

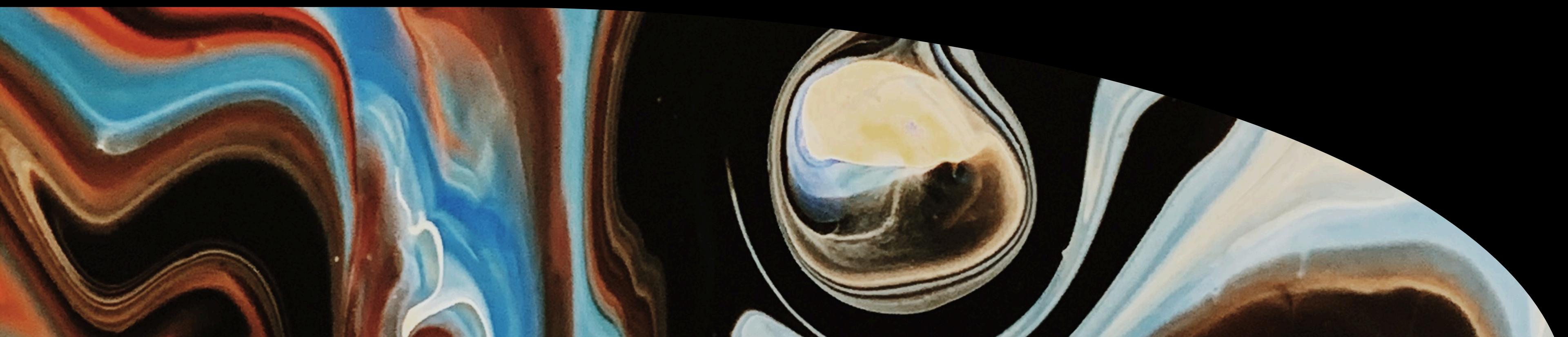


ANALISIS FAKTOR CUACA DAN HARI KERJA TERHADAP JUMLAH PENYEWAAN SEPEDA PADA BIKE SHARING DATASET MENGGUNAKAN REGRESI LINEAR BERGANDA

KELOMPOK 8

Presented by:

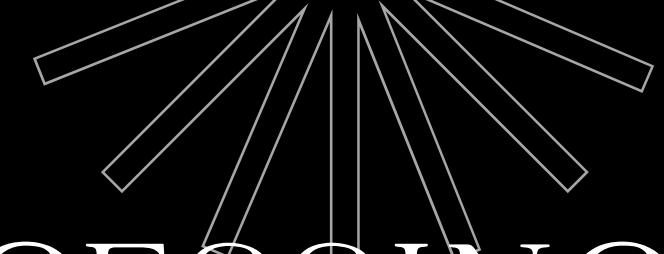
Nia Ayu Agustin (24031554081) Siti Aisyah Febriyanti (24031554158) Audy Alycia (24031554179)





Dataset yang digunakan adalah Bike Sharing Dataset dari UCI Machine Learning Repository yang berisi data penyewa selama tahun 2011-2012 di Washington D.C, Amerika Serikat. Dataset ini mencakup ribuan observasi dengan variabel yang merepresentasikan kondisi waktu, hari kerja, dan cuaca seperti suhu, kelembapan, kecepatan angin, musim, serta jumlah penyewa sepeda. Data ini tidak memiliki missing value, sehingga dapat langsung digunakan untuk analisis. Dataset ini dipilih karena lengkap dan relevan untuk dilakukan pengkajian pengaruh faktor cuaca dan hari kerja terhadap jumlah penyewa sepeda menggunakan analisis statistik deskriptif dan inferensial.

DATASET



PREPROCESSING

Dataset Bike Sharing dibaca menggunakan Python kemudian dilakukan pemilihan variabel utama yaitu jumlah sewa, suhu, kelembapan, kecepatan angin, dan hari kerja. Nama kolom disesuaikan agar lebih mudah untuk dipahami, sehingga data siap digunakan untuk analisis deskriptif dan inferensial.

```
data = df[['cnt', 'temp', 'hum', 'windspeed', 'workingday']]
```

```
data = data.rename(columns={  
    'cnt': 'jumlah_sewa',  
    'temp': 'suhu',  
    'hum': 'kelembapan',  
    'windspeed': 'kecepatan_angin',  
    'workingday': 'hari_kerja'  
})
```

```
data.head()
```

	jumlah_sewa	suhu	kelembapan	kecepatan_angin	hari_kerja
0	985	0.344167	0.805833	0.160446	0
1	801	0.363478	0.696087	0.248539	0
2	1349	0.196364	0.437273	0.248309	1
3	1562	0.200000	0.590435	0.160296	1
4	1600	0.226957	0.436957	0.186900	1



ANALISIS DESKRIPTIF

STATISTIK DESKRIPTIF DASAR

Data dianalisis menggunakan statistik deskriptif untuk menggambarkan karakteristik penyewaan sepeda berdasarkan nilai mean, median, dan standar deviasi. Rata-rata jumlah sewa per hari sebesar 4.504 sepeda dengan median 4.548 sepeda, sedangkan standar deviasi 1.937 menunjukkan variasi permintaan yang cukup tinggi.

Variable cuaca menunjukkan bahwa suhu dan kelembapan lebih bervariasi dibanding kecepatan angin. Selain itu, sebagian besar data berasal dari hari kerja, ditunjukkan oleh nilai mean 0,683995 dan median 1 pada hari kerja.

```
variabel = ['suhu', 'kelembapan', 'kecepatan_angin', 'hari_kerja', 'jumlah_sewa']

# Mean
print("\nMean:")
print(data[variabel].mean())

# Median
print("\nMedian:")
print(data[variabel].median())

# Sebaran (min, max, std)
print("\nSebaran (min, max, std):")
print(data[variabel].describe().T[['min', 'max', 'std']])
```

Mean:

suhu	0.495385
kelembapan	0.627894
kecepatan_angin	0.190486
hari_kerja	0.683995
jumlah_sewa	4504.348837

dtype: float64

Median:

suhu	0.498333
kelembapan	0.626667
kecepatan_angin	0.180975
hari_kerja	1.000000
jumlah_sewa	4548.000000

dtype: float64

Sebaran (min, max, std):

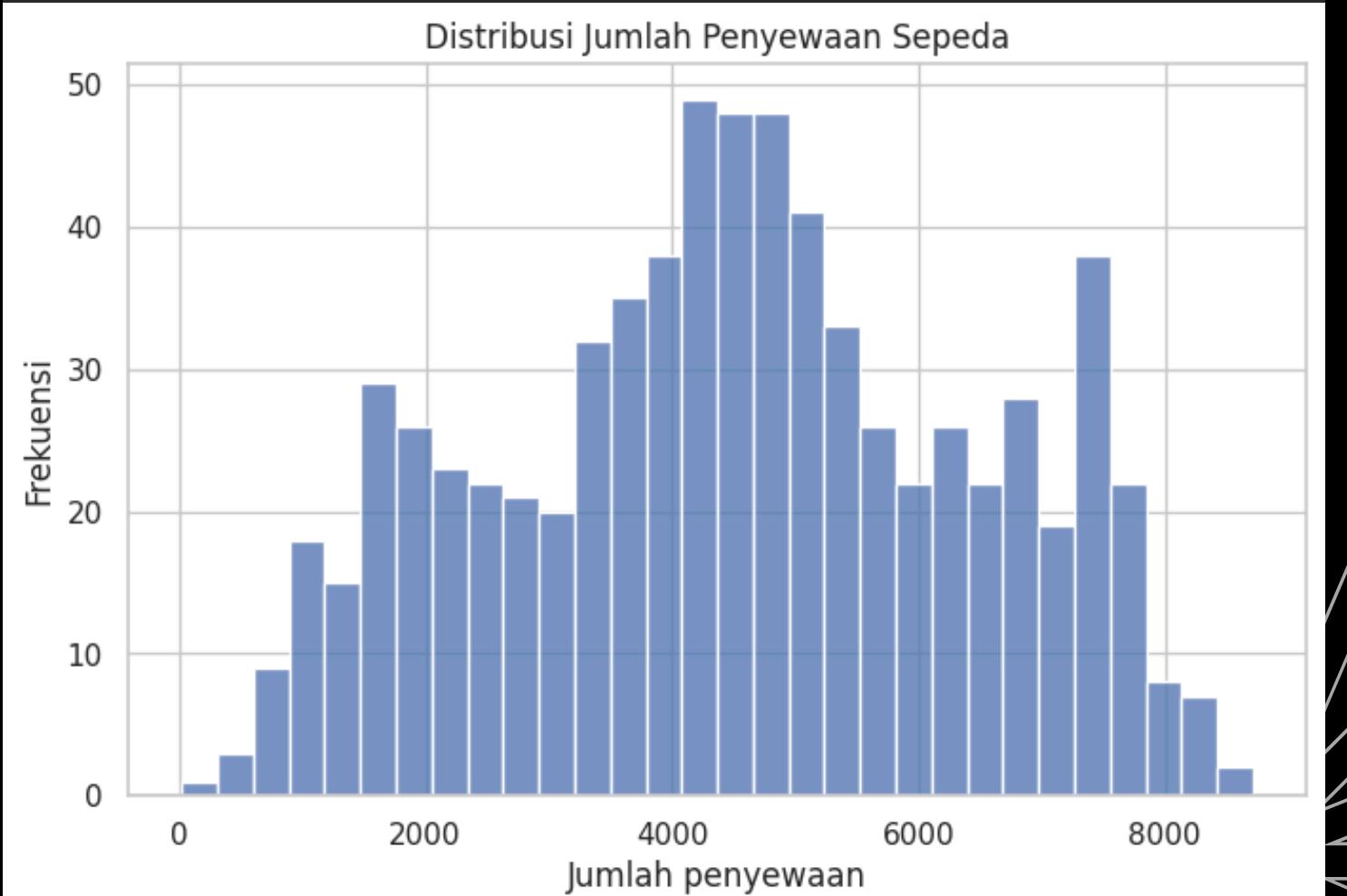
	min	max	std
suhu	0.059130	0.861667	0.183051
kelembapan	0.000000	0.972500	0.142429
kecepatan_angin	0.022392	0.507463	0.077498
hari_kerja	0.000000	1.000000	0.465233
jumlah_sewa	22.000000	8714.000000	1937.211452

DISTRIBUSI DATA

Histogram ini digunakan untuk melihat pola persebaran dan frekuensi jumlah penyewaan sepeda. Hasil grafik menunjukkan bahwa data paling banyak terkonsentrasi pada kisaran 4.000-5.000 penyewaan per hari.

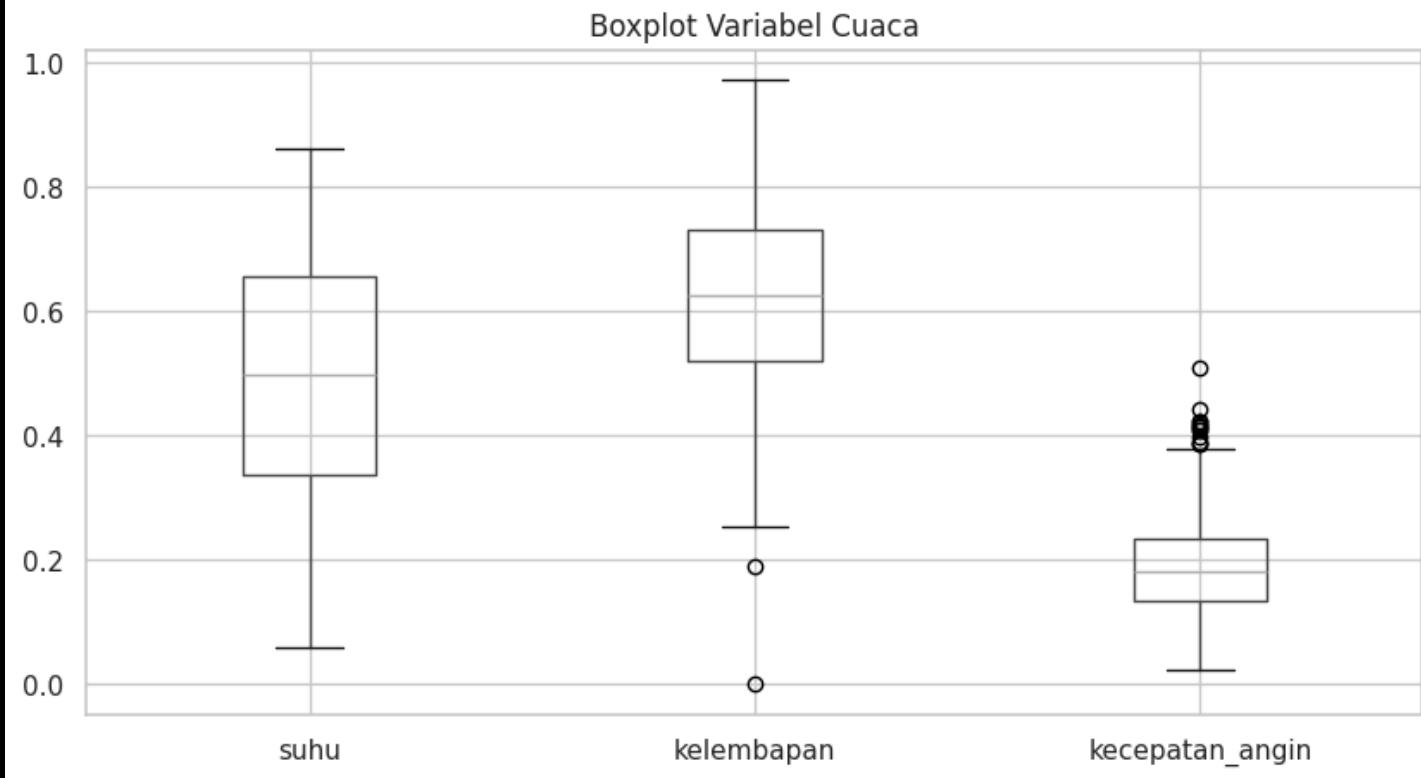
Selain itu, terlihat nilai sangat rendah dan sangat tinggi yang menandakan bahwa pada hari tertentu terjadi lonjakan atau penurunan permintaan. Hal ini menunjukkan bahwa penyewaan sepeda bersifat berubah-ubah dan tidak selalu stabil setiap harinya.

```
# Histogram jumlah penyewaan sepeda
plt.figure(figsize=(8,5))
sns.histplot(data['jumlah_sewa'], bins=30)
plt.title("Distribusi Jumlah Penyewaan Sepeda")
plt.xlabel("Jumlah penyewaan")
plt.ylabel("Frekuensi")
plt.show()
```



SEBARAN VARIABEL & HARI KERJA

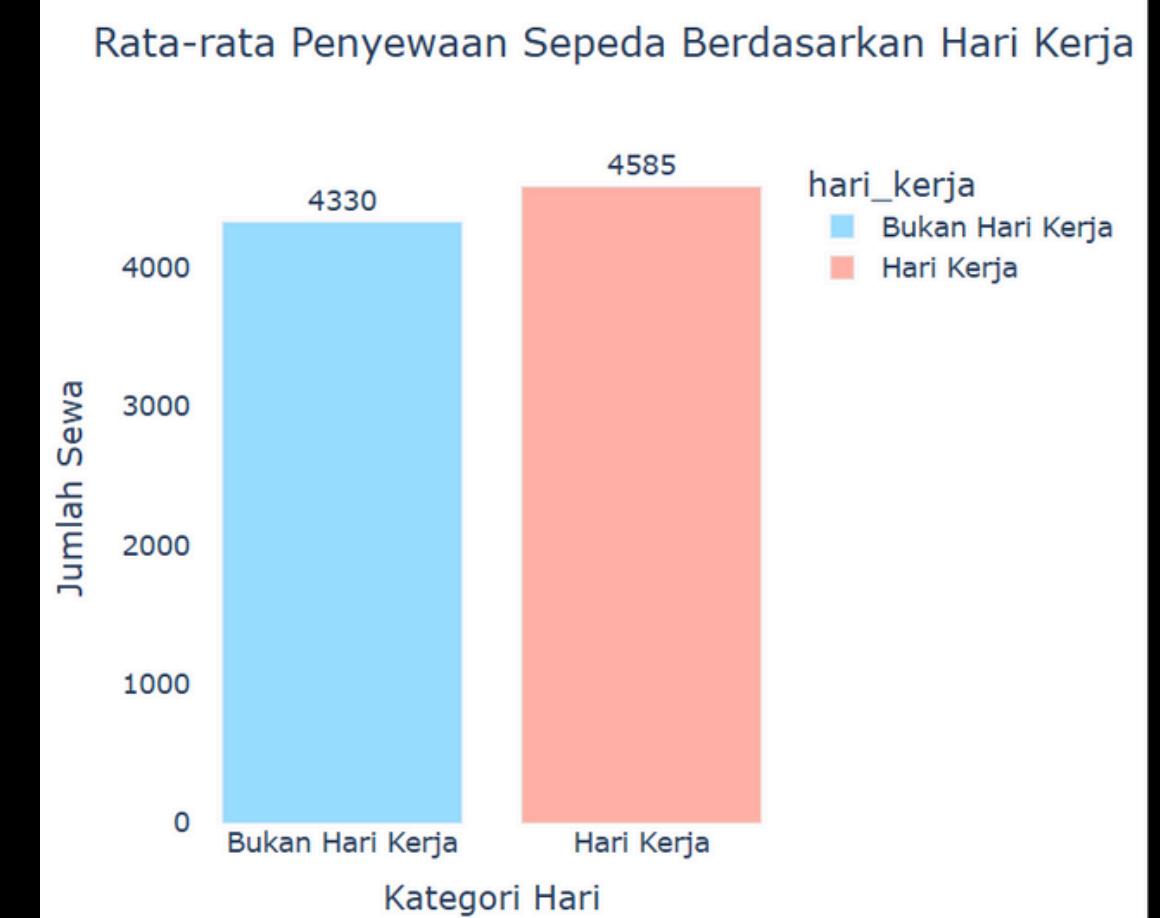
```
# Boxplot variabel cuaca
plt.figure(figsize=(10,5))
data[['suhu','kelembapan','kecepatan_angin']].boxplot()
plt.title("Boxplot Variabel Cuaca")
plt.show()
```



Boxplot menunjukkan sebaran tiga variabel cuaca. Dari grafik, kelembapan memiliki sebaran yang paling besar dan beberapa outlier, suhu berada pada rentang sedang, sedangkan kecepatan angin memiliki variasi paling kecil. Hal ini menunjukkan bahwa perubahan suhu dan kelembapan lebih memengaruhi penyewa dibandingkan angin.

```
import plotly.express as px

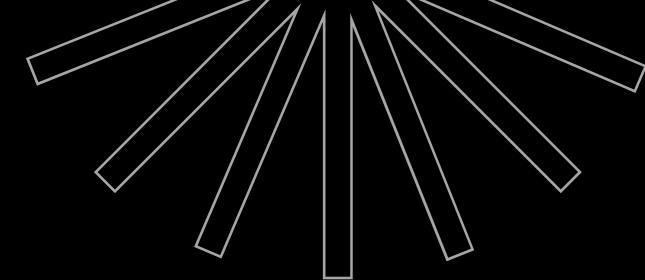
# Hitung rata-rata dulu biar grafik rapi
avg_data = data.groupby('hari_kerja')['jumlah_sewa'].mean().reset_index()
avg_data['hari_kerja'] = avg_data['hari_kerja'].map({0:'Bukan Hari Kerja', 1:'Hari Kerja'})
```



Bar chart menunjukkan rata-rata penyewa sepeda pada hari kerja lebih tinggi dibandingkan bukan hari kerja. Hal ini menandakan bahwa aktivitas kerja berpengaruh terhadap meningkatnya permintaan sepeda.



ANALISIS INFERENSIAL



ANALISIS INFERENSIAL

ANALISIS KORELASI PEARSON

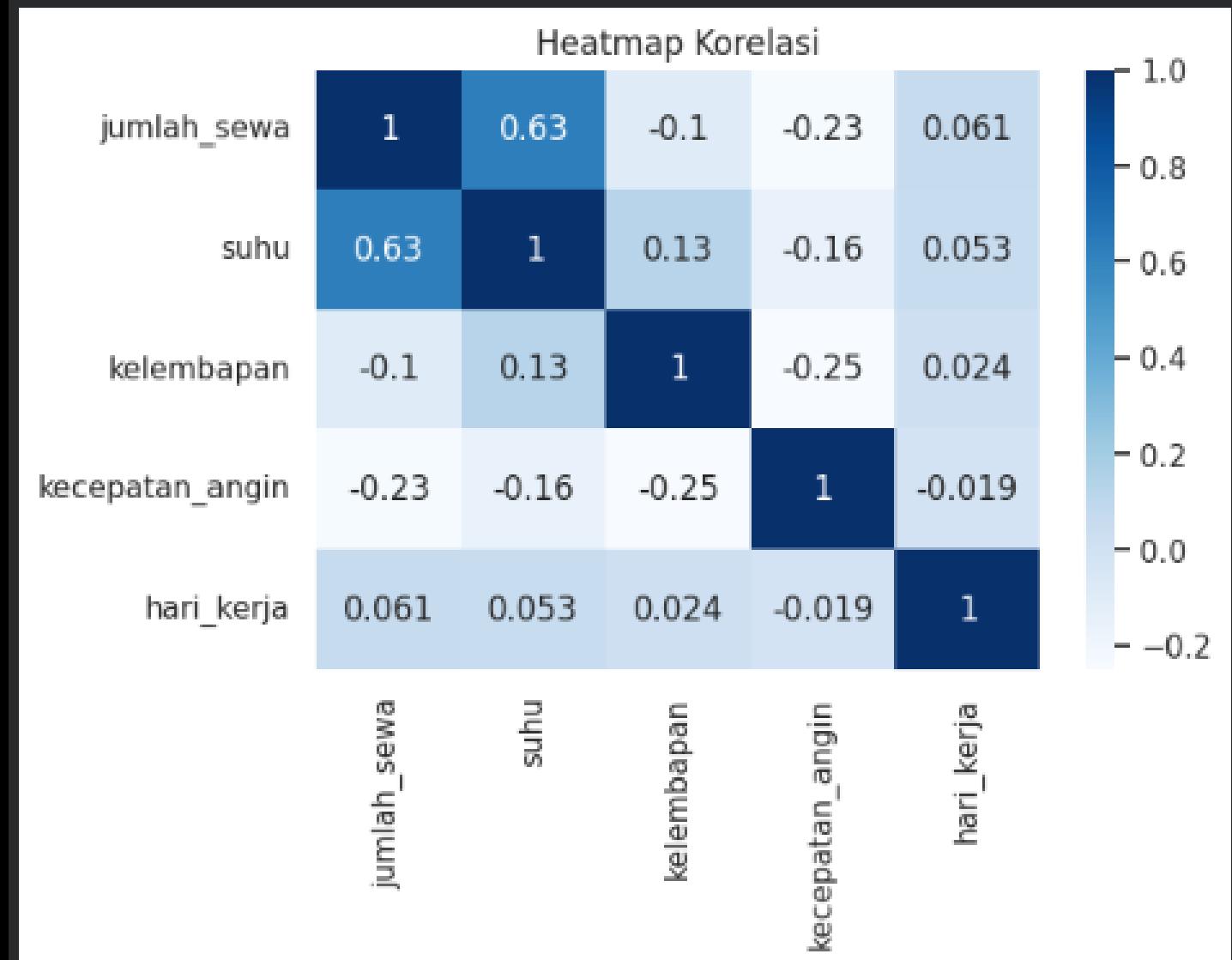
```
# Analisis Korelasi Pearson
variables = ['suhu', 'kelembapan', 'kecepatan_angin']

for var in variables:
    r, p = pearsonr(data[var], data['jumlah_sewa'])
    print(f"Korelasi {var} - jumlah_sewa: r = {r:.3f}, p = {p:.5f}")

Korelasi suhu - jumlah_sewa: r = 0.627, p = 0.00000
Korelasi kelembapan - jumlah_sewa: r = -0.101, p = 0.00645
Korelasi kecepatan_angin - jumlah_sewa: r = -0.235, p = 0.00000
```

Hasil analisis korelasi menunjukkan bahwa suhu memiliki pengaruh positif paling kuat terhadap jumlah penyewaan sepeda ($r = 0.627$), sedangkan kelembapan ($r = -0.101$) dan kecepatan angin ($r = -0.235$) berpengaruh negatif dengan kekuatan yang lebih lemah. Ketiganya signifikan secara statistik, sehingga dapat disimpulkan bahwa cuaca hangat mendorong peningkatan penyewaan sepeda, sementara kelembapan tinggi dan angin kencang cenderung menurunkannya.

```
plt.figure(figsize=(6,4))
sns.heatmap(data.corr(), annot=True, cmap="Blues")
plt.title("Heatmap Korelasi")
plt.show()
```



HEATMAP KORELASI

UJI KOEFISIEN + HIPOTESIS

Hasil regresi menunjukkan bahwa suhu memiliki koefisien positif sebesar 6609.7075 dan signifikan ($p \geq 0.000$), sehingga kenaikan suhu berkaitan dengan meningkatnya jumlah penyewaan sepeda. Kelembapan memiliki koefisien negatif – 3106.8307 dan signifikan, menunjukkan bahwa semakin lembap udara, semakin rendah jumlah penyewaan. Kecepatan angin juga berpengaruh negatif dengan koefisien –4801.7070 dan signifikan, sehingga angin kencang cenderung mengurangi penyewaan sepeda. Sementara itu, variabel hari_kerja memiliki koefisien positif 125.8049, namun p-value 0.268, sehingga pengaruhnya tidak signifikan. Nilai R-squared sebesar 0.462 berarti model mampu menjelaskan sekitar 46.2% variasi jumlah penyewaan, dan Prob(F-statistic) yang sangat kecil menunjukkan bahwa model secara keseluruhan signifikan.

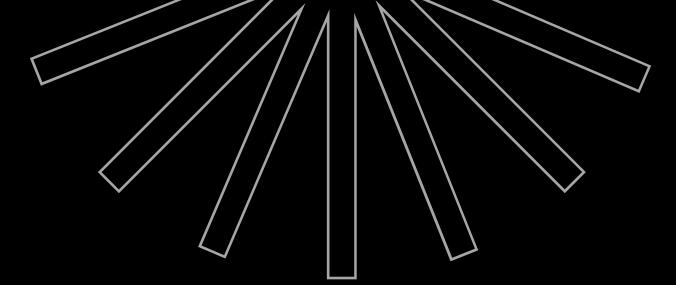
```
model = sm.OLS(y, X).fit()
print(model.summary())
```

OLS Regression Results							
Dep. Variable:	jumlah_sewa	R-squared:	0.462	Model:	OLS	Adj. R-squared:	0.459
Method:	Least Squares	F-statistic:	155.7	Date:	Tue, 02 Dec 2025	Prob (F-statistic):	3.61e-96
Time:	03:40:49	Log-Likelihood:	-6343.2	No. Observations:	731	AIC:	1.270e+04
Df Residuals:	726	BIC:	1.272e+04	Df Model:	4	Covariance Type:	nonrobust
	coef	std err	t	P> t	[0.025	0.975]	
const	4009.3688	344.524	11.637	0.000	3332.986	4685.752	
suhu	6609.7075	293.387	22.529	0.000	6033.719	7185.697	
kelembapan	-3106.8307	383.979	-8.091	0.000	-3860.673	-2352.988	
kecepatan_angin	-4801.7007	708.809	-6.774	0.000	-6193.261	-3410.141	
hari_kerja	125.8049	113.551	1.108	0.268	-97.122	348.731	

Omnibus:	10.037	Durbin-Watson:	0.404
Prob(Omnibus):	0.007	Jarque-Bera (JB):	7.868
Skew:	0.160	Prob(JB):	0.0196
Kurtosis:	2.684	Cond. No.	21.4

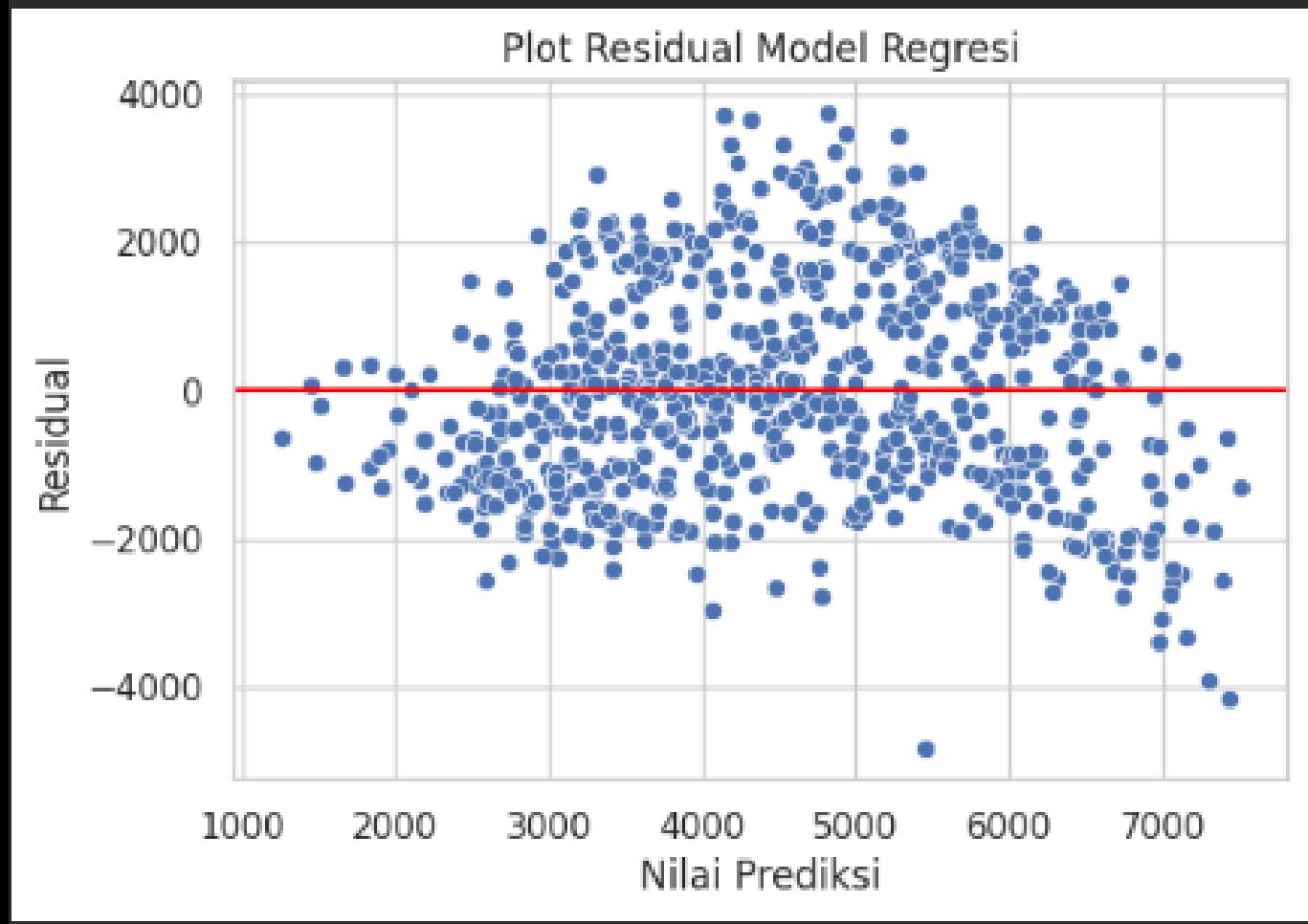
Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.



PLOT RESIDUAL MODEL

```
plt.figure(figsize=(6,4))
sns.scatterplot(x=model.fittedvalues, y=model.resid)
plt.axhline(0, color='red')
plt.title("Plot Residual Model Regresi")
plt.xlabel("Nilai Prediksi")
plt.ylabel("Residual")
plt.show()
```



Residual model tersebar acak di sekitar garis nol tanpa pola tertentu, sehingga model regresi sudah cukup layak digunakan. Namun terdapat beberapa outlier dan sedikit heteroskedastisitas karena penyebaran residu lebih lebar pada nilai prediksi tertentu, sehingga model belum sepenuhnya sempurna dalam memetakan variasi data.

INTERPRETASI KOEFISIEN DAN P-VALUE

```
coef = model.params
pvals = model.pvalues
r2 = model.rsquared

print(f"R-squared model: {r2:.3f}\n")

for var in coef.index:
    if var != "const":
        arah = "meningkatkan jumlah sewa" if coef[var] > 0 else "menurunkan jumlah sewa"
        signif = "SIGNIFIKAN" if pvals[var] < 0.05 else "tidak signifikan"
        print(f"- {var}: {arah} (coef={coef[var]:.2f}) → {signif}")
```

R-squared model: 0.462

- suhu: meningkatkan jumlah sewa (coef=6609.71) → SIGNIFIKAN
- kelembapan: menurunkan jumlah sewa (coef=-3106.83) → SIGNIFIKAN
- kecepatan_angin: menurunkan jumlah sewa (coef=-4801.70) → SIGNIFIKAN
- hari_kerja: meningkatkan jumlah sewa (coef=125.80) → tidak signifikan

Kode tersebut untuk membaca hasil regresi secara otomatis dengan menampilkan R-squared, arahan setap variabel terhadap jumlah sewa, besar koefisien serta apakah pengaruhnya signifikan atau tidak.

KESIMPULAN

Jumlah penyewaan sepeda berubah-ubah karena jumlah permintaan sepeda yang berubah-ubah setiap harinya dan tidak stabil yang terlihat dari variasi data yang tinggi dan distribusi penyewaan yang tidak merata.

Faktor cuaca menjadi penentu utama penyewaan sepeda yang dibuktikan dengan suhu berpengaruh positif, sedangkan kelembapan dan kecepatan angin berpengaruh negatif dan signifikan, artinya kondisi cuaca sangat menentukan minat masyarakat bersepeda.

Hari kerja tidak berpengaruh signifikan dalam model regresi meskipun rata-rata penyewaan lebih tinggi pada hari kerja, secara statistik variable hari kerja tidak terbukti signifikan jika dikombinasikan dengan faktor cuaca.

Model cukup baik, namun belum sempurna dengan model yang hanya mampu menjelaskan 46,2% variasi data, tetapi masih dipengaruhi oleh faktor lainnya seperti musim dan hari libur yang belum dianalisis.



KELOMPOK 8

THANK
YOU

