



PHÂN TÍCH BỘ DỮ LIỆU ỨNG DỤNG DI ĐỘNG

zomato





MỤC LỤC

01



TỔNG QUAN

PHÂN TÍCH DỮ LIỆU



02

03



TỔNG KẾT VÀ ĐƯA RA
CHIẾN LƯỢC KINH DOANH

zomato
DELIVERY PARTNER



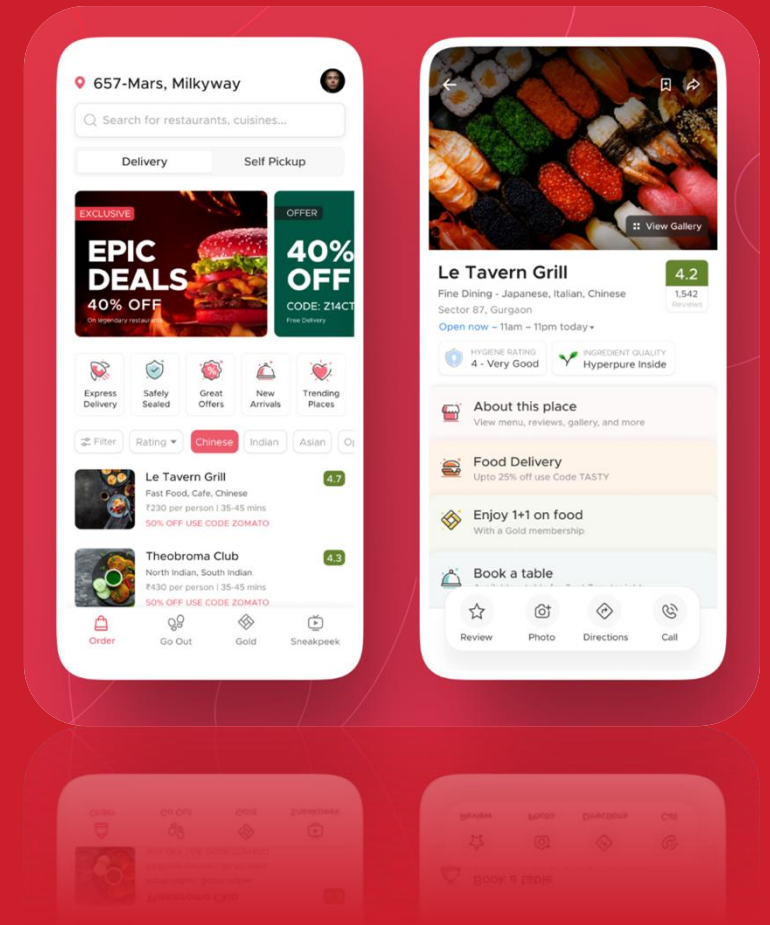
1. TỔNG QUAN

GIỚI THIỆU

zomato



Zomato là một nền tảng tìm kiếm nhà hàng, cung cấp thông tin chi tiết chuyên sâu về các nhà hàng trên toàn Ấn Độ với các đánh giá và xếp hạng từ người dùng.



Thông tin bộ dữ liệu

5 TABLES

FOOD

Chứa thông tin về tên món ăn và loại món: món chay / món mặn.

MENU

Chứa thông tin về phong cách ẩm thực, tên nhà hàng, giá các món ăn,..

ORDERS

Chứa thông tin về đơn đặt hàng của người dùng.

RESTAURANT

Chứa thông tin về các nhà hàng có trong ứng dụng.

USERS

Chứa thông tin về người dùng ứng dụng.

zomato



Trong project này, nhóm sử dụng ngôn ngữ Python cùng một số thư viện để lọc và tiền xử lý dữ liệu thô.

Các bảng sau khi được làm sạch được nhập vào Power BI để tạo ra các biểu đồ nhằm rút ra các phân tích có ý nghĩa.



Power BI

1

Xác định các cột có giá trị cho bài toán phân tích

Một số cột không phù hợp khi phân tích bài toán được loại bỏ (email, address, username, password, currency, ...)

2

Loại bỏ hoặc thêm thông tin vào các dòng chứa giá trị rỗng, giá trị không phù hợp

Điền giá trị trung bình vào các dòng có rating là rỗng

```
1 restaurant['rating'] = restaurant['rating'].map({'--': restaurant['rating'].mean()})
2 restaurant['rating'].fillna((restaurant['rating'].mean()), inplace=True)
```

[305]

Xóa dữ liệu có giá trị bất kỳ null, giá trị trùng lặp

```
1 users = users.dropna()
2 users.drop_duplicates(keep=False, inplace=True)
```

[168]

Xóa các đơn đặt hàng có doanh thu âm, thanh toán bằng USD

```
1 orders = orders[orders['sales_amount'] > 0]
2 orders = orders[orders['currency'] != 'USD']
```

[18]

Chuyển kiểu dữ liệu cột Cost về số nguyên

Lấy tên bang từ thông tin địa chỉ, tên thành phố

```
1 formatted_city = []
2 wrong_format_city = {'Bapatlachirala': 'Chirala', '': 'Uttar Pradesh',
3                       'Bharabanki': 'Barabanki', 'Bodhan-rural': 'Bodhan',
4                       'Bongaon': 'Bangaon', 'Karunagappally': 'Karunagappally',
5                       'Kothamanagalam': 'Kothamangalam', 'Kunnamkullam': 'Kunnamkulam',
6                       'Surendranagar-dudhrej': 'Surendranagar', 'Banda': 'Uttar Pradesh'}
7
8 for index, row in restaurant.iterrows():
9     city = row['city']
10
11     if city not in formatted_city:
12
13         last_city = city.rpartition(',')[0]
14
15         restaurant.loc[restaurant['city'] == city, 'city'] = last_city
16         formatted_city.extend([city])
17
18     for key, value in wrong_format_city.items():
19         if city == key:
20             restaurant.loc[restaurant['city'] == city, 'city'] = value
```

[113]

```
1 restaurant["cost"] = restaurant["cost"].str.replace("₹ ", "").astype(int)
```

[312]

Lấy tên bang từ tên thành phố bằng thư viện geopy

```
1 from geopy.geocoders import Nominatim
2
3 formatted_state_city = []
4 geolocator = Nominatim(user_agent="http")
5
6 for index, row in restaurant.iterrows():
7     city = row['city']
8
9     if city not in formatted_state_city:
10         location = geolocator.geocode(city, addressdetails=True)
11
12         full_address = location.raw['address']
13         state = full_address.get('state', '')
14
15         restaurant.loc[restaurant['city'] == city, 'city'] = state
16         formatted_state_city.extend([city])
```

[106]

3 Chuẩn hóa định dạng

Xuất dữ liệu đã được chuẩn hóa

Xuất data đã được chuẩn hóa

```
1 restaurant.to_csv('dataset/restaurant_n.csv',
2                   columns=['id', 'city', 'rating', 'rating_count'
3                             , 'cost', 'cuisine']
4                   , sep='\t', index=False)
```

[107]

Kích thước dữ liệu trước & sau tiền xử lý

```
Shape of food table:      (371561, 4)
Shape of menu table:      (1179936, 6)
Shape of orders table:    (150281, 7)
Shape of restaurant table: (148541, 12)
Shape of users table:     (100000, 12)
```

```
Shape of food table:      (371560, 2)
Shape of menu table:      (1179936, 5)
Shape of orders table:    (147061, 5)
Shape of restaurant table: (148255, 7)
Shape of users table:     (100000, 8)
```

zomato
DELIVERY PARTNER



2. Phân tích dữ liệu

2. Phân tích dữ liệu



Hướng 1: Mô tả số lượng đơn hàng và doanh thu theo thời gian

Hướng 3 : Mô phỏng số lượng đơn hàng và doanh thu theo nhân khẩu học



Hướng 2: Mô tả số lượng đơn hàng và doanh thu theo thông tin nhà hàng

Hướng 4 : Phân tích doanh thu theo khu vực



Phân Tích Dữ Liệu

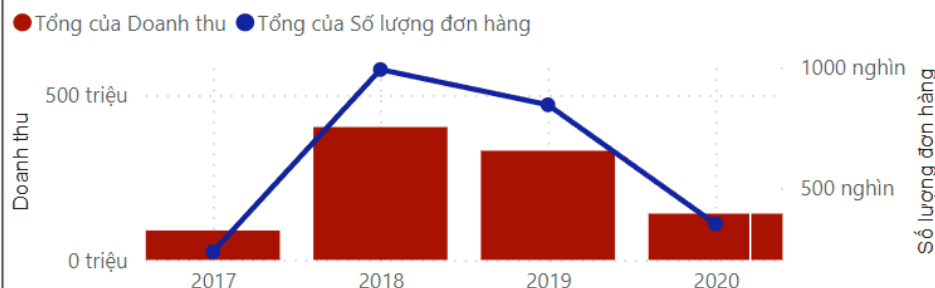
Hướng 1:
Mô tả số lượng đơn hàng và doanh thu theo thời gian



Phân loại số lượng đơn hàng và doanh thu theo thời gian

zomato

Doanh thu và Số lượng đơn hàng theo Năm

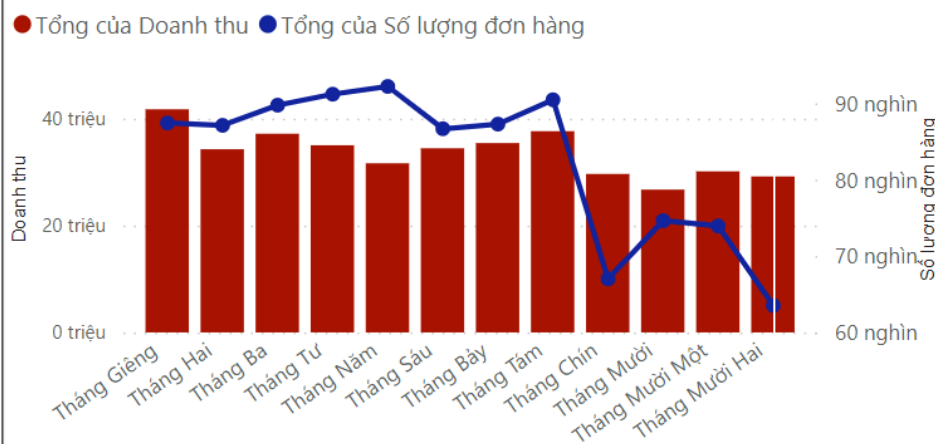


Số lượng đơn hàng và Doanh thu có xu hướng giảm từ năm 2019 đến năm 2020. Xu hướng này do ảnh hưởng của dịch Covid-19 (Bắt đầu từ tháng 11-2019)



Chuẩn bị kế hoạch đối ứng nếu có đại dịch xảy ra trong tương lai

Doanh thu và Số lượng đơn hàng theo Tháng (2018)

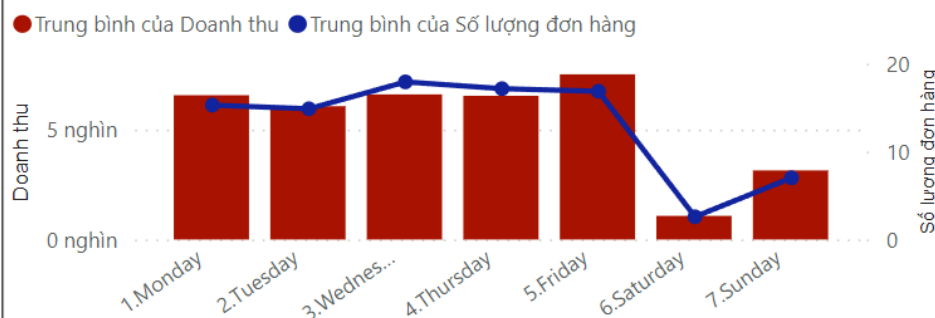


Số lượng đơn hàng và Doanh thu giảm dần từ tháng 1 đến tháng 9 và tăng mạnh từ tháng 10 đến tháng 11, đây là thời điểm có nhiều lễ hội ở Ấn Độ (Lễ hội Diwali)



Chuẩn bị đủ năng lực sản xuất và hệ thống phân phối từ tháng 9 để chuẩn bị cho mùa lễ hội

Doanh thu và Số lượng đơn hàng theo Ngày trong tuần (2018)



Lượng đặt hàng giảm mạnh vào những ngày cuối tuần

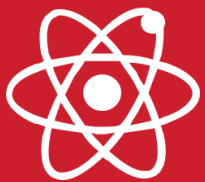


Chuẩn bị kế hoạch làm việc hằng ngày và sắp xếp công việc vào cuối tuần sao cho hợp lý (sắp xếp nhân lực, tăng ca, kế hoạch vận chuyển)

Phân Tích Dữ Liệu

Hướng 2:
Mô tả số lượng đơn hàng và doanh thu
theo thông tin nhà hàng

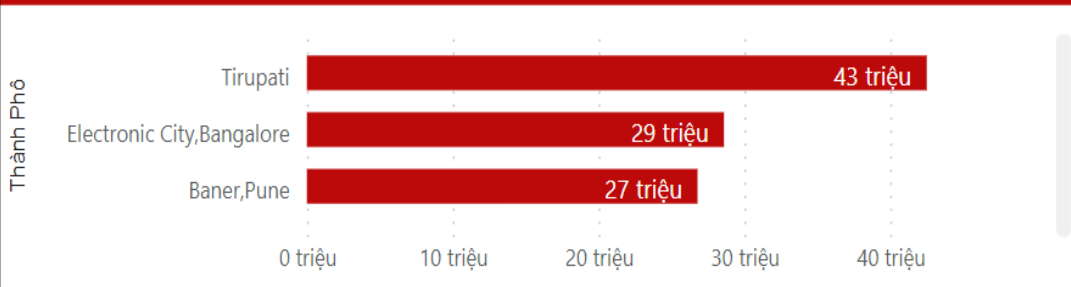




Phân loại Số lượng đơn hàng và Doanh thu theo thông tin nhà hàng

zomato

Doanh Thu theo Thành Phố



Top những thành phố có hoạt động kinh doanh sôi động



Nghiên cứu sự khác nhau nào tạo ra sự khác biệt để áp dụng cho những thành phố đang có doanh thu thấp

Doanh Thu theo Phong cách ẩm thực

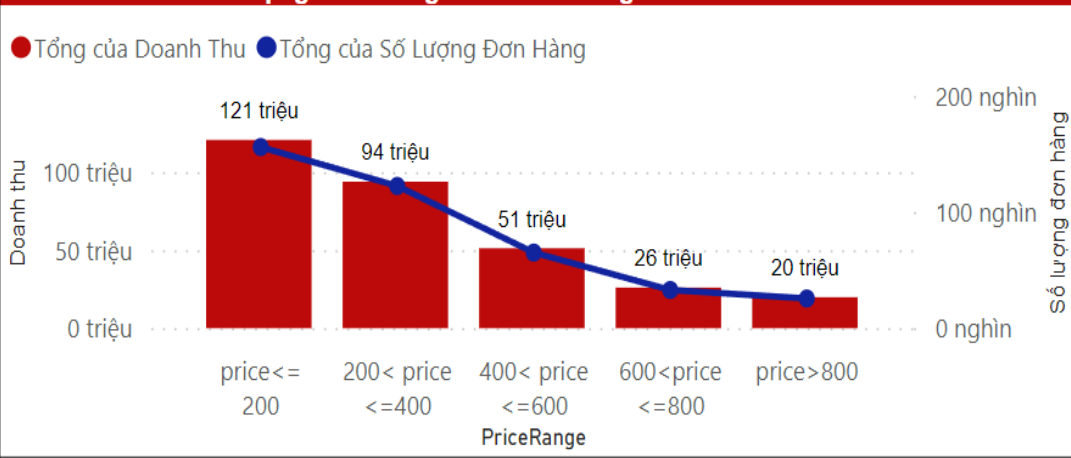


Top những nhà hàng có hoạt động kinh doanh về phong cách ẩm thực sôi động nhất



Nghiên cứu sự khác nhau nào tạo ra sự khác biệt để áp dụng cho những nhà hàng đang có doanh thu thấp

Doanh Thu và Số Lượng Đơn Hàng theo PriceRange



Những món ăn có giá thấp mang lại số lượng đơn hàng và doanh thu tốt nhất



Tập trung vào việc cải thiện chất lượng dịch vụ cho nhóm món ăn có giá thấp



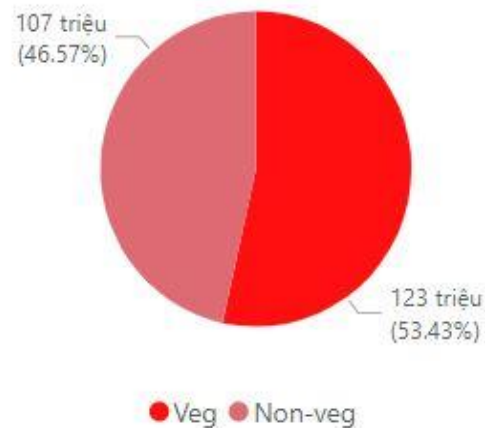
Phân loại Số lượng đơn hàng và Doanh thu theo thông tin nhà hàng

zomato

Tổng của Doanh Thu theo Month và Ăn chay và Mặn

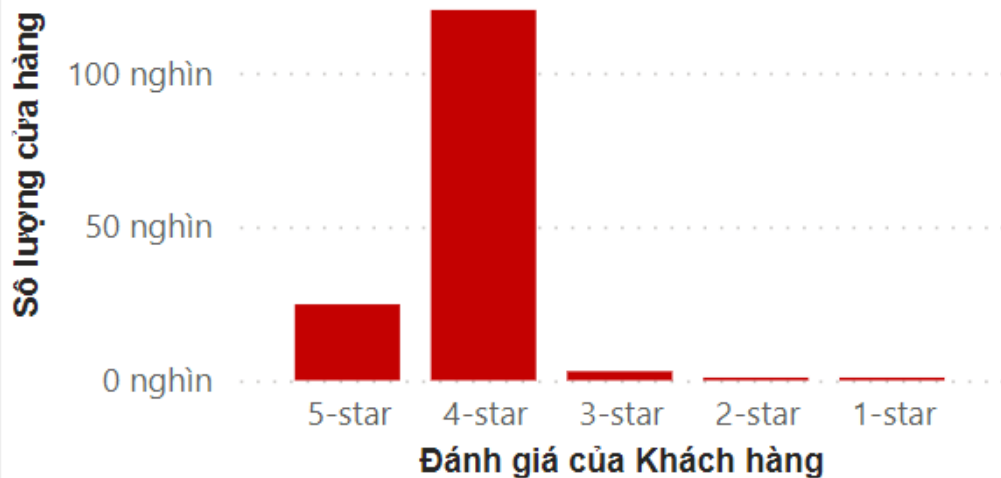


Tổng của Doanh Thu theo Ăn chay và Mặn



Xu hướng của nhóm ăn chay và không ăn chay khá cân bằng nhau và không cho thấy sự khác biệt lớn. Không cần chia chiến lược riêng cho từng nhóm

Số cửa hàng theo Đánh giá của Khách hàng



Phần lớn các nhà hàng được đánh giá 4 sao



Cần tìm hiểu lý do vì sao có những nhà hàng bị đánh giá thấp và có biện pháp cải thiện

Phân Tích Dữ Liệu

Hướng 3 :
Mô phỏng **số lượng đơn hàng** và
doanh thu theo nhân khẩu học



zomato



Phân loại Số lượng đơn hàng và Doanh thu theo nhân khẩu học

**Yếu Tố
thuộc về
nhân khẩu học
từ tập dữ liệu**



Tuổi

Trình độ

Giới Tính

**Số thành
viên gia
đình**

**Tình trạng
hôn nhân**

**Thu nhập
hàng
tháng**

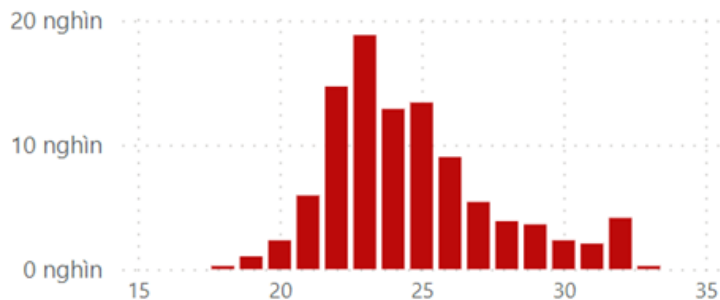
**Nghề
nghiệp**

zomato

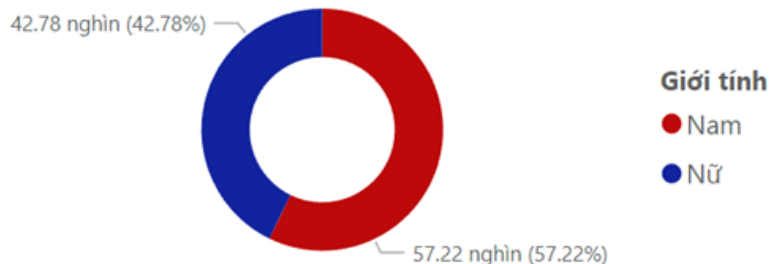


Phân loại Số lượng đơn hàng và Doanh thu theo yếu tố nhân khẩu học

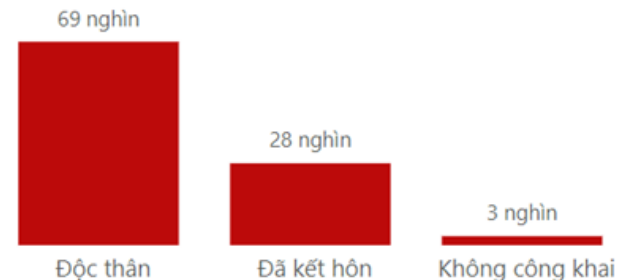
Số lượng tài khoản theo Độ tuổi



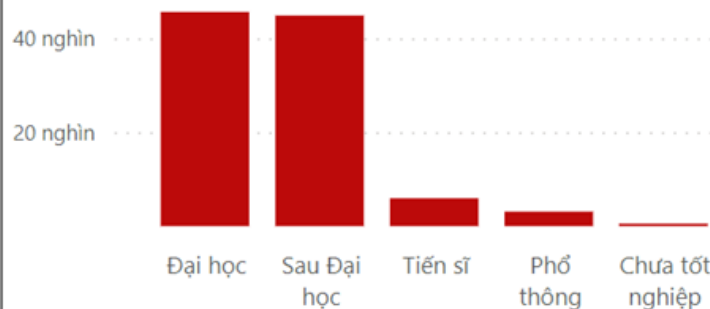
Số lượng tài khoản theo Giới tính



Số lượng tài khoản theo Tình trạng hôn nhân



Số Lượng tài khoản theo Trình độ học vấn



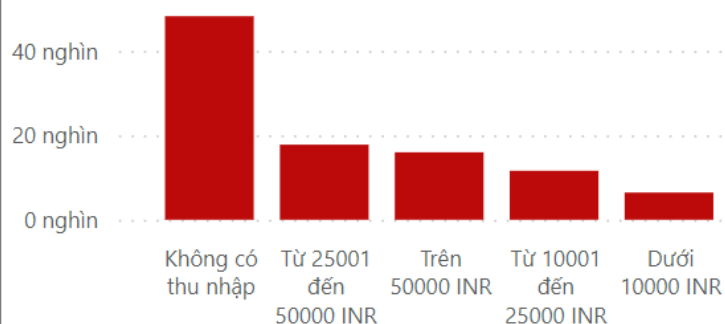
Số lượng tài khoản

100,000

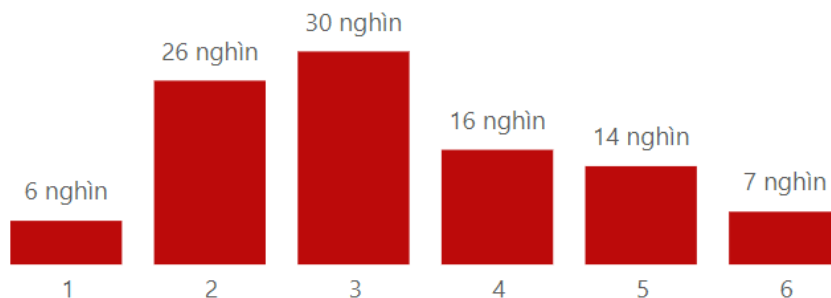
Số Lượng tài khoản theo Nghề nghiệp



Số Lượng tài khoản theo Thu nhập hàng tháng



Số Lượng tài khoản theo Số lượng thành viên gia đình





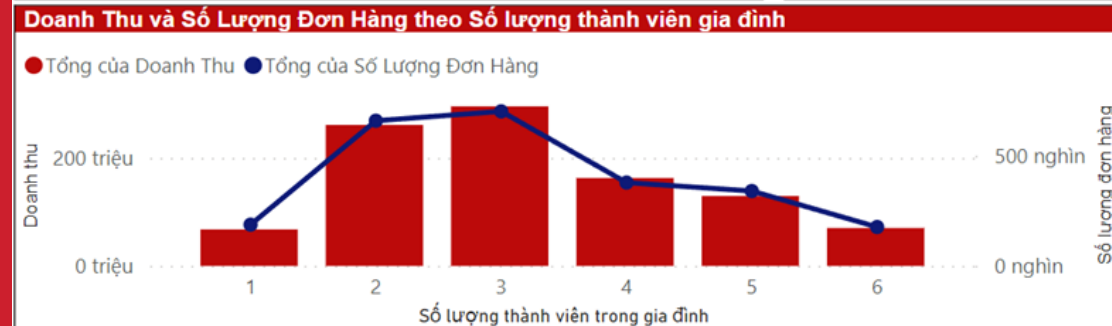
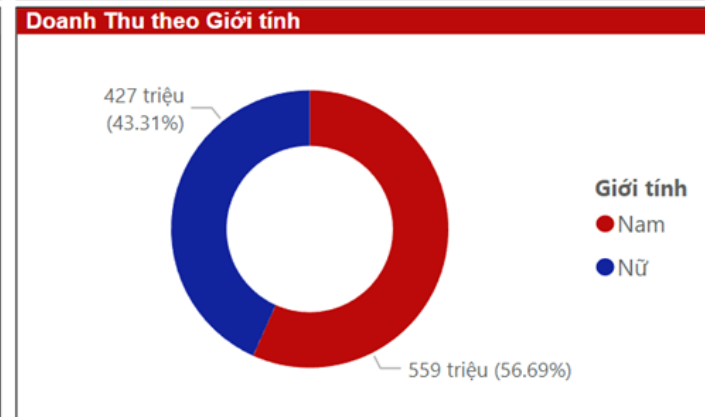
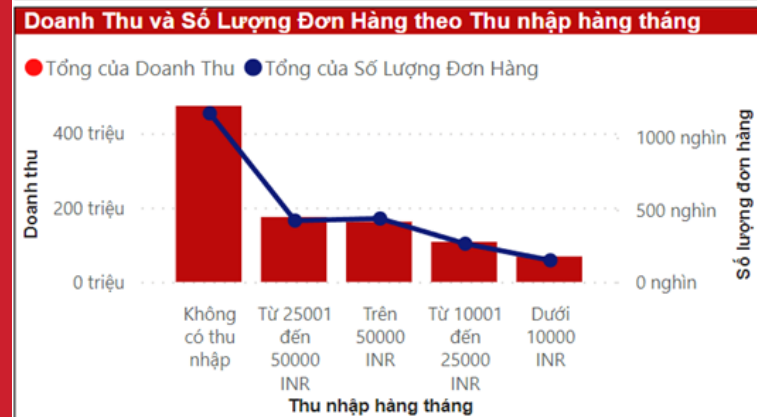
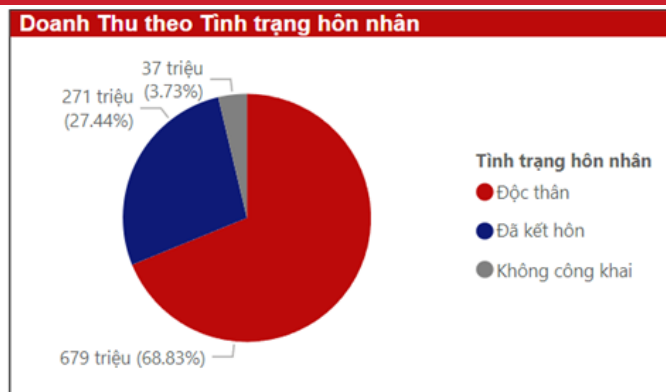
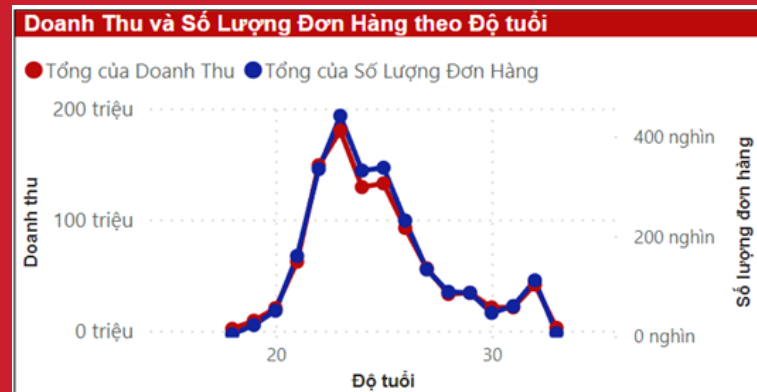
Phân loại Số lượng đơn hàng và Doanh thu theo yếu tố nhân khẩu học



Đặc điểm của tập khách hàng

- ★ 70% Khách hàng chủ yếu có độ tuổi từ 22 đến 25
- ★ Tỷ lệ khách hàng nam và nữ không chênh lệch lớn ~57% nam và ~43% nữ
- ★ 70% khách hàng chưa lập gia đình
- ★ Khoảng 50% khách hàng là sinh viên

Phân loại Số lượng đơn hàng và Doanh thu theo yếu tố nhân khẩu học



Ngày

10/4/2017

6/26/2020



Phân loại số lượng đơn hàng và Doanh thu theo yếu nhân khẩu học



Đơn hàng và Doanh thu theo tập khách hàng



Độ tuổi chi tiêu nhiều nhất cho App là từ 22 - 26 tuổi



Nhóm khách hàng đóng góp phần lớn doanh thu là nhóm có trình độ đại học và sau đại học



Nhóm khách hàng trong gia đình có từ 2 đến 3 thành viên là nhóm đóng góp nhiều nhất cho doanh thu

Phân Tích Dữ Liệu

**Hướng 4 :
Phân tích doanh thu theo khu vực**

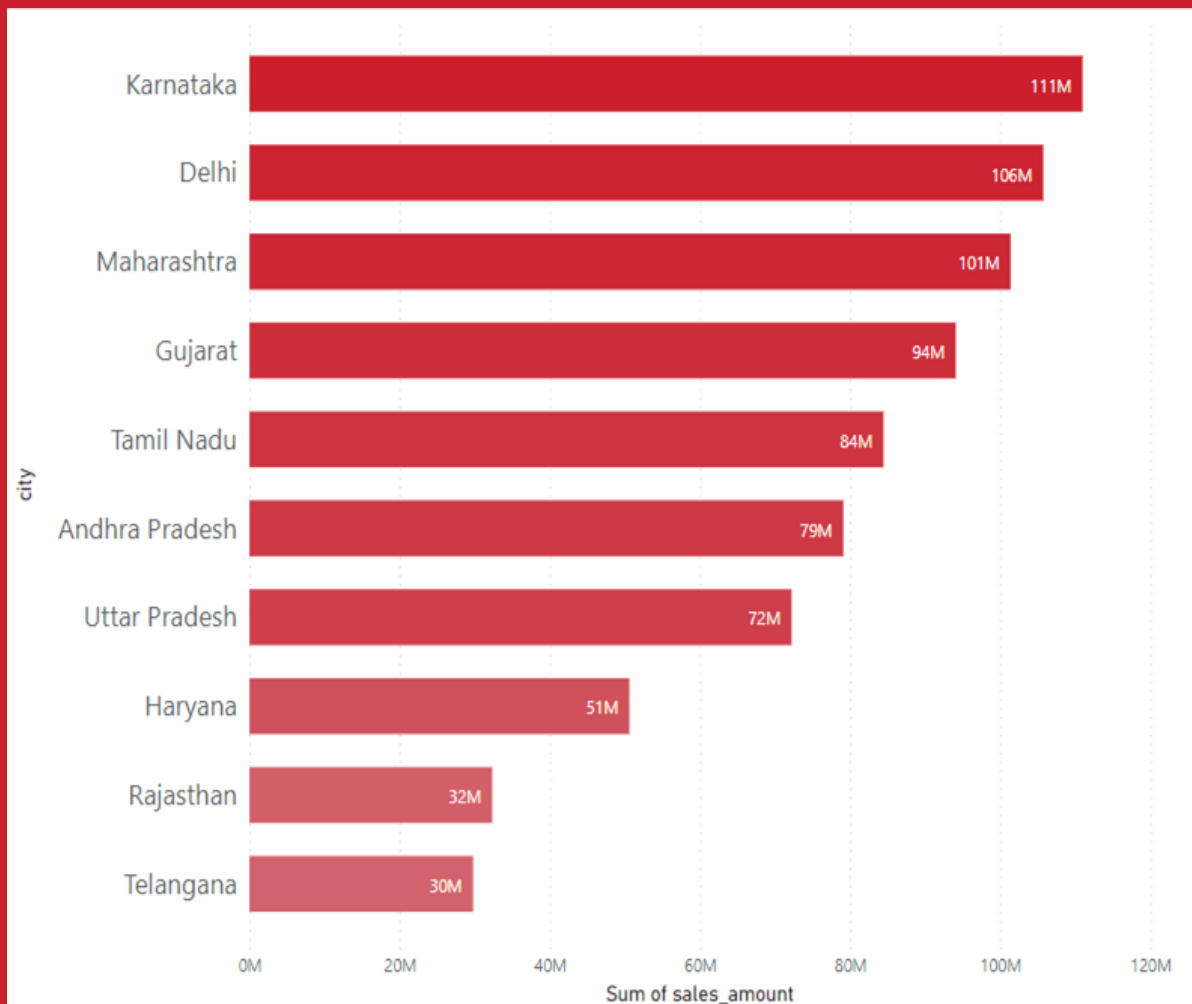
zomato

zomato

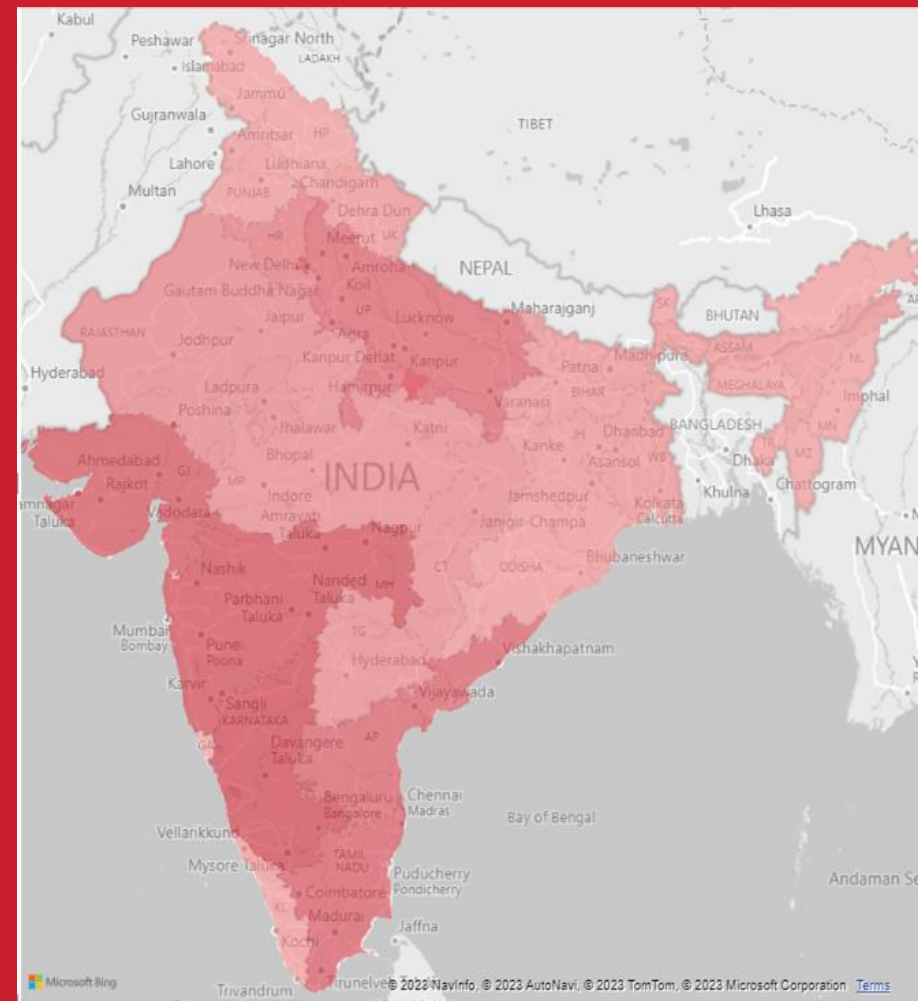
Phân tích doanh thu theo khu vực

zomato

Top 10 bang có doanh thu cao nhất



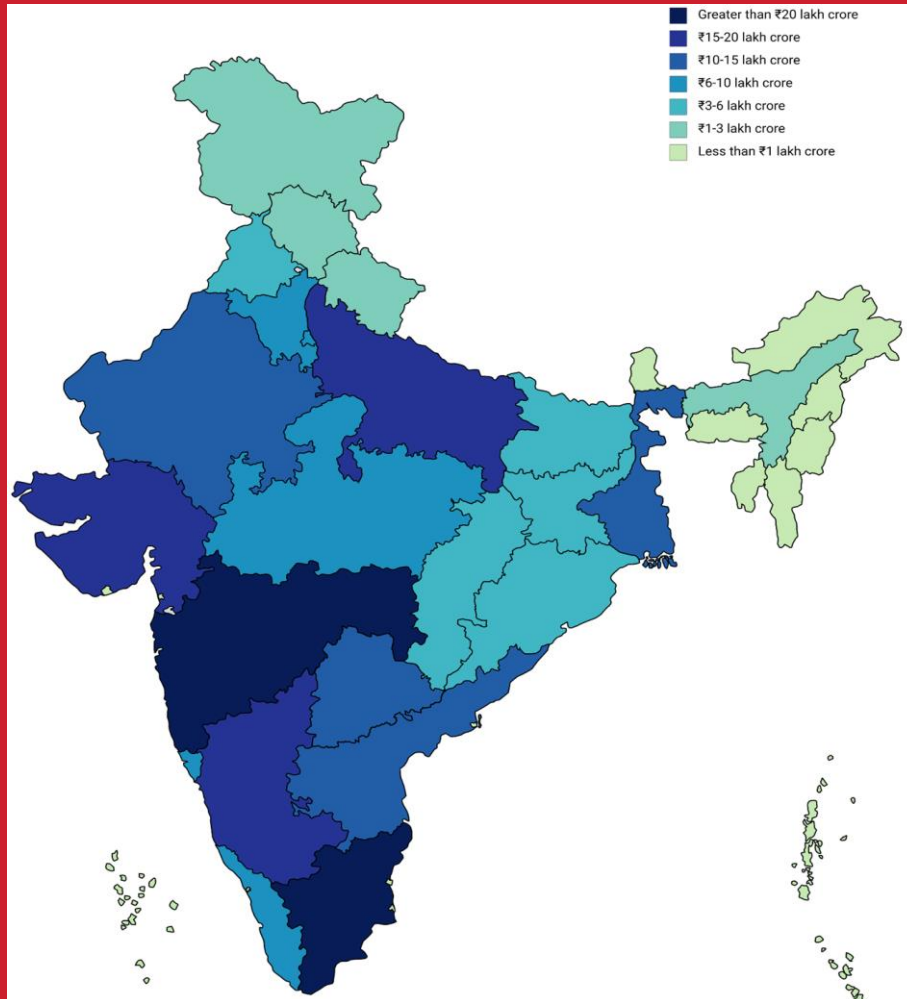
Tổng doanh thu theo bang



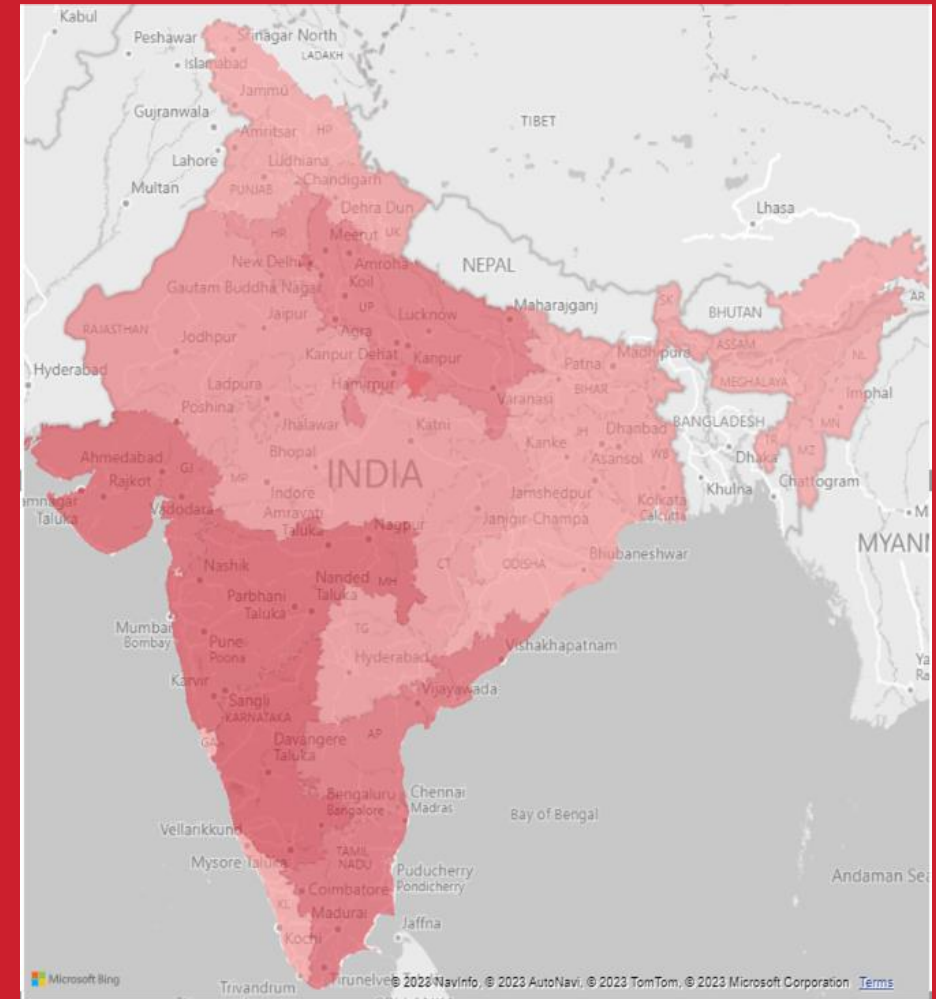
Phân tích doanh thu theo khu vực

zomato

GDP bình quân đầu người 2020 (Theo Wikipedia)



Tổng doanh thu theo bang



zomato
DELIVERY PARTNER



3. Tổng kết và đề xuất chiến lược kinh doanh

Tổng kết và đề xuất chiến lược kinh doanh

- ✓ Tập trung vào các món ăn có giá thấp và nghiên cứu để điều chỉnh giá các món ăn có giá quá cao
- ✓ Tập trung cải thiện và nâng cao chất lượng dịch vụ cho nhóm khách hàng chính và mở rộng thêm tập khách hàng trong tương lai.
- ✓ Chuẩn bị tốt mọi thứ cần thiết trước mùa lễ hội vào tháng 10 và tháng 11
- ✓ Xem xét kế hoạch công việc hằng ngày và những ngày cuối tuần để có sự sắp xếp hợp lý
- ✓ Nghiên cứu để áp dụng các điểm khác biệt tạo nên kết quả tốt ở một số nhà hàng và thành phố

Tài liệu



Nguồn Data: <https://www.kaggle.com/datasets/anas123siddiqui/zomato-database>



Xử lý dữ liệu: <https://colab.research.google.com/drive/1mYbN-WMYu0wQmxcbJTbpzBaiB5GGPH41>



PowerBI: https://drive.google.com/file/d/1zPm_ZTAjw2rzejk-H3147IzuuSL-ArHe/view?usp=sharing



***THANK YOU
FOR LISTENING !!!***

zomato