Nama : Tri Hesti Wahyuningsih

“ Tugas Sertifikasi DSBIZ (SIB BLOCKCHAIN)”

EXPLORATORY DATA ANALYSIS ON LAZADA

Lazada is the number one online shopping and selling destination in Southeast Asia.
In this notebooklet's do an Exploratory Data Analyis to review the top selling in October 2,2019and analyze the relations ship between the reviews, and prices etc which helps the people to buy.

Import Libraries

Importing Libraries

```
[ ]  1  #menghubungkan ke google
     2  from google.colab import drive
     3  drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
[ ]  1  import pandas as pd
     2  import numpy as np
     3
     4  df = pd.read_csv('/content/drive/MyDrive/AI BLOKCHAIN/Course mandiri/EDA/EDA ON LAZADA/20191002-items.csv')
     5  df = df.sort_values('itemId').reset_index(drop=True)
     6  df.head()
```

	itemId	category	name	brandName	url	price	averageRating	totalReviews	retrievedDate
0	6068	jual-flash-drives	Alienware M17xR5 Core i7-4700 - Hitam	Dell	https://www.lazada.co.id/products/alienware-m1...	30737000	5	5	2019-10-02
1	6068	shop-televisi-digital	Alienware M17xR5 Core i7-4700 - Hitam	Dell	https://www.lazada.co.id/products/alienware-m1...	30737000	5	5	2019-10-02

Get List of Column

```
[ ]  1  df.columns
```

```
Index(['itemId', 'category', 'name', 'brandName', 'url', 'price',
       'averageRating', 'totalReviews', 'retrievedDate'],
      dtype='object')
```

Identify the shape of the datatset

```
1  df.shape
```

```
(10942, 9)
```

Get bassic dataset information

```
1  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10942 entries, 0 to 10941
Data columns (total 9 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   itemId         10942 non-null  int64 
 1   category       10942 non-null  object
 2   name           10942 non-null  object
 3   brandName      10940 non-null  object
 4   url            10942 non-null  object
 5   price          10942 non-null  int64 
 6   averageRating  10942 non-null  int64 
 7   totalReviews   10942 non-null  int64 
 8   retrievedDate  10942 non-null  object
dtypes: int64(4), object(5)
memory usage: 769.5+ KB
```

make sure the item transaction data is only on October 2, 2019

```
[ ]  1  eda_year=df.groupby('retrievedDate').sum().reset_index()
```

```
1  eda_year
```

	retrievedDate	itemId	price	averageRating	totalReviews
0	2019-10-02	3213471716548	33047232096	45646	299477

```
[ ]  1  import plotly.express as px
     2  import plotly.graph_objects as go
```

let's look at the missing values

```
df.isna().sum()
```

```
itemId          0
category        0
name            0
brandName       2
url             0
price           0
averageRating   0
totalReviews    0
retrievedDate   0
dtype: int64
```

```
df[df['brandName'].isna()]
```

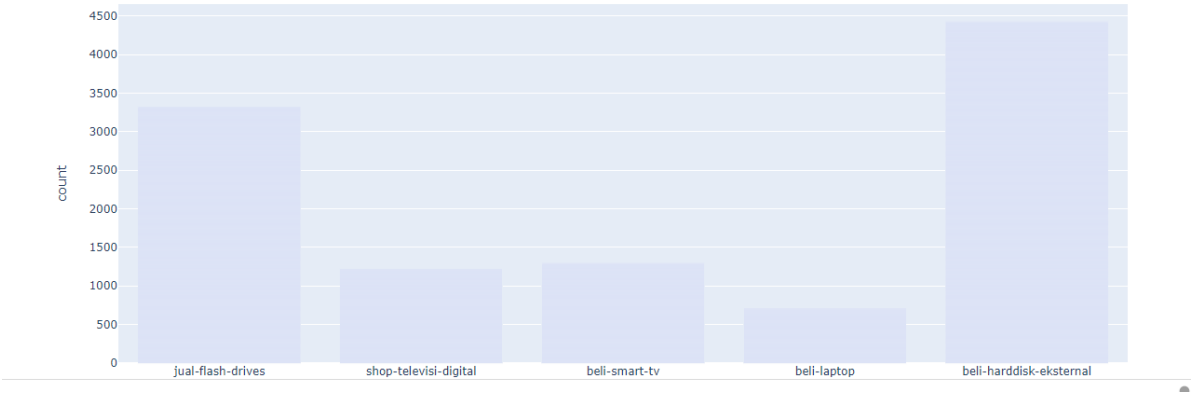
	itemId	category	name	brandName	url	price	averageRating	totalReviews	retrievedDate
390	1063107	beli-harddisk-eksternal	WDC Passport Ultra 1TB External Harddisk Porta...	NaN	https://www.lazada.co.id/products/wdc-passport...	1050000	5	1	2019-10-02
428	1663682	beli-harddisk-eksternal	Western Digital My Passport Ultra - 1 TB - Hitam	NaN	https://www.lazada.co.id/products/western-digi...	1450000	5	2	2019-10-02

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

	itemId	category	name	brandName	url	price	averageRating	totalReviews	retrievedDate
0	6068	jual-flash-drives	Alienware M17xR5 Core i7-4700 - Hitam	Dell	https://www.lazada.co.id/products/alienware-m1...	30737000	5	5	2019-10-02
1	6068	shop-televisi-digital	Alienware M17xR5 Core i7-4700 - Hitam	Dell	https://www.lazada.co.id/products/alienware-m1...	30737000	5	5	2019-10-02
2	6068	beli-smart-tv	Alienware M17xR5 Core i7-4700 - Hitam	Dell	https://www.lazada.co.id/products/alienware-m1...	30737000	5	5	2019-10-02
3	6068	beli-laptop	Alienware M17xR5 Core i7-4700 - Hitam	Dell	https://www.lazada.co.id/products/alienware-m1...	30737000	5	5	2019-10-02

Category data graph

```
px.bar(df, x='category')
```



Relationshipp between total Review & AverageRating

```
px.scatter_matrix(df,dimensions=['totalReviews','averageRating'],color='category', title='Relationship between totalReviews & averageRating')
```

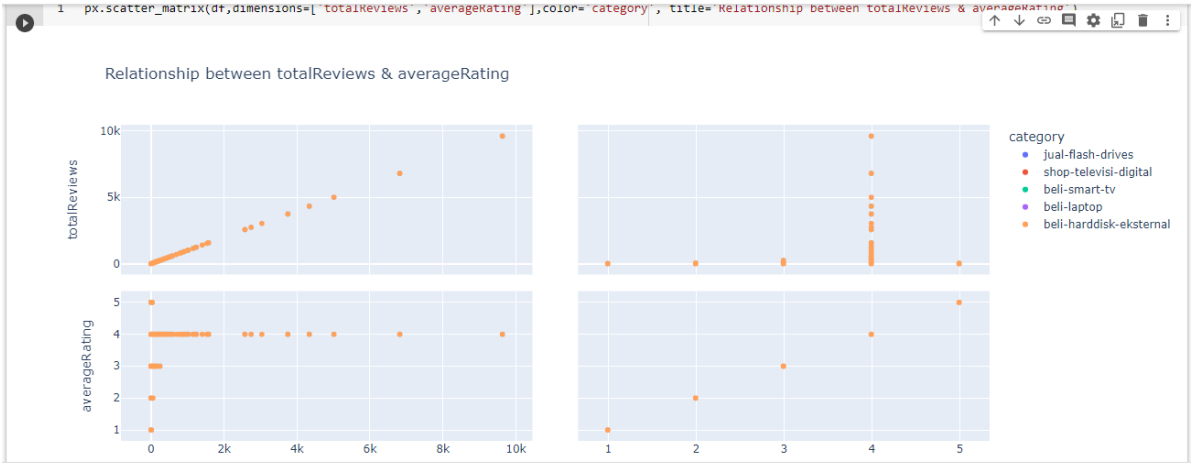


Chart between price data and averageRating

```
1  px.bar(df,x='averageRating',y='price')
```

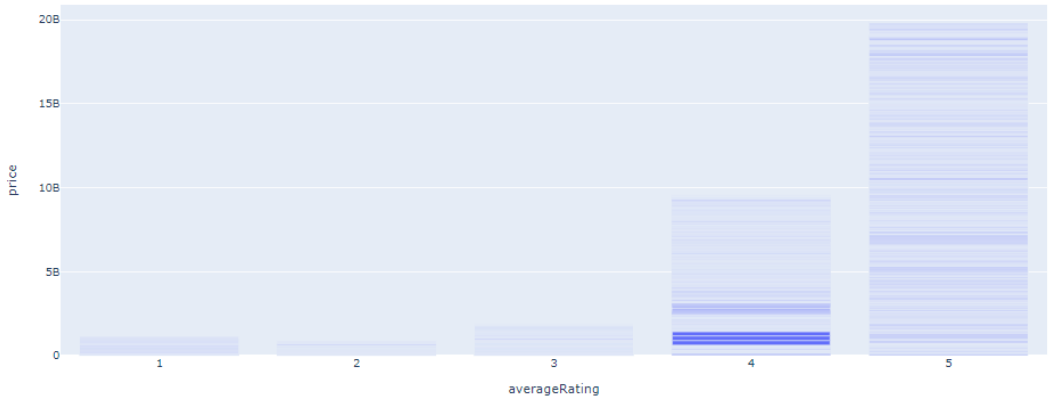
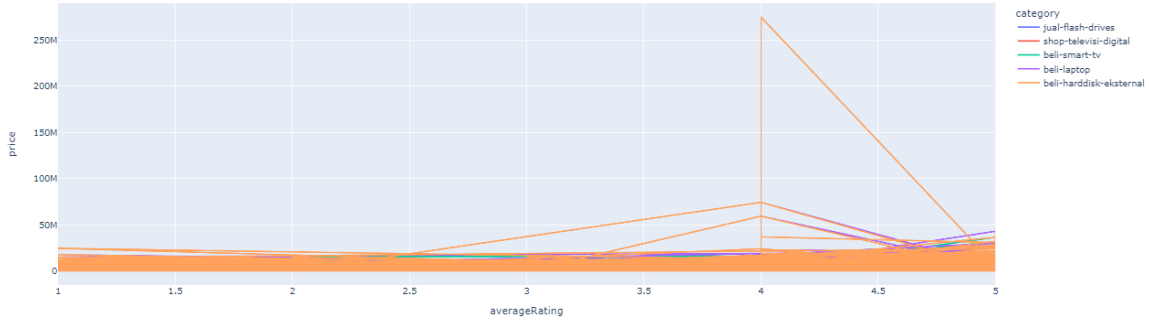


Chart between price data, avergaeRating & category

```
[ ]  1  px.line(df, x='averageRating', y='price', color='category')
```