

# SQL Final Project

December 9, 2022

## 1 Final Project Information

### 1.1 Team members:

Ruoxuan Yan: 506082695

Yijin Wang: 706085527

Ziyu Zeng: 306092539

```
[1]: import requests
import pandas as pd
```

## 2 UCLA food & reviews

```
[2]: Client_id = 'C660FLSiEvVbTy1iaAXzkA'

api_key = '
↪FCsFoBj1LGie6-HG7fTX0cIow-KG3Gb7fUDVH1HlizqnpB_K2gMRGkZ01Z9ecMBjQuA3FxpIots3HDkFSGqv-3z9iGs'
```

```
[3]: df_all = pd.DataFrame()
list_i = [0,50,100,150,200,250,300]
for i in list_i:
    url = f"https://api.yelp.com/v3/businesses/search?
↪location=UCLA&categories=food&sort_by=distance&limit=50&offset={i}"
    headers = {
        "accept": "application/json",
        "Authorization": "Bearer
↪FCsFoBj1LGie6-HG7fTX0cIow-KG3Gb7fUDVH1HlizqnpB_K2gMRGkZ01Z9ecMBjQuA3FxpIots3HDkFSGqv-3z9iGs"
    }
    response = requests.get(url, headers=headers)
    df_i = pd.DataFrame(response.json()["businesses"])
    df_all = pd.concat([df_all,df_i],ignore_index = True)
```

```
[4]: df_all['xuhao'] = range(1,len(df_all)+1)
```

```
[5]: df_all['latitude'] = 0
df_all['longitude'] = 0
df_all['address_detail'] = 0
df_all['city'] = 0
```

```

df_all['zip_code'] = 0
df_all['state'] = 0
df_all['country'] = 0
df_all['cate_alias'] = 0
df_all['cate_title'] = 0

for i in range(len(df_all)):
    df_all.loc[i, 'latitude'] = df_all['coordinates'][i]["latitude"]
    df_all.loc[i, 'longitude'] = df_all['coordinates'][i]["longitude"]
    df_all.loc[i, 'address_detail'] = df_all['location'][i]["address1"]
    df_all.loc[i, 'city'] = df_all['location'][i]["city"]
    df_all.loc[i, 'zip_code'] = df_all['location'][i]['zip_code']
    df_all.loc[i, 'state'] = df_all['location'][i]["state"]
    df_all.loc[i, 'country'] = df_all['location'][i]['country']
    df_all.loc[i, 'cate_alias'] = df_all['categories'][i][0]["alias"]
    df_all.loc[i, 'cate_title'] = df_all['categories'][i][0]['title']

```

```

[6]: df_all =
      df_all[['xuhao', 'name', 'rating', 'review_count', 'cate_alias', 'cate_title', 'price', 'distance']]
df_all

```

```

[6]:
      xuhao      name  rating  review_count \
0         1      Café Synapse      3.5         57
1         2  Court of Sciences Student Center      3.5         19
2         3      Bruin Buzz      3.5         20
3         4  Lollicup Fresh      3.5         24
4         5  Wetzels Pretzels      3.0         22
..      ...
345      346      Wings N Waffles      4.5        108
346      347      BrewWings      3.5        234
347      348  Cousins Maine Lobster - Los Angeles      3.5       1069
348      349      Good Eggs      4.5         44
349      350      Serendoggity      4.5         38

      cate_alias      cate_title  price      distance \
0         coffee      Coffee & Tea      $      188.804576
1         coffee      Coffee & Tea     $$      246.206888
2         coffee      Coffee & Tea      $      332.271880
3      bubbletea      Bubble Tea      $      340.234688
4         pretzels      Pretzels      $      340.234688
..      ...
345      foodtrucks      Food Trucks      $  18659.250794
346      chicken_wings      Chicken Wings     $$  18659.250794
347           seafood      Seafood     $$  18659.250794
348  fooddeliveryservices  Food Delivery Services   $$$  18758.357893
349           hotdog      Hot Dogs      $  18758.357893

```

		image_url	display_phone	\
0		https://s3-media1.fl.yelpcdn.com/bphoto/2-4Qox...	(310) 794-4485	
1		https://s3-media2.fl.yelpcdn.com/bphoto/0uggQm...	(310) 206-4773	
2		https://s3-media1.fl.yelpcdn.com/bphoto/R0S0qj...	(310) 206-0740	
3		https://s3-media1.fl.yelpcdn.com/bphoto/JGS3Ez...	(310) 794-2220	
4		https://s3-media1.fl.yelpcdn.com/bphoto/lirPuG...	(310) 206-0732	
..		...	...	
345		https://s3-media1.fl.yelpcdn.com/bphoto/GKVcd5...	(818) 738-0469	
346		https://s3-media3.fl.yelpcdn.com/bphoto/DFYM4V...	(562) 353-8999	
347		https://s3-media2.fl.yelpcdn.com/bphoto/X-RPyJ...	(207) 370-6112	
348		https://s3-media3.fl.yelpcdn.com/bphoto/sZ6zkc...	(415) 483-7344	
349		https://s3-media2.fl.yelpcdn.com/bphoto/eiP3Uv...	(323) 456-3034	

	is_closed		id	latitude	longitude	\
0	False	yirNnFJ8Mo8wN50y8hG6xQ	34.067151	-118.444673		
1	False	dwksGXxMa0qjiosc6uFFVg	34.067856	-118.443770		
2	False	x6M5SjgjDjxWBkxYvtLCA	34.070384	-118.444233		
3	False	U2zhu7bFJkS106NG0fajRw	34.068505	-118.444763		
4	False	H9n4r9U4Warbbns0Z0B5Qw	34.068505	-118.444763		
..	...	...	...	...		
345	False	bx98Nr15LUBntgc9kqP15w	34.058140	-118.246660		
346	False	A8u1P4p29eYwtVYbjQuRkw	34.058140	-118.246660		
347	False	1SeCdGBjwEE02MwJ7jPFeg	34.056820	-118.243870		
348	False	UH25pTDsWH8p73P6s--jIQ	34.052393	-118.243680		
349	False	e5VSkkJ9tIREnvENmo6uPA	34.052393	-118.243680		

	address_detail	city	zip_code	state	country
0	754 Charles E Young Dr S	Los Angeles	90024	CA	US
1	570 Westwood Plz	Los Angeles	90095	CA	US
2	308 Westwood Plz	Los Angeles	90095	CA	US
3	308 Westwood Plz	Los Angeles	90024	CA	US
4	308 Westwood Plz	Los Angeles	90024	CA	US
..	...	...	...	...	...
345		Los Angeles	90012	CA	US
346		Los Angeles	90012	CA	US
347		Los Angeles	90012	CA	US
348		Los Angeles		CA	US
349		Los Angeles		CA	US

[350 rows x 19 columns]

```
[7]: df_all_reviews = pd.DataFrame()
for j in range(len(df_all)):
    temp = df_all['id'][j]
    url = f"https://api.yelp.com/v3/businesses/{temp}/reviews?
    ↪limit=20&sort_by=yelp_sort"
    headers = {
```

```

"accept": "application/json",
"Authorization": f"Bearer {api_key}"
response = requests.get(url, headers=headers)
if df_all['review_count'][j] > 0 :
    df_j = pd.DataFrame(response.json()["reviews"])
    df_j['xuhao'] = j+1
    df_all_reviews = pd.concat([df_all_reviews,df_j],ignore_index = True)

```

```

[8]: df_all_reviews = df_all_reviews[['xuhao','rating','text']]
df_all_reviews

```

```

[8]:      xuhao  rating      text
0         1        5  2 words: PEET'S COFFEE.\n\nTucked away at the ...
1         1        4  Café Synapse - A small venue located across th...
2         1        5  Panini's and omelets never disappoint. Shout o...
3         2        3  No breakfast options and the seating is typica...
4         2        4  It's been too long since I first tried the Bom...
..      ...      ...      ...
966      349        4  Great local products, fast responsive customer...
967      349        5  Definitely showing my support for GoodEggs dur...
968      350        5  The vegan hot dogs were amazing! Im always loo...
969      350        5  Serendoggity is a fun food vendor for parties...
970      350        5  We hired Serendoggity for my sons birthday and...

```

[971 rows x 3 columns]

### 3 UCLA&USC coffee

```

[9]: url = "https://api.yelp.com/v3/businesses/search?
      ↪location=ucla&categories=coffee&sort_by=distance&matches_party_size_param=true&limit=50"

headers = {
    "accept": "application/json",
    "Authorization": "Bearer_
      ↪FCsFoBjlLGie6-HG7fTX0cIow-KG3Gb7fUDVHlHlizqnpB_K2gMRGkZ0lZ9ecMBjQuA3FxpIots3HDkFSGqv-3z9iGs
}

response = requests.get(url, headers=headers)
df1 = pd.DataFrame(response.json()["businesses"])
df1["school"] = 'ucla'

```

```

[10]: url = "https://api.yelp.com/v3/businesses/search?
      ↪location=usc&categories=coffee&sort_by=distance&matches_party_size_param=true&limit=50"

headers = {
    "accept": "application/json",

```

```

    "Authorization": "Bearer_
    ↪FCsFoBj1LGie6-HG7fTX0cIow-KG3Gb7fUDVH1HlizqnpB_K2gMRGkZ0lZ9ecMBjQuA3FxpIots3HDkFSGqv-3z9iGs
}

response2 = requests.get(url, headers=headers)
df2 = pd.DataFrame(response2.json()["businesses"])
df2["school"] = 'usc'

```

```

[11]: df_coffee = pd.concat([df1,df2],ignore_index = True)
df_coffee['xuhao'] = range(1,len(df_coffee)+1)

```

```

[12]: df_coffee['latitude'] = 0
df_coffee['longitude'] = 0
df_coffee['address_detail'] = 0
df_coffee['city'] = 0
df_coffee['zip_code'] = 0
df_coffee['state'] = 0
df_coffee['country'] = 0
df_coffee['cate_alias'] = 0
df_coffee['cate_title'] = 0

for i in range(len(df_coffee)):
    df_coffee.loc[i,'latitude'] = df_coffee['coordinates'][i]["latitude"]
    df_coffee.loc[i,'longitude'] = df_coffee['coordinates'][i]["longitude"]
    df_coffee.loc[i,'address_detail'] = df_coffee['location'][i]["address1"]
    df_coffee.loc[i,'city'] = df_coffee['location'][i]["city"]
    df_coffee.loc[i,'zip_code'] = df_coffee['location'][i]['zip_code']
    df_coffee.loc[i,'state'] = df_coffee['location'][i]["state"]
    df_coffee.loc[i,'country'] = df_coffee['location'][i]['country']
    df_coffee.loc[i,'cate_alias'] = df_coffee['categories'][i][0]["alias"]
    df_coffee.loc[i,'cate_title'] = df_coffee['categories'][i][0]['title']

```

```

[13]: df_coffee =
    ↪df_coffee[['xuhao','school','name','rating','review_count','cate_alias','cate_title','price
df_coffee

```

```

[13]:
   xuhao school name rating review_count \
0      1   ucla   Café Synapse    3.5         57
1      2   ucla  Court of Sciences Student Center    3.5         19
2      3   ucla   Bruin Buzz    3.5         20
3      4   ucla  Lollicup Fresh    3.5         24
4      5   ucla   Relaxtations    3.0          3
..    ...   ...
95    96   usc   Urban Espresso    4.5         43
96    97   usc   South LA Cafe    5.0        284
97    98   usc   7-Eleven    1.5          9
98    99   usc   Macchiato    4.5         87

```

```

99      100      usc                                Pico Donuts      4.5      50

```

```

      cate_alias      cate_title price      distance \
0      coffee      Coffee & Tea      $      188.804576
1      coffee      Coffee & Tea      $$      246.206888
2      coffee      Coffee & Tea      $      332.271880
3      bubbletea      Bubble Tea      $      340.234688
4      coffee      Coffee & Tea      $      345.279973
..      ...      ...      ...
95      foodtrucks      Food Trucks      $      2480.758191
96      coffee      Coffee & Tea      $      2488.152696
97      convenience      Convenience Stores      $      2596.474324
98      desserts      Desserts      $$      2624.326264
99      donuts      Donuts      $      2634.484852

```

```

                                image_url      display_phone \
0      https://s3-media1.fl.yelpcdn.com/bphoto/2-4Qox...      (310) 794-4485
1      https://s3-media2.fl.yelpcdn.com/bphoto/0uggQm...      (310) 206-4773
2      https://s3-media1.fl.yelpcdn.com/bphoto/ROS0qj...      (310) 206-0740
3      https://s3-media1.fl.yelpcdn.com/bphoto/JGS3Ez...      (310) 794-2220
4
..      ...
95      https://s3-media4.fl.yelpcdn.com/bphoto/FKlZ5K...      (562) 794-6484
96      https://s3-media2.fl.yelpcdn.com/bphoto/qaGrU1...      (213) 260-0633
97      https://s3-media3.fl.yelpcdn.com/bphoto/rBgBJj...      (213) 749-6711
98      https://s3-media2.fl.yelpcdn.com/bphoto/0g67lF...      (213) 536-5273
99      https://s3-media2.fl.yelpcdn.com/bphoto/w1XgSw...      (213) 315-5018

```

```

      is_closed      id      latitude      longitude \
0      False      yirNnFJ8Mo8wN50y8hG6xQ      34.067151      -118.444673
1      False      dwksGXxMa0qjiosc6uFFVg      34.067856      -118.443770
2      False      x6M5SjgjDjxWwBkxYvtLCA      34.070384      -118.444233
3      False      U2zhu7bFJkS106NG0fajRw      34.068505      -118.444763
4      False      Is-nXa9J7mIlgvFxydCv8A      34.068505      -118.444763
..      ...
95      False      qu8yFETeo4J8fMbf_7XV_Q      34.039830      -118.268550
96      False      yNNtnin-2Cv-hFt077oODQ      34.011879      -118.308937
97      False      dm2QQK9I3-ibsJaYUE900Q      34.027151      -118.257453
98      False      wChdSuRVNKruqtN7xUXTA      34.039065      -118.265058
99      False      3JyaTB1Qe2d6e0Z3USRqEA      34.044681      -118.276707

```

```

                                address_detail      city zip_code state country
0      754 Charles E Young Dr S      Los Angeles      90024      CA      US
1      570 Westwood Plz      Los Angeles      90095      CA      US
2      308 Westwood Plz      Los Angeles      90095      CA      US
3      308 Westwood Plz      Los Angeles      90024      CA      US
4      308 Westwood Plaza Ackerman 1st      Los Angeles      90024      CA      US

```

..	...	...	...	...	...
95		Los Angeles	90015	CA	US
96	1700 Browning Blvd	Los Angeles	90062	CA	US
97	1916 S San Pedro St	Los Angeles	90011	CA	US
98	418 W Pico Blvd	Los Angeles	90015	CA	US
99	1565 W Pico Blvd	Los Angeles	90015	CA	US

[100 rows x 20 columns]

```
[14]: df_coffee_reviews = pd.DataFrame()
for j in range(len(df_coffee)):
    temp = df_coffee['id'][j]
    url = f"https://api.yelp.com/v3/businesses/{temp}/reviews?
    ↪limit=20&sort_by=yelp_sort"
    headers = {
        "accept": "application/json",
        "Authorization": f"Bearer {api_key}"
    }
    response = requests.get(url, headers=headers)
    if df_coffee['review_count'][j] > 0 :
        df_j = pd.DataFrame(response.json()["reviews"])
        df_j['xuhao'] = j+1
        df_j['school'] = df_coffee['school'][j]
        df_coffee_reviews = pd.concat([df_coffee_reviews,df_j],ignore_index =
    ↪True)
```

```
[37]: df_coffee_reviews = df_coffee_reviews[['xuhao','school','rating','text']]
df_coffee_reviews
```

```
[37]:
```

	xuhao	school	rating	text
0	1	ucla	5	2 words: PEET'S COFFEE.\n\nTucked away at the ...
1	1	ucla	4	Café Synapse - A small venue located across th...
2	1	ucla	5	Panini's and omelets never disappoint. Shout o...
3	2	ucla	3	No breakfast options and the seating is typica...
4	2	ucla	4	It's been too long since I first tried the Bom...
..	...	...	...	...
288	99	usc	4	Short walk from La Convention Center~ 3 blocks...
289	99	usc	4	Bright, airy cafe with potted plant drinks, wa...
290	100	usc	5	Recently stopped by this donut shop to buy a f...
291	100	usc	4	This is a good little donut shop. Tasty doubl...
292	100	usc	5	Dear pico donuts,\n\nThank you for serving fre...

[293 rows x 4 columns]

## 4 Upload Data to Cloud DB

```
[16]: import pymysql
```

```
[17]: endpoint = "pachi.cmpnrvfbwnkr.us-west-1.rds.amazonaws.com"
      username = 'pachi'
      password = '55667788aabb'
```

```
[18]: connection = pymysql.connections.Connection(host=endpoint,
                                                  user=username,
                                                  password=password)
```

```
[19]: def create_database(dbname):
      connection.cursor().execute(f"CREATE DATABASE IF NOT EXISTS {dbname}")
      connection.commit()
      print("Successful")

      def create_table(dbname,tablename,formatted_sql_columns):
          connection.cursor().execute(f"USE {dbname}")
          connection.cursor().execute(f"CREATE TABLE IF NOT EXISTS {tablename}
          ↳({formatted_sql_columns})")
          connection.commit()
          print("Successful")

      def delete_table(dbname,tablename):
          connection.cursor().execute(f"USE {dbname}")
          connection.cursor().execute(f"DROP TABLE IF EXISTS {tablename}")
          connection.commit()
          print("Successful")
```

```
[117]: cols1 = list(df_all.columns)
      cols2 = list(df_all_reviews.columns)
      cols3 = list(df_coffee.columns)
      cols4 = list(df_coffee_reviews.columns)
```

```
[118]: cols1_m = []
      for i in cols1:
          modified = "`"+i+"`"
          cols1_m.append(modified)

      cols2_m = []
      for i in cols2:
          modified = "`"+i+"`"
          cols2_m.append(modified)

      cols3_m = []
      for i in cols3:
```



```

        modified = "`"+i+"`"
        cols3_m.append(modified)

cols4_m = []
for i in cols4:
    modified = "`"+i+"`"
    cols4_m.append(modified)

```

```

[119]: #formatted_sql_columns = "xuhao int,name text,rating float,review_count_
        ↪int,cate_alias text,cate_title text,price text,distance float,image_url_
        ↪text,display_phone text,is_closed text,id text,latitude float,longitude_
        ↪float,address_detail text,city text,zip_code int,state text,country text"
#formatted_sql_columns2 = "xuhao int,school text,rating float,text text"
#formatted_sql_columns3 = "xuhao int,rating float,text text"
formatted_sql_columns = " TEXT",".join(cols1_m) + " TEXT"
formatted_sql_columns2 = " TEXT",".join(cols2_m) + " TEXT"
formatted_sql_columns3 = " TEXT",".join(cols3_m) + " TEXT"
formatted_sql_columns4 = " TEXT",".join(cols4_m) + " TEXT"

```

```

[120]: dbname = "yelp"
        tablename1 = "food"
        tablename2 = "food_review"
        tablename3 = "coffee"
        tablename4 = "coffee_review"

```

```

[22]: create_database(dbname)

```

Successful

```

[121]: create_table(dbname,tablename1,formatted_sql_columns)

```

Successful

```

[122]: create_table(dbname,tablename2,formatted_sql_columns2)

```

Successful

```

[123]: create_table(dbname,tablename3,formatted_sql_columns3)

```

Successful

```

[124]: create_table(dbname,tablename4,formatted_sql_columns4)

```

Successful

```

[125]: def add_row(table_name, dbname, columns_sql, row_data):
        esses = ("%s, " * len(columns_sql))[:-2]
        fixed_columns = ",".join(columns_sql)

```



```
for i,row in df_all_reviews.iterrows():
    row = [str(i) for i in list(row)]
    add_row(tablename2,dbname,cols2_m,row)
```

[illegible]

```
[128]: for i,row in df_coffee.iterrows():
        row = [str(i) for i in list(row)]
        add_row(tablename3,dbname,cols3_m,row)
```

```
row inserted
row inserted
row inserted
row inserted
```

```
[129]: for i,row in df_coffee_reviews.iterrows():
        row = [str(i) for i in list(row)]
        add_row(tablename4,dbname,cols4_m,row)
```

[illegible]

coffee
xuhao INT
school TEXT
name TEXT
rating FLOAT
review_count INT
cate_alias TEXT
cate_title TEXT
price TEXT
distance FLOAT
image_url TEXT
display_phone TEXT
is_closed TEXT
id TEXT
latitude FLOAT
longitude FLOAT
address_detail TEXT
zip_code INT
state TEXT
country TEXT

coffee_review
xuhao INT
school TEXT
rating_ind FLOAT
text TEXT

coffee_info
xuhao INT
school TEXT
name TEXT
rating FLOAT
review_count INT
cate_alias TEXT
cate_title TEXT
price TEXT
distance FLOAT
image_url TEXT
display_phone TEXT
is_closed TEXT
id TEXT
latitude FLOAT
longitude FLOAT
address_detail TEXT
zip_code INT
Indexes

comment_info
comment_id BIGINT
xuhao INT
rating_ind FLOAT
text TEXT
Indexes

geo_info
zip_code INT
state TEXT
country TEXT
Indexes

food
xuhao INT
name TEXT
rating FLOAT
review_count INT
cate_alias TEXT
cate_title TEXT
price TEXT
distance FLOAT
image_url TEXT
display_phone TEXT
is_closed TEXT
id TEXT
latitude FLOAT
longitude FLOAT
address_detail TEXT
zip_code INT
state TEXT
country TEXT

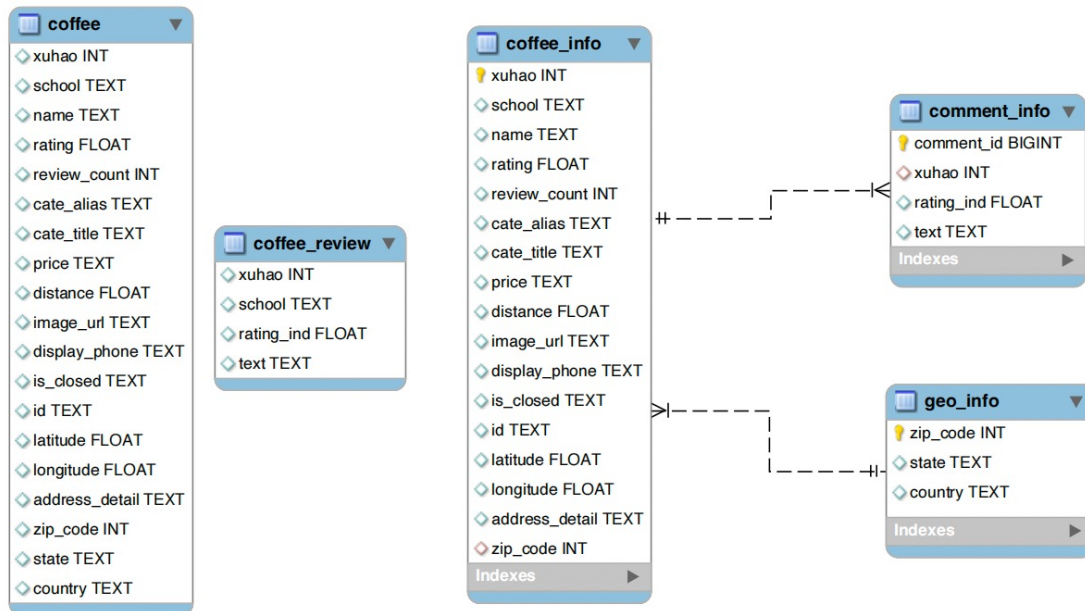
food_review
xuhao INT
rating_ind FLOAT
text TEXT

food_info
xuhao INT
name TEXT
rating FLOAT
review_count INT
cate_alias TEXT
cate_title TEXT
price TEXT
distance FLOAT
image_url TEXT
display_phone TEXT
is_closed TEXT
id TEXT
latitude FLOAT
longitude FLOAT
address_detail TEXT
zip_code INT
Indexes

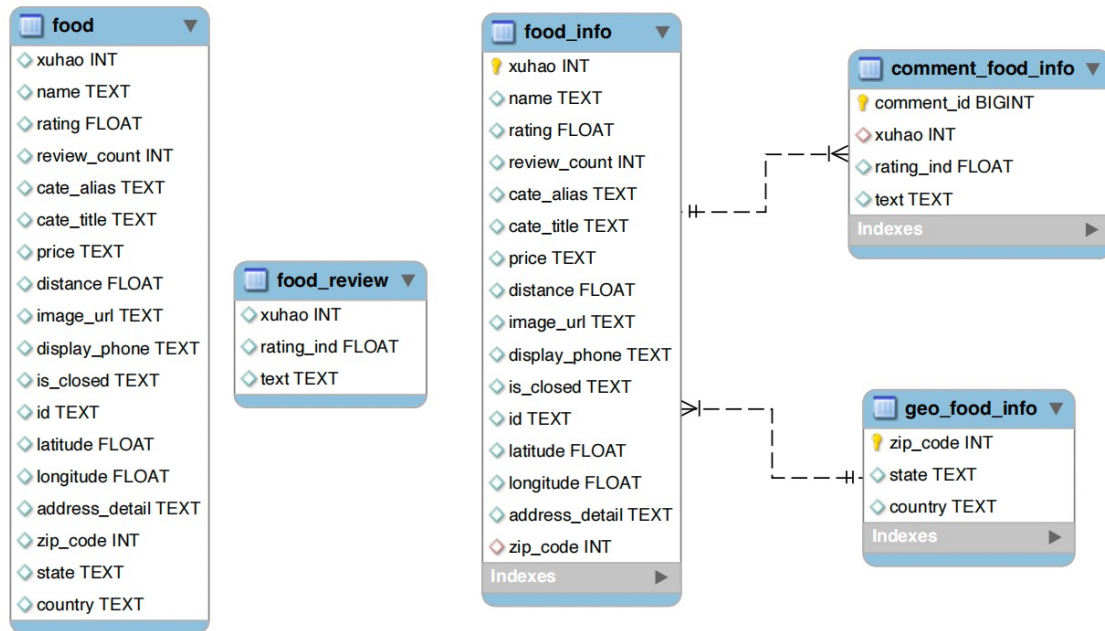
comment_food_info
comment_id BIGINT
xuhao INT
rating_ind FLOAT
text TEXT
Indexes

geo_food_info
zip_code INT
state TEXT
country TEXT
Indexes

# Coffee relational table



# Food relational table





## Food

original:

food: 350 rows \* 18 columns

food\_review: 971rows \* 3 columns

total: 9213

after change:

food\_info: 350 rows \* 16 columns

geo\_food\_info: 80 rows \* 3columns

comment\_food\_info: 293rows \* 4columns

total: 9724

## Coffee

original:

food: 100 rows \* 19 columns

food\_review: 293 rows \* 4 columns

total: 3072

after change:

food\_info: 100 rows \* 17 columns

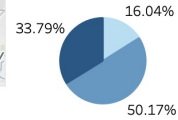
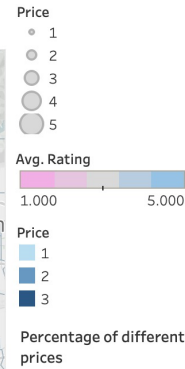
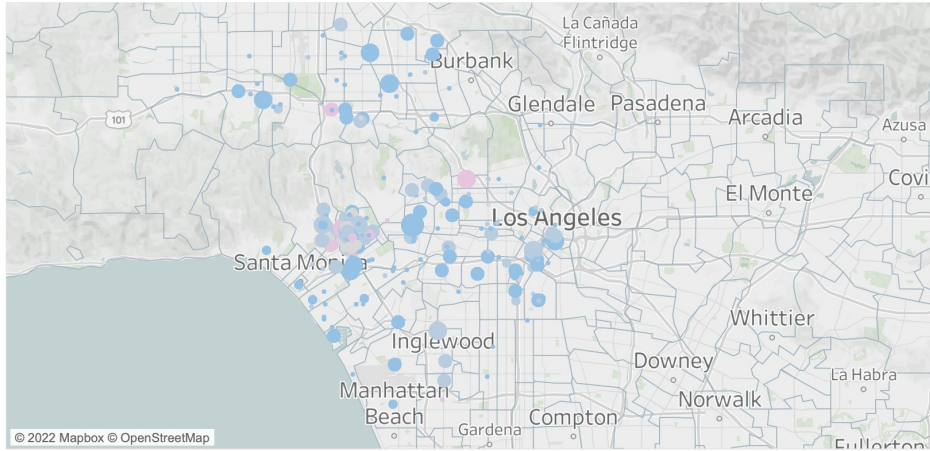
geo\_info: 80 rows \* 3columns

comment\_info: 293rows \* 4columns

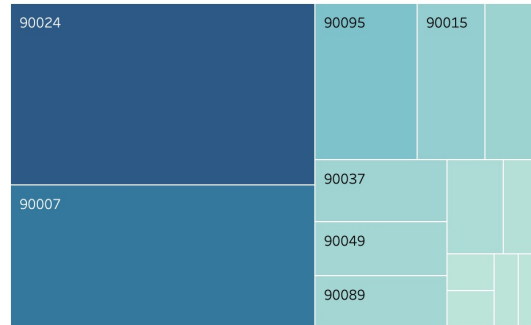
total: 2914

reduced the storage by 5.14%

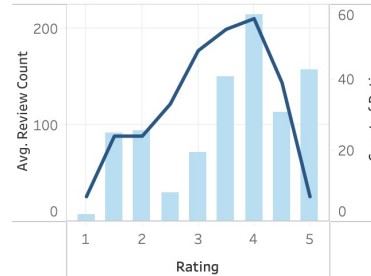
## Price and rating of restaurants around UCLA



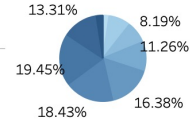
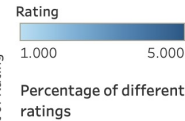
## Restaurant clustering region



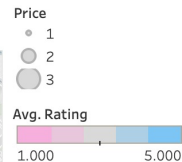
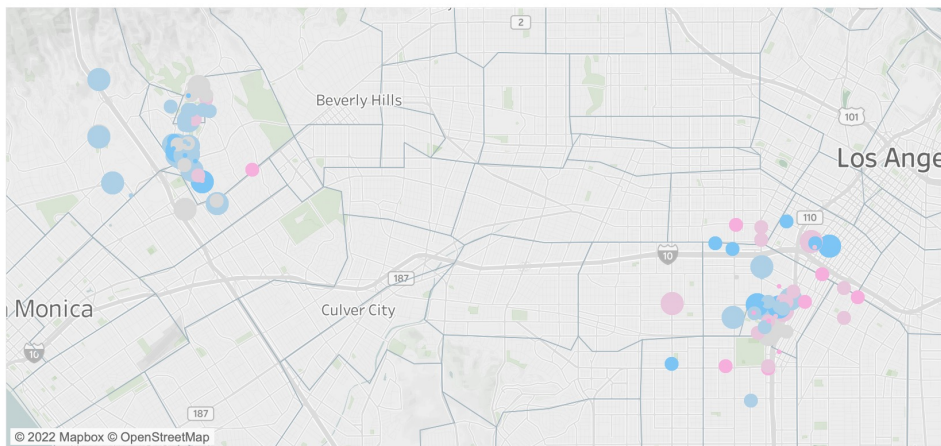
## Relationship between ratings and the average number of reviews



Measure N.. Avg. Review Co.. Count of Rating



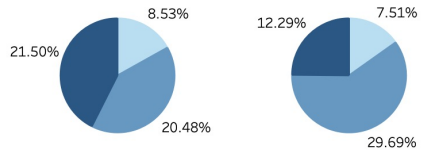
# Price and rating of coffee shops around UCLA & USC



Percentage of different prices

School

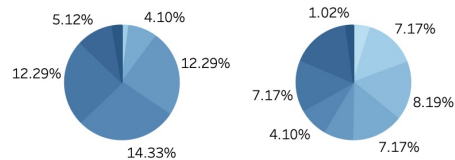
ucla usc



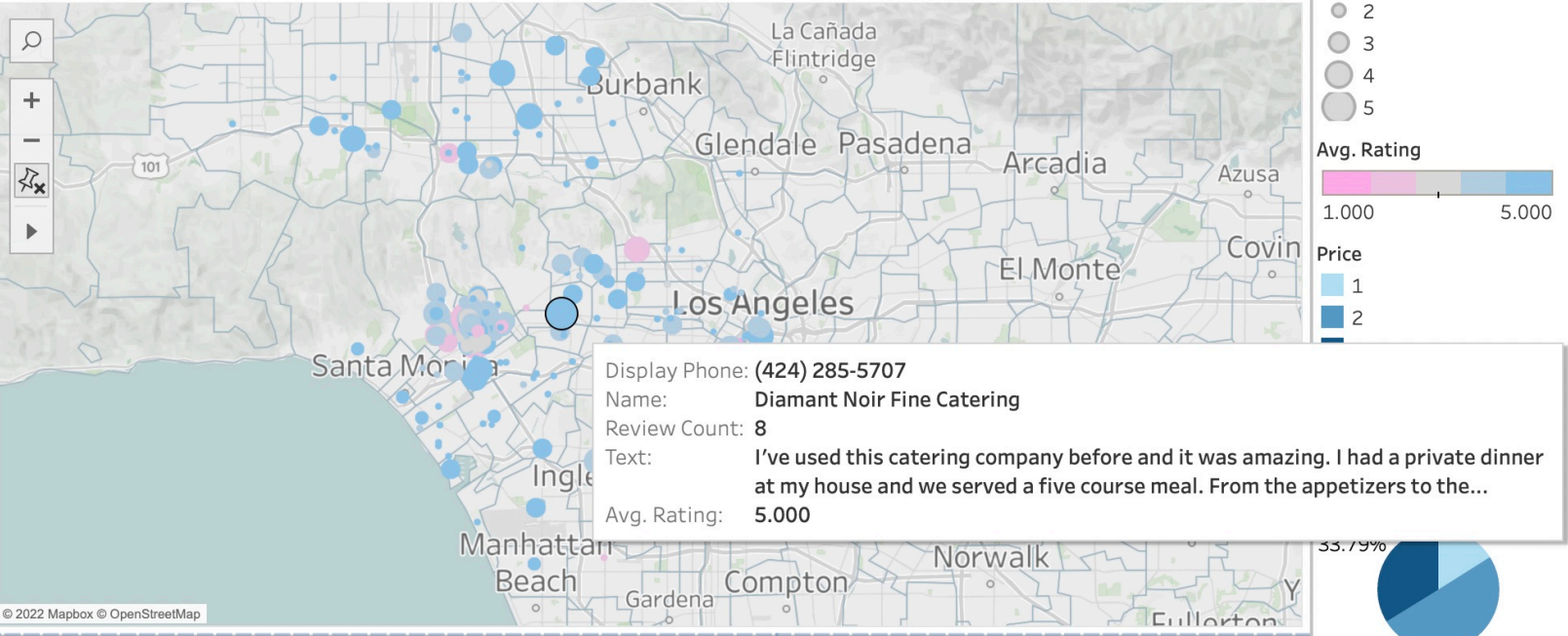
Percentage of different ratings

School

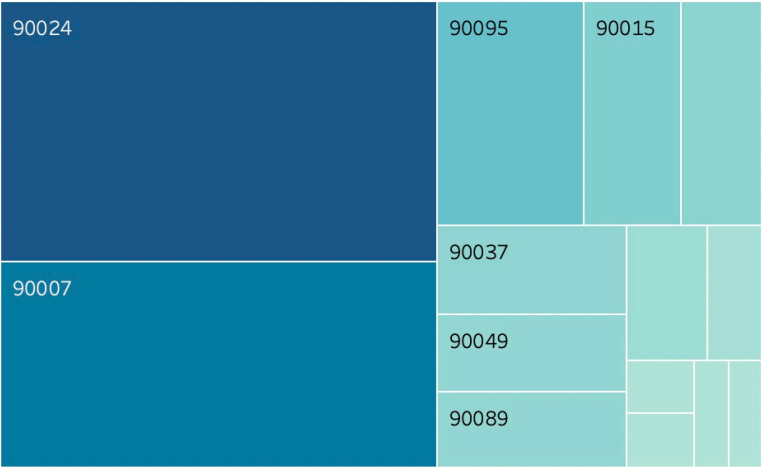
ucla usc



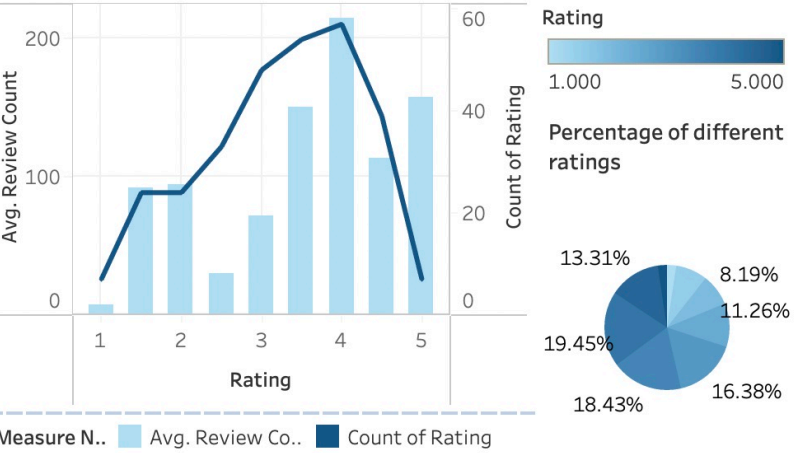
# Price and rating of restaurants around UCLA



## Restaurant clustering region



## Relationship between ratings and the average number of reviews



# Price and rating of coffee shops around UCLA & USC

