

LAPORAN SISTEM BASIS DATA KELOMPOK 10



Nama Anggota :

- | | |
|------------------------------|-----------------|
| 1. Ahmad Tajali | (2111016110009) |
| 2. Annisa Salsabila Ahdyani | (2111016220017) |
| 3. Muhammad Alfitrian Ananda | (2111016210013) |
| 4. Muhamad Michael Febrian | (2111016210012) |
| 5. Winda Agustina | (2111016320015) |

Dosen Pembimbing :
Muliadi S.Kom, M.Sc.

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
2022**

DATA WAREHOUSE

1. Definisi Data Warehouse

Data warehouse adalah sebuah database penunjang keputusan yang mengandung data yang biasanya mewakili sejarah bisnis dari suatu perusahaan. Data Historis dari data warehouse digunakan di dalam aktivitas analisis yang mendukung pengambilan keputusan dalam perusahaan tersebut. Pengelolaan data warehouse dilakukan secara terpisah dari database operasional perusahaan.

2. Karakteristik Data Warehouse

a. Subject Oriented

Yaitu mengorganisir data menurut subjek dari suatu perusahaan, misalnya konsumen, produk, dan penjualan. Difokuskan pada pemodelan dan analisis data untuk decision maker, bukan pada operasi harian atau pemrosesan transaksi. Menyediakan wawasan yang sederhana dan ringkas mengenai subjek dengan memisahkan data yg tidak relevan dalam proses pengambilan keputusan

b. Integrated

Data warehouse harus mengintegrasikan data dari sumber data yang beragam, seperti relational database, flat files, online transaction records.

c. Time-variant

Data warehouse tetap menyimpan data-data historis. Setiap struktur key dalam data warehouse mengandung elemen waktu baik eksplisit maupun implisit.

d. Non-volatile

Penyimpanan data transformasi dalam data warehouse selalu terpisah secara fisik dari lingkungan operasional. Oleh karena itu, updatedata operasional tidak terjadi pada lingkungan data warehouse, dan data warehouse tidak memerlukan pemrosesan transaksi, recovery, dan concurrency control. Hanya memerlukan dua operasi dalam mengakses data, yaitu initial loading of data dan access of data.

3. Contoh Data Warehouse pada Berbagai Industri

- 1) Pada perusahaan Retail yang menggunakan data warehouse untuk melacak pengiriman, memprediksi perilaku konsumen.
- 2) Penerbangan, memeriksa rute penerbangan yang memungkinkan.

- 3) Industri Perbankan yang melakukan analisis performa, mengukur efisiensi antrian.
- 4) Pusat Kesehatan, menentukan perawatan pasien, memprediksi hasil kesehatan.
- 5) Sektor Publik, memeriksa catatan pajak, fasilitas kesehatan.

4. Keuntungan Data Warehouse

- a. Kemungkinan kembalinya nilai investasi dalam jumlah besar
Perusahaan harus menanamkan modal dalam jumlah yang cukup besar untuk menjamin penerapan data warehouse pada perusahaan tersebut dapat berhasil. Namun penelitian yang dilakukan oleh International Data Corporation (IDC) pada tahun 1996 menunjukkan bahwa dalam jangka waktu kurang lebih tiga tahun nilai Return of Investment (ROI), dari perusahaan yang menerapkan data warehouse, meningkat sebesar 401%.
- b. Keunggulan bersaing (competitive advantage)
Besarnya ROI untuk perusahaan yang mengimplementasikan data warehouse merupakan bukti akan keunggulan bersaing yang luar biasa sebagai efek dari teknologi ini. Keunggulan bersaing diperoleh karena decision maker dapat mengakses data yang sebelumnya tidak tampak, tidak diketahui dan belum dimanfaatkan, seperti permintaan konsumen, tren pada konsumen.
- c. Meningkatkan produktivitas pembuat keputusan
Data warehouse meningkatkan produktivitas decision maker dengan menciptakan database yang konsisten, subject-oriented, dan mendukung data historis. Dengan kata lain, data warehouse mengintegrasikan data dari berbagai sistem yang bertentangan menjadi suatu bentuk yang menyediakan gambaran perusahaan yang konsisten

BUSINESS INTELLIGENCE

1. Pengertian Business Intelligence

Business Intelligence adalah konsep mengumpulkan data, menyimpan data dan menyeleksi data guna menyediakan informasi untuk membantu dan meningkatkan kualitas pengambilan keputusan bisnis bagi perusahaan. Sehingga dapat dikatakan bahwa sistem informasi adalah sebagai tempat entry data, sedangkan aplikasi business intelligence sebagai tempat analisa data.

2. Contoh Business Intelligence

Salah satu contohnya yaitu pengembangan sistem informasi penjualan barang pada suatu perusahaan menggunakan konsep BI. Business intelligence tools untuk monitoring penjualan pada perusahaan akan terbagi menjadi 2 halaman utama berdasarkan role user yang berinteraksi dengan sistem, yaitu pimpinan perusahaan dan admin perusahaan, dimana setiap user memiliki menu sesuai tanggung jawabnya masing-masing.

Berikut contoh lain menurut tahun perkembangannya

- 1) Tahun 1980 : Printed Green Bar Reports

Komputasi berbasis Sistem di Komputer, perpaduan dan Pengembangan dari SIM (Sistem Informasi Manajemen) dengan Riset Operasi. Digunakan dalam sistem pendukung keputusan, sistem untuk mencetak laporan kafetaria hijau dari printer dot matrix.

- 2) Tahun 1990 : The OLAP/OLTP Breakthrough

BI Diarahkan dan Dikembangkan ke Metode OLAP dan OLTP. Digunakan dalam rancangan data warehouse, Sistem OLTP yang memiliki berbagai database untuk keperluan bisnis.

- 3) Tahun 2000 : The Rise Of The Web

BI berbasis web pengguna/ web user yang mempunyai saluran informasi secara real-time. Digunakan dalam Analisis berbasis prediksi faktual, Laporan berbasis web yang memuat diagram serta grafik, penggabungan dan peringkasan data laporan lebih efisien.

- 4) Tahun 2010 : Portable and Swift

BI berbasis mobile. Mengintegrasikan teknologi terdahulu ke dalam perangkat mobile. Digunakan dalam Laporan bisnis berbasis perangkat mobile, Visualisasi data untuk perangkat mobile.

3. Manfaat Business Intelligence

- a. Meningkatkan nilai data dan informasi organisasi

Seluruh data dan informasi dapat diintegrasikan sedemikian rupa sehingga menghasilkan pengambilan keputusan yang lengkap. Informasi-informasi yang dulunya tidak dicakupkan sebagai salah satu faktor pengambilan keputusan dapat dengan mudah dilakukan dengan menggunakan BI. Data dan informasi yang dihasilkan pun juga menjadi lebih mudah diakses dimengerti (user friendly).

- b. Memudahkan pemantauan kinerja organisasi

Dalam mengukur kinerja suatu organisasi, seringkali dipergunakan ukuran yang disebut Key Performance Indicator (KPI). KPI tidak selalu diukur dengan satuan uang, namun dapat juga berdasarkan kecepatan pelaksanaan suatu layanan. BI dapat dengan mudah menunjukkan pencapaian KPI suatu organisasi dengan mudah, cepat, dan tepat.

- c. Meningkatkan nilai investasi teknologi informasi yang sudah ada

Business Intelligence (BI) tidak selalu harus mengubah atau menggantikan sistem informasi yang sudah digunakan sebelumnya. Sebaliknya, BI hanya menambahkan layanan pada sistem-sistem tersebut sehingga data dan informasi yang sudah ada dapat menghasilkan informasi yang komprehensif dan memiliki kegunaan yang lebih baik.

- d. Menciptakan pegawai yang memiliki akses informasi yang baik (well-informed workers)

Dalam melaksanakan pekerjaannya sehari-hari, seluruh level dari suatu organisasi (mulai dari pegawai/bawahan sampai dengan pimpinan) selalu berkaitan dan atau membutuhkan akses data dan informasi. Business Intelligence (BI) mempermudah seluruh level pegawai dalam mengakses data dan informasi yang diperlukan sehingga membantu membuat suatu keputusan.

- e. Meningkatkan efisiensi biaya

Business Intelligence (BI) dapat meningkatkan efisiensi karena mempermudah seseorang dalam melakukan pekerjaan, hemat waktu, dan mudah pemanfaatannya. Waktu yang dibutuhkan untuk mencari data dan mendapatkan informasi yang dibutuhkan semakin singkat dan cara untuk mendapatkannya pun tidak memerlukan pengetahuan (training) yang rumit.

STUDI KASUS

FIFA WORLD CUP 2022

A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69																															

Ini adalah contoh data warehouse yang belum diolah, bisa dilihat di file excel *Fifa World Cup 2022.csv*

Untuk mengolah data tersebut kami menggunakan Python, berikut source code dan outputnya

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
sns.set(color_codes=True)
```

```
datas = pd.read_csv('Fifa Worldcup 2022.csv')
datas.head ()
```

	S1. No	Match No.	Team	Against	Group	Goal	Possession (%)	Inside Penalty Area	Outside Penalty Area	Assists	...	Fouls Against	Offsides	Passes	Passes Completed	Crosses	Crosses Completed	Corners	Free Kicks	Penalties Scored	Pts
0	1	1	Qatar	Ecuador	A	0	40	0	0	0	...	15	3	453	387	10	5	1	19	0	0
1	2	1	Ecuador	Qatar	A	2	46	2	0	1	...	15	4	484	419	26	10	3	17	1	3
2	3	2	England	Iran	B	6	69	6	0	6	...	9	2	810	733	29	9	8	16	0	3
3	4	2	Iran	England	B	2	20	2	0	1	...	14	2	232	156	11	3	0	10	1	0
4	5	3	Senegal	Netherlands	A	0	39	0	0	0	...	13	2	391	326	22	8	6	14	0	0

5 rows x 38 columns

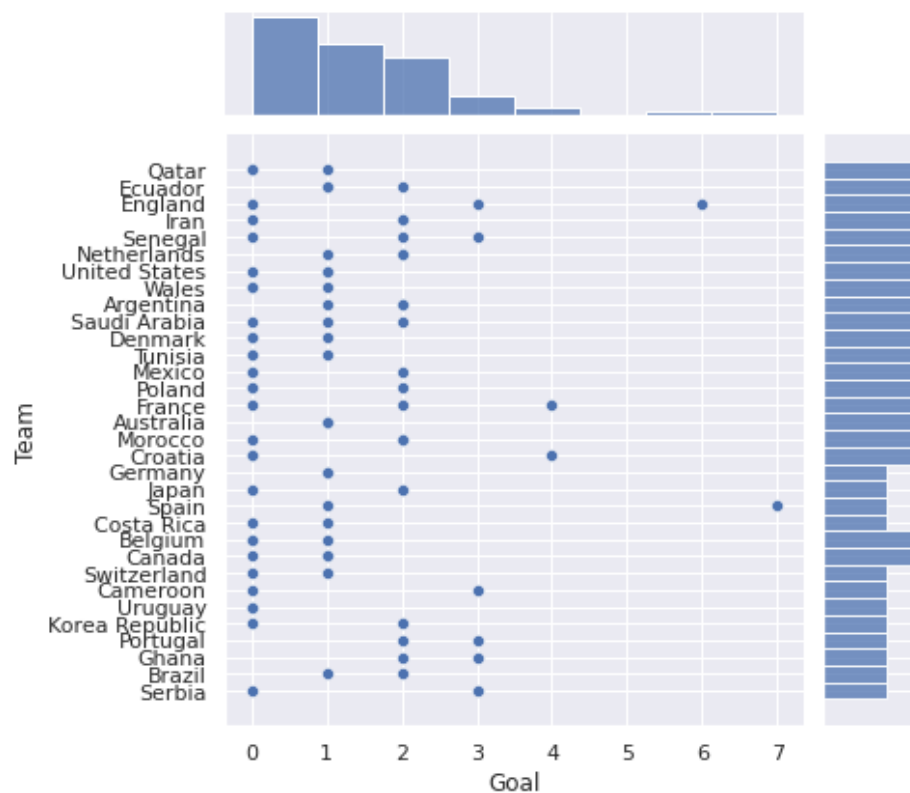
Pertama buat variabel dan masukkan file excel tadi ke dalam python, maksudnya kita import terlebih dahulu, dan seperti gambar diatas beginilah outputnya.

```
[30] ## Goal dari Semua Team
sns.jointplot(x=datas['Goal'],y=datas['Team'])
## Asisst Dari Semua Team
sns.jointplot(x=datas['Assists'],y=datas['Team'])
```

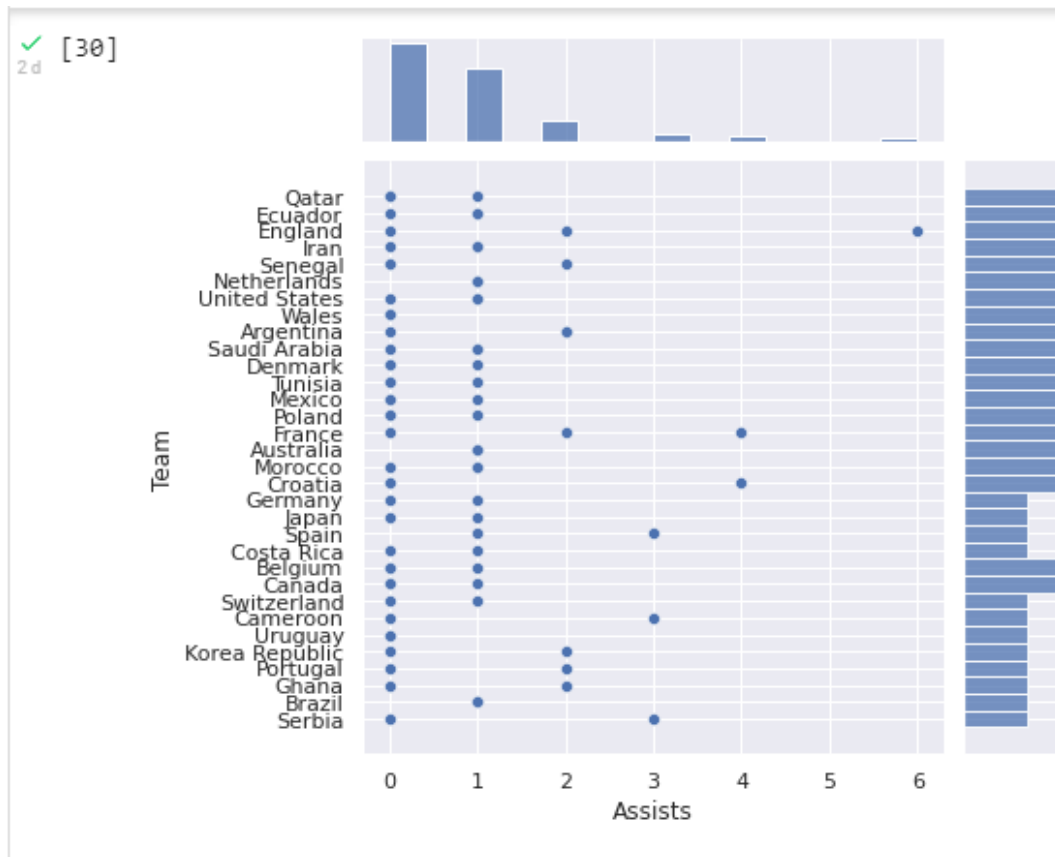
Lalu selanjutnya source code untuk menampilkan grafik data “Goal” dan “Assists” dari semua Team yang ada dan berikut output grafik datanya

✓ [30] <seaborn.axisgrid.JointGrid at 0x7f964c2e0190>

2d



Data Goal



Data Assists

✓ [33] 4 d

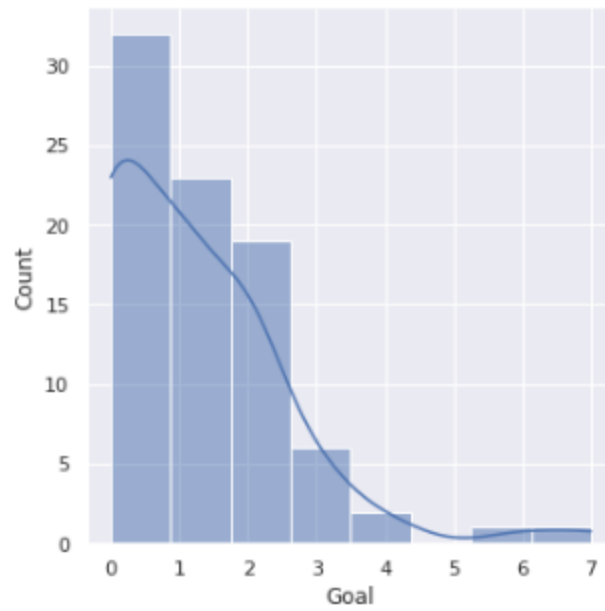
```

## Penyebaran Goal & Assist
sns.displot(datas.Goal, kde=True)
sns.displot(datas.Assists, kde=True)

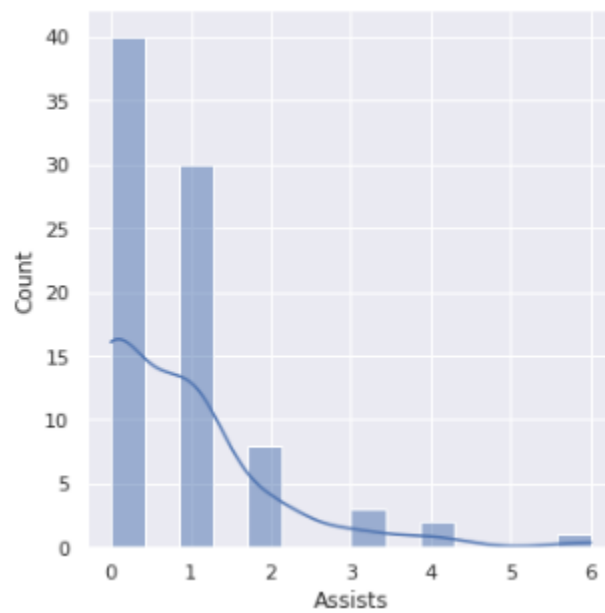
## Goal By Pinalty
sns.displot(datas['Penalties Scored'],)

```

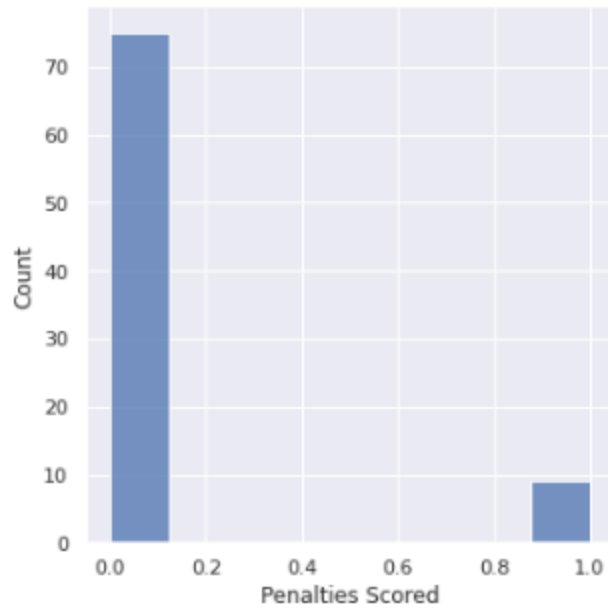
Lalu ini untuk menampilkan penyebaran dari data “Goal” dan “Assists” serta data “Goal” dari “Penalti”



Grafik Data Goal



Grafik Data Assists



Grafik Data Goal dari penalti

Lalu untuk pengambilan data dari banyaknya kesempatan dan goal yang tercipta, dibawah source code nya

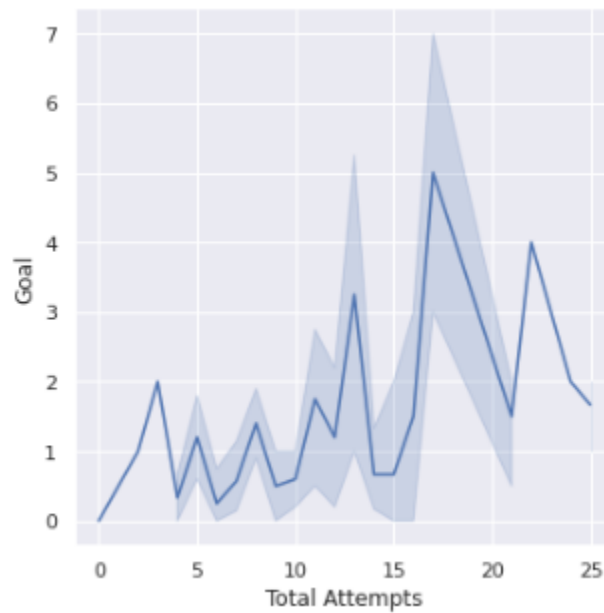
```
## Banyaknya Attempts dan Goal yang tercipta
sns.jointplot(x=datas['Total Attempts'],y=datas['Goal'],
              kind="reg", truncate=False,
              color="m")

## By Line
sns.relplot(data=datas, x="Total Attempts", y="Goal", kind='line')
```

<seaborn.axisgrid.FacetGrid at 0x7f964bdde610>



Grafik data goal dari total kesempatan/peluang yang tercipta



Grafik Line data goal dari total kesempatan/peluang yang tercipta

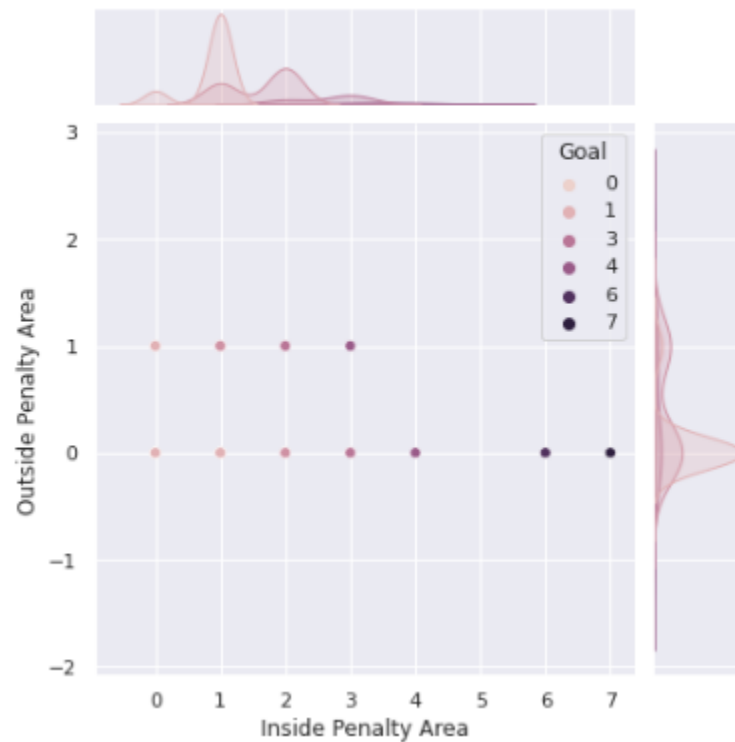
Lalu menampilkan grafik data dari banyaknya gol yang tercipta pada kotak gawang/penalti area “inside” dan “outside”nya, dibawah source codenya

```
2d ✓ ## Grafik banyaknya gol yang tercipta di inside dan outside
sns.jointplot(x=datas['Inside Penalty Area'],y=datas['Outside Penalty Area'], hue=datas['Goal'])

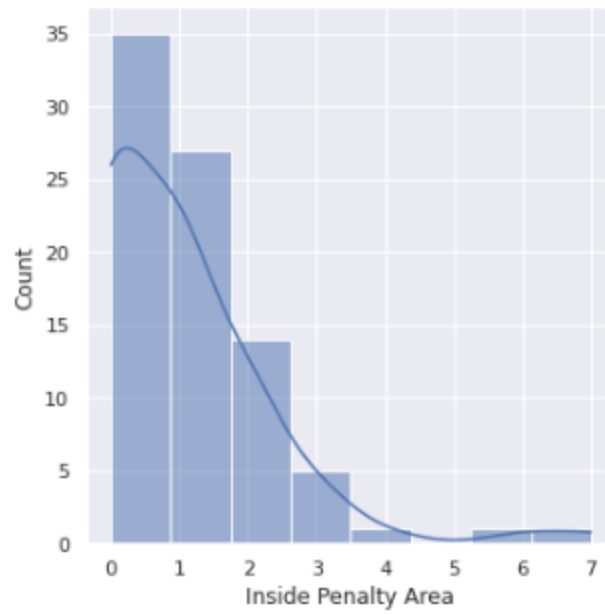
## Inside
sns.displot((datas['Inside Penalty Area']), kde=True)

## Outside
sns.displot((datas['Outside Penalty Area']), kde=True)
```

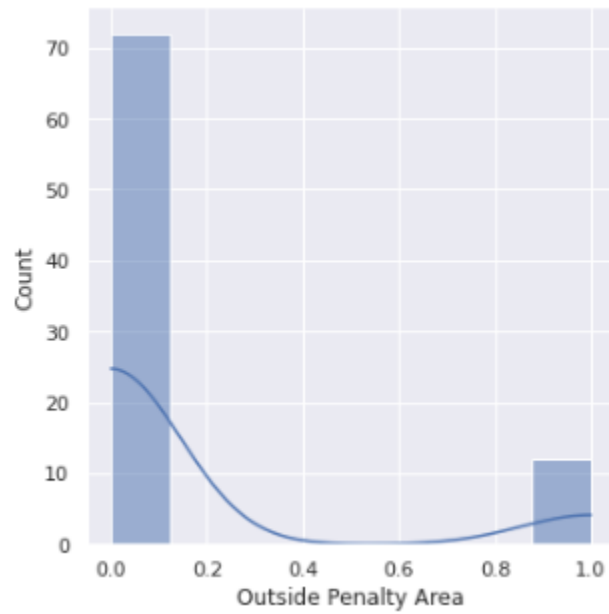
<seaborn.axisgrid.FacetGrid at 0x7f964ead1130>



Grafik data inside dan outside penalty area/kotak penalti




Grafik data inside

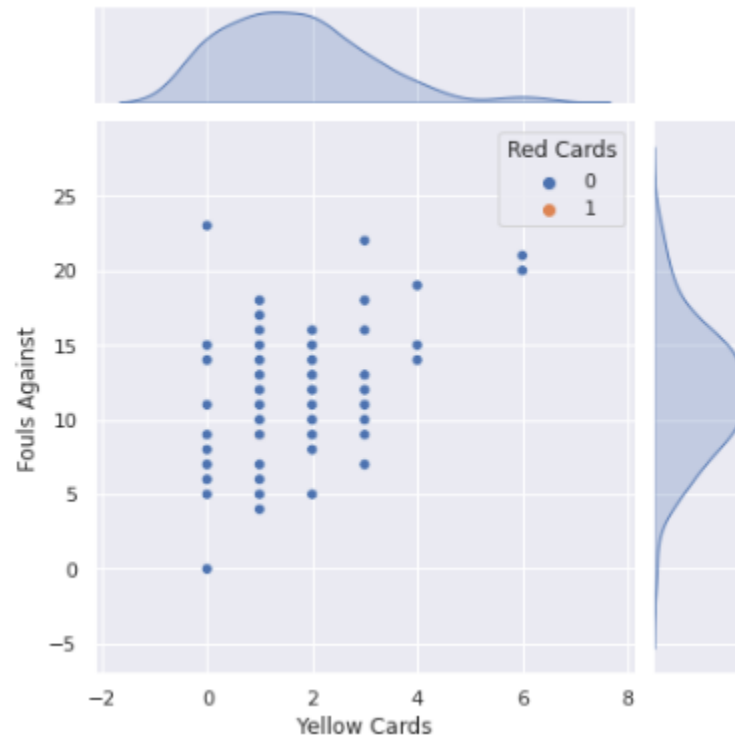


Grafik data Outside

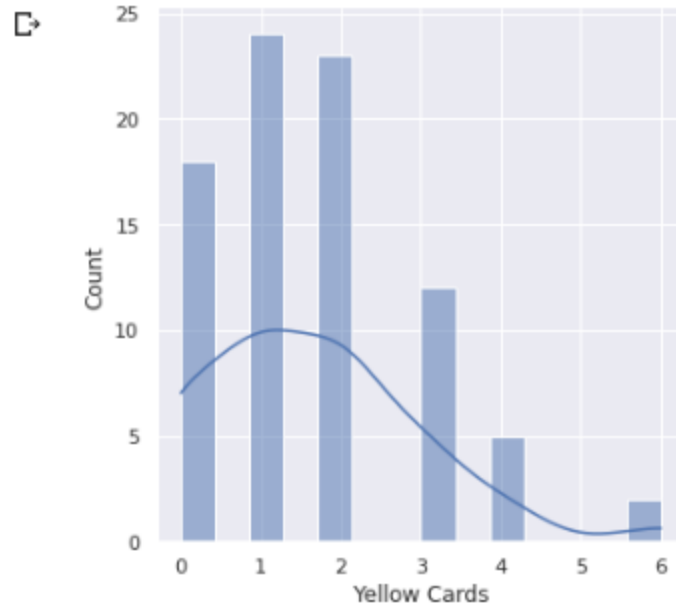
Menampilkan foul atau pelanggaran yang terjadi pada semua team yaitu ada kartu kuning, merah dan offsides, dibawah source codenya

```
3d ✓  ## Banyak Foul yang terjadi yang menghasilkan merah dan kuning  
sns.jointplot(x=(datas['Yellow Cards']),y=datas['Fouls Against'], hue=datas['Red Cards'])  
  
## Yellow Cards  
sns.displot((datas['Yellow Cards']), kde=True)  
## Red Cards  
sns.displot((datas['Red Cards']), kde=True)  
## Offside  
sns.displot((datas['Offsides']), kde=True)
```

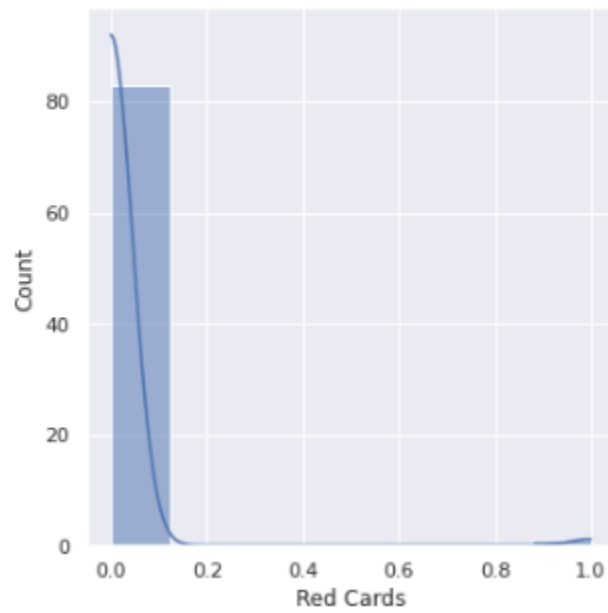
<seaborn.axisgrid.FacetGrid at 0x7f964bb6bc10>



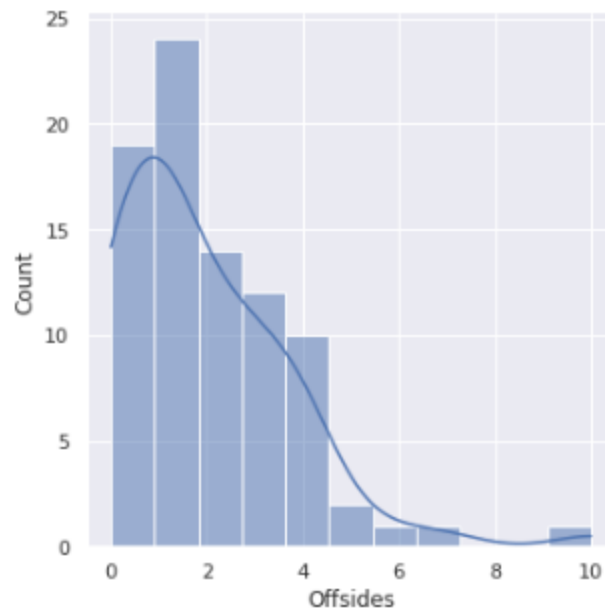
Grafik data pelanggaran / Fouls



Pelanggaran Kartu Kuning



Pelanggaran Kartu Merah




Pelanggaran OFFSIDES

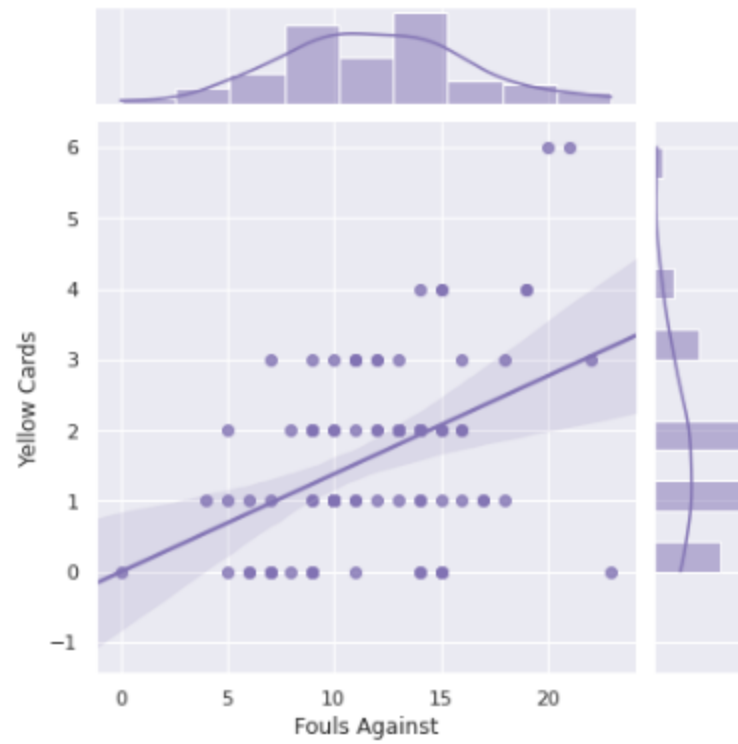
Biasanya di permainan bola, bila melakukan pelanggaran kartu kuning atau merah maka akan mengurangi points, berikut kita coba menampilkan grafiknya, dibawah source codenya

```
✓ 2 d ▶ from pandas.io.formats import style

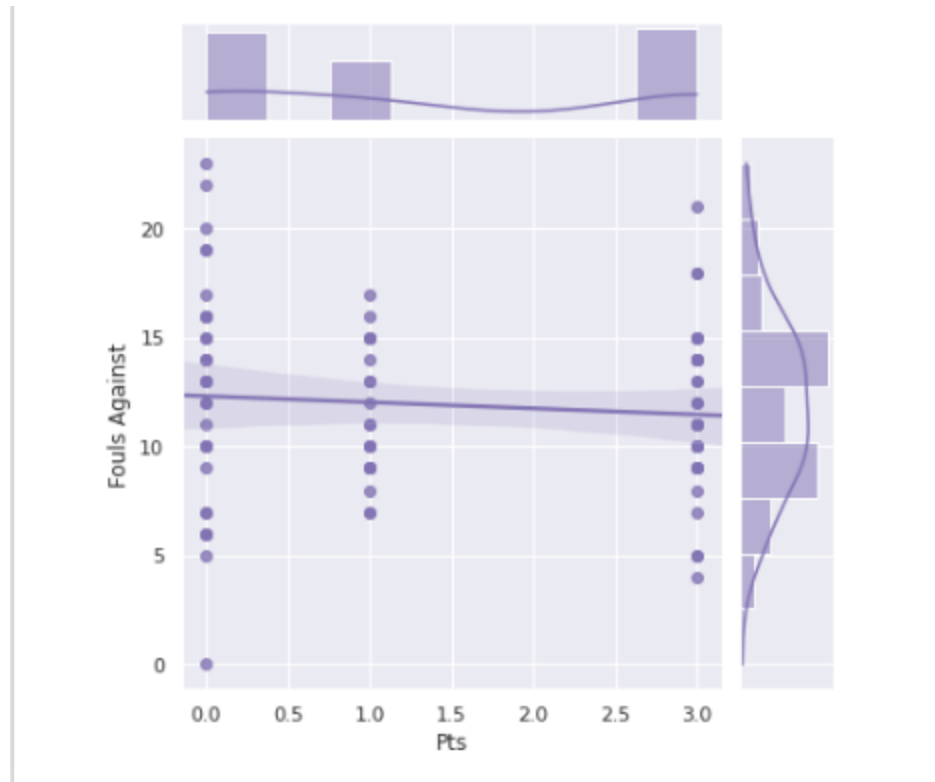
## Foul yang menghasilkan Yellow dan Red
sns.jointplot(x=(datas['Fouls Against']),y=datas['Yellow Cards'],
              kind="reg", truncate=False,
              color="m")

## Foul terhadap Point
sns.jointplot(x=(datas['Pts']),y=datas['Fouls Against'],
              kind="reg", truncate=False,
              color="m")
```

 <seaborn.axisgrid.JointGrid at 0x7f964b951550>



Ini grafik datanya



Ini data points nya

Kita juga bisa melihat pengelompokkan goal yang ada di setiap group, dibawah source codenya

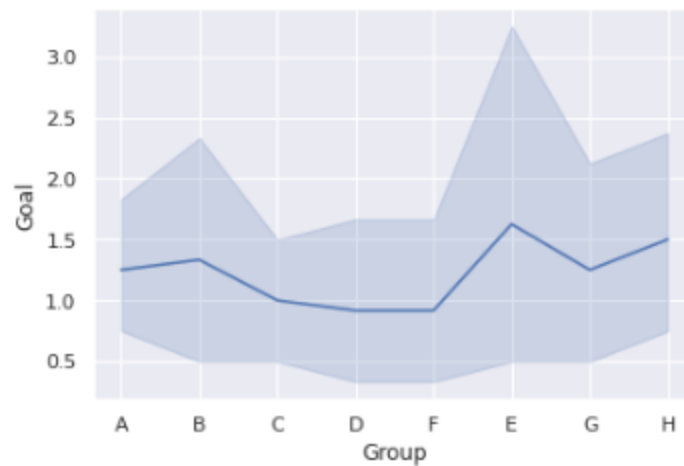
```

✓ 1 d # Plot
sns.set_theme(style="darkgrid")

# Plot the responses for different events and regions
sns.lineplot(x=(datas['Group']),y=datas['Goal'])

```

<matplotlib.axes._subplots.AxesSubplot at 0x7f964b75c250>



Ini grafik line datanya.

MACHINE LEARNING : FIFA WORLD CUP 2022

Sepakbola memang tidak akan pernah lepas dari prediksi dan kemungkinan, namun penuh dengan ketidakpastian. Tidak ada yang tau tim mana yang akan menang, tidak ada yang tahu siapa yang akan mencetak gol kegawang. Namun semua itulah yang membuat sepak bola menjadi menyenangkan. Disini saya mencoba mem “prediksi” pemenang FIFA WORLD CUP 2022 menggunakan machine learning.

Data yang digunakan adalah riwayat semua negara yang ikut piala dunia dari tahun 1930 - 2018, lalu data masing-masing group dan masing-masing pertandingan pada fase group untuk nantinya menentukan pemenang dan runner up group dan berlanjut sampai final.

Karena keterbatasan data dan ingin membuat machine learning yang sederhana, saya memutuskan untuk menggunakan distribusi Poisson. Distribusi Poisson adalah distribusi probabilitas diskrit yang menggambarkan jumlah peristiwa yang terjadi dalam interval waktu atau wilayah peluang yang tetap. Jika kita menganggap sebuah gol sebagai peristiwa yang mungkin terjadi dalam 90 menit

pertandingan sepak bola, kita dapat menghitung peluang jumlah gol yang dapat dicetak dalam pertandingan oleh Tim A dan Tim B.

Tapi itu tidak cukup. Kita masih harus memenuhi asumsi distribusi Poisson.

1. Jumlah peristiwa dapat dihitung (pertandingan dapat memiliki 1, 2, 3 gol atau lebih)
2. Terjadinya peristiwa bersifat independen (terjadinya satu tujuan tidak boleh mempengaruhi kemungkinan tujuan lain).
3. Tingkat di mana peristiwa terjadi adalah konstan (probabilitas gol yang terjadi dalam interval waktu tertentu harus persis sama untuk setiap interval waktu lain dengan panjang yang sama)
4. Dua peristiwa tidak dapat terjadi pada waktu yang sama persis (dua tujuan tidak dapat terjadi pada waktu yang sama)

Tanpa ragu asumsi 1 dan 4 terpenuhi, tetapi 2 dan 3 sebagian benar. Yang mengatakan, mari kita asumsikan bahwa asumsi 2 dan 3 selalu benar.

Rumus distribusi Poisson :

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Lambda : median gol dalam 90 menit (Tim A dan Tim B)

x : jumlah gol dalam pertandingan yang bisa dicetak oleh Tim A dan Tim B

Untuk menghitung lambda, kita membutuhkan rata-rata gol yang dicetak/dibobol oleh masing-masing tim nasional.

A. Melakukan import untuk library dan memasukkannya ke variabel berikut:

```
In [4]: import pandas as pd
import pickle
from scipy.stats import poisson
```

```
In [5]: dict_table = pickle.load(open('dict_table', 'rb'))
df_historical_data = pd.read_csv('clean_fifa_worldcup_matches.csv')
df_fixture = pd.read_csv('clean_fifa_worldcup_fixture.csv')
```

```
display(dict_table)

{'Group A': Pos      Team Pld W D L GF GA GD Pts
0 1 Qatar (H)      0 0 0 0 0 0 0 0
1 2 Ecuador      0 0 0 0 0 0 0 0
2 3 Senegal      0 0 0 0 0 0 0 0
3 4 Netherlands  0 0 0 0 0 0 0 0,
'Group B': Pos      Team Pld W D L GF GA GD Pts
0 1 England      0 0 0 0 0 0 0 0
1 2 Iran          0 0 0 0 0 0 0 0
2 3 United States 0 0 0 0 0 0 0 0
3 4 Wales        0 0 0 0 0 0 0 0,
'Group C': Pos      Team Pld W D L GF GA GD Pts
0 1 Argentina    0 0 0 0 0 0 0 0
1 2 Saudi Arabia 0 0 0 0 0 0 0 0
2 3 Mexico       0 0 0 0 0 0 0 0
3 4 Poland       0 0 0 0 0 0 0 0,
'Group D': Pos      Team Pld W D L GF GA GD Pts
0 1 France       0 0 0 0 0 0 0 0
1 2 Australia    0 0 0 0 0 0 0 0
2 3 Denmark      0 0 0 0 0 0 0 0
3 4 Tunisia      0 0 0 0 0 0 0 0,
'Group E': Pos      Team Pld W D L GF GA GD Pts
0 1 Spain        0 0 0 0 0 0 0 0
1 2 Costa Rica   0 0 0 0 0 0 0 0
2 3 Germany      0 0 0 0 0 0 0 0
3 4 Japan        0 0 0 0 0 0 0 0,
'Group F': Pos      Team Pld W D L GF GA GD Pts
0 1 Belgium      0 0 0 0 0 0 0 0
1 2 Canada       0 0 0 0 0 0 0 0
2 3 Morocco      0 0 0 0 0 0 0 0
3 4 Croatia      0 0 0 0 0 0 0 0,
'Group G': Pos      Team Pld W D L GF GA GD Pts
0 1 Brazil       0 0 0 0 0 0 0 0
1 2 Serbia       0 0 0 0 0 0 0 0
2 3 Switzerland  0 0 0 0 0 0 0 0
3 4 Cameroon     0 0 0 0 0 0 0 0,
'Group H': Pos      Team Pld W D L GF GA GD Pts
0 1 Portugal     0 0 0 0 0 0 0 0
1 2 Ghana        0 0 0 0 0 0 0 0
2 3 Uruguay      0 0 0 0 0 0 0 0
3 4 South Korea  0 0 0 0 0 0 0 0}]
```

Data tabel group

```
display(df_historical_data)
```

	HomeTeam	AwayTeam	Year	HomeGoals	AwayGoals	TotalGoals
0	France	Mexico	1930	4	1	5
1	Uruguay	Argentina	1930	4	2	6
2	Uruguay	Yugoslavia	1930	6	1	7
3	Argentina	United States	1930	6	1	7
4	Paraguay	Belgium	1930	1	0	1
...
895	Brazil	Costa Rica	2018	2	0	2
896	Serbia	Switzerland	2018	1	2	3
897	Serbia	Brazil	2018	0	2	2
898	France	Peru	2018	1	0	1
899	Brazil	Belgium	2018	1	2	3

900 rows x 6 columns

Data histori piala dunia

```
display(df_fixture)
```

	home	score	away	year
0	Qatar	Match 1	Ecuador	2022
1	Senegal	Match 2	Netherlands	2022
2	Qatar	Match 18	Senegal	2022
3	Netherlands	Match 19	Ecuador	2022
4	Ecuador	Match 35	Senegal	2022
...
59	Winners Match 51	Match 59	Winners Match 52	2022
60	Winners Match 57	Match 61	Winners Match 58	2022
61	Winners Match 59	Match 62	Winners Match 60	2022
62	Losers Match 61	Match 63	Losers Match 62	2022
63	Winners Match 61	Match 64	Winners Match 62	2022

64 rows x 4 columns

Data pertandingan piala dunia 2022

B. Mengolah histori data piala dunia

Perhitungan kekuatan tim

Pisah df jadi df_home dan df_away

```
df_home = df_historical_data[['HomeTeam', 'HomeGoals', 'AwayGoals']]
df_away = df_historical_data[['AwayTeam', 'HomeGoals', 'AwayGoals']]
```

```
df_home = df_home.rename(columns={'HomeTeam': 'Team', 'HomeGoals': 'GoalsScored', 'AwayGoals': 'GoalsConceded'})
df_away = df_away.rename(columns={'AwayTeam': 'Team', 'HomeGoals': 'GoalsConceded', 'AwayGoals': 'GoalsScored'})
```

```
df_team_strength = pd.concat([df_home, df_away], ignore_index=True).groupby('Team').mean()
df_team_strength
```

	GoalsScored	GoalsConceded
Team		
Algeria	1.000000	1.461538
Angola	0.333333	0.666667
Argentina	1.691358	1.148148
Australia	0.812500	1.937500
Austria	1.482759	1.620690
...
Uruguay	1.553571	1.321429
Wales	0.800000	0.800000
West Germany	2.112903	1.241935
Yugoslavia	1.666667	1.272727
Zaire	0.000000	4.666667

85 rows × 2 columns

C. Membuat fungsi untuk memprediksi pemenang antar pertandingan

Kelas untuk memprediksi point

```
def predict_points(home, away):
    if home in df_team_strength.index and away in df_team_strength.index:
        #goalsScored * goalsConceded
        lamb_home = df_team_strength.at[home, 'GoalsScored'] * df_team_strength.at[away, 'GoalsConceded']
        lamb_away = df_team_strength.at[away, 'GoalsScored'] * df_team_strength.at[home, 'GoalsConceded']
        prob_home, prob_away, prob_draw = 0, 0, 0
        for x in range(0, 11):
            for y in range(0, 11):
                p = poisson.pmf(x, lamb_home) * poisson.pmf(y, lamb_away)
                if x == y:
                    prob_draw += p
                elif x > y:
                    prob_home += p
                else:
                    prob_away += p
        points_home = 3 * prob_home + prob_draw
        points_away = 3 * prob_away + prob_draw
        return (points_home, points_away)
    else:
        return(0, 0)
```

predict_points menghitung berapa banyak poin yang akan didapat tim tuan rumah dan tandang. Untuk melakukannya, saya menghitung lambda untuk setiap tim menggunakan rumus $\text{average_goals_scored} * \text{average_goals_conceded}$.

Kemudian saya mensimulasikan semua kemungkinan skor pertandingan dari 0–0 hingga 10–10 (skor terakhir itu hanyalah batas jangkauan gol saya). Setelah saya memiliki λ dan x , saya menggunakan rumus distribusi Poisson untuk menghitung p .

Prob_home, prob_draw, dan prob_away mengakumulasi nilai p jika, katakanlah, pertandingan berakhir masing-masing dengan skor 1–0 (tuan rumah menang), 1–1 (seri), atau 0–1 (tandang menang). Akhirnya, poin dihitung dengan rumus di bawah ini.

```
points_home = 3* prob_home + prob_draw  
points_away = 3* prob_away + prob_draw
```

Jika kita menggunakan predict_points untuk memprediksi pertandingan Inggris vs Amerika Serikat, kita akan mendapatkan hasil seperti :

```
predict_points('England', 'United States')  
(2.2356147635326007, 0.5922397535606193)
```

Artinya, Inggris mendapat 2,23 poin, sedangkan AS mendapat 0,59. Saya mendapatkan desimal karena menggunakan probabilitas.

Jika kita menerapkan fungsi predict_points ini ke semua pertandingan di babak penyisihan grup, kita akan mendapatkan posisi 1 dan 2 dari setiap grup, sehingga pertandingan berikutnya di babak sistem gugur.

D. Membuat fungsi untuk memprediksi point tiap group

Group Stage

```
df_fixture_group_48 = df_fixture[:48].copy()
df_fixture_knockout = df_fixture[48:56].copy()
df_fixture_quarter = df_fixture[56:60].copy()
df_fixture_semi = df_fixture[60:62].copy()
df_fixture_final = df_fixture[62:].copy()
```

```
display(df_fixture_group_48.head())
```

	home	score	away	year
0	Qatar	Match 1	Ecuador	2022
1	Senegal	Match 2	Netherlands	2022
2	Qatar	Match 18	Senegal	2022
3	Netherlands	Match 19	Ecuador	2022
4	Ecuador	Match 35	Senegal	2022

```
for group in dict_table:
    teams_in_group = dict_table[group]['Team'].values
    df_fixture_group_6 = df_fixture_group_48[df_fixture_group_48['home'].isin(teams_in_group)]
    for index, row in df_fixture_group_6.iterrows():
        home, away = row['home'], row['away']
        points_home, points_away = predict_points(home, away)
        dict_table[group].loc[dict_table[group]['Team'] == home, 'Pts'] += points_home
        dict_table[group].loc[dict_table[group]['Team'] == away, 'Pts'] += points_away
    dict_table[group] = dict_table[group].sort_values('Pts', ascending=False).reset_index()
dict_table[group] = dict_table[group][['Team', 'Pts']]
dict_table[group] = dict_table[group].round(0)
```

```
dict_table['Group H']
```

	Team	Pts
0	Portugal	6.0
1	Uruguay	5.0
2	Ghana	4.0
3	South Korea	2.0

E. Membuat fungsi untuk memprediksi 16 besar

Untuk 16 besar, tidak perlu memprediksi poin, tetapi pemenang setiap braket yang telah diprediksi. Inilah sebabnya mengapa saya membuat fungsi `get_winner` baru berdasarkan fungsi `predict_points` sebelumnya.

16 Besar

```
for group in dict_table:
    group_winner = dict_table[group].loc[0, 'Team']
    runners_up = dict_table[group].loc[1, 'Team']

    df_fixture_knockout.replace({f'Winners {group}': group_winner,
                                f'Runners-up {group}': runners_up}, inplace=True)
df_fixture_knockout['winner'] = '?'
df_fixture_knockout
```

	home	score	away	year	winner
48	Netherlands	Match 49	Wales	2022	?
49	Argentina	Match 50	Denmark	2022	?
50	France	Match 52	Poland	2022	?
51	England	Match 51	Senegal	2022	?
52	Germany	Match 53	Belgium	2022	?
53	Brazil	Match 54	Uruguay	2022	?
54	Croatia	Match 55	Spain	2022	?
55	Portugal	Match 56	Switzerland	2022	?

```
def get_winner(df_fixture_updated):
    for index, row in df_fixture_updated.iterrows():
        home, away = row['home'], row['away']
        points_home, points_away = predict_points(home, away)
        if points_home > points_away:
            winner = home
        else:
            winner = away
        df_fixture_updated.loc[index, 'winner'] = winner
    return df_fixture_updated
```

```
get_winner(df_fixture_knockout)
```

	home	score	away	year	winner
48	Netherlands	Match 49	Wales	2022	Netherlands
49	Argentina	Match 50	Denmark	2022	Argentina
50	France	Match 52	Poland	2022	France
51	England	Match 51	Senegal	2022	England
52	Germany	Match 53	Belgium	2022	Germany
53	Brazil	Match 54	Uruguay	2022	Brazil
54	Croatia	Match 55	Spain	2022	Spain
55	Portugal	Match 56	Switzerland	2022	Portugal

```
df_fixture_quarter
```

	home	score	away	year
56	Winners Match 53	Match 58	Winners Match 54	2022
57	Winners Match 49	Match 57	Winners Match 50	2022
58	Winners Match 55	Match 60	Winners Match 56	2022
59	Winners Match 51	Match 59	Winners Match 52	2022

Sederhananya, jika `points_home` lebih besar dari `points_away` pemenangnya adalah tim tuan rumah, jika tidak, pemenangnya adalah tim tandang.

F. Quarter Final

Quarter Final

```
def update_table(df_fixture_round_1, df_fixture_round_2):
    for index, row in df_fixture_round_1.iterrows():
        winner = df_fixture_round_1.loc[index, 'winner']
        match = df_fixture_round_1.loc[index, 'score']
        df_fixture_round_2.replace({f'Winners {match}': winner}, inplace=True)
    df_fixture_round_2['winner'] = '?'
    return df_fixture_round_2
```

```
update_table(df_fixture_knockout, df_fixture_quarter)
```

	home	score	away	year	winner
56	Germany	Match 58	Brazil	2022	?
57	Netherlands	Match 57	Argentina	2022	?
58	Spain	Match 60	Portugal	2022	?
59	England	Match 59	France	2022	?

```
get_winner(df_fixture_quarter)
```

	home	score	away	year	winner
56	Germany	Match 58	Brazil	2022	Brazil
57	Netherlands	Match 57	Argentina	2022	Netherlands
58	Spain	Match 60	Portugal	2022	Portugal
59	England	Match 59	France	2022	France

G. Semi-Final

Semi Final

```
update_table(df_fixture_quarter, df_fixture_semi)
```

	home	score	away	year	winner
60	Netherlands	Match 61	Brazil	2022	?
61	France	Match 62	Portugal	2022	?

```
get_winner(df_fixture_semi)
```

	home	score	away	year	winner
60	Netherlands	Match 61	Brazil	2022	Brazil
61	France	Match 62	Portugal	2022	France

H. Final

Final

```
update_table(df_fixture_semi, df_fixture_final)
```

	home	score	away	year	winner
62	Losers Match 61	Match 63	Losers Match 62	2022	?
63	Brazil	Match 64	France	2022	?

```
get_winner(df_fixture_final)
```

	home	score	away	year	winner
62	Losers Match 61	Match 63	Losers Match 62	2022	Losers Match 62
63	Brazil	Match 64	France	2022	Brazil

Pemenang Piala Dunia Qatar 2022 adalah Brazil