```
In [2]:  import os
         import pandas as pd
         pd.options.display.float_format = '{:20,.2f}'.format
         pd.set_option('display.max_rows', 5000)
         pd.set_option('display.max_columns', 5000)
         pd.set_option('display.width', 1000)
         pd.set_option('display.max_colwidth', -1)
```

```
In [9]:  fodors_zagats_path = (
             r'/home/ubuntu/jupyter/ServerX/1_Standard Data Integration/Sample D
             r'/Unprocessed Data/customer_samples/fodors_zagats_restaurants'
         )
```

Fodors data

```
In [16]:  fodor_fields = ['id', 'name', 'addr', 'city', 'phone', 'type']
```

```
In [32]:  fodors_data = pd.read_csv(
              os.path.join(fodors_zagats_path, 'fodors.csv'),
              sep = ',',
              quotechar = "'",
              usecols  = fodor_fields,
          )[fodor_fields]
```

```
In [33]:  fodors_data.rename(
              columns =
              {
                  'id': 'record_id',
                  'addr': 'address',
                  'type': 'category'
              },
              inplace = True
          )
```

```
In [34]:  fodors_data.head()
```

Out[34]:

	record_id	name	address	city	phone	category
0	534	arnie morton\s of chicago'	435 s. la cienega blv.	los angeles	310/246-1501	american
1	535	art\s delicatessen'	12224 ventura blvd.	studio city	818/762-1221	american
2	536	hotel bel-air	701 stone canyon rd.	bel air	310/472-1211	californian
3	537	cafe bizou	14016 ventura blvd.	sherman oaks	818/788-3536	french
4	538	campanile	624 s. la brea ave.	los angeles	213/938-1447	american

Zagats data

```
In [28]:  zagats_fields = ['id', 'name', 'addr', 'city', 'phone', 'type']
```

```
In [35]:  zagats_data = pd.read_csv(
              os.path.join(fodors_zagats_path, 'zagats.csv'),
              sep = ',',
              quotechar = "'",
              usecols  = zagats_fields,
          )[zagats_fields]
```

```
In [36]:  zagats_data.rename(
              columns =
              {
                  'id': 'record_id',
                  'addr': 'address',
                  'type': 'category'
              },
              inplace = True
          )
```

```
In [37]:  zagats_data.head()
```

Out[37]:

	record_id	name	address	city	phone	category
0	1	apple pan the	10801 w. pico blvd.	west la	310-475-3585	american
1	2	asahi ramen	2027 sawtelle blvd.	west la	310-479-2231	noodle shops
2	3	baja fresh	3345 kimber dr.	westlake village	805-498-4049	mexican
3	4	belvedere the	9882 little santa monica blvd.	beverly hills	310-788-2306	pacific new wave
4	5	benita\s frites'	1433 third st. promenade	santa monica	310-458-2889	fast food

Matches Fodors Zagats

```
In [38]:  matches_fod_zag_data = pd.read_csv(
              os.path.join(fodors_zagats_path, 'matches_fodors_zagats.csv'),
              sep = ',',
              quotechar = "'",
              #usecols  = zagats_fields,
          )#[zagats_fields]
```

```
In [39]:  matches_fod_zag_data.columns
```

Out[39]: Index(['fodors_id', 'zagats_id'], dtype='object')

```
In [63]: matches_fod_zag_data.head(20)
```

Out[63]:

	fodors_id	zagats_id
0	534	219
1	535	220
2	536	221
3	537	222
4	538	223
5	539	224
6	540	225
7	541	226
8	542	227
9	543	228
10	544	229

```
In [68]: zagats_data[zagats_data['record_id'] == 236]
```

Out[68]:

	record_id	name	address	city	phone	category	source
235	236	philippe the original	1001 n. alameda st.	chinatown	213-628-3781	cafeterias	zagats

```
In [ ]:
```

```
In [65]: fodors_data[fodors_data['record_id'] == 551]
```

Out[65]:

	record_id	name	address	city	phone	category	source
17	551	philippe\s the original'	1001 n. alameda st.	los angeles	213/628-3781	american	fodors

```
In [41]: customer_all_path = (r'/home/ubuntu/jupyter/ServerX/1_Standard Data Int(
                             r'/Processed Data/customer_samples/')
```

```
In [42]: fodors_data['source'] = 'fodors'
         zagats_data['source'] = 'zagats'
```

```
In [43]: fodors_zagats_all = pd.concat([fodors_data, zagats_data])
```

```
In [44]: fodors_zagats_all.rename(columns={'record_id': 'Id'}, inplace=True)
```

```
In [46]: fodors_zagats_all.to_csv(customer_all_path + 'fodors_zagats_all.csv', s(
```

```
In [49]: customer_all_path
```

Out[49]: '/home/ubuntu/jupyter/Hadoco/1_Standard Data Integration/Sample Datase
ts/Processed Data/customer_samples/'

```
In [48]: fodors_zagats_all.head()
```

Out[48]:

	Id	name	address	city	phone	category	source
0	534	arnie morton\s of chicago'	435 s. la cienega blv.	los angeles	310/246-1501	american	fodors
1	535	art\s delicatessen'	12224 ventura blvd.	studio city	818/762-1221	american	fodors
2	536	hotel bel-air	701 stone canyon rd.	bel air	310/472-1211	californian	fodors
3	537	cafe bizou	14016 ventura blvd.	sherman oaks	818/788-3536	french	fodors
4	538	campanile	624 s. la brea ave.	los angeles	213/938-1447	american	fodors