# Chipotle快餐数据 - 了解数据



## 1. 从文件中加载数据，并查看数据前3行

```python
# 导入pandas库并命名为pd
import pandas as pd
```

```python
!ls data
```

```
chipotle.tsv
```

```python
# 加载数据
chipo = pd.read_csv('data/chipotle.tsv', sep='\t')
```

```
chipo.head(3)
```

```
<tr style="text-align: right;">
  <th></th>
  <th>order_id</th>
  <th>quantity</th>
  <th>item_name</th>
  <th>choice_description</th>
  <th>item_price</th>
</tr>
```

```
<tr>
  <th>0</th>
  <td>1</td>
  <td>1</td>
  <td>Chips and Fresh Tomato Salsa</td>
  <td>NaN</td>
  <td>$2.39</td>
</tr>
<tr>
  <th>1</th>
  <td>1</td>
  <td>1</td>
  <td>Izze</td>
  <td>[Clementine]</td>
  <td>$3.39</td>
</tr>
<tr>
  <th>2</th>
  <td>1</td>
  <td>1</td>
  <td>Nantucket Nectar</td>
  <td>[Apple]</td>
  <td>$3.39</td>
</tr>
```

```python
# 查看各列的数据类型
chipo.dtypes
```

```
1  order_id                int64
2  quantity                int64
3  item_name              object
4  choice_description     object
5  item_price             object
6  dtype: object
```

```
1  # 将item_price列转换为数字类型
2  chipo.item_price = chipo.item_price.apply(lambda x:
   float(x[1:]))
```

```
1  chipo.dtypes
```

```
1  order_id                int64
2  quantity                int64
3  item_name              object
4  choice_description     object
5  item_price            float64
6  dtype: object
```

```
1  chipo.head(10)
```

```html
<tr style="text-align: right;">
    <th></th>
    <th>order_id</th>
    <th>quantity</th>
    <th>item_name</th>
    <th>choice_description</th>
    <th>item_price</th>
</tr>
```

```html
<tr>
  <th>0</th>
  <td>1</td>
  <td>1</td>
  <td>Chips and Fresh Tomato Salsa</td>
  <td>NaN</td>
  <td>2.39</td>
</tr>
<tr>
  <th>1</th>
  <td>1</td>
  <td>1</td>
  <td>Izze</td>
  <td>[Clementine]</td>
  <td>3.39</td>
</tr>
<tr>
  <th>2</th>
  <td>1</td>
  <td>1</td>
  <td>Nantucket Nectar</td>
  <td>[Apple]</td>
  <td>3.39</td>
</tr>
<tr>
  <th>3</th>
  <td>1</td>
  <td>1</td>
  <td>Chips and Tomatillo-Green Chili Salsa</td>
  <td>NaN</td>
  <td>2.39</td>
</tr>
<tr>
  <th>4</th>
```

```
35      <td>2</td>
36      <td>2</td>
37      <td>Chicken Bowl</td>
38      <td>[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
     </td>
39      <td>16.98</td>
40   </tr>
41   <tr>
42      <th>5</th>
43      <td>3</td>
44      <td>1</td>
45      <td>Chicken Bowl</td>
46      <td>[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...
     </td>
47      <td>10.98</td>
48   </tr>
49   <tr>
50      <th>6</th>
51      <td>3</td>
52      <td>1</td>
53      <td>Side of Chips</td>
54      <td>NaN</td>
55      <td>1.69</td>
56   </tr>
57   <tr>
58      <th>7</th>
59      <td>4</td>
60      <td>1</td>
61      <td>Steak Burrito</td>
62      <td>[Tomatillo Red Chili Salsa, [Fajita Vegetables...
     </td>
63      <td>11.75</td>
64   </tr>

65   <tr>
```

```
66    <th>8</th>
67    <td>4</td>
68    <td>1</td>
69    <td>Steak Soft Tacos</td>
70    <td>[Tomatillo Green Chili Salsa, [Pinto Beans, Ch...
   </td>
71    <td>9.25</td>
72  </tr>
73  <tr>
74    <th>9</th>
75    <td>5</td>
76    <td>1</td>
77    <td>Steak Burrito</td>
78    <td>[Fresh Tomato Salsa, [Rice, Black Beans, Pinto...
   </td>
79    <td>9.25</td>
80  </tr>
```

## 2. 查看数据集的基本信息

```
1  chipo.info()
```

```
 1  <class 'pandas.core.frame.DataFrame'>
 2  RangeIndex: 4622 entries, 0 to 4621
 3  Data columns (total 5 columns):
 4  order_id               4622 non-null int64
 5  quantity               4622 non-null int64
 6  item_name              4622 non-null object
 7  choice_description     3376 non-null object
 8  item_price             4622 non-null float64
 9  dtypes: float64(1), int64(2), object(2)
10  memory usage: 180.6+ KB
```

## 3. 查看数据的行列数和列名

```
 1  print ('(row, column) = ', chipo.shape)
```

```
 1  (row, column) =   (4622, 5)
```

```
 1  # 列名
 2  print('columns = ', chipo.columns)
```

```
 1  columns =  Index(['order_id', 'quantity', 'item_name', 'choice_description',
 2        'item_price'],
 3      dtype='object')
```

```
 1  # 查看索引信息
 2  chipo.index
```

```
 1  RangeIndex(start=0, stop=4622, step=1)
```

# 4. 查询商品种类的数量，以及下单最多的几种商品

```python
1  print ('Total items: ', len(chipo.item_name.unique()))
2  print ('------------------------------------------------------
   ----------------------------')
3  print ('Total items: ',
   chipo.item_name.value_counts().count())
4  print ('------------------------------------------------------
   ----------------------------')
5  print ('Total items = ',
   len(chipo.groupby(['item_name']).groups))
```

```
1  Total items:  50
2  ----------------------------------------------------------
   --------------------
3  Total items:  50
4  ----------------------------------------------------------
   --------------------
5  Total items =   50
```

```python
1  # 获取销量最多的商品名
2  chipo.item_name.value_counts().head().index
```

```
1  Index(['Chicken Bowl', 'Chicken Burrito', 'Chips and
   Guacamole',
2         'Steak Burrito', 'Canned Soft Drink'],
3        dtype='object')
```

```python
1  # 获取销量最多的商品销售量
2  chipo.item_name.value_counts().head()
```

```
1  Chicken Bowl           726
2  Chicken Burrito        553
3  Chips and Guacamole    479
4  Steak Burrito          368
5  Canned Soft Drink      301
6  Name: item_name, dtype: int64
```

## 5. 查询最受欢迎的食物及其销量，销售额

```python
1  # 查询最受欢迎的食物
2  favoriteItems =
   chipo.groupby(['item_name']).sum().sort_values('quantity',
   ascending=False)
3  # 删除order_id列
4  favoriteItems=favoriteItems.drop('order_id', axis=1)
5  favoriteItems.head()
```

```html
1  <tr style="text-align: right;">
2    <th></th>
3    <th>order_id</th>
4    <th>quantity</th>
5    <th>item_price</th>
6  </tr>
7  <tr>
8    <th>item_name</th>
9    <th></th>
10   <th></th>
11   <th></th>
12  </tr>
```

```html
<tr>
  <th>Chicken Bowl</th>
  <td>713926</td>
  <td>761</td>
  <td>7342.73</td>
</tr>
<tr>
  <th>Chicken Burrito</th>
  <td>497303</td>
  <td>591</td>
  <td>5575.82</td>
</tr>
<tr>
  <th>Chips and Guacamole</th>
  <td>449959</td>
  <td>506</td>
  <td>2201.04</td>
</tr>
<tr>
  <th>Steak Burrito</th>
  <td>328437</td>
  <td>386</td>
  <td>3851.43</td>
</tr>
<tr>
  <th>Canned Soft Drink</th>
  <td>304753</td>
  <td>351</td>
  <td>438.75</td>
</tr>
```

# 6. 查询最受欢迎食物（下单数，销量，销售额）

```python
favoriteItems = chipo.groupby(['item_name']).agg({'order_id' : 'count', 'quantity' : 'sum', 'item_price' : 'sum'}).sort_values('quantity', ascending=False)
favoriteItems.head()
```

```html
<tr style="text-align: right;">
  <th></th>
  <th>order_id</th>
  <th>quantity</th>
  <th>item_price</th>
</tr>
<tr>
  <th>item_name</th>
  <th></th>
  <th></th>
  <th></th>
</tr>
```

```
<tr>
  <th>Chicken Bowl</th>
  <td>726</td>
  <td>761</td>
  <td>7342.73</td>
</tr>
<tr>
  <th>Chicken Burrito</th>
  <td>553</td>
  <td>591</td>
  <td>5575.82</td>
</tr>
<tr>
  <th>Chips and Guacamole</th>
  <td>479</td>
  <td>506</td>
  <td>2201.04</td>
</tr>
<tr>
  <th>Steak Burrito</th>
  <td>368</td>
  <td>386</td>
  <td>3851.43</td>
</tr>
<tr>
  <th>Canned Soft Drink</th>
  <td>301</td>
  <td>351</td>
  <td>438.75</td>
</tr>
```

## 7. 查询订单的商品总数量

```
1  chipo.quantity.sum()
```

```
1  4972
```

## 8. 统计总收入

```
1  chipo.item_price.sum()
```

```
1  34500.160000000003
```

## 9. 统计订单总数

```
1  chipo.order_id.value_counts().count()
```

```
1  1834
```

## 10. 订单平均消费

```
1  chipo.item_price.sum() /
   chipo.order_id.value_counts().count()
2  # chipo.groupby(by=['order_id']).sum()['item_price'].mean()
```

```
1  18.811428571428575
```

## 11. 按商品名称过滤查询数据

```
1  chipo[(chipo['item_name'] == 'Chicken Bowl') &
   (chipo['quantity'] > 1)].head()
```

```html
<tr style="text-align: right;">
  <th></th>
  <th>order_id</th>
  <th>quantity</th>
  <th>item_name</th>
  <th>choice_description</th>
  <th>item_price</th>
</tr>
```

```html
<tr>
  <th>4</th>
  <td>2</td>
  <td>2</td>
  <td>Chicken Bowl</td>
  <td>[Tomatillo-Red Chili Salsa (Hot), [Black Beans...
</td>
  <td>16.98</td>
</tr>
<tr>
  <th>154</th>
  <td>70</td>
  <td>2</td>
  <td>Chicken Bowl</td>
  <td>[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...
</td>
  <td>17.50</td>
</tr>
<tr>
  <th>282</th>
  <td>124</td>
  <td>2</td>
  <td>Chicken Bowl</td>
  <td>[Fresh Tomato Salsa, [Rice, Black Beans, Chees...
</td>
  <td>17.50</td>
</tr>
<tr>
  <th>409</th>
  <td>178</td>
  <td>3</td>
  <td>Chicken Bowl</td>

  <td>[[Fresh Tomato Salsa (Mild), Tomatillo-Green C...
```

```
      </td>
31      <td>32.94</td>
32    </tr>
33    <tr>
34      <th>415</th>
35      <td>181</td>
36      <td>2</td>
37      <td>Chicken Bowl</td>
38      <td>[Tomatillo Red Chili Salsa]</td>
39      <td>17.50</td>
40    </tr>
```

## 12. 查询消费金额最大的订单信息 (3个)

```
1  maxOrder =
   chipo.groupby(['order_id']).sum().sort_values('item_price',
   ascending=False)
2  maxOrder.head(3)
```

```
<tr style="text-align: right;">
  <th></th>
  <th>quantity</th>
  <th>item_price</th>
</tr>
<tr>
  <th>order_id</th>
  <th></th>
  <th></th>
</tr>
```

```
<tr>
  <th>926</th>
  <td>23</td>
  <td>205.25</td>
</tr>
<tr>
  <th>1443</th>
  <td>35</td>
  <td>160.74</td>
</tr>
<tr>
  <th>1483</th>
  <td>14</td>
  <td>139.00</td>
</tr>
```

```python
arrOrderId = maxOrder.head(3).index[0:3].values
arrOrderId
```

```
array([ 926, 1443, 1483], dtype=int64)
```

```
chipo[chipo['order_id'].isin(arrOrderId)]
```

```html
<tr style="text-align: right;">
  <th></th>
  <th>order_id</th>
  <th>quantity</th>
  <th>item_name</th>
  <th>choice_description</th>
  <th>item_price</th>
</tr>
```

```html
<tr>
   <th>2304</th>
   <td>926</td>
   <td>1</td>
   <td>Steak Burrito</td>
   <td>[Fresh Tomato Salsa, [Rice, Sour Cream, Lettuce]]
</td>
   <td>9.25</td>
</tr>
<tr>
   <th>2305</th>
   <td>926</td>
   <td>1</td>
   <td>Chicken Bowl</td>
   <td>[Roasted Chili Corn Salsa, [Fajita Vegetables,...
</td>
   <td>8.75</td>
</tr>
<tr>
   <th>2306</th>
   <td>926</td>
   <td>1</td>
   <td>Chicken Bowl</td>
   <td>[Roasted Chili Corn Salsa, [Fajita Vegetables,...
</td>
   <td>8.75</td>
</tr>
<tr>
   <th>2307</th>
   <td>926</td>
   <td>1</td>
   <td>Chicken Bowl</td>

   <td>[Roasted Chili Corn Salsa, [Fajita Vegetables,...
```

```
    </td>
31    <td>8.75</td>
32  </tr>
33  <tr>
34    <th>2308</th>
35    <td>926</td>
36    <td>1</td>
37    <td>Steak Bowl</td>
38    <td>[Fresh Tomato Salsa, [Rice, Black Beans, Lettu...
    </td>
39    <td>9.25</td>
40  </tr>
41  <tr>
42    <th>2309</th>
43    <td>926</td>
44    <td>1</td>
45    <td>Chicken Bowl</td>
46    <td>[Fresh Tomato Salsa, [Rice, Black Beans, Chees...
    </td>
47    <td>8.75</td>
48  </tr>
49  <tr>
50    <th>2310</th>
51    <td>926</td>
52    <td>1</td>
53    <td>Steak Burrito</td>
54    <td>[Roasted Chili Corn Salsa, [Rice, Cheese, Sour...
    </td>
55    <td>9.25</td>
56  </tr>
57  <tr>
58    <th>2311</th>
59    <td>926</td>

60    <td>1</td>
```

```
    <td>Chicken Burrito</td>
    <td>[Fresh Tomato Salsa, [Rice, Black Beans, Chees...
</td>
    <td>8.75</td>
</tr>
<tr>
    <th>2312</th>
    <td>926</td>
    <td>1</td>
    <td>Chicken Bowl</td>
    <td>[Fresh Tomato Salsa, [Rice, Lettuce]]</td>
    <td>8.75</td>
</tr>
<tr>
    <th>2313</th>
    <td>926</td>
    <td>1</td>
    <td>Chicken Bowl</td>
    <td>[Fresh Tomato Salsa, [Rice, Cheese, Sour Cream...
</td>
    <td>8.75</td>
</tr>
<tr>
    <th>2314</th>
    <td>926</td>
    <td>1</td>
    <td>Chicken Salad Bowl</td>
    <td>[Roasted Chili Corn Salsa, [Rice, Sour Cream]]</td>
    <td>8.75</td>
</tr>
<tr>
    <th>2315</th>
    <td>926</td>

    <td>1</td>
```

```
93    <td>Steak Bowl</td>
94    <td>[Fresh Tomato Salsa, [Rice, Black Beans, Chees...
</td>
95    <td>9.25</td>
96  </tr>
97  <tr>
98    <th>2316</th>
99    <td>926</td>
100   <td>1</td>
101   <td>Chicken Burrito</td>
102   <td>[Roasted Chili Corn Salsa, [Rice, Black Beans,...
</td>
103   <td>8.75</td>
104 </tr>
105 <tr>
106   <th>2317</th>
107   <td>926</td>
108   <td>1</td>
109   <td>Steak Bowl</td>
110   <td>[Roasted Chili Corn Salsa, [Rice, Black Beans,...
</td>
111   <td>9.25</td>
112 </tr>
113 <tr>
114   <th>2318</th>
115   <td>926</td>
116   <td>1</td>
117   <td>Chicken Bowl</td>
118   <td>[Roasted Chili Corn Salsa, [Fajita Vegetables,...
</td>
119   <td>8.75</td>
120 </tr>
121 <tr>

122   <th>2319</th>
```

```
123    <td>926</td>
124    <td>1</td>
125    <td>Steak Bowl</td>
126    <td>[Fresh Tomato Salsa, [Rice, Cheese]]</td>
127    <td>9.25</td>
128  </tr>
129  <tr>
130    <th>2320</th>
131    <td>926</td>
132    <td>1</td>
133    <td>Chicken Burrito</td>
134    <td>[Fresh Tomato Salsa, [Rice, Cheese, Lettuce]]</td>
135    <td>8.75</td>
136  </tr>
137  <tr>
138    <th>2321</th>
139    <td>926</td>
140    <td>1</td>
141    <td>Chicken Bowl</td>
142    <td>[Roasted Chili Corn Salsa, [Rice, Pinto Beans,...
     </td>
143    <td>8.75</td>
144  </tr>
145  <tr>
146    <th>2322</th>
147    <td>926</td>
148    <td>1</td>
149    <td>Chicken Bowl</td>
150    <td>[Roasted Chili Corn Salsa, [Rice, Cheese]]</td>
151    <td>8.75</td>
152  </tr>
153  <tr>
154    <th>2323</th>

155    <td>926</td>
```

```
156    <td>1</td>
157    <td>Barbacoa Burrito</td>
158    <td>[Fresh Tomato Salsa, [Rice, Black Beans, Chees...
       </td>
159    <td>9.25</td>
160  </tr>
161  <tr>
162    <th>2324</th>
163    <td>926</td>
164    <td>1</td>
165    <td>Chicken Burrito</td>
166    <td>[Tomatillo Red Chili Salsa, [Rice, Cheese, Sou...
       </td>
167    <td>8.75</td>
168  </tr>
169  <tr>
170    <th>2325</th>
171    <td>926</td>
172    <td>1</td>
173    <td>Steak Bowl</td>
174    <td>[Tomatillo Red Chili Salsa, [Rice, Black Beans...
       </td>
175    <td>9.25</td>
176  </tr>
177  <tr>
178    <th>2326</th>
179    <td>926</td>
180    <td>1</td>
181    <td>Veggie Bowl</td>
182    <td>[Roasted Chili Corn Salsa, [Rice, Black Beans,...
       </td>
183    <td>8.75</td>
184  </tr>

185  <tr>
```

```
186    <th>3598</th>
187    <td>1443</td>
188    <td>15</td>
189    <td>Chips and Fresh Tomato Salsa</td>
190    <td>NaN</td>
191    <td>44.25</td>
192 </tr>
193 <tr>
194    <th>3599</th>
195    <td>1443</td>
196    <td>7</td>
197    <td>Bottled Water</td>
198    <td>NaN</td>
199    <td>10.50</td>
200 </tr>
201 <tr>
202    <th>3600</th>
203    <td>1443</td>
204    <td>1</td>
205    <td>6 Pack Soft Drink</td>
206    <td>[Coke]</td>
207    <td>6.49</td>
208 </tr>
209 <tr>
210    <th>3601</th>
211    <td>1443</td>
212    <td>3</td>
213    <td>Veggie Burrito</td>
214    <td>[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...
       </td>
215    <td>33.75</td>
216 </tr>
217 <tr>

218    <th>3602</th>
```

```
    <td>1443</td>
    <td>4</td>
    <td>Chicken Burrito</td>
    <td>[Fresh Tomato Salsa, [Rice, Black Beans, Chees...
</td>
    <td>35.00</td>
</tr>
<tr>
    <th>3603</th>
    <td>1443</td>
    <td>3</td>
    <td>Steak Burrito</td>
    <td>[Fresh Tomato Salsa, [Rice, Black Beans, Chees...
</td>
    <td>27.75</td>
</tr>
<tr>
    <th>3604</th>
    <td>1443</td>
    <td>2</td>
    <td>Bottled Water</td>
    <td>NaN</td>
    <td>3.00</td>
</tr>
<tr>
    <th>3700</th>
    <td>1483</td>
    <td>1</td>
    <td>Chicken Bowl</td>
    <td>[Roasted Chili Corn Salsa, [Rice, Black Beans,...
</td>
    <td>8.75</td>
</tr>

<tr>
```

```
250    <th>3701</th>
251    <td>1483</td>
252    <td>1</td>
253    <td>Steak Bowl</td>
254    <td>[Roasted Chili Corn Salsa, [Fajita Vegetables,...
       </td>
255    <td>11.75</td>
256  </tr>
257  <tr>
258    <th>3702</th>
259    <td>1483</td>
260    <td>1</td>
261    <td>Chicken Burrito</td>
262    <td>[Tomatillo Red Chili Salsa, [Fajita Vegetables...
       </td>
263    <td>8.75</td>
264  </tr>
265  <tr>
266    <th>3703</th>
267    <td>1483</td>
268    <td>1</td>
269    <td>Steak Bowl</td>
270    <td>[Roasted Chili Corn Salsa, [Fajita Vegetables,...
       </td>
271    <td>9.25</td>
272  </tr>
273  <tr>
274    <th>3704</th>
275    <td>1483</td>
276    <td>1</td>
277    <td>Steak Burrito</td>
278    <td>[Roasted Chili Corn Salsa, [Rice, Black Beans,...
       </td>

279    <td>11.75</td>
```

```
280  </tr>
281  <tr>
282    <th>3705</th>
283    <td>1483</td>
284    <td>1</td>
285    <td>Steak Soft Tacos</td>
286    <td>[Fresh Tomato Salsa, [Fajita Vegetables, Black...
     </td>
287    <td>9.25</td>
288  </tr>
289  <tr>
290    <th>3706</th>
291    <td>1483</td>
292    <td>1</td>
293    <td>Chicken Salad Bowl</td>
294    <td>[Roasted Chili Corn Salsa, [Pinto Beans, Chees...
     </td>
295    <td>8.75</td>
296  </tr>
297  <tr>
298    <th>3707</th>
299    <td>1483</td>
300    <td>1</td>
301    <td>Steak Bowl</td>
302    <td>[Roasted Chili Corn Salsa, [Black Beans, Chees...
     </td>
303    <td>9.25</td>
304  </tr>
305  <tr>
306    <th>3708</th>
307    <td>1483</td>
308    <td>1</td>
309    <td>Chicken Bowl</td>
310    <td>[Fresh Tomato Salsa, [Rice, Black Beans, Sour ...
```

```
      </td>
311    <td>8.75</td>
312 </tr>
313 <tr>
314    <th>3709</th>
315    <td>1483</td>
316    <td>1</td>
317    <td>Chicken Bowl</td>
318    <td>[Roasted Chili Corn Salsa, [Rice, Black Beans,...
      </td>
319    <td>11.25</td>
320 </tr>
321 <tr>
322    <th>3710</th>
323    <td>1483</td>
324    <td>1</td>
325    <td>Steak Soft Tacos</td>
326    <td>[Fresh Tomato Salsa, Guacamole]</td>
327    <td>11.75</td>
328 </tr>
329 <tr>
330    <th>3711</th>
331    <td>1483</td>
332    <td>1</td>
333    <td>Carnitas Bowl</td>
334    <td>[Fresh Tomato Salsa, [Fajita Vegetables, Rice,...
      </td>
335    <td>9.25</td>
336 </tr>
337 <tr>
338    <th>3712</th>
339    <td>1483</td>
340    <td>1</td>

341    <td>Chicken Salad Bowl</td>
```

```
342    <td>[Roasted Chili Corn Salsa, [Black Beans, Chees...
    </td>
343    <td>8.75</td>
344  </tr>
345  <tr>
346    <th>3713</th>
347    <td>1483</td>
348    <td>1</td>
349    <td>Steak Bowl</td>
350    <td>[Roasted Chili Corn Salsa, [Rice, Black Beans,...
    </td>
351    <td>11.75</td>
352  </tr>
```