```python
import pandas as pd
```

```python
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```python
yelp= pd.read_csv("pizza.csv")
yelp
```

	business_id	name	neighborhood	address	
0	sxQrSzv4SS4b6o3tgmWS7A	"Chuck E Cheese's"	NaN	"1035 Washington Pike"	Bridge
1	Ql0iiHl2xpIwtkuC255Wiw	"John's Incredible Pizza Company"	Eastside	"3700 S Maryland Pkwy"	Las Ve
2	pbUsYtULpwmhZXlzeTfcww	"Stone & Barrel"	NaN	"24218 S Oakwood Blvd"	Sun La
3	gWsWtppVufrGfGWw1HcjOA	"Dream Lanes"	NaN	"13 Atlas Ct"	Mad
4	T0ju0drMLfXJPNLnEsEYgw	"Pizza Hot Wings"	Scarborough	"3007 Sheppard Avenue E"	Tor
...	...	...	...	...	
6062	7JkPRXmtRmspIX_k3zMNvg	"Teatro Pizzeria and Wine Bar"	NaN	"32409 N Scottsdale Rd"	Scotts
6063	w2lE4nbufBqwrub7CW80xg	"Goodfellas Pizza"	Streetsville	"209 Queen Street S"	Mississa
				"8801 N	

```python
yelp.describe()
```

	latitude	longitude	stars	review_count	is_open
count	6067.000000	6067.000000	6067.000000	6067.000000	6067.000000
mean	39.536523	-88.233598	3.348772	49.129553	0.792484
std	5.007670	24.455903	0.829963	121.137196	0.405562
min	-34.515952	-115.350952	1.000000	3.000000	0.000000
25%	35.200027	-111.939013	3.000000	6.000000	1.000000
50%	40.440941	-81.358555	3.500000	15.000000	1.000000
75%	43.648540	-79.554400	4.000000	44.000000	1.000000
max	59.436505	14.092636	5.000000	3741.000000	1.000000

```
yelp.columns
```

```
Index(['business_id', 'name', 'neighborhood', 'address', 'city', 'state',
       'postal_code', 'latitude', 'longitude', 'stars', 'review_count',
       'is_open', 'categories'],
      dtype='object')
```

```
yelp.dtypes
```

```
business_id      object
name             object
neighborhood     object
address          object
city             object
state            object
postal_code      object
latitude        float64
longitude       float64
stars           float64
review_count      int64
is_open           int64
categories       object
dtype: object
```

```
yelp = yelp.replace({'\$':''}, regex = True)
```

```
plt.scatter(yelp['stars'], yelp['review_count'])
#plot of relation between review count and stars
```

Saved successfully!

```
<matplotlib.collections.PathCollection at 0x7f451f4dfa90>
```



```
yelp.sort_values(by='stars', ascending=False)
```

	business_id	name	neighborhood	address	ci
3473	Perap0eAyCbUx4r4rQTSnA	"Pie Express"	Oakland	"148 Oakland Ave"	Pittsbu
4649	ihWw0eSKMQfO5cXLFS5itQ	"Homestyle Pizza"	NaN	"3350 Millers Run Rd"	C
3063	ueHTq9P9jbb5Ar0_zo6YKw	"Restaurant Dorion"	NaN	"297 Boul Harwood"	Vaudre Dori
3537	NuT1ejKruFnBFQWC6Xg82w	"Papa Murphy's"	Spring Valley	"7210 S. Durango Dr., Suite A"	Las Veg
5976	epyhvtf5JH5sAJggINVglg	"J&B's Pizza, BBQ And Deli"	NaN	"9950 E Broadway Rd"	Me
...	...	...	...	...	
5197	YHN7wVaOFp9wkl_w7NXLXw	"Kens Burgers and Pizza"	Southeast	"7141 S Eastern Ave, Ste D"	Las Veg
2198	Lvr6DdVkFl8IMGfRvUcXQA	"Little Caesars	NaN	"4340 W McDowell Rd Ste	Phoe

Saved successfully! ✕

	business_id	name	neighborhood	address	city
231	q-YQIvBSNZxYJl1xuB0H_w	"CM2 Pizzeria & Bakeshop"	NaN	"11485 N 136th St"	Scottsdale
279	j65Mnw8aFJkYgkt_UR5BOw	"Matt's Sub Shack and Pizza"	NaN	"812 Little Deer Creek Valley Rd"	Russellton
284	F6eEu0qhYpS99e1ag3q0Bw	"Braw Burgers And Pizza"	Newington	"54A Clerk Street"	Edinburgh
344	893VryJbZcCm5V9xon_aLA	"Those Guys Pies"	Northwest	"3369 Thom Blvd, rose grilled"	Las Vegas
352	Lkq-3a2oZUPDSUWBRzUXWg	"Senor Pizza"	NaN	"1635 E Baseline Rd"	Phoenix
...	...	...	...	...	...
5897	VsRAIb4k5CjEF-3l_bac0g	"Akropolis Gyro & Pizza"	NaN	"1690 W Sunset Rd, Ste 104"	Henderson

```python
star = yelp.loc[yelp['stars'] == 5]
```

```python
star.sort_values(by='review_count', ascending=False)
```
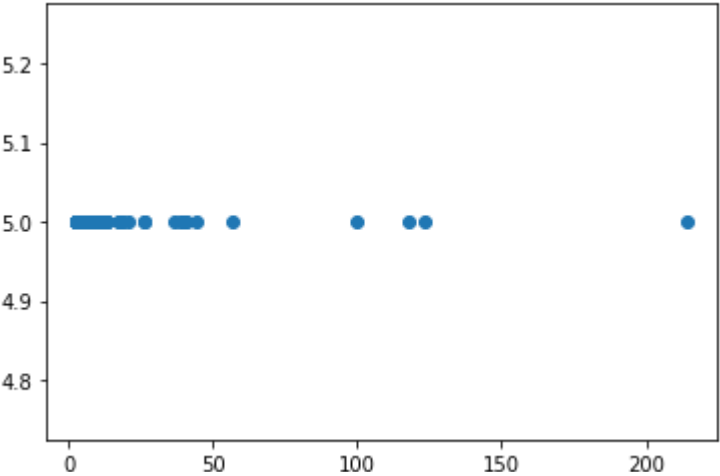
Saved successfully!   ✕

	business_id	name	neighborhood	address	cit
5897	VsRAIb4k5CjEF-3l_bac0g	"Akropolis Gyro & Pizza"	NaN	"1690 W Sunset Rd, Ste 104"	Hendersc
3475	FvXZcRB8bocNMDvFUnoWhg	"In Forno Pizza"	NaN	"35840 Chester Rd"	Avc
1538	E_jLyf_YuGgMP_rw8tvNSA	"Niko's Pizza Las Vegas"	Spring Valley	"4555 S Fort Apache Rd, Ste 112"	Las Vega
1648	Llm_iXzE0-8_XKwI2e4JdA	"Saffron JAK"	NaN	"814 E Union Hills Dr, Ste C-6"	Phoen
		"Do			

```python
plt.scatter(star['review_count'], star['stars'])
```

    <matplotlib.collections.PathCollection at 0x7f451f44d490>


```python
star.describe()
```

Saved successfully!   ✕

	latitude	longitude	stars	review_count	is_open	
count	123.000000	123.000000	123.0	123.000000	123.000000	

```
yelp.sort_values(by='review_count', ascending=False)
```

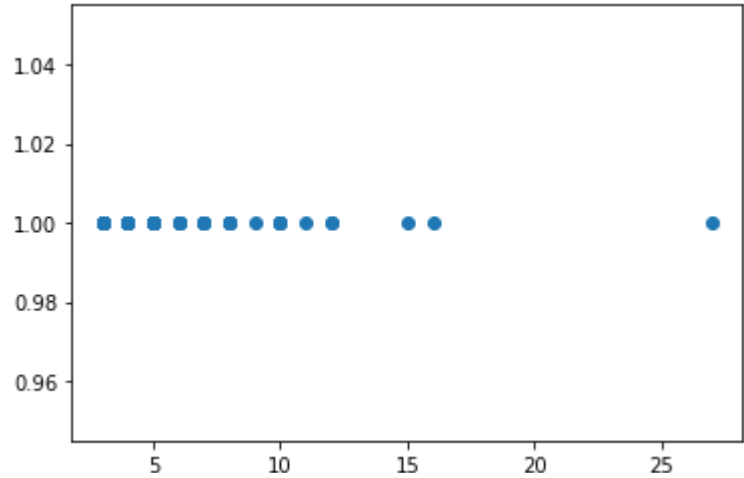|	business_id	name	neighborhood	address	c:
2326	iCQpiavjjPzJ5_3gPD5Ebg	"Secret Pizza"	The Strip	"The Cosmopolitan Of Las Vegas, 3708 Las Vegas...	Ve
3704	3kdSl5mo9dWC4clrQjEDGg	"Egg & I"	Westside	"4533 W Sahara Ave, Ste 5"	Ve
5613	pSQFynH1VxkfSmehRXlZWw	"Pizzeria Bianco"	NaN	"623 E Adams St"	Phoe
48	JzOp695tclcNCNMuBl7oxA	"Four Peaks Brewing"	NaN	"1340 E 8th St, Ste 104"	Ten
2633	3GEEy7RP6e4bT4LAiWFMFQ	"Pizza Rock"	Downtown	"201 N 3rd St"	Ve
...	...	...	...	...	
4480	VOvcCzgnhyJ63r3Xk5gvpQ	"Vocelli Pizza"	NaN	"1025 Washington Pike, Ste 27"	Bridge
4473	Uq2mlRObAytzkTloerN03Q	"Pisa Pizza"	NaN	"2015 N Power Rd"	M

```
yelp.loc[yelp['stars'] == 1]
```

Saved successfully!   ✕

	business_id	name	neighborhood	address	ci
416	OIpS3O5ifR0wFu5PdvMM5w	"Pizza Hut"	NaN	"1165 Mentor Ave"	Painesv
451	O5vXNIody0SMddowzrAMXQ	"Pizza Hut"	NaN	"3734 W T Harris Blvd"	Charlc
489	F2CdVtudYJlSgTAqWDX8gA	"Pizza Hut"	NaN	"2324 Ardmore Blvd"	For H
506	CdqDmKlSVrTBigSZnpQYug	"Pizza Hut"	NaN	"10050 W Bell Rd, Ste 22"	Sun C
509	XgsGLjUpVhfTqvkapVxYHw	"Pizza Hut"	NaN	"3044 Eastway Dr"	Charlc

```
lowstar = yelp.loc[yelp['stars'] == 1]
#list all of the restaurants with 1 star
```

| 5460 | nzUv52JpOQz98Fk9GB4AKA | "Pizza Pizza" | NaN | Keele | North Y |

```
plt.scatter(lowstar['review_count'], lowstar['stars'])
```

    <matplotlib.collections.PathCollection at 0x7f451f433a10>


```
lowstar.describe()
```

Saved successfully!   ✕

	latitude	longitude	stars	review_count	is_open
count	69.000000	69.000000	69.0	69.000000	69.000000
mean	39.681012	-88.002616	1.0	5.753623	0.782609
std	4.140976	19.182637	0.0	3.979380	0.415493
min	33.414660	-115.282641	1.0	3.000000	0.000000

```
lowstar.sort_values(by='review_count', ascending=False)
```

	business_id	name	neighborhood	address	city
451	O5vXNIody0SMddowzrAMXQ	"Pizza Hut"	NaN	"3734 W T Harris Blvd"	Charlotte
489	F2CdVtudYJlSgTAqWDX8gA	"Pizza Hut"	NaN	"2324 Ardmore Blvd"	Forest Hills
3968	AyuE2AyyKSZj97SNLAHM9w	"Sbarro"	Downtown Core	"Toronto Eaton Center, 220 Yonge Street"	Toronto
5910	S0yP4IZmwVpeqxSe0R2Fvw	"Subway"	The Strip	"2890 Las Vegas Blvd S"	Las Vegas
5237	mILfcRQxAippB-Y6Bm4uiw	"Pizza Hut"	NaN	"3530 Mt Holly Hunterville Rd"	Charlotte
...	...	...	...	...	...
				"1837 Finch	

```
pizzahut=·yelp.loc[yelp['name']·==·'"Pizza·Hut"']
```

```
yelp.loc[yelp['name'] == '"Pizza Hut"']
```

Saved successfully!   ✕

	business_id	name	neighborhood	address	city
362	Gq4J8WD_mSY-o50R1PEKjw	"Pizza Hut"	NaN	"Fife Leisure Scheme, Whimbrel Place"	Dunfermline
365	NxZt-j74NpvIU5hdb2KJTA	"Pizza Hut"	NaN	"2727 E McKellips Rd, Ste 106"	Mesa
373	7HjVdfoTqs4YbKqmoyKOeQ	"Pizza Hut"	Stockbridge	"110 Raeburn Place"	Edinburgh
374	WFM2FE0THduTbnjgGdGDLQ	"Pizza Hut"	NaN	"80 Allegheny River Blvd"	Verona
375	AZw_8ngMd8vnAbzKX7zPlw	"Pizza Hut"	NaN	"1317 Savoy Plaza Ln"	Savoy
...	...	...	...	...	...
		"Pizza		"8225 W Indian	

```
pizzahut.describe()
```

	latitude	longitude	stars	review_count	is_open
count	300.000000	300.000000	300.000000	300.000000	300.000000
mean	38.767055	-91.884392	2.261667	12.160000	0.940000
std	4.687250	21.080315	0.743330	10.658318	0.237884
min	33.234238	-115.324682	1.000000	3.000000	0.000000
25%	35.112221	-112.079480	2.000000	4.000000	1.000000
50%	36.304323	-81.764806	2.000000	9.000000	1.000000
75%	41.683552	-80.012082	2.500000	15.000000	1.000000
max	57.481913	9.173350	5.000000	71.000000	1.000000

Saved successfully!   '] == '"Little Caesars Pizza"']

```
littleceasers.describe()
```

	latitude	longitude	stars	review_count	is_open
count	71.000000	71.000000	71.000000	71.000000	71.000000
mean	36.943075	-101.588974	2.542254	12.690141	0.985915
std	3.998433	15.423775	0.791682	17.302842	0.118678
min	33.231942	-115.282305	1.000000	3.000000	0.000000
25%	33.481192	-112.349743	2.000000	4.000000	1.000000
50%	36.012544	-111.910525	2.500000	8.000000	1.000000
75%	41.299636	-81.707586	3.000000	13.000000	1.000000
max	45.480618	-73.842224	5.000000	116.000000	1.000000

```python
Pizzanova= yelp.loc[yelp['name'] == '"Pizza Nova"']
```

```python
Pizzanova.describe()
```

	latitude	longitude	stars	review_count	is_open
count	42.000000	42.000000	42.000000	42.00000	42.0
mean	43.749309	-79.464962	3.119048	5.50000	1.0
std	0.104866	0.140029	0.810070	3.63083	0.0
min	43.559505	-79.787493	1.500000	3.00000	1.0
25%	43.666233	-79.522047	2.500000	3.00000	1.0
50%	43.753711	-79.428149	3.000000	4.00000	1.0
75%	43.805087	-79.369621	3.500000	6.00000	1.0
max	44.018416	-79.267795	5.000000	20.00000	1.0

```python
Pizzapizza= yelp.loc[yelp['name'] == '"Pizza Pizza"']
```

```python
Pizzapizza.describe()
```

Saved successfully!   ✕

	latitude	longitude	stars	review_count	is_open	
count	85.000000	85.000000	85.000000	85.000000	85.000000	
mean	43.866197	-78.944980	2.370588	5.635294	0.917647	

```
yelp.loc[yelp['review_count'] > 1000]
#list all of the restauarnt with over 1000 review count
```

	business_id	name	neighborhood	address	
48	JzOp695tclcNCNMuBl7oxA	"Four Peaks Brewing"	NaN	"1340 E 8th St, Ste 104"	T
971	0FUtlsQrJI7LhqDPxLumEw	"Joe's Farm Grill"	NaN	"3000 E Ray Rd, Bldg 1"	C
1499	pHJu8tj3sI8eC5aIHLFEfQ	"Nora's Italian Cuisine"	NaN	"5780 W Flamingo Rd"	V
1622	N0apJkxIem2E8irTBRKnHw	"Wolfgang Puck Bar & Grill Las Vegas"	The Strip	"3799 Las Vegas Blvd S"	V
1631	UYh1N1xQhOh-a7nX92xPzQ	"Brio Tuscan Grille"	Southeast	"6653 Las Vegas Blvd"	V
1662	LwQB9H3jZ9wTk24Lr-AnZQ	"Spago by Wolfgang Puck"	The Strip	"The Forum Shops At Caesars, 3500 Las Vegas Bl..."	V
2326	iCQpiavjjPzJ5_3gPD5Ebg	"Secret Pizza"	The Strip	"The Cosmopolitan Of Las Vegas, 3708 Las Vegas..."	V
2633	3GEEy7RP6e4bT4LAiWFMFQ	"Pizza Rock"	Downtown	"201 N 3rd St"	V
3704	3kdSl5mo9dWC4clrQjEDGg	"Egg & I"	Westside	"4533 W Sahara Ave, Ste 5"	V
5613	pSQFynH1VxkfSmehRXlZWw	"Pizzeria Bianco"	NaN	"623 E Adams St"	Ph
	...w	"Cibo"	NaN	"603 N 5th Ave"	Ph

Saved successfully! ✕

```
yelp.loc[yelp['state'] == "NV"]
#list amount of pizza restuants per state "NV" is interchangable to any valid states such as
```

	business_id	name	neighborhood	address	ci
1	Ql0iiHl2xpIwtkuC255Wiw	"John's Incredible Pizza Company"	Eastside	"3700 S Maryland Pkwy"	Las Veg
8	pfmr8R3WH8RXqW0W6D8ffQ	"Lombardi's Romagna Mia"	The Strip	"3663 Las Vegas Blvd S"	Las Veg
13	aQUo8irLBywAZN26ln_Q1w	"WILD"	Downtown	"150 N Las Vegas Blvd, Ste 120"	Las Veg
24	bJP4l_BGq2CudEu0m-wNjg	"Artisan Fine Dining Room"	NaN	"Artisan Hotel, 1501 W Sahara Ave"	Las Veg
26	LhaOYo_5j5W_JIY5fYPKuQ	"Brooklyn's Restaurant"	NaN	"10 Via Brianza"	Henders
...	...	...	...	...	
6010	#NAME?	"Double Play Sports Bar"	Southeast	"9495 Las Vegas Blvd S"	Las Veg
		"Fellini's	The Strip	"2000 Las	

```
NV= yelp.loc[yelp['state'] == "NV"]
```

```
NV.describe()
```

Saved successfully!   ✕

	latitude	longitude	stars	review_count	is_open
count	763.000000	763.000000	763.000000	763.000000	763.000000
mean	36.125157	-115.173321	3.268676	109.507208	0.701180
std	0.076479	0.086122	0.851447	232.196786	0.458041

```
AZ= yelp.loc[yelp['state'] == "AZ"]
```

	latitude	longitude	stars	review_count	is_open
25%	36.069509	-115.243285	2.500000	14.000000	0.000000

```
AZ.describe()
```

	latitude	longitude	stars	review_count	is_open
count	1327.000000	1327.000000	1327.000000	1327.000000	1327.000000
mean	33.491173	-111.999381	3.365863	86.009043	0.725697
std	0.129254	0.187601	0.792177	155.695089	0.446331
min	33.137323	-112.497764	1.000000	3.000000	0.000000
25%	33.388714	-112.132860	3.000000	12.000000	0.000000
50%	33.480899	-111.978699	3.500000	33.000000	1.000000
75%	33.602500	-111.873812	4.000000	96.000000	1.000000
max	33.873916	-111.506818	5.000000	2035.000000	1.000000

```
yelp['state'].value_counts()
```

```
AZ    1327
ON    1075
OH     765
NV     763
PA     739
QC     450
NC     420
WI     177
BW     149
IL      88
EDH     63
SC      26
MLN      4
NYK      4
1        3
CHE      3
```

Saved successfully! ✕

```
NY       2
WHT      1
CA       1
```

```
       WLN          1
       Name: state, dtype: int64
```

`yelp.isnull()`

	business_id	name	neighborhood	address	city	state	postal_code	la
0	False	False	True	False	False	False	False	
1	False	False	False	False	False	False	False	
2	False	False	True	False	False	False	False	
3	False	False	True	False	False	False	False	
4	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	
6062	False	False	True	False	False	False	False	
6063	False	False	False	False	False	False	False	
6064	False	False	True	False	False	False	False	
6065	False	False	False	False	False	False	False	
6066	False	False	True	False	False	False	False	

`yelp.isnull().sum()`

```
       business_id        0
       name               0
       neighborhood    3885
       address            0
       city               0
       state              0
       postal_code        4
       latitude           0
       longitude          0
       stars              0
       review_count       0
       is_open            0
       categories         0
       dtype: int64
```

Saved successfully!   ✕

`X= yelp[['review_count']]`

`X`

	review_count
0	11
1	268
2	23
3	8
4	3
...	...
6062	45
6063	116
6064	337
6065	10
6066	13

6067 rows × 1 columns

y

	stars
0	3.0
1	3.0
2	4.0
3	3.0
4	3.5
...	...
6062	3.5
6063	3.5
6064	4.0
6065	3.0
6066	3.5

Saved successfully! ✕

```python
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size =0.2, random_state=10)
```

```
len(X_train)
```

```
4853
```

```
len(X_test)
```

```
1214
```

```
len(y_train)
```

```
4853
```

```
from sklearn.linear_model import LinearRegression
clf = LinearRegression()
```

```
clf.fit(X_train,y_train)
```

```
LinearRegression()
```

```
clf.predict(X_test)
```

```
array([[3.29254607],
       [3.62439866],
       [3.43410556],
       ...,
       [3.28906509],
       [3.29370639],
       [3.29486671]])
```

```
clf.predict(y_test)
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:493: FutureWarning: The feature
Feature names unseen at fit time:
- stars
Feature names seen at fit time, yet now missing:
- review_count

  warnings.warn(message, FutureWarning)
array([[3.28790477],
       [3.29080558],
       [3.29022542],
```

Saved successfully!   ✕

```
[3.20004320],
       [3.28906509]])
```

```
clf_pred = clf.predict(X_test)
```

```
y_test
```

	stars
3679	2.0
4163	4.5
5724	4.0
4495	4.5
3328	4.0
...	...
1523	2.5
1144	4.5
176	3.0
2926	3.5
3832	3.0

1214 rows × 1 columns

```
clf.score(X_test, y_test)
```

```
0.03491541103556728
```

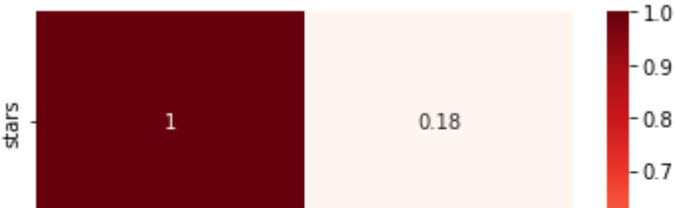```
clf.score(X_train, y_train)
```

```
0.02917716057971298
```

```
yelp[['stars','review_count']].corr()
```

```
sns.heatmap(yelp[['stars','review_count']].corr(), annot=True, cmap = 'Reds')
plt.show()
```

Saved successfully!   ✕

```
remove = ['longitude' , 'neighborhood', 'latitude']

yelp.drop(remove, inplace =True, axis =1)
```



```
yelp.duplicated()
```

```
0       False
1       False
2       False
3       False
4       False
        ...
6062    False
6063    False
6064    False
6065    False
6066    False
Length: 6067, dtype: bool
```

```
yelp.drop_duplicates()
```

	business_id	name	address	city	state	pc
0	sxQrSzv4SS4b6o3tgmWS7A	"Chuck E Cheese's"	"1035 Washington Pike"	Bridgeville	PA	
1	Ql0iiHl2xpIwtkuC255Wiw	"John's Incredible Pizza Company"	"3700 S Maryland Pkwy"	Las Vegas	NV	
2	pbUsYtULpwmhZXlzeTfcww	"Stone & Barrel"	"24218 S Oakwood Blvd"	Sun Lakes	AZ	
3	gWsWtppVufrGfGWw1HcjOA	"Dream Lanes"	"13 Atlas Ct"	Madison	WI	
	w	"Pizza Hot Wings"	"3007 Sheppard Avenue E"	Toronto	ON	

Saved successfully! ✕

```
yelp['stars'].describe()
```

```
count    6067.000000
mean        3.348772
std         0.829963
min         1.000000
25%         3.000000
50%         3.500000
75%         4.000000
max         5.000000
Name: stars, dtype: float64
```
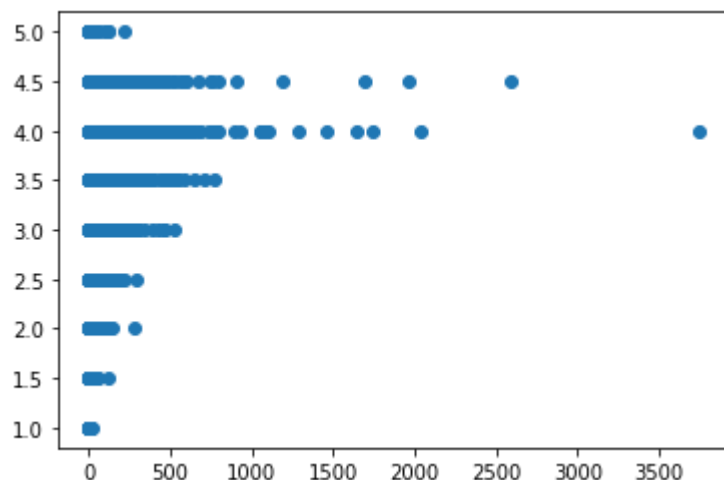
```
yelp['review_count'].describe()
```

```
count    6067.000000
mean       49.129553
std       121.137196
min         3.000000
25%         6.000000
50%        15.000000
75%        44.000000
max      3741.000000
Name: review_count, dtype: float64
```

```
%matplotlib inline
plt.scatter(yelp['review_count'], yelp['stars'])
```

```
<matplotlib.collections.PathCollection at 0x7f451f286e10>
```


Saved successfully!   ✕

✓ 0s   completed at 9:21 PM   ● ✕

Saved successfully!   ✕