Final Project Presentation
By Group 4
March 6, 2018

MASY1-GC3500
Chia-Wei Lin, Minyi Huang, Liyingjian Liu
School of Professional Studies
New York University

Introduction

About Yelp

Purpose: To connect people with great local businesses

Founded in *2004*

Unique Visitors via mobile **app**: 29M monthly average

Reviews by the end of Q4 2017: 148M

Yelp makes money by **selling ads** to local businesses.

yelp

Large market opportunity

20+ million
Local business locations
in the U.S.

- Claimed local business locations: 4.2 million
- Paying advertising accounts: 163, 000

yelp

Yelp Data Model

Project Challenge

The Dataset

5,200,000 reviews

1,000,000 reviews loaded

174,000 businesses

200,000 pictures

11 metropolitan areas

1,100,000 tips by 1,300,000 users

Over 1.2 million business attributes like hours, ~~parking, availability, and ambience~~

Aggregated check-ins over time for each of the 174,000 businesses

yelp

Key Findings

Key Findings

1. Business Review

2. Business Wideness

3. Category

4. Restaurant

5. City

yelp

Business Review

Business Review Ranking (by Review Count)

BUSINESS NAME	REVIEW NUM	STARS	USEFUL	FUNNY	COOL
Starbucks	**382**	**3**	**382**	**382**	**382**
McDonald's	**269**	**2**	**269**	**269**	**269**
Chipotle Mexican Grill	**190**	**3**	**190**	**190**	**190**
Pita Jungle	149	3.8	149	149	149
Dunkin' Donuts	138	2.6	138	138	138
Olive Garden Italian Restaurant	135	3.3	135	135	135
In-N-Out Burger	133	4.1	133	133	133
Yard House	127	3.8	127	127	127
The Cheesecake Factory	125	3.5	125	125	125
Subway	125	2.7	125	125	125

REVIEW_NUM

yelp
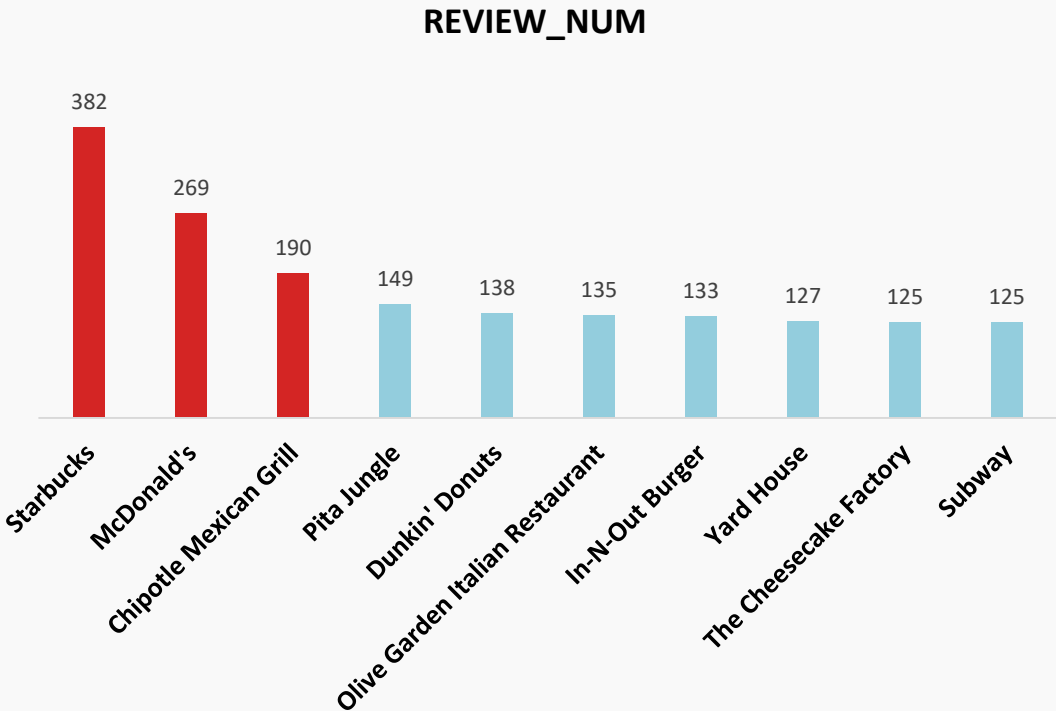
Code: Business Review Ranking (by Review Count)

```sql
select *
from (
    select replace(business_name,'"', '') business_name,
    count(user_name) review_num, round(avg(gbu.stars), 1) stars,
    count(gre.useful) useful,    count(gre.funny) funny, count(gre.cool) cool

    from gp_review gre, gp_business gbu, gp_user gus
    where gre.business_id = gbu.business_id
    and gre.user_id = gus.user_id

    group by replace(business_name,'"', '')
    order by count(user_name) desc
    )
where rownum <= 10
```

yelp

Business Review Ranking > 100

BUSINESS NAME	REVIEW NUM	STARS	USEFUL	FUNNY	COOL
In-N-Out Burger	**133**	**4.1**	**133**	**133**	**133**
Grimaldi's Pizzeria	**119**	**3.9**	**119**	**119**	**119**
Pita Jungle	**149**	**3.8**	**149**	**149**	**149**
Yard House	127	3.8	127	127	127
Cheesecake Factory	125	3.5	125	125	125
Smashburger	119	3.5	119	119	119
Olive Garden Italian Restaurant	135	3.3	135	135	135
Jimmy John's	108	3.3	108	108	108
Starbucks	382	3	382	382	382
Chipotle Mexican Grill	190	3	190	190	190

Review > 100 Order by Stars

yelp

Code: Business Review Ranking (Review > 100)

```sql
select *
from (
    select replace(business_name,'"', '') business_name,
    count(user_name) review_num, round(avg(gbu.stars), 1) stars,
    count(gre.useful) useful,   count(gre.funny) funny, count(gre.cool) cool

    from gp_review gre, gp_business gbu, gp_user gus

    where gre.business_id = gbu.business_id
    and gre.user_id = gus.user_id

    group by replace(business_name,'"', '')
    having count(user_name) >= 100
    order by round(avg(gre.stars), 1) desc
    )
where rownum <= 10
```

yelp

Business Review Ranking > 50

BUSINESS NAME	REVIEW NUM	STARS	USEFUL	FUNNY	COOL
John Mull's Meats & Road Kill Grill	68	4.5	68	68	68
Snooze An AM Eatery	66	4.5	66	66	66
The Great Greek	57	4.5	57	57	57
Backyard Taco	56	4.5	56	56	56
Harumi Sushi	55	4.5	55	55	55
Detroit Coney Grill	52	4.5	52	52	52
The Golden Tiki	51	4.5	51	51	51
Dutch Bros Coffee	70	4.4	70	70	70
Viva Chicken	50	4.4	50	50	50
MOD Pizza	61	4.3	61	61	61

Review > 50 Order by Stars

yelp

Code: Business Review Ranking (Review > 50)

```sql
select *
from (
    select replace(business_name,'"', '') business_name,
    count(user_name)      review_num, round(avg(gbu.stars), 1) stars,
    count(gre.useful) useful,    count(gre.funny) funny, count(gre.cool) cool

    from gp_review gre, gp_business gbu, gp_user gus

    where gre.business_id = gbu.business_id
    and gre.user_id = gus.user_id

    group by replace(business_name,'"', '')
    having count(user_name) >= 50
    order by round(avg(gre.stars), 1) desc
    )
where rownum <= 10
```

Business Wideness

Top Business and its Wideness

BUSINESS NAME	BUSINESS WIDNESS
Oxi Fresh Carpet Cleaning	**15116**
Uber	**12093**
George Street Photo & Video	**12092**
Clean Air Lawn Care	6046
iCracked	6046
Southwest Gas Corporation	6046
Handyman Services	6046
Neighborhood Garage Door Services	6046
Megabus	6046
ExecuCar	6046

BUSINESS_WIDNESS

yelp
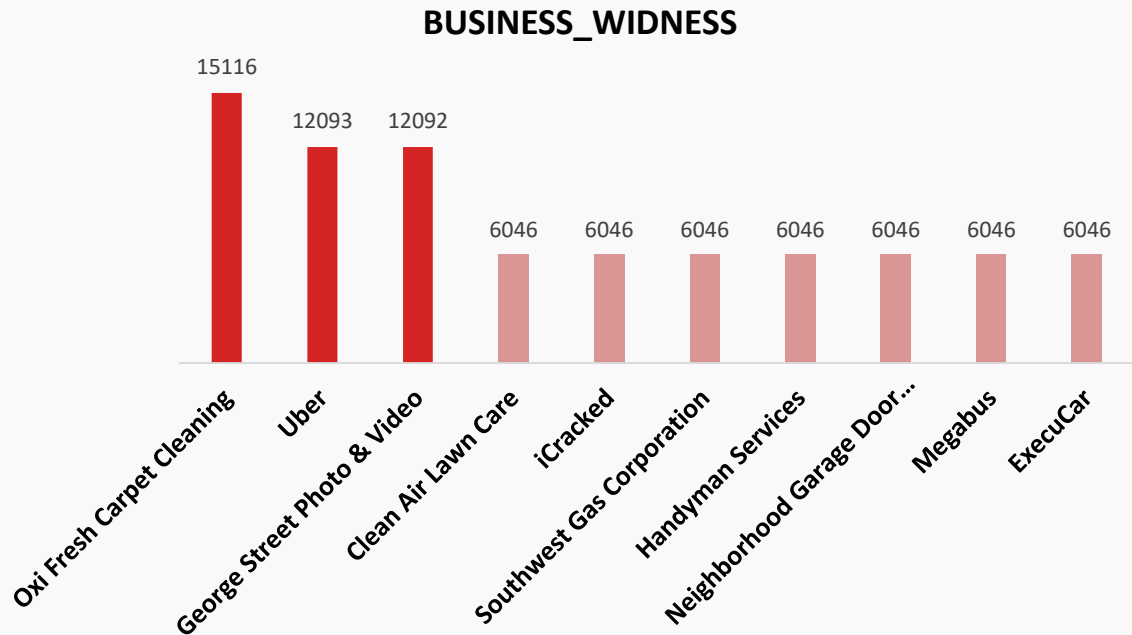
Code: Business Wideness

```sql
select *
from (
        select replace(gbu.business_name, '"', '') business_name,
        count(glo.address) as business_widness
        from gp_business gbu, gp_business_location gbl, gp_location glo
        where gbu.business_id = gbl.business_id
        and gbl.location_id = glo.location_id
        group by replace(gbu.business_name,'"', '')
        order by business_widness desc
    )
where rownum <= 10
```

Category

Business in Category

CATEGORY NAME	NUMBER
Restaurants	**25611**
Shopping	**13025**
Food	**11515**
Beauty & Spas	8069
Home Services	7713
Health & Medical	6797
Nightlife	5733
Local Services	5306
Automotive	5180
Bars	4987

Busines Number in Categories

yelp
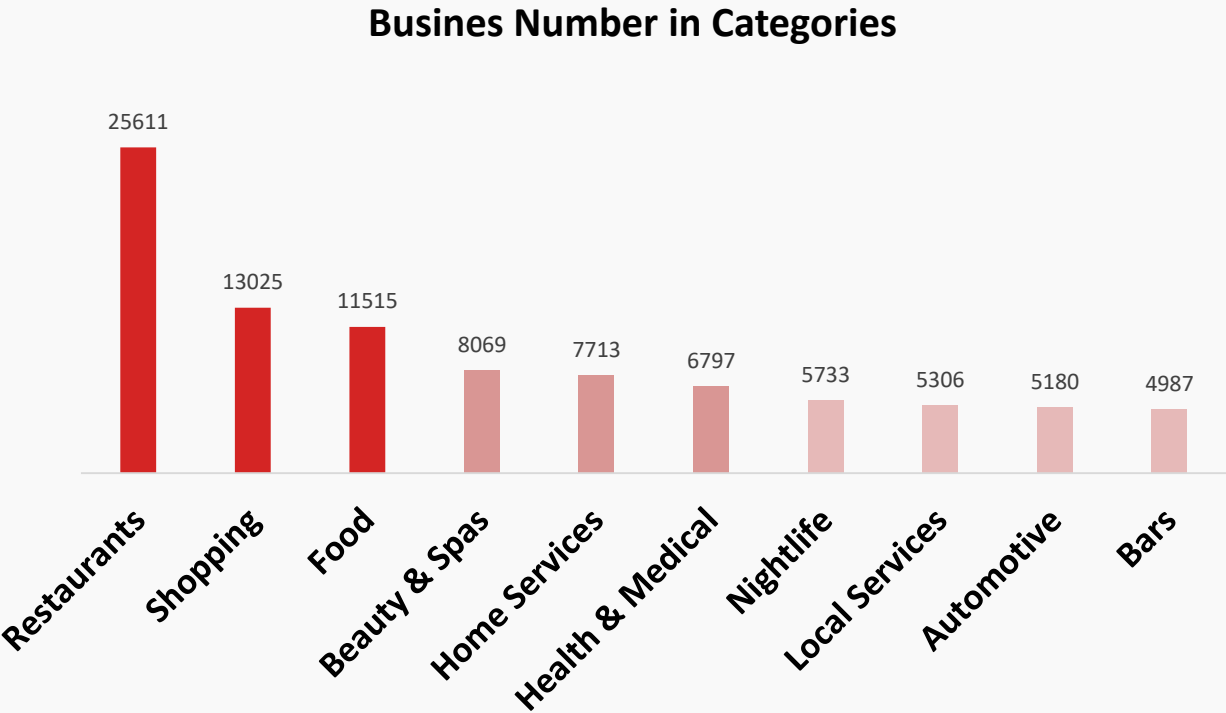
Code: Business in Category

```sql
select *
from (
        select category_name, count (business_id) as "number"
        from gp_business
        join gp_business_category using (business_id)
        join gp_category using (category_id)

        group by category_name
        order by count (business_id) desc
        )
where rownum <= 10
```

yelp

Business in Category (by Stars)

CATEGORY NAME	REVIEW NUMBER	STARS
Beauty & Spas	9128	4.1
Home Services	6563	4
Food	20157	3.9
Health & Medical	5474	3.9
Automotive	5503	3.8
American (New)	9670	3.8
Shopping	9012	3.8
Sandwiches	6695	3.8
Breakfast & Brunch	8403	3.8
Italian	6262	3.8

Category in Review & Star

yelp
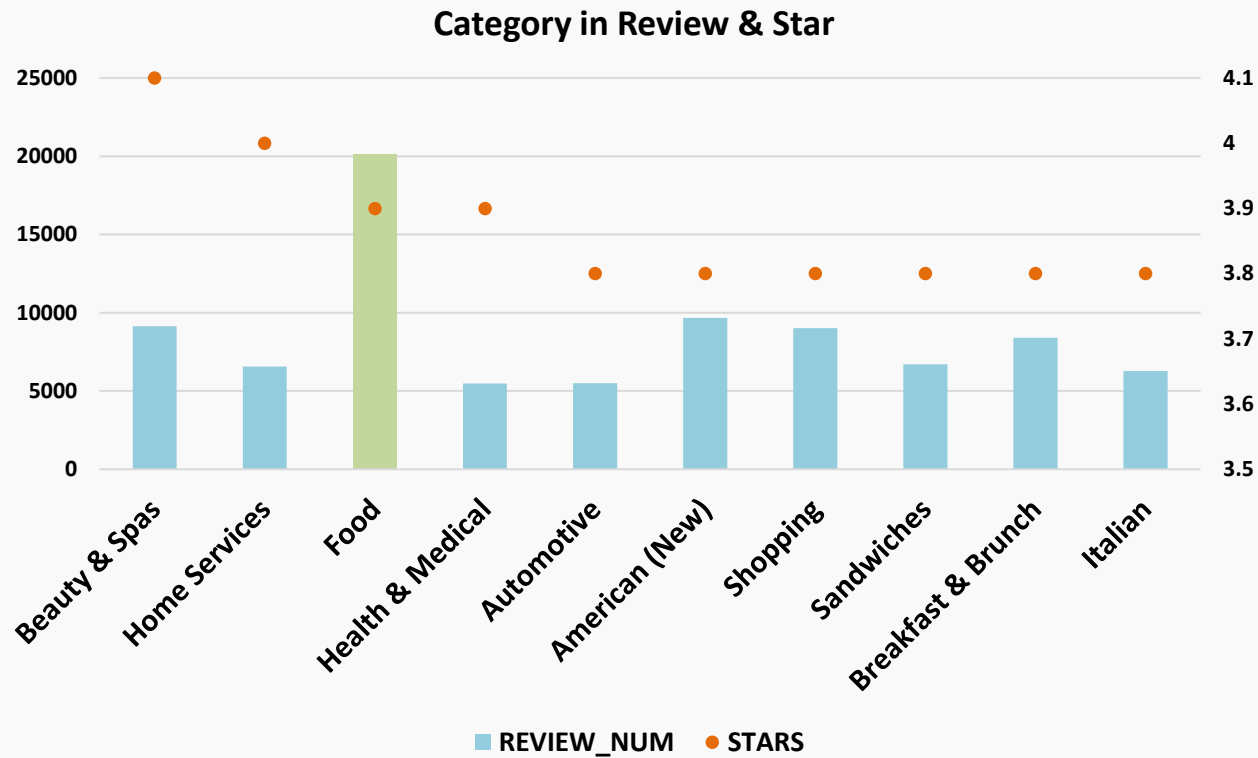
Code: Business in Category (by Stars)

```sql
select *
from (
        select category_name, count (user_id) as review_num, round(avg(gre.stars), 1) stars
        from gp_business gbu, gp_business_category gbc, gp_category gca, gp_review gre

        where gre.business_id = gbu.business_id
        and gbu.business_id = gbc.business_id
        and gbc.category_id = gca.category_id

        group by category_name
        having count (user_id) >= 5000
        order by stars desc
    )
where rownum <= 10
```

yelp

Business in Category (by Review Count)

CATEGORY NAME	REVIEW NUMBER	STARS
Restaurants	**66060**	**3.7**
Food	**20157**	**3.9**
Nightlife	**16501**	**3.7**
Bars	15487	3.7
American (Traditional)	10607	3.5
American (New)	9670	3.8
Beauty & Spas	9128	4.1
Shopping	9012	3.8
Breakfast & Brunch	8403	3.8
Pizza	6986	3.7

Category with Most Reviews

yelp
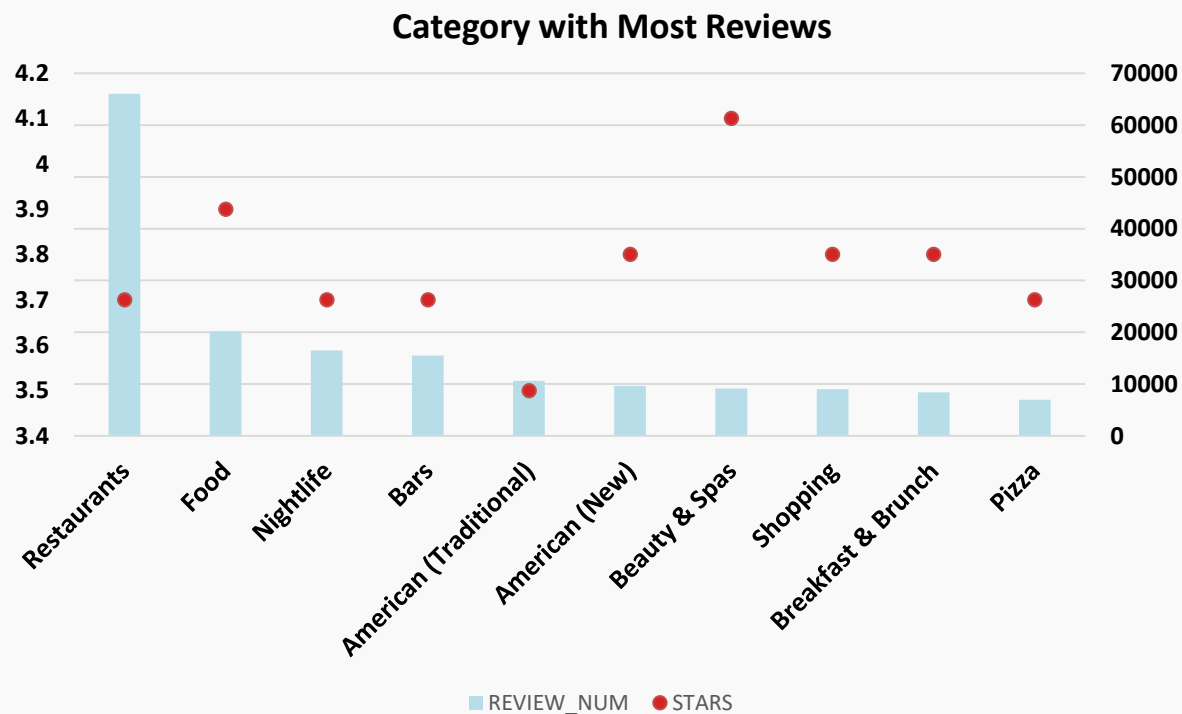
Code: Business in Category (by Review Count)

```
create view  gp_cat_mostreview as

select *
from (

    select category_name, count (user_id) as review_num, round(avg(gre.stars), 1) stars
    from gp_business gbu, gp_business_category gbc, gp_category gca, gp_review gre

    where gre.business_id = gbu.business_id
    and gbu.business_id = gbc.business_id
    and gbc.category_id = gca.category_id

    group by category_name
    order by count (user_id) desc
    )
where rownum <= 10
```

yelp

Restaurant with Most Review

BUSINESS NAME	REVIEW NUMBER
McDonald's	**269**
Chipotle Mexican Grill	**190**
Pita Jungle	**149**
Olive Garden Italian Restaurant	135
In-N-Out Burger	133
Yard House	127
Taco Bell	125
Subway	125
The Cheesecake Factory	125
Buffalo Wild Wings	123

REVIEW_NUM bar chart: McDonald's 269, Chipotle Mexican Grill 190, Pita Jungle 149, Olive Garden Italian... 135, In-N-Out Burger 133, Yard House 127, Taco Bell 125, Subway 125, The Cheesecake Factory 125, Buffalo Wild Wings 123

yelp
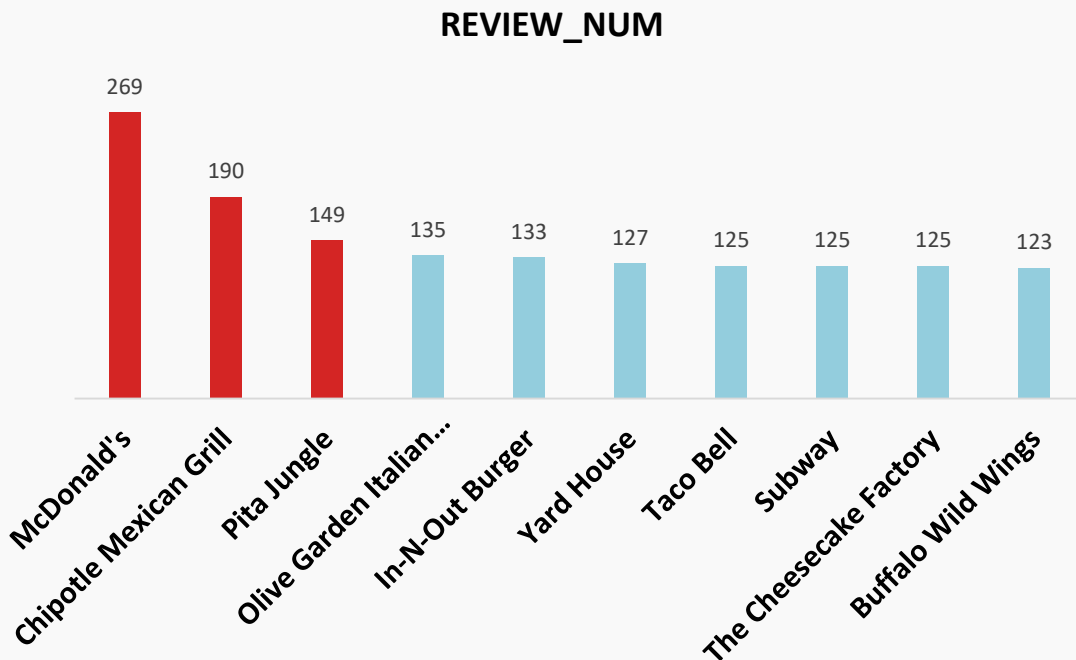
Code: Restaurants with Most Review

```sql
select *
from (

    select replace (gbu.business_name, '"', '') business_name, gbu.review_count
    from gp_business gbu, gp_business_category gbc, gp_category gca

    where gbu.business_id = gbc.business_id
    and gbc.category_id = gca.category_id
    and category_name = 'Restaurants'

    order by review_count desc
    )
where rownum <= 10
```

yelp

Top 3 Restaurants Locations Distribution - McDonald's

CITY	BRANCH WIDNESS
Las Vegas	76
Phoenix	63
Toronto	37
Charlotte	22
Mesa	18
Brampton	16
Mississauga	15
Glendale	14
Cleveland	13
Scottsdale	13
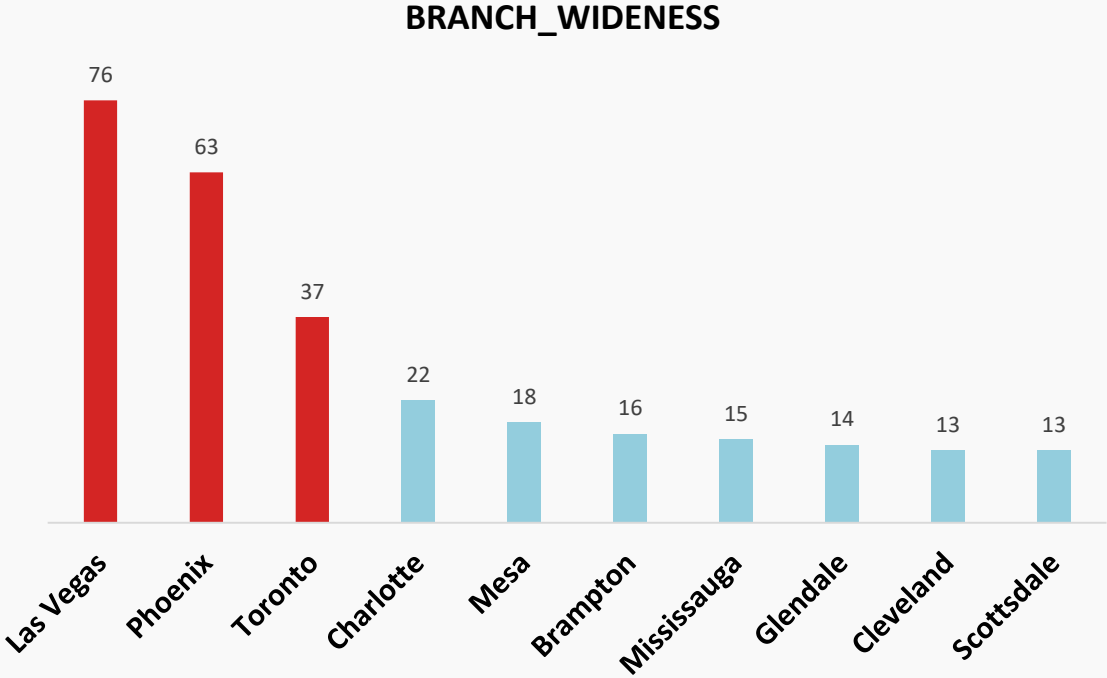
BRANCH_WIDENESS

yelp

Code: McDonald's

```sql
select city, count(gbu.business_id) branch_widness
from gp_business gbu, gp_business_location gbl, gp_location glo

where gbu.business_id = gbl.business_id
and gbl.location_id = glo.location_id
and business_name like '"McDonald%'

group by city
order by branch_widness desc;
```

Top 3 Restaurants Locations Distribution - Chipotle Mexican Grill

CITY	BRANCH WIDNESS
Glendale	**21**
Phoenix	**12**
Las Vegas	**12**
Pittsburgh	**12**
Pineville	6
Charlotte	6
Scottsdale	5
Toronto	3
Tempe	3
Peoria	2

BRANCH_WIDENESS

Code: Chipotle Mexican Grill

```sql
select city, count(gbu.business_id) branch_widness
from gp_business gbu, gp_business_location gbl, gp_location glo

where gbu.business_id = gbl.business_id
and gbl.location_id = glo.location_id
and business_name = '"Chipotle Mexican Grill"'

group by city
order by branch_widness desc;
```
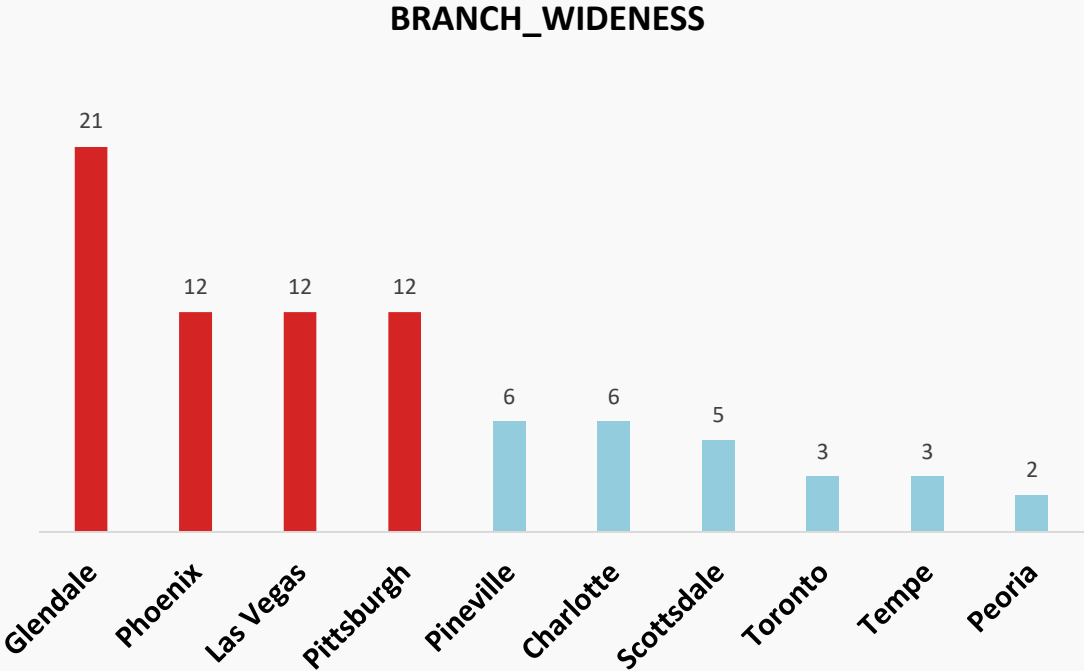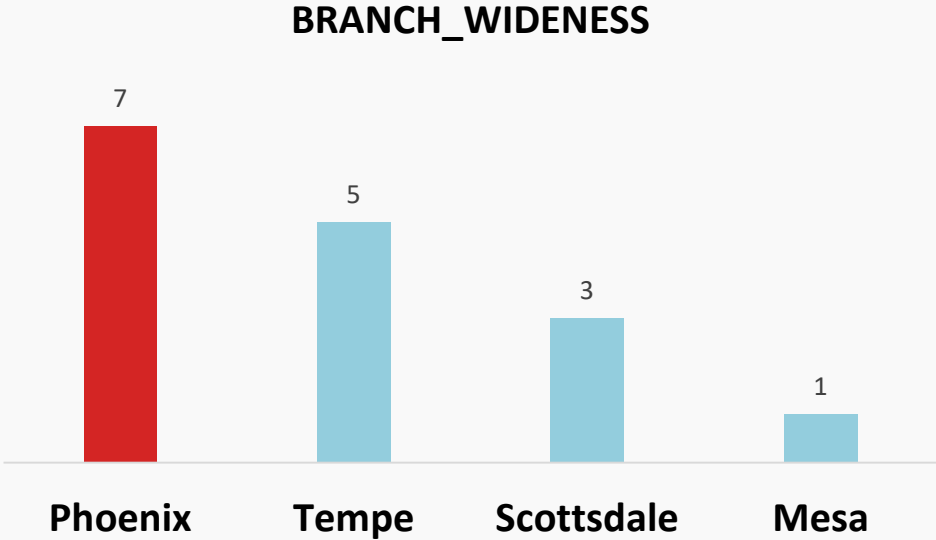
Top 3 Restaurants Locations Distribution - PITA Jungle

CITY	BRANCH WIDNESS
Phoenix	7
Tempe	5
Scottsdale	3
Mesa	1

BRANCH_WIDENESS

Code: PITA Jungle

```sql
select city, count(gbu.business_id) branch_widness
from gp_business gbu, gp_business_location gbl, gp_location glo

where gbu.business_id = gbl.business_id
and gbl.location_id = glo.location_id
and business_name = '"Pita Jungle"'

group by city
order by branch_widness desc
```

yelp

City

Business in City

CITY	STATE	BUSINESS_NUM
Las Vegas	**NV**	**2,288,905**
Phoenix	**AZ**	**1,650,035**
Scottsdale	**AZ**	**650,215**
Mesa	AZ	541,387
Charlotte	NC	460,456
Toronto	ON	359,546
Henderson	NV	351,481
Gilbert	AZ	337,193
Chandler	AZ	308,993
Pittsburgh	PA	274,058

BUSINESS_NUM

yelp
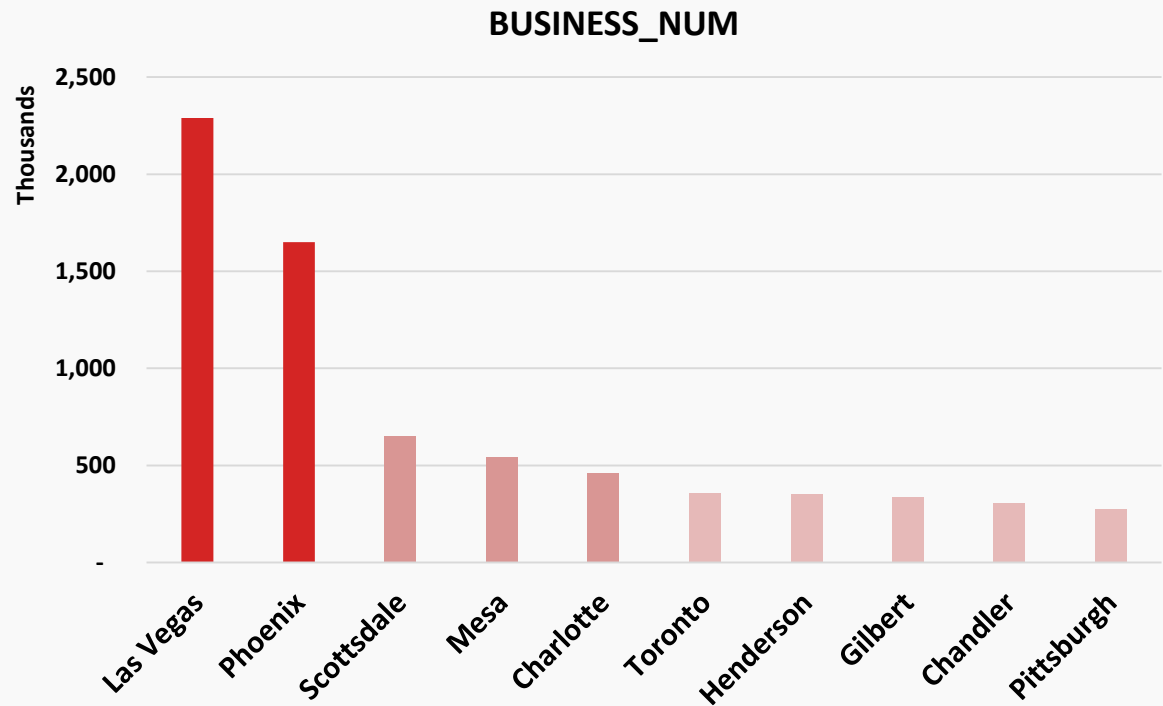
Code: Business in City

```sql
select city, state_name state, count(gbu.business_id) as business_num

from gp_business gbu, gp_business_location gbl, gp_location glo, gp_state gst,
     gp_business_category gbc, gp_category gca

where gbu.business_id = gbl.business_id
        and gbl.location_id = glo.location_id
        and glo.state_id = gst.state_id
        and gbu.business_id = gbc.business_id
        and gbc.category_id = gca.category_id
        and category_name in ('Restaurants', 'Shopping', 'Food')

group by city, state_name
order by count(gbu.business_id) desc
```
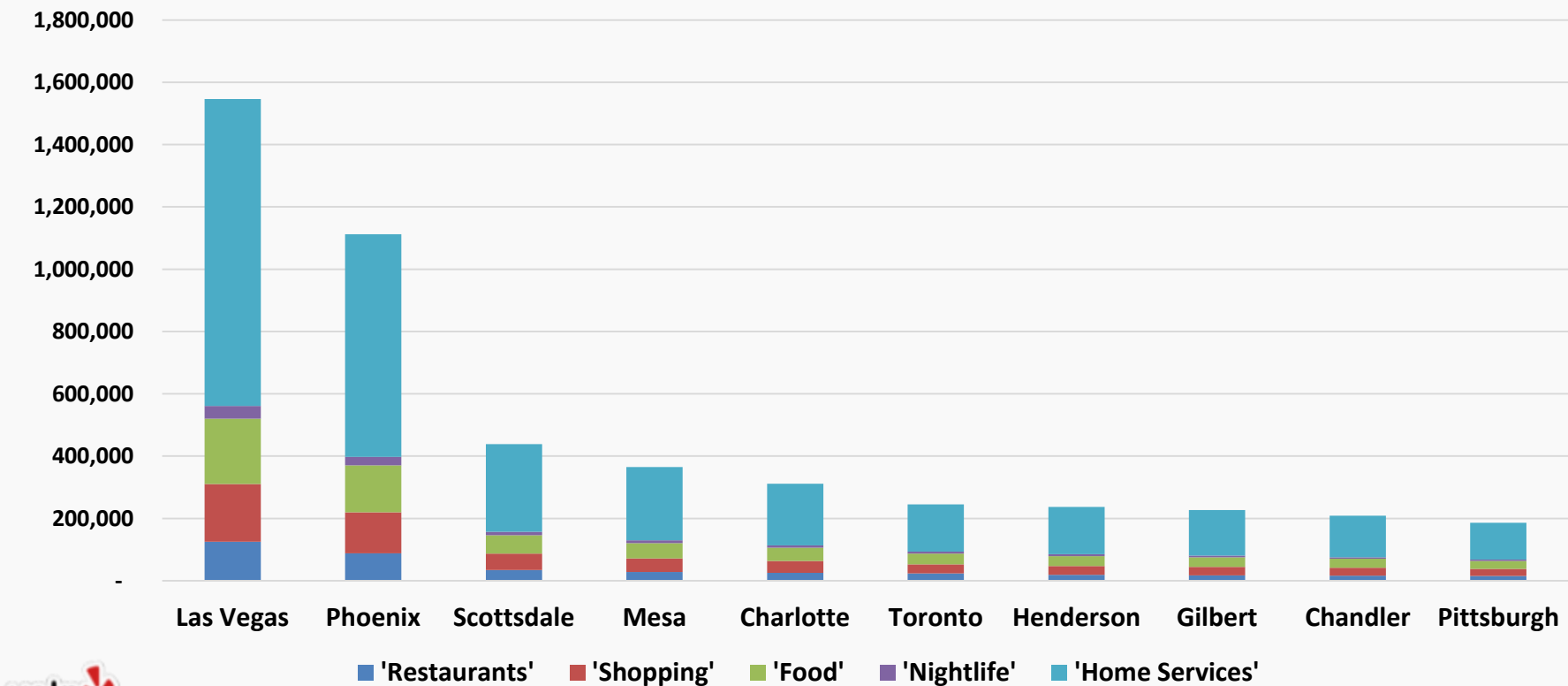
Business in City in Five Categories

CITY	Restaurants	Shopping	Food	Nightlife	Home Services
Las Vegas	125,028	185,743	209,392	41,253	984,643
Phoenix	87,576	132,252	150,409	27,710	714,500
Scottsdale	34,390	52,853	59,078	11,043	280,945
Mesa	28,546	43,305	49,261	9,018	234,663
Charlotte	25,530	37,745	42,466	8,014	197,765
Toronto	23,109	29,709	34,132	6,794	151,537
Henderson	18,931	28,380	32,063	6,047	151,634
Gilbert	17,599	26,911	30,689	5,605	146,334
Chandler	16,497	25,658	28,169	5,173	133,175
Pittsburgh	15,823	22,383	25,375	4,913	117,331

Business in City

Top City with Most Number of Business in Five Categories

Legend: 'Restaurants' 'Shopping' 'Food' 'Nightlife' 'Home Services'

yelp

Code: Business in Five Categories

```sql
select *
from
(select city, category_name
from gp_business gbu
join gp_business_category using (business_id)
join gp_category using (category_id)
join gp_business_location using (business_id)
join gp_location using (location_id)

where city in ('Las Vegas' , 'Phoenix','Scottsdale','Mesa',
        'Charlotte','Toronto','Henderson','Gilbert','Chandler','Pittsburgh'))

pivot (
        count (*)
        for category_name in ( 'Restaurants', 'Shopping', 'Food','Nightlife', 'Home Services') )
```

Conclusion

Conclusion

1. Most businesses are **chain stores**, so they have lots of reviews.

2. When a business has **more reviews**, it usually has **lower star ranks**.

3. Less review, higher star ranks.

4. Services based business have the most branches.

5. Yelp focus on **restaurants**, that's why restaurants have the most reviews.

yelp

Reference

https://www.kaggle.com/yelp-dataset/yelp-dataset
https://www.yelp.com/about
https://www.yelp.com/dataset

The entire final project represents our group own work and original work in accordance with University regulations.

yelp