

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

Amazon S3 console view of the **zomato-bigdata** bucket. The **Objects** tab is selected, showing a list of objects. One object is present:

Name	Type	Last modified	Size	Storage class
zomato.csv	csv	April 24, 2023, 23:04:33 (UTC-04:00)	547.5 MB	Standard

## Amazon Sage Maker

Amazon SageMaker console view of **Notebook instances**. The table shows the following instance:

Name	Instance	Creation time	Status	Actions
zomato-del-2	ml.t2.medium	4/24/2023, 11:05:44 PM	InService	<a href="#">Open Jupyter</a>   <a href="#">Open JupyterLab</a>

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

/

Name

Last Modified

Deliverable\_final.ipynb5 minutes ago

iii

Launcher

Launcher

Deliverable\_final.ipynb

conda\_python3

estimated\_cost\_for\_2\_ppi

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

[44]: data.head()

	name	online_order	book_table	rate	votes	location	rest_type	cuisines	estimated_cost_for_2_ppi	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\_ppi','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()

	online_order	book_table	votes	location	rest_type	cuisines	estimated_cost_for_2_ppi	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	998	27	6	51	3000.0	0	9062

[ ]:

Simple

01

Initialized (additional servers needed)

conda\_python3 | idle

Mode: Command

Ln 1, Col 51

Deliverable\_final.ipynb 1