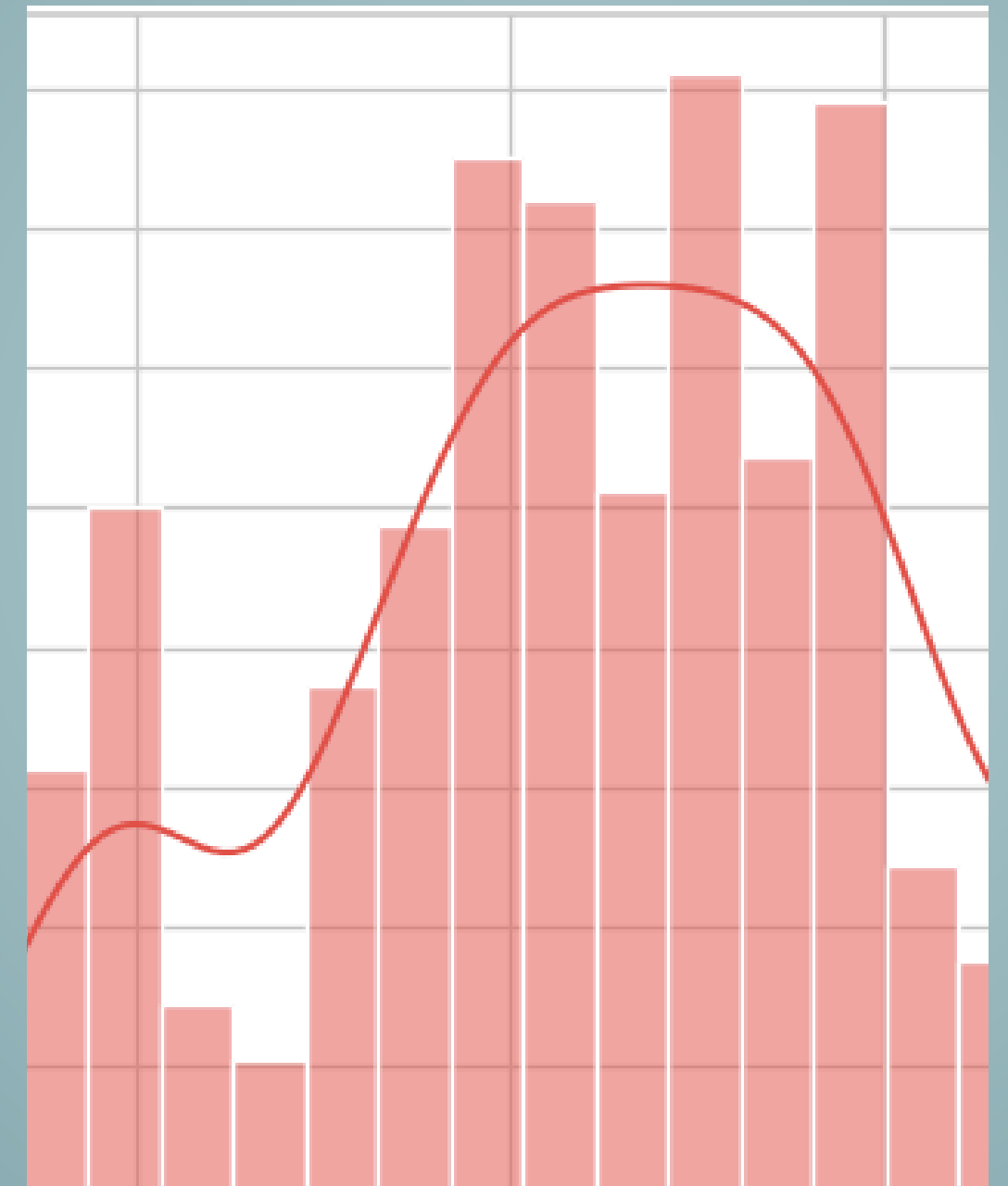

DATA ANALYST

Project

Dinda Mestika

<https://github.com/dindamestika/Data-Analyst-Position-Salary-Analysis>



Analysis Target

Analisis Gaji Data Analyst di Berbagai Company

- Distribusi estimasi gaji untuk posisi Data Analyst
- Perbandingan gaji berdasarkan lokasi (misalnya, apakah gaji di New York lebih tinggi dibanding kota lain?)
- Perbandingan gaji berdasarkan industri (misalnya, apakah Data Analyst di sektor teknologi mendapatkan gaji lebih tinggi dibanding sektor kesehatan?)

Data Quality Check

Meta Data Table

Job Title → Nama posisi pekerjaan.

Salary Estimate → Estimasi gaji untuk posisi tersebut.

Job Description → Deskripsi pekerjaan.

Rating → Rating perusahaan berdasarkan ulasan.

Company Name → Nama perusahaan yang membuka lowongan.

Location → Lokasi pekerjaan.

Headquarters → Lokasi kantor pusat perusahaan.

Size → Ukuran perusahaan berdasarkan jumlah karyawan.

Founded → Tahun perusahaan didirikan.

Type of Ownership → Jenis kepemilikan perusahaan (misalnya nonprofit, private company).

Industry → Industri tempat perusahaan beroperasi.

Sector → Sektor bisnis perusahaan.

Revenue → Pendapatan tahunan perusahaan.

Competitors → Nama pesaing utama perusahaan.

Easy Apply → Apakah lowongan ini memiliki fitur Easy Apply (True/False).

Data Quality Check

Unnamed: 0		Job Title	Salary Estimate	Job Description	Rating	Company Name	Location	Headquarters	Size	Founded	Type of ownership	Industry	Sector	Revenue	Competitors	Easy Apply
0	0	Data Analyst, Center on Immigration and Justic...	\$37K-\$66K (Glassdoor est.)	Are you eager to roll up your sleeves and ham...	3.2	Vera Institute of Justice\n3.2	New York, NY	New York, NY	201 to 500 employees	1961	Nonprofit Organization	Social Assistance	Non-Profit	\$100 to \$500 million (USD)	-1	True
1	1	Quality Data Analyst	\$37K-\$66K (Glassdoor est.)	Overview\n\nProvides analytical and technical ...	3.8	Visiting Nurse Service of New York\n3.8	New York, NY	New York, NY	10000+ employees	1893	Nonprofit Organization	Health Care Services & Hospitals	Health Care	\$2 to \$5 billion (USD)	-1	-1
2	2	Senior Data Analyst, Insights & Analytics Team...	\$37K-\$66K (Glassdoor est.)	We're looking for a Senior Data Analyst who ha...	3.4	Squarespace\n3.4	New York, NY	New York, NY	1001 to 5000 employees	2003	Company - Private	Internet	Information Technology	Unknown / Non-Applicable	GoDaddy	-1
3	3	Data Analyst	\$37K-\$66K (Glassdoor est.)	Requisition NumberRR-0001939\nRemote:Yes\nWe c...	4.1	Celerity\n4.1	New York, NY	McLean, VA	201 to 500 employees	2002	Subsidiary or Business Segment	IT Services	Information Technology	\$50 to \$100 million (USD)	-1	-1
4	4	Reporting Data Analyst	\$37K-\$66K (Glassdoor est.)	ABOUT FANDUEL GROUP\n\nFanDuel Group is a worl...	3.9	FanDuel\n3.9	New York, NY	New York, NY	501 to 1000 employees	2009	Company - Private	Sports & Recreation	Arts, Entertainment & Recreation	\$100 to \$500 million (USD)	DraftKings	True

Data Quality Check

Penyesuaian *features data type*

#	Column	Non-Null Count	Dtype
0	Unnamed: 0	2253 non-null	int64
1	Job Title	2253 non-null	object
2	Salary Estimate	2253 non-null	object
3	Job Description	2253 non-null	object
4	Rating	2253 non-null	float64
5	Company Name	2252 non-null	object
6	Location	2253 non-null	object
7	Headquarters	2253 non-null	object
8	Size	2253 non-null	object
9	Founded	2253 non-null	int64
10	Type of ownership	2253 non-null	object
11	Industry	2253 non-null	object
12	Sector	2253 non-null	object
13	Revenue	2253 non-null	object
14	Competitors	2253 non-null	object
15	Easy Apply	2253 non-null	object

dtypes: float64(1), int64(2), object(13)

- Konversi "Salary Estimate" menjadi angka (Min Salary & Max Salary)
- Konversi "Competitors" menjadi kategori
- Konversi "Easy Apply" menjadi Boolean
- Konversi "Founded" menjadi usia perusahaan

Pre-Processing

Tidak ada *duplicate data* dan tidak ada *missing value*

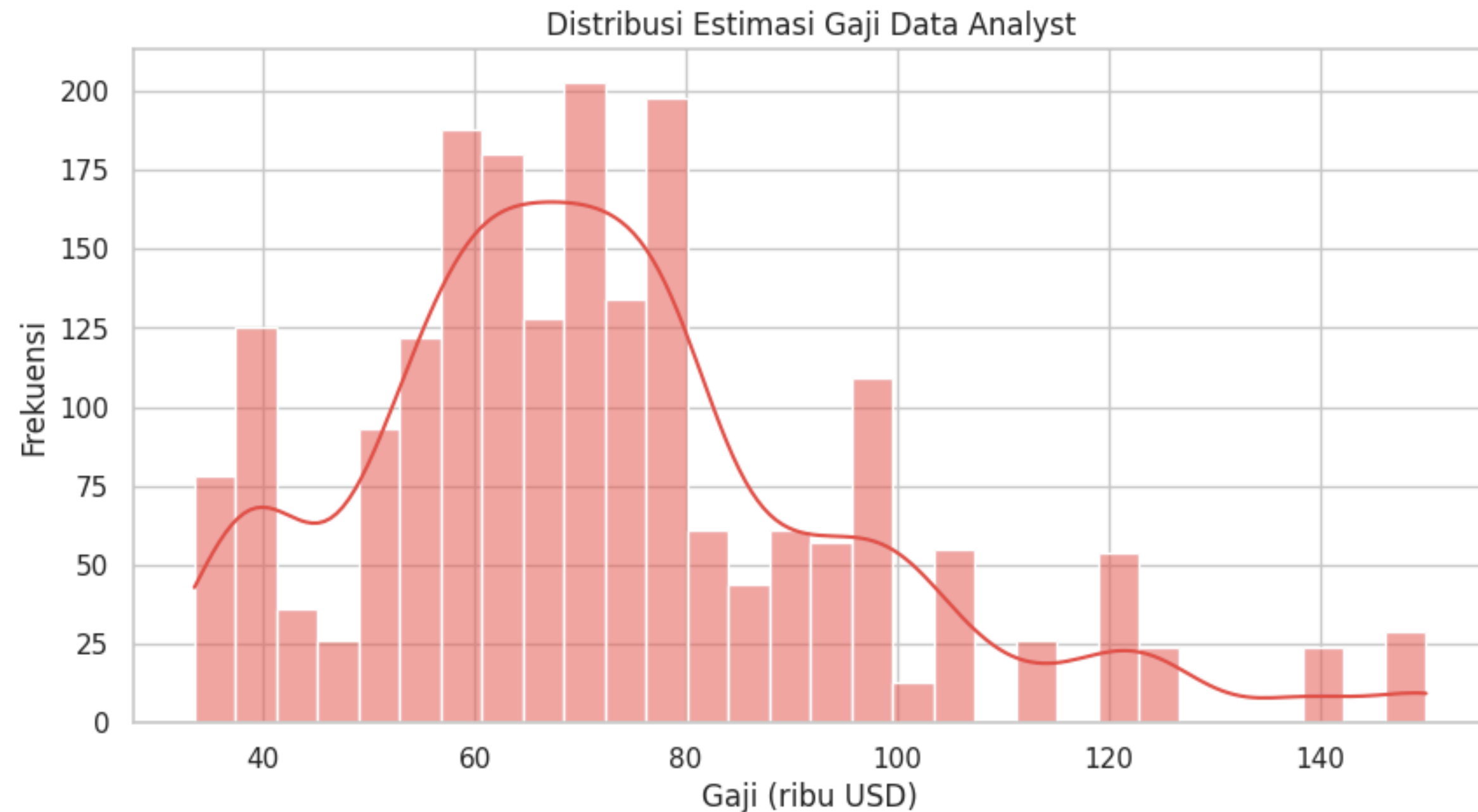
```
▼ Duplicate  
[ ] df.duplicated().sum()  
0
```

duplicate

missing value

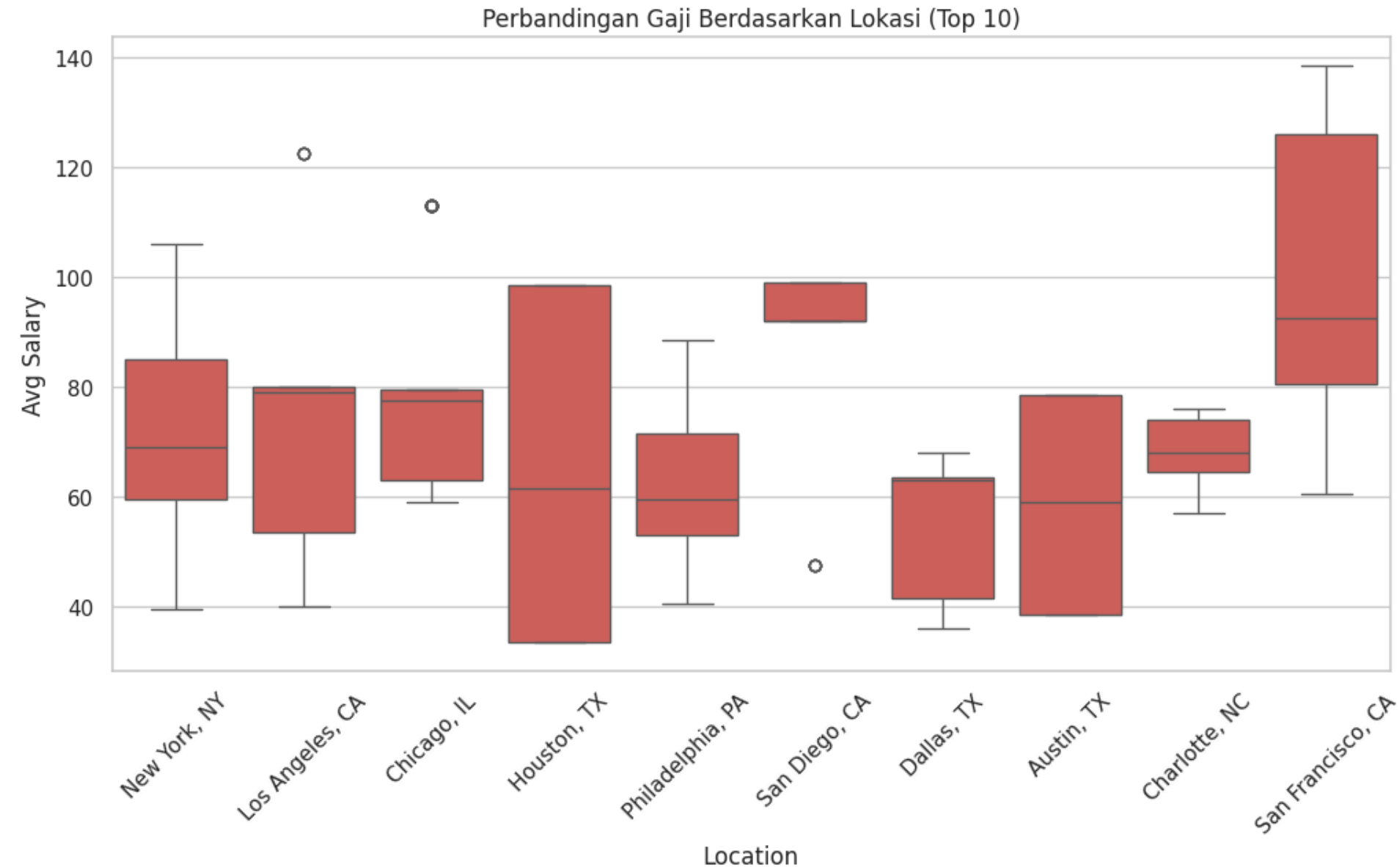
0
Unnamed: 0 0
Type of ownership 0
Max Salary 0
Min Salary 0
Easy Apply 0
Competitors 0
Revenue 0
Sector 0
Industry 0
Founded 0
Job Title 0
Size 0
Headquarters 0
Location 0
Company Name 0
Rating 0
Job Description 0
Salary Estimate 0
Company Age 0

Data Insight



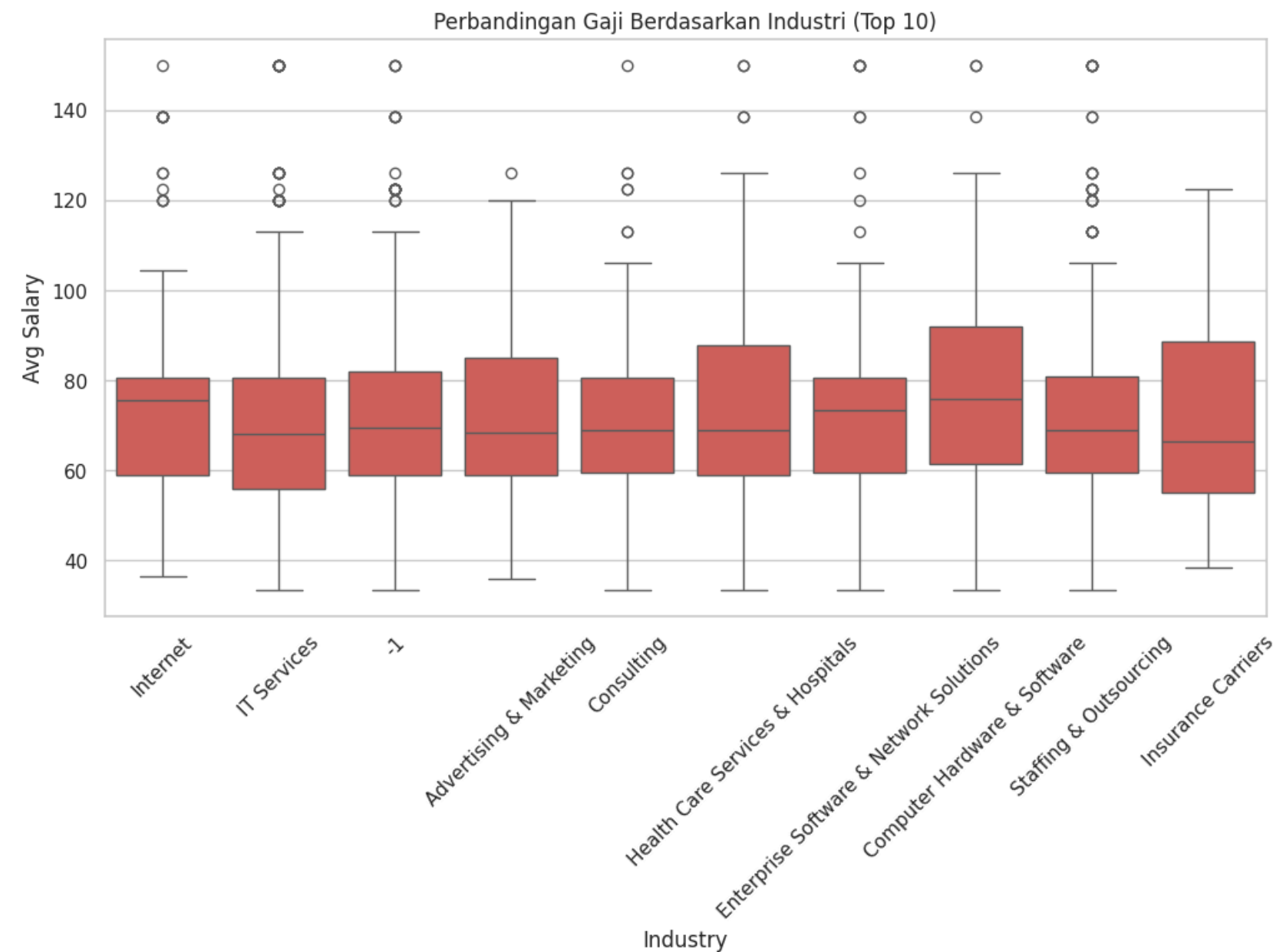
- Gaji Data Analyst sebagian besar berada di rentang 60-80 ribu USD.
- Sebagian kecil Data Analyst bisa mendapatkan gaji di atas 100 ribu USD
- Masih ada Data Analyst yang dibayar lebih rendah (30-50 ribu USD)

Data Insight



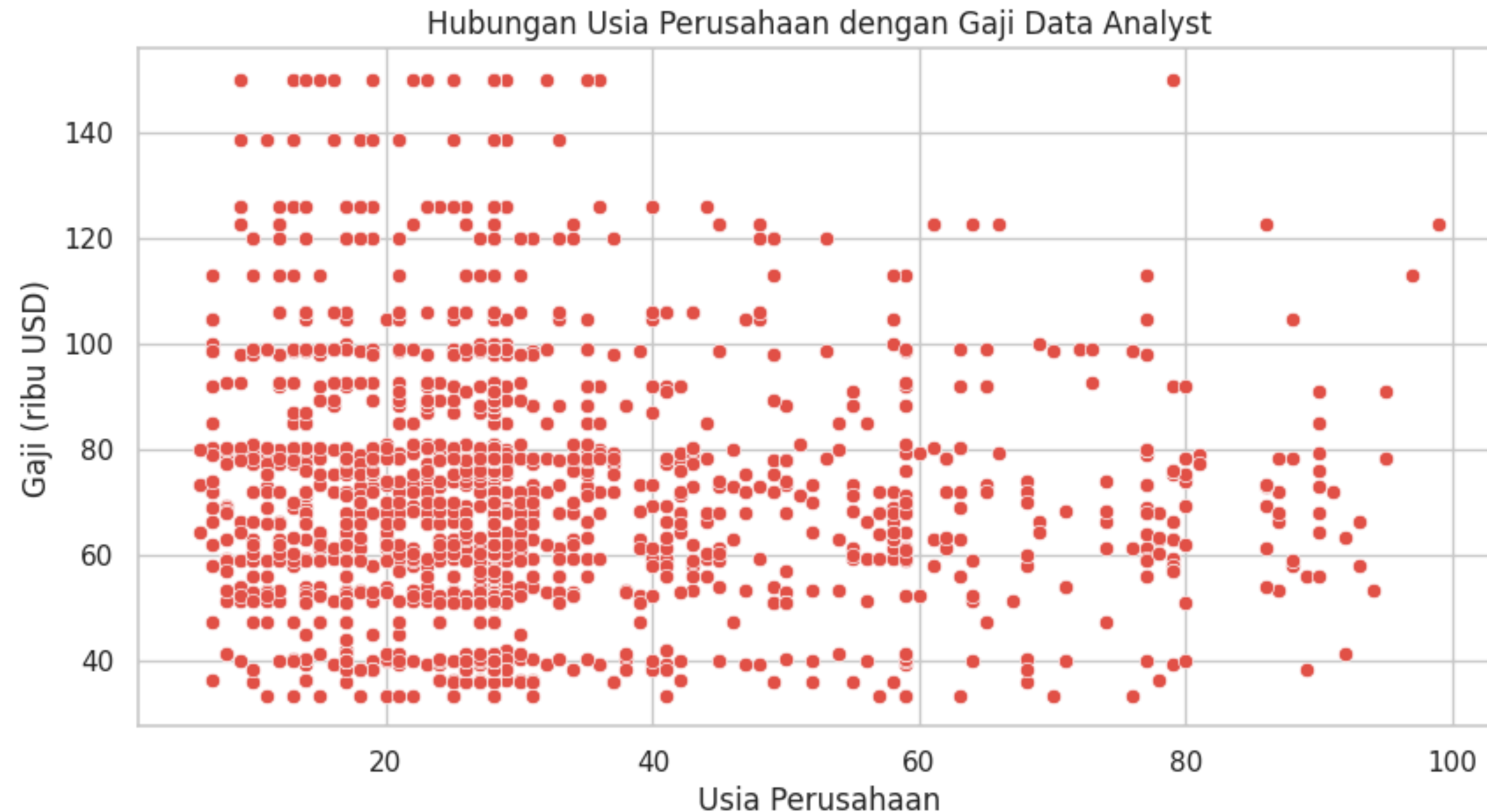
- San Francisco, CA adalah kota terbaik untuk Data Analyst yang mencari gaji tinggi
- Houston, TX memiliki rentang gaji yang luas, yang bisa berarti peluang lebih banyak tetapi juga risiko lebih tinggi tergantung pada industri
- Dallas, TX dan Philadelphia, PA cenderung memiliki gaji lebih rendah
- Chicago, IL dan Charlotte, NC menghasilkan gaji lebih stabil tanpa banyak fluktuasi.

Data Insight



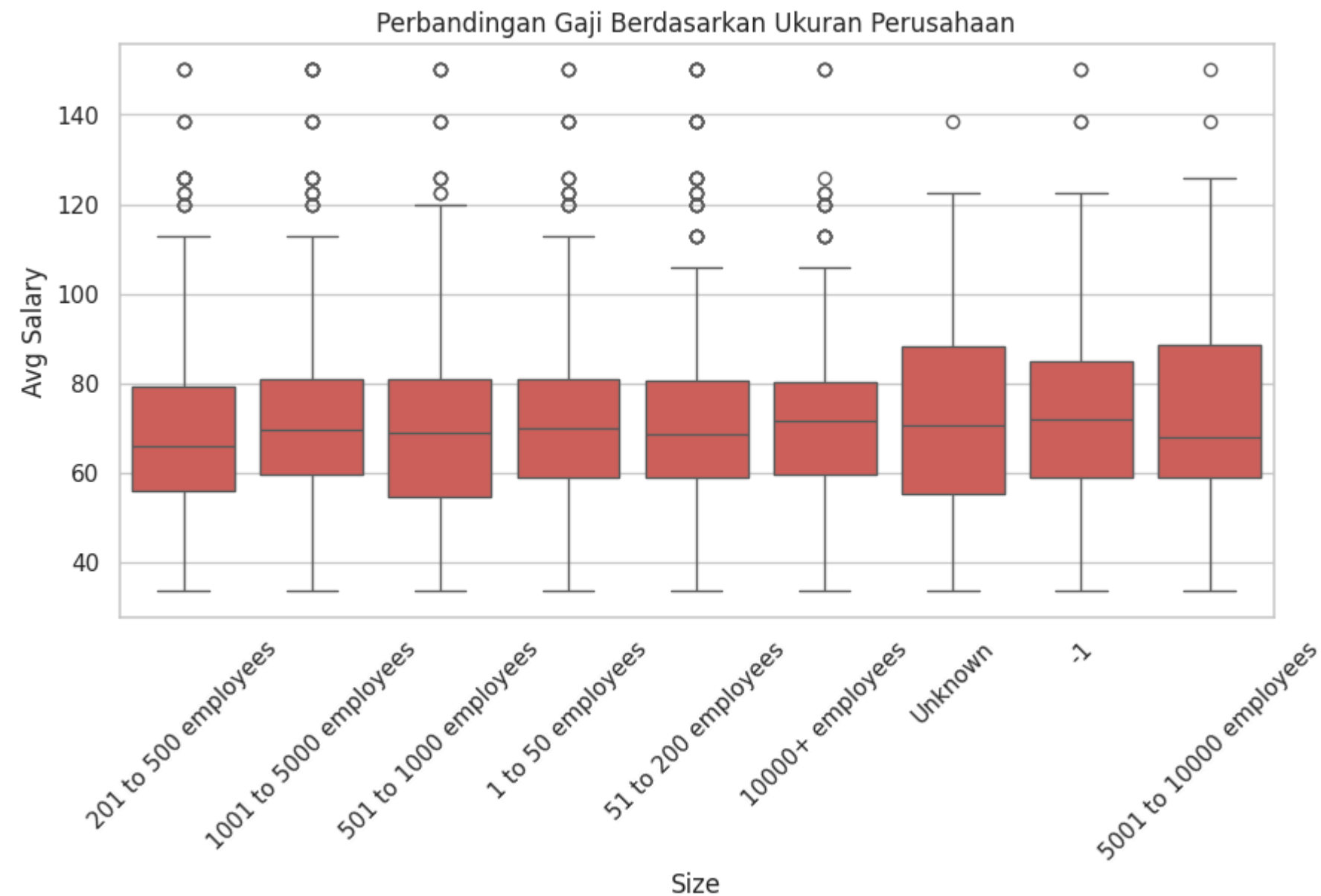
- Sektor industri Enterprise Software & Network Solutions, Consulting, dan Computer Hardware & Software menghasilkan gaji paling tinggi
- Sektor IT Services dan Insurance Carriers menghasilkan gaji paling stabil
- Banyak outlier dengan gaji tinggi dalam beberapa industri, menunjukkan bahwa dengan pengalaman dan keahlian yang tepat, ada peluang untuk mendapatkan gaji di atas rata-rata.

Data Insight



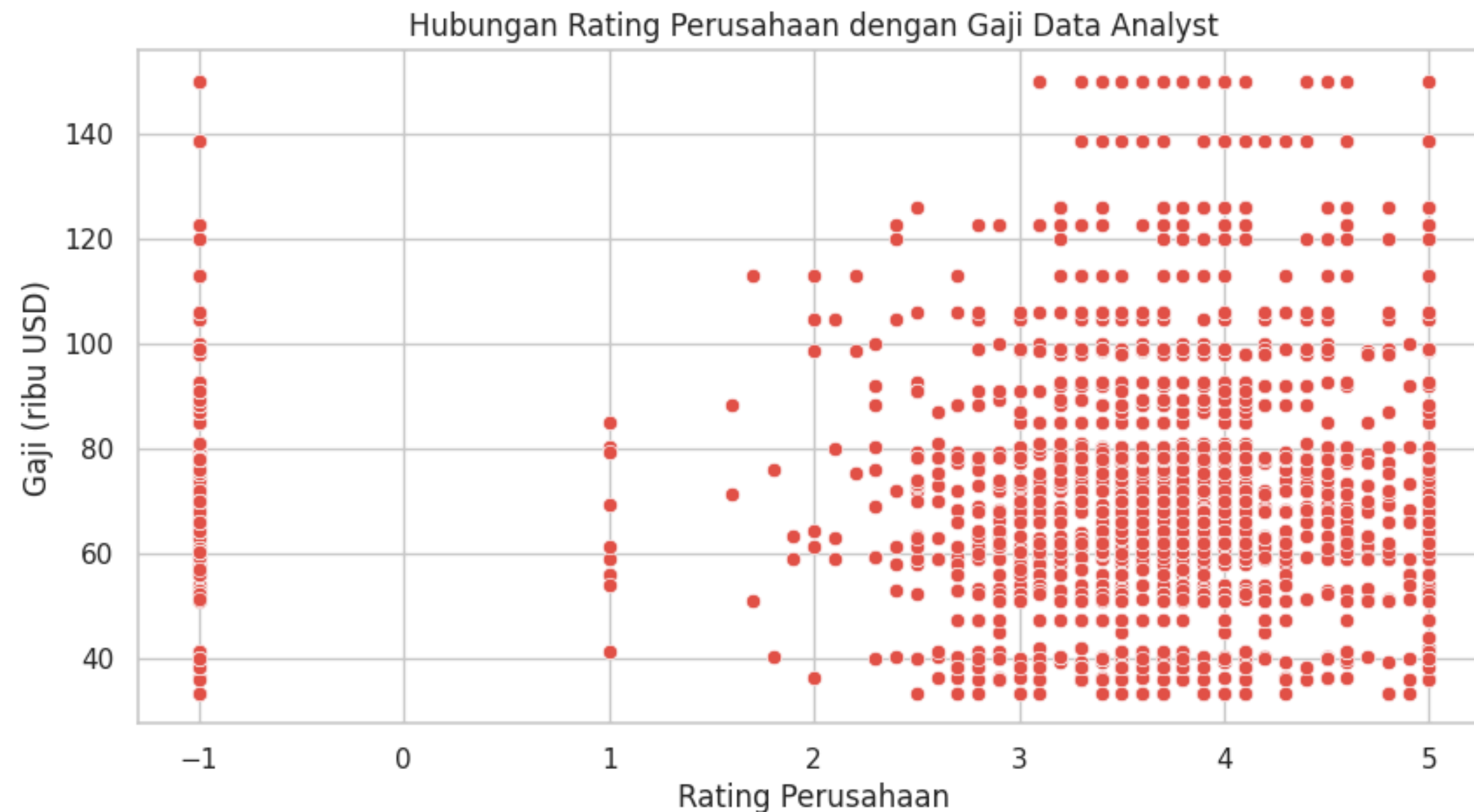
- Gaji Data Analyst bisa tinggi di perusahaan baru maupun lama.
- Perusahaan muda (startup) bisa menawarkan gaji tinggi, tetapi ada juga risiko mendapatkan gaji lebih rendah.
- Perusahaan lama cenderung lebih stabil dalam rentang gaji, tetapi tetap ada peluang mendapatkan gaji tinggi.

Data Insight



- Ukuran perusahaan tidak memiliki dampak besar terhadap gaji Data Analyst.
- Baik perusahaan kecil maupun besar memiliki peluang untuk menawarkan gaji tinggi.

Data Insight



- Gaji Data Analyst tidak terlalu bergantung pada rating perusahaan.
- Perusahaan dengan rating rendah masih bisa menawarkan gaji kompetitif.

Modelling

▼ Data Splitting

- Data Train 60%
- Data Validation 20%
- Data Testing 20%

```
[ ] df = pd.read_csv('DataAnalyst.csv')  
df.shape
```

```
↗ (2253, 16)
```

```
[ ] from sklearn.model_selection import train_test_split
```

```
[ ] df_full_train, df_test = train_test_split(df, test_size=0.2, random_state=1)  
df_train, df_valid = train_test_split(df_full_train, test_size=0.25, random_state=1)
```

```
[ ] df_train.shape, df_valid.shape, df_test.shape
```

```
↗ ((1351, 16), (451, 16), (451, 16))
```

```
[ ] 1351/2253, 451/2253, 451/2253
```

```
↗ (0.5996449178872614, 0.20017754105636928, 0.20017754105636928)
```

▼ Random Forest

```
[ ] from sklearn.ensemble import RandomForestClassifier
```

```
[ ] model = RandomForestClassifier(random_state=42)  
model.fit(X_train_final, y_train_final)
```

```
↗  
▼ RandomForestClassifier ⓘ ⓘ  
RandomForestClassifier(random_state=42)
```

```
[ ] from sklearn.metrics import accuracy_score
```

```
y_valid_pred_class = model.predict(X_valid_final)  
accuracy = accuracy_score(y_valid_final, y_valid_pred_class)  
print(f'Accuracy: {accuracy:.3f}')
```

```
↗ Accuracy: 0.991
```

Menggunakan model Random Forest
dengan akurasi model 99%

▼ Performance Stability Check

```
[ ] X_full_train_final = pd.concat([X_train_final, X_valid_final])  
    y_full_train_final = pd.concat([y_train_final, y_valid_final])
```

```
[ ] model = RandomForestClassifier(random_state=42)  
    model.fit(X_full_train_final, y_full_train_final)
```



RandomForestClassifier ⓘ ⓘ
RandomForestClassifier(random_state=42)

```
[ ] from sklearn.metrics import accuracy_score  
  
    y_test_pred_class = model.predict(X_test_final) # Prediksi label kelas  
    accuracy = accuracy_score(y_test_final, y_test_pred_class)  
    print(f'Accuracy: {accuracy:.3f}')
```



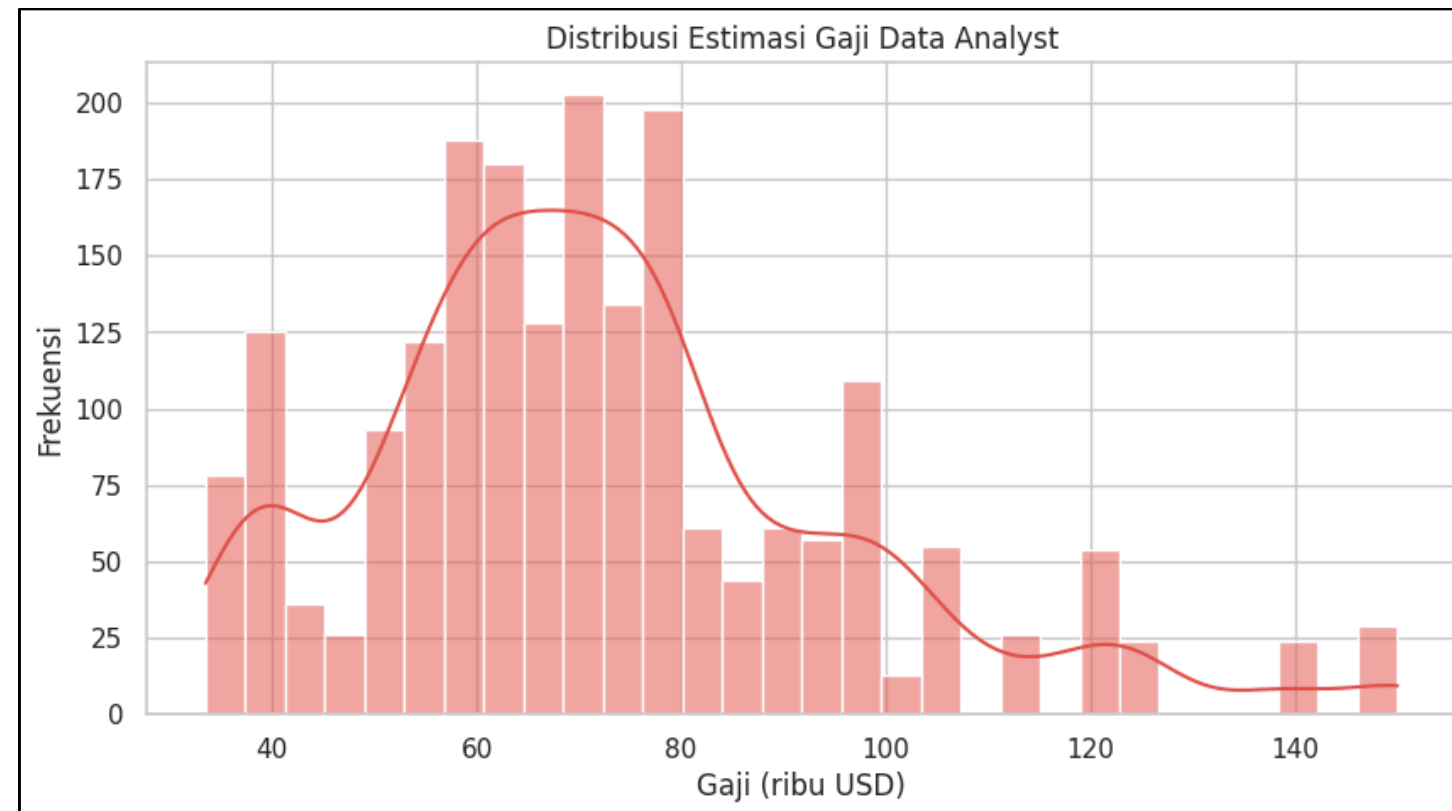
Accuracy: 0.998

Pemeriksaan Performa

Akurasi performa 99.8%

Result Interpretation

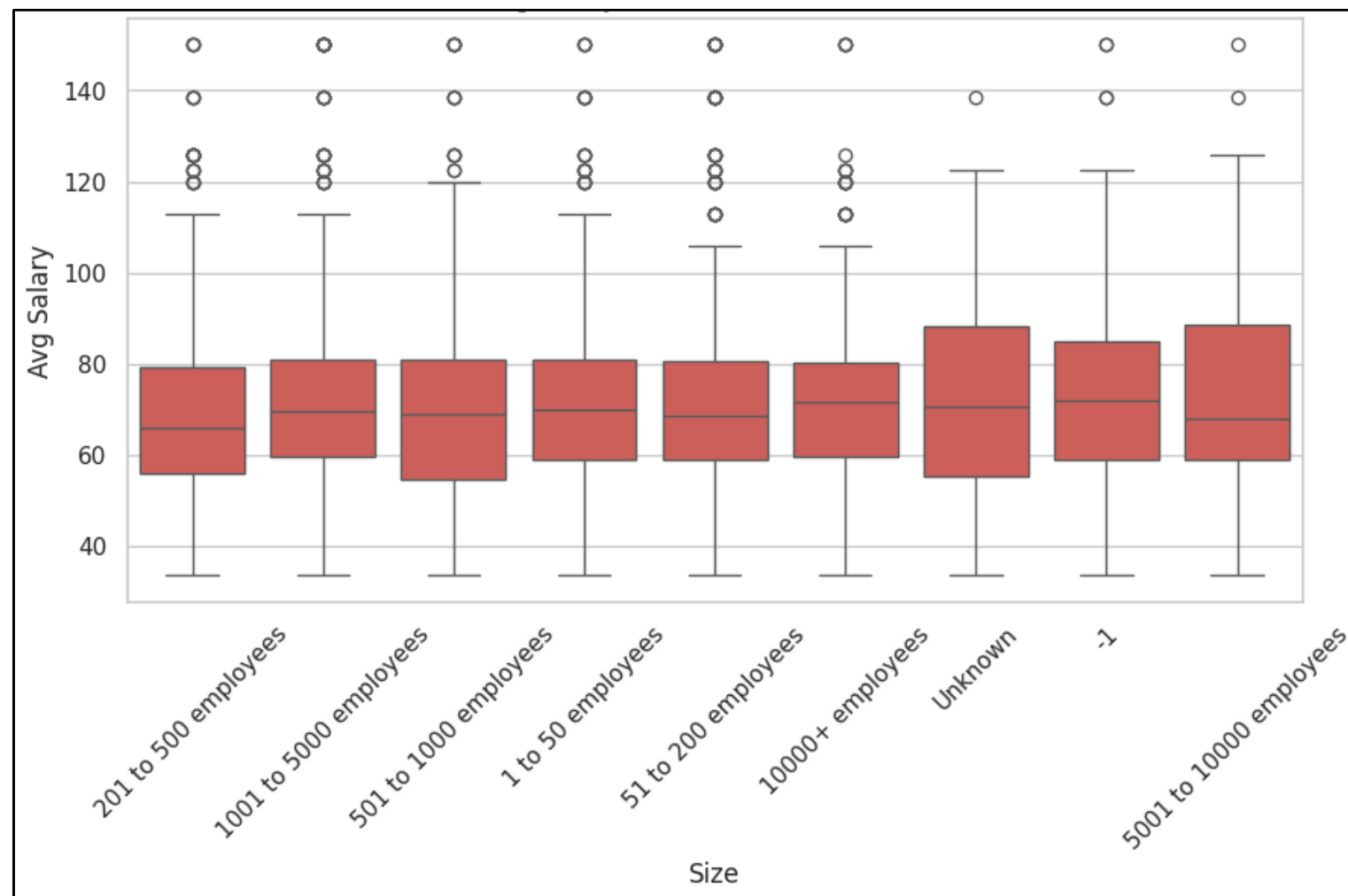
Distribusi Gaji Data Analyst Secara Umum



- Gaji Data Analyst memiliki rentang yang cukup luas, mulai dari sekitar 40 ribu USD hingga lebih dari 140 ribu USD per tahun.
- Sebagian besar gaji berada di rentang 60 hingga 80 ribu USD, menunjukkan bahwa ini adalah kisaran umum untuk posisi Data Analyst.
- Ada beberapa outlier yang memiliki gaji jauh lebih tinggi (di atas 120 ribu USD), yang kemungkinan dipengaruhi oleh faktor tambahan seperti pengalaman, keterampilan spesifik, atau lokasi pekerjaan.

Result Interpretation

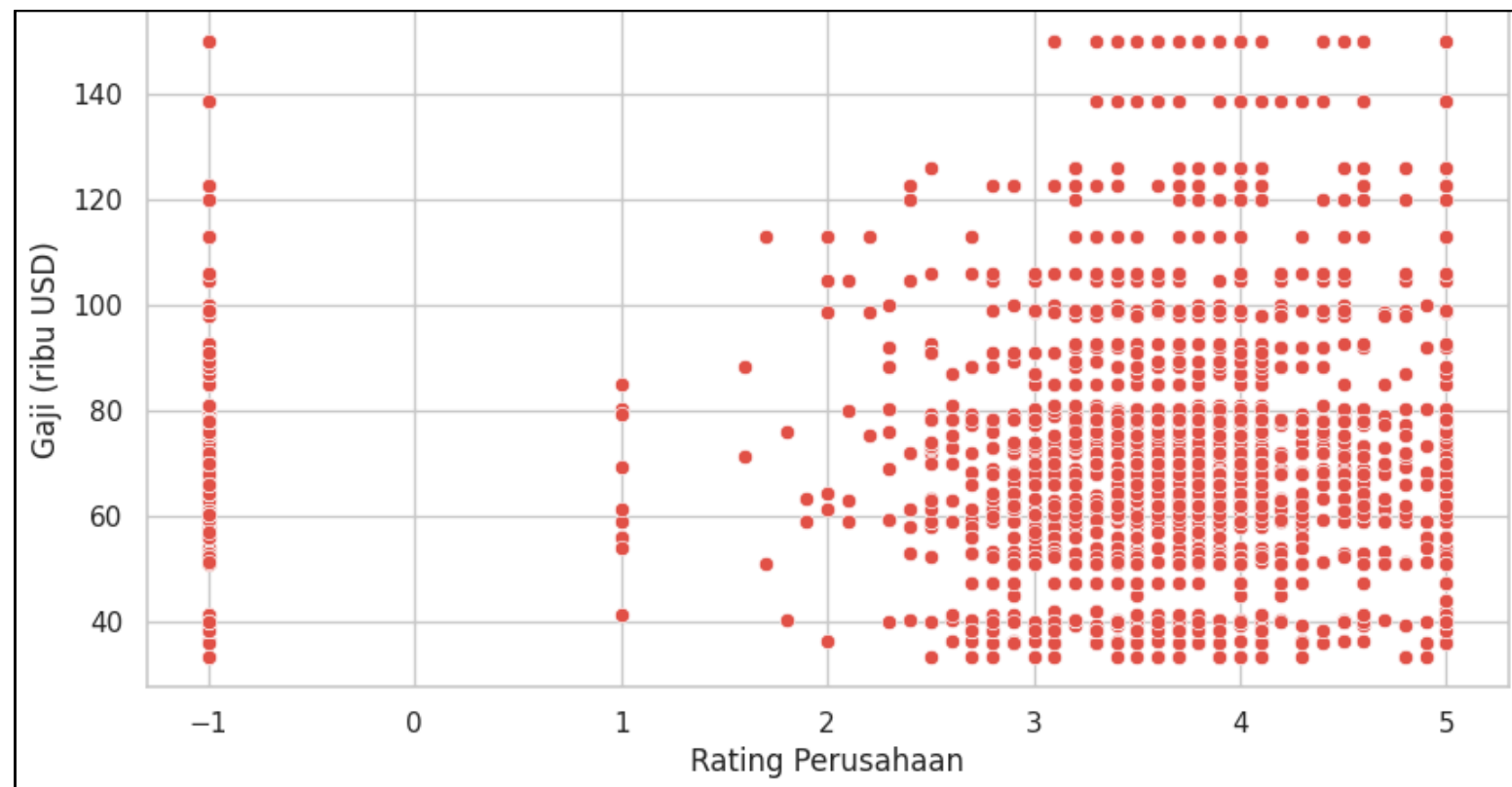
Perbandingan Gaji Berdasarkan Ukuran Perusahaan



- Ukuran perusahaan tidak secara signifikan mempengaruhi gaji Data Analyst.
- Baik perusahaan kecil (1-50 karyawan) maupun perusahaan besar (10.000+ karyawan) memiliki distribusi gaji yang serupa.
- Namun, beberapa perusahaan besar cenderung memiliki gaji maksimal yang lebih tinggi, meskipun median gaji tetap mirip dengan perusahaan kecil.
- Ini menunjukkan bahwa kesempatan mendapatkan gaji tinggi lebih mungkin di perusahaan besar, tetapi bukan jaminan.

Result Interpretation

Hubungan Rating Perusahaan dengan Gaji Data Analyst



- Tidak ada korelasi yang jelas antara rating perusahaan dan gaji Data Analyst.
- Perusahaan dengan rating rendah hingga tinggi (1-5) memiliki distribusi gaji yang serupa, dengan beberapa perusahaan rating rendah tetap menawarkan gaji yang kompetitif.
- Perusahaan dengan rating tinggi mungkin menawarkan lingkungan kerja lebih baik, tetapi tidak selalu berarti gaji lebih tinggi.
- Banyak outlier dengan gaji tinggi ditemukan di semua rating, menunjukkan bahwa faktor lain seperti industri atau lokasi lebih berpengaruh terhadap gaji dibanding rating perusahaan.

Conclusion

- Gaji Data Analyst memiliki distribusi yang cukup luas, tetapi mayoritas berada di kisaran 60-80 ribu USD per tahun.
- Ukuran perusahaan tidak terlalu berpengaruh terhadap gaji. Meskipun perusahaan besar memiliki beberapa posisi dengan gaji lebih tinggi, median gaji tetap mirip dengan perusahaan kecil.
- Rating perusahaan juga bukan faktor utama dalam menentukan gaji. Perusahaan dengan rating rendah masih bisa menawarkan gaji yang kompetitif.
- Faktor lain seperti industri, lokasi kerja, dan pengalaman kemungkinan lebih menentukan variasi gaji dibanding ukuran atau rating perusahaan.



Data Analyst Enthusiast

Thank You

+62-853-7412-5601

Padang, West Sumatera, Indonesia

mestikadinda@gmail.com

<https://github.com/dindamestika/Data-Analyst-Position-Salary-Analysis>