

```
#Deleting Unnnecessary Columns
zomato=zomato_real.drop(['url','dish_liked','phone'],ax
is=1) #Dropping the column "dish_liked", "phone", "url"
and saving the new dataset as "zomato"
```

```
#Removing the Duplicates
zomato.duplicated().sum()
zomato.drop_duplicates(inplace=True)
```

```
#Remove the NaN values from the dataset
zomato.isnull().sum()
zomato.dropna(how='any',inplace=True)
```

```
#Changing the column names
zomato = zomato.rename(columns={'approx_cost(for two
people)':'cost','listed_in(type)':'type',
'listed_in(city)':'city'})
```

```
#Some Transformations
zomato['cost'] = zomato['cost'].astype(str) #Changing
the cost to string
zomato['cost'] = zomato['cost'].apply(lambda x:
x.replace(',','.')) #Using lambda function to replace
',' from cost
zomato['cost'] = zomato['cost'].astype(float)
#Removing '/5' from Rates
zomato = zomato.loc[zomato.rate != 'NEW']
zomato = zomato.loc[zomato.rate !=
='-'].reset_index(drop=True)
remove_slash = lambda x: x.replace('/5', '') if
isinstance(x, str) else x
```

```
zomato.rate =
zomato.rate.apply(remove_slash).str.strip().astype('flo
at')
```

```
# Adjust the column names
zomato.name = zomato.name.apply(lambda x:x.title())
zomato.online_order.replace(('Yes','No'),(True,
False),inplace=True)
zomato.book_table.replace(('Yes','No'),(True,
False),inplace=True)
```

```
## Computing Mean Rating
restaurants = list(zomato['name'].unique())
zomato['Mean Rating'] = 0
```

```
for i in range(len(restaurants)):
    zomato['Mean Rating'][zomato['name'] ==
restaurants[i]] = zomato['rate'][zomato['name'] ==
restaurants[i]].mean()
```

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler(feature_range = (1,5))
zomato[['Mean Rating']] =
scaler.fit_transform(zomato[['Mean Rating']]).round(2)
```

Now the next step is to perform some text preprocessing steps which include:

1. Lower casing
2. Removal of Punctuations
3. Removal of Stopwords
4. Removal of URLs
5. Spelling correction