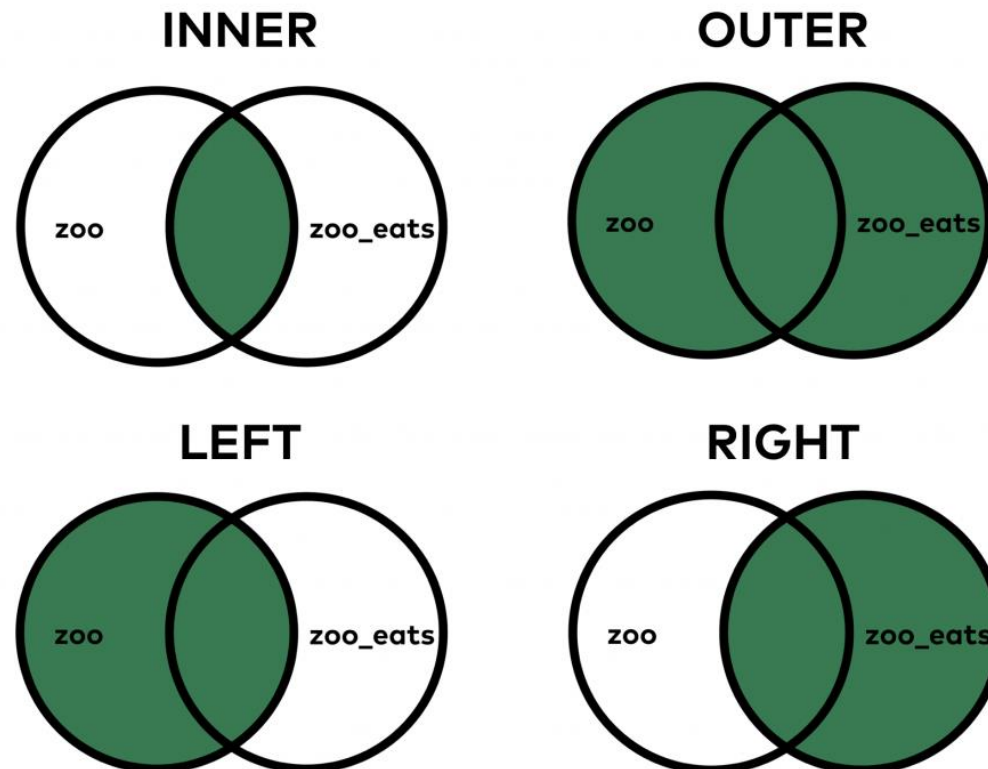
ID	Value
1	50.30
1	123.30
1	132.90
2	50.30
2	123.30
2	132.90
2	88.90
3	50.30
3	123.30



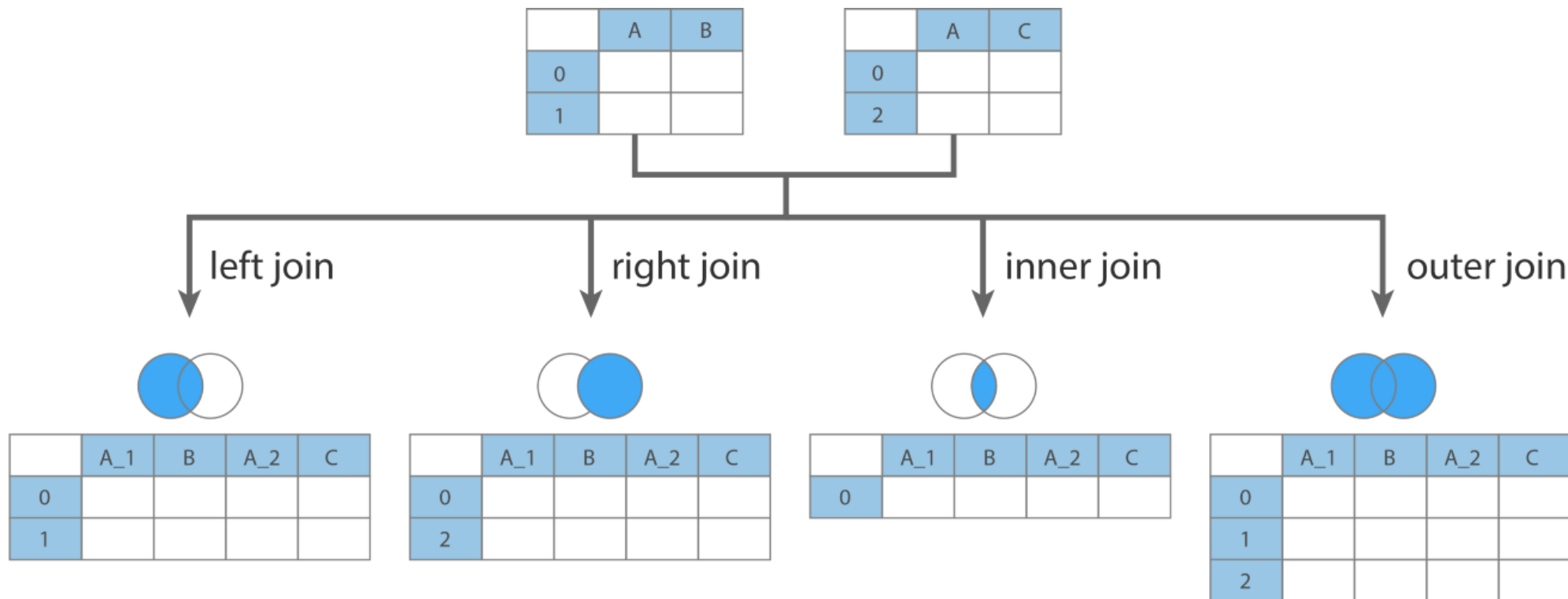
ID	Value
1	306.50
2	395.40
3	173.60



## ادغام دیتافریم ها: Join



## ادغام دیتافریم ها: Join



# مصورسازی داده ها



وزارت علوم، تحقیقات و فناوری

دانشگاه خاتم

Khatam University



فاناب  
FANAP

## • ابزارهای مصورسازی در پایتون

 pandas

 matplotlib

 seaborn

# با تشکر

