

COMPARING THE RECOMMENDATION SYSTEM OF LEADING ECOMMERCE WEBSITES AMAZON,BIGBASKET , FLIPKART ON INTEGRATED DATA

IMPORING THE LIBRARIES AND PACKAGES

```
!pip install scrapy
!pip install tabulate

!pip install scrapy\n!pip install tabulate
import os
import logging
import json
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import scrapy
from scrapy.crawler import CrawlerProcess
from glob import glob
import glob, os
from ipywidgets import widgets
from IPython.display import display, clear_output
from tabulate import tabulate
pd.set_option('display.max_colwidth', None)
import warnings
warnings.filterwarnings('ignore')
CONST_USER_AGENT = 'Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1)'

CONST_PRODUCT = 'Product'
CONST_URL = 'Url'
CONST_PLATFORM = 'Platform'

CONST_MAX_PAGE_COUNT = 'Max_Page_Count'
CONST_PRODUCT_NAME = 'Product_Name'
CONST_PRODUCT_COST = 'Product_Cost'

CONST_PRODUCT_BEVERAGES = 'Beverages'
CONST_PRODUCT_BEVERAGES_TEA = 'Tea'
CONST_PRODUCT_BEVERAGES_COFFEE = 'Coffee'
CONST_PRODUCT_GROCERIES = 'Grocery'
CONST_PRODUCT_MASALA_AND_SPICES = 'Masala & Spices'

CONST_PLATFORM_BIGBASKET = 'BigBasket'
```

```
CONST_PLATFORM_DISPATCHER = 'Dispatcher'
CONST_PLATFORM_AMAZON = 'Amazon'
CONST_PLATFORM_FLIPKART = 'Flipkart'
```

```
CONST_DF_FORMAT_CSV = '.csv'
CONST_DF_FORMAT_JSON = '.json'
```

```
CONST_NOT_APPLICABLE = 'N/A'
```

```
Collecting zope.interface>=5.1.0
  Downloading zope.interface-5.5.0-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.whl
    |██████████| 254 kB 63.3 MB/s
Collecting itemloaders>=1.0.1
  Downloading itemloaders-1.0.6-py3-none-any.whl (11 kB)
Collecting queuelib>=1.4.2
  Downloading queuelib-1.6.2-py2.py3-none-any.whl (13 kB)
Collecting service-identity>=18.1.0
  Downloading service_identity-21.1.0-py2.py3-none-any.whl (12 kB)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (from -)
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (from -)
Collecting PyDispatcher>=2.0.5
  Downloading PyDispatcher-2.0.6.tar.gz (38 kB)
Collecting w3lib>=1.17.0
  Downloading w3lib-2.0.1-py3-none-any.whl (20 kB)
Collecting Twisted>=18.9.0
  Downloading Twisted-22.10.0-py3-none-any.whl (3.1 MB)
    |██████████| 3.1 MB 43.8 MB/s
Collecting itemadapter>=0.1.0
  Downloading itemadapter-0.7.0-py3-none-any.whl (10 kB)
Requirement already satisfied: cffi>=1.12 in /usr/local/lib/python3.7/dist-packages (from -)
Requirement already satisfied: pycparser in /usr/local/lib/python3.7/dist-packages (from -)
Collecting jmespath>=0.9.5
  Downloading jmespath-1.0.1-py3-none-any.whl (20 kB)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from py)
Requirement already satisfied: pyasn1 in /usr/local/lib/python3.7/dist-packages (from py)
Requirement already satisfied: attrs>=19.1.0 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: pyasn1-modules in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: typing-extensions>=3.6.5 in /usr/local/lib/python3.7/d
Collecting hyperlink>=17.1.1
  Downloading hyperlink-21.0.0-py2.py3-none-any.whl (74 kB)
    |██████████| 74 kB 3.2 MB/s
Collecting Automat>=0.8.0
  Downloading Automat-22.10.0-py2.py3-none-any.whl (26 kB)
Collecting constantly>=15.1
  Downloading constantly-15.1.0-py2.py3-none-any.whl (7.9 kB)
Collecting incremental>=21.3.0
  Downloading incremental-22.10.0-py2.py3-none-any.whl (16 kB)
Requirement already satisfied: idna>=2.5 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/d
Requirement already satisfied: requests>=2.1.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: filelock>=3.0.8 in /usr/local/lib/python3.7/dist-packa
Collecting requests-file>=1.4
  Downloading requests_file-1.5.1-py2.py3-none-any.whl (3.7 kB)
Requirement already satisfied: urllib3!=1.25.0,!>=1.25.1,<1.26,>=1.21.1 in /usr/local/
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: idna>=2.5 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/d
Requirement already satisfied: requests>=2.1.0 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: filelock>=3.0.8 in /usr/local/lib/python3.7/dist-packa
Collecting requests->=1.4
  Downloading requests_file-1.5.1-py2.py3-none-any.whl (3.7 kB)
Requirement already satisfied: urllib3!=1.25.0,!>=1.25.1,<1.26,>=1.21.1 in /usr/local/
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pac
```

```
Building wheels for collected packages: PyDispatcher
  Building wheel for PyDispatcher (setup.py) ... done
    Created wheel for PyDispatcher: filename=PyDispatcher-2.0.6-py3-none-any.whl size=1
      Stored in directory: /root/.cache/pip/wheels/c9/d6/6a/de198d890277cde60ca3dbebe7ae5!
Successfully built PyDispatcher
Installing collected packages: w3lib, cssselect, zope.interface, requests-file, parse
Successfully installed Automat-22.10.0 PyDispatcher-2.0.6 Twisted-22.10.0 constantly-
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/p
Requirement already satisfied: tabulate in /usr/local/lib/python3.7/dist-packages (0.8.7)
ERROR: Invalid requirement: 'scrapyn!pip'
```

PREPROCESSING AMAZON DATASET

```
df_amazon = pd.read_csv("/content/drive/MyDrive/product.csv")
df_amazon.head()
```

Unnamed: 0	asin	avg.rating	avg.helpful.ratio	also_bought	also_viewed
0	7806397051	2.50	NaN	['B00KR26VFE', 'B00E7LQHZ0', 'B00BMW24TU', 'B00K67AQN8', 'B008GOR6O0', 'B00KVBZAXK', 'B008B2Z94Q', 'B00IIFVJZ4', 'B00FYUX3Z0', 'B00HWR1OZ4', 'B00EYZR3E8', 'B00IHQKRCA', 'B00KS4FDLO', 'B00GBTX1GO', 'B00FYVOG1Y', 'B009WJCEMM', 'B00ID5PRQQ', 'B00IIG7UQ0', 'B0068Y0EXK', 'B00JFLRCOW', 'B009NAE91E', 'B00B4YVU4G', 'B00K48JDWY', 'B006HCJXBM', 'B009VXY05S', 'B00I99OON8', 'B009WI3RZQ', 'B00IJ3LG94', 'B0067F8BBM', 'B002TPQPEE', 'B00IBVE79K', 'B00K05ZR08', 'B00K407VOO', 'B00GMDIKV0', 'B00KWSCOZO', 'B00EL1HZ34', 'B00HKA110G', 'B00KL53KGA', 'B00B8P5CTA', 'B00EYYW2HW', 'B00IKBGQAE', 'B00KHKXPHS', 'B00HPS4664', 'B00I9ZCV6E', 'B00EOFEEKF8', 'B00F16KHIA', 'B00L0R8KK4', 'B00G8F2RNE', 'B00942W51C', 'B00KXOLTD0', 'B00DAYGJWV', 'B004X1QY20', 'B00BLDA3J8']	['B008GOR6O0', 'B00EOFEEKF8', 'B00IIFVJZ4', 'B009C7RZNC', 'B00K05ZR08', 'B009NAE9A0', 'B00A66T112', 'B00C8R39TM', 'B009VYF0MY', 'B00K0683R6', 'B004M7WKKK', 'B00K0683OE', 'B008MU8TL2', 'B00HKA110G', 'B00K0683PS', 'B00BMW24TU', 'B0061KQEJVW', 'B00E7LQHZ0', 'B00KR26VFE', 'B00IHR9EVE', 'B00IHQKRCA', 'B00GWDKRG6', 'B00DFQOIJ', 'B00GWDZBHQ', 'B002QFGQAK', 'B007RKC27U', 'B00JE46HM8', 'B00IIG7UQ0', 'B009NAE91E', 'B008D5JRL2', 'B00IHR9EV4']

'B00HL9BG9C', 'B00CZ49RSK',
'B00GLM2FAO', 'B00J3HHJUA',
'B00JEGCI6U', 'B009WJCEMM',
'B00K7P74GQ', 'B004I3LU7C',
'B004W5RCNM', 'B00LE9YL0G',
'B00JK99BNE', 'B00EYZR3E8',
'B005TKZS76', 'B00FQ8OL6U',
'B00E57MCMS', 'B00L54D9NU',
'B001TK1I4M', 'B00GBTX1GO',
'B00CQ1DKZS', 'B00I9GTBD4',
'B00FORNVGO', 'B00IVHDGMS',
'B00GQU5YY0', 'B009P732GI',
'B00GZIHP3G', 'B008B2Z94Q',
'B00BX7DGIC', 'B004UO2GMW',
'B004GNMP0U', 'B00KHOF6V2',
'B00DQ2ILQY', 'B008D5D9NY',
'B00IZZ7RTO', 'B00IHQQGKM',
'B00CBT0TYU', 'B00AU05YVU',
'B008NQJA5E', 'B00HWR1OZ4',
'B0092LDWVW', 'B00K67AQN8',
'B00HQ8MA70', 'B002QFGKUQ',
'B003QAPTAW', 'B0067F8BBM',
'B004M7WKKK', 'B00FYVOG1Y']

[B0054GLD1U',
'B003BRZCUC',
'B0054GBXOW',
'B006VDOPPQ',
 'B005R3IPJI',
'B003BRZ7B6',
'B0074SE5CA',
'B000X5LUYC',
'B009HOU1TU',
'B000VQEQQXA',
 'B000CHGZS'

1	2	9759091062	3.09	NaN	'B000Y5IEQ10', 'B002TNIN2S', 'B0074RLEZC', 'B007CMS5C4', 'B0071H5C76', 'B007M8TB08', 'B0072CVZ0S', 'B002SKMNGO', 'B006YZ8S4', 'B006V6T1X4', 'B00DJWN0FI', 'B00CZ9RDG8', 'B000OV3ZDO', 'B00H04Z0BI', 'B0089MUOOS', 'B003BS4CUC', 'B00BKFNQI2', 'B000VQEUK4', 'B006ZBP8NM', 'B005JNKXF0', 'B002W23NQ2', 'B00J7RIXEM', 'B005IJTKFY', 'B00886V412', 'B0071H6ZD6', 'B005GMY88W', 'B004RANK5Q', 'B009JJA49U', 'B0071H640U', 'B002UFJ66M', 'B0093HPZ2E', 'B000MQ6X8A', 'B002W01P82', 'B004N56WWI', 'B0089PZZKS', 'B005IJQAIE', 'B00HXTD60W', 'B003Y5VLX8', 'B00CVDTMI0', 'B007R4ZJ8A', 'B000VXVEGK', 'B000O2P3RE', 'B0011SUGFE', 'B00FM7PGTQ', 'B00176YZUM', 'B00DH1O15O', 'B00BTCZJW2', 'B008GOR6O0', 'B005JECMQM', 'B000W3WP1W', 'B00CE2JC9W', 'B004TSFBNK', 'B0053L2E1O', 'B00CPQHKYG', 'B003COAFPQ', 'B005GMYPMG', 'B0013OULOM', 'B0054GBXOW', 'B0054GLD1U', 'B006VDOPPQ', 'B003BRZCUC', 'B009HOU1TU', 'B00GAXR43M', 'B0074RLEZC', 'B003BRZ7B6', 'B002TNIN2S', 'B002SKMNGO', 'B0072CVZ0S', 'B006Q4JNCU', 'B000VQEQXA', 'B00H04Z0BI', 'B00DJWN0FI', 'B003BS4CUC', 'B002TOLEMS', 'B0074SE5CA', 'B007R4ZJ8A', 'B000TBVO9G', 'B000NB3VOS', 'B005R3IPJI', 'B00BTCZJW2', 'B000X5LUYC', 'B005IJTKFY', 'B006YZ8S4', 'B008I7GTWA', 'B005DXHDGX', 'B001T8OZGC', 'B00DKB2CYI', 'B008BJ5GMO', 'B005EM6NYQ', 'B000VQEUK4', 'B000XQO85O', 'B00KLGFH1A', 'B008HCTX3S', 'B0071H5C76', 'B006ZBP8NM', 'B00J8WR3B0', 'B0002Z8UZ0', 'B006V6T1X4', 'B006ZYU5YQ', 'B00F12F5QQ', 'B007TO3BFQ', 'B0009R1526', 'B009FGHKNA', 'B000PKENQ1O1'
---	---	------------	------	-----	--

'B00GZ0WFQG',
 'B004H7GB9G',
 'B000MIR7TM',
 'B002EP9OUG',
 'B00K2T5RZ2',
 'B00F12F5QQ',
 'B00B78672O',
 'B00447F1YK',
 'B00174LGSI',
 'B008R7BT5S',
 'B0060G4IW0',
 'B00JH74XPA',
 'B00J20ES34',
 'B00BUU7KVG',
 'B003CU26SY',
 'B003GLR1E8',
 'B00D3KXYS4',
 'B005OZC2MU',
 'B00K0RKB2K',
 'B002BS43HU',
 'B006NMY0MS',
 'B003CTTZQQ',
 'B004TSFE28',
 'B000GD5LY6',
 'B002QFGKUQ',
 'B008BJ5GMO']

['B0072CSVB4',
 'B005YWBOHW',
 'B00CGOUL2A',
 'B00GIWPK2W',
 '9788072208',
 '9788072305',
 'B00C8RZP6C',
 'B00B1YH7PA',
 'B00HPYVCM4',
 'B008AZ0E80',
 'B00GFE0VX6',
 'B0079W1YUW',
 'B0079VUA5S',
 'B004OC8HQO',
 'B00B9KT9FW',
 'B00KTB1ZF4',
 'B00KSL75E0',
 'B002XQ1SQ4',
 'B003HUP0H8',
 'B008KS9VBS',
 'B006C5OHSI',
 'B00727PMYS',
 'B00BUUCRGO',
 'B00CMR3A6A',
 'B00DST09YM',
 'B006C5OHSI',
 'B006P14842',
 'B00KB93G06',
 'B00D8H2FNW',
 'B005S0S5X0']

```
[B00CGOUL2A',
 '9788072208'],
 [B00KSLWAGI',
 'B00CMQZCE4',
 'B004564IZI',
 'B007Q1ODRM',
 'B00CLYHF9W',
 'B001JQ6072',
 'B0072CYOAG',
 'B00L24H15A',
 'B00H00AT5E',
 'B008RTYOK8',
 'B0009OAIIG',
 'B006P14842',
 'B0079VWZ3S',
 'B000JL7Z5S',
 'B00JUCIN4O',
 'B007ZML10O',
 'B000P43JXQ',
 'B0079W22QM',
 'B0079VYMDY',
 'B0072CUKQS',
 'B00L6XS1II',
 'B000U66I0A',
 'B0079VZFSAs',
 'B00E9EU1LG',
 'B000CNTT9I',
 'B000C1Z3LS']
```

```
[B007P7OPQQ',
 'B0017JT658',
 'B0084HM1DA',
 'B000MX5NVQ',
 'B000AQF50Y',
 'B0013FPN1W',
 'B00AH5IOGA',
 'B0072CTK5K',
 'B000XE5260',
 'B00KSLJ5WA',
 'B0079VZFSAs',
 'B0072CVBZC',
 'B001RCIPMU',
 'B0072CS07A',
 'B0017LQNQ6',
 'B00JUCIN4O',
 'B00HZPKTY',
 'B007VXH2DM',
 'B004VUBNQU',
 'B009NPOY14',
 'B00CGOUL2A',
 'B00KCX9BK0',
 'B0072CQU2G',
 'B00KJ0Q34S',
 'B0022V2PKI',
 'B004B2J8UG',
 [B005M2AQRI',
 'B000VOHKK8',
 'B0017JT658',
 'B000KS1A7Y',
 'B000XE49GO']
```

3	4	9790790961	4.60	NaN
---	---	------------	------	-----

'B0072CYS9I',
 'B001362KHQ',
 'B001JQ6O72',
 'B000VOHKK8',
 'B0079VYY3M',
 'B0072CWP3O',
 'B0049CXC8W',
 'B000VOL914',
 'B000KRZD4Q',
 'B0072CUD8I',
 'B002CT01M4',
 'B009RA8HJ0',
 'B00BZBO7OI',
 'B0009N36NG',
 'B0091XL4B6',
 'B00152SA1S',
 'B002Z7FVTQ',
 'B00L290S28',
 'B000UICHT6U',
 'B002LZ0IPY',
 'B00023QRCK',
 'B00456E6N2',
 'B002LKZ1JW',
 'B0009N33Z2',
 'B000E7UFM2',
 'B00IPZE0XU',
 'B0079W3WIE',
 'B000C1ZDTU',
 'B000C2YXVI',
 'B00KC977OG',
 'B000E7WHLY',
 'B0072CXVTQ',
 'B005M2AQRI',
 'B0072CWRAA',
 'B000FCTITM',
 'B0079W1SDA',
 'B000E7UFS6',
 'B000C235XK',
 'B002LL4UO8',
 'B002LZ0I2W',
 'B0079VS8M0',
 'B0009N35D2',
 'B0009N5EZO',
 'B00EZ1RXGY',
 'B00FY3T3TM',
 'B0072CY7J4',
 'B0079W0FW0',
 'B0049OP9VS',
 'B00161IKD6',
 'B003VX1KF2',
 'B0072CTS2K',
 'B0079VXYZG',
 'B0007XGBVC',
 'B004P0UUBK',
 'B0079VTWBG',
 'B002LL308U',
 'B0072CWDVI']

```
-----,
'B0072CVHE2',
'B000VOPG0E',
'B000XE87IA',
'B003XLX046',
'B000JL7Z5S',
'B004564IZI',
'B00DAUYQX4',
'B0009N34GU',
'B0072CQJBS',
'B00EYBI67U',
'B0079FW5WA',
'B0079VTHJ8',
'B008HLP3WY',
```

```
df_amazon['avg.rating'].unique()
```

```
array([2.5 , 3.09, 5. , 4.6 , 4.5 , 4.29, 4.09, 4. , 4.4 , 3.33, 3.67,
       3.8 , 2. , 4.75, 3. , 4.57, 4.55, 4.22, 4.8 , 4.83, 4.33, 3.4 ,
       3.5 , 3.94, 3.63, 4.64, 4.62, 4.08, 4.63, 4.2 , 4.25, 4.67, 3.98,
       2.25, 4.65, 4.19, 3.55, 3.54, 3.88, 3.96, 4.04, 4.17, 4.12, 3.75,
       3.92, 3.95, 4.32, 3.93, 2.67, 3.83, 4.38, 3.56, 3.6 , 4.18, 1. ,
       3.81, 3.91, 4.23, 4.21, 4.86, 4.3 , 4.41, 3.25, 2.75, 4.35, 4.42,
       4.06, 3.86, 3.71, 4.71, 4.44, 4.36, 4.56, 3.57, 3.62, 3.2 , 4.43,
       3.45, 2.83, 1.5 , 1.67, 4.54, 4.7 , 4.9 , 3.7 , 3.89, 4.15, 4.27,
       4.39, 3.38, 3.14, 3.3 , 4.82, 3.73, 4.26, 3.29, 4.14, 4.95, 4.11,
       3.43, 2.33, 2.38, 3.79, 3.78, 3.69, 3.82, 3.27, 3.85, 4.73, 3.87,
       4.53, 4.69, 2.8 , 4.47, 3.12, 3.76, 4.79, 4.91, 4.46, 3.9 , 4.77,
       4.1 , 4.24, 4.74, 4.37, 3.36, 4.78, 3.42, 4.92, 4.07, 4.59, 2.56,
       4.58, 2.6 , 4.13, 3.77, 4.31, 3.53, 3.58, 4.51, 2.4 , 4.61, 4.88,
       4.45, 3.74, 4.16, 3.46, 3.18, 3.17, 4.34, 3.64, 4.05, 2.86, 4.93,
       4.68, 4.72, 3.65, 4.48, 4.02, 3.11, 2.71, 2.2 , 4.89, 4.81, 4.28,
       3.99, 3.22, 1.33, 4.87, 3.21, 3.06, 2.57, 2.88, 2.9 , 3.1 , 2.92,
       3.26, 3.48, 3.59, 4.01, 3.47, 3.72, 4.52, 4.85, 4.84, 1.75, 3.97,
       3.66, 4.03, 3.35, 4.66, 4.76, 2.91, 1.86, 3.41, 3.61, 3.84, 3.19,
       3.44, 1.83, 1.25, 1.6 , 2.89, 3.16, 2.73, 3.68, 3.04, 4.49, 3.23,
       3.51, 2.24, 3.39, 3.13, 4.97, 3.52, 4.96, 2.79, 4.94, 2.47, 4.98])
```

ДУУСЛУЧЖАУ, ДУУУ ВИРУБВВ,

```
df_amazon.drop(df_amazon[df_amazon['price'] == ''].index, inplace = True)
```

ДУУУУЛХВВА,

```
df_amazon.dtypes
```

Unnamed: 0	int64
asin	object
avg.rating	float64
avg.helpful.ratio	float64
also_bought	object
also_viewed	object
brand	object
categories	object
description	object
price	float64
salesRank	object
title	object
dtype:	object

```
df_amazon['avg.rating'].unique()
```

```
array([2.5 , 3.09, 5. , 4.6 , 4.5 , 4.29, 4.09, 4. , 4.4 , 3.33, 3.67,
       3.8 , 2. , 4.75, 3. , 4.57, 4.55, 4.22, 4.8 , 4.83, 4.33, 3.4 ,
       3.5 , 3.94, 3.63, 4.64, 4.62, 4.08, 4.63, 4.2 , 4.25, 4.67, 3.98,
       2.25, 4.65, 4.19, 3.55, 3.54, 3.88, 3.96, 4.04, 4.17, 4.12, 3.75,
       3.92, 3.95, 4.32, 3.93, 2.67, 3.83, 4.38, 3.56, 3.6 , 4.18, 1. ,
       3.81, 3.91, 4.23, 4.21, 4.86, 4.3 , 4.41, 3.25, 2.75, 4.35, 4.42,
       4.06, 3.86, 3.71, 4.71, 4.44, 4.36, 4.56, 3.57, 3.62, 3.2 , 4.43,
       3.45, 2.83, 1.5 , 1.67, 4.54, 4.7 , 4.9 , 3.7 , 3.89, 4.15, 4.27,
       4.39, 3.38, 3.14, 3.3 , 4.82, 3.73, 4.26, 3.29, 4.14, 4.95, 4.11,
       3.43, 2.33, 2.38, 3.79, 3.78, 3.69, 3.82, 3.27, 3.85, 4.73, 3.87,
       4.53, 4.69, 2.8 , 4.47, 3.12, 3.76, 4.79, 4.91, 4.46, 3.9 , 4.77,
       4.1 , 4.24, 4.74, 4.37, 3.36, 4.78, 3.42, 4.92, 4.07, 4.59, 2.56,
       4.58, 2.6 , 4.13, 3.77, 4.31, 3.53, 3.58, 4.51, 2.4 , 4.61, 4.88,
       4.45, 3.74, 4.16, 3.46, 3.18, 3.17, 4.34, 3.64, 4.05, 2.86, 4.93,
       4.68, 4.72, 3.65, 4.48, 4.02, 3.11, 2.71, 2.2 , 4.89, 4.81, 4.28,
       3.99, 3.22, 1.33, 4.87, 3.21, 3.06, 2.57, 2.88, 2.9 , 3.1 , 2.92,
       3.26, 3.48, 3.59, 4.01, 3.47, 3.72, 4.52, 4.85, 4.84, 1.75, 3.97,
       3.66, 4.03, 3.35, 4.66, 4.76, 2.91, 1.86, 3.41, 3.61, 3.84, 3.19,
       3.44, 1.83, 1.25, 1.6 , 2.89, 3.16, 2.73, 3.68, 3.04, 4.49, 3.23,
       3.51, 2.24, 3.39, 3.13, 4.97, 3.52, 4.96, 2.79, 4.94, 2.47, 4.98])
```

```
df_amazon[['avg.rating', 'price']] = df_amazon[['avg.rating', 'price']].astype(np.float)
df_amazon['SpecialPrice'] = df_amazon['price']
df_amazon['IsAvailable'] = True
df_amazon['Platform'] = CONST_PLATFORM_AMAZON
df_amazon.columns
```

```
Index(['Unnamed: 0', 'asin', 'avg.rating', 'avg.helpful.ratio', 'also_bought',
       'also_viewed', 'brand', 'categories', 'description', 'price',
       'salesRank', 'title', 'SpecialPrice', 'IsAvailable', 'Platform'],
      dtype='object')
```

```
df_amazon.head()
```

Unnamed: 0	asin	avg.rating	avg.helpful.ratio	also_bought	also_viewed
0	7806397051	2.50	NaN	["B00KR26VFE", 'B00E7LQHZ0', 'B00BMW24TU', 'B00K67AQN8', 'B008GOR6O0', 'B00KVBZAXK', 'B008B2Z94Q', 'B00IIFVJZ4', 'B00FYUX3Z0', 'B00HWR1OZ4', 'B00EYZR3E8', 'B00IHQKRCA', 'B00KS4FDLO', 'B00GBTX1GO', 'B00FYVOG1Y', 'B009WJCEMM', 'B00ID5PRQQ', 'B00IIG7UQ0', 'B0068Y0EXK', 'B00JFLRCOW', 'B009NAE91E', 'B00B4YVU4G', 'B00K48JDWY', 'B006HCJXBM', 'B009VXY05S', 'B00I99OON8', 'B009WI3RZQ', 'B00IJ3LG94', 'B0067F8BBM', 'B002TPQPEE', 'B00IBVE79K', 'B00K05ZR08', 'B00K407V0O', 'B00GMDIKV0', 'B00KWSKOZO', 'B00EL1HZ34', 'B00HKA110G', 'B00KL53KGA', 'B00B8P5CTA', 'B00EYYW2HW', 'B00IKBGQAE', 'B00KHKXPHS', 'B00HPS4664', 'B00I9ZCV6E', 'B00EOFKEF8', 'B00F16KHIA', 'B00L0R8KK4', 'B00G8F2RNE', 'B00942W51C', 'B00KXOLTD0', 'B00DAYGJWV', 'B004X1QY20', 'B00BLDA3J8'.]	["B008GOR6O0", 'B00EOFKEF8', 'B00IIFVJZ4', 'B009C7RZNC', 'B00K05ZR08', 'B009NAE9A0', 'B00A66T112', 'B00C8R39TM', 'B009VYF0MY', 'B00K0683R6', 'B004M7WKKK', 'B00K0683OE', 'B008MU8TL2', 'B00HKA110G', 'B00K0683PS', 'B00BMW24TU', 'B0061KQEJVW', 'B00E7LQHZ0', 'B00KR26VFE', 'B00IHR9EVE', 'B00IHQKRCA', 'B00GWDKRG6', 'B00DFQOIJ', 'B00GWDZBHQ', 'B002QFGQAK', 'B007RKC27U', 'B00JE46HM8', 'B00IIG7UQ0', 'B009NAE91E', 'B008D5JRL2', 'B00IHR9EV4'.]

'B00HL9BG9C', 'B00CZ49RSK',
 'B00GLM2FAO', 'B00J3HHJUA',
 'B00JEGCI6U', 'B009WJCEMM',
 'B00K7P74GQ', 'B004I3LU7C',
 'B004W5RCNM', 'B00LE9YL0G',
 'B00JK99BNE', 'B00EYZR3E8',
 'B005TKZS76', 'B00FQ8OL6U',
 'B00E57MCMS', 'B00L54D9NU',
 'B001TK1I4M', 'B00GBTX1GO',
 'B00CQ1DKZS', 'B00I9GTBD4',
 'B00FORNVGO', 'B00IVHDGMS',
 'B00GQU5YY0', 'B009P732GI',
 'B00GZIHP3G', 'B008B2Z94Q',
 'B00BX7DGIC', 'B004UO2GMW',
 'B004GNMP0U', 'B00KHOF6V2',
 'B00DQ2ILQY', 'B008D5D9NY',
 'B00IZZ7RTO', 'B00IHQQGKM',
 'B00CBT0TYU', 'B00AU05YVU',
 'B008NQJA5E', 'B00HWR1OZ4',
 'B0092LDWVW', 'B00K67AQN8',
 'B00HQ8MA70', 'B002QFGKUQ',
 'B003QAPTAW', 'B0067F8BBM',
 'B004M7WKKK', 'B00FYVOG1Y']
 'B00H1SW3RC',
 'B008FX7B9M',
 'B0061KQEJVW',
 'B00IIFOEIS',
 'B006GTKSHY',
 'B00D8U570W',
 'B00IQK5K4C',
 'B00IZ6KIE4',
 'B00FORNYMU',
 'B00LAYIFD4',
 'B00ACOFN0Q',
 'B00BZ4QDW4',
 'B005C8MMPQ',
 '1111306923',
 'B006Z5K3FG',
 'B0054SIB1S',
 'B002UBVZYC',
 'B00J921XM4',
 'B00CZ4W8UE',
 'B00EU89CKC',
 'B00J3IFAA0',
 'B005O30Q0M']

["B0054GLD1U",
 'B003BRZCUC',
 'B0054GBXOW',
 'B006VDOPPQ',
 'B005R3IPJI',
 'B003BRZ7B6',
 'B0074SE5CA',
 'B000X5LUYC',
 'B009HOU1TU',
 'B000VQEQA',
 'B000C1F0781'

1	2	9759091062	3.09	NaN	'B00HXTD60W', 'B003Y5VLX8', 'B00CVDTMI0', 'B007R4ZJ8A', 'B000VXVEGK', 'B000O2P3RE', 'B0011SUGFE', 'B00FM7PGTQ', 'B00176YZUM', 'B00DH1O15O', 'B00BTCZJW2', 'B008GOR6O0', 'B005JECMQM', 'B000W3WP1W', 'B00CE2JC9W', 'B004TSFBNK', 'B0053L2E1O', 'B00CPQHKYG', 'B003COAFPQ', 'B005GMYPMG', 'B0013OULOM', 'B0054GBXOW', 'B0054GLD1U', 'B006VDOPPQ', 'B003BRZCUC', 'B009HOU1TU', 'B00GAXR43M', 'B0074RLEZC', 'B003BRZ7B6', 'B002TNIN2S', 'B002SKMNGO', 'B0072CVZ0S', 'B006Q4JNCU', 'B000VQEQXA', 'B00H04Z0BI', 'B00DJWN0FI', 'B003BS4CUC', 'B00BKFNQI2', 'B000VQEUK4', 'B006ZBP8NM', 'B005JNKXF0', 'B002W23NQ2', 'B00J7RIXEM', 'B005IJTKFY', 'B00886V412', 'B0071H6ZD6', 'B005GMY88W', 'B004RANK5Q', 'B009JJA49U', 'B0071H640U', 'B002UFJ66M', 'B0093HPZ2E', 'B000MQ6X8A', 'B002W01P82', 'B004N56WWI', 'B0089PZZKS', 'B005IJQAIE', 'B008I7GTWA', 'B005DXHDGX', 'B001T8OZGC', 'B00DKB2CYI', 'B008BJ5GMO', 'B005EM6NYQ', 'B000VQEUK4', 'B000XQO85O', 'B00KLGFH1A', 'B008HCTX3S', 'B0071H5C76', 'B006ZBP8NM', 'B00J8WR3B0', 'B0002Z8UZ0', 'B006V6T1X4', 'B006ZYU5YQ', 'B00F12F5QQ', 'B007TO3BFQ', 'B0009R1526', 'B009FGHKNA', 'B000PKENQI0'
---	---	------------	------	-----	--

'B00GZ0WFQG',	'B000OV3ZDO',
'B004H7GB9G',	'B007M8TB08',
'B000MIR7TM',	'B001E919LU',
'B002EP9OUG',	'B0072CUBRQ',
'B00K2T5RZ2',	'B0054GG4AA']
'B00F12F5QQ',	
'B00B78672O',	
'B00447F1YK',	
'B00174LGSI',	
'B008R7BT5S',	
'B0060G4IW0',	
'B00JH74XPA',	
'B00J20ES34',	
'B00BUU7KVG',	
'B003CU26SY',	
'B003GLR1E8',	
'B00D3KXYS4',	
'B005OZC2MU',	
'B00K0RKB2K',	
'B002BS43HU',	
'B006NMY0MS',	
'B003CTTZQQ',	
'B004TSFE28',	
'B000GD5LY6',	
'B002QFGKUQ',	
'B008BJ5GMO'	

[`B0072CSVB4',
 `B005YWBOHW',
 `B00CGOUL2A',
 `B00GIWPK2W',
 `9788072208',
 `9788072305',
 `B00C8RZP6C',
 `B00B1YH7PA',
 `B00HPYVCM4',
 `B008AZ0E80',
 `B00GFE0VX6',
 `B0079W1YUW',
 `B0079VUA5S',
 `B004OC8HQO',
 `B00B9KT9FW',
 `B00KTB1ZF4',
 `B00KSL75E0',
 `B002XQ1SQ4',
 `B003HUP0H8',
 `B008KS9VBS',
 `B006C5OHSI',
 `B00727PMYS',
 `B00BUUCRGO',
 `B00CMR3A6A',
 `B00DST09YM',
 `B00KB93G06',
 `B00D8H2FNW',
 `B005SO55X0']

```
[B00CGOUL2A',
 '9788072208'],
 [B00KSLWAGI',
 'B00CMQZCE4',
 'B004564IZI',
 'B007Q1ODRM',
 'B00CLYHF9W',
 'B001JQ6072',
 'B0072CYOAG',
 'B00L24H15A',
 'B00H00AT5E',
 'B008RTYOK8',
 'B0009OAIIG',
 'B006P14842',
 'B0079VWZ3S',
 'B000JL7Z5S',
 'B00JUCIN4O',
 'B007ZML10O',
 'B000P43JXQ',
 'B0079W22QM',
 'B0079VYMDY',
 'B0072CUKQS',
 'B00L6XS1II',
 'B000U66I0A',
 'B0079VZFSAs',
 'B00E9EU1LG',
 'B000CNTT9I',
 'B000C1Z3LS']
```

```
[B007P7OPQQ',
 'B0017JT658',
 'B0084HM1DA',
 'B000MX5NVQ',
 'B000AQF50Y',
 'B0013FPN1W',
 'B00AH5IOGA',
 'B0072CTK5K',
 'B000XE5260',
 'B00KSLJ5WA',
 'B0079VZFSAs',
 'B0072CVBZC',
 'B001RCIPMU',
 'B0072CS07A',
 'B0017LQNQ6',
 'B00JUCIN4O',
 'B00HZPKTY',
 'B007VXH2DM',
 'B004VUBNQU',
 'B009NPOY14',
 'B00CGOUL2A',
 'B00KCX9BK0',
 'B0072CQU2G',
 'B00KJ0Q34S',
 'B0022V2PKI',
 'B004B2J8UG',
 [B005M2AQRI',
 'B000VOHKK8',
 'B0017JT658',
 'B000KS1A7Y',
 'B000XE49GO']
```

3	4	9790790961	4.60	NaN
---	---	------------	------	-----

'B0072CYS9I',
 'B001362KHQ',
 'B001JQ6O72',
 'B000VOHKK8',
 'B0079VYY3M',
 'B0072CWP3O',
 'B0049CXC8W',
 'B000VOL914',
 'B000KRZD4Q',
 'B0072CUD8I',
 'B002CT01M4',
 'B009RA8HJ0',
 'B00BZBO7OI',
 'B0009N36NG',
 'B0091XL4B6',
 'B00152SA1S',
 'B002Z7FVTQ',
 'B00L290S28',
 'B000UICHT6U',
 'B002LZ0IPY',
 'B00023QRCK',
 'B00456E6N2',
 'B002LKZ1JW',
 'B0009N33Z2',
 'B000E7UFM2',
 'B00IPZE0XU',
 'B0079W3WIE',
 'B000C1ZDTU',
 'B000C2YXVI',
 'B00KC977OG',
 'B000E7WHLY',
 'B0072CXVTQ',
 'B005M2AQRI',
 'B0072CWRAA',
 'B000FCTITM',
 'B0079W1SDA',
 'B000E7UFS6',
 'B000C235XK',
 'B002LL4UO8',
 'B002LZ0I2W',
 'B0079VS8M0',
 'B0009N35D2',
 'B0009N5EZO',
 'B00EZ1RXGY',
 'B00FY3T3TM',
 'B0072CY7J4',
 'B0079W0FW0',
 'B0049OP9VS',
 'B00161IKD6',
 'B003VX1KF2',
 'B0072CTS2K',
 'B0079VXYZG',
 'B0007XGBVC',
 'B004P0UUBK',
 'B0079VTWBG',
 'B002LL308U',
 'B0072CWDVI']

					'B0072CVHE2', 'B000VOPG0E', 'B000XE87IA', 'B003XLX046', 'B000JL7Z5S', 'B004564IZI', 'B00DAUYQX4', 'B0009N34GU', 'B0072CQJBS', 'B00EYBI67U', 'B0079FW5WA', 'B0079VTHJ8', 'B008HLP3WY', 'B00HUAOE2S', 'B000OP7LTY', 'B00F14IKQQ', 'B00CXCCI8A']
					['B000E7YM8K', 'B0019M21OQ', 'B0006V31FY', 'B0072CU2PM', 'B000BIDC0Q', 'B000GHWPI2', 'B007P7R3OW', 'B001366JMI', 'B005FGHPM0', 'B00GII2H8Q', 'B002JISTS6', 'B001D1ISOK', 'B0079VVTAS', ['B0019M21OQ', 'B000E7YM8K', 'B0006V31FY', 'B008D5JK9Q', 'B007P7R3OW']]
4	5	9790794231	4.50	NaN	'B0079W1VJQ', 'B00CITNW92', 'B004C7SNBU', 'B000JC51IU', 'B007VTRDBW', 'B004V457OK', 'B002LZ2I9S', 'B000ZKHV7A', 'B0072C116V2'

```
df_amazon.isna().any()
```

Unnamed: 0	False
asin	False
avg.rating	False
avg.helpful.ratio	True
also_bought	False
also_viewed	False
brand	True
categories	False
description	True
price	True
salesRank	False
title	False
SpecialPrice	True

```
IsAvailable      False
Platform        False
dtype: bool
```

```
df=df_amazon.dropna(axis=0)
```

```
df
```

Unnamed: 0	asin	avg.rating	avg.helpful.ratio	also_bought	also_v
21	22	B000052Y25	5.00	1.00	['B0087GR95S', 'B0088W2E60', 'B0088W7PUK', 'B0085XVQZW', 'B005J398JW', 'B00DNE5YZQ', 'B000G017TC', 'B000XDCGOM', 'B0001EKQRW', 'B006FWZ5CA', 'B0033ZIQVO', 'B0001EKUNC', 'B006UFNH22', 'B005U5KFTG', 'B000LNWUIG', 'B004PCRSYK', 'B003MUSRGY', 'B00DNE664E', 'B004LT9EDU', 'B003MUQOR8', 'B0006FMK98', 'B0000Y3GF2', 'B006IC0LXK', 'B00E8J36FU', 'B003MUNI0E', 'B0088W93LY', 'B0052ZKKWG', 'B000GCW9F6', 'B0033GLYQM', 'B008NB43UQ', 'B0016AVIQS', 'B0028R4SRO', 'B00DE2XKWQ', 'B0088W7Z7S', 'B00HHZALDC', 'B000BI4DDQ', 'B00GO5PEX8', 'B0087GRCL4', 'B00AFCP9FO', 'B002J29FKI', 'B005MT7I5O', 'B0088W7WX0', 'B000ZJ2842', 'B008M8TYU4', 'B00606RORK', 'B004B88BQW', 'B00BARE3SC', 'B002XO4LWE', 'B0064FTVZ0', 'B00008O2X5', 'B003U5XH70', 'B008OSIAFC', 'B003XWG880']

					'B002C74898', 'B00FEMSAWO', 'B00A9VX0IY', 'B00ICBW1QK', 'B004J0094Y', 'B0088W8DLA', 'B00BJ8RM7G', 'B0088W8NC4', 'B0088W48NM', 'B00KGGMR4A', 'B00LD8OE3C', 'B001FA1NTG', 'B00JIRRTYG', 'B00J7PLJYA', 'B0016P9Z9A']	'B0016B9', 'B00J2E9', 'B0088W9', 'B0088W9', 'B001FOL', 'B00GO5I', 'B00EE1I', 'B003U5I', 'B0016B9'
22	23	B000052Y5F	3.00	1.00	['B001HSM', 'B0044NANVK', 'B001HSMOEE', 'B000V7VZ3I', 'B000GG850U', 'B008AZQX4Y', 'B00BZF95R8', 'B008B0K0CO', 'B00006K13S', 'B001HSMPAM', 'B00CAWCDBU', 'B003ZTRR6I', 'B000F34TCC', 'B008KNX7QS', 'B000RZTMUW', 'B001LFEUYK', 'B00I79YSUY', 'B000WCPFKQ', 'B001V9QS6O', 'B00839QBOE', 'B001GAOHTI', 'B000UZED0S', 'B008D5KHIY', 'B000GCI8ZG', 'B00269WST4', 'B005WDS5W0', 'B0009XFYSG', 'B002ZG8H8O', 'B00HMCUOD2', 'B000QYNNQ8', 'B004AI9CM0', 'B00260H4HO', 'B0047UT5DW', 'B003840WFS', 'B0057D7ZBM', 'B005C55HPG', 'B001KYVL08', 'B000OT9UNK', 'B0002UDK4Q', 'B000TV59EC', 'B000K2VE1S']	 'B0044NA', 'B008AZC', 'B003ZTT', 'B001V9C', 'B00K1D', 'B00CAWC', 'B001HSM', 'B008B0I', 'B008KNP', 'B00JLNG', 'B00BZF', 'B00006', 'B0016A', 'B0002U', 'B000V7', 'B00D64', 'B008BH4', 'B000AMF', 'B004SEI', 'B0013OV', 'B00013Y', 'B001EF', 'B0077S', 'B00DYSX', 'B002RWU', 'B000GG', 'B000GW', 'B001G7', 'B00392I', 'B008B0M', 'B001G7C', 'B000GC', 'B001UQ', 'B0040IW', 'B0091FA', 'B000OT', 'B00HMCL', 'B000N7I', 'B000GC', 'B00121I'

					DVUVKZ1F1S , 'B000KTB19K', 'B0072CRMHI', 'B005D0DTS2', 'B000VPTHP8', 'B000YHLS8W', 'B000V1LX0O', 'B000GGKQMU', 'B005Y4T5C6', 'B000RMFGGY', 'B002GU5YCQ', 'B001OOLF82', 'B00028N8LI', 'B00BI8OG2Q', 'B00EDR6YB0', 'B001OK1YUA', 'B000EEH80M', 'B00ICBW1QK'] ['B00B04DBL', 'B000W3I', 'B000WII', 'B002BX7', 'B005CY', 'B00I13', 'B000R3E', 'B005CY', 'B0002FC', 'B0037L', 'B00A0I', 'B00H4P', 'B00DTW', 'B00201', 'B000VPI', 'B004X8, ['B0002FCCXG', 'B002BX74QM', 'B000R36HKC', 'B000W30RP8', 'B005CYKEL4', 'B009PIR36W', 'B004DBU9MG', 'B00A0IWTDI', 'B002UQHYEC', 'B0020100FM', 'B00CLVJ03O', 'B00A0IWTDI', 'B001RDH4G2', 'B005CYJI1Q', 'B00I13T8EC', 'B008BWIQM8', 'B003CP12O8', 'B000KTB19K', 'B00B04205I'] 'B0013IX', 'B009PI', 'B00CHL', 'B0098QI', 'B00GTB', 'B000MJ', 'B00JNXC', 'B009PIF', 'B00A0IW', 'B0088W8', 'B00B42I'
69	70	B0000530LV	5.00	1.00	

					'B00A0IV 'B00GCGI 'B00AZJC 'B00JME 'B000EOI 'B001EIY 'B005CYI 'B00KA6C 'B009084 'B007LG 'B000NPW 'B00B22I 'B004VGS 'B000052 'B00D3KX 'B0072C
71	72	B0000530NA	5.00	1.00	[{'B002UUT39M', 'B004CR3 'B0088W7Z00', 'B00DGCI 'B00J6V р KHJS', 'B0088W 'B004CR3D4W', 'B00E4MI 'B00D5ETV08', 'B005M3 'B007Y5F1PI', 'B00CLB 'B0076PEAX0', 'B002UU 'B0088W8RMU', 'B00I7JE 'B001F51TR2', 'B00J7PA 'B00JHLLQ6U', 'B00HDQ0 'B008X8R9NM', 'B00GNU9 'B0088W46W0', 'B00J6VI 'B000052YQ0', 'B005M2C 'B0012J30PU', 'B000T9C 'B003LZ6KZA', 'B000052 'B00JJRLG7G', 'B005PML 'B00HDQ0NVK', 'B007M8 'B007939EOE', 'B00DDS0 'B0088W8RF2', 'B001GG 'B005PMLVTW', 'B000052 'B000052YQ2', 'B0076PI 'B001F51TNQ', 'B007M8 'B004969H0U', 'B00IBD 'B0088W49TU', 'B0009RF 'B0012J135O', 'B00J7P 'B0081NLR0K', 'B0088W 'B00496819S', 'B0081NI 'B000F65JSC', 'B008X8F 'B009GEUPAG', 'B00JSL 'B007PBX496', 'B00HDQ0 'B00CSKSD90', 'B003LZ 'B001JQLNMI', 'B00BT7 'B007FYEC80', 'B0011D1 'B001F51TOU', 'B00D5E 'B002UUT3YM', 'B002DPI 'B0088W2E4W', 'B009GEL 'B00HWVQBSU', 'B0047EF 'B004CASKMO', 'B0016BF 'B004AS6GGU', 'B00D1XV 'B0033WSTXM'. 'B001W8V

'B00496DJBS',	'B001F5
'B005M31I4C',	'B008W
'B007UHHPZ4',	'B008W8
'B001F51TDG',	'B000F6
'B00HJRN9NW',	'B003KW2
'B009CUYUVY',	'B002K8
'B00DQSVXLI',	'B008QUC
'B00CIUMLLG',	'B00543
'B0028K3J7Q',	'B0088W
'B002Z7EOOY',	'B0088W
'B00IOXKUG4',	'B0055FF
'B00JGPND4K',	'B00CSKS
'B00JJDIH9A',	'B0034LI
'B00CSKSH6O',	'B0088W
'B00D5Z5OBM',	'B005ZIK
'B000T9CBQS',	'B00IN36
'B002F8F546',	'B009GE
'B005ZIC6Q8',	'B00JHLL
'B005CYJ79E']	'B0072C1

['B00D8FK9F0',	'B00D8F
'B002JDUMFO',	'B0075H
'B0075HE6B0',	'B0055S
'B00LDY8TU0',	'B00CRHF
'B006WQC0BI',	'B0014ZW
'B006WQC022',	'B00DGQV8
'B0072CV0UI',	'B00C9EK
'B0055S4BZ2',	'B001DO3
'B0047F71SS',	'B008CT
'B0055S4CRY',	'B0047F
'B00C9EKOHY',	'B0089MI
'B00AYWN8OY',	'B002PBHQP4
'B000NP7K1E',	'B00DGCEUMI'
'B004YWRUY',	'B0014ZWHZQ'
'B005H1525A',	'B001EIMH78E'
'B0014ZWHZQ',	'B005H1541C'
'B00EIMH78E',	'B0015571ZO'
'B005H1541C',	'B003A8EKTG'
'B0015571ZO',	'B00CRHR0BG'
'B003A8EKTG',	'B001EUS8G8'
'B00CRHR0BG',	'B0047F9820'
'B001EUS8G8',	'B002KDWHHA'
'B0047F9820',	'B00ECU0A18'
'B002KDWHHA',	'B001DO77YY'
'B00ECU0A18',	'B008EKY8VK'
'B001DO77YY',	'B0047F70B6'
'B008EKY8VK',	'B000C93O4M6'
'B0047F70B6',	'B008CT12G2'
'B000C93O4M6',	'B009X62WM6'
'B008CT12G2',	'B0073ANH14'
'B009X62WM6',	'B00DG1B6V2'
'B0073ANH14',	'B00JHLLQ6U'
'B00DG1B6V2',	'B000D0016U'
'B00JHLLQ6U',	'B00LDY8TU0'
'B000D0016U',	'B00CSKSH6O'

84

85

B00005333I

3.50

1.00

'B0005EU
 'B003RF
 'B00ECU
 'B000F6:
 'B0009R
 'B0047F
 'B00B1I'
 'B000NP'
 'B00H3R7
 'B00496F
 'B00IGL/
 'B0047F
 'B00CSK:
 'B00EIMI
 'B007M8:
 'B00E96:
 'B008FZ
 'B005H1
 'B0055S4
 'B00GP2N
 'B00CSKS
 'B001DPM
 'B005H1
 'B0073AI
 'B00GM3
 'B00CYZF
 'B002JL:
 'B006WQO
 'B004VM:
 'B00479:
 'B00BFBN
 'B00DGCI
 'B000KOM08Q',
 'B00027DMI8',
 'B00CSKSD90',
 'B0055170UC',
 'B0013NHSYE',
 'B0088W2WEO',
 'B00J8UCO5C',
 'B00HF6SJYV',
 'B000092OX4']

...

...

...

...

...

...

['B00IU2C148',
 'B00IZQ66V8',
 'B00JJOEV9Y',
 'B00KVO2JWC',
 'B00DCF7QW0',
 'B00L8FQZ3M',
 'B00L8HA5L8',
 'B00L8FDIZU',
 'B00LSXYC6G',
 'B00KEU5K1A',
 'B00L3RCOI0',
 'B00LTM6S40',
 'B00JQ12U0M',
 ...]

					'B00LDGV15I', 'B00L3GF2BW', 'B00KVE7R0Q', 'B00JZRX9SE', 'B00KVWJ5GC', 'B00LGCLOX8', 'B00L3MG3S2', 'B00KFUH00I', 'B00M04HZIO', 'B00LGVUUA2', 'B00KWP8ZIM', 'B00LGZDW3U', 'B00KH6F6TM', 'B00L49RXLA', 'B00KSLCU72', 'B00L4A4QXM', 'B00L2NX8YE', 'B00KCRYIWM', 'B00K4S1L1K', 'B00LOWR484', 'B00LOQBO1S', 'B00L3N9J46', 'B00KVQELFI', 'B00KSHYIM6', 'B00K1GQW5K', 'B00K1H6AY2', 'B00JZVMTOU', 'B00KHADRZS', 'B00KGEN2KK', 'B00JV1GLTI', 'B00KYB12ZM', 'B00L1F79LQ', 'B00LORXVUE', 'B00LG63DOM', 'B00JNV6CEK', 'B00L5L7ZPQ', 'B00L1H4I0Y', 'B00JQL8XLC', 'B00JQL8PU6', 'B00KEXR5IS', 'B00L5KTZ0K', 'B00JLXNJY6', 'B00JLL8DOE', 'B00KCZO2WK', 'B00L3GRNBO', 'B00JVVZS56', 'B00JLWFQHU', 'B00JOUZOTY', 'B00KC7I2GU', 'B00K6XM8KQ', 'B00JQKCW9C', 'B00H3TT74G', 'B00G1PRRRRC', 'B00LGMWDX0', 'B00KB5XOGG', 'B00JPHP3B0', 'B00KB8FGSM'.	['B00DCF7 'B00KVO2 'B00IZQ 'B00IU2 'B00GOQI 'B00H3T 'B00L1H 'B00IO2I 'B00F5E 'B00JQ12 'B00IO2I 'B00BT0C 'B00L2N 'B00L3R 'B002VP 'B00BMLZ 'B00I9QD 'B00KVC 'B00KSTZ 'B00KK6I 'B00BT7C 'B00JDJI 'B009XRM 'B00B1A 'B0092KJ' 'B003VJ 'B00DE8I 'B00B1AI 'B00GOQF 'B00DRM 'B00II3I 'B00IZT> 'B00JJAY 'B00H9A4 'B00JD9W 'B00E4C 'B00GOQF 'B00JUA/ 'B004NH 'B00ING1 'B0035IC 'B00LMI4 'B00JTHL 'B0094F' 'B001NP! 'B00K9WZ 'B00JCS. 'B00JEJ/ 'B00L3J0 'B00FRF -----
11280	11281	B00JT74V1O	4.84	0.98		

[...]
'B00KE7T7FS',
'B00KSFQN3K',
'B00J4X7KBG',
'B00LB18EKK',
'B00KCJPPXG',
'B00IOUDVYA',
'B00IIBB7A0',
'B00L3LB0IG',
'B00L3LBUGI',
'B00L3K91OW',
'B00L5LS07E',
'B00JPJE6CU',
'B00JQL8WCC',
'B00I2OR6YE',
'B00IMWA4RC',
'B002OSICSI',
'B00H4IOT5I',
'B00K37OI6C',
'B00I9KKAHQ',
'B00K32ICRS',
'B00K8ATI7S',
'B00KIX796W',
'B00KXE9RWU']

[
'B00KXG8JHC',
'B00KQBR9FM',
'B00KZNOHTW',
'B00KQVTX06',
'B00KQASCDG',
'B00KHGIK54',
'B00I1ISDZM',
'B00KQ4PEBU',
'B00K6SM5N6',
'B00L07J5BW',
'B00KHH2VOY',
'B00KD73PBQ',
'B00GF1NZLY',
'B00KCFAZTE',
'B00LALMG3W',
'B00KCTER3U',
'B00KY4Z1I8',
'B00KAL5JAU',
'B00HZ7IRT2',
'B00GXBEV0A',
'B00JL2TURM',
'B00JEKYNZA',
'B00K7PO996',
'B00KVO2JWC',
'B00KNJOLSA',
'B00I2495QG',
'B00JJVG6HC',
'B00K7ILPYU',
'B00JYY56ZW',
'B00IT1HKV4',
'B00G5WO2R4',
'B00FX1QCQC',
...]

11284 11285 B00JYKBZ3I 5.00 0.89

'B00IRQNH1K',
 'B00IT69F62',
 'B00JPV3MIM',
 'B00GGSDWH8', ['B00JBjh
 'B00K2O3NV2', 'B00KXGj
 'B00E3OXGQY', 'B00KCTE
 'B00FJQFJX8', 'B00700
 'B00G75X158', 'B00K58H
 'B00FFANXMM', 'B00GYF
 'B00C8YCFU4', 'B00GCQM
 'B00ITA41T4', 'B0036B
 'B00I5NX0VA', 'B0069Y
 'B00COCNZ86', 'B00KAL
 'B00H0LVCVI', 'B00GFBW
 'B00DXNXG2U', 'B00C7D'
 'B00HJD8NLY', 'B0083FI
 'B00JX979CQ', 'B004NML
 'B00GLE8SNK', 'B00IZQ
 'B00HHR0SYW', 'B00GZ0W
 'B00HMTK8OU', 'B00CME
 'B00GCQMSVA',
 'B00FPLTWJE',
 'B00GFBWWWC',
 'B00GN8XLUE',
 'B00FBTSLY',
 'B00IS4CPPI',
 'B00GYF65TK',
 'B00COCREVU',
 'B00IW110B2',
 'B00IDWP4IA',
 'B00IMHN0B4',
 'B00K2N9PQU',
 'B00GNU6MFI',
 'B009PARMMA',
 'B006K4O66E',
 'B00HARZKVU',
 'B00IP42FBA',
 'B00ELAP3X4',
 'B00HBR41LY',
 'B00GXE29EM',
 'B00H93NJLS',
 'B00H8RVMC8',
 'B00GDV55B4',
 'B00G3ZHOLY',
 'B00C7DYBX0',
 'B00FT8ZHHE',
 'B00IBS9QC6',
 'B00GWFJAD0',
 'B00CJSVBOA',
 'B00FFR59VS',
 'B00BBSPMIU',
 'B00CNVBCLK',
 'B00GLDL9HS',
 'B000F6RFX4',
 'B00DDT116M']

['B00JZ0:

11286	11287	B00JZ0L7VC	4.36	0.97
--------------	-------	------------	------	------

'B00GOJM
 'B00JZSD
 'B008O4Y
 'B00BIYF
 'B00GSO'
 'B00BQG
 'B00KDL
 'B00JSJF
 'B00DI9Y
 'B00K14'
 'B00KCV
 'B00JY4F
 'B009S9E
 'B00LZYV
 'B00H8ZI
 'B008KJI
 'B0031ZI
 'B00F50I
 'B00JZKJ
 'B009S3F
 'B00JS3I
 'B0072CL
 'B008MQC
 'B00K4Q
 'B00CGN!
 'B00DLA'
 'B00FIZM!
 'B00JSU'
 'B00K9S:
 'B00GI1'
 'B00ANTC
 'B006FLL'
 'B00DLAX
 'B00IXQI
 'B00L9F
 'B00005
 'B007XPB
 'B00A1O7
 'B00JND4
 'B00E8LL
 'B001T8I
 'B00AR03
 'B001T8
 'B00K8LI
 'B00HNXV
 'B0001MK
 'B0007IF
 'B00GNEF
 'B00HDZI
 'B00GQQL
 'B00JG6L
 'B00JQL3
 'B0039P3
 'B00CGRM
 'B003XW1
 'B005QTE
 'B00E8Z

'B00AWCM

[`B00IAM
`B00K1GC
`B00JZVM
`B00CWJL
`B00KTA
`B00IBSS
`B00JH1L
`B00L3L
`B00L3K9
`B006ZA
`B004LJ
`B00KEX
`B009A2
`B00JQ12
`B000EPA
`B007004
`B004GE
`B00HJB
`B00KH6F
`B006I9
`B00ID22
`B001CL
`B00ADXI
`B001CM
`B00J9SU
`B0057GI
`B008R1F
`B00DTV
`B00JC7I
`B00KHH2
`B00168I
`B004OTV
`B00JL2T
`B00KCF

11289 11290 B00K1H6AY2 5.00 0.81

'B00KCJPPXG',	'B00H3VC
'B00KE7T7FS',	'B00ICI
'B00JQ12U0M',	'B0072C
'B00LGVWMWX0',	'B00EU3C
'B00KTUWSSI',	'B001MA7
'B00KCRYIWM',	'B00KHV
'B00L5KTZ0K',	'B002CSF
'B00JQL8WCC',	'B001GIC
'B00L2NX8YE',	'B00BVG
'B00L2F5KP2',	'B007G5C
'B00KXE9RWU',	'B008LL
'B00KFNJY6S',	'B009NP
'B00KGEM2KK',	'B001G7I
'B00KVQELFI',	'B00EE7S
'B00KSHYIM6',	'B00I4ZE
'B00KLZO2JE',	'B005M3
'B00KHV6TIE',	'B00JIRF
'B00KZ0RPB2',	'B00005
'B00KHADRZS',	'B0089MU
'B00L1PP5LW',	'B0047PT
'B00KTAJAIY',	
'B00KYB12ZM',	
'B00KYC0DQU',	
'B00IIBB7A0',	
'B00IU079YM',	
'B00K6IGTI8',	
'B00L5LSO7E',	
'B00LGW2XFG',	
'B00K6XM8KQ',	
'B00KGHKJSA',	
'B00JT74V1O',	
'B00KI2RQHU',	
'B00L2M52DK',	
'B00K8J1JKI',	
'B00J6YIISC',	
'B00K32ICRS',	
'B00KK62714',	
'B00CA8636K',	
'B00HPKFMSI',	
'B00KDN5YAA',	
'B00KLT6O3M'	

[B00GP18
'B00G2
'B00KN9
'B001E1
'B00FJQ
'B001HT'
'B00JB5C
'B00LAP1
'B00K37
'B000JE'
'B004MNF
'B001TU'
'B003GSK
'B0015N.'
'B000CC

```
'B00KQBRYFIV',
'B00KCTER3U',
'B00KQASCDG',
'B00JEKYNZA',
'R00KD73PRO'
```

```
'B004BAF',
'B00DW2O',
'B00K1',
'B005BD7'
```

```
df.shape
```

```
(720, 15)
```

```
'B00KCFAZTE'.
```

```
D000VVV.
```

```
df.isnull().sum()
```

Unnamed:	0
asin	0
avg.rating	0
avg.helpful.ratio	0
also_bought	0
also_viewed	0
brand	0
categories	0
description	0
price	0
salesRank	0
title	0
SpecialPrice	0
IsAvailable	0
Platform	0
dtype:	int64

PREPROCESSING BIGBASKET DATASET

```
'R00F00D39I'
```

```
D00070
```

```
'B00HHF0HF4'.
```

```
D00070
```

```
df_bigbasket =pd.read_csv("/content/drive/MyDrive/BigBasket Products.csv")
df_bigbasket.head()
```

index	product	category	sub_category	brand	sale_price	market_price
0	1	Garlic Oil - Vegetarian Capsule 500 mg	Beauty & Hygiene	Hair Care	Sri Sri Ayurveda	220.0
1	2	Water Bottle - Orange	Kitchen, Garden & Pets	Storage & Accessories	Mastercook	180.0

```
df_bigbasket['rating'].unique()
```

```
array([4.1, 2.3, 3.4, 3.7, 4.4, 3.3, 3.6, 4. , 3.5, 4.3, 2.2, 5. , 3.8,
       4.2, 4.9, 3.9, 1.5, 2.9, 2.6, 3. , 4.5, nan, 1. , 4.6, 4.8, 3.1,
       2. , 1.8, 2.8, 3.2, 4.7, 2.5, 2.4, 2.7, 2.1, 1.7, 1.9, 1.2, 1.3,
       1.4, 1.6])
```

```
df_bigbasket['Category'] = CONST_PRODUCT_BEVERAGES
df_bigbasket['Platform'] = CONST_PLATFORM_BIGBASKET
```

```
df_bigbasket.columns
```

```
Index(['index', 'product', 'category', 'sub_category', 'brand', 'sale_price',
       'market_price', 'type', 'rating', 'description', 'Category',
       'Platform'],
      dtype='object')
```

```
df_bigbasket.dtypes
```

```
index          int64
product        object
category       object
sub_category   object
brand          object
sale_price     float64
market_price   float64
type           object
rating         float64
description    object
Category       object
Platform       object
dtype: object
```

```
df_bigbasket[df_bigbasket['rating'] != CONST_NOT_APPLICABLE].head()
```

index	product	category	sub_category	brand	sale_price	market_price
0	1 Garlic Oil - Vegetarian Capsule 500 mg	Beauty & Hygiene	Hair Care	Sri Sri Ayurveda	220.0	220.0
1	2 Water Bottle - Orange	Kitchen, Garden & Pets	Storage & Accessories	Mastercook	180.0	180.0

PREPROCESSING FLIPKART DATASET

- Electronics, Household

```
df_flipkart = pd.read_csv("/content/drive/MyDrive/flipkart_com-ecommerce_sample.csv")
df_flipkart.head()
```

	uniq_id	crawl_timestamp	product_url
0	c2d766ca982eca8304150849735ffef9	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-cycling-shorts/p/itmeh2ffvzetthbb?pid=SRTEH2FF9KEDEFGF

1 7f7036a6d550aaa89d34c77bd39a5e48 2016-03-25 http://www.flipkart.com/fabhomedecor-fabric-double-sofa-bed/p/itmeh3qgfamccfp?pid=SBEEH3QGU7MFYJFY
22:59:23 +0000

2 f449ec65dc0c041b6ae5e6a32717d01b 2016-03-25
22:59:23 +0000 http://www.flipkart.com/aw-
bellies/p/itmeh4grgfbkexnt?
pid=SHOEH4GRSUBJGZXE

3 0973b37acd0c664e3de26e97e5571454 2016-03-25
22:59:23 +0000 http://www.flipkart.com/alisha-solid-
women-s-cycling-
shorts/p/itmeh2f6sdgah2pq?
pid=SRTEH2F6HUZMQ6SJ

4	bc940ea42ee6bef5ac7cea3fb5cfbee7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sicons-all-purpose-arnica-dog-shampoo/p/itmeh3zyw2vhgsp5?pid=PSOEH3ZYDMSYARJ5
---	----------------------------------	------------------------------	---

```
df_flipkart.shape
```

```
(20000, 15)
```

```
df_flipkart['Platform'] = CONST_PLATFORM_FLIPKART
```

```
df_flipkart.columns
```

```
Index(['uniq_id', 'crawl_timestamp', 'product_url', 'product_name',
       'product_category_tree', 'pid', 'retail_price', 'discounted_price',
       'image', 'is_FK_Advantage_product', 'description', 'product_rating',
       'overall_rating', 'brand', 'product_specifications', 'Platform'],
      dtype='object')
```

▼ COMBINING ALL THE PLATFORM DATA TOGETHER

```
df_all_platforms = pd.concat([df_amazon, df_bigbasket, df_flipkart], axis=0)
df_all_platforms.columns
```

```
Index(['Unnamed: 0', 'asin', 'avg.rating', 'avg.helpful.ratio', 'also_bought',
       'also_viewed', 'brand', 'categories', 'description', 'price',
       'salesRank', 'title', 'SpecialPrice', 'IsAvailable', 'Platform',
       'index', 'product', 'category', 'sub_category', 'sale_price',
       'market_price', 'type', 'rating', 'Category', 'uniq_id',
       'crawl_timestamp', 'product_url', 'product_name',
       'product_category_tree', 'pid', 'retail_price', 'discounted_price',
       'image', 'is_FK_Advantage_product', 'product_rating', 'overall_rating',
       'product_specifications'],
      dtype='object')
```

```
df_all_platforms.dtypes
```

Unnamed: 0	float64
------------	---------

```

asin                      object
avg.rating                 float64
avg.helpful.ratio          float64
also_bought                object
also_viewed                object
brand                      object
categories                 object
description                object
price                      float64
salesRank                  object
title                      object
SpecialPrice               float64
IsAvailable                object
Platform                   object
index                      float64
product                     object
category                    object
sub_category                object
sale_price                  float64
market_price                float64
type                        object
rating                      float64
Category                    object
uniq_id                     object
crawl_timestamp             object
product_url                 object
product_name                object
product_category_tree       object
pid                         object
retail_price                float64
discounted_price            float64
image                        object
is_FK_Advantage_product    object
product_rating               object
overall_rating               object
product_specifications      object
dtype: object

```

```

df_all_platforms['SpecialPrice'] = df_all_platforms['SpecialPrice'].replace(',', '', regex=True)
df_all_platforms['avg.rating'] = df_all_platforms['avg.rating'].replace(CONST_NOT_APPLICABLE, np.nan)

```

```

df_all_platforms['rating'] = df_all_platforms['rating'].replace(CONST_NOT_APPLICABLE, 0, regex=True)
df_all_platforms[['rating']] = df_all_platforms[['rating']].astype(np.float)
df_all_platforms[['avg.rating']] = df_all_platforms[['avg.rating']].astype(np.float)
df_all_platforms.dtypes

```

```

Unnamed: 0                  float64
asin                        object
avg.rating                   float64
avg.helpful.ratio            float64
also_bought                  object
also_viewed                  object
brand                       object

```

```
categories          object
description        object
price              float64
salesRank          object
title              object
SpecialPrice       float64
IsAvailable        object
Platform           object
index              float64
product            object
category           object
sub_category       object
sale_price         float64
market_price       float64
type               object
rating             float64
Category           object
uniq_id            object
crawl_timestamp   object
product_url        object
product_name       object
product_category_tree object
pid                object
retail_price       float64
discounted_price   float64
image              object
is_FK_Advantage_product object
product_rating     object
overall_rating     object
product_specifications object
dtype: object
```

```
df_all_platforms.to_csv('df_all_platforms.csv', index=False)
df_all_platforms[df_all_platforms['Platform'] == CONST_PLATFORM_AMAZON ].head()
```

Unnamed: 0	asin	avg.rating	avg.helpful.ratio		also_bought	also_viewed
					['B00KR26VFE', 'B00E7LQHZ0', 'B00BMW24TU', 'B00K67AQN8', 'B008GOR6O0', 'B00KVBZAXK', 'B008B2Z94Q', 'B00IIFVJZ4', 'B00FYUX3Z0', 'B00HWR1OZ4', 'B00EYZR3E8', 'B00IHQKRCA', 'B00KS4FDLO', 'B00GBTX1GO', 'B00FYVOG1Y', 'B009WJCEMM', 'B00ID5PRQQ', 'B00IIG7UQ0', 'B0068Y0EXK', 'B00JFLRCOW', 'B009NAE91E', 'B00B4YVU4G', 'B00K48JDWY', 'B006HCJXBM', 'B009VXY05S', 'B00I99OON8', 'B009WI3RZQ', 'B00IJ3LG94', 'B0067F8BBM', 'B002TPQPEE', 'B00IBVE79K', 'B00K05ZR08', 'B00K407VOO', 'B00GMDIKV0', 'B00KWSOCZO', 'B00EL1HZ34', 'B00HKA110G', 'B00KL53KGA', 'B00B8P5CTA', 'B00EYYW2HW', 'B00IKBGQAE', 'B00KHKXPHS', 'B00HPS4664', 'B00I9ZCV6E', 'B00EOFEEKF8', 'B00F16KHIA', 'B00L0R8KK4', 'B00G8F2RNE', 'B00942W51C', 'B00KXOLTD0', 'B00DAYGJWV', 'B004X1QY20', 'B00BLDA3J8']	['B008GOR6O0', 'B00EOFEEKF8', 'B00IIFVJZ4', 'B009C7RZNC', 'B00K05ZR08', 'B009NAE9A0', 'B00A66T112', 'B00C8R39TM', 'B009VYF0MY', 'B00K0683R6', 'B004M7WKKK', 'B00K0683OE', 'B008MU8TL2', 'B00HKA110G', 'B00K0683PS', 'B00BMW24TU', 'B0061KQEJVW', 'B00E7LQHZ0', 'B00KR26VFE', 'B00IHR9EVE', 'B00IHQKRCA', 'B00GWDKRG6', 'B00DFQOIJ', 'B00GWDZBHQ', 'B002QFGQAK', 'B007RKC27U', 'B00JE46HM8', 'B00IIG7UQ0', 'B009NAE91E', 'B008D5JRL2', 'B00IHR9EV4']
0	1.0	7806397051	2.50	NaN		

'B00HL9BG9C', 'B00CZ49RSK',
'B00GLM2FAO', 'B00J3HHJUA',
'B00JEGCI6U', 'B009WJCEMM',
'B00K7P74GQ', 'B004I3LU7C',
'B004W5RCNM', 'B00LE9YL0G',
'B00JK99BNE', 'B00EYZR3E8',
'B005TKZS76', 'B00FQ8OL6U',
'B00E57MCMS', 'B00L54D9NU',
'B001TK1I4M', 'B00GBTX1GO',
'B00CQ1DKZS', 'B00I9GTBD4',
'B00FORNVGO', 'B00IVHDGMS',
'B00GQU5YY0', 'B009P732GI',
'B00GZIHP3G', 'B008B2Z94Q',
'B00BX7DGIC', 'B004UO2GMW',
'B004GNMP0U', 'B00KHOF6V2',
'B00DQ2ILQY', 'B008D5D9NY',
'B00IZZ7RTO', 'B00IHQQGKM',
'B00CBT0TYU', 'B00AU05YVU',
'B008NQJA5E', 'B00HWR1OZ4',
'B0092LDWVW', 'B00K67AQN8',
'B00HQ8MA70', 'B002QFGKUQ',
'B003QAPTAW', 'B0067F8BBM',
'B004M7WKKK', 'B00FYVOG1Y']
'B00H1SW3RC',
'B008FX7B9M',
'B0061KQEVIEW',
 'B00IIFOEIS',
'B006GTKSHY',
'B00D8U570W',
'B00IQK5K4C',
 'B00IZ6KIE4',
'B00FORNYMU',
 'B00LAYIFD4',
'B00ACOFN0Q',
'B00BZ4QDW4',
'B005C8MMPQ',
 '1111306923',
'B006Z5K3FG',
 'B0054SIB1S',
'B002UBVZYC',
'B00J921XM4',
'B00CZ4W8UE',
'B00EU89CKC',
 'B00J3IFAA0',
'B005O30Q0M']

['B0054GLD1U',
'B003BRZCUC',
'B0054GBXOW',
'B006VDOPPQ',
 'B005R3IPJI',
'B003BRZ7B6',
'B0074SE5CA',
'B000X5LUYC',
'B009HOU1TU',
'B000VQEQQXA',
 'B000CHF07G'

1	2.0	9759091062	3.09	NaN
---	-----	------------	------	-----

'B000Y5IEQV0',
 'B002TNIN2S',
 'B0074RLEZC',
 'B007CMS5C4',
 'B0071H5C76',
 'B007M8TB08',
 'B0072CVZ0S',
 'B002SKMNGO',
 'B006YZ8S4',
 'B006V6T1X4',
 'B00DJWN0FI',
 'B00CZ9RDG8',
 'B000OV3ZDO',
 'B00H04Z0BI',
 'B0089MUOOS',
 'B003BS4CUC',
 'B00BKFNQI2',
 'B000VQEUK4',
 'B006ZBP8NM',
 'B005JNKXF0',
 'B002W23NQ2',
 'B00J7RIXEM',
 'B005IJTKFY',
 'B00886V412',
 'B0071H6ZD6',
 'B005GMY88W',
 'B004RANK5Q',
 'B009JJA49U',
 'B0071H640U',
 'B002UFJ66M',
 'B0093HPZ2E',
 'B000MQ6X8A',
 'B002W01P82',
 'B004N56WWI',
 'B0089PZZKS',
 'B005IJQAIE',
 'B00HXTD60W',
 'B003Y5VLX8',
 'B00CVDTMI0',
 'B007R4ZJ8A',
 'B000VXVEGK',
 'B000O2P3RE',
 'B0011SUGFE',
 'B00FM7PGTQ',
 'B00176YZUM',
 'B00DH1O15O',
 'B00BTCZJW2',
 'B008GOR6O0',
 'B005JECMQM',
 'B000W3WP1W',
 'B00CE2JC9W',
 'B004TSFBNK',
 'B0053L2E1O',
 'B00CPQHKYG',
 'B003COAFPQ',
 'B005GMYPMG',
 'B0013OULOM',
 '['B0054GBXOW',
 'B0054GLD1U',
 'B006VDOPPQ',
 'B003BRZCUC',
 'B009HOU1TU',
 'B00GAXR43M',
 'B0074RLEZC',
 'B003BRZ7B6',
 'B002TNIN2S',
 'B002SKMNGO',
 'B0072CVZ0S',
 'B006Q4JNCU',
 'B000VQEQXA',
 'B00H04Z0BI',
 'B00DJWN0FI',
 'B003BS4CUC',
 'B002TOLEMS',
 'B0074SE5CA',
 'B007R4ZJ8A',
 'B000TBVO9G',
 'B000NB3VOS',
 'B005R3IPJI',
 'B00BTCZJW2',
 'B000X5LUYC',
 'B005IJTKFY',
 'B006YZ8S4',
 'B008I7GTWA',
 'B005DXHDGX',
 'B001T8OZGC',
 'B00DKB2CYI',
 'B008BJ5GMO',
 'B005EM6NYQ',
 'B000VQEUK4',
 'B000XQO85O',
 'B00KLGFH1A',
 'B008HCTX3S',
 'B0071H5C76',
 'B006ZBP8NM',
 'B00J8WR3B0',
 'B0002Z8UZ0',
 'B006V6T1X4',
 'B006ZYU5YQ',
 'B00F12F5QQ',
 'B007TO3BFQ',
 'B0009R1526',
 'B009FGHKNA',
 'B000PKENQI01'

'B00GZ0WFQG',
 'B004H7GB9G',
 'B000MIR7TM',
 'B002EP9OUG',
 'B00K2T5RZ2',
 'B00F12F5QQ',
 'B00B78672O',
 'B00447F1YK',
 'B00174LGSI',
 'B008R7BT5S',
 'B0060G4IW0',
 'B00JH74XPA',
 'B00J20ES34',
 'B00BUU7KVG',
 'B003CU26SY',
 'B003GLR1E8',
 'B00D3KXYS4',
 'B005OZC2MU',
 'B00K0RKB2K',
 'B002BS43HU',
 'B006NMY0MS',
 'B003CTTZQQ',
 'B004TSFE28',
 'B000GD5LY6',
 'B002QFGKUQ',
 'B008BJ5GMO']

['B0072CSVB4',
 'B005YWBOHW',
 'B00CGOUL2A',
 'B00GIWPK2W',
 '9788072208',
 '9788072305',
 'B00C8RZP6C',
 'B00B1YH7PA',
 'B00HPYVCM4',
 'B008AZ0E80',
 'B00GFE0VX6',
 'B0079W1YUW',
 'B0079VUA5S',
 'B004OC8HQO',
 'B00B9KT9FW',
 'B00KTB1ZF4',
 'B00KSL75E0',
 'B002XQ1SQ4',
 'B003HUP0H8',
 'B008KS9VBS',
 'B006C5OHSI',
 'B00727PMYS',
 'B00BUUCRGO',
 'B00CMR3A6A',
 'B00DST09YM',
 'B006C5OHSI',
 'B006P14842',
 'B00KB93G06',
 'B00D8H2FNW',
 'B005S0S5X0']

```
[B00CGOUL2A',
 '9788072208'],
 [B00KSLWAGI',
 'B00CMQZCE4',
 'B004564IZI',
 'B007Q1ODRM',
 'B00CLYHF9W',
 'B001JQ6072',
 'B0072CYOAG',
 'B00L24H15A',
 'B00H00AT5E',
 'B008RTYOK8',
 'B0009OAIIG',
 'B006P14842',
 'B0079VWZ3S',
 'B000JL7Z5S',
 'B00JUCIN4O',
 'B007ZML10O',
 'B000P43JXQ',
 'B0079W22QM',
 'B0079VYMDY',
 'B0072CUKQS',
 'B00L6XS1II',
 'B000U66I0A',
 'B0079VZFSAs',
 'B00E9EU1LG',
 'B000CNTT9I',
 'B000C1Z3LS']
```

```
[B007P7OPQQ',
 'B0017JT658',
 'B0084HM1DA',
 'B000MX5NVQ',
 'B000AQF50Y',
 'B0013FPN1W',
 'B00AH5IOGA',
 'B0072CTK5K',
 'B000XE5260',
 'B00KSLJ5WA',
 'B0079VZFSAs',
 'B0072CVBZC',
 'B001RCIPMU',
 'B0072CS07A',
 'B0017LQNQ6',
 'B00JUCIN4O',
 'B00HZPKTY',
 'B007VXH2DM',
 'B004VUBNQU',
 'B009NPOY14',
 'B00CGOUL2A',
 'B00KCX9BK0',
 'B0072CQU2G',
 'B00KJ0Q34S',
 'B0022V2PKI',
 'B004B2J8UG',
 [B005M2AQRI',
 'B000VOHKK8',
 'B0017JT658',
 'B000KS1A7Y',
 'B000XE49GO']
```

3	4.0	9790790961	4.60	NaN
---	-----	------------	------	-----

'B0072CYS9I',
 'B001362KHQ',
 'B001JQ6O72',
 'B000VOHKK8',
 'B0079VYY3M',
 'B0072CWP3O',
 'B0049CXC8W',
 'B000VOL914',
 'B000KRZD4Q',
 'B0072CUD8I',
 'B002CT01M4',
 'B009RA8HJ0',
 'B00BZBO7OI',
 'B0009N36NG',
 'B0091XL4B6',
 'B00152SA1S',
 'B002Z7FVTQ',
 'B00L290S28',
 'B000UICHT6U',
 'B002LZ0IPY',
 'B00023QRCK',
 'B00456E6N2',
 'B002LKZ1JW',
 'B0009N33Z2',
 'B000E7UFM2',
 'B00IPZE0XU',
 'B0079W3WIE',
 'B000C1ZDTU',
 'B000C2YXVI',
 'B00KC977OG',
 'B000E7WHLY',
 'B0072CXVTQ',
 'B005M2AQRI',
 'B0072CWRAA',
 'B000FCTITM',
 'B0079W1SDA',
 'B000E7UFS6',
 'B000C235XK',
 'B002LL4UO8',
 'B002LZ0I2W',
 'B0079VS8M0',
 'B0009N35D2',
 'B0009N5EZO',
 'B00EZ1RXGY',
 'B00FY3T3TM',
 'B0072CY7J4',
 'B0079W0FW0',
 'B0049OP9VS',
 'B00161IKD6',
 'B003VX1KF2',
 'B0072CTS2K',
 'B0079VXYZG',
 'B0007XGBVC',
 'B004P0UUBK',
 'B0079VTWBG',
 'B002LL308U',
 'B0072CWDVI']

					'B0072CVHE2', 'B000VOPG0E', 'B000XE87IA', 'B003XLX046', 'B000JL7Z5S', 'B004564IZI', 'B00DAUYQX4', 'B0009N34GU', 'B0072CQJBS', 'B00EYBI67U', 'B0079FW5WA', 'B0079VTHJ8', 'B008HLP3WY', 'B00HUAOE2S', 'B000OP7LTY', 'B00F14IKQQ', 'B00CXCC18A']
4	5.0	9790794231	4.50	NaN	[['B000E7YM8K', 'B0019M21OQ', 'B0006V31FY', 'B0072CU2PM', 'B000BIDC0Q', 'B000GHWPI2', 'B007P7R3OW', 'B001366JMI', 'B005FGHPM0', 'B00GII2H8Q', 'B002JISTS6', 'B001D1ISOK', 'B0079VVTAS', ['B0019M21OQ', 'B000E7YM8K', 'B0006V31FY', 'B0079W1VJQ', 'B00CITNW92', 'B008D5JK9Q', 'B007P7R3OW']]

```
df_all_platforms[df_all_platforms[ 'Platform' ] == CONST_PLATFORM_BIGBASKET ].head()
```

Unnamed:
0 asin avg.rating avg.helpful.ratio also_bought also_viewed brand

0 NaN NaN NaN NaN NaN NaN NaN Sri Sri
Ayurveda

1 NaN NaN NaN NaN NaN NaN NaN Mastercook

2 NaN NaN NaN NaN NaN NaN NaN Trm

3	NaN	NaN	NaN	NaN	NaN	NaN	Nakoda
---	-----	-----	-----	-----	-----	-----	--------

```
4      NaN  NaN      NaN      NaN      NaN      NaN      Nivea
```

```
df_all_platforms[df_all_platforms['Platform'] == CONST_PLATFORM_FLIPKART ].head()
```

Unnamed:
0 asin avg.rating avg.helpful.ratio also_bought also_viewed
brar

0 NaN NaN NaN NaN NaN NaN Alist

1	NaN	NaN	NaN	NaN	NaN	NaN	FabHomeDecor
---	-----	-----	-----	-----	-----	-----	--------------

2	NaN	NaN	NaN	NaN	NaN	NaN	A1
---	-----	-----	-----	-----	-----	-----	----

3	NaN	NaN	NaN	NaN	NaN	NaN	Alist
---	-----	-----	-----	-----	-----	-----	-------

DATA ANALYSIS ON INTEGRATED DATA

```
df_all_platforms.shape
```

```
(58901, 37)
```

```
df_all_platforms.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 58901 entries, 0 to 19999
Data columns (total 37 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Unnamed: 0        11346 non-null   float64
 1   asin              11346 non-null   object  
 2   avg.rating        11346 non-null   float64
 3   avg.helpful.ratio 792 non-null    float64
 4   also_bought       11346 non-null   object  
 5   also_viewed       11346 non-null   object  
 6   brand              53020 non-null   object  
 7   categories         11346 non-null   object  
 8   description        58102 non-null   object  
 9   price              10941 non-null   float64
 10  salesRank         11346 non-null   object  
 11  title              11346 non-null   object  
 12  SpecialPrice      10941 non-null   float64
 13  IsAvailable       11346 non-null   object  
 14  Platform            58901 non-null   object  
 15  index              27555 non-null   float64
 16  product             27554 non-null   object  
 17  category            27555 non-null   object  
 18  sub_category        27555 non-null   object  
 19  sale_price          27555 non-null   float64
 20  market_price        27555 non-null   float64
 21  type                27555 non-null   object  
 22  rating              18929 non-null   float64
 23  Category            27555 non-null   object  
 24  uniq_id             20000 non-null   object  
 25  crawl_timestamp     20000 non-null   object  
 26  product_url          20000 non-null   object  
 27  product_name         20000 non-null   object  
 28  product_category_tree 20000 non-null   object  
 29  pid                  20000 non-null   object  
 30  retail_price         19922 non-null   float64
 31  discounted_price     19922 non-null   float64
 32  image                19997 non-null   object  
 33  is_FK_Advantage_product 20000 non-null   object  
 34  product_rating        20000 non-null   object  
 35  overall_rating        20000 non-null   object  
 36  product_specifications 19986 non-null   object
```

```
dtypes: float64(11), object(26)
memory usage: 17.1+ MB
```

```
df_all_platforms.describe()
```

	Unnamed: 0	avg.rating	avg.helpful.ratio	price	SpecialPrice	i
count	11346.000000	11346.000000	792.000000	10941.000000	10941.000000	27555.0
mean	5673.500000	4.163377	0.736503	16.617354	16.617354	13778.0
std	3275.452411	0.770957	0.319704	19.305909	19.305909	7954.5
min	1.000000	1.000000	0.000000	0.010000	0.010000	1.0
25%	2837.250000	3.800000	0.500000	6.680000	6.680000	6889.5
50%	5673.500000	4.330000	0.860000	11.590000	11.590000	13778.0
75%	8509.750000	4.750000	1.000000	20.000000	20.000000	20666.5
max	11346.000000	5.000000	1.000000	499.000000	499.000000	27555.0



```
df_all_platforms.sample(10)
```

Unnamed: 0	asin	avg.rating	avg.helpful.ratio	also_bought	also_vi
---------------	------	------------	-------------------	-------------	---------

2729	NaN	NaN	NaN	NaN	NaN
-------------	-----	-----	-----	-----	-----

15335	NaN	NaN	NaN	NaN	NaN
--------------	-----	-----	-----	-----	-----

14373	NaN	NaN	NaN	NaN	NaN
-------	-----	-----	-----	-----	-----

2184	NaN	NaN	NaN	NaN	NaN
------	-----	-----	-----	-----	-----

['B0029PT4KG',
 'B004O4SIYS',
 'B0000635WI',
 'B000JVA80C',
 'B005N7YWX6',
 'B002DPUY3S',
'B000NDCHNW',
 'B00BH0RZ78',
 'B000OSAH8I',
 'B005AYE8UE',
'B0083VODME',
 '0679885722',
 '0307336182',
 '0761148574',
'B008N06RAG',
 'B0002E7DIQ',
 '1440505365',
 'B004LU29TK',
'B00HPS5EK6',
 'B00519BD0G',
 'B003D6TSC4',
 '1889392448',
'B00HHGPKZK',
 'B003D7E2XI',
'B000KJZOWU', ['B000JVA
'B00HSME26Q', 'B0029PT4
 'B000I615Q4', 'B004O4SIYS
 'B004BH15K2', 'B005JFX',
 'B0032AM8HI', 'B001IATZ
 'B00820HS1E', 'B004PK2'
'B00DSR6EBG', 'B002DPU'
 'B008X7JBPM', 'B00DUH0'
 '1561487171', 'B006L1O9
'B001VC995S', 'B008N06F
 '0738217670', 'B00DLRN5
 '0789210770', 'B000JVCF
 '0743251547', 'B00J9E8M
'B001X4Z1N8', 'B009Q5Q
'B0016BCS9S', 'B000NDCH
 'B0000041EV', 'B00BH0R
 'B00112RH9I', 'B001X4Z
 '1599214547', 'B00HPS5
'B00AQ4WJM2', 'B002K4L0
 'B000EILS0Y', 'B004Y9D1
 '1449418465', 'B005N7YV
'B001IATZFM', 'B00L9A9
 '1594745730'. 'B00J4GRI

							'B001A0M3LI',	'B002JUL
							'B00HHGTM4K',	'B00I1ATY'
10670	10671.0	B00DM14TYC	3.92	NaN			'B000I8WNT0',	'B00K22SC'
							'B000OZI8HS',	'B0016BCI'
							'B007JYZW08',	'B004PK2'
							'B009LA4VOG',	'B00JT3S'
							'0830755772',	'B00112F'
							'141652472X',	'B0088W8'
							'B0051NTHJG',	'B000063'
							'B0083QLONK',	'B00HUHJM'
							'B0000536EK',	'B002B46I'
							'B00FRS2C2E',	'B0037W6'
							'B0062CW1BG',	'B00E36RN'
							'0770436471',	'B000OSA'
							'0761133267',	'B000OZIK'
							'B000RIWARG',	'B00519BI'
							'B002B46H6Y',	'B004W7XC'
							'B000EM8308',	'B004EDYI'
							'B000I5UASY',	'B0019QZC'
							'B00C3EXTDQ',	'B004PK0'
							'B00ICNAN5Y',	'B00EMBS2'
							'B00HHGTMNG',	'B006YTXC'
							'B003D76KPQ',	'B004P6B'
							'B007KLQJV6',	'B005MLH2'
							'B00DVKJXFE',	'B001UHSV'
							'B0079PY5DM',	'B002PK0I'
							'B009Q5Q126',	'B000PR5I'
							'B007YSB2L2',	'B007HBF2'
							'B00DJP2T8E',	'B00GSGS'
							'B0097F3YES',	
							'B00HBLEJH6',	
							'B00CDKT3G2',	
							'B003DGZRNS',	
							'B009FEO7M4',	
							'B009AVCADK',	
							'B002K6F7CS',	
							'B00I9A4W7K',	
							'B0050KTWK4',	
							'0399166262',	
							'B00AIVN87I',	
							'B004ZH2UE8',	
							'B002XNK2VO',	
							'B009RPH0VQ',	
							'B00JXJAJAA',	
							'B001W2HZ1C',	
							'B00537128O',	
							'B003IUGHSI',	
							'B005DFD1EE',	
							'B00474CUTO',	
							'B008AUKXIG',	
							'B00DUH0T4C',	
							'B000V1BDIG',	
							'B004BH503U']	

19026	NaN	NaN	NaN	NaN	NaN
--------------	-----	-----	-----	-----	-----

5958	NaN	NaN	NaN	NaN	NaN
-------------	-----	-----	-----	-----	-----

10186

NaN

NaN

NaN

NaN

NaN

12855

NaN

NaN

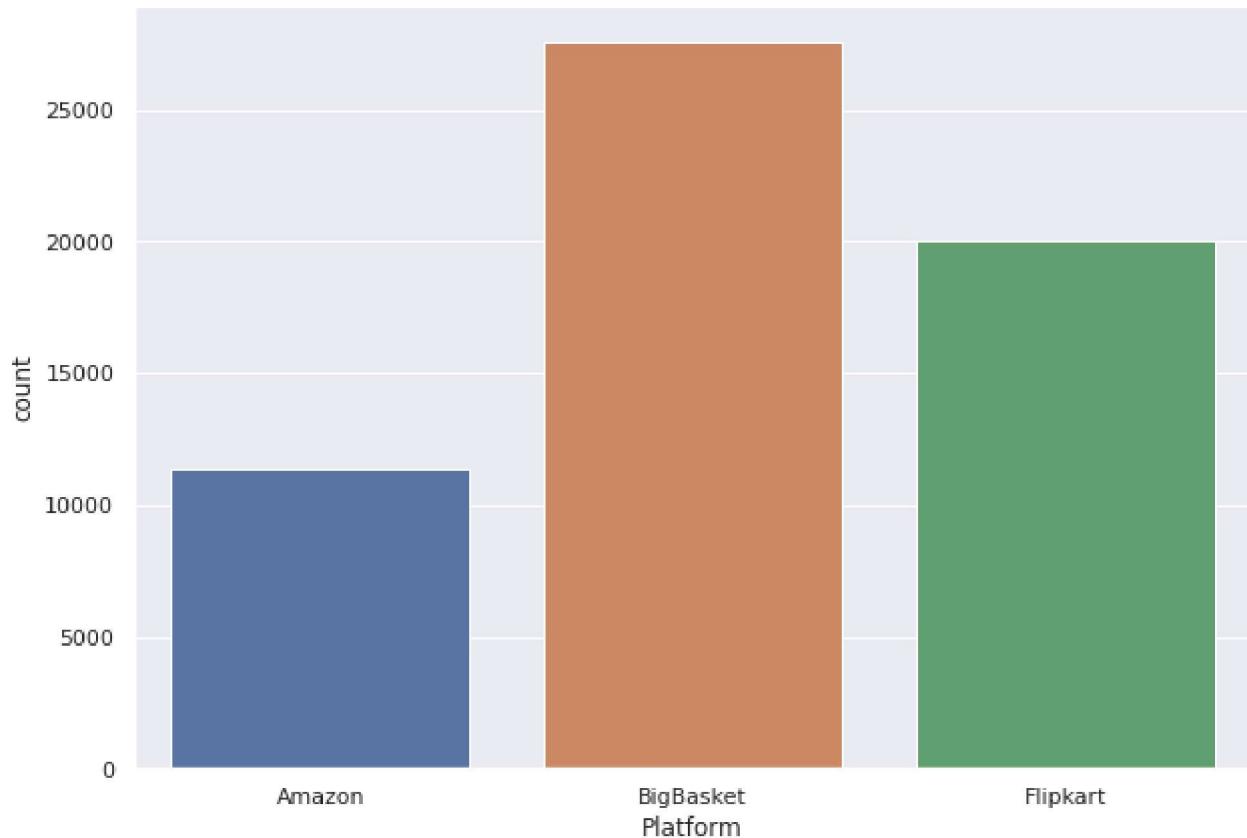
NaN

NaN

NaN

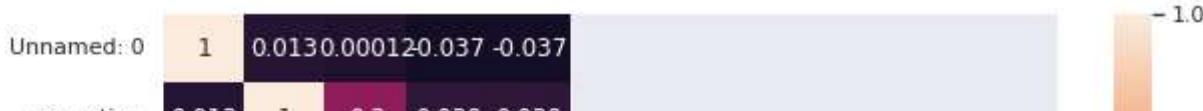
VISUALIZATION ON DATASET

```
fig = plt.figure(figsize=(10,7))
sns.set()
ax = sns.countplot(x="Platform", data=df_all_platforms)
```



```
corr = df_all_platforms.corr()
fig = plt.figure(figsize=(10,7))
sns.heatmap(corr, xticklabels=corr.columns, yticklabels=corr.columns, annot=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f470f041e10>
```



```
df_all_platforms.isna().any()
```

Unnamed: 0	True
asin	True
avg.rating	True
avg.helpful.ratio	True
also_bought	True
also_viewed	True
brand	True
categories	True
description	True
price	True
salesRank	True
title	True
SpecialPrice	True
IsAvailable	True
Platform	False
index	True
product	True
category	True
sub_category	True
sale_price	True
market_price	True
type	True
rating	True
Category	True
uniq_id	True
crawl_timestamp	True
product_url	True
product_name	True
product_category_tree	True
pid	True
retail_price	True
discounted_price	True
image	True
is_FK_Advantage_product	True
product_rating	True
overall_rating	True
product_specifications	True
dtype: bool	

```
df_all_platforms.isna().sum()
```

Unnamed: 0	47555
asin	47555
avg.rating	47555
avg.helpful.ratio	58109
also_bought	47555
also_viewed	47555

```

brand           5881
categories      47555
description     799
price           47960
salesRank       47555
title           47555
SpecialPrice    47960
IsAvailable     47555
Platform         0
index           31346
product          31347
category         31346
sub_category     31346
sale_price       31346
market_price     31346
type             31346
rating           39972
Category         31346
uniq_id          38901
crawl_timestamp 38901
product_url      38901
product_name     38901
product_category_tree 38901
pid              38901
retail_price     38979
discounted_price 38979
image             38904
is_FK_Advantage_product 38901
product_rating   38901
overall_rating   38901
product_specifications 38915
dtype: int64

```

```
df_all_platforms.isna().sum().sum()
```

```
1385331
```

RECOMMENDATION SYSTEM FOR THE INTEGRATED DATASET

```

def search(sender):
    ''' recommending user based on special price and ratings. '''
    search_item = sender.value.lower()
    df_filtered = df_all_platforms[df_all_platforms['product_name'].astype(str) \
        .str.contains(search_item)].sort_values(by=['SpecialPrice'], ascending=[True])
    .groupby(by=['Platform']).first()
    df = df_filtered.sort_values(by=['rating'], ascending=[False]).copy()
    df.reset_index(inplace=True)
    df_top_product = df.iloc[1]
    msg = str.format('Recommended Platform : {0}', df_top_product.Platform)
    print(tabulate([[msg]], tablefmt='fancy_grid'))
    display(df)
    display(df[['Platform', 'rating']].head(4).style.hide_index().highlight_max(color='lightgreen'))

```

```
fig = plt.figure(figsize=(10,7))
ax = sns.barplot(x="Platform", y="SpecialPrice", data=df, capsize=.2)
# recommending based on price and ratings
text = widgets.Text()
print(tabulate([[ 'Type the product name in below box and press enter to search.']], tablefmt=display(text)
text.on_submit(search)
```

→ +-----+
| Type the product name in below box and press enter to search. |
+-----+

Garlic Oil - Vegetarian Capsule 500 mg

```
IndexError Traceback (most recent call last)
<ipython-input-39-0d387aa845fa> in search(sender)
    7     df = df_filtered.sort_values(by=['rating'], ascending=[False]).copy()
    8     df.reset_index(inplace=True)
--> 9     df_top_product = df.iloc[1]
    10    msg = str.format('Recommended Platform : {0}', df_top_product.Platform)
    11    print(tabulate([[msg]], tablefmt='fancy_grid'))
```

◆ 2 frames

```
/usr/local/lib/python3.7/dist-packages/pandas/core/indexing.py in
_validate_integer(self, key, axis)
1498         len_axis = len(self.obj_.get_axis(axis))
1499         if key >= len_axis or key < -len_axis:
-> 1500             raise IndexError("single positional indexer is out-of-bounds")
1501
1502     # -----
```

IndexError: single positional indexer is out-of-bounds

SEARCH STACK OVERFLOW

Colab paid products - [Cancel contracts here](#)

✓ 0s completed at 3:42 PM

