



Bài giảng môn học:  
**Khoa học dữ liệu (7080509)**

# Chương 4: Một số thư viện Python quan trọng trong khoa học dữ liệu – Phần 1

Đặng Văn Nam

[dangvannam@hmg.edu.vn](mailto:dangvannam@hmg.edu.vn)

# Nội dung phần 1

---

**4.1 Quy trình thực hiện một dự án KHDL**

**4.2 Giới thiệu một số thư viện Python quan trọng cho KHDL**

**4.3 Thư viện Numpy**

**4.4 Thư viện Pandas**

**4.5 Thư viện Matplotlib**

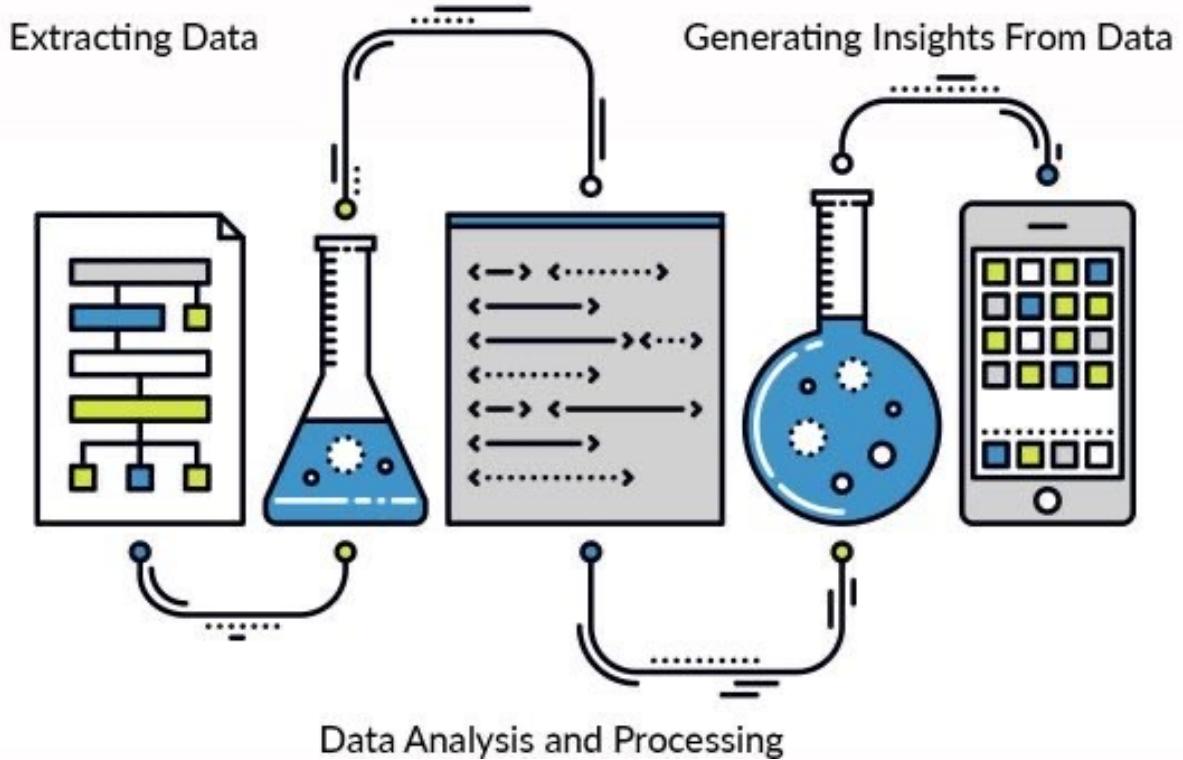
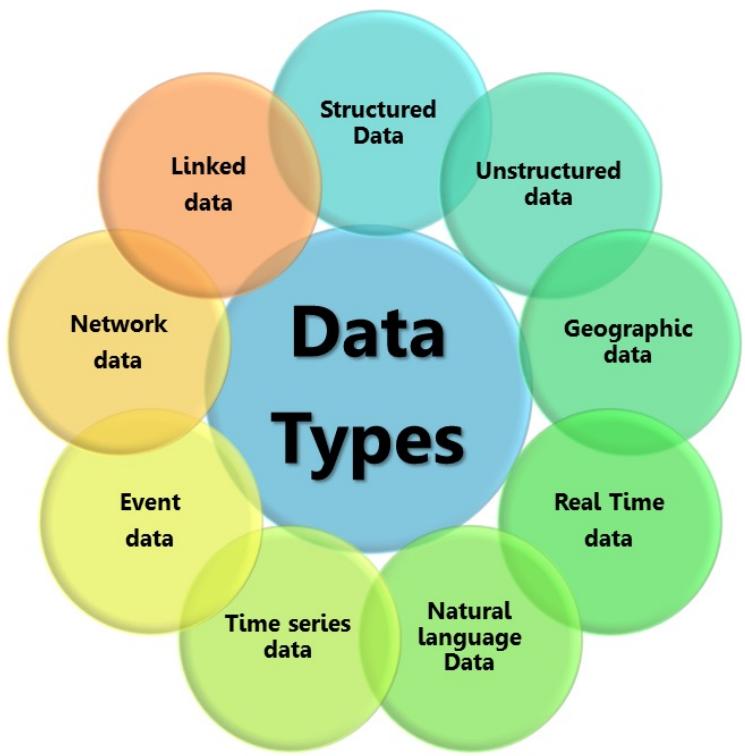


# 1. Quy trình của một dự án về Khoa học dữ liệu



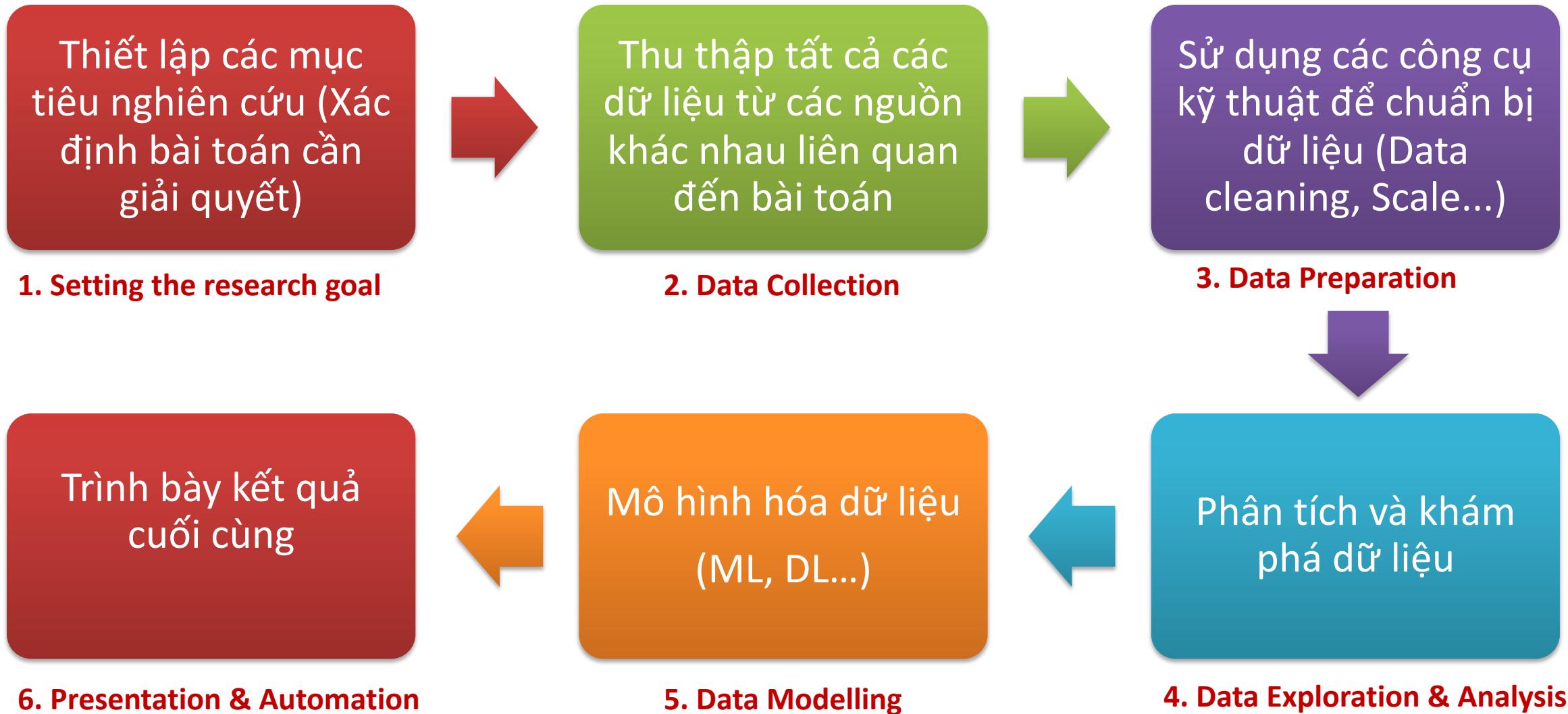
**“DATA IS THE NEW GOLD”**





We'll help you get **insights**

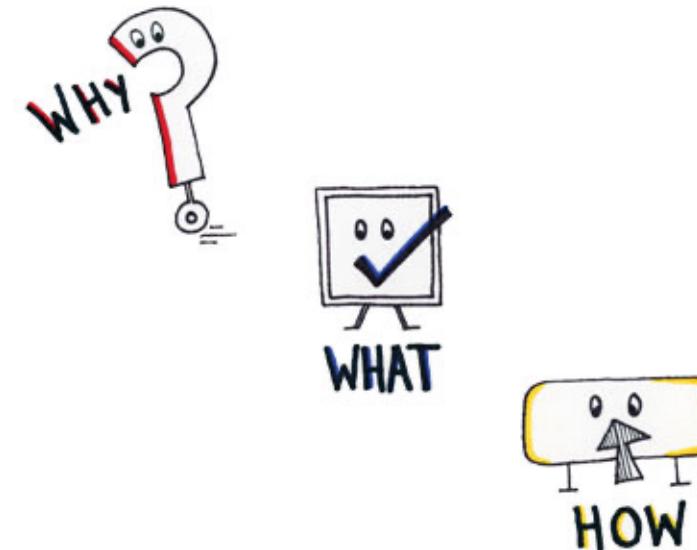
# Quy trình tổng quát



# B1: Thiết lập mục tiêu nghiên cứu

- **Setting the research goal**

- Một dự án data science bắt đầu bằng việc tìm hiểu xác định mục tiêu (what), lý do cần thực hiện (why) và thực hiện nó như thế nào (how)



- Trả lời ba câu hỏi này là mục tiêu của giai đoạn đầu tiên, để mọi người biết phải làm gì và có thể đồng ý về hướng hành động tốt nhất.

## B1: Thiết lập mục tiêu nghiên cứu (2)

---

- **Kết quả của quá trình này bao gồm:**
  - Mục tiêu nghiên cứu
  - Hiểu nhiệm vụ và bối cảnh dự án
  - Dự án cần dữ liệu gì? lấy như thế nào? Ở đâu? Số lượng như thế nào?
  - Tài nguyên nào bạn muốn sử dụng (nhân lực, thời gian)
  - Sản phẩm của dự án là gì, sẽ được sử dụng ở đâu?
  - Đâu là thước đo dự án có thành công hay không?
  - Một kế hoạch hành động với thời gian biểu (timetable)

# B1: Thiết lập mục tiêu nghiên cứu (3)

---

- **Mục tiêu nghiên cứu**
  - Mục tiêu nghiên cứu có thể được bắt nguồn từ nhu cầu hoặc nhiệm vụ từ công ty, xã hội hoặc phát sinh trong quá trình thực hiện một dự án
    - Tìm danh sách khách hàng tiềm năng đang cần vay vốn
    - Từ tag người vào ảnh có trước (facebook)
    - Dự đoán khả năng bị bệnh của một người
    - Hệ thống recommend mua hàng cho khách hàng (Amazon)
    - Đánh giá hoạt động sản xuất, kinh doanh của các cửa hàng
    - ...

# B1: Thiết lập mục tiêu nghiên cứu (4)

- **Ví dụ**
  - Mục tiêu: Phân tích hoạt động kinh doanh của chuỗi cửa hàng điện tử trong năm 2019?
  - Trả lời các câu hỏi:
    - Phân tích hoạt động kinh doanh theo từng tháng, xác định tháng nào trong năm có doanh số bán hàng cao nhất – thấp nhất?
    - Cửa hàng nào hoạt động hiệu quả trong năm 2019?
    - Thói quen của khách hàng?
    - Sản phẩm nào được bán chạy nhất?



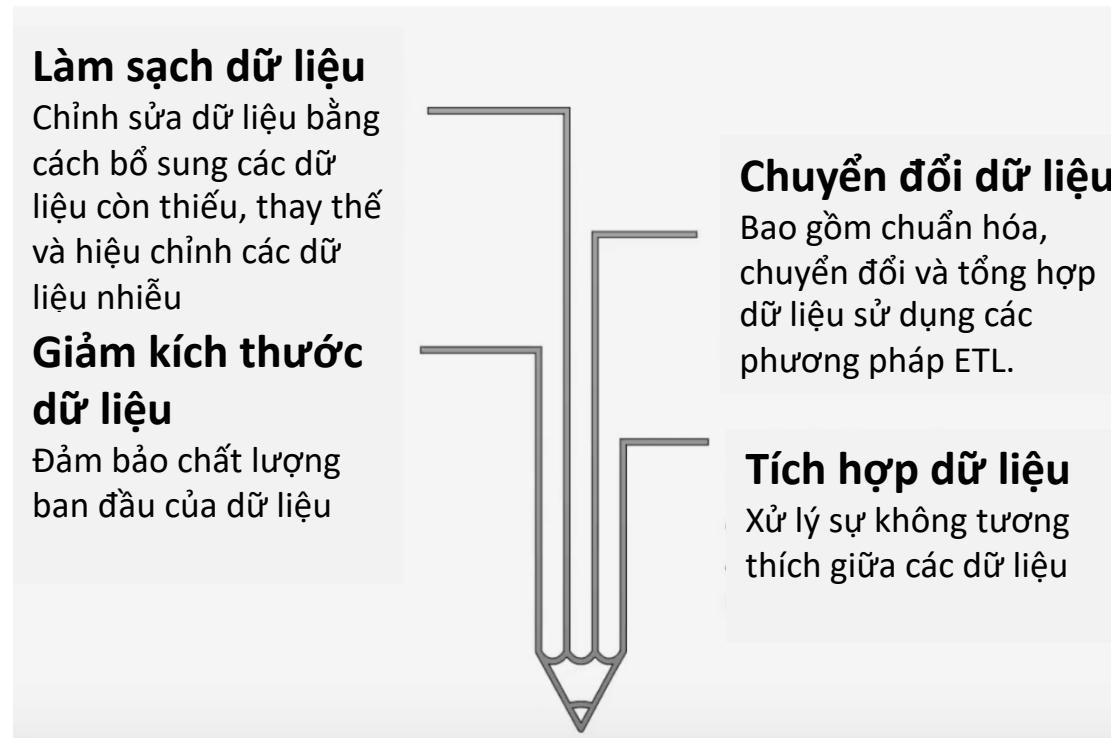
## B2: Thu thập dữ liệu

- **Thu thập dữ liệu (Data collection)**
  - Mục tiêu: thu thập tất cả các dữ liệu cần cho dự án
  - Dữ liệu có thể được lưu trữ ở nhiều dạng, từ các tệp văn bản đơn giản đến các bảng trong cơ sở dữ liệu
  - **Thu thập toàn bộ dữ liệu bán hàng tại tất cả các cửa hàng trong năm 2019**



### B3. Chuẩn bị dữ liệu

- Dữ liệu thu thập được đều là dữ liệu thô, chứa rất nhiều nhiễu (noise), dữ liệu thiếu, dữ liệu sai định dạng, không nhất quán....Dữ liệu này không thể phân tích được → Cần phải được tiền xử lý và làm sạch.
- Là bước **quan trọng**, chiếm **nhiều thời gian** và **nguồn lực nhất** trong bất kỳ một dự án khoa học dữ liệu nào (80%)



# B3: Chuẩn bị dữ liệu (2)

## • Chuẩn bị dữ liệu

	A	B	C	D	E	F
9137	149941	Wired Headphones	2	11.99	01/26/19 09:59	25 Jefferson St, Portland, OR 97035
9138	149942	Macbook Pro Laptop	1	1700	2001-05-19 8:13	122 River St, Boston, MA 02215
9139	149943	AAA Batteries (4-pack)	1	2.99	01/25/19 02:30	918 12th St, San Francisco, CA 94016
9140	149944	20in Monitor	1	109.99	2001-07-19 21:34	653 Church St, New York City, NY 10001
9141	149945	AAA Batteries (4-pack)	2	2.99	01/15/19 20:40	110 Main St, Atlanta, GA 30301
9142	149946	AA Batteries (4-pack)	1	3.84	01/17/19 16:24	836 River St, New York City, NY 10001
9143	149947	Apple Airpods Headphones	1	150	01/31/19 17:37	265 Cedar St, Seattle, WA 98101
9144	149948	Bose SoundSport Headphones	1	99.99	2001-09-19 0:05	833 Hickory St, New York City, NY 10001
9145	149949	USB-C Charging Cable	1	11.95	01/15/19 13:40	599 Cedar St, San Francisco, CA 94016
9146						
9147	149950	Apple Airpods Headphones	1	150	01/25/19 12:49	174 Jackson St, Atlanta, GA 30301
9148	149951	USB-C Charging Cable	1	11.95	01/25/19 11:34	660 Adams St, Austin, TX 73301
9149	149952	20in Monitor	1	109.99	2001-06-19 18:10	200 Walnut St, Los Angeles, CA 90001

Dữ liệu bị thiếu  
(missing)

Dữ liệu không hợp lệ  
(Improper)

	A	B	C	D	E	F
9308	150101	USB-C Charging Cable	1	11.95	01/28/19 16:06	393 Sunset St, Seattle, WA 98101
9309	150102	Wired Headphones	1	11.99	2001-04-19 16:44	223 4th St, San Francisco, CA 94016
9310	150103	AA Batteries (4-pack)	2	3.84	2001-06-19 19:29	121 South St, Dallas, TX 75001
9311	150104	USB-C Charging Cable	1	11.95	01/26/19 15:45	748 Cedar St, New York City, NY 10001
9312	150105	AAA Batteries (4-pack)	1	2.99	2001-05-19 12:56	11 7th St, San Francisco, CA 94016
	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
9314	150106	AA Batteries (4-pack)	1	3.84	01/24/19 11:59	818 West St, New York City, NY 10001
9315	150107	Wired Headphones	1	11.99	2001-06-19 17:56	692 Forest St, Dallas, TX 75001
9316	150108	34in Ultrawide Monitor	1	379.99	2001-06-19 13:39	638 Willow St, Boston, MA 02215
9317	150109	AA Batteries (4-pack)	1	3.84	2001-06-19 16:08	570 Spruce St, San Francisco, CA 94016

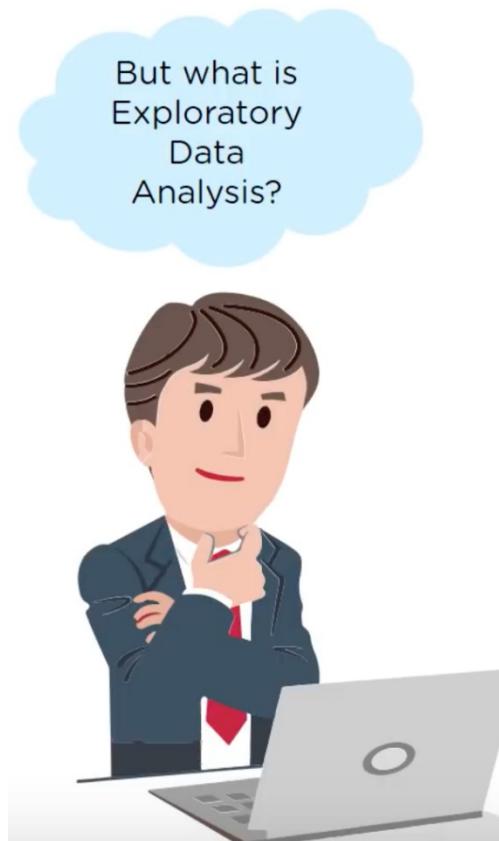
## B4: Khám phá dữ liệu



- Khám phá dữ liệu để **hiểu rõ** hơn về **mối quan hệ giữa các biến** và nhận biết được các **thông tin** được truyền tải **từ dữ liệu**.
- Lựa chọn các **mô hình phù hợp** và các **biến quan trọng** để đưa vào mô hình.

# B4: Khám phá dữ liệu (2)

## Phân tích dữ liệu khai phá



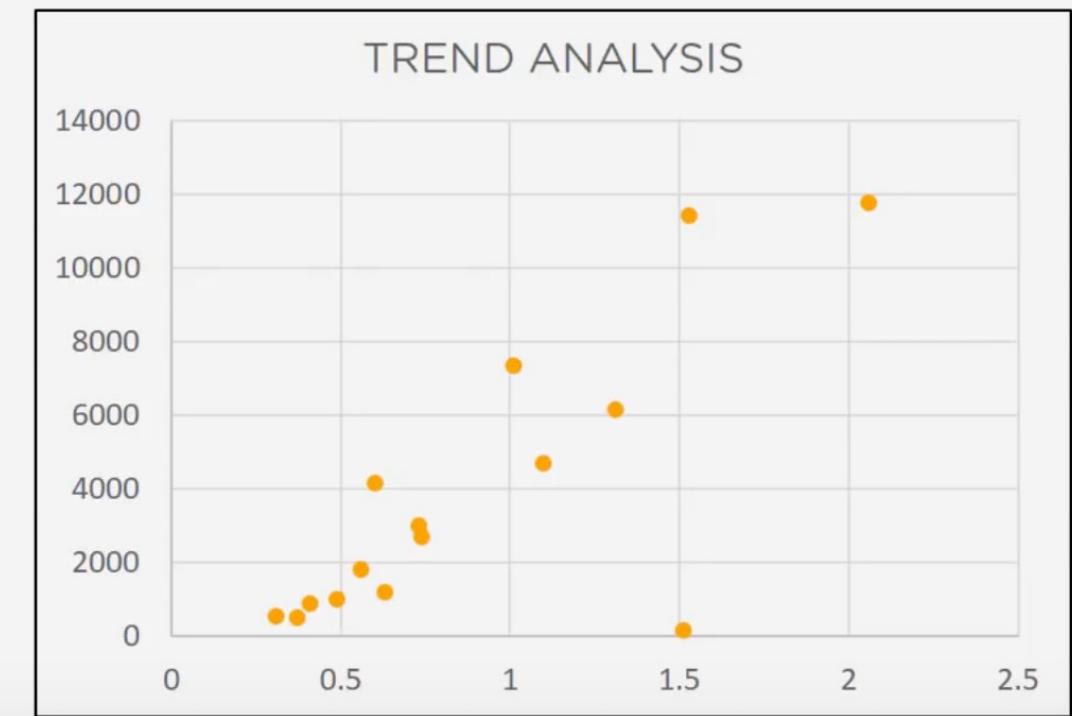
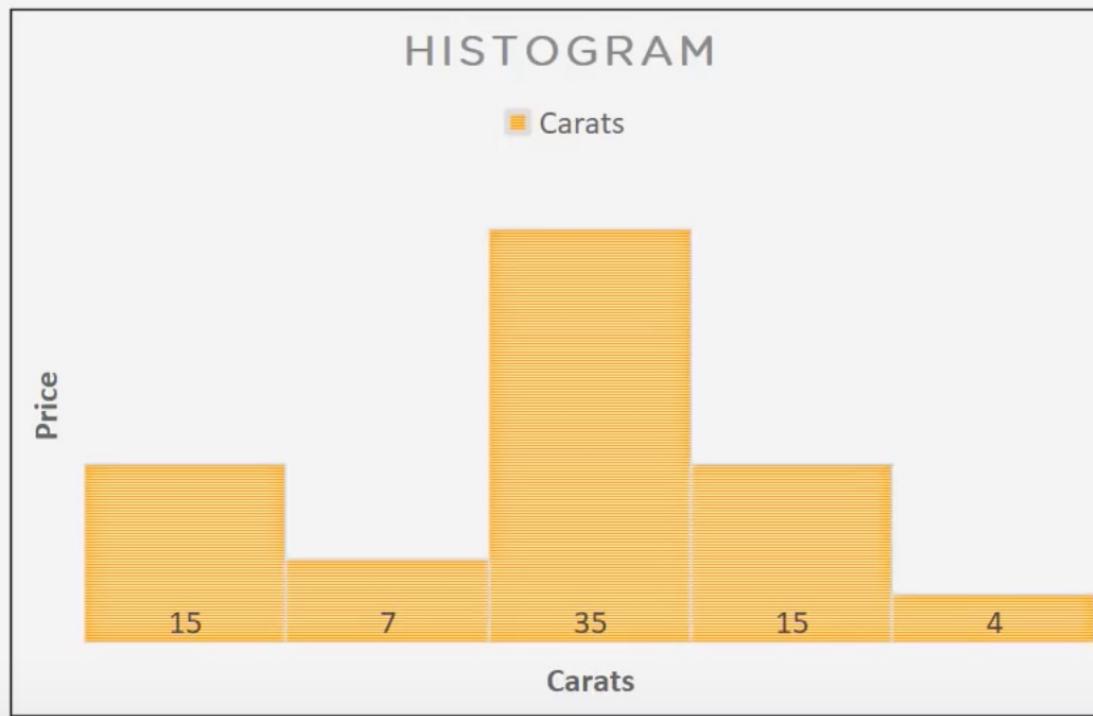
Exploratory Data Analysis (EDA) là phương pháp phân tích dữ liệu chủ yếu sử dụng kỹ thuật về biểu đồ, đồ thị.

Mục tiêu:

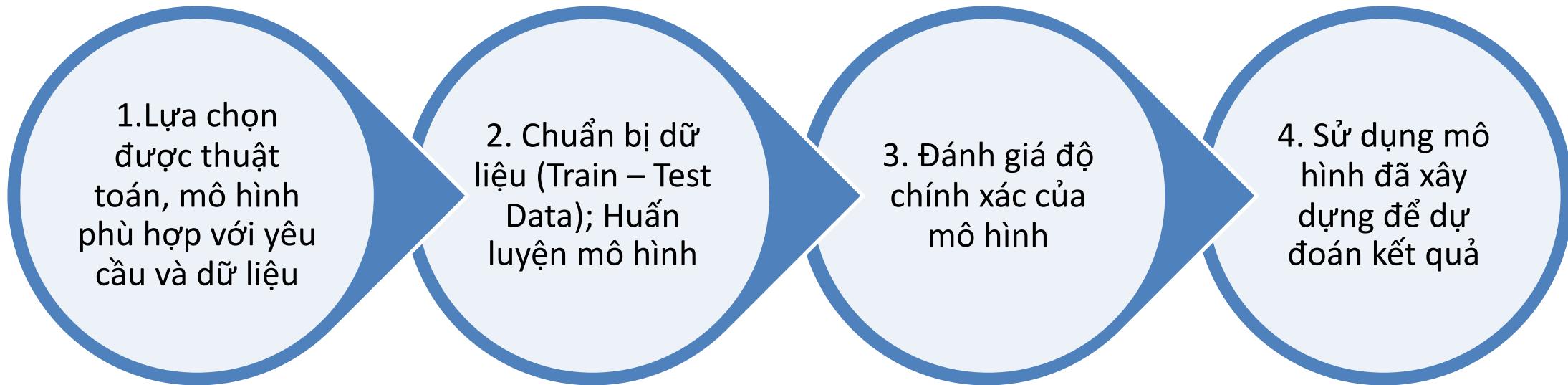
- Hiểu về kiểu dữ liệu và trả lời các câu hỏi về dữ liệu.
- Hiểu rõ cách thức phân phối dữ liệu
- Xác định các trường hợp ngoại lệ (errors)
- Xác định các quy luật có trong dữ liệu (pattern)

## B4: Khám phá dữ liệu (3)

- Phân tích khám phá dữ liệu EDA
  - Histogram
  - Trend Analysis



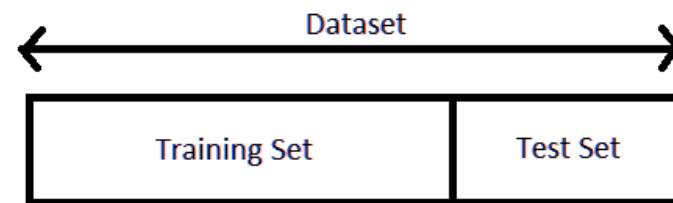
## B5: Xây dựng mô hình



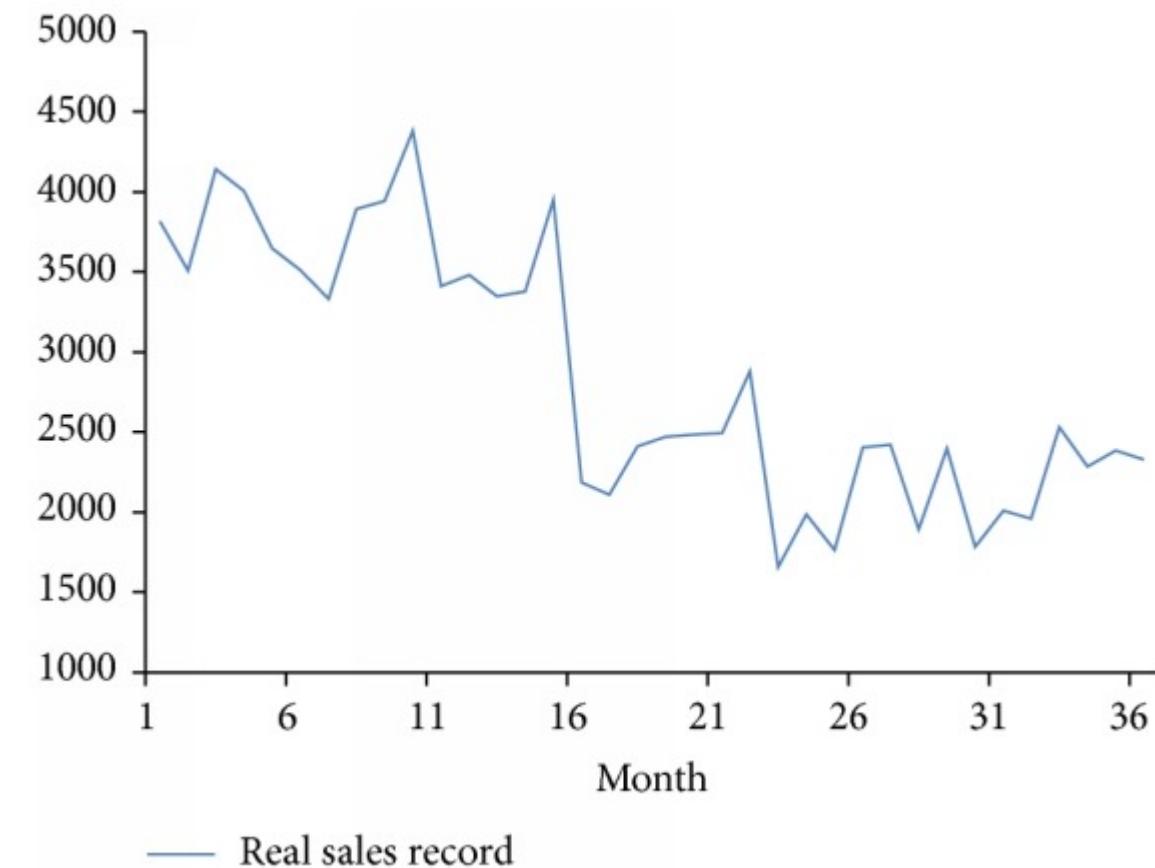
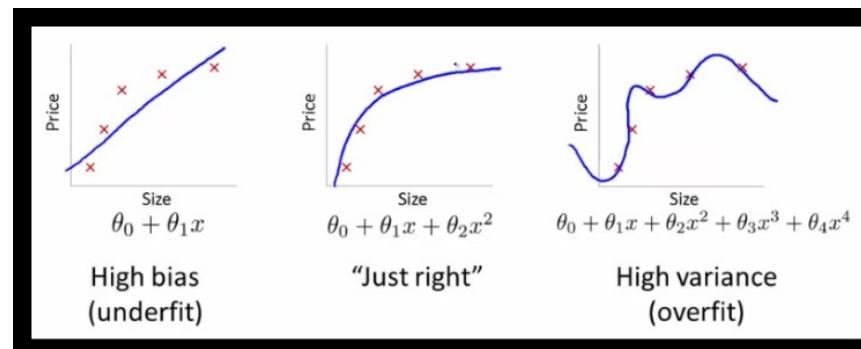
# B5: Xây dựng mô hình (2)

## Lựa chọn thuật toán, mô hình:

- Dựa vào dữ liệu của bài toán cụ thể, để lựa chọn được thuật toán phù hợp.
  - Dữ liệu mẫu (train data) được sử dụng để chạy mô hình.



## Huấn luyện mô hình:



# B6: Trình diễn kết quả



Trình bày kết quả của dự án với khách hàng, công ty



hợp với các công cụ (ứng dụng) khác



**Ví dụ:**  
**Thực hiện dự  
án phân tích  
tập dữ liệu  
kinh doanh**



# Bước 1. Xác định mục tiêu của dự án

Công ty bao gồm 10 cửa hàng kinh doanh đồ điện tử, được đặt tại 10 thành phố của 8 bang nước Mỹ.

1. Austin (TX - Texas)
2. Dallas (TX - Texas)
3. San Francisco (CA - California)
4. Los Angeles (CA - California)
5. New York City (NY - New York)
6. Portland (ME - Maine)
7. Portland (OR - Oregon)
8. Atlanta (GA - Georgia)
9. Seattle (WA - Washington)
10. Boston (MA - Massachusetts)



# Bước 1. Xác định mục tiêu của dự án (t)

## Mục tiêu:

- Đánh giá hoạt động kinh doanh của chuỗi cửa hàng trong năm 2019
- Thực hiện phân tích tìm ra các thông tin có ích (insights) từ dữ liệu để cải thiện, nâng cao hoạt động kinh doanh tại các cửa hàng

**Câu hỏi 1:** Cửa hàng ở thành phố nào hoạt động hiệu quả nhất?

**Câu hỏi 2:** Tháng nào trong năm có doanh số bán hàng cao nhất - thấp nhất?

**Câu hỏi 3:** Khách hàng mua sản phẩm thường tập trung vào khung giờ nào trong ngày?

**Câu hỏi 4:** Khách hàng thường mua các Sản phẩm nào cùng với nhau?

**Câu hỏi 5:** Sản phẩm nào bán được số lượng nhiều nhất? Tại sao?



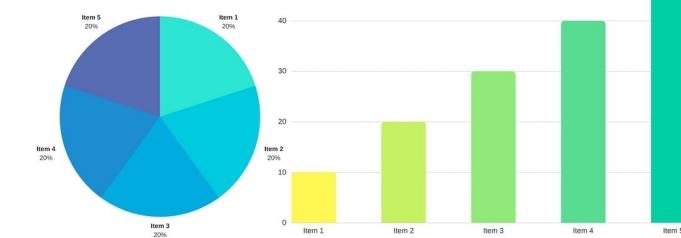
## Bước 2. Thu thập và mô tả dữ liệu

Dữ liệu bán hàng bao gồm 12 file của chuỗi cửa hàng kinh doanh này; Mỗi file dữ liệu lưu lại toàn bộ các đơn hàng đã bán được theo từng tháng tương ứng của năm 2019.

- \* Dữ liệu tháng 1: Sales\_January\_2019.csv
- \* Dữ liệu tháng 2: Sales\_February\_2019.csv
- \* Dữ liệu tháng 3: Sales\_March\_2019.csv
- \* Dữ liệu tháng 4: Sales\_April\_2019.csv
- \* Dữ liệu tháng 5: Sales\_May\_2019.csv
- \* Dữ liệu tháng 6: Sales\_June\_2019.csv
- \* Dữ liệu tháng 7: Sales\_July\_2019.csv
- \* Dữ liệu tháng 8: Sales\_August\_2019.csv
- \* Dữ liệu tháng 9: Sales\_September\_2019.csv
- \* Dữ liệu tháng 10: Sales\_October\_2019.csv
- \* Dữ liệu tháng 11: Sales\_November\_2019.csv
- \* Dữ liệu tháng 12: Sales\_December\_2019.csv



### SALES ANALYSIS



## Bước 2. Thu thập và mô tả dữ liệu (t)

Mỗi file dữ liệu bao gồm 6 thông tin:

1. Order ID: Mã đơn hàng
2. Product: Tên sản phẩm đã bán theo từng đơn hàng
3. Quantity Ordered: Số lượng sản phẩm bán
4. Price Each: Giá của mỗi sản phẩm
5. Order Date: Thời gian bán hàng
6. Purchase Address: Địa chỉ mua hàng



Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
141234	iPhone	1	700	01/22/19 21:25	944 Walnut St, Boston, MA 02215
141235	Lightning Charging Cable	1	14.95	01/28/19 14:15	185 Maple St, Portland, OR 97035
141236	Wired Headphones	2	11.99	01/17/19 13:33	538 Adams St, San Francisco, CA 94016
141237	27in FHD Monitor	1	149.99	01/05/19 20:33	738 10th St, Los Angeles, CA 90001
141238	Wired Headphones	1	11.99	01/25/19 11:59	387 10th St, Austin, TX 73301

(Nếu hóa đơn mua nhiều sản phẩm, có nhiều bản ghi với cùng mã Order ID)

# Bước 3. Chuẩn bị dữ liệu

Tập dữ liệu tổng hợp là tập dữ liệu thô cần được chuẩn hóa và làm sạch:

1. Xử lý các dòng dữ liệu rỗng, không có giá trị trong tập dữ liệu tổng hợp.
2. Xử lý các dòng tiêu đề bị lặp lại
3. Chuyển đổi kiểu dữ liệu các cột số
4. Bổ sung thêm các cột dữ liệu mới phục vụ phân tích

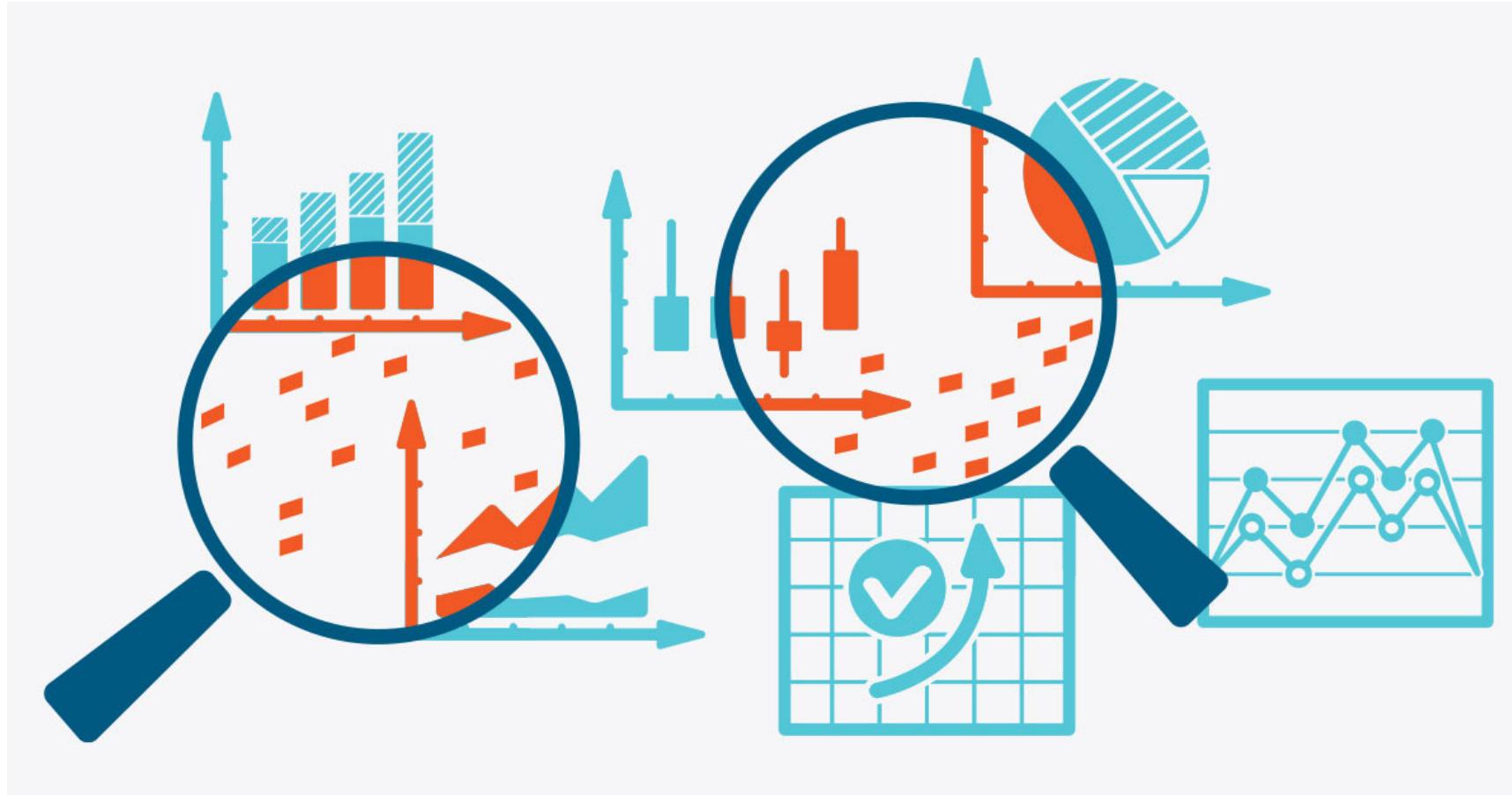


# Bước 3. chuẩn bị dữ liệu (t)

Tập dữ liệu sau khi được tiền chuẩn hóa và làm sạch:

Order ID	Product	Quantity	Price Each	Order Date	Purchase Address	Month	Hour	City	Sales
141234	iPhone	1	700	2019-01-22 21:25	944 Walnut St, Boston, MA 02215	1	21	Boston (MA)	700
141235	Lightning Charging Cable	1	14.95	2019-01-28 14:15	185 Maple St, Portland, OR 97035	1	14	Portland (OR)	14.95
141236	Wired Headphones	2	11.99	2019-01-17 13:33	538 Adams St, San Francisco, CA 94016	1	13	San Francisco (CA)	23.98
141237	27in FHD Monitor	1	149.99	2019-01-05 20:33	738 10th St, Los Angeles, CA 90001	1	20	Los Angeles (CA)	149.99
141238	Wired Headphones	1	11.99	2019-01-25 11:59	387 10th St, Austin, TX 73301	1	11	Austin (TX)	11.99
141239	AAA Batteries (4-pack)	1	2.99	2019-01-29 20:22	775 Willow St, San Francisco, CA 94016	1	20	San Francisco (CA)	2.99
141240	27in 4K Gaming Monitor	1	389.99	2019-01-26 12:16	979 Park St, Los Angeles, CA 90001	1	12	Los Angeles (CA)	389.99
141241	USB-C Charging Cable	1	11.95	2019-01-05 12:04	181 6th St, San Francisco, CA 94016	1	12	San Francisco (CA)	11.95
141242	Bose SoundSport Headphones	1	99.99	2019-01-01 10:30	867 Willow St, Los Angeles, CA 90001	1	10	Los Angeles (CA)	99.99
141243	Apple Airpods Headphones	1	150	2019-01-22 21:20	657 Johnson St, San Francisco, CA 94016	1	21	San Francisco (CA)	150
141244	Apple Airpods Headphones	1	150	2019-01-07 11:29	492 Walnut St, San Francisco, CA 94016	1	11	San Francisco (CA)	150
141245	Macbook Pro Laptop	1	1700	2019-01-31 10:12	322 6th St, San Francisco, CA 94016	1	10	San Francisco (CA)	1700
141246	AAA Batteries (4-pack)	3	2.99	2019-01-09 18:57	618 7th St, Los Angeles, CA 90001	1	18	Los Angeles (CA)	8.97
141247	27in FHD Monitor	1	149.99	2019-01-25 19:19	512 Wilson St, San Francisco, CA 94016	1	19	San Francisco (CA)	149.99
141248	Flatscreen TV	1	300	2019-01-03 21:54	363 Spruce St, Austin, TX 73301	1	21	Austin (TX)	300
141249	27in FHD Monitor	1	149.99	2019-01-05 17:20	440 Cedar St, Portland, OR 97035	1	17	Portland (OR)	149.99
141250	Vareebadd Phone	1	400	2019-01-10 11:20	471 Center St, Los Angeles, CA 90001	1	11	Los Angeles (CA)	400
141251	Apple Airpods Headphones	1	150	2019-01-24 8:13	414 Walnut St, Boston, MA 02215	1	8	Boston (MA)	150
141252	USB-C Charging Cable	1	11.95	2019-01-30 9:28	220 9th St, Los Angeles, CA 90001	1	9	Los Angeles (CA)	11.95
141253	AA Batteries (4-pack)	1	3.84	2019-01-17 0:09	385 11th St, Atlanta, GA 30301	1	0	Atlanta (GA)	3.84
141254	AAA Batteries (4-pack)	1	2.99	2019-01-08 11:51	238 Sunset St, Seattle, WA 98101	1	11	Seattle (WA)	2.99
141255	USB-C Charging Cable	1	11.95	2019-01-09 20:55	764 11th St, Los Angeles, CA 90001	1	20	Los Angeles (CA)	11.95
141256	Google Phone	1	600	2019-01-29 10:40	675 Washington St, Portland, OR 97035	1	10	Portland (OR)	600
141257	Apple Airpods Headphones	1	150	2019-01-12 18:51	338 Highland St, San Francisco, CA 94016	1	18	San Francisco (CA)	150

## Bước 4. Phân tích, đánh giá hoạt động kinh doanh



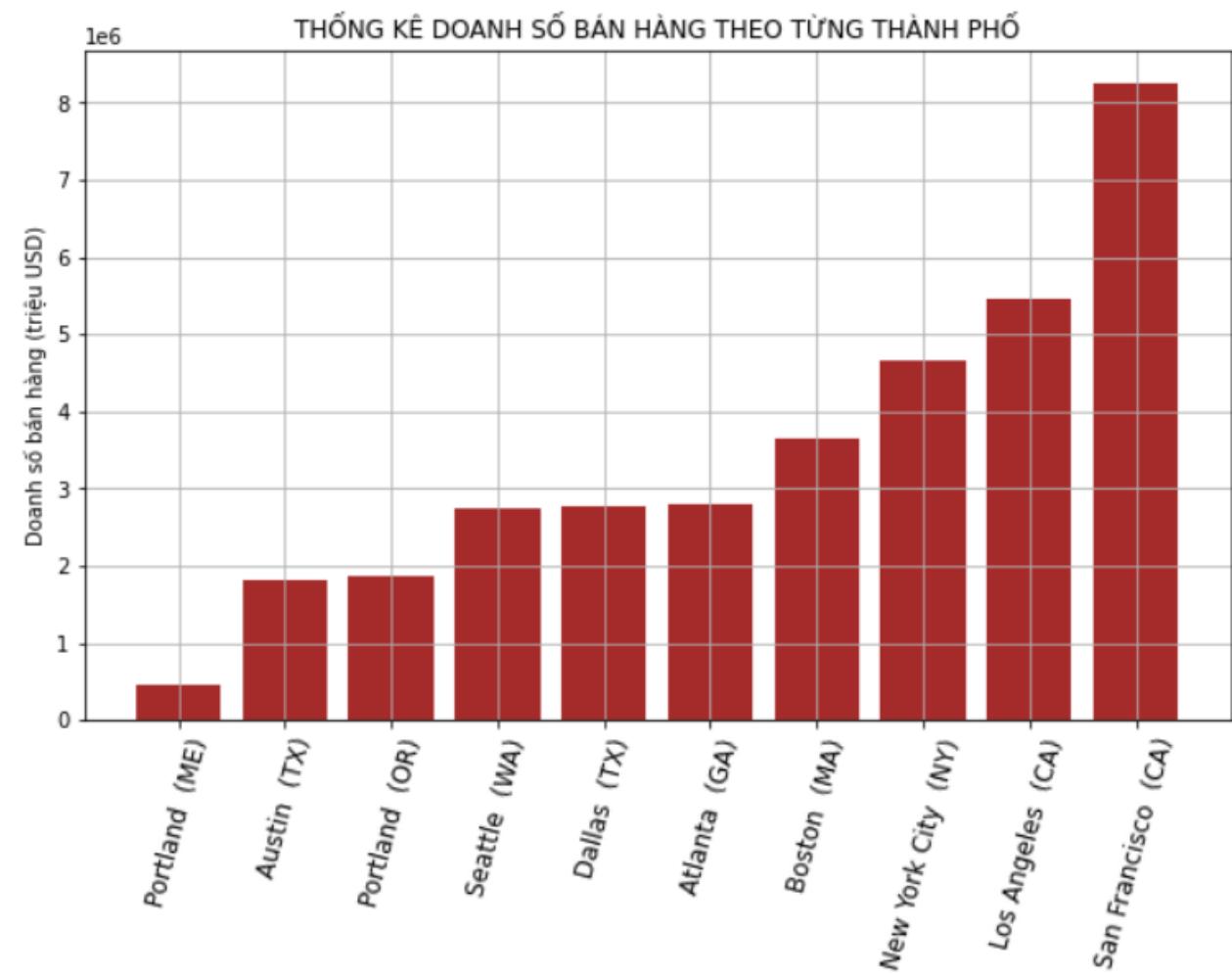
[Chi tiết các bước thực hiện trong file jupyter notebook...](#)

# Bước 4. Phân tích, đánh giá hoạt động kinh doanh

Câu hỏi 1: Cửa hàng ở thành phố nào bán được hàng nhiều nhất?

Sales	City
4.497583e+05	Portland (ME)
1.819582e+06	Austin (TX)
1.870732e+06	Portland (OR)
2.747755e+06	Seattle (WA)
2.767975e+06	Dallas (TX)
2.795499e+06	Atlanta (GA)
3.661642e+06	Boston (MA)
4.664317e+06	New York City (NY)
5.452571e+06	Los Angeles (CA)
8.262204e+06	San Francisco (CA)

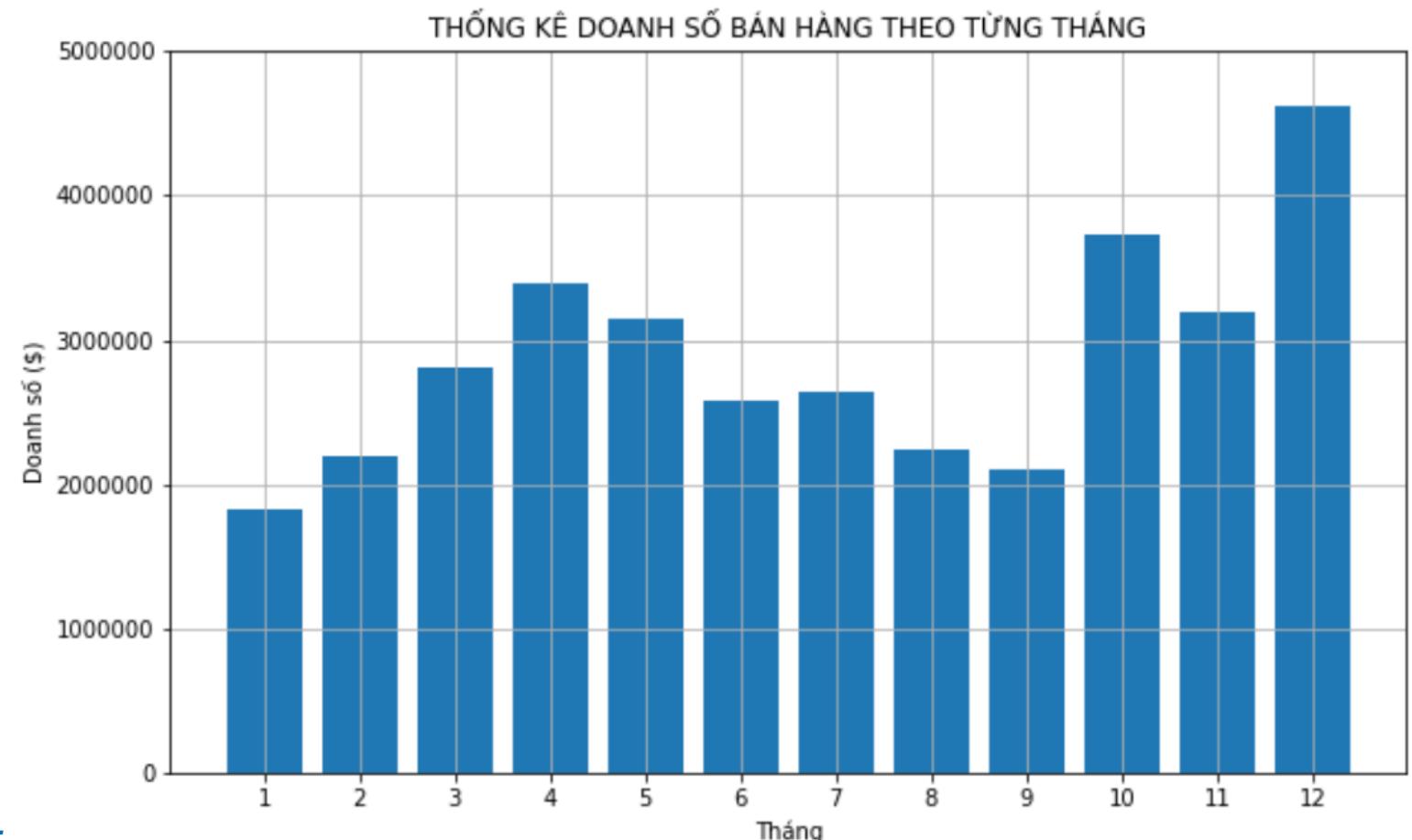
W  
Insight!



[Chi tiết các bước thực hiện trong file jupyter notebook...](#)

# Bước 4. Phân tích, đánh giá hoạt động kinh doanh

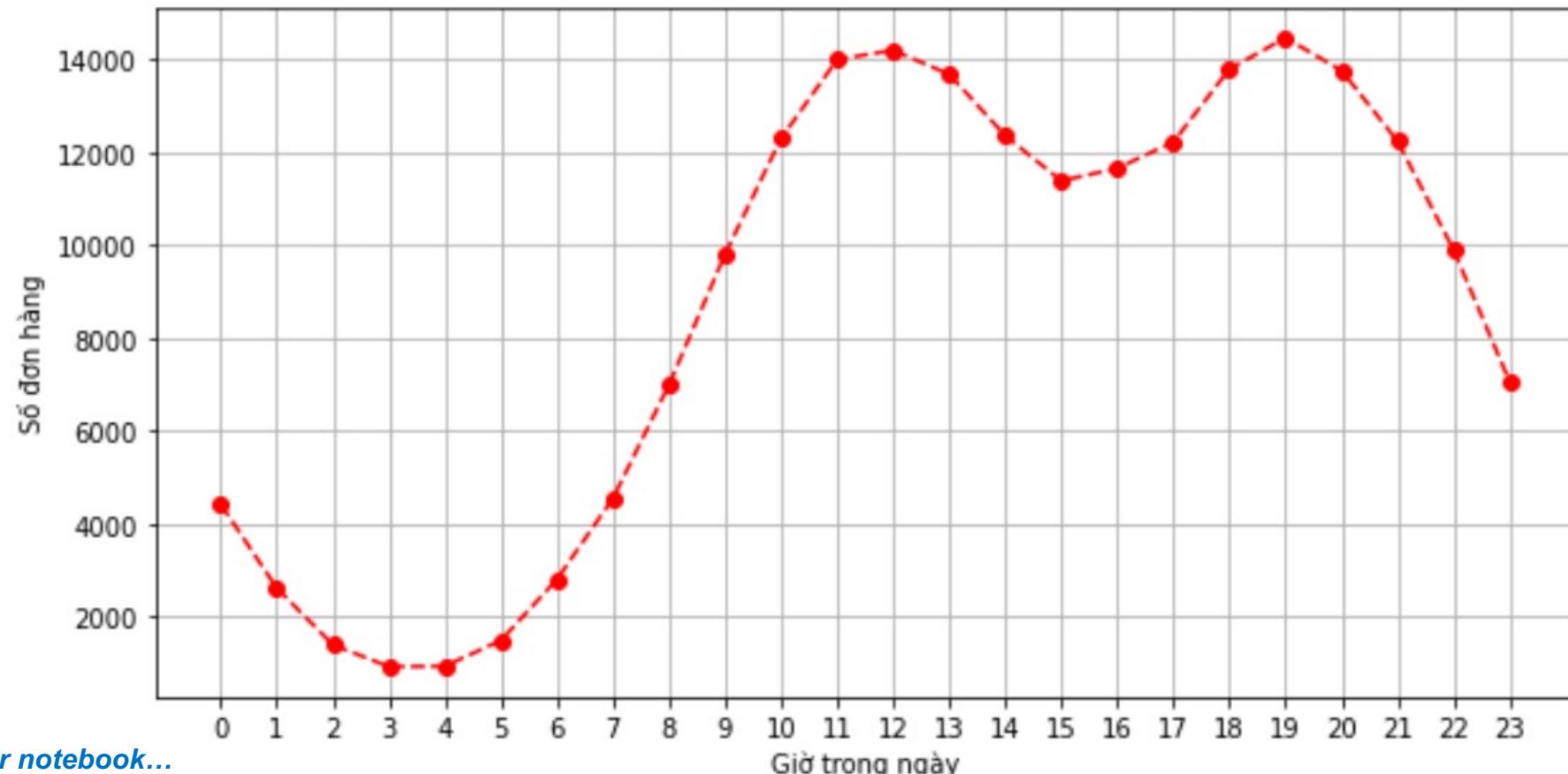
Câu hỏi 2: Tháng nào trong năm có doanh số bán hàng cao nhất - thấp nhất?



[Chi tiết các bước thực hiện trong file jupyter notebook...](#)

# Bước 4. Phân tích, đánh giá hoạt động kinh doanh

Câu hỏi 3: Khách hàng mua sản phẩm thường tập trung vào khung giờ nào trong ngày?



[Chi tiết các bước thực hiện trong file jupyter notebook...](#)

# Bước 4. Phân tích, đánh giá hoạt động kinh doanh

Câu hỏi 4: Trong các hóa đơn của Khách hàng, Sản phẩm nào thường được mua cùng với nhau?



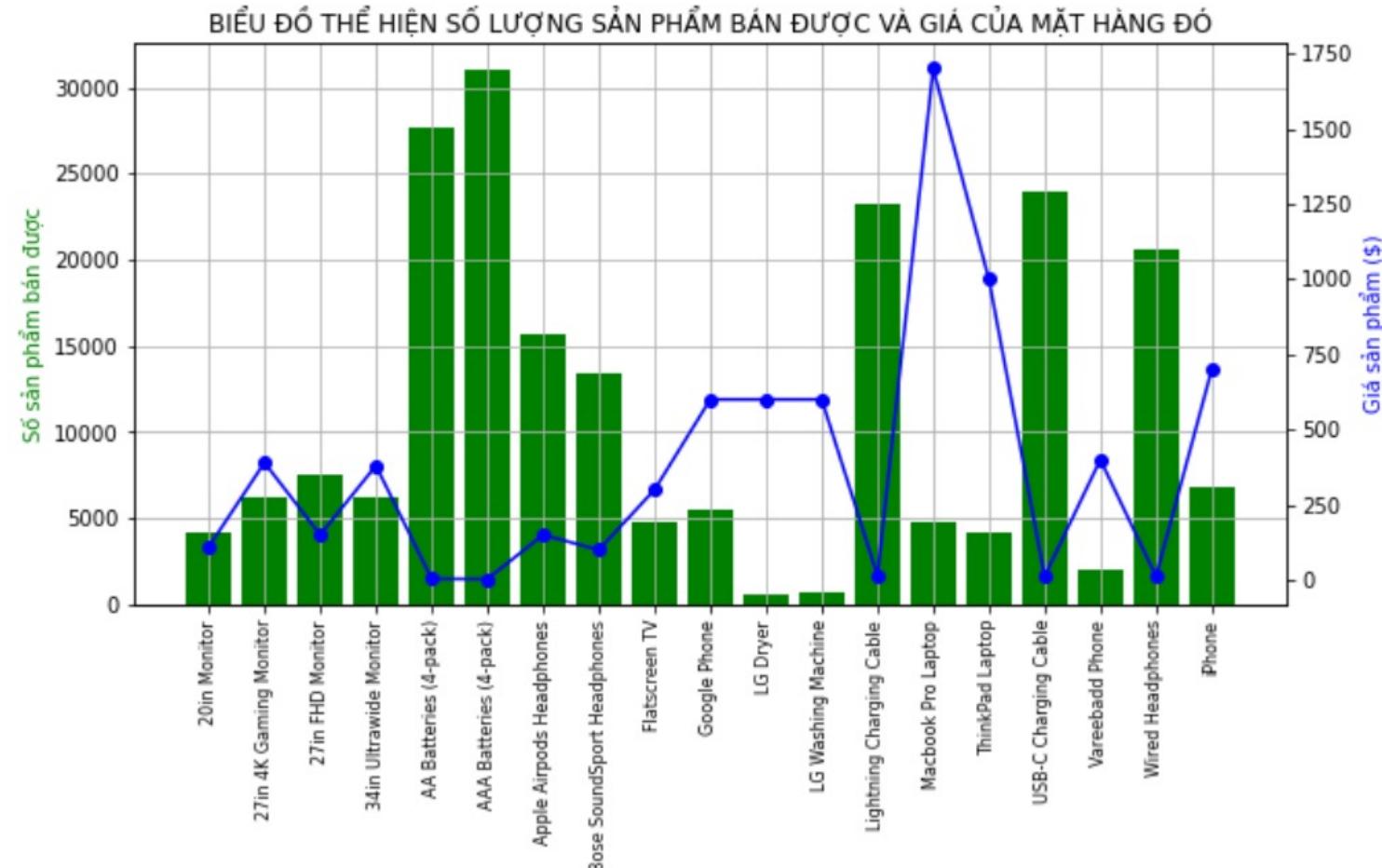
```
('iPhone', 'Lightning Charging Cable') 1005  
(('Google Phone', 'USB-C Charging Cable') 987  
(('iPhone', 'Wired Headphones') 447  
(('Google Phone', 'Wired Headphones') 414  
(('Vareebadd Phone', 'USB-C Charging Cable') 361  
(('iPhone', 'Apple Airpods Headphones') 360  
(('Google Phone', 'Bose SoundSport Headphones') 220  
(('USB-C Charging Cable', 'Wired Headphones') 160  
(('Vareebadd Phone', 'Wired Headphones') 143  
(('Lightning Charging Cable', 'Wired Headphones') 92  
(('Lightning Charging Cable', 'Apple Airpods Headphones') 81  
(('Vareebadd Phone', 'Bose SoundSport Headphones') 80  
(('USB-C Charging Cable', 'Bose SoundSport Headphones') 77  
(('Apple Airpods Headphones', 'Wired Headphones') 69  
(('Lightning Charging Cable', 'USB-C Charging Cable') 58
```

[Chi tiết các bước thực hiện trong file jupyter notebook...](#)

# Bước 4. Phân tích, đánh giá hoạt động kinh doanh

Câu hỏi 5: Sản phẩm nào của chuỗi cửa hàng bán được số lượng nhiều nhất?  
Tại sao?

Insight!

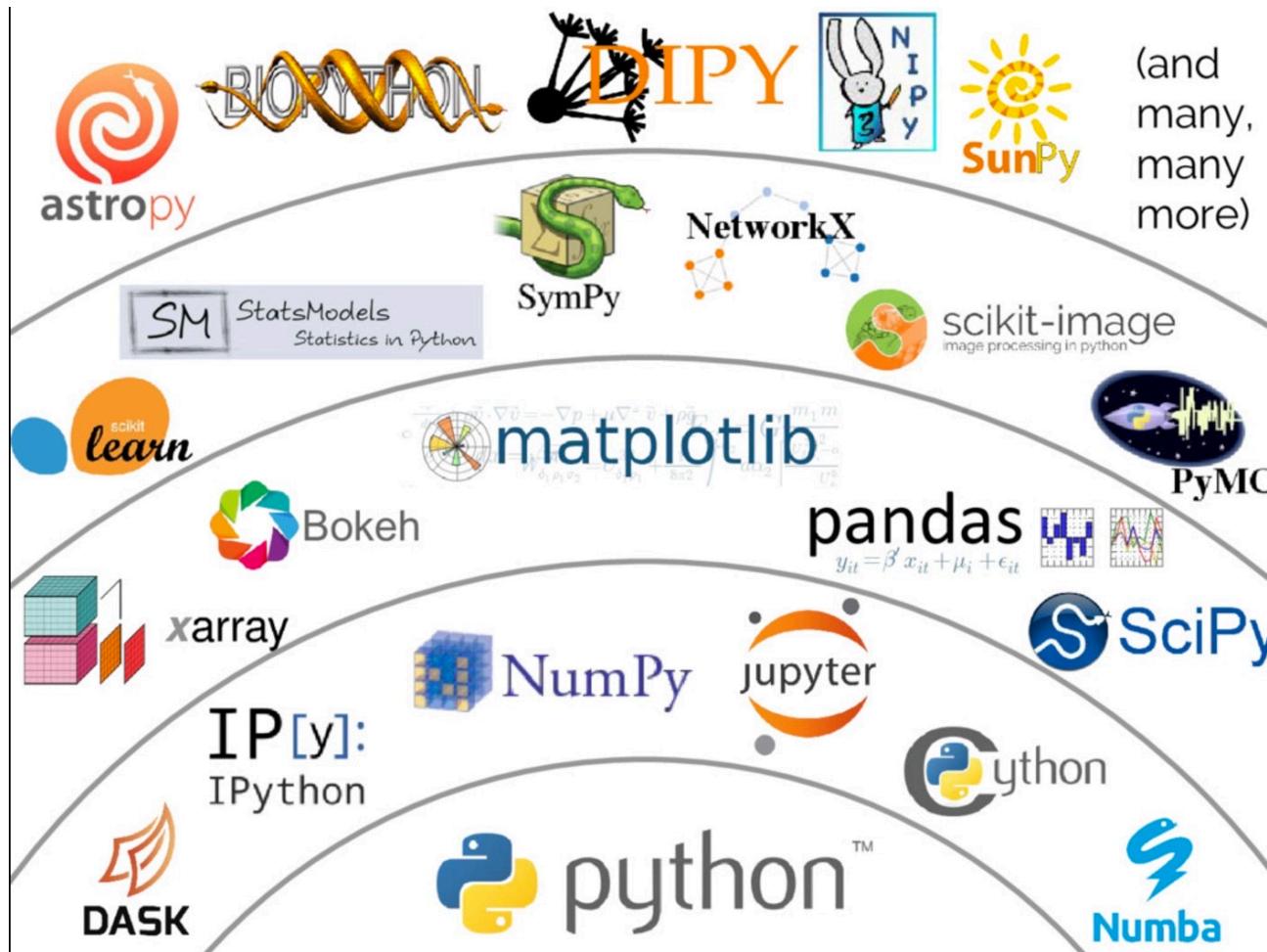


Chi tiết các bước thực hiện trong file jupyter notebook...

## 2. Một số thư viện quan trọng trong KHDL

# Thư viện Python

- Python có hệ thống thư viện rất phong phú, hỗ trợ nhiều lĩnh vực khác nhau.



- Do đó, tùy thuộc vào lĩnh vực nghiên cứu cụ thể, để lựa chọn và sử dụng các thư viện cho phù hợp.

# Thư viện Python cho khoa học dữ liệu

## Top 5 Python Libraries for Data Science



Report from Cloud Academy suggests that the top technical skill in demand for data engineers is python. 67 percent of job posts mentioned python.

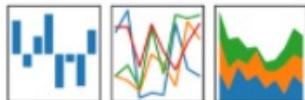


### 1.) NUMPY

Through NumPy, you can use it as an efficient multi-dimensional container of generic data. It also contains sophisticated (broadcasting) functions, tools for integrating C/C++ and Fortran code, useful linear algebra, Fourier transform, and random number capabilities

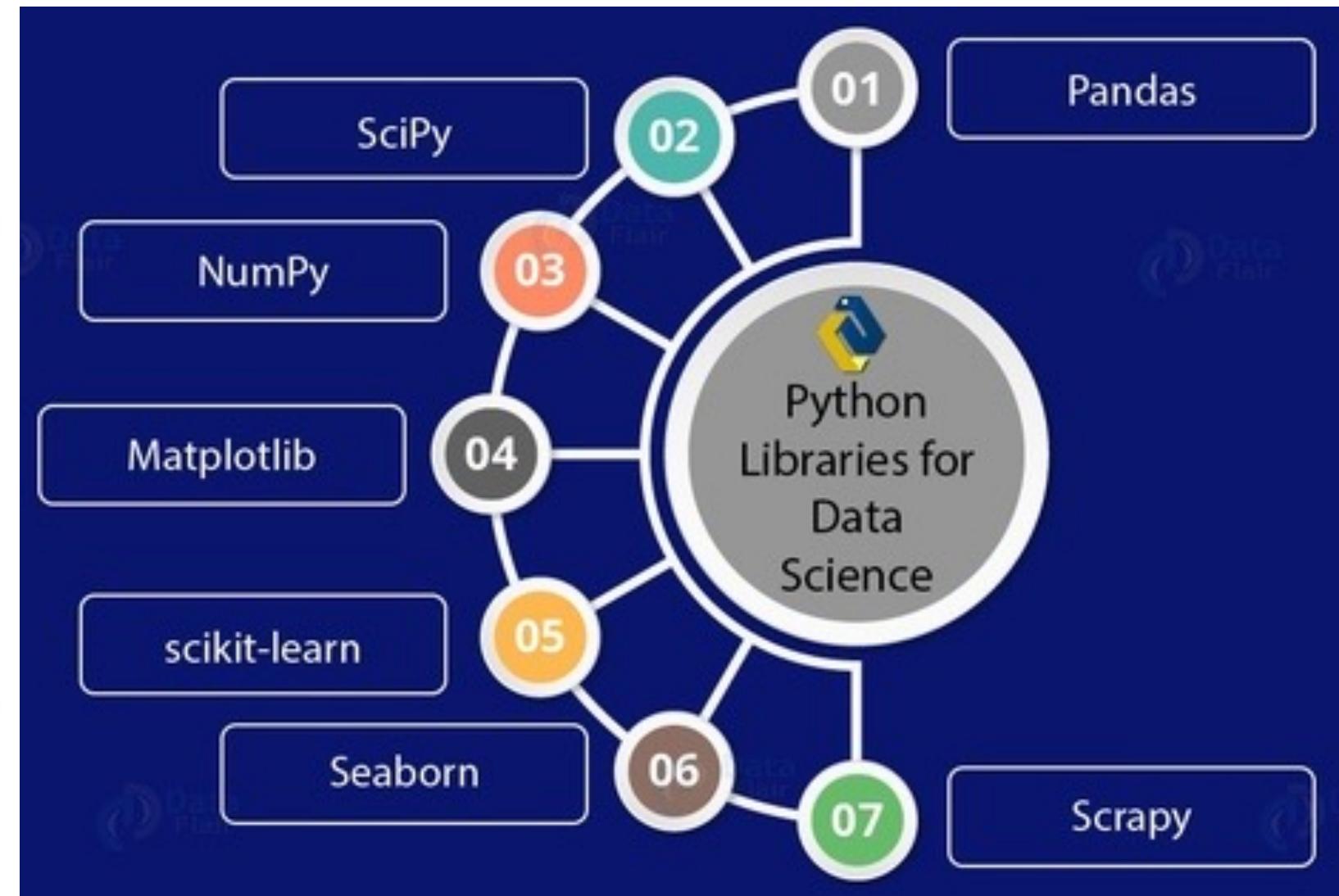
### 2.) PANDAS

Pandas provides high-performance, easy-to-use data structures and data analysis tools for python. You can store and manage data from tables by performing manipulation over rows and columns.



### 3.) SCIKIT-LEARN

Scikit-Learn is a powerful library for machine learning in Python. It contains simple and efficient tools for data mining and data analysis. It is built on NumPy, SciPy and matplotlib.



# Thư viện Python cho khoa học dữ liệu



# Thư viện Python cho khoa học dữ liệu

01

**Pandas:** Sử dụng cho quản lý và tương tác với dữ liệu có cấu trúc, được sử dụng rộng rãi trong việc thu thập và tiền xử lý dữ liệu.

02

**SciPy:** Dựa trên Numpy, cung cấp các công cụ mạnh cho khoa học và kỹ nghệ, như biến đổi fourier rời rạc, đại số tuyến tính, tối ưu hóa và ma trận thưa

03

**Numpy:** Thư viện chuyên xử lý dữ liệu số (nhiều chiều), bao gồm cả các hàm đại số tuyến tính cơ bản, biến đổi fourier, sinh số ngẫu nhiên nâng cao,...

04

**Matplotlib:** Thư viện này được sử dụng để trực quan hóa dữ liệu (Data Visualization), chuyên vẽ các biểu đồ, hỗ trợ rất nhiều loại biểu đồ khác nhau...

# Thư viện Python cho khoa học dữ liệu



**Scikit-learn:** Thư viện chuyên về học máy; thư viện này có sẵn nhiều công cụ hiệu quả cho học máy và thiết lập các mô hình thống kê như các thuật toán phân lớp, hồi quy, phân cụm và giảm chiều dữ liệu...

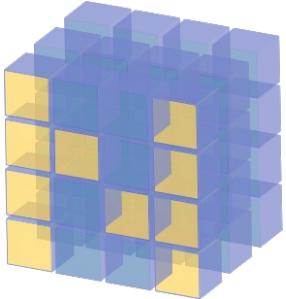


**Seaborn:** Thư viện này dựa trên Matplotlib, cung cấp các công cụ hiển thị dữ liệu một cách trực quan, hiệu quả. Mục tiêu của thư viện này là sử dụng việc trực quan hóa dữ liệu như là trọng tâm của khám phá và hiểu dữ liệu

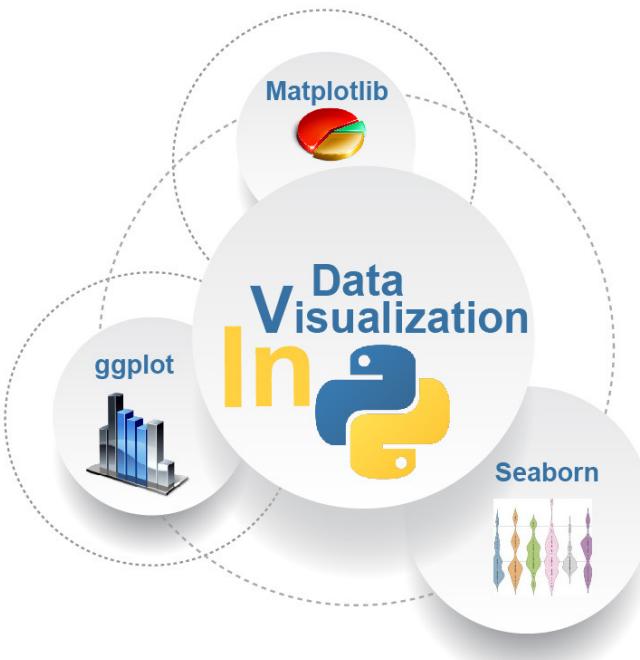


**Scrapy:** Thư viện này chuyên về việc thu thập thông tin trên Web, rất phù hợp với việc lấy các dữ liệu theo mẫu.

# Một số thư viện trong khóa học

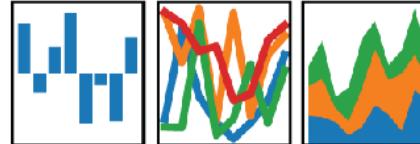


## NumPy



## pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



	BandName	WavelengthMax	WavelengthMin
0	CoastalAerosol	450	430
1	Blue	510	450
2	Green	590	530
3	Red	670	640
4	NearInfrared	880	850
5	ShortWaveInfrared_1	1650	1570
6	ShortWaveInfrared_2	2290	2110
7	Cirrus	1380	1360

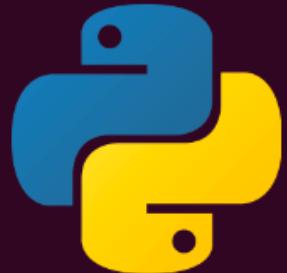
```
1 #Gọi các thư viện để sử dụng
2 import numpy as np
3 import pandas as pd
4 import matplotlib as mat
5 import matplotlib.pyplot as plt
6 print('Numpy: ',np.__version__)
7 print('Pandas: ',pd.__version__)
8 print('Matplotlib: ',mat.__version__)
```

Numpy: 1.19.5

Pandas: 1.2.4

Matplotlib: 3.3.4

# Một số thư viện trong khóa học



## How to Install Python PIP

```
1 #Cú pháp cài đặt: !pip install <tên thư viện>
2 #Cài đặt thư viện scrapy vào hệ thống:
3 !pip install --user scrapy
4 #hoặc
5 !pip install scrapy
```

```
1 #Cú pháp cài đặt: !pip install <tên thư viện>
2 #Cài đặt thư viện scrapy vào hệ thống:
3 !pip3 install --user scrapy
4 #hoặc
5 !pip3 install scrapy
```

```
1 #Update thư viện Lên phiên bản cao hơn:
2 !pip install --user <tên thư viện> --upgrade
```