Analisis Data Netflix Originals Films & IMDB Score

Laporan Tugas Besar

Disusun sebagai syarat tugas besar

Oleh:

Kartini Copa -16521489

Eleanora Felicia - 16521491

Syasya Umaira - 16521537

Jason Rivalino – 16521541

KU1102-72 Pengenalan Komputasi

SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA

INSTITUT TEKNOLOGI BANDUNG

NOVEMBER 2021

KATA PENGANTAR

Puji dan syukur kami panjatkan kehadirat tuhan yang Maha Esa, karena berkat karunia-Nya kami dapat menyelesaikan Tugas Besar ini yang berjudul ‘Analisis Data Netflix Originalas Films & IMDB Score’. Adapun tujuan dibuat tugas ini adalah agar mahasiswa dapat mengenali persoalan komputasi dalam persoalan yang di berikan, menggunakan dekomposisi dan abstraksi dalam persoalan, serta menghasilkan artifak komputasi yang relevan. Mahasiswa juga mampu berkolaborasi dalam kelompok pemecahan persoalan komputasi dan agar mahasiswa mampu berkomunikasi dengan berbagai pihak dalam rangka mengekspresikan dan bertukar ide mengenai pemecahan persoalan komputasi.

Kami berharap agar materi yang kami bawakan ini dapat bermanfaat dan mudah untuk dimengerti dengan baik bagu pembaca. Selain itu, kami juga berharap materi yang kami bawakan dapat membuat pembaca semakin memahami deskripsi data dan file, karakteristik data, statistik data, visualisasi, dan korelasi data.

Kami nyadari bahwa laporan yang kami bawakan ini masih jauh dari kata sempurna. Maka dari itu, kami mengharapkan kritik, saran, serta masukan yang membangun dari pembaca sebagai motivasi bagi penulis gun menyempurnakan materi ini.

Kami ingin mengucapkan terima kasih kepada semua pihak yang telah membantu kami menyelesaikan Tugas Besar ini, terutama dosen kami yaitu Bapak Rizal Dwi Prayogo yang sudah membimbing kami dalam menyelesaikan Tugas ini. Semoga laporan ini dapat bermanfaat dan dapat memberikan pengetahuan-pengetahuan baru bagi pembaca.

DAFTAR ISI

TUGAS 3

A. Deskripsi data dan file

Dataset yang kami analisis adalah Netfilx Original Film & IMDB Score. Tentu saja data ini membahas skor IMDB film original Netflix dengan format data csv. Dari data dapat diketahui genre, hari, bulan , tahun, skor IMDB, dan korelasi antara data yang satu dengan yang lain. Data ini berisikan semua film Netflix Original yang dirilis sebelum tanggal 1 Juni 2021. Selain itu, data ini juga mencakup film dokumenter dan film special Netflix. Data tersebut diambil dari halaman Wikipedia yang kemudian diintegrasikan dengan dataset yang terdiri dari skor IMDB. Skor IMDB dipilih oleh para pengguna dan sebagian besar film tersebut memiliki 1.000 lebih ulasan. Dataset Netfilx Original Film & IMDB Score dapat dilihat di link https: //www.kaggle.com/luiscorter/netflix-original-films-imdb-scores. Dimensi data adalah 6 kolom dan 153 baris. Ukuran file dataset Netflix Original Film & IMDB Score adalah 38,6 Kb. Bahasa pemrograman yang kami gunakan adalah python. Kami menggunakan pandas untuk loading data; load pandas dan dataframe dari file data.csv. Kemudian, load *library* matplotlib.pyplot untuk plotting; menampilkan bar, chart dan histogram. Kami juga menggunakan *library* numpy untuk mengetahui informasi terkait data.

TUGAS 4

A. Karakteristik data

Dalam data Netflix Original Films & IMDB Scores terdapat 6 atribut/kolom, yaitu:

1. Atribut Title

Atribut title adalah atribut yang memaparkan judul film netflix dan termasuk tipe data kategorikal-nominal.

Title terdiri atas nilai data berupa berbagai judul film yang ada dalam range nilai IMDb lebih dari 7 (memuat judul film).

Persen data kosong: 0%

Spreadsheet tool yang digunakan untuk mendapatkan: melakukan sorting

2. Atribut Genre

Atribut genre adalah atribut yang memaparkan pembagian tipe film dan termasuk tipe data kategorikal-nominal.

Genre terdiri atas nilai data berupa berbagai jenis tipe film yang ada dari data ini (terdiri atas 31 genre film yang berbeda, ada film yang merupakan gabungan dari dua genre atau lebih seperti gabungan genre animation/science fiction).

Persen data kosong: 0%

Spreadsheet tool yang digunakan untuk mendapatkan: melakukan sorting

3. Atribut Premiere

Atribut premiere adalah atribut yang memaparkan tanggal film tersebut ditampilkan atau ditunjukkan untuk pertama kalinya di Netflix dan termasuk tipe data kategorikal-nominal.

Premiere terdiri atas nilai data berupa tanggal, bulan, dan tahun dari film itu dirilis. Film yang dirilis memiliki rentang waktu antara tahun 2015 hingga tahun 2021 sebelum tanggal 1 Juni.

Persen data kosong: 0%

Spreadsheet tool yang digunakan untuk mendapatkan: melakukan sorting

4. Atribut Runtime

Atribut runtime adalah atribut yang menunjukkan berapa lama film tersebut ditayangkan atau diputar dan termasuk tipe data kuantitatif-continues.

Runtime terdiri atas nilai data berupa menit , yaitu berapa lama film diputar dalam hitungan menit.

Persen data kosong: 0%

Spreadsheet tool yang digunakan untuk mendapatkan: melakukan sorting

5. Atribut IMDb Score

Atribut IMDb Score adalah atribut yang menunjukkan rating atau peringkat film terbaik dan termasuk tipe data kuantitatif-continues.

IMDb Score pada umumnya terdiri atas nilai data dengan rentang 1-10. Pada data netflix ini hanya terdapat rentang nilai 7-9.

Persen data kosong: 0%

Spreadsheet tool yang digunakan untuk mendapatkan: melakukan sorting

6. Atribut Language

Atribut language adalah atribut yang menunjukkan bahasa yang digunakan pada film tersebut dan termasuk jenis kategorikal-nominal.

Language terdiri atas nilai data berupa berbagai jenis Bahasa pada film yang ada dalam data ini (terdiri atas 21 bahasa film yang berbeda, ada film yang merupakan gabungan dari dua bahasa atau lebih seperti gabungan bahasa English/Spanish).

Persen data kosong: 0.66%

Spreadsheet tool yang digunakan untuk mendapatkan: melakukan sorting.

TUGAS 5

A. Sampel Data

```
import pandas as pd
df=pd.read_csv("Netflixoriginal.csv")
print(df.loc[20:29]) #menampilkan data pada baris ke 21 hingga 30 (perhitungan baris dimulai dari 0)
print()              #[select beberapa data dalam range baris tertentu]
print(df.loc[(df["Runtime"]<=90)&(df["IMDb Score"]>=8.5)]) #menampilkan data film yang durasinya kurang dari 90 menit
print()                                                    #dan IMDb Scorenya lebih dari 8.5 [select data dengan range tertentu]
print(df.sort_values(["IMDb Score"], ascending=[0])) #menampilkan urutan data berdasarkan IMDb Score terbesar hingga terkecil
print()                                              #[sort data terbesar hingga terkecil]
print(df.sort_values(["Runtime"], ascending=[1])) #menampilkan urutan data berdasarkan durasi film terkecil hingga terbesar
print()                                           #[sort data terkecil hingga terbesar]
print(df.loc[(df["IMDb Score"]>=8.0),["Title","Language"]]) #hanya menampilkan judul dan bahasa pada film dengan IMDb Score lebih dari 8.0
print()                                                     #[sampel data pada setiap kolom]
```

Penjelasan: program menampilkan data dalam bentuk csv dengan mengimport pandas terlebih dahulu. Pada sampel terdapat 5 kondisi. Kondisi pertama menampilkan data pada baris tertentu. Kondisi kedua menampilkan durasi dan IMDB score tertentu. Kondisi ketiga dan keempat menampilkan data dari yang terbesar ke yang terkecil dalam hal ini durasi dan IMDB score. Kondisi yang terakhir suatu nilai menampilkan kolom tertentu dalam hal ini IMDB score.

```
C:\Users\karti\OneDrive\Dokumen\NETFLIX TUGAS 5>C:/Users/karti/AppData/Local/Programs/Python/Python310/python.exe "c:/Users/karti/OneDrive/Dokumen/NETFLIX TUGAS 5/netfilxtugas5.1.py"
                                Title            Genre          Premiere  Runtime  IMDb Score Language
20                  Angela's Christmas        Animation   November 30, 2018       30         7.1  English
21              Angela's Christmas Wish        Animation    December 1, 2020       47         7.1  English
22                               Beats            Drama      June 19, 2019      110         7.1  English
23                    Circus of Books      Documentary      April 22, 2020       92         7.1  English
24  Dance Dreams: Hot Chocolate Nutcracker      Documentary   November 27, 2020       80         7.1  English
25              Derren Brown: Sacrifice  Mentalism special    October 19, 2018       49         7.1  English
26              El Pepe: A Supreme Life      Documentary   December 27, 2019       73         7.1  Spanish
27                             End Game      Documentary        May 4, 2018       40         7.1  English
28                               Evelyn      Documentary  September 10, 2019       96         7.1  English
29                                Ferry      Crime drama        May 14, 2021      106         7.1    Dutch

                                Title       Genre         Premiere  Runtime  IMDb Score    Language
150  Emicida: AmarElo - It's All For Yesterday  Documentary  December 8, 2020       89        8.6  Portuguese
151   David Attenborough: A Life on Our Planet  Documentary   October 4, 2020       83        9.0     English

                                Title                          Genre          Premiere  Runtime  IMDb Score                Language
151    David Attenborough: A Life on Our Planet                Documentary    October 4, 2020       83        9.0                 English
150   Emicida: AmarElo - It's All For Yesterday                Documentary   December 8, 2020       89        8.6               Portuguese
149                  Springsteen on Broadway                  One-man show  December 16, 2018      153        8.5                 English
148  Winter on Fire: Ukraine's Fight for Freedom              Documentary    October 9, 2015       91        8.4  English/Ukranian/Russian
147       Taylor Swift: Reputation Stadium Tour               Concert Film  December 31, 2018      125        8.4                 English
..                                  ...                            ...               ...      ...        ...                     ...
15                          Rose Island                          Comedy   December 9, 2020      117        7.0                 Italian
16                The Christmas Chronicles  Christmas/Fantasy/Adventure/Comedy  November 22, 2018      104        7.0                 English
17                                The Dirt                         Biopic      March 22, 2019      108        7.0                 English
18                     The Night Comes for Us                Action-thriller   October 19, 2018      121        7.0               Indonesian
0          Feminists: What Were They Thinking?                  Documentary   October 12, 2018       86        7.0                 English

[152 rows x 6 columns]

                                Title                Genre          Premiere  Runtime  IMDb Score Language
66                               Zion          Documentary    August 10, 2018       11         7.2  English
128            If Anything Happens I Love You  Animation / Short  November 20, 2020       12         7.8  English
64   The Road to El Camino: A Breaking Bad Movie          Making-of   October 29, 2019       13         7.2  English
2    I'm No Longer Here: A Discussion with Guillerm...  Aftershow / Interview   November 3, 2020       14         7.0  English
120                             Anima  Musical / Short      June 27, 2019       15         7.7  English
..                                ...              ...               ...      ...         ...      ...
113  Rolling Thunder Revue: A bob Dylan Story by Ma...      Documentary      June 12, 2019      144         7.6  English
111                             Ludo  Anthology/Dark comedy  November 12, 2020      149         7.6    Hindi
77                      Raat Akeli Hai         Thriller      July 31, 2020      149         7.3    Hindi
149            Springsteen on Broadway      One-man show  December 16, 2018      153         8.5  English
129                      The Irishman      Crime drama  November 27, 2019      209         7.8  English

[152 rows x 6 columns]

                                Title                Language
135  Struggle: The Life and Lost Art of Szukaiski         English
136                  Chasing Coral                    English
137               My Octopus Teacher                  English
138                  Rising Phoenix                   English
139                             13th                  English
140  Disclosure: Trans Lives on Screen                English
141                           Klaus                   English
142                      Seaspiracy                   English
143  The Three Deaths of Marisela Escobedo            Spanish
144            Cuba and the Cameraman                 English
145             Dancing with the Birds                English
146  Ben Platt: Live from Radio City Music Hall        English
147       Taylor Swift: Reputation Stadium Tour        English
148  Winter on Fire: Ukraine's Fight for Freedom  English/Ukranian/Russian
149            Springsteen on Broadway                English
150  Emicida: AmarElo - It's All For Yesterday      Portuguese
151   David Attenborough: A Life on Our Planet         English
```

Penjelasan: ini merupakan output dari beberapa kondisi yang di jelaskan sebelumnya.

B. Statistik

Atribut data: rata-rata, standar deviasi, *percentile* (10%, 25%, 50%, 75%, 90), nilai minimum dan maksimum data kuantitatif

```python
#statistik data
from numpy.lib.function_base import quantile
import pandas as pd
import numpy as np
from numpy import percentile
df=pd.read_csv("Netflixoriginal.csv")
std=np.std(df) #menghitung standar deviasi pada data numerik yaitu runtime dan IMDb Score
print("Standar Deviasi:")
print(std)
print()
mean=np.mean(df) #menghitung rata-rata pada data numerik yaitu runtime dan IMDb Score
print("Rata-rata:")
print(mean)
print()
q1,q2,q3,q4,q5=np.percentile(df["Runtime"],[10,25,50,75,90]) #menghitung simpangan kuartil untuk durasi film
print("Simpangan kuartil (Percentile) 10% untuk Runtime:",q1)
print("Simpangan kuartil (Percentile) 25% untuk Runtime:",q2)
print("Simpangan kuartil (Percentile) 50% untuk Runtime:",q3)
print("Simpangan kuartil (Percentile) 75% untuk Runtime:",q4)
print("Simpangan kuartil (Percentile) 90% untuk Runtime:",q5)
print()
min1=(df.min()["Runtime"])
min2=(df.min()["IMDb Score"])
print("Nilai minimum:") #menghitung nilai minimum pada data numerik yaitu runtime dan IMDb Score
print("Waktu film minimum adalah",min1,"menit")
print("IMDb Score minimum adalah",min2)
print()
max1=(df.max()["Runtime"])
max2=(df.max()["IMDb Score"])
print("Nilai maksimum:") #menghitung nilai maksimum pada data numerik yaitu runtime dan IMDb Score
print("Waktu film maksimum adalah",max1,"menit")
print("IMDb Score maksimum adalah",max2)
```

Penjelasan: berdasarkan coding diatas kami mencari standar deviasi, rata-rata, *percentile* (10%, 25%, 50%, 75%, 90), nilai minimum dan maksimum data kuantitatif. Langkah awal adalah dengan mengimport pandas dan numpy, lalu dari numpy import ke *percentile.* Setelah itu, gunakan formula python untuk mencari nilai yang diinginkan.

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL
C:\Users\hp\AppData\Roaming\Python\Python310\site-packages\numpy\core\fromnumeric.py:3579: FutureWarning: Dropping of nuisance columns in DataFrame
reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError.  Select only valid columns before calling the re
duction.
  return std(axis=axis, dtype=dtype, out=out, ddof=ddof, **kwargs)
Standar Deviasi:
Runtime       34.435389
IMDb Score     0.398238
dtype: float64

C:\Users\hp\AppData\Roaming\Python\Python310\site-packages\numpy\core\fromnumeric.py:3438: FutureWarning: Dropping of nuisance columns in DataFrame
reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError.  Select only valid columns before calling the re
duction.
  return mean(axis=axis, dtype=dtype, out=out, **kwargs)
Rata-rata:
Runtime       89.993421
IMDb Score     7.412500
dtype: float64

Simpangan kuartil (Percentile) 10% untuk Runtime: 39.0
Simpangan kuartil (Percentile) 25% untuk Runtime: 75.5
Simpangan kuartil (Percentile) 50% untuk Runtime: 96.0
Simpangan kuartil (Percentile) 75% untuk Runtime: 110.0
Simpangan kuartil (Percentile) 90% untuk Runtime: 128.70000000000002

Nilai minimum:
Waktu film minimum adalah 11 menit
IMDb Score minimum adalah 7.0

Nilai maksimum:
Waktu film maksimum adalah 209 menit
IMDb Score maksimum adalah 9.0
```

Penjelasan: setelah di running akan muncul output seperti gambar diatas. Data yang keluar hanya ada dua karena dari sampel data hanya memiliki dua data numerik, yaitu runtime/durasi dan IMDb score.

TUGAS 6

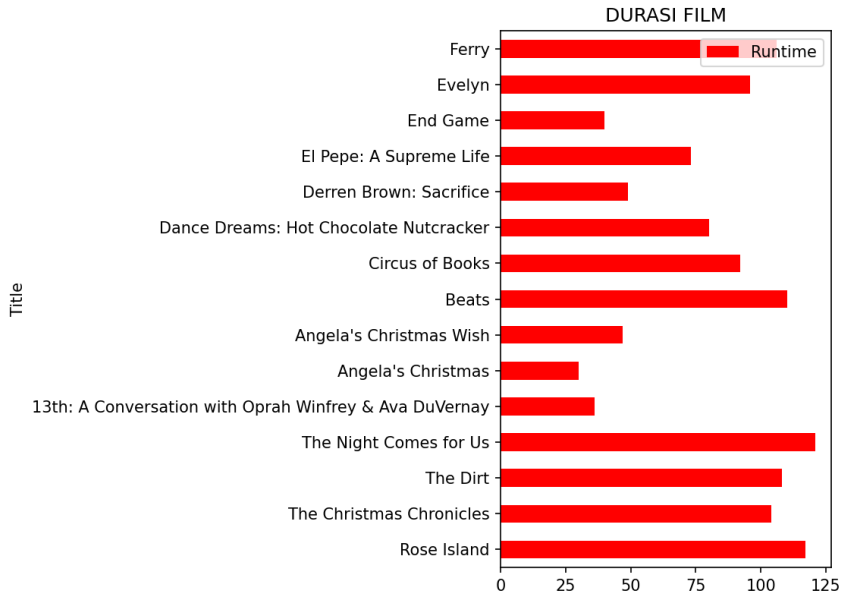
A. Visualisasi

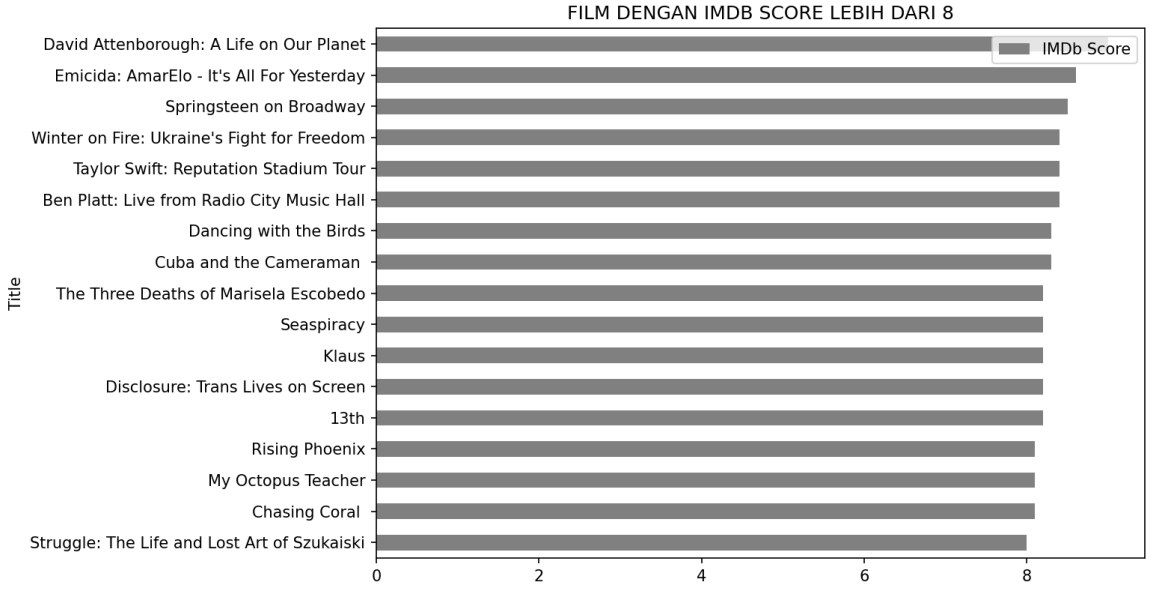
1. Kategori perbandingan dan penampilan perubahan waktu

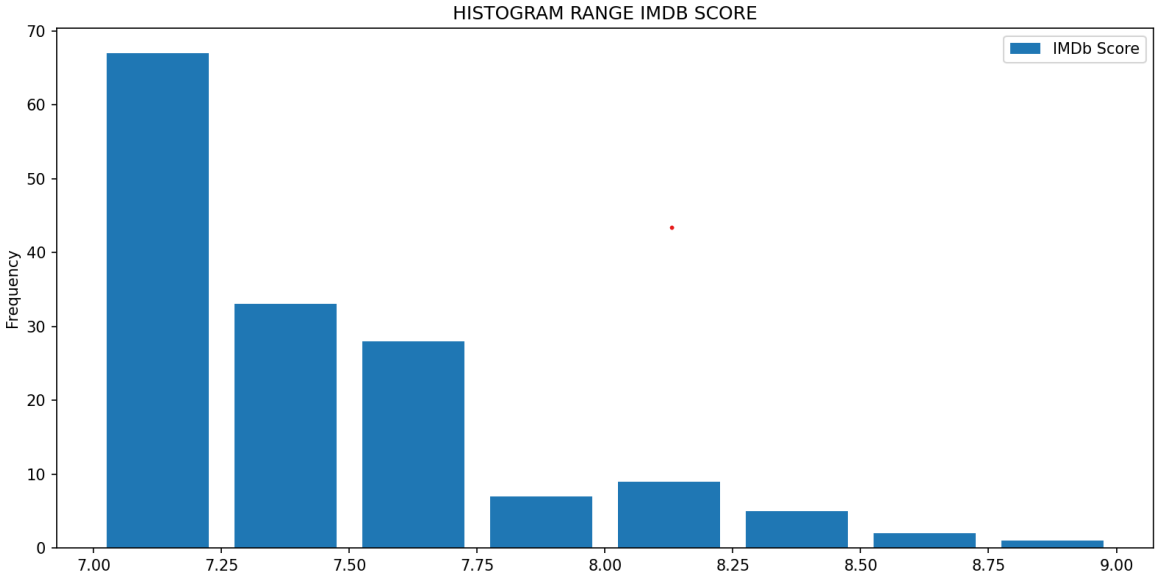
```python
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv("Netflixoriginal.csv")
df.loc[15:29].plot(kind="barh",x="Title",y="Runtime",title="DURASI FILM",color="red")
plt.show() #comparing categories horizontal bar chart
df.loc[(df["IMDb Score"]>=8)].plot(kind="barh",x="Title",y="IMDb Score",title="FILM DENGAN IMDB SCORE LEBIH DARI 8",color="gray")
plt.show() #comparing categories horizontal bar chart
print()
df[["IMDb Score"]].plot(kind="hist",bins=[7,7.25,7.5,7.75,8,8.25,8.5,8.75,9],rwidth=0.8,title="HISTOGRAM RANGE IMDB SCORE")
plt.show() #comparing categories histogram
print()
df.loc[(df["Runtime"]<=25)].plot(kind="line",x="IMDb Score",y=["Runtime"],title="LINE CHART RATING FILM DENGAN WAKTU KURANG DARI 25 MENIT",color="black")
plt.show() #showing over times line chart
print()
df.loc[(df["Runtime"]>=135)].plot(kind="area",x="IMDb Score",y=["Runtime"],title="AREA CHART RATING FILM DENGAN WAKTU LEBIH DARI 135 MENIT",color="orange")
plt.show() #showing over times area chart
print()
```

Penjelasan: pertama impor Pandas untuk membaca file csv. Lalu df loc 15:29 karena perhitungan baris dimulai dari nol maka ditampilkan data dari baris ke 16 sampai 30 jenis grafiknya adalah horizontal barchart dengan sumbu horizontal adalah runtime atau durasi film lalu sumbu vertikal adalah judul film. Program ketiga df loc untuk imdb score lebih dari 8, jenis grafiknya adalah horizontal barchart dengan sumbu horizontalnya adalah imdb score dan sumbu tegaknya adalah judul film. Program keempat menampilkan imdb score dengan tipe grafik histogram dengan range 7 hingga 9 dengan sumbu horizontal adalah range imdb score dan sumbu vertikal adalah banyak nya film dengan imdb score tertentu. Program kelima menampilkan imdb score atau rating film yang runtime nya kurang dari dan sama dengan 25 menit dengan grafik line chart dan sumbu horizontalnya adalah IMDb score dan sumbu vertikalnya adalah runtime film. Program keenam menampilkan imdb score atau rating film yang runtime nya lebih dari dan sama dengan 135 menit dengan tipe area chart, sumbu horizontalnya adalah IMDb score dan sumbu vertikalnya adalah runtime film.

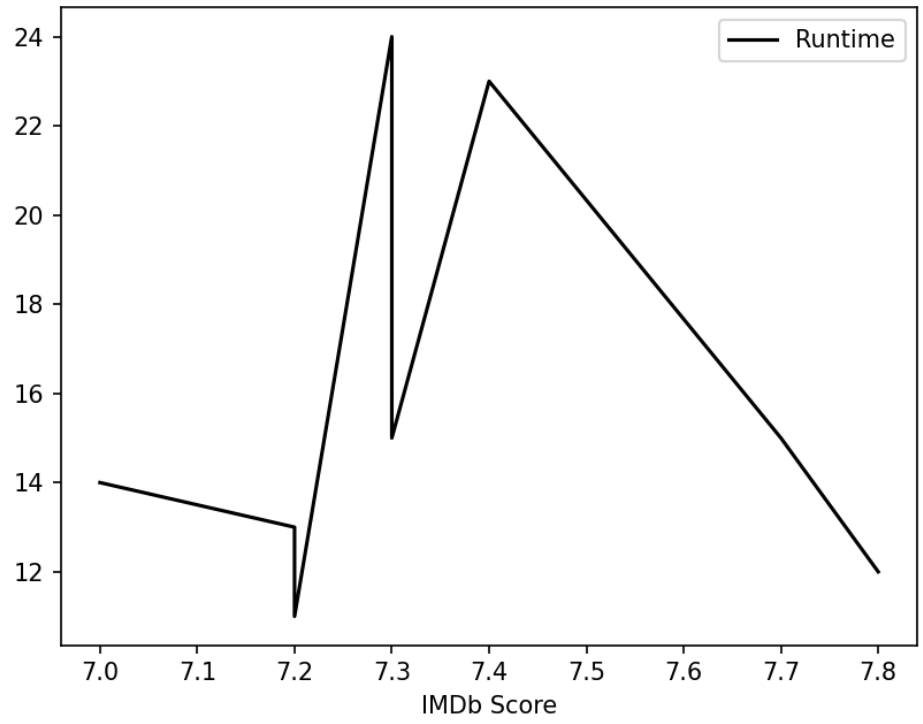
Berikut adalah outputnya:



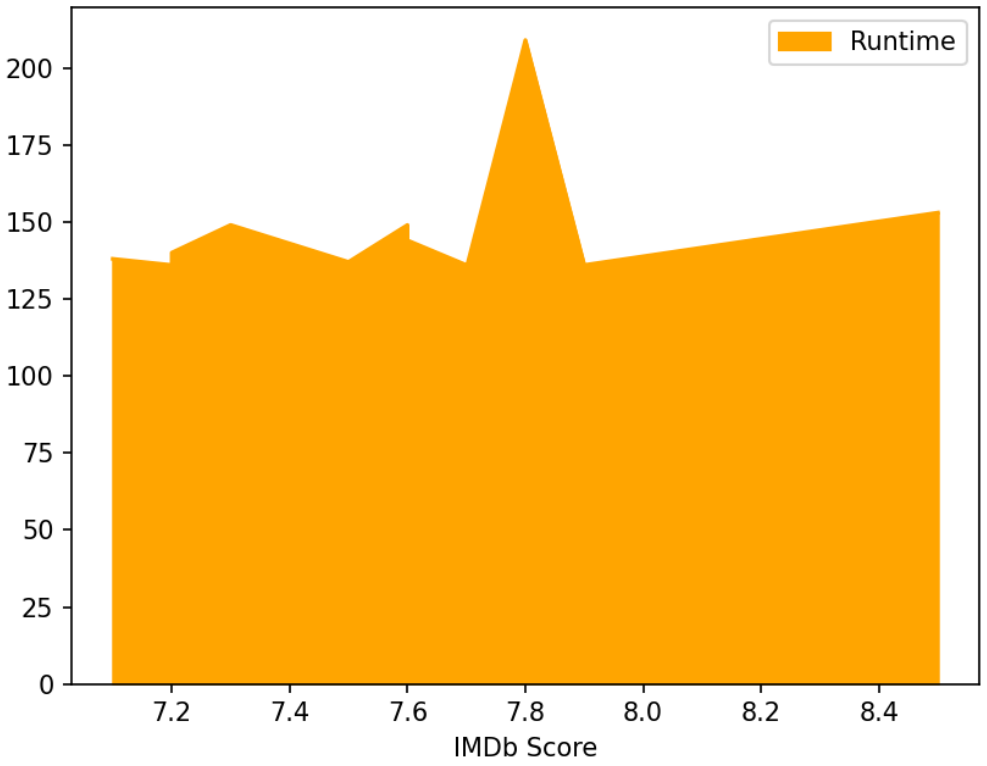

FILM DENGAN IMDB SCORE LEBIH DARI 8
IMDb Score
Title
David Attenborough: A Life on Our Planet
Emicida: AmarElo - It's All For Yesterday
Springsteen on Broadway
Winter on Fire: Ukraine's Fight for Freedom
Taylor Swift: Reputation Stadium Tour
Ben Platt: Live from Radio City Music Hall
Dancing with the Birds
Cuba and the Cameraman
The Three Deaths of Marisela Escobedo
Seaspiracy
Klaus
Disclosure: Trans Lives on Screen
13th
Rising Phoenix
My Octopus Teacher
Chasing Coral
Struggle: The Life and Lost Art of Szukaiski
0
2
4
6
8


HISTOGRAM RANGE IMDB SCORE
IMDb Score
Frequency
70
60
50
40
30
20
10
0
7.00
7.25
7.50
7.75
8.00
8.25
8.50
8.75
9.00

LINE CHART RATING FILM DENGAN WAKTU KURANG DARI 25 MENIT



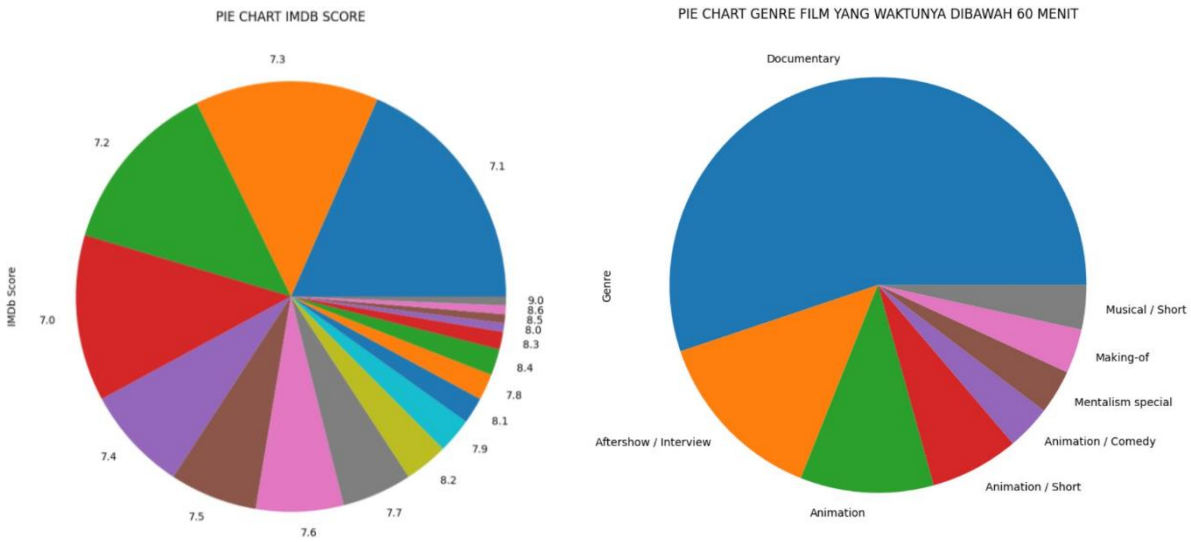
AREA CHART RATING FILM DENGAN WAKTU LEBIH DARI 135 MENIT

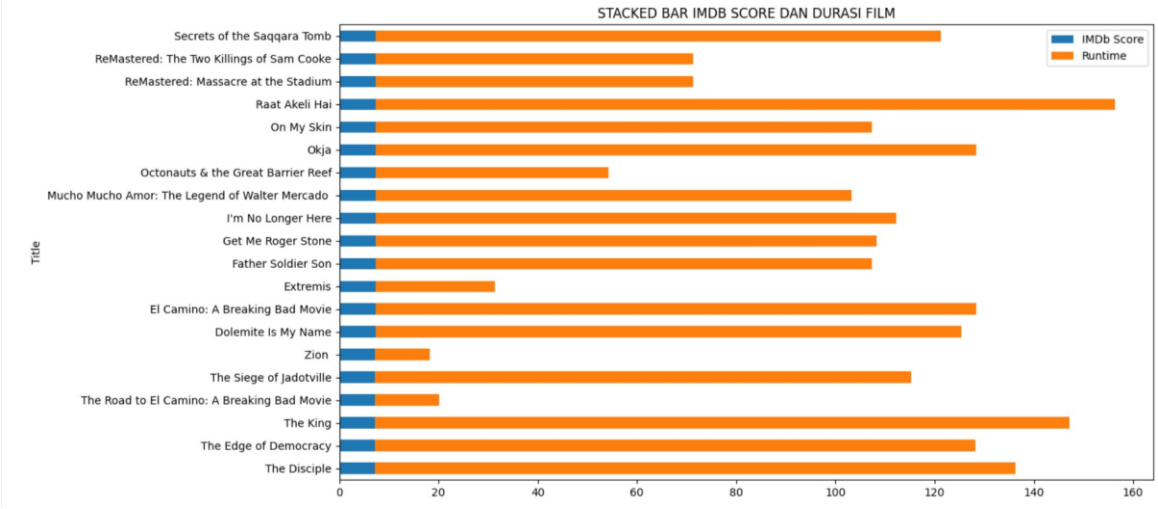


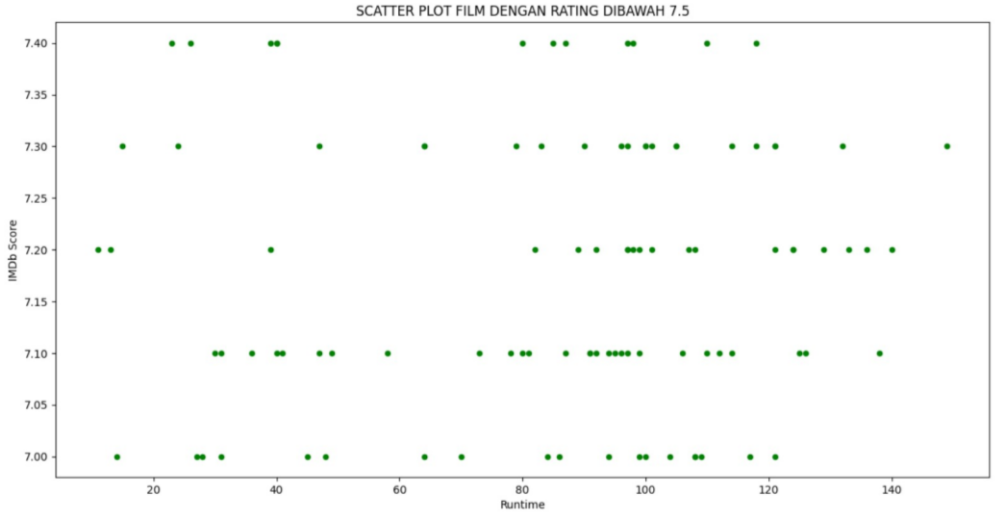
2. Kategori penampilan hierarki dan plotting relationships

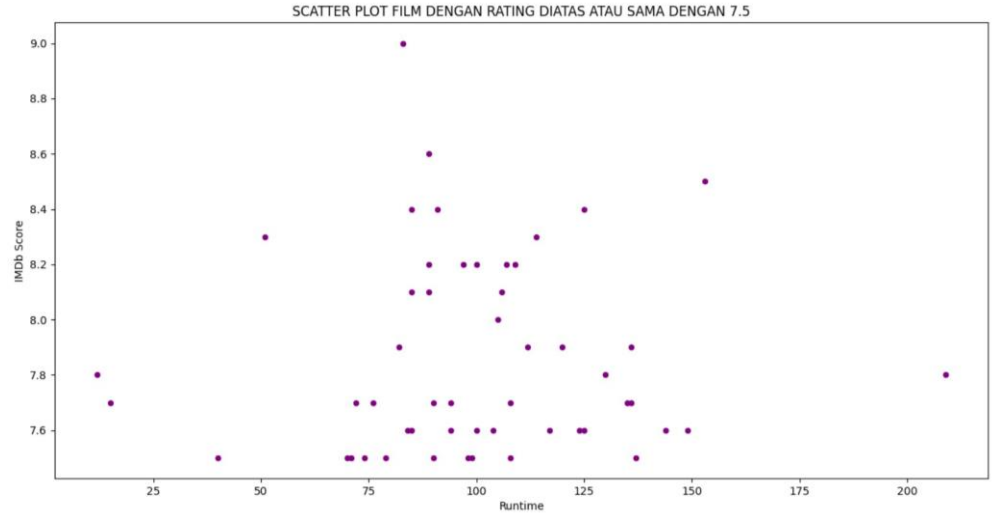
```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv("Netflixoriginal.csv")
df2=df["IMDb Score"].value_counts() #whole-part relationships pie chart
df2.plot(kind="pie",title="PIE CHART IMDB SCORE")
plt.show()
print()
df2=(df.loc[(df["Runtime"]<=60)])["Genre"].value_counts() #whole-part relationships pie chart with selection
df2.plot(kind="pie",title="PIE CHART GENRE FILM YANG WAKTUNYA DIBAWAH 60 MENIT")
plt.show()
print()
df.loc[61:80].plot(kind="barh",x="Title",y=["IMDb Score","Runtime"], stacked = True,title="STACKED BAR IMDB SCORE DAN DURASI FILM")
plt.show() #whole-part relationships stacked bar
print()
df.loc[(df["IMDb Score"]<7.5)].plot(kind="scatter", x="Runtime", y="IMDb Score",title="SCATTER PLOT FILM DENGAN RATING DIBAWAH 7.5",color="green")
plt.show() #plotting relationships scatter plot
print()
df.loc[(df["IMDb Score"]>=7.5)].plot(kind="scatter", x="Runtime", y="IMDb Score",title="SCATTER PLOT FILM DENGAN RATING DIATAS ATAU SAMA DENGAN 7.5",color="purple")
plt.show() #plotting relationships scatter plot
print()
```

Berikut outputnya:




STACKED BAR IMDB SCORE DAN DURASI FILM
IMDb Score
Runtime
Title
Secrets of the Saqqara Tomb
ReMastered: The Two Killings of Sam Cooke
ReMastered: Massacre at the Stadium
Raat Akeli Hai
On My Skin
Okja
Octonauts & the Great Barrier Reef
Mucho Mucho Amor: The Legend of Walter Mercado
I'm No Longer Here
Get Me Roger Stone
Father Soldier Son
Extremis
El Camino: A Breaking Bad Movie
Dolemite Is My Name
Zion
The Siege of Jadotville
The Road to El Camino: A Breaking Bad Movie
The King
The Edge of Democracy
The Disciple
0
20
40
60
80
100
120
140
160


SCATTER PLOT FILM DENGAN RATING DIBAWAH 7.5
IMDb Score
Runtime
7.40
7.35
7.30
7.25
7.20
7.15
7.10
7.05
7.00
20
40
60
80
100
120
140


SCATTER PLOT FILM DENGAN RATING DIATAS ATAU SAMA DENGAN 7.5
IMDb Score
Runtime
9.0
8.8
8.6
8.4
8.2
8.0
7.8
7.6
25
50
75
100
125
150
175
200

Whole-part relationship sendiri merupakan data yang menunjukkan proporsi bagian-bagian pada suatu variabel dari keseluruhannya. Beberapa contoh data yang memakai whole-part relationship antara lain diagram pie chart dan staked bar chart.

- Diagram pie chart

Pie chart merupakan diagram yang menunjukkan proporsi/persentase dari kategori-kategori dalam suatu variabel

Pada program python dari kelompok kami, ditampilkan dua jenis diagram pie chart. Berikut penjelasan-penjelasan detailnya:

Pada diagram pie chart yang pertama, kelompok kami menampilkan diagram dengan deskripsi sebagai berikut:

Legenda: IMDb Score (menunjukkan persebaran nilai pada diagram pie)

Judul: PIE CHART IMDB SCORE

Informasi yang didapatkan dari diagram ini adalah persebaran nilai IMDb Score pada film yang ada di file csvnya dimana pada diagram ini didominasi oleh nilai 7.0-7.3 dan yang terkecil yaitu nilai 9.0 dimana hanya terdapat 1 film saja yang mendapatkan nilai ini.

Pada diagram pie chart yang kedua, kelompok kami menampilkan diagram dengan deskripsi sebagai berikut:

Kondisional: Runtime (durasi film) dibawah 60 menit

Legenda: Genre (menunjukkan persebaran Genre pada diagram pie)

Judul: PIE CHART GENRE FILM YANG WAKTUNYA DIBAWAH 60 MENIT

Informasi yang didapatkan dari diagram ini adalah persebaran genre pada film yang memiliki durasi waktu kurang dari 60 menit dengan genre sesuai yang ada pada file csvnya dimana pada diagram ini didominasi oleh genre Documentary dan terdapat 7 genre lainnya yang tidak terlalu mendominasi.

- Stacked Bar chart

Stacked bar chart merupakan grafik yang menampilkan bagian-bagian dari total nilai untuk suatu kategori dalam satu bar/batang

Berikut penjelasan detail dari stacked bar chart yang ada program python dari kelompok kami:

Jenis stacked bar: barh (bar horizontal)

Sumbu x: title (judul film)

Sumbu y: IMDb Score, Runtime

Legenda: IMDb Score, Runtime (menunjukkan panjang nilai diagram dari gabungan nilai IMDb Score dan Runtime)

Judul: STACKED BAR IMDB SCORE DAN DURASI FILM

Informasi yang didapatkan dari grafik ini adalah gabungan dari nilai IMDb Score dan juga Runtime (durasi film). Pada grafik, nilai IMDb Score ditunjukkan oleh warna biru dan nilai Runtime ditunjukkan oleh warna oranye, untuk nilai IMDb Score dikarenakan nilai yang ada berdekatan semua antara satu film dengan yang lain maka perbedaan yang ada tidak terlalu signifikan terlihat. Sedangkan untuk Runtime dikarenakan terdapat banyak film dengan durasi yang berbeda-beda, maka grafik yang ada juga menjadi mencolok perbedaannya.

Plotting relationships

- Scatter Plot

Scatter plot merupakan grafik yang terdiri atas titik-titik nilai yang dipetakan di atas koordinat x dan y yang merepresentasikan nilai dari 2 variable

Pada program python dari kelompok kami, ditampilkan dua jenis scatter plot. Berikut penjelasan-penjelasan detailnya:

Scatter plot yang pertama merupakan scatter plot yang menunjukkan perbandingan antara nilai sumbu x yaitu runtime dan sumbu y yaitu IMDb Score. Untuk scatter yang pertama, terdapat kondisional yaitu scatter plotnya hanya menunjukkan film dengan rating IMDb Score dibawah 7.5 sehingga judul dari grafik ini adalah SCATTER PLOT FILM DENGAN RATING DIBAWAH 7.5

Scatter plot yang pertama merupakan scatter plot yang menunjukkan perbandingan antara nilai sumbu x yaitu runtime dan sumbu y yaitu IMDb Score. Untuk scatter yang pertama, terdapat kondisional yaitu scatter plotnya hanya menunjukkan film dengan rating IMDb Score dibawah 7.5 sehingga judul dari grafik ini adalah SCATTER PLOT FILM DENGAN RATING DIBAWAH 7.5. Selain itu, pewarnaan dari titik plot juga diatur dalam program ini dimana program kami menggunakan warna hijau untuk titik-titiknya dengan fungsi color=green

Informasi yang didapatkan dari grafik ini adalah menunjukkan persebaran film berdasarkan Runtimenya terhadap rating dari IMDb Scorenya sehingga dapat dipakai untuk melakukan analisa korelasi grafik

Scatter plot yang kedua merupakan scatter plot yang menunjukkan perbandingan antara nilai sumbu x yaitu runtime dan sumbu y yaitu IMDb Score. Untuk scatter yang kedua, terdapat kondisional yaitu scatter plotnya hanya menunjukkan film dengan rating IMDb Score lebih dari sama dengan 7.5 sehingga judul dari grafik ini adalah SCATTER PLOT FILM DENGAN RATING DIATAS ATAU SAMA DENGAN 7.5. Selain itu, pewarnaan dari titik plot juga diatur dalam program ini dimana program kami menggunakan warna ungu untuk titik-titiknya dengan fungsi color=purple

Informasi yang didapatkan dari grafik ini adalah menunjukkan persebaran film berdasarkan Runtimenya terhadap rating dari IMDb Scorenya sehingga dapat dipakai untuk melakukan analisa korelasi grafik

TUGAS 7

A. Korelasi

```
1  import pandas as pd
2  import matplotlib.pyplot as plt
3  df=pd.read_csv("Netflixoriginal.csv")
4  df.plot(kind="scatter", x="Runtime", y="IMDb Score",title="SCATTER PLOT SEMUA FILM",color="purple")
5  print(df["IMDb Score"].corr(df["Runtime"]))
6  plt.show()
7  print()
```

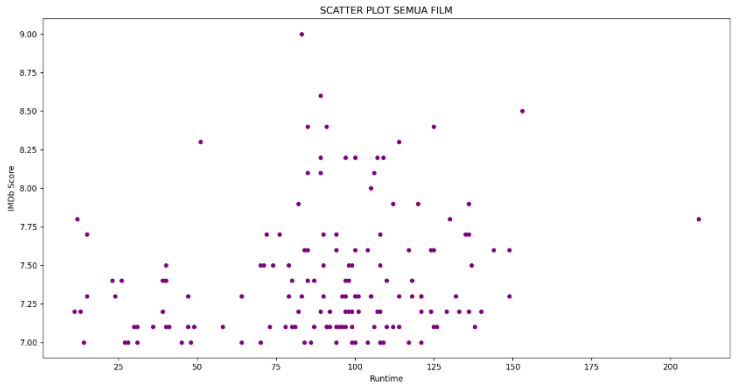
Penjelasan: Dalam tugas 7 membuat korelasi antar data kuantitatif, dalam data set yang kami punya hanya memiliki dua data kuantitatif, yaitu runtime/durasi dan IMDb score. Langkah pertama yaitu kita import panda dan inport matplotlib.pyplot. lalu, kami plot x menjadi runtime/durasi dan y menjadi IMDb scorenya. Kemudian kami print IMDb score dan runtime.

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL

Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

PS C:\Users\hp\OneDrive\Documents\ITB\PENGKOMP\tugas 7> & "C:/Program Files/Python310/python.exe" "c:/Users/hp/OneDrive/Documents/ITB/PENGKOMP/tugas
 7/netflixtugas7.py"
0.1840353010290945
```



Penjelasan: Berdasarkan grafik scatter plot dari data ini, grafik menunjukkan bahwa titik cenderung berada dalam posisi yang acak dan tidak membentuk suatu garis sama sekali yang menunjukkan kenaikan atau penurunan pada data. Selain itu, correlative coefficient yang dihasilkan antara hubungan IMDb Score dengan Runtime hanya sekitar 0,18 yang berarti koefisien sudah hampir mendekati angka 0. Berdasarkan data scatter plot dan angka koefisien yang dihasilkan, dapat disimpulkan bahwa antara grafik IMDb Score dengan grafik Runtime tidak ada hubungan satu sama lain atau No Correlation. Artinya, kenaikan IMDb Score tidak akan berpengaruh sama sekali dengan kenaikan atau penurunan yang ada pada Runtime Film

Kesimpulan

Proses menganalisis data menambah pemahaman tentang fungsi dari *library* yang ada pada python. *Library* pandas untuk import file csv, eksplorasi data dan data pre-processing menggunakan library NumPy. Selain itu, visualisasi persebaran data menggunakan *library* Matplotlib.

Referensi

https://cdn-edunex.itb.ac.id/31889-Introduction-to-Computation/59129-Data-Visualization/1636353376894_KU1102_DA_3_PY_VisualisasiData.pdf

https://cdn-edunex.itb.ac.id/31889-Introduction-to-Computation/59126-Descriptive-Analytics--Visualization/1636352402157_KU1102_DA_2_PY_DescriptiveAnalyticsStatistics.pdf

https://youtu.be/jXkfKHztJ7A

Link Video Presentasi:
https://drive.google.com/file/d/1ljel2jiN2d4C2POium9ZQki8GzwfEOC3/view?usp=sharing

PEMBAGIAN TUGAS

- Tugas 3: Kartini Copa - 16521489
- Tugas 4: Eleanora Felicia - 16521491
- Tugas 5.1: Kartini Copa - 16521489
- Tugas 5.2: Syasya Umaira - 16521537
- Tugas 6.1: Eleanora Felicia – 16521491
- Tugas 6.2: Jason Rivalino - 16521541
- Tugas 7: Syasya Umaira - 16521537