

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
❧❧●●❧❧



TASK 2.2

MÔN: ĐỒ ÁN KỸ THUẬT LẬP TRÌNH

GVHD: Lưu Quang Huân

Lớp L01

Danh sách thành viên

STT	HỌ VÀ TÊN	MSSV
1	Mai Nguyễn Minh Hiếu	1812171
2	Ngô Quang Anh	1810014
3	Hồ Ngọc Trí	1814454

1) Các công việc thực hiện

-Mai Nguyễn Minh Hiếu :

crawl thêm dữ liệu để được ~10.000 cũng như là tập hợp dữ liệu của các thành viên trong nhóm lại cũng như là sạch dữ liệu.

-Ngô Quang Anh :

Phân tích dữ liệu và viết báo cáo

-Hồ Ngọc Trí :

Phân tích dữ liệu và đưa dữ liệu lên Database

2) Làm sạch dữ liệu

Ở đây ta có thể thấy dữ liệu thô khi được crawl về từ trang web

Out[15]:

	Title	Comment	Camera		Locate	Link_image
0	Oui's Bánh Mì - Thiên Phước	0	0	319 C15 Khu Cao Ốc Thuận Việt, Thiên Phước, P. 15		https://images.foody.vn/res/g105/1049106/prof/...
1	Bánh Mì Huỳnh Hoa - Bánh Mì Pate	344	872		Quận 11	https://images.foody.vn/res/g2/13831/prof/s640...
2	Bánh Mì 110	42	90		TP. HCM	https://images.foody.vn/res/g5/41612/prof/s640...
3	Bánh Mì Cô Diệp	55	66	26 Lê Thị Riêng, P. Bến Thành		https://images.foody.vn/res/g3/20512/prof/s640...
4	Bánh Nấm - Bánh Mì - Bến Vân Đồn	155	377		Quận 1	https://images.foody.vn/res/g21/203291/prof/s6...
...
2803	Bánh Mì & Xôi Daddy	24	159		NaN	https://images.foody.vn/res/g20/190702/prof/s6...
2804	Bánh Mì Chảo - 189 Đinh Tiên Hoàng	154	653		2	https://images.foody.vn/res/g5/48258/prof/s640...
2805	A Túng - Bánh Mì Bò Nướng Bơ Cambodia - Cống Q...	331	1174	145 Cống Quỳnh, P. Nguyễn Cư Trinh		https://images.foody.vn/res/g21/206873/prof/s6...
2806	Bánh Mì Anh Phan - Bánh Mì Gà & Thịt Nguội	50	127		Quận 1	https://images.foody.vn/res/g7/62817/prof/s640...
2807	Bánh Mì Bà Nội - Bánh Mì Thịt Nướng	36	35		TP. HCM	https://images.foody.vn/res/g97/965039/prof/s6...

2808 rows x 5 columns

từ dữ liệu ta sẽ lược bỏ các thành phần NaN cũng như là phần link ảnh không cần thiết cho việc phân tích nên cũng lược bỏ.

[38]:

	Title	Comment	Camera	Locate
0	Oui's Bánh Mì - Thiên Phước	0	0	319 C15 Khu Cao Ốc Thuận Việt, Thiên Phước, P. 15
1	Bánh Mì Huỳnh Hoa - Bánh Mì Pate	344	872	Quận 11
2	Bánh Mì 110	42	90	TP. HCM
3	Bánh Mì Cô Diệp	55	66	26 Lê Thị Riêng, P. Bến Thành
4	Bánh Năm - Bánh Mì - Bến Vân Đồn	155	377	Quận 1
...
2802	Bánh Mì Đại Long	3	27	TP. HCM
2804	Bánh Mì Chảo - 189 Đinh Tiên Hoàng	154	653	2
2805	A Tùng - Bánh Mì Bò Nướng Bơ Cambodia - Cống Q...	331	1174	145 Cống Quỳnh, P. Nguyễn Cư Trinh
2806	Bánh Mì Anh Phan - Bánh Mì Gà & Thịt Nguội	50	127	Quận 1
2807	Bánh Mì Bà Nội - Bánh Mì Thịt Nướng	36	35	TP. HCM

2577 rows x 4 columns

dữ liệu đã được giảm từ 2808 -> 2577

3)Phân tích dữ liệu

đến với dữ liệu chung

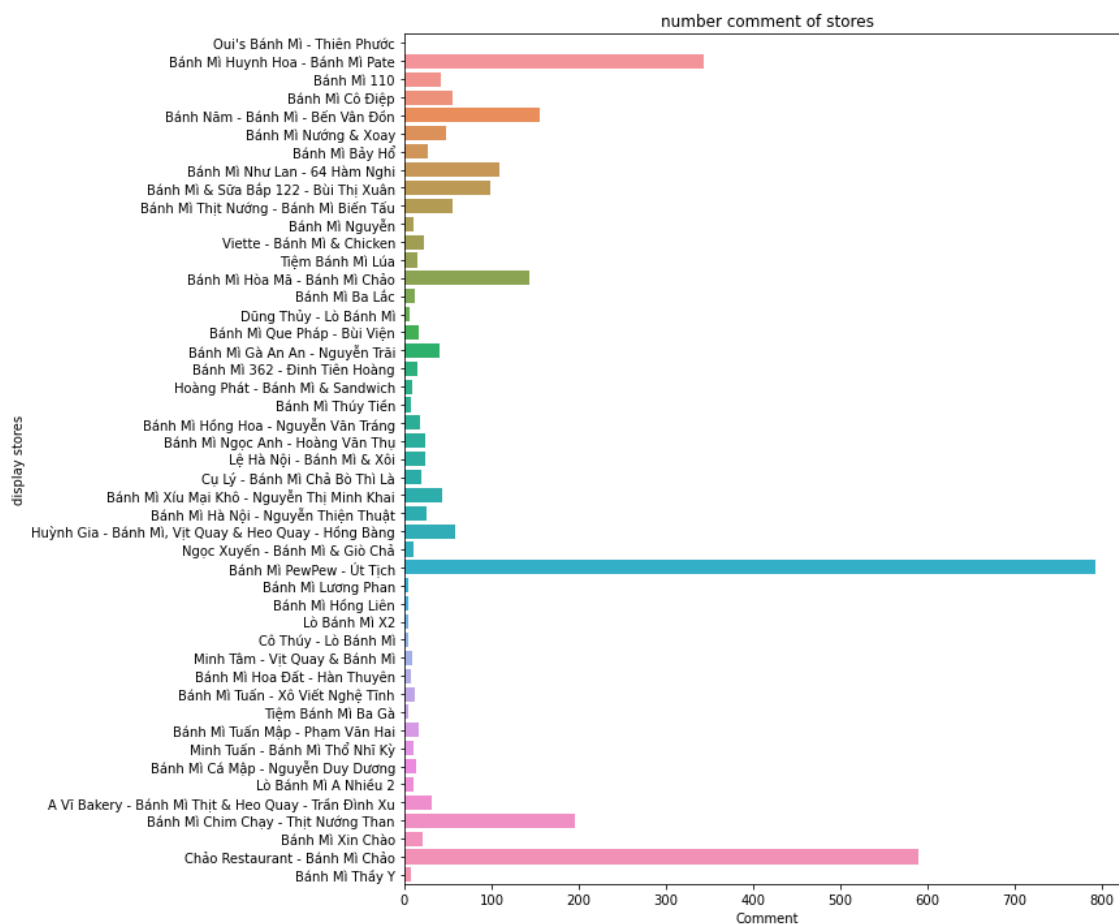
[4]:

	Comment	Camera
count	2577.000000	2577.000000
mean	49.161040	175.483120
std	92.218903	336.033141
min	0.000000	0.000000
25%	4.000000	10.000000
50%	7.000000	27.000000
75%	36.000000	127.000000
max	793.000000	2551.000000

nếu ta có thể thấy Tổng số dữ liệu sau khi được lọc thì còn : 2577 data

tổng trung bình Comment 49 khá là cao cho mỗi cửa hàng cũng như là có 175 Photos được thêm vào cửa hàng.

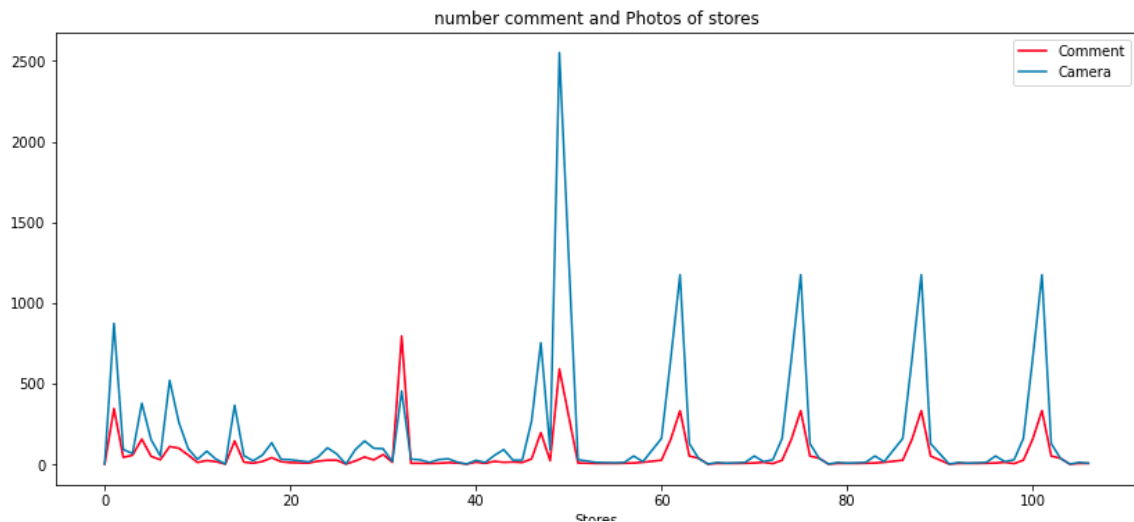
Ở đây ta có thể thấy được có cửa hàng không có Comment hay Photos nhưng cũng có cửa hàng có tận 793 Comment hay 2551 Photos dẫn đến sự chênh lệch dữ liệu quá lớn.



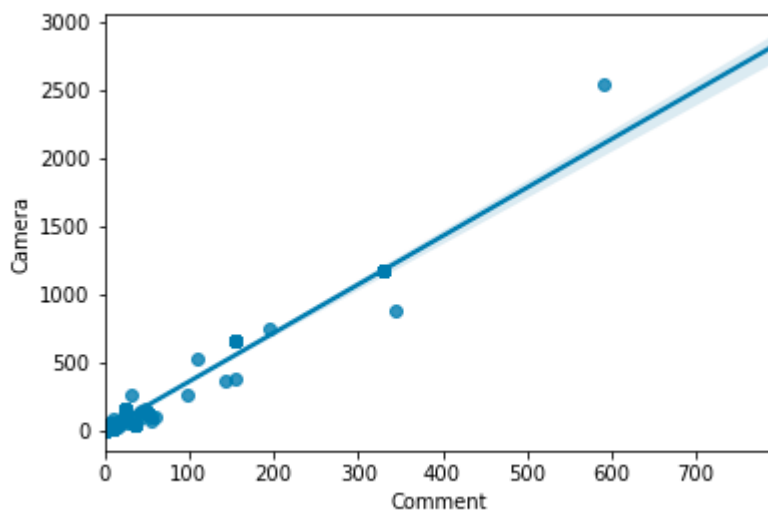
đây là các 50 cửa hàng bánh mì được hệ thống chọn đầu tiên. Theo biểu đồ ta thấy 'Comment' cho các cửa hàng là không đồng đều, việc không đồng đều này có thể là do vị trí địa lý, hay cũng có thể là do độ nổi tiếng của các cửa hàng. Đặc biệt trong đó có cửa hàng

"Bánh Mì PewPew - Út Tịch" là chiếm lượng 'Comment' nổi trội hơn nhất (~800).

Tiếp theo ta sẽ cùng xem lượng ảnh của cửa hàng được chụp cũng như lượng Comment:



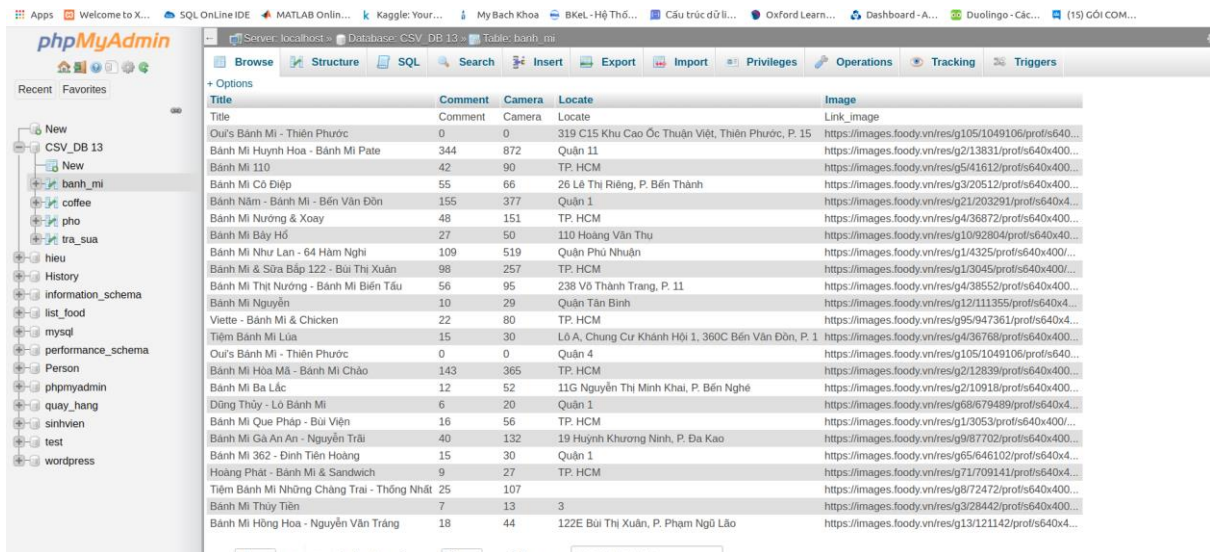
ta thấy lượng "Comment" ít hơn hẳn so với các lượng Photo đã được chụp. Chúng tôi khách hàng thích review bằng ảnh hơn so với Comment cũng như là thể hiện được chân thật phong cảnh của cửa hàng. Lượng ảnh ở những cửa hàng nhất định luôn có những đột biến có khi lên đến (>2500 Photos)



ở đây là lượng phân bố Comment và các Photos ta có thể thấy có chiều hướng tăng dần khi mà các lượng Comment càng nhiều thì lượng Photos của 1 cửa hàng cũng được tăng đáng kể (không kể đến các giá trị quá lớn vì nó có thể gây nhiễu) nhưng nhìn chung thì ta có thể thấy lượng Comment cũng có ảnh hưởng đến cửa hàng

4) Đưa dữ liệu lên trên Database

Ngoài ra nhóm cũng đã thành công trong việc đưa các dữ liệu lên trên Database với tổng ~10.000 data để chuẩn bị cho việc ứng dụng sau này.



Title	Comment	Camera	Locate	Image
Ou's Bánh Mì - Thiên Phước	0	0	319 C15 Khu Cao Ốc Thuận Việt, Thiên Phước, P. 15	https://images.foody.vn/res/g105/1049106/profs640x400...
Bánh Mì Huỳnh Hoa - Bánh Mì Pate	344	872	Quận 11	https://images.foody.vn/res/g2/13831/profs640x400...
Bánh Mì 110	42	90	TP. HCM	https://images.foody.vn/res/g5/41612/profs640x400...
Bánh Mì Cò Đép	55	66	26 Lê Thị Riêng, P. Bến Thành	https://images.foody.vn/res/g3/20512/profs640x400...
Bánh Nấm - Bánh Mì - Bến Văn Dòn	155	377	Quận 1	https://images.foody.vn/res/g21/203291/profs640x400...
Bánh Mì Nướng & Xoay	48	151	TP. HCM	https://images.foody.vn/res/g4/36872/profs640x400...
Bánh Mì Bầy Hồ	27	50	110 Hoàng Văn Thu	https://images.foody.vn/res/g109/2804/profs640x400...
Bánh Mì Như Lan - 64 Hàm Nghi	109	519	Quận Phú Nhuận	https://images.foody.vn/res/g1/4325/profs640x400...
Bánh Mì & Sữa Bắp 122 - Bùi Thị Xuân	98	257	TP. HCM	https://images.foody.vn/res/g1/3045/profs640x400...
Bánh Mì Thịt Nướng - Bánh Mì Biển Tầu	56	95	238 Võ Thành Trang, P. 11	https://images.foody.vn/res/g4/38552/profs640x400...
Bánh Mì Nguyễn	10	29	Quận Tân Bình	https://images.foody.vn/res/g12/111355/profs640x400...
Viette - Bánh Mì & Chicken	22	80	TP. HCM	https://images.foody.vn/res/g95/947361/profs640x400...
Tiệm Bánh Mì Lúa	15	30	Lô A, Chung Cư Khánh Hội 1, 360C Bến Văn Dòn, P. 1	https://images.foody.vn/res/g4/36768/profs640x400...
Ou's Bánh Mì - Thiên Phước	0	0	Quận 4	https://images.foody.vn/res/g105/1049106/profs640x400...
Bánh Mì Hòa Mã - Bánh Mì Chảo	143	365	TP. HCM	https://images.foody.vn/res/g2/12839/profs640x400...
Bánh Mì Ba Lắc	12	52	11G Nguyễn Thị Minh Khai, P. Bến Nghé	https://images.foody.vn/res/g1/3053/profs640x400...
Dũng Thủy - Lò Bánh Mì	6	20	Quận 1	https://images.foody.vn/res/g68/679489/profs640x400...
Bánh Mì Que Pháp - Bùi Viện	16	56	TP. HCM	https://images.foody.vn/res/g1/3053/profs640x400...
Bánh Mì Gà An An - Nguyễn Trãi	40	132	19 Huỳnh Khương Ninh, P. Đa Kao	https://images.foody.vn/res/g9/87702/profs640x400...
Bánh Mì 362 - Đinh Tiên Hoàng	15	30	Quận 1	https://images.foody.vn/res/g65/646102/profs640x400...
Hoàng Phát - Bánh Mì & Sandwich	9	27	TP. HCM	https://images.foody.vn/res/g1/709141/profs640x400...
Tiệm Bánh Mì Những Chàng Trai - Thống Nhất	25	107		https://images.foody.vn/res/g8/72472/profs640x400...
Bánh Mì Thủy Tiên	7	13	3	https://images.foody.vn/res/g3/28442/profs640x400...
Bánh Mì Hồng Hoa - Nguyễn Văn Tráng	18	44	122E Bùi Thị Xuân, P. Phạm Ngũ Lão	https://images.foody.vn/res/g13/121142/profs640x400...

5) Dự kiến công việc tuần tới

- Crawl thêm các dữ liệu cần thiết cho việc ứng dụng hơn
- Tìm hiểu ứng dụng của data crawl được
- ứng dụng cơ bản với các data crawl được trên Database