

**Đề thi:**

# PYTHON FOR MACHINE LEARNING, DATA SCIENCE AND VISUALIZATION

Thời gian: 120 phút

Ngày thi: 16/08/2020

----- ĐƯỢC sử dụng tài liệu -----

- 1 Tạo DataFrame df chứa danh sách các sinh viên khoa CN, khoa TO và khoa VL. Các danh sách này được đọc từ các tập tin sinh\_vien\_CN.csv, sinh\_vien\_TO.csv và sinh\_vien\_VL.csv
- 2 Cho biết số cột, số dòng, các dữ liệu head và tail của DataFrame df.

	masv	ho	ten	gioitinh	ngaysinh	email	didong	cmnd	hocbong	makh
0	C0001	Khương Thảo	Loan	False	1999-01-04 00:00:00	ktloan@gmail.com	987314518	586900775484	2000000	CN
1	C0002	Đặng Bạch	Ngọc	True	1999-11-11 00:00:00	dbngoc@gmail.com	987587327	274387352269	0	CN
2	C0003	Phạm Văn Minh	Thiện	True	1999-11-23 00:00:00	pvmthien@gmail.com	987858734	835772714136	0	CN
3	C0004	Đinh Thị Thanh	Dung	False	1999-01-19 00:00:00	dttdung@gmail.com	987508413	373064334392	1000000	CN
4	C0005	Trần Mạnh	Thiện	True	1999-08-19 00:00:00	tmthien@gmail.com	987579776	233384596844	1000000	CN

- 3 Tạo lại index cho DataFrame df với giá trị bắt đầu từ 0

	masv	ho	ten	gioitinh	ngaysinh	email	didong	cmnd	hocbong	makh
995	V0996	Trần Mạnh	Thăng	True	1999-06-21 00:00:00	tmthang@gmail.com	987250062	768656897616	2000000	VL
996	V0997	Trần Vĩnh	Thiện	True	1999-07-23 00:00:00	tvthien@gmail.com	987716502	468267857113	3000000	VL
997	V0998	Khương Thảo	Hạnh	False	1999-08-16 00:00:00	kthanh@gmail.com	913696608	702859343545	0	VL
998	V0999	Đinh Ngọc	Ngọc	True	1999-05-25 00:00:00	dnngoc@gmail.com	913689521	187831202224	3000000	VL
999	V1000	Lý Mạnh	Bảo	True	1999-06-03 00:00:00	lmbao@gmail.com	987656186	865173188581	0	VL

- 4 Tạo DataFrame df\_ket\_qua từ tập tin ket\_qua.csv

	masv	mamh	diem
0	C0001	CSD1	7.0
1	C0001	CSD2	6.0
2	C0001	CTDL	9.0
3	C0001	KTLT	5.5
4	C0001	LQL1	7.0



- 5 Tạo DataFrame `df_dtb` bằng cách nhóm `df_ket_qua` theo `masv` và tính điểm trung bình cho mỗi sinh viên.

	masv	diem
0	C0001	6.5
1	C0002	7.2
2	C0003	7.7
3	C0004	6.9
4	C0005	7.7

- 6 Trộn `df` và `df_dtb` theo `masv` và có kết quả của `df` như sau:

	masv	ho	ten	gioitinh	ngaysinh	email	didong	cmnd	hocbong	makh	diem
0	C0001	Khương Thảo	Loan	False	1999-01-04 00:00:00	ktloan@gmail.com	987314518	586900775484	2000000	CN	6.5
1	C0002	Đặng Bạch	Ngọc	True	1999-11-11 00:00:00	dbngoc@gmail.com	987587327	274387352269	0	CN	7.2
2	C0003	Phạm Văn Minh	Thiện	True	1999-11-23 00:00:00	pvmthien@gmail.com	987858734	835772714136	0	CN	7.7
3	C0004	Đinh Thị Thanh	Dung	False	1999-01-19 00:00:00	dttdung@gmail.com	987508413	373064334392	1000000	CN	6.9
4	C0005	Trần Mạnh	Thiện	True	1999-08-19 00:00:00	tmthien@gmail.com	987579776	233384596844	1000000	CN	7.7

- 7 In ra các sinh viên của khoa CN và có `diem > 8`

	masv	ho	ten	gioitinh	ngaysinh	email	didong	cmnd	hocbong	makh	diem
43	C0044	Bạch Thị	Thúy	False	1999-11-09 00:00:00	btthuy@gmail.com	987681542	483340458851	0	CN	8.05
51	C0052	Lê Thị Thảo	Ly	False	1999-02-14 00:00:00	lttly@gmail.com	168434476	874368763621	1000000	CN	8.55

- 8 In ra các sinh viên của khoa CN và có `diem > 8`. Sắp xếp tăng theo `ten` và giảm theo `diem`

	masv	ho	ten	gioitinh	ngaysinh	email	didong	cmnd	hocbong	makh	diem
153	C0154	Cao Văn Minh	Bảo	True	1999-05-11 00:00:00	cvmbao@gmail.com	913598140	530460808272	3000000	CN	8.60
293	C0294	Phạm Nam	Bảo	True	1999-12-30 00:00:00	pnbao@gmail.com	987586638	968772828515	1000000	CN	8.05
435	C0436	Lâm Thị Thanh	Dung	False	1999-11-24 00:00:00	ltdung@gmail.com	168753165	971746450069	2000000	CN	8.20
271	C0272	Thái Nam	Dũng	True	1999-08-12 00:00:00	tndung@gmail.com	913486853	701413814886	0	CN	8.20
108	C0109	Bạch Thị	Hoa	False	1999-10-14 00:00:00	bthoa@gmail.com	913174720	347889227616	1000000	CN	8.20
356	C0357	Lê Thị	Hoa	False	1999-08-24 00:00:00	lthoa@gmail.com	168168028	407376423068	3000000	CN	8.10



9 In ra các nam sinh viên không có học bổng.

	masv	ho	ten	gioitinh	ngaysinh	email	didong	cmnd	hocbong	makh	diem
1	C0002	Đặng	Ngọc	True	1999-11-11 00:00:00	dbngoc@gmail.com	987587327	274387352269	0	CN	7.20
2	C0003	Phạm Văn	Thiên	True	1999-11-23 00:00:00	pvmthien@gmail.com	987858734	835772714136	0	CN	7.70
25	C0026	Lâm Vĩnh	Hiệu	True	1999-07-26 00:00:00	lvhieu@gmail.com	987880747	304474621035	0	CN	6.80

10 In ra tổng học bổng của mỗi khoa.

	makh	hocbong
0	CN	711000000
1	TO	140000000
2	VL	616000000

11 Thống kê học bổng của mỗi khoa.

	makh	sum	max	min	mean
0	CN	711000000	3000000	0	1422000
1	TO	140000000	3000000	0	1400000
2	VL	616000000	3000000	0	1540000

12 Thống kê điểm của mỗi khoa.

	makh	count	max	min	mean
0	CN	500	8.75	5.50	7.2188
1	TO	100	8.35	6.10	7.2990
2	VL	400	8.65	5.15	7.1875

13 Cho biết các sinh viên có điểm lớn nhất, điểm nhỏ nhất.

	masv	ho	ten	gioitinh	ngaysinh	email	didong	cmnd	hocbong	makh	diem
832	V0833	Trần Phú	Minh	False	1999-05-24 00:00:00	tpminh@gmail.com	168861731	190672558617	2000000	VL	5.15

	masv	ho	ten	gioitinh	ngaysinh	email	didong	cmnd	hocbong	makh	diem
354	C0355	Văn Thị	Mai	False	1999-11-10 00:00:00	vtmai@gmail.com	987327541	855354458643	0	CN	8.75

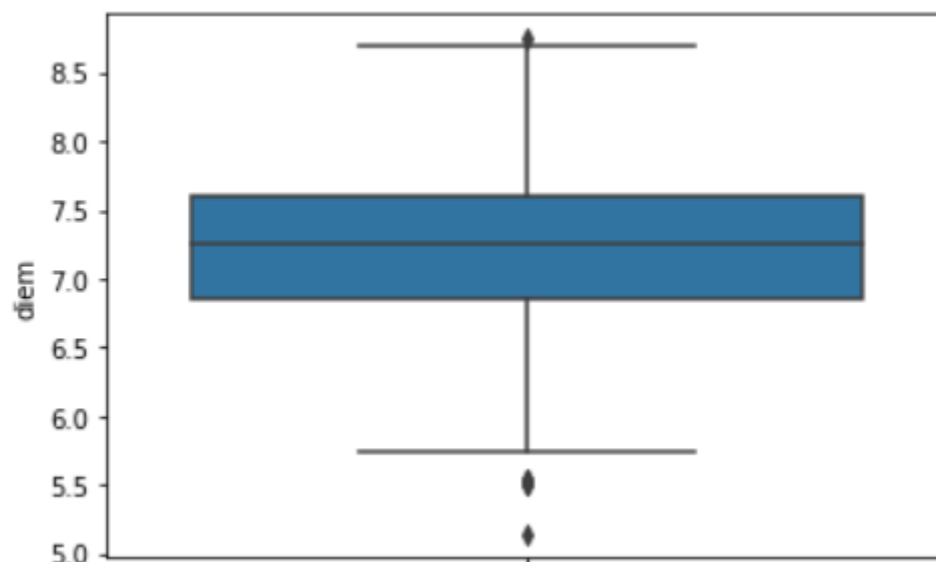
14 Trong DataFrame df tạo thêm cột ketqua với kết quả là giỏi nếu điểm  $\geq 8$  và kết quả là khá nếu điểm  $< 8$ .

	masv	ho	ten	makh	diem	ketqua
0	C0001	Khương Thảo	Loan	CN	6.50	khá
1	C0002	Đặng Bạch	Ngọc	CN	7.20	khá
2	C0003	Phạm Văn Minh	Thiện	CN	7.70	khá
3	C0004	Đinh Thị Thanh	Dung	CN	6.90	khá
4	C0005	Trần Mạnh	Thiện	CN	7.70	khá
5	C0006	Trần Văn Minh	Hiệu	CN	7.80	khá
6	C0007	Lê Thị	Mai	CN	7.65	khá

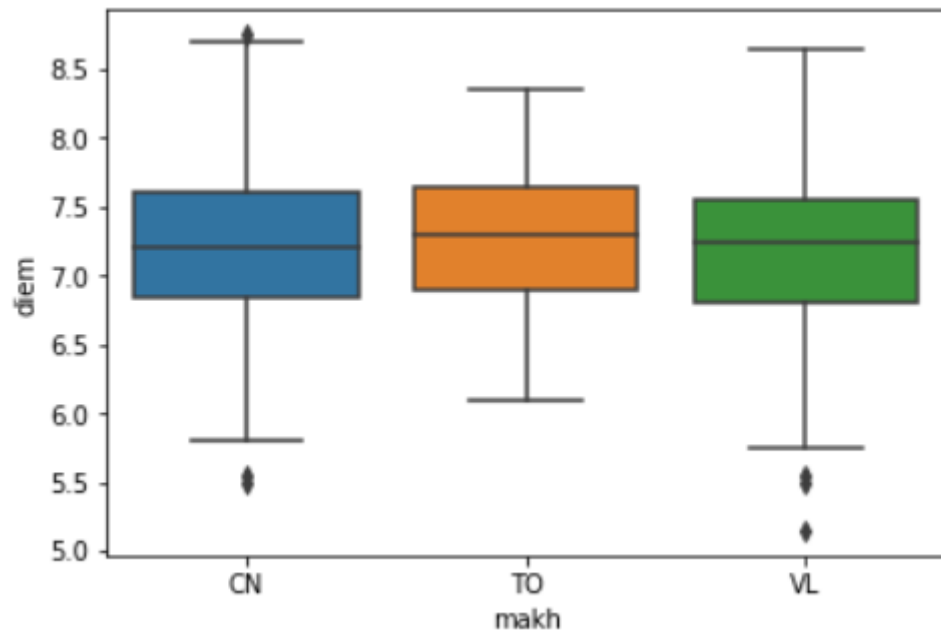
15 Thống kê theo mẫu:

	makh	ketqua	masv
0	CN	giỏi	39
1	CN	khá	461
2	TO	giỏi	12
3	TO	khá	88
4	VL	giỏi	33
5	VL	khá	367

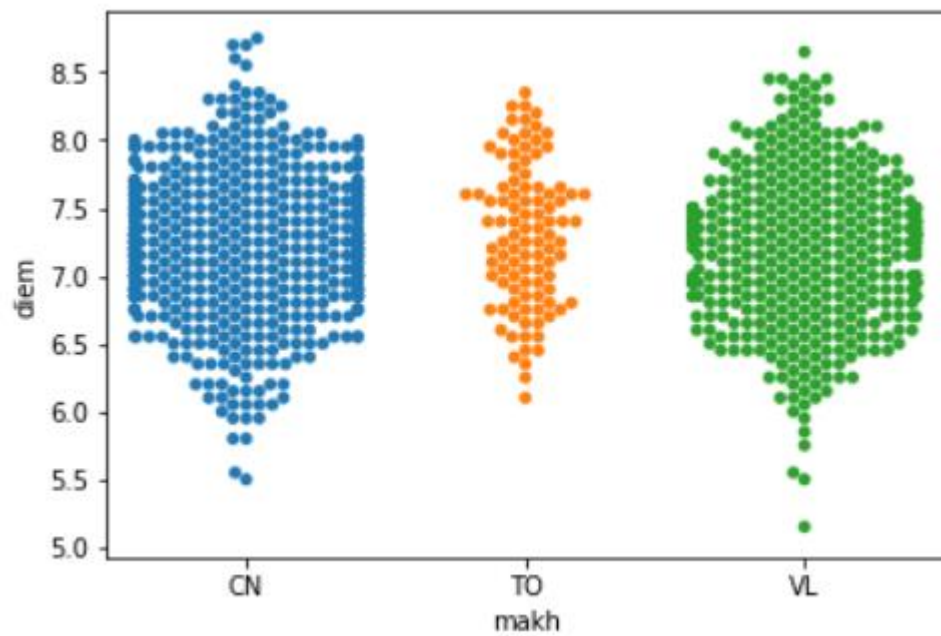
16 Vẽ biểu đồ boxplot với trục y là điểm, dữ liệu là DataFrame df.



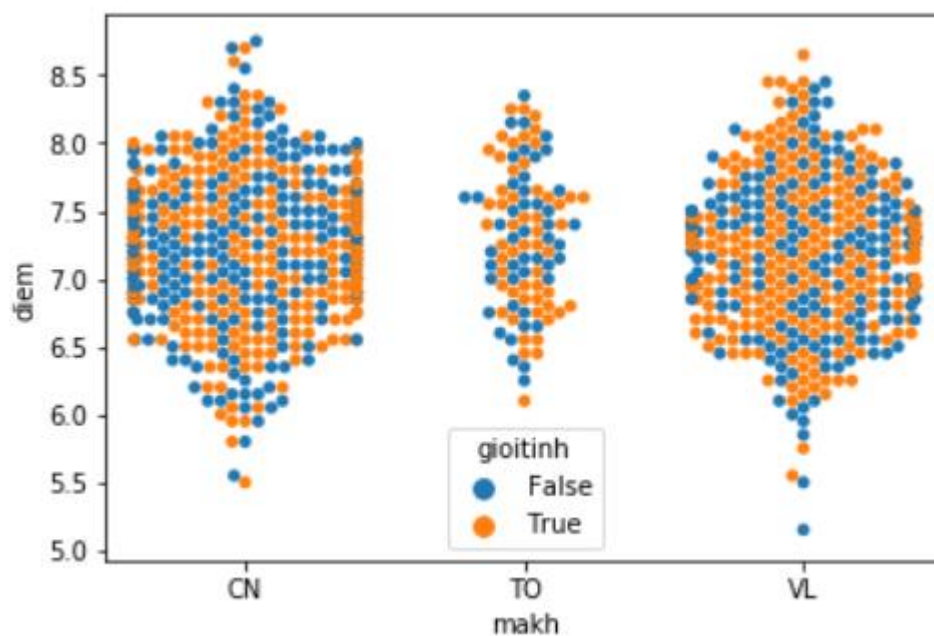
17 Vẽ biểu đồ boxplot với trục y là điểm, trục x là makh.



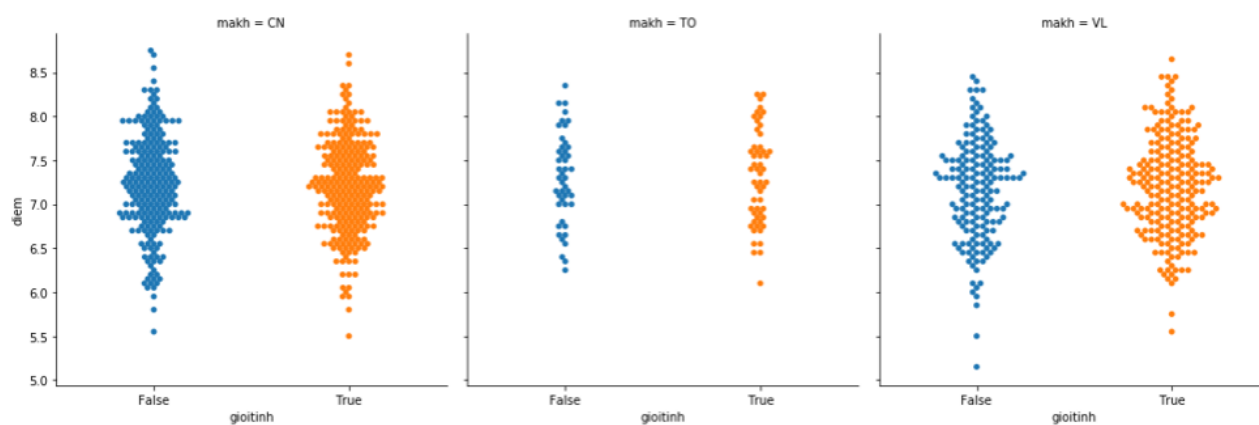
18 Vẽ biểu đồ theo mẫu sau:



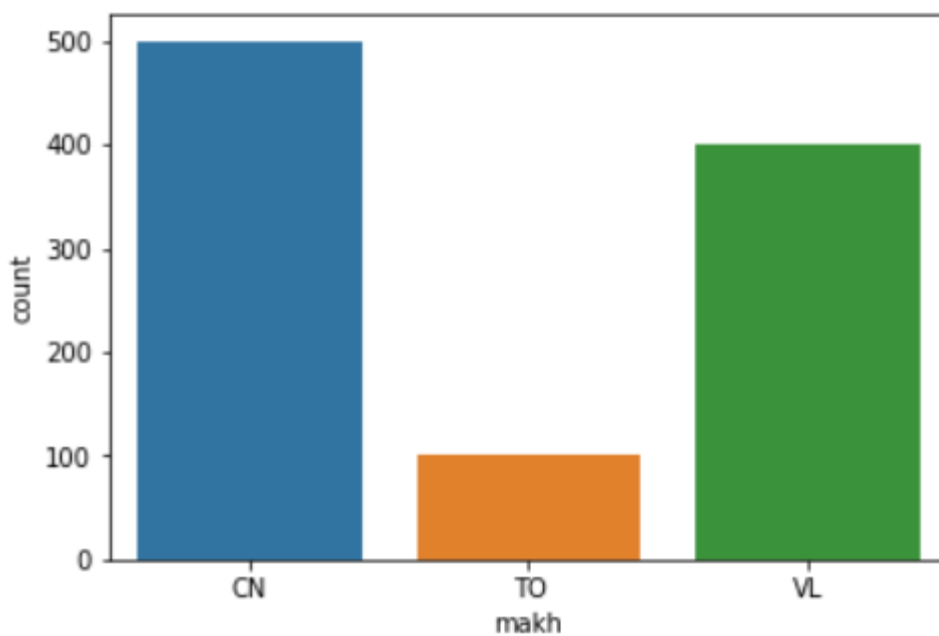
19 Vẽ biểu đồ theo mẫu sau:



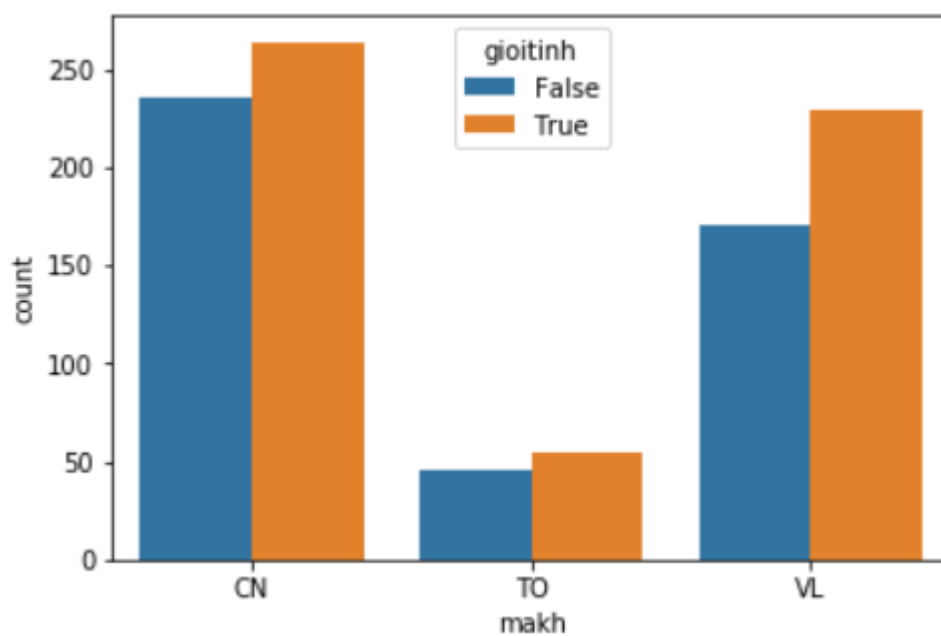
20 Vẽ biểu đồ theo mẫu sau:



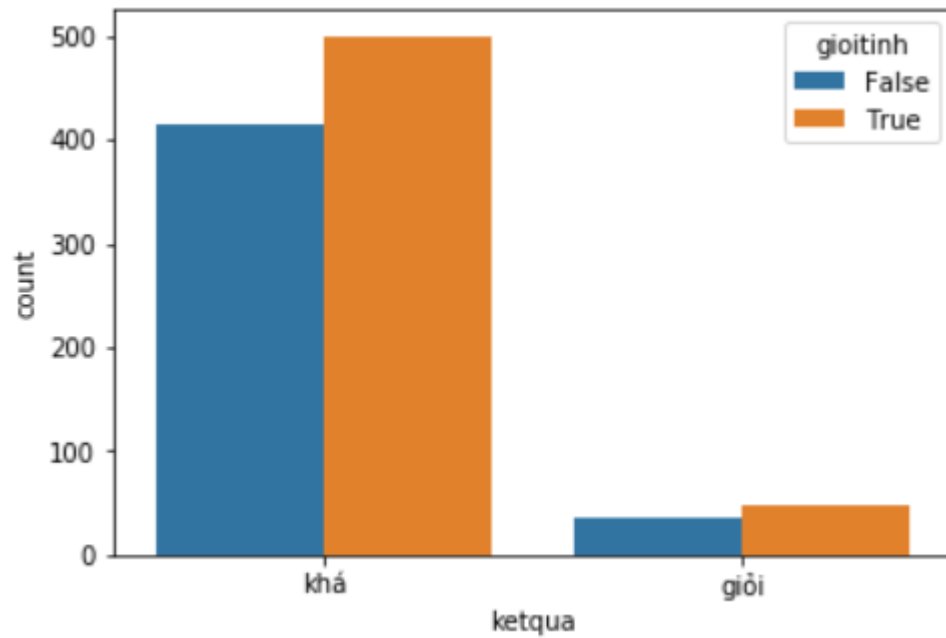
21 Vẽ biểu đồ countplot với x là makh.



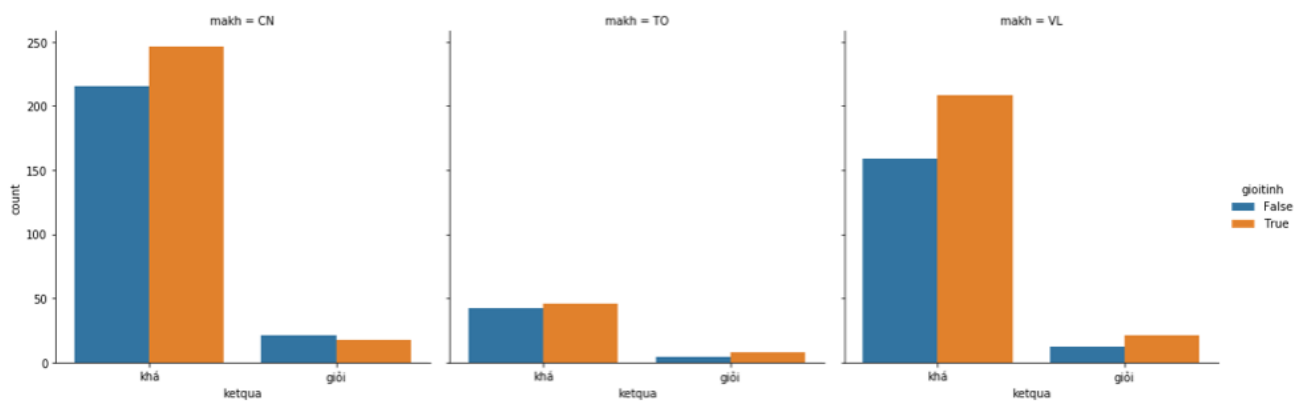
22 Vẽ biểu đồ theo mẫu sau:



23 Vẽ biểu đồ theo mẫu sau:



24 Vẽ biểu đồ theo mẫu sau:



25 Vẽ biểu đồ theo mẫu sau:



