```
import pandas as pd
```

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

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

	business_id	name	neighborhood	address	city	state
0	sxQrSzv4SS4b6o3tgmWS7A	"Chuck E Cheese's"	NaN	"1035 Washington Pike"	Bridgeville	PA
1	Ql0iiHl2xpIwtkuC255Wiw	"John's Incredible Pizza Company"	Eastside	"3700 S Maryland Pkwy"	Las Vegas	NV
2	pbUsYtULpwmhZXlzeTfcww	"Stone & Barrel"	NaN	"24218 S Oakwood Blvd"	Sun Lakes	AZ
3	gWsWtppVufrGfGWw1HcjOA	"Dream Lanes"	NaN	"13 Atlas Ct"	Madison	WI
4	T0ju0drMLfXJPNLnEsEYgw	"Pizza Hot Wings"	Scarborough	"3007 Sheppard Avenue E"	Toronto	ON
...	...	...	...	...	...	...
6062	7JkPRXmtRmspIX_k3zMNvg	"Teatro Pizzeria and Wine Bar"	NaN	"32409 N Scottsdale Rd"	Scottsdale	AZ
6063	w2IE4nbufBqwrub7CW80xg	"Goodfellas Pizza"	Streetsville	"209 Queen Street S"	Mississauga	ON
6064	ToFm2DhhTdr0Kv0A7Sdj6g	"Timo Wine Bar"	NaN	"8801 N Central Ave"	Phoenix	AZ

```
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
```

```
<matplotlib.collections.PathCollection at 0x7fa52dd50050>
```

3500
3000
2500
2000
1500
1000

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

	business_id	name	neighborhood	address	city	state
3473	Perap0eAyCbUx4r4rQTSnA	"Pie Express"	Oakland	"148 Oakland Ave"	Pittsburgh	PA
4649	ihWw0eSKMQfO5cXLFS5itQ	"Homestyle Pizza"	NaN	"3350 Millers Run Rd"	Cecil	PA
3063	ueHTq9P9jbb5Ar0_zo6YKw	"Restaurant Dorion"	NaN	"297 Boul Harwood"	Vaudreuil-Dorion	QC
3537	NuT1ejKruFnBFQWC6Xg82w	"Papa Murphy's"	Spring Valley	"7210 S. Durango Dr., Suite A"	Las Vegas	NV
5976	epyhvtf5JH5sAJggINVglg	"J&B's Pizza, BBQ And Deli"	NaN	"9950 E Broadway Rd"	Mesa	AZ
...	...	...	...	...	...	...
5197	YHN7wVaOFp9wkl_w7NXLXw	"Kens Burgers and Pizza"	Southeast	"7141 S Eastern Ave, Ste D"	Las Vegas	NV
2198	Lvr6DdVkFI8IMGfRvUcXQA	"Little Caesars Pizza"	NaN	"4340 W McDowell Rd, Ste 6"	Phoenix	AZ

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

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

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

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

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

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

	business_id	name	neighborhood	address	city	state

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

	business_id	name	neighborhood	address	city	state
416	OIpS3O5ifR0wFu5PdvMM5w	"Pizza Hut"	NaN	"1165 Mentor Ave"	Painesville	OH
451	O5vXNIody0SMddowzrAMXQ	"Pizza Hut"	NaN	"3734 W T Harris Blvd"	Charlotte	NC
489	F2CdVtudYJlSgTAqWDX8gA	"Pizza Hut"	NaN	"2324 Ardmore Blvd"	Forest Hills	PA
506	CdqDmKlSVrTBigSZnpQYug	"Pizza Hut"	NaN	"10050 W Bell Rd, Ste 22"	Sun City	AZ
509	XgsGLjUpVhfTqvkapVxYHw	"Pizza Hut"	NaN	"3044 Eastway Dr"	Charlotte	NC
...	...	...	...	...	...	...
5460	nzUv52JpOQz98Fk9GB4AKA	"Pizza Pizza"	NaN	"3358 Keele Street"	North York	ON
5478	n93mgR2PhSXmB_znHJBBGA	"Pizza Hut"	NaN	"10 Jacob Keffer Parkway	Concord	ON

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

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

	business_id	name	neighborhood	address	city	state	p
451	O5vXNIody0SMddowzrAMXQ	"Pizza Hut"	NaN	"3734 W T Harris Blvd"	Charlotte	NC	
489	F2CdVtudYJlSgTAqWDX8gA	"Pizza Hut"	NaN	"2324 Ardmore Blvd"	Forest Hills	PA	
3968	AyuE2AyyKSZj97SNLAHM9w	"Sbarro"	Downtown Core	"Toronto Eaton Center, 220 Yonge Street"	Toronto	ON	
5910	S0yP4IZmwVpeqxSe0R2Fvw	"Subway"	The Strip	"2890 Las Vegas Blvd S"	Las Vegas	NV	

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

	business_id	name	neighborhood	address	city	stat
48	JzOp695tclcNCNMuBl7oxA	"Four Peaks Brewing"	NaN	"1340 E 8th St, Ste 104"	Tempe	A
971	0FUtlsQrJI7LhqDPxLumEw	"Joe's Farm Grill"	NaN	"3000 E Ray Rd, Bldg 1"	Gilbert	A
1499	pHJu8tj3sI8eC5aIHLFEfQ	"Nora's Italian Cuisine"	NaN	"5780 W Flamingo Rd"	Las Vegas	N
		"Wolfgang Puck Bar &		"3799 Las	Las	

```
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	city	state
1	Ql0iiHl2xpIwtkuC255Wiw	"John's Incredible Pizza Company"	Eastside	"3700 S Maryland Pkwy"	Las Vegas	NV
8	pfmr8R3WH8RXqW0W6D8ffQ	"Lombardi's Romagna Mia"	The Strip	"3663 Las Vegas Blvd S"	Las Vegas	NV
13	aQUo8irLBywAZN26ln_Q1w	"WILD"	Downtown	"150 N Las Vegas Blvd, Ste 120"	Las Vegas	NV
24	bJP4l_BGq2CudEu0m-wNjg	"Artisan Fine Dining Room"	NaN	"Artisan Hotel, 1501 W Sahara Ave"	Las Vegas	NV
26	LhaOYo_5j5W_JIY5fYPKuQ	"Brooklyn's Restaurant"	NaN	"10 Via Brianza"	Henderson	NV
...	...	...	...	...	...	...
6010	#NAME?	"Double Play Sports Bar"	Southeast	"9495 Las Vegas Blvd S"	Las Vegas	NV
6019	LN0JGAl8Rr_r_5t_X8Kz6g	"Fellini's Ristorante"	The Strip	"2000 Las Vegas Blvd S"	Las Vegas	NV

```
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
FIF       2
HLD       2
C         2
NY        2
WHT       1
CA        1
WLN       1
Name: state, dtype: int64
```

```
yelp.isnull()
```

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

6067 rows × 13 columns

```
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
```

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

```
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

```
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)
```

6067 rows × 1 columns

```
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]])
```

```
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)
```

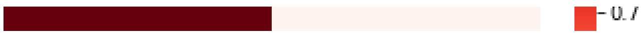
```
0.03491541103556728
```

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

#plot the correlation matrix of salary, balance and age in data dataframe.
sns.heatmap(yelp[['stars','review_count']].corr(), annot=True, cmap = 'Reds')
plt.show()
```

```
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	postal_code
0	sxQrSzv4SS4b6o3tgmWS7A	"Chuck E Cheese's"	"1035 Washington Pike"	Bridgeville	PA	15017
1	Ql0iiHl2xpIwtkuC255Wiw	"John's Incredible Pizza Company"	"3700 S Maryland Pkwy"	Las Vegas	NV	89119
2	pbUsYtULpwmhZXlzeTfcww	"Stone & Barrel"	"24218 S Oakwood Blvd"	Sun Lakes	AZ	85248
3	gWsWtppVufrGfGWw1HcjOA	"Dream Lanes"	"13 Atlas Ct"	Madison	WI	53714
4	T0ju0drMLfXJPNLnEsEYgw	"Pizza Hot Wings"	"3007 Sheppard Avenue E"	Toronto	ON	M1T 3J5
...	...	...	...	...	...	...

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

```
count    6067.000000
mean        3.348772
std         0.829963
```

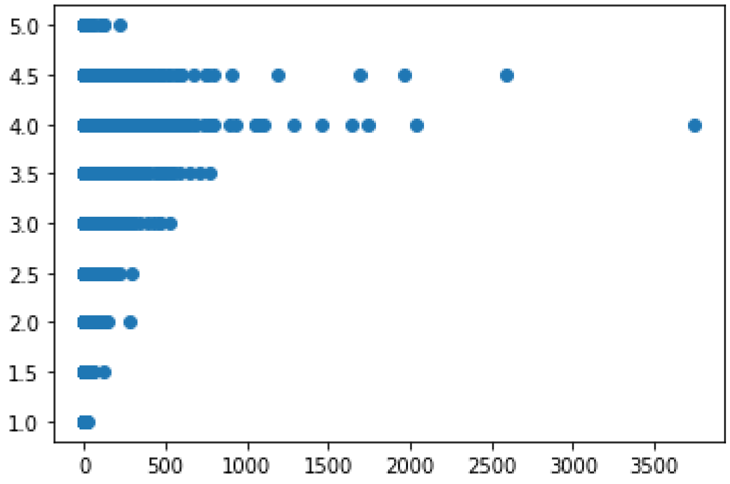
```
min          1.000000
25%          3.000000
50%          3.500000
75%          4.000000
max          5.000000
Name: stars, dtype: float64
```

```
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 0x7fa52a19bb10>
```



✓ 0s completed at 9:54 PM ● ✕