

Đề thi:**PYTHON FOR MACHINE LEARNING, DATA SCIENCE AND VISUALIZATION**

Thời gian: 120 phút

Ngày thi: 24/09/2020

----- ĐƯỢC sử dụng tài liệu -----

- 1 Tạo DataFrame employees chứa danh sách các nhân viên. Danh sách này được đọc từ tập tin employees.xls
- 2 Xem các thông tin của employees và xem 5 nhân viên đầu tiên.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 107 entries, 0 to 106
Data columns (total 11 columns):
EMPLOYEE_ID      107 non-null int64
FIRST_NAME       107 non-null object
LAST_NAME        107 non-null object
EMAIL            107 non-null object
PHONE_NUMBER     107 non-null object
HIRE_DATE        107 non-null object
JOB_ID           107 non-null object
SALARY           107 non-null int64
COMMISSION_PCT   35 non-null float64
MANAGER_ID       106 non-null float64
DEPARTMENT_ID    106 non-null float64
dtypes: float64(3), int64(2), object(6)
memory usage: 6.7+ KB
```

	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE_NUMBER	HIRE_DATE	JOB_ID	SALARY	COMMISSION_PCT	MANAGER_ID	DEPARTMENT_ID
0	100	Steven	King	SKING	515.123.4567	2003-06-17 00:00:00	AD_PRES	24000	NaN	NaN	
1	101	Neena	Kochhar	NKOCHHAR	515.123.4568	2005-09-21 00:00:00	AD_VP	17000	NaN	100.0	
2	102	Lex	De Haan	LDEHAAN	515.123.4569	2001-01-13 00:00:00	AD_VP	17000	NaN	100.0	
3	103	Alexander	Hunold	AHUNOLD	590.423.4567	2006-01-03 00:00:00	IT_PROG	9000	NaN	102.0	
4	104	Bruce	Ernst	BERNST	590.423.4568	2007-05-21 00:00:00	IT_PROG	6000	NaN	103.0	

- 3 Tạo các DataFrame departments chứa danh sách các nhân viên và jobs chứa danh sách các công việc. Các danh sách này được đọc từ các tập tin departments.xls và job.xls

	DEPARTMENT_ID	DEPARTMENT_NAME	MANAGER_ID	LOCATION_ID
0	10	Administration	200.0	1700
1	20	Marketing	201.0	1800
2	30	Purchasing	114.0	1700
3	40	Human Resources	203.0	2400
4	50	Shipping	121.0	1500

	JOB_ID	JOB_TITLE	MIN_SALARY	MAX_SALARY
0	AC_ACCOUNT	Public Accountant	4200	9000
1	AC_MGR	Accounting Manager	8200	16000
2	AD_ASST	Administration Assistant	3000	6000
3	AD_PRES	President	20080	40000
4	AD_VP	Administration Vice President	15000	30000

- 4 Tạo DataFrame emp là danh sách các nhân viên của DataFrame employees chỉ với các cột theo mẫu sau:

```
emp.head()
```

	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	HIRE_DATE	SALARY	DEPARTMENT_ID	JOB_ID
0	100	Steven	King	2003-06-17 00:00:00	24000	90.0	AD_PRES
1	101	Neena	Kochhar	2005-09-21 00:00:00	17000	90.0	AD_VP
2	102	Lex	De Haan	2001-01-13 00:00:00	17000	90.0	AD_VP
3	103	Alexander	Hunold	2006-01-03 00:00:00	9000	60.0	IT_PROG
4	104	Bruce	Ernst	2007-05-21 00:00:00	6000	60.0	IT_PROG

- 5 Trong DataFrame emp hãy thêm vào hai cột: DEPARTMENT_NAME (tên phòng) và JOB_TITLE (tên công việc) theo mẫu sau:

```
emp.head()
```

	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	HIRE_DATE	SALARY	DEPARTMENT_ID	JOB_ID	DEPARTMENT_NAME	JOB_TITLE
0	100	Steven	King	2003-06-17 00:00:00	24000	90.0	AD_PRES	Executive	President
1	101	Neena	Kochhar	2005-09-21 00:00:00	17000	90.0	AD_VP	Executive	Administration Vice President
2	102	Lex	De Haan	2001-01-13 00:00:00	17000	90.0	AD_VP	Executive	Administration Vice President
3	103	Alexander	Hunold	2006-01-03 00:00:00	9000	60.0	IT_PROG	IT	Programmer
4	104	Bruce	Ernst	2007-05-21 00:00:00	6000	60.0	IT_PROG	IT	Programmer

Từ đây trở đi DataFrame emp sẽ được sử dụng để làm các câu hỏi tiếp theo

- 6 Liệt kê các nhân viên của các phòng 50 và 90, sắp giảm theo department_id và tăng theo salary.

	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	HIRE_DATE	SALARY	DEPARTMENT_ID	JOB_ID	DEPARTMENT_NAME	JOB_TITLE
	101	Neena	Kochhar	2005-09-21 00:00:00	17000	90.0	AD_VP	Executive	Administration Vice President
	102	Lex	De Haan	2001-01-13 00:00:00	17000	90.0	AD_VP	Executive	Administration Vice President
	100	Steven	King	2003-06-17 00:00:00	24000	90.0	AD_PRES	Executive	President
	132	TJ	Olson	2007-04-10 00:00:00	2100	50.0	ST_CLERK	Shipping	Stock Clerk
	128	Steven	Markle	2008-03-08 00:00:00	2200	50.0	ST_CLERK	Shipping	Stock Clerk



- 7 Liệt kê các nhân viên của phòng có department_name là Sales và có salary ≥ 5000 và salary ≤ 7000
- 8 Cho biết phòng nào có tổng lương lớn nhất

DEPARTMENT_ID	DEPARTMENT_NAME	SALARY
7	80.0	Sales 304500

- 9 Cho biết nhân viên không thuộc phòng ban nào cả

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	HIRE_DATE	SALARY	DEPARTMENT_ID	JOB_ID	DEPARTMENT_NAME	JOB_TITLE
78	178	Kimberely	Grant	2007-05-24 00:00:00	7000	NaN	SA_REP	NaN Sales Representative

- 10 In thông tin thống kê của salary

```
count      107.000000
mean      6461.831776
std       3909.579731
min       2100.000000
25%       3100.000000
50%       6200.000000
75%       8900.000000
max       24000.000000
Name: SALARY, dtype: float64
```

- 11 Liệt kê các nhân viên có lương (salary) > trung bình các lương, sắp tăng theo department_name và sắp giảm theo salary

	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	SALARY	DEPARTMENT_NAME	JOB_TITLE
0	205	Shelley	Higgins	12008	Accounting	Accounting Manager
1	206	William	Gietz	8300	Accounting	Public Accountant
2	100	Steven	King	24000	Executive	President
3	101	Neena	Kochhar	17000	Executive	Administration Vice President
4	102	Lex	De Haan	17000	Executive	Administration Vice President
5	108	Nancy	Greenberg	12008	Finance	Finance Manager
6	109	Daniel	Faviet	9000	Finance	Accountant
7	110	John	Chen	8200	Finance	Accountant
8	112	Jose Manuel	Urman	7800	Finance	Accountant
9	111	Ismael	Sciarra	7700	Finance	Accountant

- 12 Cho biết nhân viên có lương lớn nhất, nhân viên có lương nhỏ nhất



- 13 Cho biết các nhân viên có lương lớn nhất của mỗi phòng, sắp tăng theo department_id:

	EMPLOYEE_ID	FIRST_NAME	SALARY
DEPARTMENT_ID			
10.0	200	Jennifer	4400
20.0	201	Michael	13000
30.0	114	Den	11000
40.0	203	Susan	6500
50.0	121	Adam	8200
60.0	103	Alexander	9000
70.0	204	Hermann	10000
80.0	145	John	14000
90.0	100	Steven	24000
100.0	108	Nancy	12008
110.0	205	Shelley	12008

- 14 Tạo cột BONUS có giá trị là SALARY + 500 và in ra 5 nhân viên đầu tiên theo mẫu sau:

	EMPLOYEE_ID	FIRST_NAME	SALARY	BONUS
0	100	Steven	24000	24500
1	101	Neena	17000	17500
2	102	Lex	17000	17500
3	103	Alexander	9000	9500
4	104	Bruce	6000	6500

- 15 Tạo cột NAM_VAO_LAM là năm vào làm (trích năm từ cột HIRE_DATE). Xóa cột HIRE_DATE

	EMPLOYEE_ID	FIRST_NAME	NAM_VAO_LAM
0	100	Steven	2003
1	101	Neena	2005
2	102	Lex	2001
3	103	Alexander	2006
4	104	Bruce	2007

16 Thống kê theo mẫu sau:

	DEPARTMENT_NAME	count	max	min	mean	sum
0	Accounting	2	12008	8300	10154.000000	20308
1	Administration	1	4400	4400	4400.000000	4400
2	Executive	3	24000	17000	19333.333333	58000
3	Finance	6	12008	6900	8601.333333	51608
4	Human Resources	1	6500	6500	6500.000000	6500
5	IT	5	9000	4200	5760.000000	28800
6	Marketing	2	13000	6000	9500.000000	19000
7	Public Relations	1	10000	10000	10000.000000	10000
8	Purchasing	6	11000	2500	4150.000000	24900
9	Sales	34	14000	6100	8955.882353	304500
10	Shipping	45	8200	2100	3475.555556	156400

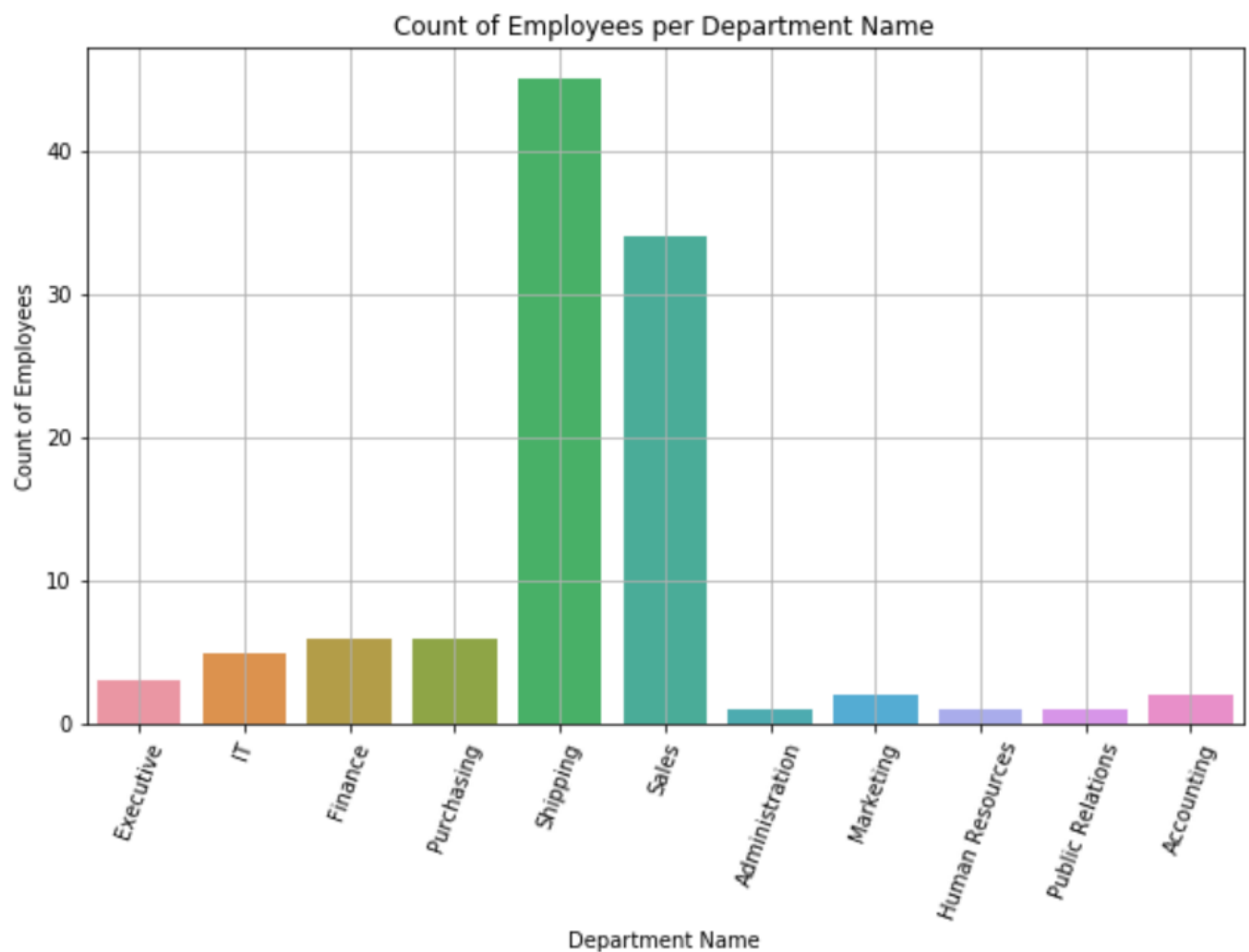
17 Thống kê theo mẫu sau:

	NAM_VAO_LAM	COUNT
0	2005	29
1	2006	24
2	2007	19
3	2008	11
4	2004	10
5	2002	7
6	2003	6
7	2001	1

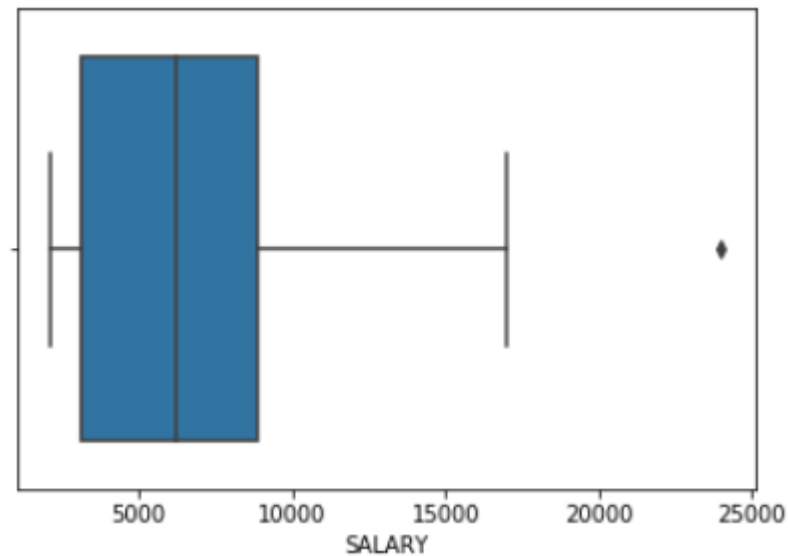
18 Thống kê theo mẫu sau:

	DEPARTMENT_NAME	JOB_TITLE	COUNT
0	Accounting	Accounting Manager	1
1	Accounting	Public Accountant	1
2	Administration	Administration Assistant	1
3	Executive	Administration Vice President	2
4	Executive	President	1
5	Finance	Accountant	5
6	Finance	Finance Manager	1
7	Human Resources	Human Resources Representative	1
8	IT	Programmer	5
9	Marketing	Marketing Manager	1
10	Marketing	Marketing Representative	1

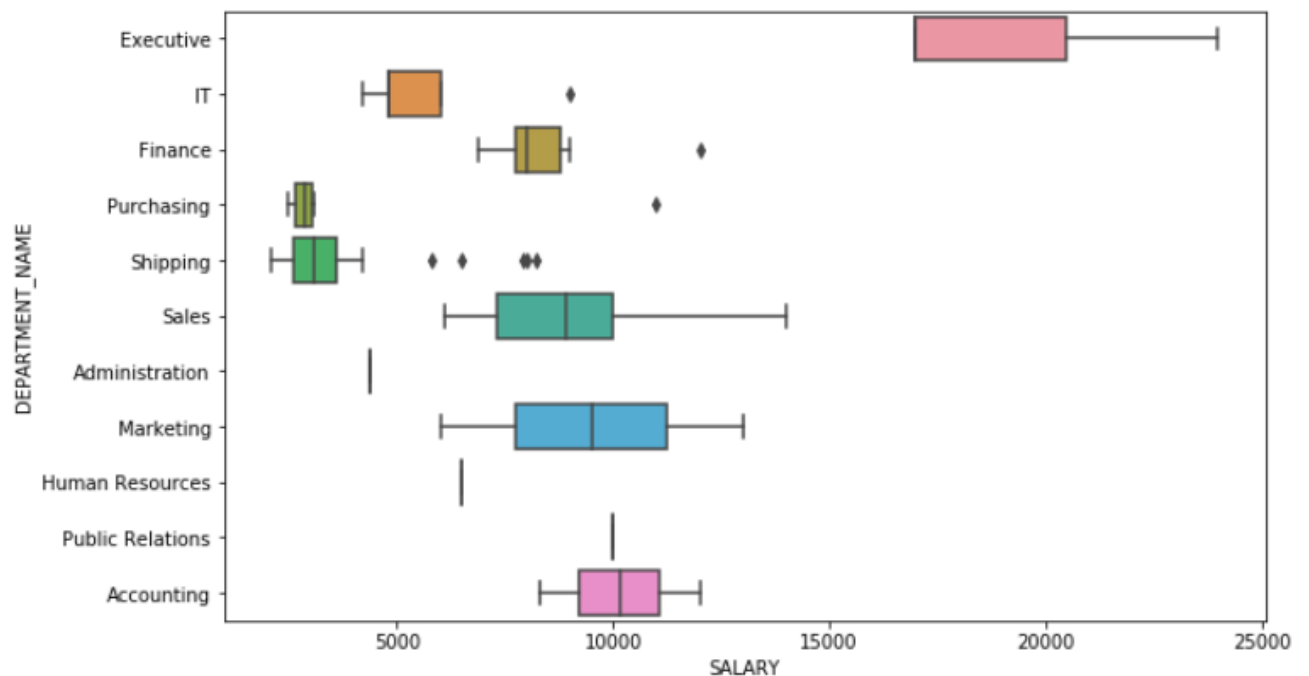
19 Vẽ biểu đồ theo mẫu sau: (đếm số nhân viên theo phòng)



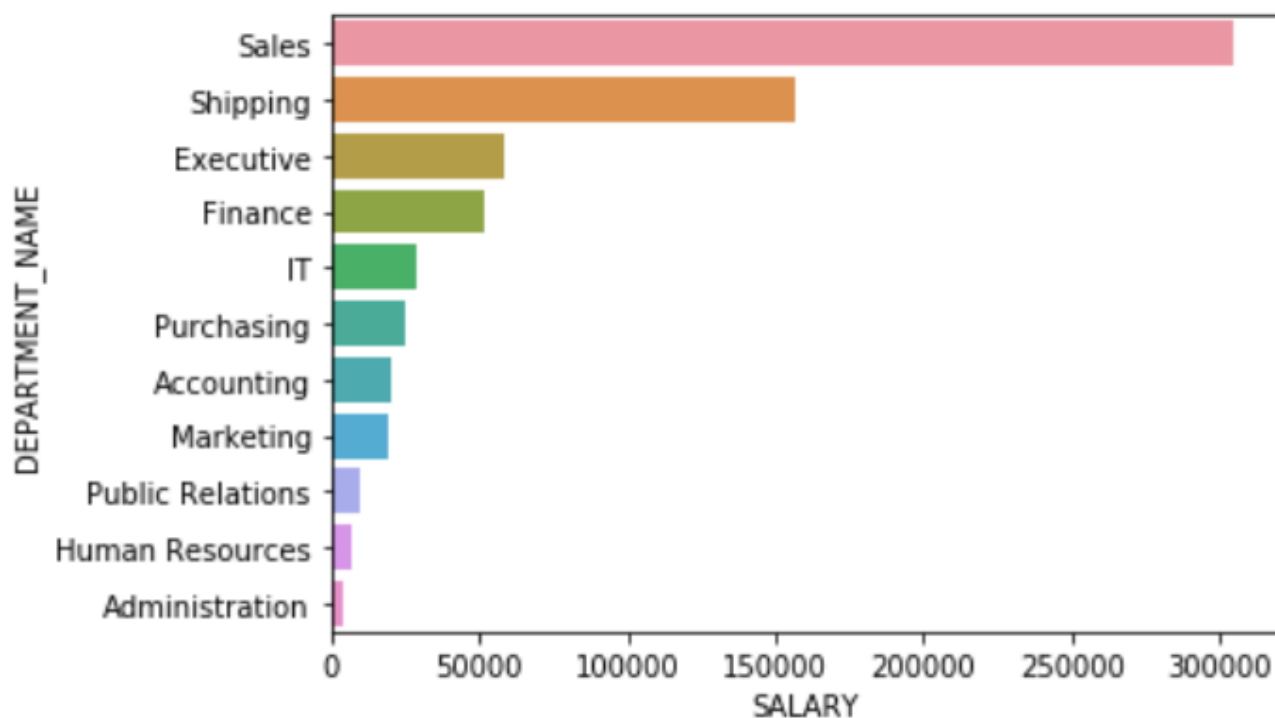
20 Vẽ biểu đồ theo mẫu sau: vẽ lương các nhân viên toàn công ty



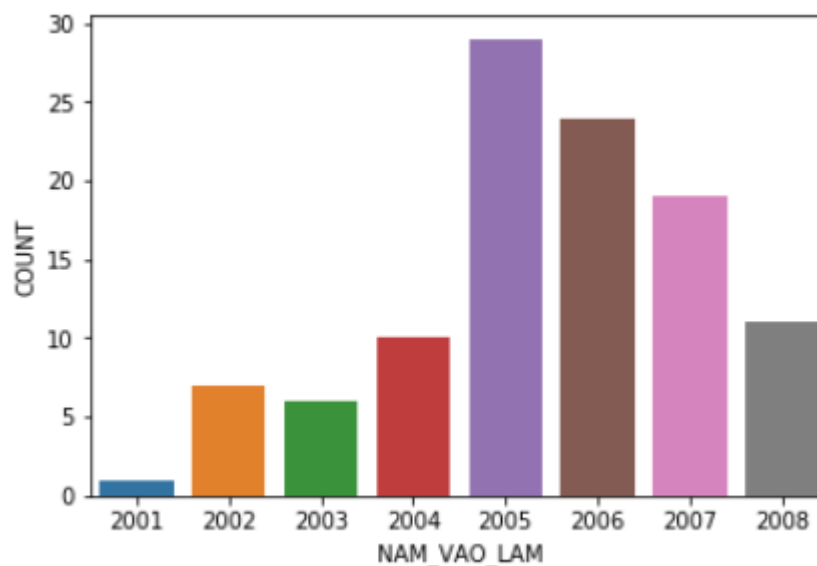
21 Vẽ biểu đồ theo mẫu sau: vẽ lương các nhân viên theo mỗi phòng



22 Vẽ biểu đồ theo mẫu sau: vẽ tổng lương theo mỗi phòng



23 Vẽ biểu đồ theo mẫu sau: vẽ số nhân viên vào làm theo năm



24 Vẽ biểu đồ theo mẫu sau: top 5 theo số nhân viên của mỗi phòng