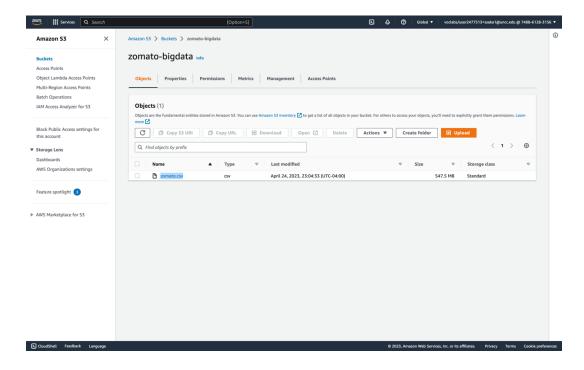Git Hub: https://github.com/Vishnureddy466/bigData
S3 Bucket

Amazon Sage Maker

No two features are highly correlated so no need of dropping columns further

```
[44]: data.head()
```

[44]:

	name	online_order	book_table	rate	votes	location	rest_type	cuisines	estimated_cost_for_2_ppl	reviews_list	menu_item	type
0	Jalsa	1	1	4.1	775	1	1	51	800.0	[('Rated 4.0', 'RATED\n A beautiful place to ...	9062	0
1	Spice Elephant	1	0	4.1	787	1	1	51	800.0	[('Rated 4.0', 'RATED\n Had been here for din...	9062	0
2	San Churro Cafe	1	0	3.8	918	1	6	51	800.0	[('Rated 3.0', "RATED\n Ambience is not that ...	9062	0
3	Addhuri Udupi Bhojana	0	0	3.7	88	1	4	47	300.0	[('Rated 4.0', "RATED\n Great food and proper...	9062	0
4	Grand Village	0	0	3.8	166	4	1	51	600.0	[('Rated 4.0', 'RATED\n Very good restaurant ...	9062	0

Splitting Dataset into train and test datasets

```
[45]: x = newdata.loc[:,['online_order','book_table','votes','location','rest_type','cuisines','estimated_cost_for_2_ppl','type','menu_item']]
      y = newdata['rate']
```

```
[46]: from sklearn.model_selection import train_test_split
      x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=.2)
      x_train.head()
```

[46]:

	online_order	book_table	votes	location	rest_type	cuisines	estimated_cost_for_2_ppl	type	menu_item
12166	1	0	20	6	1	34	550.0	2	9062
31143	1	1	236	23	1	51	800.0	4	7681
38513	1	0	6	37	6	51	400.0	3	8148
46589	0	0	0	40	6	51	400.0	2	9062
47226	0	0	988	27	6	51	3000.0	0	9062

```
[ ]:
```