

ANALISIS GAJI PEKERJAAN SEBAGAI DATA SCIENTIST DARI TAHUN 2020 - 2024 DI SELURUH DUNIA



OLEH AJI SAKTI | FADHLI JAHFAL | RAFID FARHAN | RIYAN PUTRA



LATAR BELAKANG



TINJAUAN PUSTAKA

- **Data science dan perannya dalam di era digital**
- **Konsep dan Teori Gaji dalam Lingkungan Kerja**
- **Peran Python dalam Analisis Data**
- **Perumusan Hipotesis Penelitian**



METODE PENELITIAN



HASIL PENELITIAN DAN PEMBAHASAN

STEP 1 | Import library dan Data set

Import Library

```
#menggunakan Library Pandas dan Seaborn
import pandas as pd
import seaborn as sns
```

Import Data Set

```
df = pd.read_csv('https://docs.google.com/spreadsheets/d/e/2PACX-1vTJ7xBQatjweuBlGqckV4_KNsZ1uFtH--dMPewKRSQTEocqKDJ7SVJEnnaMirctLOGuVIk6GtDkpYr7/pub?gid=80702269&single=true&outp
#df = merupakan struktur data utama dalam Pandas yang memungkinkan untuk menyimpan data, manipulasi, analisis, dan visualisasi data dengan mudah(df=Data Frame)
#Memasukan data ke dalam program/google colab
```

Keterangan :

Untuk meng-import data set perhatikan extensi nya apa sesuai kan dengan extensi dari data set nya, untuk data set yang kita pakai menggunakan extensi CSV jadi kita menggunakan **pd.read csv**

STEP 1 | Import library dan Data set

Menampilkan beberapa baris dari data set

```
df.head()

#Untuk menampilkan beberapa baris data yg sudah di insert sebelumnya
```

Output

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	company_size
0	2021	MI	FT	Data Scientist	30400000	CLP	40038	CL	100	CL	L
1	2021	MI	FT	BI Data Analyst	11000000	HUF	36259	HU	50	US	L
2	2020	MI	FT	Data Scientist	11000000	HUF	35735	HU	50	HU	L
3	2021	MI	FT	ML Engineer	8500000	JPY	77364	JP	50	JP	S
4	2022	SE	FT	Lead Machine Learning Engineer	7500000	INR	95386	IN	50	IN	L

STEP 2 | MERUBAH DATA KOLOM/BARIS AGAR LEBIH MUDAH DI ANALISIS

Mengubah nama data di dalam kolom, dengan tujuan agar mudah dibaca dan dianalisis

```
df.rename(columns={ #Mengubah nama data di dalam kolom, dengan tujuan agar mudah dibaca dan dianalisis
    'Unnamed: 0': 'Index',
    'work_year': 'Tahun_kerja',
    'experience_level': 'Level_pengalaman',
    'employment_type': 'Jenis_pekerjaan',
    'job_title': 'Jabatan',
    'salary': 'Gaji',
    'salary_currency': 'Mata_uang',
    'salary_in_usd': 'Gaji_dalam_usd',
    'employee_residence': 'Residence_karyawan',
    'remote_ratio': 'Remoteness_ratio',
    'company_location': 'Lokasi_perusahaan',
    'company_size': 'Ukuran_perusahaan'
}, inplace=True)
```

STEP 2 | MERUBAH DATA KOLOM/BARIS AGAR LEBIH MUDAH DI ANALISIS

mengecek kembali apakah berhasil berubah/tidak

```
df.head()# mengecek kembali apakah berhasil berubah/tidak
```

Output
Sebelum diubah

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	company_size
0	2021	MI	FT	Data Scientist	30400000	CLP	40038	CL	100	CL	L
1	2021	MI	FT	BI Data Analyst	11000000	HUF	36259	HU	50	US	L
2	2020	MI	FT	Data Scientist	11000000	HUF	35735	HU	50	HU	L
3	2021	MI	FT	ML Engineer	8500000	JPY	77364	JP	50	JP	S
4	2022	SE	FT	Lead Machine Learning Engineer	7500000	INR	95386	IN	50	IN	L

Sesudah diubah

	Tahun_kerja	Level_pengalaman	Jenis_pekerjaan	Jabatan	Gaji	Mata_uang	Gaji_dalam_usd	Residence_karyawan	Remoteness_ratio	Lokasi_perusahaan	Ukuran_perusahaan
0	2021	MI	FT	Data Scientist	30400000	CLP	40038	CL	100	CL	L
1	2021	MI	FT	BI Data Analyst	11000000	HUF	36259	HU	50	US	L
2	2020	MI	FT	Data Scientist	11000000	HUF	35735	HU	50	HU	L
3	2021	MI	FT	ML Engineer	8500000	JPY	77364	JP	50	JP	S
4	2022	SE	FT	Lead Machine Learning Engineer	7500000	INR	95386	IN	50	IN	L

STEP 2 | MERUBAH DATA KOLOM/BARIS AGAR LEBIH MUDAH DI ANALISIS

Mengubah nama data di dalam baris, dengan tujuan agar lebih mudah di analisis

```
df['Level_pengalaman'].replace({ #Mengubah nama data di dalam baris, dengan tujuan agar lebih mudah di analisis
    'SE': 'Senior',
    'MI': 'Mid',
    'EN': 'Entry',
    'EX': 'Executive'
}, inplace=True)
```

#PERBEDAAN UTAMA NYA RENAME(UNTUK MENGUBAH NAMA DATA DI DALAM KOLOM DAN INDEX TABEL), SEDANGKAN UNTUK REPLACE(UNTUK MENGUBAH NAMA DATA DI DALAM TABEL)

```
df['Jenis_pekerjaan'].replace({ #Mengubah nama data di dalam baris, dengan tujuan agar lebih mudah di analisis
    'FT': 'Full Time',
    'PT': 'Part Time',
    'CT': 'Contract',
    'Fl': 'Freelance'
}, inplace=True)
```

STEP 2 | MERUBAH DATA KOLOM/BARIS AGAR LEBIH MUDAH DI ANALISIS

mengecek kembali apakah berhasil berubah/tidak

```
df.head()# mengecek kembali apakah berhasil berubah/tidak
```

Output
Sebelum diubah

work_year		experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	company_size
0	2021	MI	FT	Data Scientist	30400000	CLP	40038	CL	100	CL	L
1	2021	MI	FT	BI Data Analyst	11000000	HUF	36259	HU	50	US	L
2	2020	MI	FT	Data Scientist	11000000	HUF	35735	HU	50	HU	L
3	2021	MI	FT	ML Engineer	8500000	JPY	77364	JP	50	JP	S
4	2022	SE	FT	Lead Machine Learning Engineer	7500000	INR	95386	IN	50	IN	L

Sesudah diubah

Tahun_kerja		Level_pengalaman	Jenis_pekerjaan	Jabatan	Gaji	Mata_uang	Gaji_dalam_usd	Residence_karyawan	Remoteness_ratio	Lokasi_perusahaan	Ukuran_perusahaan
0	2021	Mid	Full Time	Data Scientist	30400000	CLP	40038	CL	100	CL	L
1	2021	Mid	Full Time	BI Data Analyst	11000000	HUF	36259	HU	50	US	L
2	2020	Mid	Full Time	Data Scientist	11000000	HUF	35735	HU	50	HU	L
3	2021	Mid	Full Time	ML Engineer	8500000	JPY	77364	JP	50	JP	S
4	2022	Senior	Full Time	Lead Machine Learning Engineer	7500000	INR	95386	IN	50	IN	L

STEP 2 | MERUBAH DATA KOLOM/BARIS AGAR LEBIH MUDAH DI ANALISIS

MEMFILTER DATA DENGAN JABATAN KHUSUS UNTUK DATA SCIENTIST, SESUAI DENGAN TUJUAN ANALISIS INI

```
#MEMFILTER DATA DENGAN JABATAN KHUSUS UNTUK DATA SCIENTIST, SESUAI DENGAN TUJUAN ANALISIS INI
data_scientist= df.loc[ (df['Jabatan']=='Data Scientist'),#SYNTAX (loc [])DIGUNAKAN UNTUK MENGAKSES BARIS DAN KOLOM BERDASARKAN LABEL(NAMABARIS/NAMAKOLOM)
['Tahun_kerja',
'Jabatan',
'Jenis_pekerjaan',
'Level_pengalaman',
'Gaji_dalam_usd',
'Mata_uang',
'Lokasi_perusahaan',
#DENGAN KRITERIA TERTENTU(SESUAI TUJUAN)
```

MEMFILTER KOLOM/BARIS YG HANYA DIPERLUKAN SAJA UNTUK ANALISIS

```
#MEMFILTER KOLOM/BARIS YG HANYA DIPERLUKAN SAJA UNTUK ANALISIS
]].sort_values(by='Gaji_dalam_usd', ascending=False)#SYNTAX (SORT_VALUES()) MERUPAKAN MENG SORTIR DATA /MENGURUTKAN DATA DARI YANG TERTINGGI MENUJU
#KE PALING KECIL, KARENA (ASCENDING=FALSE) ARTINYA DIURUTKAN DARI YG TERBESAR KE YG TERKECIL
#JIKA (ASCENDING=TRUE)MAKA DIURUTKAN DARI MULAI YG TERKECIL KE YG TERBESAR
data_scientist #PRINT
```

STEP 2 | MERUBAH DATA KOLOM/BARIS AGAR LEBIH MUDAH DI ANALISIS

menampilkan data yang sudah tadi di filter

data_scientist

mengecek kembali apakah berhasil berubah/tidak

	Tahun_kerja	Jabatan	Jenis_pekerjaan	Level_pengalaman	Gaji_dalam_usd	Mata_uang	Lokasi_perusahaan	
76	2023	Data Scientist	Full Time	Senior	750000	USD	US	
81	2024	Data Scientist	Full Time	Senior	720000	USD	US	
78	2024	Data Scientist	Full Time	Senior	720000	USD	US	
125	2020	Data Scientist	Full Time	Senior	412000	USD	US	
139	2024	Data Scientist	Full Time	Mid	385000	USD	US	
...	
54	2022	Data Scientist	Full Time	Entry	17805	INR	IN	
7	2022	Data Scientist	Full Time	Entry	17684	HUF	HU	
52	2023	Data Scientist	Full Time	Mid	17025	INR	IN	
60	2021	Data Scientist	Full Time	Mid	16904	INR	IN	
14830	2023	Data Scientist	Full Time	Entry	16000	USD	EC	
3015 rows × 7 columns								

STEP 3 |

CEK APAKAH ADA DATA YANG KOTOR

menampilkan info dari data set

```
df.info() #menampilkan info dari data set
```

Output

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 14838 entries, 0 to 14837  
Data columns (total 11 columns):  
#   Column                Non-Null Count  Dtype  
---  ---  
0   Tahun_kerja           14838 non-null  int64  
1   Level_pengalaman      14838 non-null  object  
2   Jenis_pekerjaan       14838 non-null  object  
3   Jabatan               14838 non-null  object  
4   Gaji                  14838 non-null  int64  
5   Mata_uang             14838 non-null  object  
6   Gaji_dalam_usd        14838 non-null  int64  
7   Residence_karyawan    14838 non-null  object  
8   Remoteness_ratio      14838 non-null  int64  
9   Lokasi_perusahaan     14838 non-null  object  
10  Ukuran_perusahaan     14838 non-null  object  
dtypes: int64(4), object(7)  
memory usage: 1.2+ MB
```

tidak ada data yang null, maka untuk keseluruhan data sudah bersih

STEP 3 | CEK APAKAH ADA DATA YANG KOTOR

menampilkan info dari index pekerjaan data scientist

```
data_scientist.info()
```

Output

```
<class 'pandas.core.frame.DataFrame'>
Index: 3015 entries, 76 to 14830
Data columns (total 7 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Tahun_kerja           3015 non-null   int64
 1   Jabatan                3015 non-null   object
 2   Jenis_pekerjaan       3015 non-null   object
 3   Level_pengalaman      3015 non-null   object
 4   Gaji_dalam_usd        3015 non-null   int64
 5   Mata_uang              3015 non-null   object
 6   Lokasi_perusahaan     3015 non-null   object
dtypes: int64(2), object(5)
memory usage: 188.4+ KB
```

INDEX DATA SCIENTIST HANYA MEMILIKI
3015 ROW DENGAN 7 KOLOM

STEP 3 | DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

mengecek kembali apakah ada data kotor

```
df.isnull().sum()  
  
#TIDAK ADA YANG NULL, MAKA DIPASTIKAN DATA SUDAH BERSIH
```

Output

	0
Tahun_kerja	0
Level_pengalaman	0
Jenis_pekerjaan	0
Jabatan	0
Gaji	0
Mata_uang	0
Gaji_dalam_usd	0
Residence_karyawan	0
Remoteness_ratio	0
Lokasi_perusahaan	0
Ukuran_perusahaan	0
dtype: int64	

TIDAK ADA YANG NULL, MAKA DIPASTIKAN
DATA SUDAH BERSIH

STEP 4 | ANALISIS

DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

MEMFILTER DENGAN SYARAT JENIS PEKERJAAN=FULLTIME , LEVEL PENGALAMAN=EXECUTIVE

```
Executive=data_scientist.loc[(data_scientist['Jenis_pekerjaan'] == 'Full Time') &(data_scientist['Level_pengalaman']== 'Executive'),
['Tahun_kerja',
 'Jabatan',
 'Level_pengalaman',
 'Gaji_dalam_usd',
 'Mata_uang',
 'Lokasi_perusahaan',
]].sort_values(by='Gaji_dalam_usd', ascending=False)#DIURUTKAN DARI BESAR KE KECIL
Executive
#MEMFILTER DENGAN SYARAT JENIS PEKERJAAN=FULLTIME, LEVEL PENGALAMAN=EXECUTIVE
```

Output

	Tahun_kerja	Jabatan	Level_pengalaman	Gaji_dalam_usd	Mata_uang	Lokasi_perusahaan
464	2023	Data Scientist	Executive	300000	USD	US
474	2023	Data Scientist	Executive	299500	USD	US
942	2023	Data Scientist	Executive	258750	USD	US
943	2023	Data Scientist	Executive	258750	USD	US
1150	2023	Data Scientist	Executive	250000	USD	US
1191	2023	Data Scientist	Executive	249600	USD	US
1296	2024	Data Scientist	Executive	246100	USD	US
1770	2023	Data Scientist	Executive	228856	USD	GB
1769	2023	Data Scientist	Executive	228856	USD	GB
1898	2023	Data Scientist	Executive	225000	USD	US

Disini kita memfilter jabatan data scientist dengan jenis pekerjaan fulltime dan pengalaman executive

STEP 4 | ANALISIS

DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

Menghitung rata rata dari tahun kerja dan gaji

```
Executive.groupby('Tahun_kerja')['Gaji_dalam_usd'].mean()# MENGHITUNG RATA RATA DARI DATA DIATAS DENGAN SYNTAX GROUPBY DAN MEAN
```

Output

Tahun_kerja	Gaji_dalam_usd
2022	144500.000000
2023	188562.344828
2024	165500.000000

dtype: float64

berdasarkan syntac diatas kita ingin mengelompokkan data kolom Tahun_kerja dan dikelompok dengan kolom Gaji_dalam_usd lalu kita cari rata-ratanya berdasarkan data yang kita kelompokkan

STEP 4 | ANALISIS

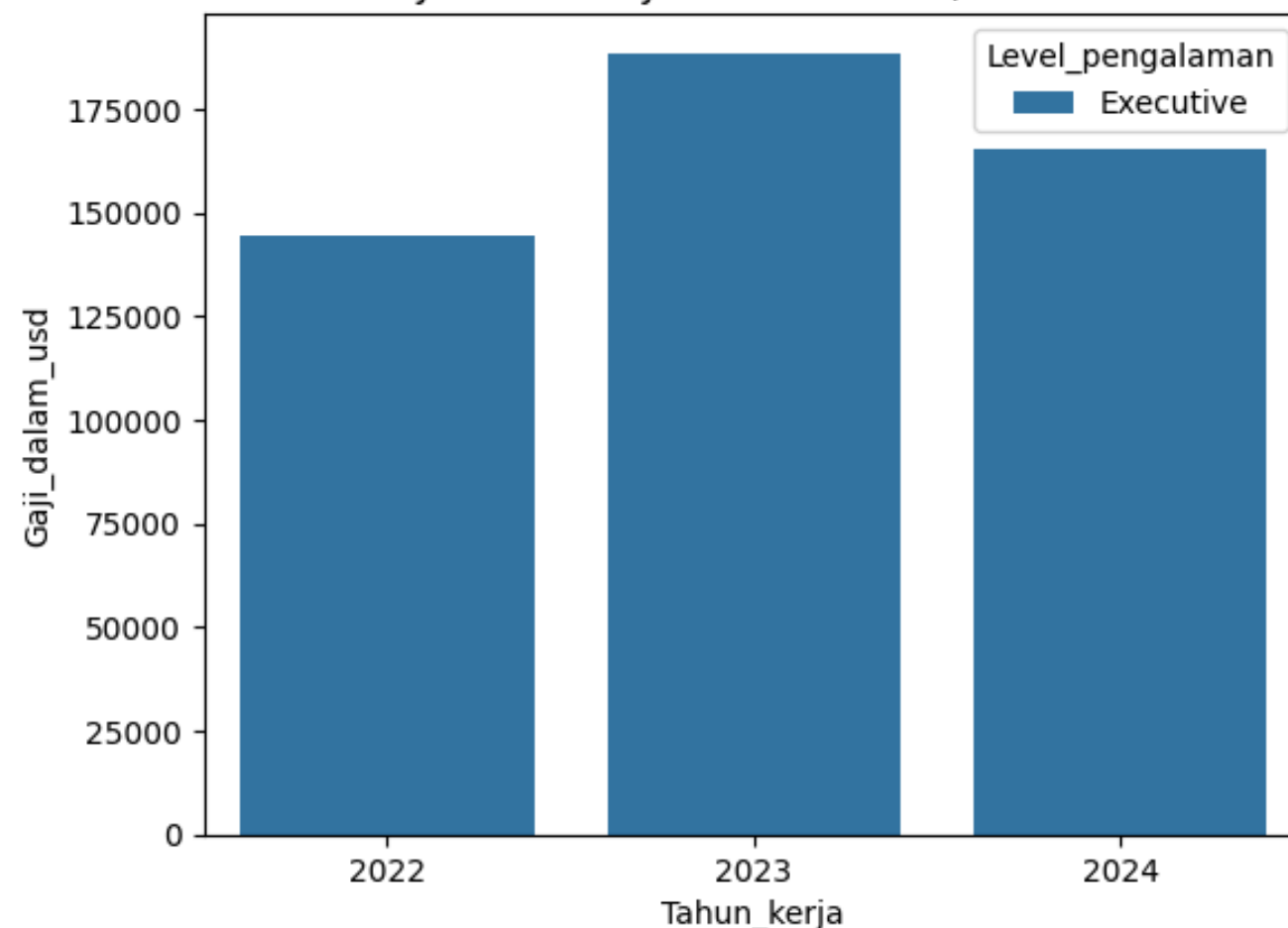
DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

Visualisasi data dari tahun kerja sebagai X dan gaji sebagai Y

```
sns.barplot(  
    data=Executive,  
    x='Tahun_kerja',  
    y='Gaji_dalam_usd',  
    hue='Level_pengalaman',  
    ci=None  
  
).set(title=' GAJI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME (LEVEL PENGLAMAN KERJA EXECUTIVE)')#judul  
#VISUALISASI DATA DENGAN MENGGUNAKAN LIBRARY BARPLOT DENGAN DATA EXECUTIVE
```

Output

GAJI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME (LEVEL PENGLAMAN KERJA EXECUTIVE)



Gaji paling tinggi di level pengalaman executive didapatkan pada tahun 2023 dengan besaran gaji sebesar +- 175000

STEP 4 | ANALISIS

DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

MEMFILTER DENGAN SYARAT JENIS PEKERJAAN=FULLTIME, LEVEL PENGALAMAN=SENIOR

```
#DATA GAJI DI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME (LEVEL PENGLAMAN KERJA SENIOR)

Senior=data_scientist.loc[(data_scientist['Jenis_pekerjaan'] == 'Full Time') &(data_scientist['Level_pengalaman']== 'Senior'),
['Tahun_kerja',
 'Jabatan',
 'Level_pengalaman',
 'Gaji_dalam_usd',
 'Mata_uang',
 'Lokasi_perusahaan',
]].sort_values(by='Gaji_dalam_usd', ascending=False)
Senior
#MEMFILTER DENGAN SYARAT JENIS PEKERJAAN=FULLTIME, LEVEL PENGALAMAN=senior
```

Output

	Tahun_kerja	Jabatan	Level_pengalaman	Gaji_dalam_usd	Mata_uang	Lokasi_perusahaan
76	2023	Data Scientist	Senior	750000	USD	US
78	2024	Data Scientist	Senior	720000	USD	US
81	2024	Data Scientist	Senior	720000	USD	US
125	2020	Data Scientist	Senior	412000	USD	US
170	2023	Data Scientist	Senior	370000	USD	US
...
14645	2022	Data Scientist	Senior	37824	EUR	ES
14642	2022	Data Scientist	Senior	37824	EUR	ES
14641	2022	Data Scientist	Senior	37824	EUR	ES
14775	2024	Data Scientist	Senior	31250	GBP	GB
4367	2021	Data Scientist	Senior	20171	TRY	TR

2140 rows x 6 columns

Disini kita memfilter jabatan data scientist dengan jenis pekerjaan fulltime dan pengalaman senior

STEP 4 | ANALISIS

DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

Menghitung rata rata dari tahun kerja dan gaji

```
#Menghitung Rata-rata dari Level pengalaman kerja
Senior.groupby('Tahun_kerja')['Gaji_dalam_usd'].mean().sort_values(ascending=False)
```

Output

Gaji_dalam_usd	
Tahun_kerja	
2023	173480.984352
2020	172916.250000
2024	162411.981595
2022	151931.518519
2021	87071.250000
dtype: float64	

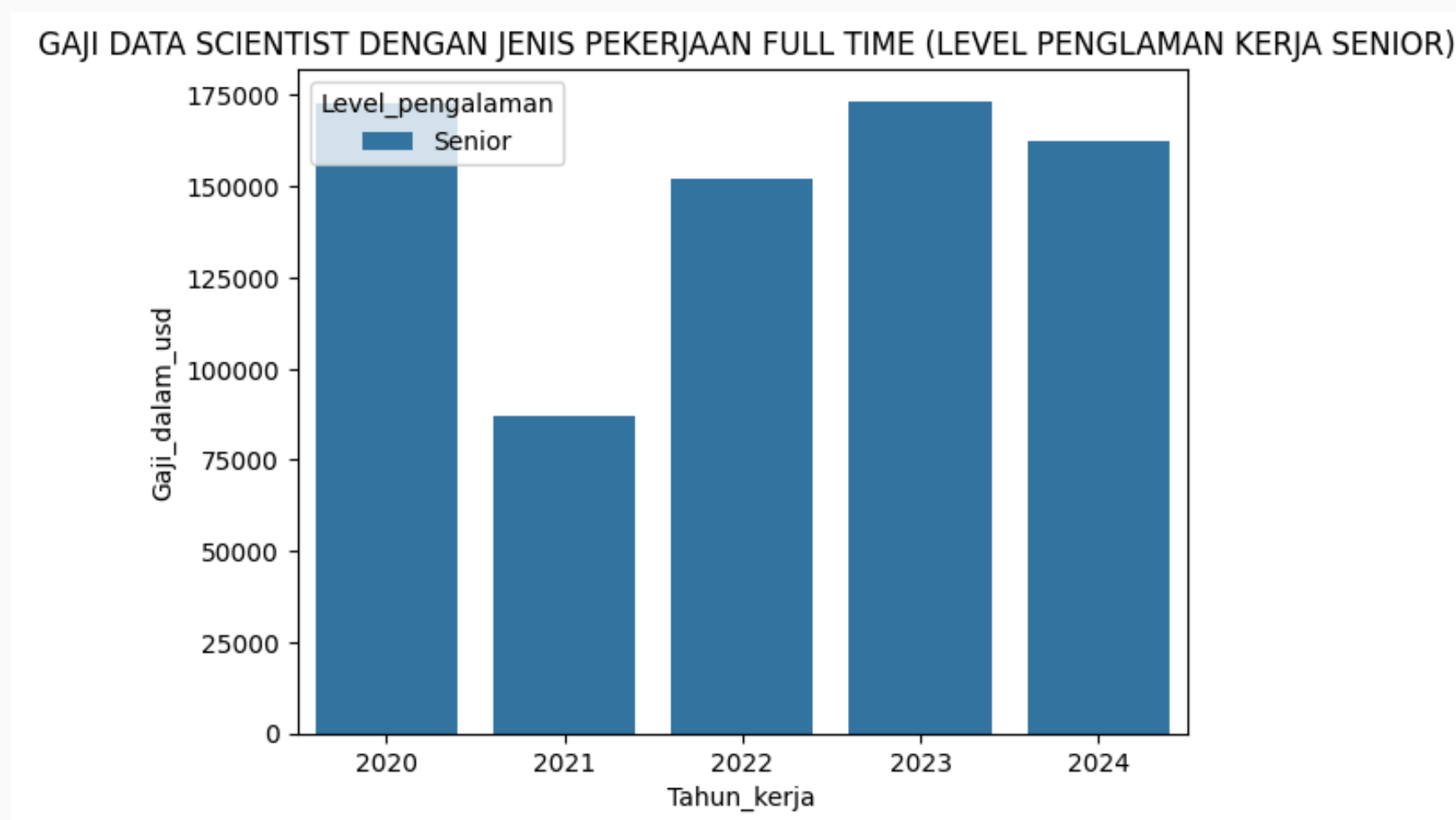
STEP 4 | ANALISIS

DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

Visualisasi data dari tahun kerja sebagai X dan gaji sebagai Y

```
sns.barplot(  
    data=Senior,  
    x='Tahun_kerja',  
    y='Gaji_dalam_usd',  
    hue='Level_pengalaman',  
    ci=None  
).set(title=' GAJI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME (LEVEL PENGLAMAN KERJA SENIOR)')
```

Output



STEP 4 | ANALISIS

DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

MEMFILTER DENGAN SYARAT JENIS PEKERJAAN=FULLTIME, LEVEL PENGALAMAN= MID (MENENGAH)

```
Mid=data_scientist.loc[(data_scientist['Jenis_pekerjaan'] == 'Full Time') &(data_scientist['Level_pengalaman']== 'Mid'),
['Tahun_kerja',
'Jabatan',
'Level_pengalaman',
'Gaji_dalam_usd',
'Mata_uang',
'Lokasi_perusahaan',
]].sort_values(by='Gaji_dalam_usd', ascending=False)
Mid
```

Output

	Tahun_kerja	Jabatan	Level_pengalaman	Gaji_dalam_usd	Mata_uang	Lokasi_perusahaan
139	2024	Data Scientist	Mid	385000	USD	US
347	2024	Data Scientist	Mid	310000	USD	US
348	2023	Data Scientist	Mid	310000	USD	US
503	2023	Data Scientist	Mid	296980	USD	US
851	2023	Data Scientist	Mid	260600	USD	US
...
14804	2021	Data Scientist	Mid	25532	EUR	DE
14780	2022	Data Scientist	Mid	25000	USD	TR
65	2023	Data Scientist	Mid	24613	THB	TH
52	2023	Data Scientist	Mid	17025	INR	IN
60	2021	Data Scientist	Mid	16904	INR	IN

701 rows × 6 columns

Disini kita memfilter jabatan data scientist dengan jenis pekerjaan fulltime dan pengalaman mid (menengah)

STEP 4 | ANALISIS

DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

Menghitung rata rata dari tahun kerja dan gaji

```
Mid.groupby('Tahun_kerja')['Gaji_dalam_usd'].mean().sort_values(ascending=False)
```

Output

Tahun_kerja	Gaji_dalam_usd
2024	135024.984227
2023	124881.494545
2022	99188.440000
2021	83348.173913
2020	71256.000000

dtype: float64

Berdasarkan syntac diatas kita ingin mencari rata-rata pada Data scientist level mid dan kita urutkan dari nilai terbesar sampai terkecil

STEP 4 | ANALISIS

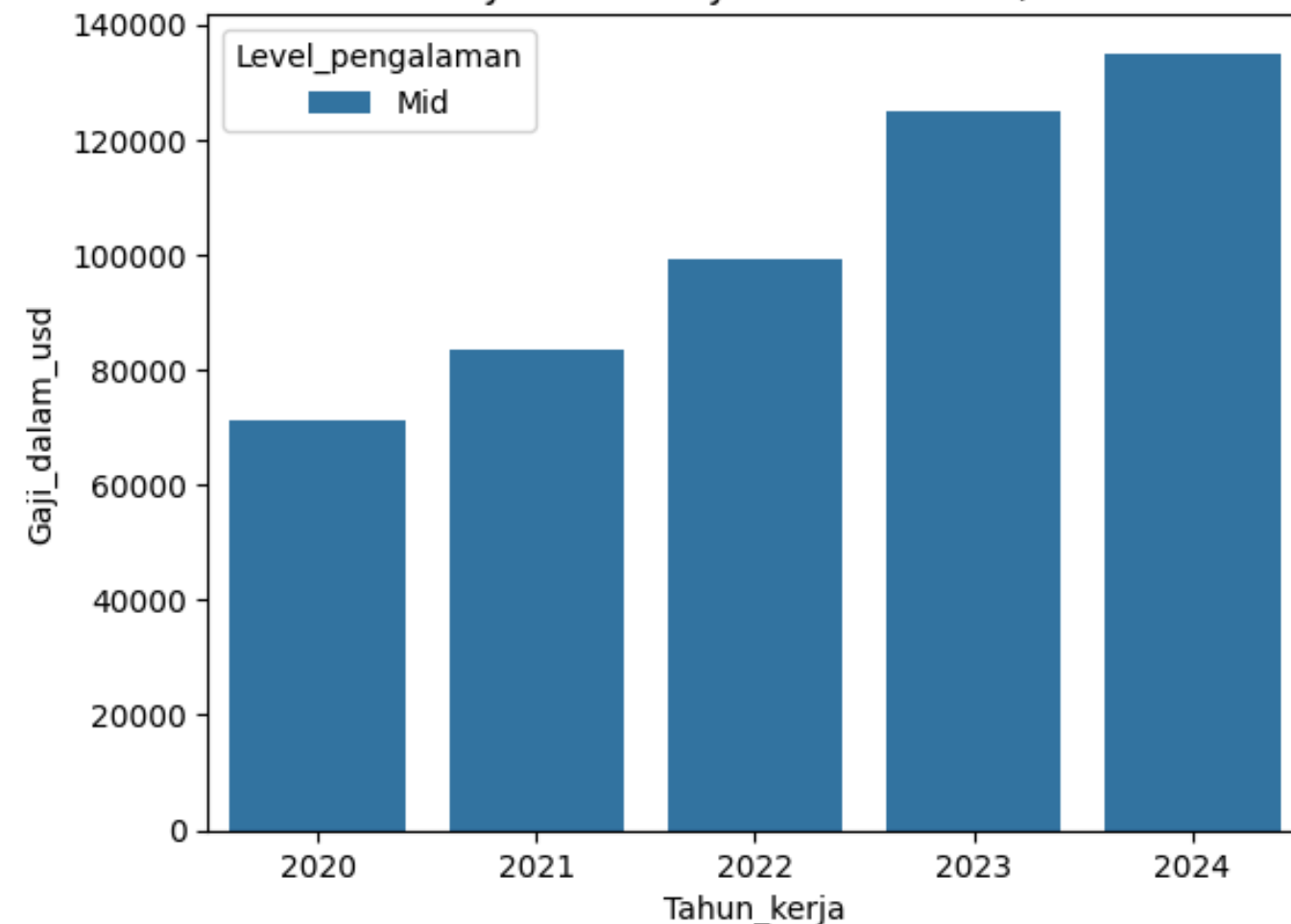
DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

Visualisasi data dari tahun kerja sebagai X dan gaji sebagai Y

```
sns.barplot(  
    data=Mid,  
    x='Tahun_kerja',  
    y='Gaji_dalam_usd',  
    hue='Level_pengalaman',  
    ci=None  
).set(title=' GAJI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME (LEVEL PENGLAMAN KERJA MID)')
```

Output

GAJI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME (LEVEL PENGLAMAN KERJA MID)



Posisi Data Scientist level Mid memiliki prospek yang sangat positif, peningkatan gaji yang konsisten menunjukkan permintaan yang terus meningkat, dan level Mid menjadi posisi yang menjanjikan dengan peningkatan kompensasi yang stabil.

STEP 4 | ANALISIS

DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

MEMFILTER DENGAN SYARAT JENIS PEKERJAAN=FULLTIME, LEVEL PENGALAMAN= ENTRY (BAWAH)

```
Entry=data_scientist.loc[(data_scientist['Jenis_pekerjaan'] == 'Full Time') &(data_scientist['Level_pengalaman']== 'Entry'),
['Tahun_kerja',
'Jabatan',
'Level_pengalaman',
'Gaji_dalam_usd',
'Mata_uang',
'Lokasi_perusahaan',
]].sort_values(by='Gaji_dalam_usd', ascending=False)
Entry
```

Output

	Tahun_kerja	Jabatan	Level_pengalaman	Gaji_dalam_usd	Mata_uang	Lokasi_perusahaan
3629	2023	Data Scientist	Entry	190000	USD	US
4350	2022	Data Scientist	Entry	180000	USD	US
4335	2022	Data Scientist	Entry	180000	USD	US
4556	2024	Data Scientist	Entry	175100	USD	US
4557	2024	Data Scientist	Entry	175100	USD	US
...
11156	2023	Data Scientist	Entry	19910	BRL	BR
14824	2023	Data Scientist	Entry	19434	EUR	GR
54	2022	Data Scientist	Entry	17805	INR	IN
7	2022	Data Scientist	Entry	17684	HUF	HU
14830	2023	Data Scientist	Entry	16000	USD	EC
127 rows × 6 columns						

Disini kita memfilter jabatan data scientist dengan jenis pekerjaan fulltime dan pengalaman entry (menengah)

STEP 4 | ANALISIS

DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

Menghitung rata rata dari tahun kerja dan gaji

```
Entry.groupby('Tahun_kerja')['Gaji_dalam_usd'].mean().sort_values(ascending=False)
```

Output

Tahun_kerja	Gaji_dalam_usd
2023	93537.952381
2024	92477.000000
2022	79514.851852
2021	63604.333333
2020	61646.200000
dtype: float64	

Berdasarkan syntac diatas kita ingin menghitung rata-rata dari Gaji_dalam_usd per Tahun_kerja, lalu difilter dari yang terbesar sampai yang terkecil

STEP 4 | ANALISIS

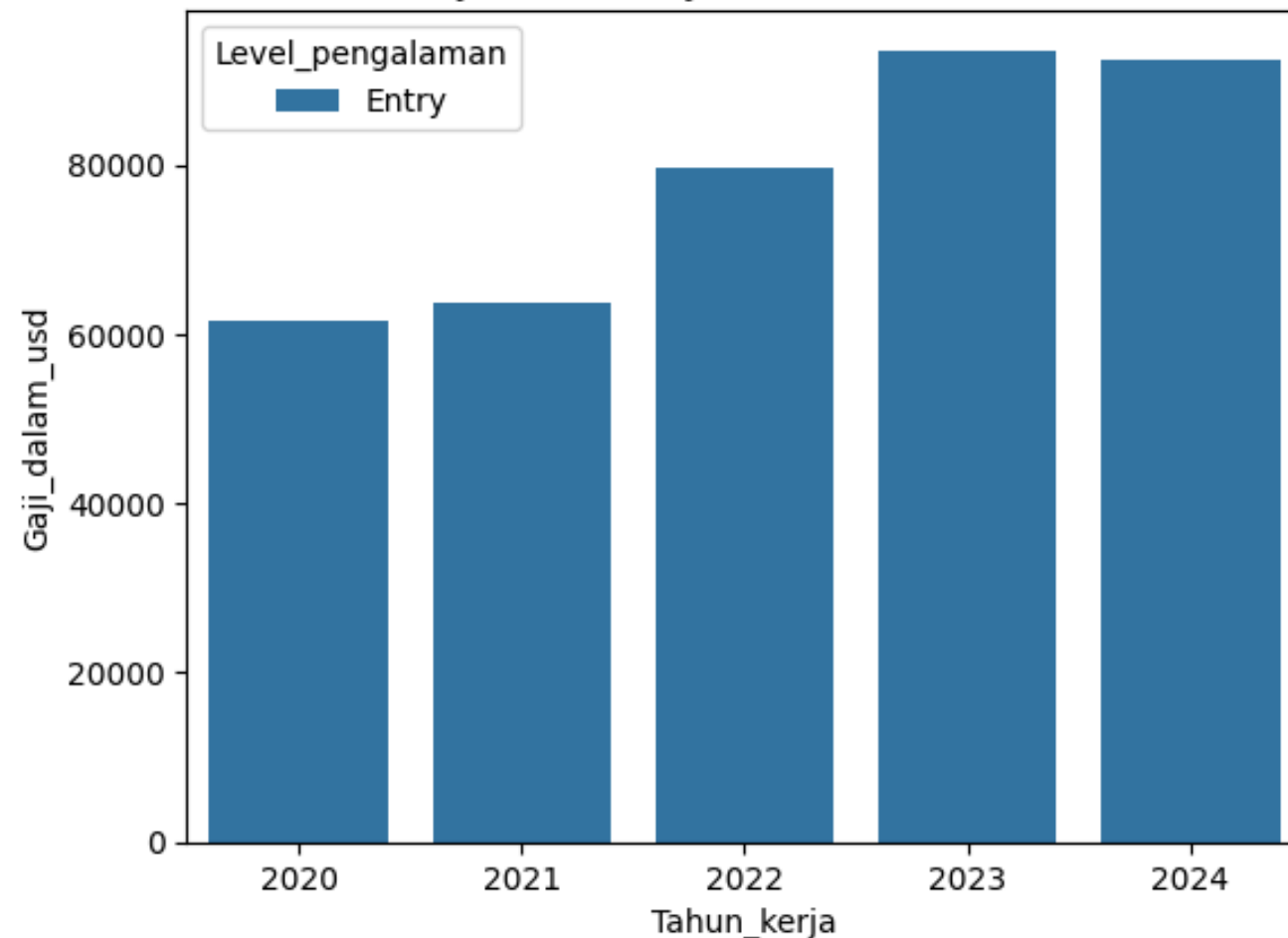
DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME

Visualisasi data dari tahun kerja sebagai X dan gaji sebagai Y

```
sns.barplot(  
    data=Entry,  
    x='Tahun_kerja',  
    y='Gaji_dalam_usd',  
    hue='Level_pengalaman',  
    ci=None  
).set(title=' GAJI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME (LEVEL PENGLAMAN KERJA Entry)')
```

Output

GAJI DATA SCIENTIST DENGAN JENIS PEKERJAAN FULL TIME (LEVEL PENGLAMAN KERJA Entry)



Posisi Entry Level menunjukkan gaji awal yang cukup kompetitif, tren kenaikan positif menunjukkan industri yang berkembang, stabilitas gaji di tahun terakhir menunjukkan standarisasi posisi entry level.

STEP 5| ANALISIS

DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN CONTRACT

MEMFILTER DENGAN SYARAT JENIS PEKERJAAN=CONTRACT, LEVEL PENGALAMAN= SENIOR

```
Senior=data_scientist.loc[(data_scientist['Jenis_pekerjaan'] == 'Contract') &(data_scientist['Level_pengalaman']== 'Senior'),
['Tahun_kerja',
'Jabatan',
'Jenis_pekerjaan',
'Level_pengalaman',
'Gaji_dalam_usd',
'Mata_uang',
'Lokasi_perusahaan',
]].sort_values(by='Gaji_dalam_usd', ascending=False)
Senior
```

Output

	Tahun_kerja	Jabatan	Jenis_pekerjaan	Level_pengalaman	Gaji_dalam_usd	Mata_uang	Lokasi_perusahaan
3512	2024	Data Scientist	Contract	Senior	191027	USD	US
9519	2024	Data Scientist	Contract	Senior	120869	USD	US

Disini kita memfilter jabatan data scientist dengan jenis pekerjaan contract dan pengalaman senior

STEP 5 | ANALISIS

DATA GAJI SEBAGAI DATA SCIENTIST DENGAN JENIS PEKERJAAN

CONTRACT

Menghitung rata rata dari tahun kerja dan gaji

```
Senior.groupby('Tahun_kerja')['Gaji_dalam_usd'].mean()
```

Output

```
      Gaji_dalam_usd
Tahun_kerja
2024          155948.0
dtype: float64
```

Berdasarkan syntac diatas kita ingin mencari rata-rata dari jenis pekerjaan Contract dan level pekerjaan senior

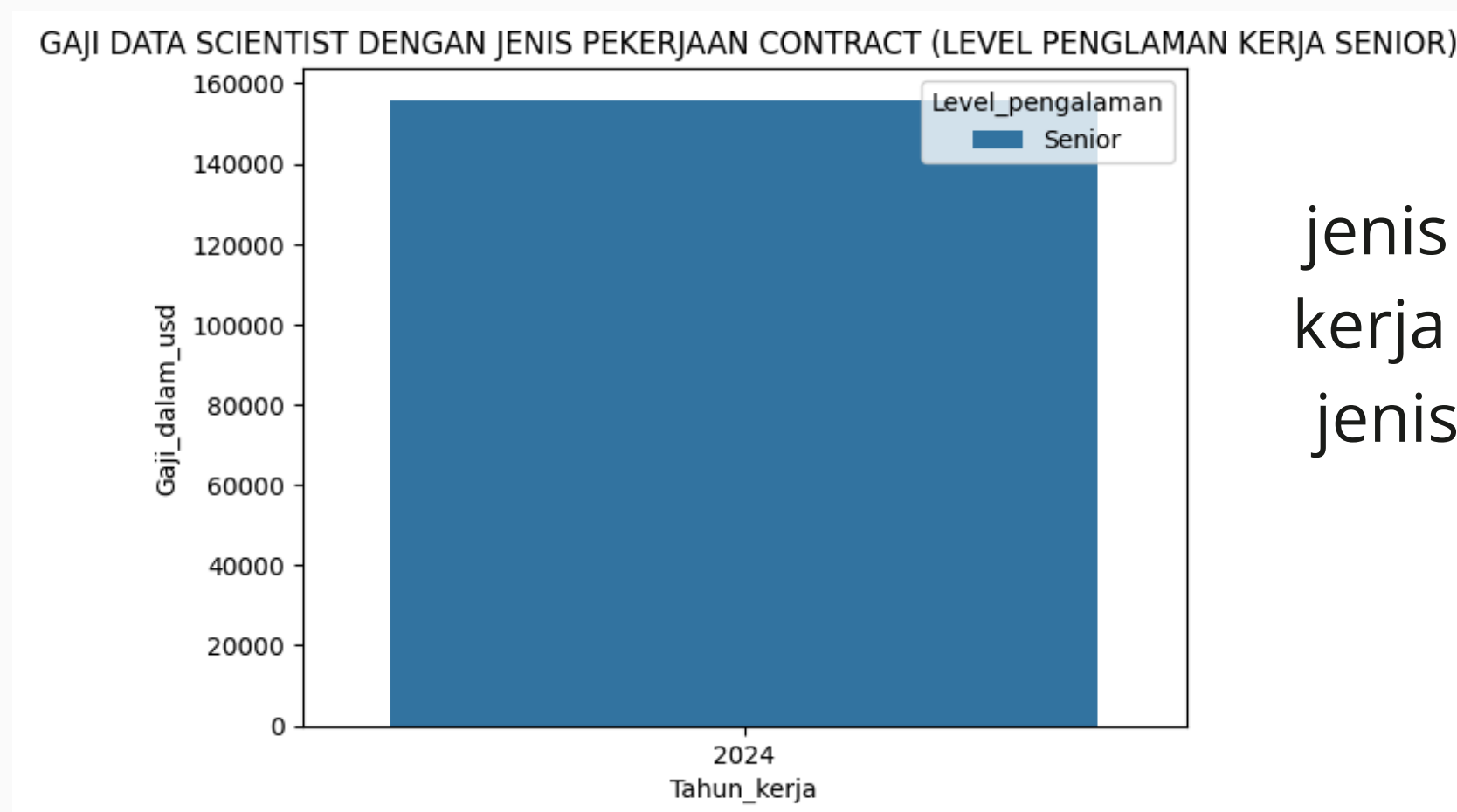
STEP 5 |

ANALISIS STATISTIK

Visualisasi data dari tahun kerja sebagai X dan gaji sebagai Y

```
sns.barplot(  
    data=Senior,  
    x='Tahun_kerja',  
    y='Gaji_dalam_usd',  
    hue='Level_pengalaman',  
    ci=None  
).set(title=' GAJI DATA SCIENTIST DENGAN JENIS PEKERJAAN CONTRACT (LEVEL PENGLAMAN KERJA SENIOR)')
```

Output



jenis pekerjaan kontrak memiliki jumlah tenaga kerja yang masih minimal, jauh berbeda dengan jenis pekerjaan full time, yang rata-rata gajinya hanya 160.000 USD pada tahun 2024.

STEP 6|

ANALISIS STATISTIK

menghitung rata rata gaji sebagai data scientist berdasarkan level pengalaman dari tahun 2020 - 2024, dengan keseluruhan level pengalaman diurutkan dari yang terbesar

```
rata_rata= data_scientist.groupby('Level_pengalaman')['Gaji_dalam_usd'].mean().sort_values(ascending=False)
rata_rata
#menghitung rata rata gaji sebagai data scientist berdasarkan pengalaman dari tahun 2020 - 2024, diurutkan dari yang terbesar
```

Output

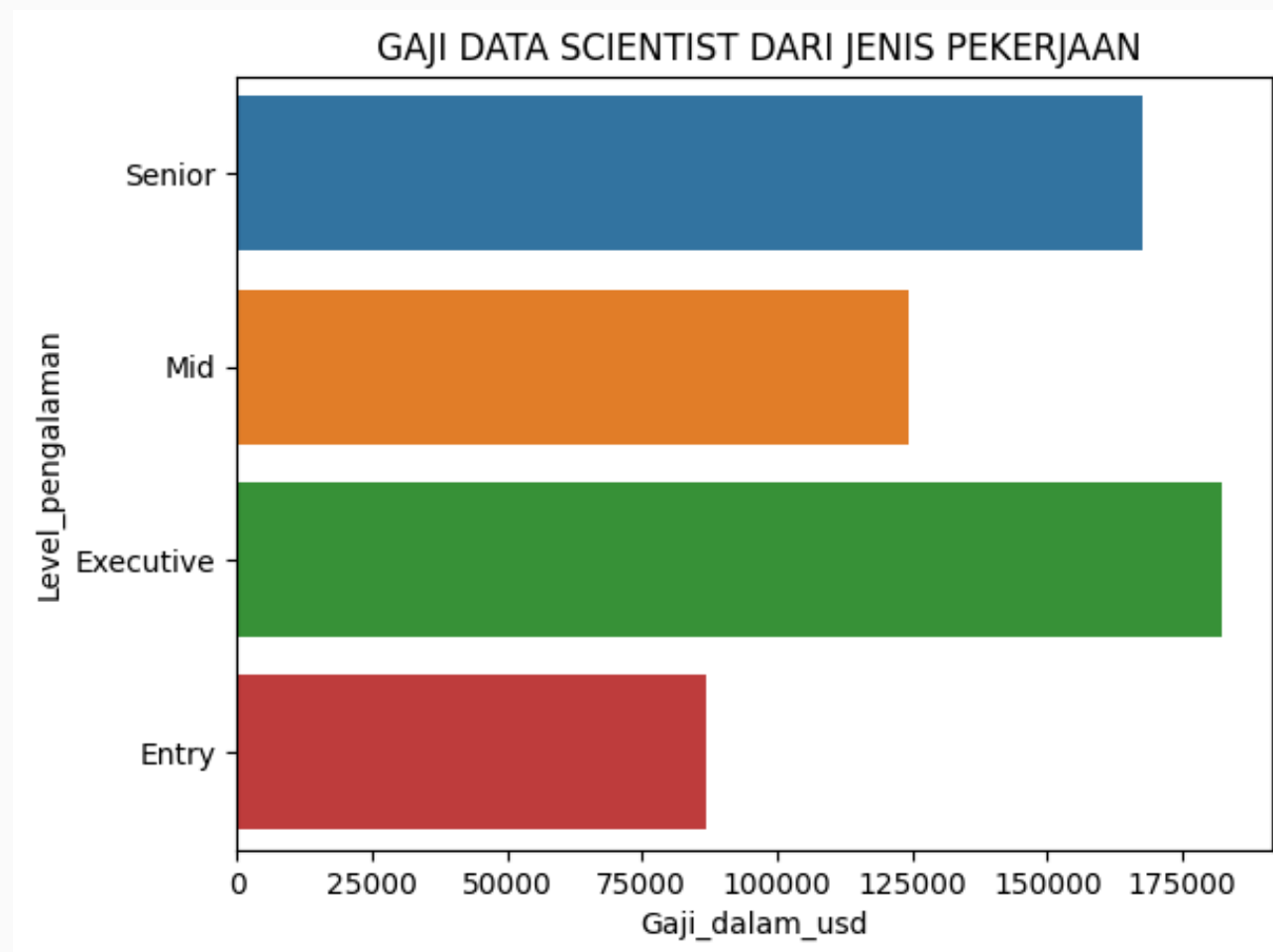
Level_pengalaman	Gaji_dalam_usd
Executive	182440.756757
Senior	167625.925303
Mid	124480.467236
Entry	87028.373134
dtype: float64	

STEP 6 | ANALISIS STATISTIK

Visualisasi data dari tahun kerja sebagai X dan gaji sebagai Y

```
sns.barplot(  
    data=data_scientist,  
    y='Level_pengalaman',  
    x='Gaji_dalam_usd',  
    hue='Level_pengalaman',  
    ci=None  
).set(title=' GAJI DATA SCIENTIST DARI JENIS PEKERJAAN')
```

Output



Industri Data Science memiliki jenjang karir yang jelas dengan peningkatan kompensasi yang signifikan; perbedaan gaji antar tingkat menunjukkan nilai pengalaman dalam industri, gaji yang menarik bahkan untuk tingkat masuk dan struktur gaji menunjukkan industri yang matang dengan jalur karir yang jelas.

STEP 6 |

ANALISIS STATISTIK

Menghitung percentile (10%, 25%, 50%, 75%, 90%) dari setiap pengalaman

```
percentile=data_scientist.groupby('Level_pengalaman')['Gaji_dalam_usd'].quantile([0.1,0.25,0.5,0.75,0.9]) #menghitung percentile
percentile
```

Output

Gaji_dalam_usd		
Level_pengalaman		
Entry	0.10	32974.00
	0.25	51141.00
	0.50	81000.00
	0.75	119800.00
	0.90	150000.00

Mid	0.10	61472.30
	0.25	86144.25
	0.50	120000.00
	0.75	160000.00
	0.90	194080.00

Senior	0.10	101050.00
	0.25	130000.00
	0.50	160350.00
	0.75	198440.00
	0.90	239000.00

Executive	0.10	112166.40
	0.25	144540.00
	0.50	181900.00
	0.75	225000.00
	0.90	253500.00

STEP 6 | ANALISIS STATISTIK

mengubah Pandas Series menjadi Pandas DataFrame. karena seaborn membutuhkan data dalam bentuk dataframe, bukan series.

```
percentile = percentile.reset_index()
```

Berdasarkan syntac diatas kita ingin mengubah indeks dari DataFrame percentile menjadi kolom biasa. Ini berguna jika indeks sebelumnya memiliki makna khusus, seperti tanggal atau kategori, dan kita ingin memperlakukannya sebagai data biasa.

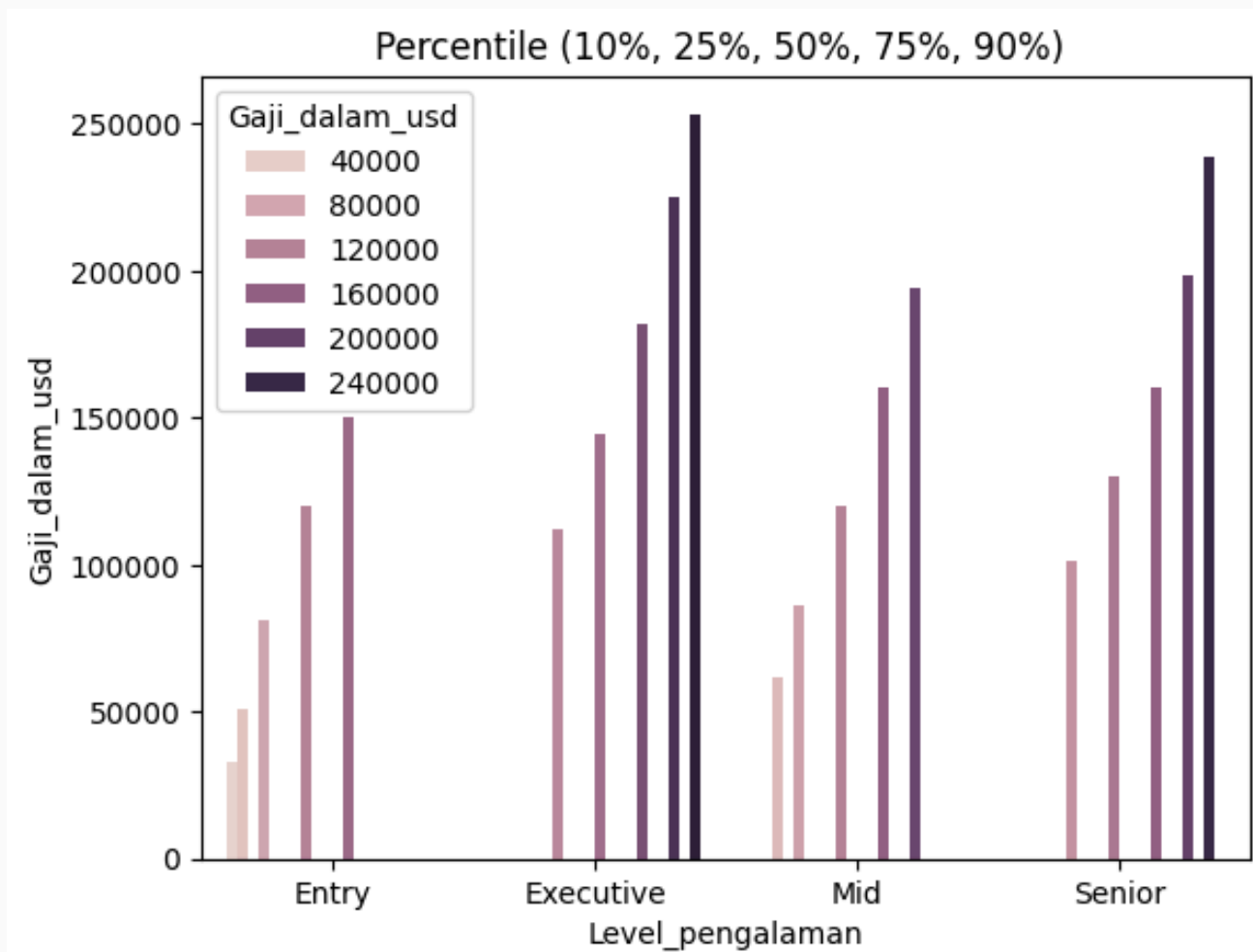
STEP 6 |

ANALISIS STATISTIK

Visualisasi data dari level pengalaman sebagai X dan gaji sebagai Y

```
sns.barplot(  
    data=percentile,  
    x='Level_pengalaman',  
    y='Gaji_dalam_usd',  
    hue='Gaji_dalam_usd',  
    ci=None  
).set(title='Percentile (10%, 25%, 50%, 75%, 90%) ')
```

Output



STEP 6 |

ANALISIS STATISTIK

Menghitung ekstremum minimal dan maximal

```
#menghitung data ekstremum dari data diatas

#mengitung min dan max dari gaji sebagai data scientist
minmax=data_scientist.groupby('Level_pengalaman')['Gaji_dalam_usd'].agg(['min','max']).sort_values(by='min', ascending=False)
minmax
```

Output

	min	max
Level_pengalaman		
Executive	78000	300000
Senior	20171	750000
Mid	16904	385000
Entry	16000	190000

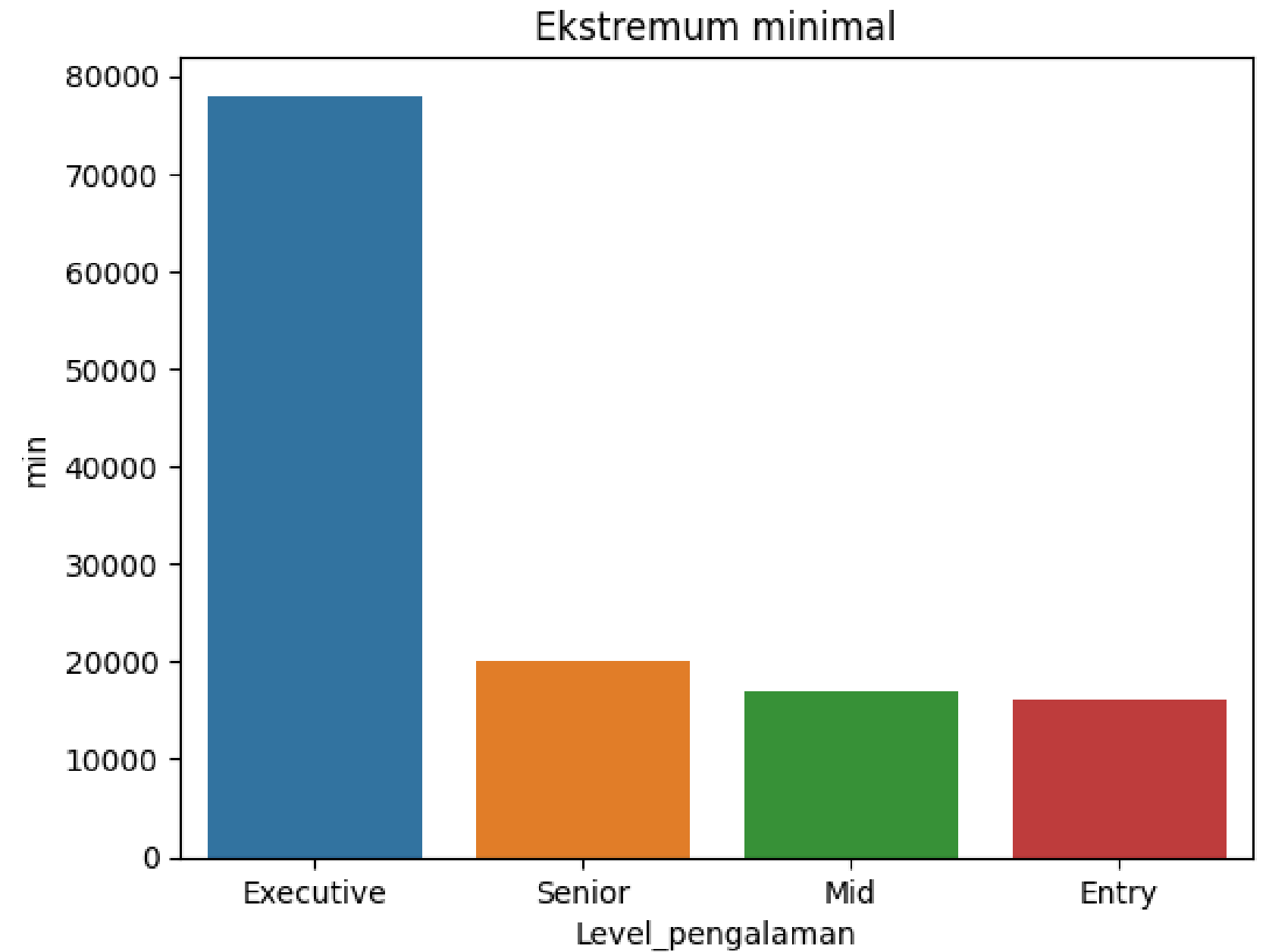
STEP 6 |

ANALISIS STATISTIK

visualisasi dalam data ekstremum minimal

```
sns.barplot(  
    data=minmax,  
    x='Level_pengalaman',  
    y='min',  
    hue='Level_pengalaman',  
    ci=None  
).set(title='Ekstremum minimal')
```

Output

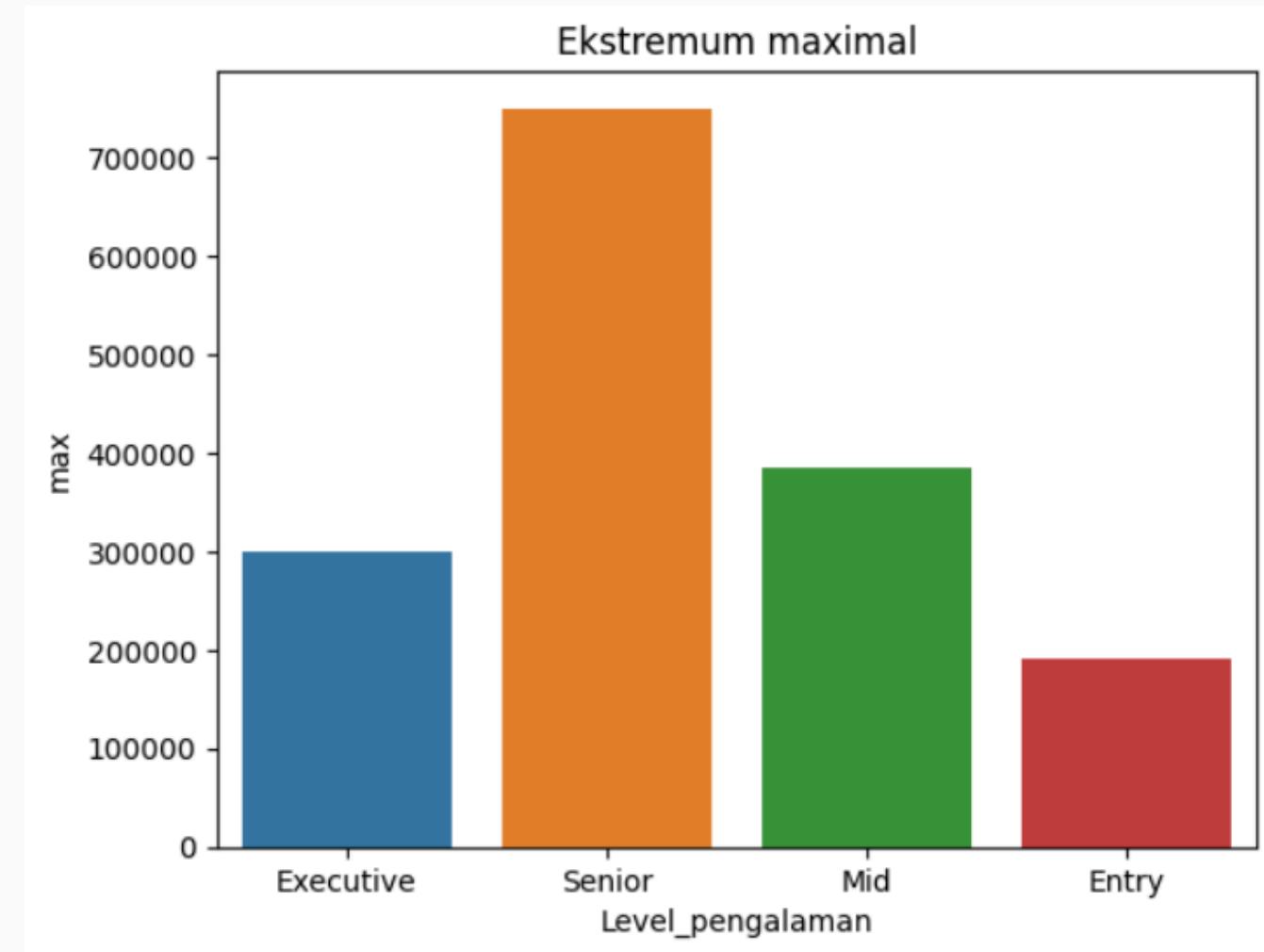


STEP 6 | ANALISIS STATISTIK

visualisasi dalam data ekstremum maximal

```
sns.barplot(  
    data=minmax,  
    x='Level_pengalaman',  
    y='max',  
    hue='Level_pengalaman',  
    ci=None  
).set(title='Ekstremum maximal')
```

Output



STEP 6 | ANALISIS STATISTIK

MENGHITUNG DISTRIBUSI FREKUENSI

```
[35] frequency_table = data_scientist['Tahun_kerja'].value_counts().reset_index()  
frequency_table.columns = ['Tahun_kerja', 'Frekuensi']  
print(frequency_table)# MENGHITUNG JUMLAH PEKERJA SEBAGAI DATA SCIENTIST DARI
```

Output

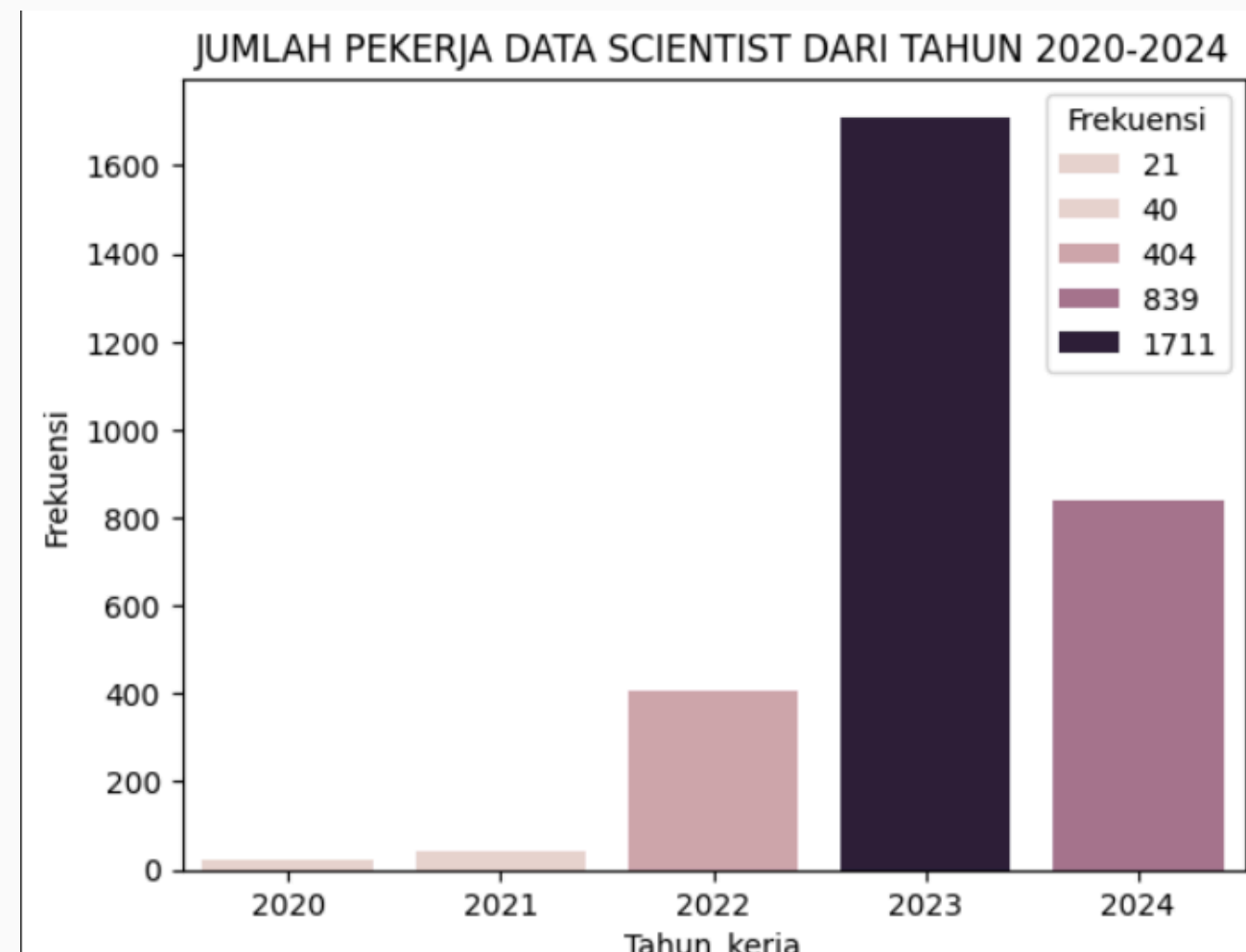
	Tahun_kerja	Frekuensi
0	2023	1711
1	2024	839
2	2022	404
3	2021	40
4	2020	21

STEP 6 | ANALISIS STATISTIK

visualisasi jumlah pekerja data scientist dari tahun 2020-2024

```
sns.barplot(  
    data=frequency_table,  
    x='Tahun_kerja',  
    y='Frekuensi',  
    hue='Frekuensi',  
    ci=None  
).set(title='JUMLAH PEKERJA DATA SCIENTIST DARI TAHUN 2020-2024')
```

Output :



STEP 6 | ANALISIS STATISTIK

Menghitung standar deviasi tahun 2020

```
standard_deviation = data_scientist.loc[data_scientist['Tahun_kerja'] == 2020, 'Gaji_dalam_usd'].std()  
print(standard_deviation)
```

Output

```
81436.64871089616
```

Berdasarkan syntac diatas kita dapat mengetahui bahwa jumlah 81436.64871089616 adalah deviasi standar dari gaji data scientist pada tahun 2020. Ini berarti bahwa gaji rata-rata mereka sekitar 81.436,65 USD lebih rendah dari gaji rata-rata.

STEP 6 | ANALISIS STATISTIK

Menghitung standar deviasi tahun 2021

```
standard_deviation = data_scientist.loc[data_scientist['Tahun_kerja'] == 2021, 'Gaji_dalam_usd'].std()  
print(standard_deviation)
```

Output

```
40573.319274722766
```

Berdasarkan syntac diatas Kita dapat mengetahui bahwa jumlah 40573.319274 adalah deviasi standar dari gaji dalam dolar untuk data scientist yang bekerja pada tahun 2021. Dengan kata lain, gaji rata-rata data scientist pada tahun 2021 adalah sekitar 40.573,32 dolar.

STEP 6 | ANALISIS STATISTIK

Menghitung standar deviasi tahun 2022

```
standard_deviation = data_scientist.loc[data_scientist['Tahun_kerja'] == 2022, 'Gaji_dalam_usd'].std()  
print(standard_deviation)
```

Output

```
53191.03555578297
```

Berdasarkan syntac diatas kita dapat mengetahui bahwa jumlah 53191.035555 adalah deviasi standar dari gaji data scientist pada tahun 2023. Artinya, gaji rata-rata data scientist pada tahun 2023 sekitar 53.191.55 USD lebih rendah dari gaji rata-rata mereka.

STEP 6 | ANALISIS STATISTIK

Menghitung standar deviasi tahun 2023

```
standard_deviation = data_scientist.loc[data_scientist['Tahun_kerja'] == 2023, 'Gaji_dalam_usd'].std()  
print(standard_deviation)
```

Output

```
61794.28254661448
```

STEP 6 | ANALISIS STATISTIK

Menghitung standar deviasi tahun 2024

```
standard_deviation = data_scientist.loc[data_scientist['Tahun_kerja'] == 2024, 'Gaji_dalam_usd'].std()  
print(standard_deviation)
```

Output

```
62228.69676062714
```

Berdasarkan syntac diatas kami dapat mengetahui bahwa gaji sebagai data scientist pada tahun 2024 rata-rata 62.228.70 USD lebih rendah dari gaji rata-rata.

STEP 6 | ANALISIS STATISTIK

Menghitung standar deviasi tahun 2020 - 2024

```
#menghitung standar deviasi  
standard_deviation = data_scientist['Gaji_dalam_usd'].std()  
print(standard_deviation)
```

Output

```
62296.32302064552
```

Berdasarkan syntac diatas kami dapat menyimpulkan bahwa: Dalam konteks ini, nilai standar deviasi sebesar 62296.32 USD menunjukkan variasi yang signifikan dalam gaji para data scientist dalam data. Artinya, ada beberapa data scientist yang menerima gaji yang jauh di atas rata-rata, dan ada juga yang menerima gaji yang jauh di bawah rata-rata.

STEP 7 | ANALISIS KORELASI

melakukan uji korelasi terhadap gaji dan tahun kerja

```
correlation = df['Gaji_dalam_usd'].corr(df['Tahun_kerja'])  
print(correlation)
```

Output

```
0.09087347570331308
```

Hubungan antara kedua variabel sangat kuat, karena nilai 0,9 hampir sama dengan 1. Artinya, tahun kerja memiliki pengaruh yang signifikan terhadap gaji dalam dolar di dataset ini

STEP 7 |

KESIMPULAN

Hasil analisis tentang peningkatan gaji data scientist dari tahun 2020 hingga 2024 menunjukkan tren peningkatan gaji umum di semua level pengalaman(Executive, Senior, Mid, dan Entry) karena kebutuhan akan pengolahan data yang kompleks dan pengambilan keputusan berbasis data semakin meningkat.

Secara keseluruhan, temuan analisis ini memberikan gambaran penting tentang bagaimana kompensasi pekerjaan data scientist berkembang. Selain itu, mereka memiliki kemampuan untuk membantu bisnis, tenaga kerja, dan pemangku kebijakan memahami dinamika pasar tenaga kerja di bidang ini.

*"Data itu seperti sampah.
Sebaiknya Anda tahu apa yang
akan Anda lakukan dengannya
sebelum Anda
mengumpulkannya."*

~ Mark Twain

TERIMA KASIH
