DOKUMENTASI TUGAS AKHIR DATA MINING 4408

NIM : A11.2021.13434  
NAMA : Nico Darmawan

Code :

import matplotlib.pyplot as plt

import matplotlib.ticker as tick

import pandas  as pd

import numpy   as np

import seaborn as sns

import re

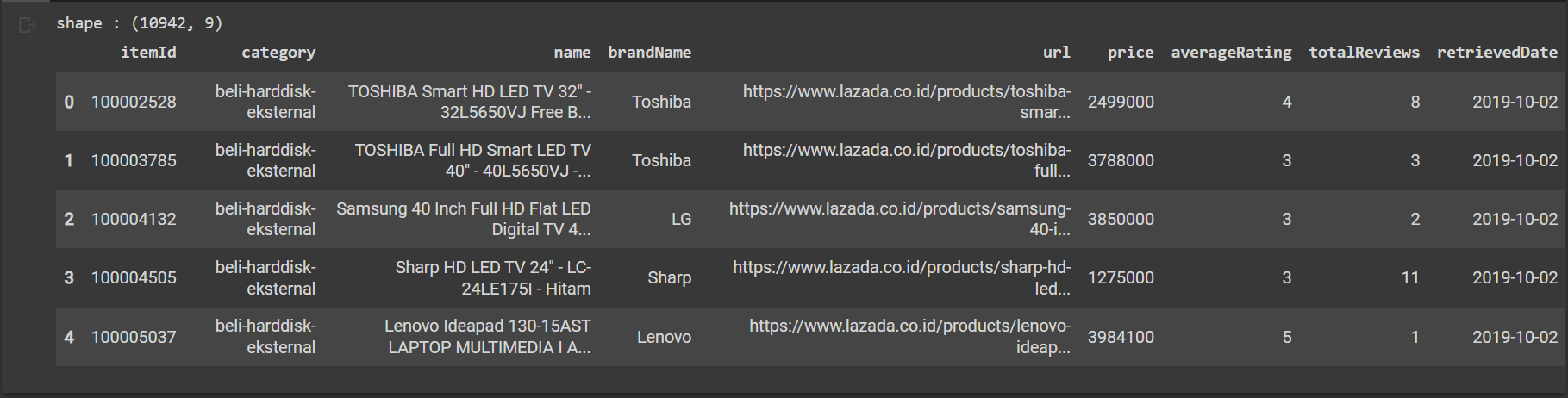
import os

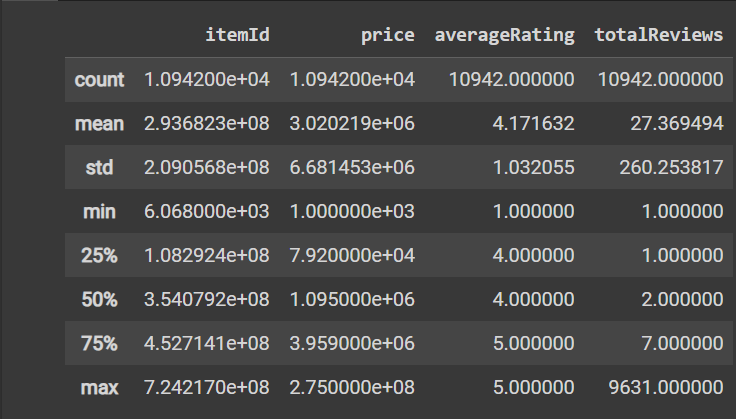
**Load Data**

df\_items = pd.read\_csv('20191002-items.csv')

df\_reviews = pd.read\_csv('20191002-reviews.csv')

print('shape : {}'.format(df\_items.shape))

df\_items.head()

df\_items.describe()

print('shape : {}'.format(df\_reviews.shape))

df\_reviews.head(4)

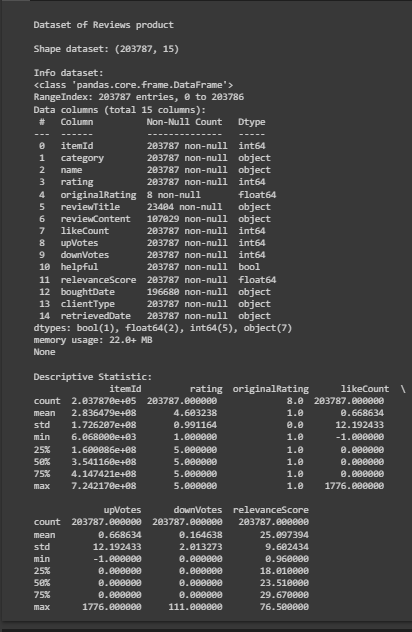
print('\nDataset of Reviews product \n')

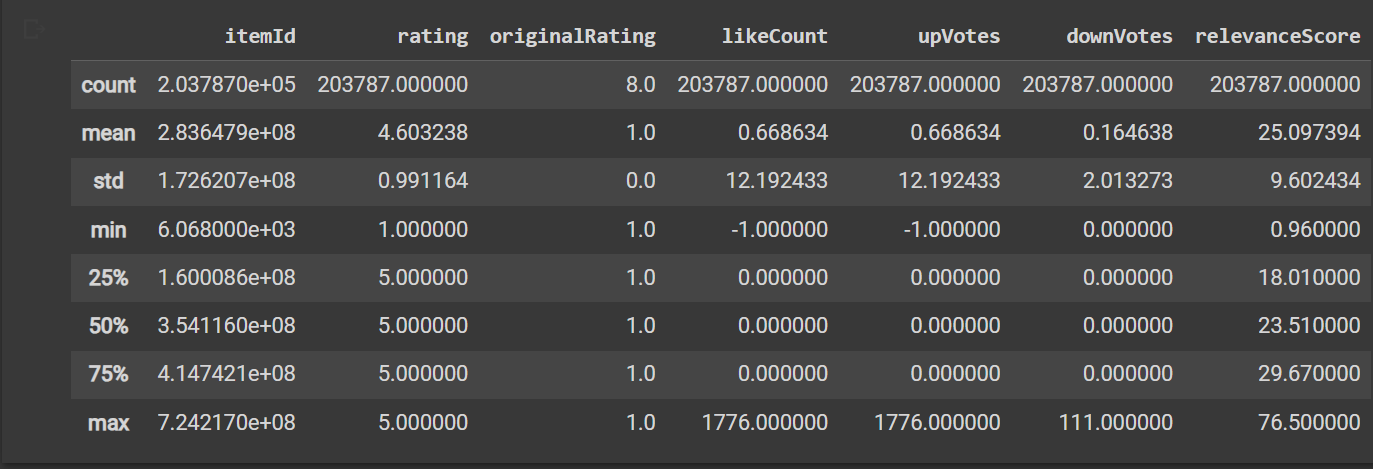
print('Shape dataset:', df\_reviews.shape)

print('\nInfo dataset:')

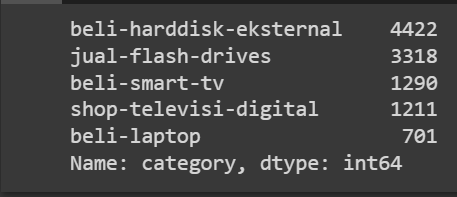
print(df\_reviews.info())

print('\nDescriptive Statistic:\n', df\_reviews.describe())



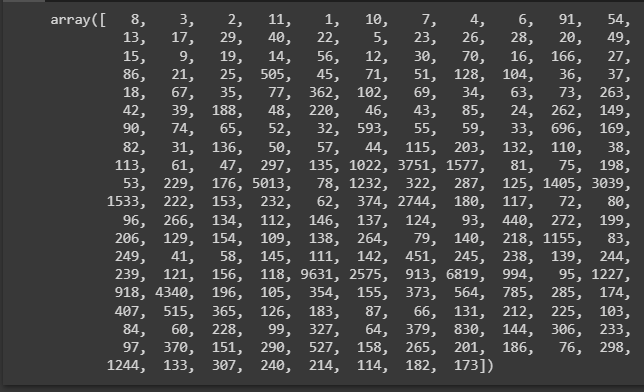
df\_reviews.describe()

df\_items.category.value\_counts()



df\_items['brandName'].unique()



df\_items['totalReviews'].unique()

kategori\_items = ['A','B','C','D','E']

data = [4422 , 3318 , 1290 , 1211 , 701]

warna = ['pink', 'cyan', '#e03364', 'yellowgreen','skyblue']

plt.pie(data,

       labels  = kategori\_items,

       colors  = warna,

       autopct = '%1.1f%%',

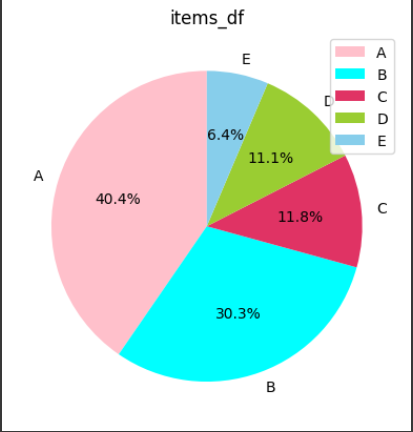
       startangle=90

       )

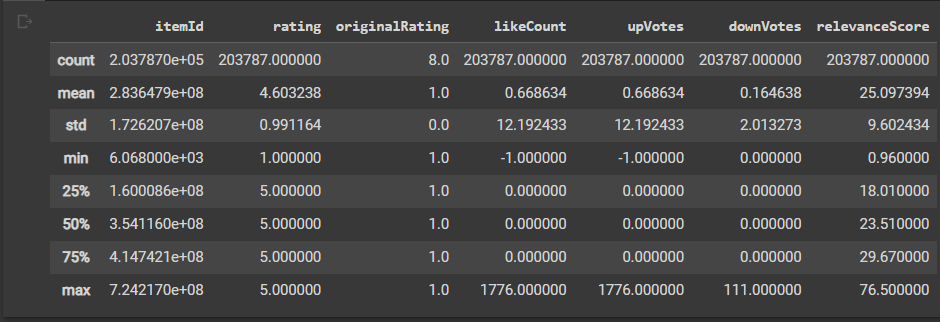
plt.legend()

plt.title ('items\_df')

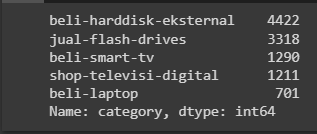
plt.show()



df\_reviews.describe()



df\_items.category.value\_counts()



df\_items = df\_items.head(10)

fig = plt.figure(figsize=(7,3))

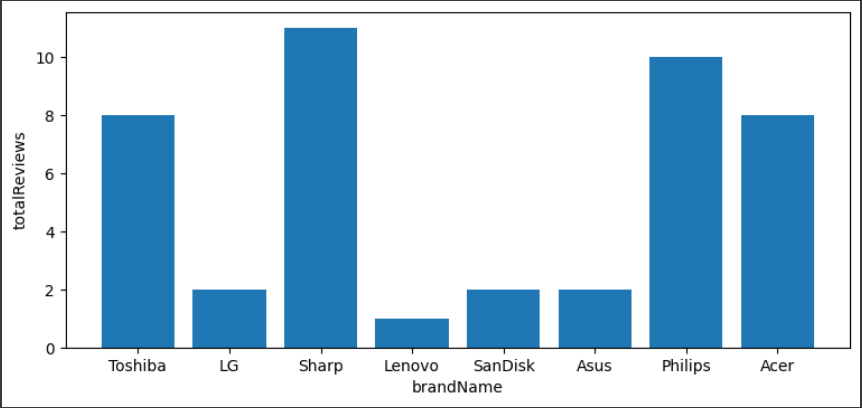
ax  = fig.add\_axes([0,0,1,1])

ax.set\_xlabel('brandName')

ax.set\_ylabel('totalReviews')

ax.bar(df\_items['brandName'],df\_items['totalReviews'])

plt.show()



df\_items = df\_items.groupby('brandName').sum()['averageRating'].reset\_index()

fig = plt.figure()

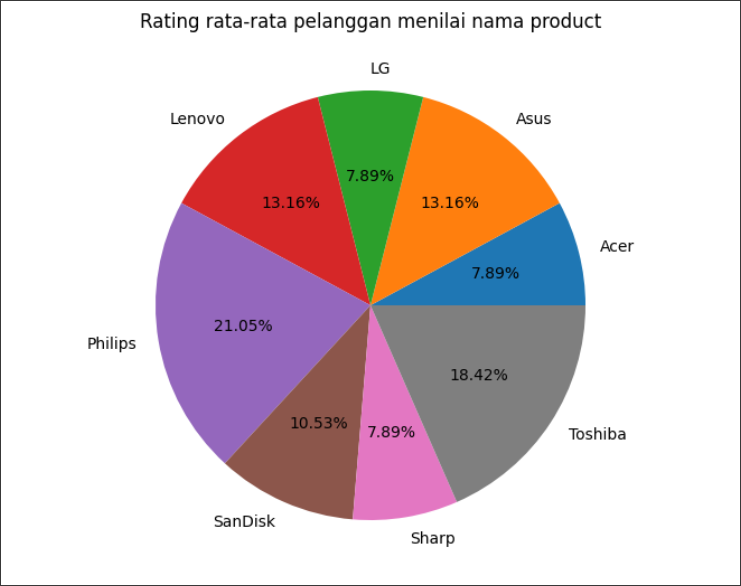
ax  = fig.add\_axes([0,0,1,1])

ax.axis('equal')

ax.set\_title(label= 'Rating rata-rata pelanggan menilai nama product')

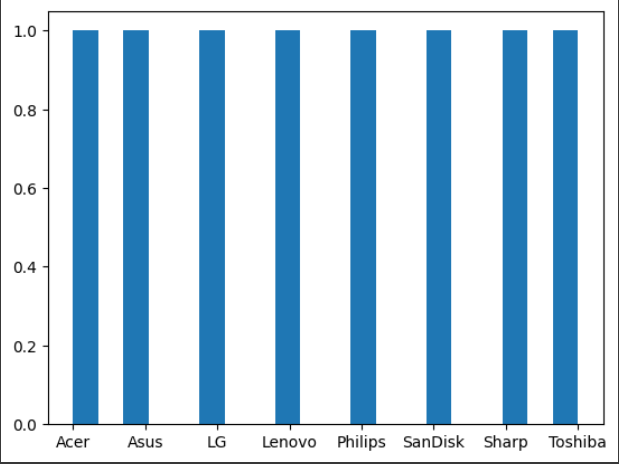
ax.pie(df\_items['averageRating'],labels = df\_items['brandName'],autopct='%1.2f%%')

plt.show()

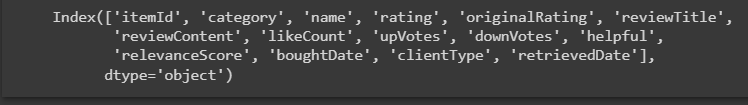


plt.hist(df\_items['brandName'].dropna(), bins=20)

plt.show()



df\_reviews.columns



cat = ['relevanceScore']

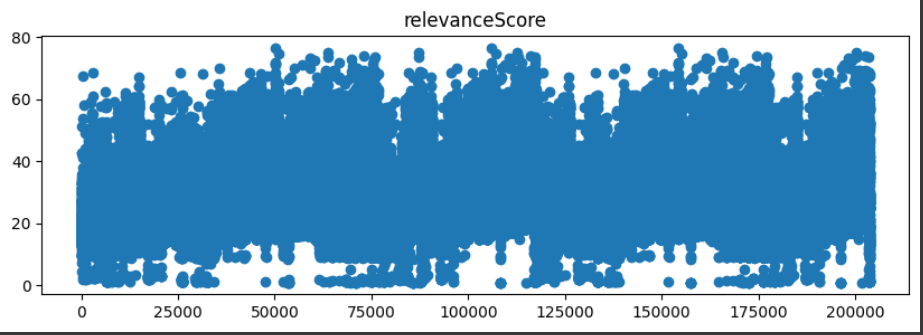
for col in cat:

  plt.figure(figsize=(10,3))

  plt.scatter(df\_reviews.index,df\_reviews[col])

  plt.title(col)

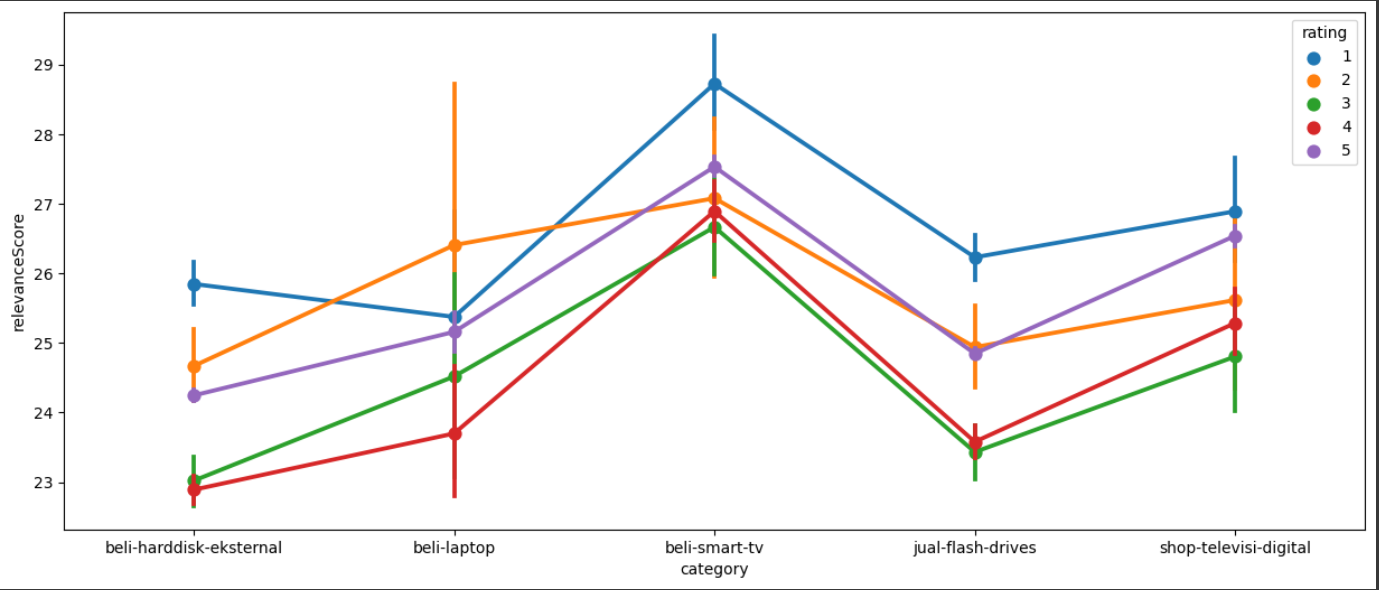
  plt.show()



plt.figure(figsize=(15,6))

sns.pointplot(x="category", y="relevanceScore", hue="rating", data=df\_reviews)

plt.show()



plt.figure(figsize=(15,6))

sns.pointplot(x="clientType", y="relevanceScore", hue="rating", data=df\_reviews)

plt.show()

