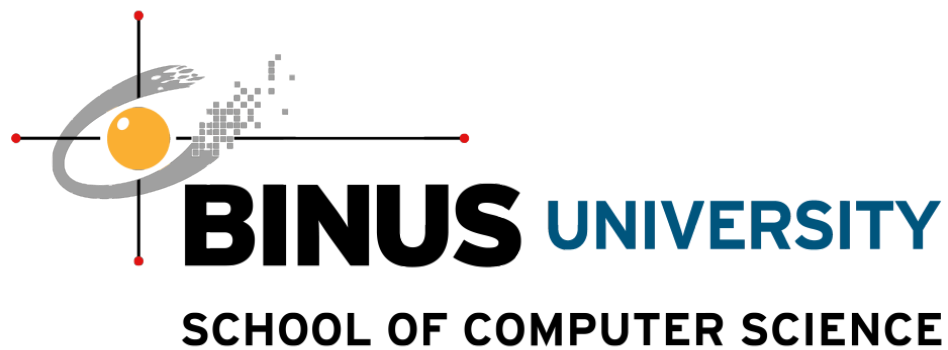


Morocco Electricity Consumption Analysis



Big Data Processing - COMP6579001

Oleh

Rasxelz Cornelius Khu	2702237316
Steven Andrew Dharma Wira	2702255792
Varen Permata Sari	2702240084
Raditya Pastika Heriyanto	2702219950
Brandon Feivel	2702246131
Laurance Sebastian Gunawan	2702253843

BAB I

Introduction

1.1 Latar Belakang

Konsumsi listrik merupakan salah satu indikator utama dalam menganalisis kebutuhan energi di suatu wilayah. Seiring dengan meningkatnya permintaan energi listrik, pemahaman terhadap pola konsumsi menjadi sangat penting untuk mengoptimalkan distribusi energi, pengelolaan sumber daya, serta perencanaan infrastruktur yang berkelanjutan.

Maroko, sebagai negara berkembang yang mengandalkan kombinasi sumber energi konvensional dan terbarukan, menghadapi tantangan dalam memenuhi permintaan listrik yang terus meningkat. Oleh karena itu, analisis konsumsi listrik secara menyeluruh diperlukan untuk mendukung pengambilan keputusan yang tepat dalam sektor energi.

Proyek ini berfokus pada analisis konsumsi listrik di Maroko selama tahun 2017, menggunakan dataset dari Maven Analytics. Dataset tersebut mencakup data konsumsi listrik yang dicatat setiap 10 menit di tiga zona berbeda, serta variabel lingkungan seperti suhu, kelembaban, dan kecepatan angin. Studi ini bertujuan untuk mengidentifikasi pola musiman, mengevaluasi pengaruh faktor lingkungan, serta menganalisis perbedaan konsumsi listrik antar zona.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, penelitian ini dirancang untuk menjawab pertanyaan-pertanyaan utama berikut:

1. Bagaimana pola konsumsi listrik di Maroko pada tingkat harian, mingguan, dan tahunan?
2. Sejauh mana faktor lingkungan seperti suhu, kelembaban, dan kecepatan angin mempengaruhi konsumsi listrik?
3. Apakah terdapat perbedaan signifikan dalam konsumsi listrik antara ketiga zona yang dianalisis?

1.3 Tujuan Projek

1. Analisis Pola Konsumsi Berdasarkan Zona

Melakukan perbandingan pola konsumsi listrik di tiga zona berbeda untuk menentukan apakah terdapat perbedaan signifikan dalam perilaku penggunaan energi. Analisis ini akan membantu mengidentifikasi zona dengan kebutuhan energi yang unik atau menemukan pola konsumsi yang seragam di seluruh wilayah.

2. Analisis Korelasi dengan Faktor Lingkungan

Menyelidiki hubungan antara konsumsi listrik dan variabel lingkungan seperti suhu, kelembaban, dan kecepatan angin. Tujuannya adalah untuk menentukan faktor lingkungan mana yang paling mempengaruhi penggunaan listrik serta bagaimana perubahan kondisi lingkungan mempengaruhi konsumsi energi.

3. Analisis Pola Musiman

- Tingkat Per Jam: Mengidentifikasi fluktuasi konsumsi listrik sepanjang hari untuk menemukan waktu puncak penggunaan dan periode dengan permintaan rendah.
- Tingkat Mingguan: Menganalisis tren konsumsi mingguan untuk mendeteksi pola berulang atau anomali yang mungkin terkait dengan aktivitas rutin mingguan.
- Tingkat Tahunan: Mengevaluasi pola tahunan untuk memahami bagaimana perubahan musim mempengaruhi konsumsi energi secara keseluruhan.

BAB II

Metodologi dan Alur Pekerjaan

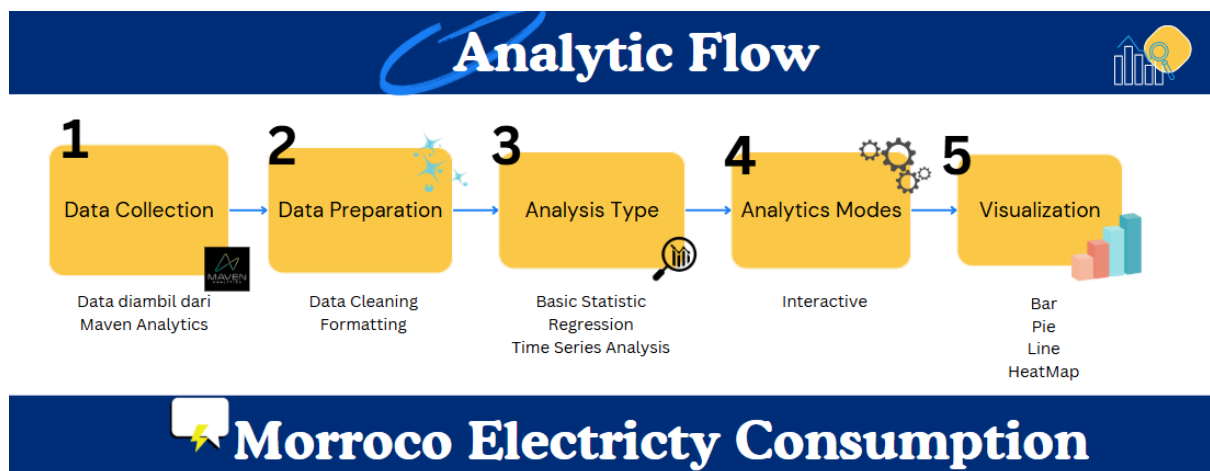
2.1 Dataset

Dataset diambil dari Maven Analytics dengan sumber UCI Machine Learning Repository. Dataset berbentuk .csv dan memiliki 61.755 baris data dengan 9 atribut yaitu:

- Datetime: Tanggal dan waktu pencatatan (tipe *datetime*)
- Temperature: Suhu lingkungan (*int*)
- Humidity: Kelembaban lingkungan (*int*)
- WindSpeed: Kecepatan angin (*int*)
- GeneralDiffuseFlows: Aliran difus umum (*int*)
- PowerConsumption_Zone1: Konsumsi listrik total (*int*)
- PowerConsumption_Zone2: Konsumsi listrik di Zona 1 (*int*)
- PowerConsumption_Zone3: Konsumsi listrik di Zona 2 (*int*)

2.2 Analysis Flow

Environment yang digunakan adalah Google Colab, dengan bahasa pemrograman Python. Berikut adalah analytic flow dari proyek ini:



Data yang diambil dari Maven Analytics, akan melalui proses data cleaning dengan menghapus missing value dan formatting untuk memastikan data bertipe date dalam format yang sesuai. Tipe analisis yang digunakan ada 3 yaitu basic statistic, regression, dan time series analysis. Basic statistic mencakup max, min, mean, correlation, dan sum yang akan divisualisasikan dengan box plot untuk setiap zona sebagai analisis deskriptif mendasar. Melanjutkan analisis korelasi, simple linear regression dan metode pearson digunakan untuk menganalisis hubungan antara faktor lingkungan (temperatur, angin, dan kelembaban) terhadap penggunaan listrik. Korelasi akan divisualisasikan dengan heatmap dan scatter plot. Time series juga digunakan untuk menganalisa pola musiman dalam penggunaan listrik. Berikut penjelasan proses dan code yang dilakukan.

1. Import Library

```
[ ] import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

- pandas dan numpy: untuk manipulasi, transformasi, dan analisis data numerik serta statistik dasar.
- matplotlib dan seaborn: untuk membuat grafik visual seperti line plot, box plot, scatter plot, dan heatmap.

2. Data Loading & Preprocessing

```
[ ] from google.colab import drive
drive.mount('/content/drive')

df = pd.read_csv('/content/drive/MyDrive/powerconsumption.csv')
df['Datetime'] = pd.to_datetime(df['Datetime'])
df.set_index('Datetime', inplace=True)

df.head()
```

Dataset disimpan pada Google Drive pribadi dan dihubungkan dengan Google Colab untuk menghindari proses unggah ulang setiap kali proyek dijalankan. Kemudian data preprocessing dilakukan dengan tahapan meliputi:

- Konversi Kolom Datetime: Kolom Datetime dikonversi ke dalam format waktu (datetime object) agar dapat digunakan dalam analisis deret waktu.
- Pembersihan Data: Menghapus missing values untuk memastikan integritas data.
- Outlier Handling: Outlier tetap dipertahankan, karena dalam konteks konsumsi listrik, nilai ekstrim bisa mencerminkan puncak pemakaian aktual, bukan kesalahan data (noise).

Setelah tahap ini, data siap untuk dianalisis lebih lanjut. Potongan sampel data ditampilkan sebagai verifikasi keberhasilan preprocessing.

	Temperature	Humidity	WindSpeed	GeneralDiffuseFlows	DiffuseFlows	PowerConsumption_Zone1	PowerConsumption_Zone2	PowerConsumption_Zone3
Datetime								
2017-01-01 00:00:00	6.559	73.8	0.083	0.051	0.119	34055.69620	16128.87538	20240.96386
2017-01-01 00:10:00	6.414	74.5	0.083	0.070	0.085	29814.68354	19375.07599	20131.08434
2017-01-01 00:20:00	6.313	74.5	0.080	0.062	0.100	29128.10127	19006.68693	19668.43373
2017-01-01 00:30:00	6.121	75.0	0.083	0.091	0.096	28228.86076	18361.09422	18899.27711
2017-01-01 00:40:00	5.921	75.7	0.081	0.048	0.085	27335.69620	17872.34043	18442.40964

```
print(df.isnull().sum())
df.dropna(inplace=True)

Temperature      0
Humidity         0
WindSpeed       0
GeneralDiffuseFlows  0
DiffuseFlows    0
PowerConsumption_Zone1  0
PowerConsumption_Zone2  0
PowerConsumption_Zone3  0
Hour            0
DayOfWeek       0
Month           0
dtype: int64
```

3. Basic Statistic Analysis

```
[ ] zones = ['PowerConsumption_Zone1', 'PowerConsumption_Zone2', 'PowerConsumption_Zone3']
for zone in zones:
    print(f"\nStatistik untuk {zone}:")
    print(f" Total (Sum)      : {df[zone].sum():,.2f}")
    print(f" Rata-rata (Mean): {df[zone].mean():,.2f}")
    print(f" Median      : {df[zone].median():,.2f}")
    print(f" Modus (Mode)  : {df[zone].mode().iloc[0]:,.2f}")
    print(f" Minimum      : {df[zone].min():,.2f}")
    print(f" Maksimum     : {df[zone].max():,.2f}")
    print(f" Std Dev      : {df[zone].std():,.2f}")
```

sebagai analisis deskriptif mendasar, setiap zona akan dihitung total, rata-rata, median, min, max penggunaan listrik. Berikut hasilnya, agar data lebih menarik, digunakan visualisasi boxplot dan pie chart.

Zone	Sum	Mean	Median	Std	min	max	mode
1	1.695394 e+09	32344.97 0564	32265.92 0340	7130.562 564	13895.69 6200	52204.39 512	34560.0
2	1.102964 e+09	21042.50 9082	20823.16 8405	5201.465 892	8560.081 466	37408.86 076	21600.0
3	9.348607 e+08	17835.40 6218	16415.11 7470	6622.165 099	5935.174 070	47598.32 636	17280.0

```
[ ] plt.figure(figsize=(10,6))
sns.boxplot(data=df[['PowerConsumption_Zone1', 'PowerConsumption_Zone2', 'PowerConsumption_Zone3']])
plt.title('Distribusi Konsumsi Listrik per Zona')
plt.ylabel('Konsumsi Listrik')
plt.show()
```

```
[ ] zones = ['Zone 1', 'Zone 2', 'Zone 3']
totals = [1695393977.06, 1102964156.06, 934860652.34]

plt.figure(figsize=(7, 7))
plt.pie(totals, labels=zones, autopct='%1.1f%%', startangle=140, colors=['steelblue', 'darkorange', 'seagreen'])
plt.title('Persentase Total Konsumsi Listrik per Zona')
plt.axis('equal')
plt.show()
```

4. Analisis Korelasi Faktor Lingkungan

```
▶ korelasi = df[['Temperature', 'Humidity', 'WindSpeed',  
               'PowerConsumption_Zone1', 'PowerConsumption_Zone2', 'PowerConsumption_Zone3']].corr(method='pearson')
```

Dalam dataset disediakan 3 faktor lingkungan yaitu temperatur, kecepatan angin, dan kelembaban. Untuk mengetahui faktor apa yang paling mempengaruhi penggunaan listrik, maka digunakan perhitungan pearson. Teknik pearson adalah teknik untuk menghitung korelasi antar 2 variabel, dengan hasil koefisien.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \cdot \sqrt{\sum (y_i - \bar{y})^2}}$$

Jika koefisien mendekati -1 maka artinya tidak ada korelasi, jika mendekati 1 maka artinya ada korelasi kuat antar variabel. Korelasi ditampilkan dalam heatmap menggunakan skala warna, dimana warna merah menunjukkan korelasi negatif, biru menunjukkan korelasi positif, dan warna netral menunjukkan korelasi rendah atau tidak ada korelasi.

```
▶ plt.figure(figsize=(10, 6))  
sns.heatmap(korelasi, annot=True, cmap='coolwarm', fmt=".2f")  
plt.title('Heatmap Korelasi Faktor Lingkungan dan Konsumsi Listrik')  
plt.show()
```

Hasil heatmap menunjukkan bahwa temperatur adalah faktor yang paling berpengaruh. maka analisis akan dilanjutkan untuk melihat hubungan keduanya menggunakan regresi linear. Model ini mengasumsikan bahwa hubungan antara kedua variabel bersifat linier. Model regresi linier digunakan untuk membentuk garis trend, yang divisualisasikan pada scatter plot dengan garis regresi sebagai indikasi hubungan linier antara variabel.

```
▶ plt.figure(figsize=(8, 5))  
scatter = plt.scatter(df['Temperature'], df['PowerConsumption_Zone1'],  
                     c=df['Temperature'], cmap='viridis', s=10)  
plt.colorbar(label='Suhu (°C)')  
  
z = np.polyfit(df['Temperature'], df['PowerConsumption_Zone1'], 1)  
p = np.poly1d(z)  
plt.plot(df['Temperature'], p(df['Temperature']), color='red', linewidth=2, label='Regresi Linier')  
plt.title('Hubungan Suhu dan Konsumsi Listrik (Zone 1)')  
plt.xlabel('Suhu (°C)')  
plt.ylabel('Konsumsi Listrik (Watt)')  
plt.grid(True)  
plt.legend()  
plt.tight_layout()  
plt.show()
```

5. Time Series Analysis

```
[ ] df['Hour'] = df.index.hour
    df['DayOfWeek'] = df.index.day_name()
    df['Month'] = df.index.month_name()
```

Untuk melakukan time series analysis, diperlukan kolom atribut baru yang mewakili satuan waktu. Time series akan dianalisis berdasarkan perhari, perminggu, dan perbulan untuk melihat pola musiman di masing-masing zona. Hasil akan divisualisasikan dengan line plot untuk memperjelas pola pemakaian listrik.

```
[ ] hourly = df.groupby('Hour')[['PowerConsumption_Zone1', 'PowerConsumption_Zone2', 'PowerConsumption_Zone3']].mean()

plt.figure(figsize=(15,6))
sns.lineplot(data=hourly)
plt.title('Rata-rata Konsumsi Listrik per Jam')
plt.xlabel('Jam (0-23)')
plt.ylabel('Konsumsi Listrik (Watt atau MW)')
plt.grid(True)
plt.xticks(range(24))
plt.show()
```

```
[ ] weekday_order = ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday']
weekly = df.groupby('DayOfWeek')[['PowerConsumption_Zone1', 'PowerConsumption_Zone2', 'PowerConsumption_Zone3']].mean().reindex(weekday_order)

plt.figure(figsize=(10,6))
sns.lineplot(data=weekly)
plt.title('Rata-rata Konsumsi Listrik per Hari')
plt.xlabel('Hari')
plt.ylabel('Konsumsi Listrik')
plt.grid(True)
plt.show()
```

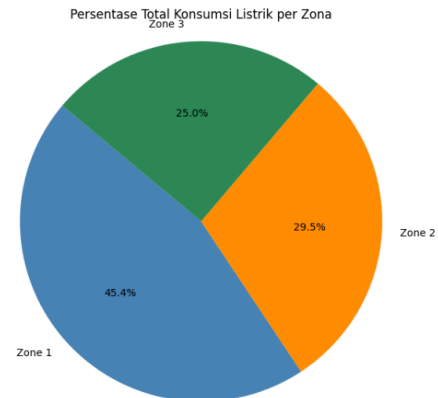
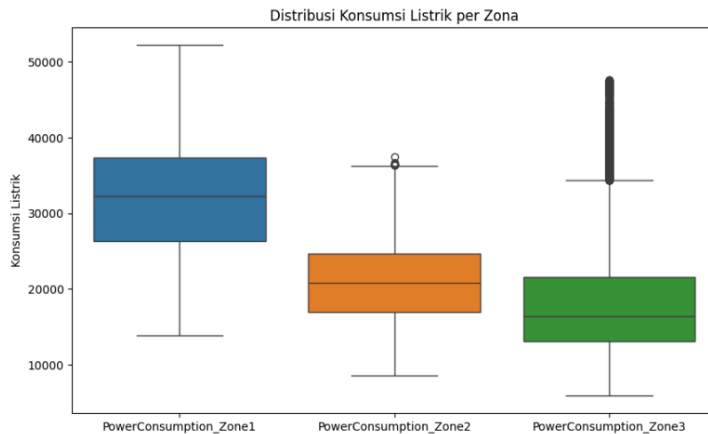
```
[ ] monthly_order = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December']
monthly = df.groupby('Month')[['PowerConsumption_Zone1', 'PowerConsumption_Zone2', 'PowerConsumption_Zone3']].mean().reindex(monthly_order)

plt.figure(figsize=(12,6))
sns.lineplot(data=monthly)
plt.title('Rata-rata Konsumsi Listrik per Bulan')
plt.xlabel('Bulan')
plt.ylabel('Konsumsi Listrik')
plt.grid(True)
plt.xticks(rotation=45)
plt.show()
```


BAB III

Diskusi dan Hasil Analisis

Berdasarkan hasil analisis data yang telah dilakukan, berikut ini adalah pemaparan dan interpretasi hasil dalam beberapa aspek utama:



3.1 Basic Statistic

3.1.1 Zona 1 (Biru) :

- Merupakan zona dengan konsumsi listrik tertinggi dibandingkan dua zona lainnya.
- Nilai median konsumsi berada pada ~32.265 watt, yang sejalan dengan rata-rata.
- Data banyak terkonsentrasi di sekitar nilai tengah (interquartile range yang relatif sempit).
- Jumlah outlier relatif sedikit, menunjukkan pola konsumsi yang stabil.
- Kesimpulan: Zona 1 menunjukkan karakteristik konsumsi listrik yang stabil dan tinggi secara konsisten. Hal ini mengindikasikan bahwa area ini kemungkinan merupakan kawasan industri atau komersial.

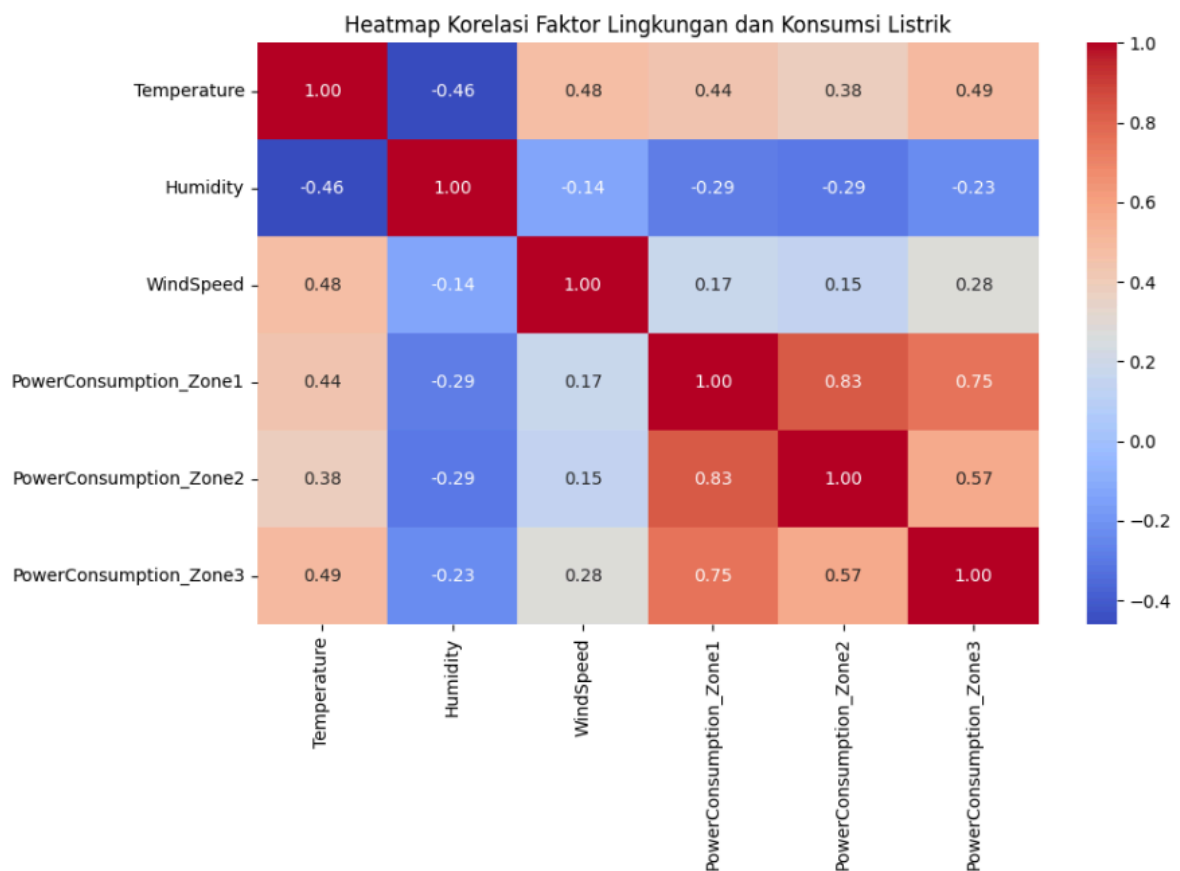
3.1.2 Zone 2 (Orange):

- Tingkat konsumsi sedang, lebih rendah dari Zona 1 namun masih lebih tinggi dibandingkan Zona 3.
- Median konsumsi sekitar ~20.823 watt.
- Ditemukan beberapa outlier dengan nilai tinggi, tetapi tidak terlalu ekstrem.
- Kesimpulan: Pola konsumsi cukup stabil dengan fluktuasi ringan. Zona ini kemungkinan merupakan kombinasi area semi-komersial dan permukiman.

3.1.3 Zone 3 (Hijau):

- Memiliki jumlah outlier paling banyak, terutama di sisi atas, menandakan adanya lonjakan konsumsi dalam waktu tertentu.
- Median konsumsi adalah yang terendah, sekitar ~16.415 watt.
- Whiskers pada boxplot lebih panjang, menandakan variasi konsumsi yang tinggi.
- Kesimpulan: Zona 3 menunjukkan karakteristik konsumsi yang tidak stabil dengan banyak fluktuasi, kemungkinan besar merupakan kawasan residensial atau daerah yang aktivitas listriknya berubah-ubah.

3.2 Analisis Korelasi



3.2.1 Pengaruh Temperatur

Zona	Korelasi	Interpretasi
Zone 1	0.44	Korelasi positif moderat. Kenaikan suhu diikuti oleh kenaikan konsumsi listrik, kemungkinan disebabkan oleh penggunaan pendingin ruangan.
Zone 2	0.38	Korelasi positif lemah hingga moderat. Pengaruh suhu tetap ada, namun tidak sekuat Zona 1.
Zone 3	0.49	Korelasi positif moderat. Temperatur merupakan faktor yang paling berpengaruh terhadap konsumsi listrik di zona ini.

Kesimpulan: Semakin tinggi suhu udara, konsumsi listrik cenderung meningkat di seluruh zona, dengan pengaruh paling kuat terjadi di Zona 3.

3.2.2 Pengaruh Kelembaban

Zona	Korelasi	Interpretasi
Zone 1	-0.29	Lemah negatif
Zone 2	-0.29	Lemah negatif
Zone 3	-0.23	Lemah negatif

Kesimpulan: Kelembaban memiliki pengaruh negatif yang lemah terhadap konsumsi listrik. Semakin lembab udara, konsumsi cenderung sedikit menurun. Hal ini bisa jadi karena udara yang lembab terasa lebih sejuk, sehingga kebutuhan akan pendingin berkurang.

3.2.3 Pengaruh Kecepatan Angin

Zona	Korelasi	Interpretasi
Zone 1	0.17	Sangat lemah positif
Zone 2	0.15	Sangat lemah positif
Zone 3	0.28	Lemah positif

Kesimpulan: Kecepatan angin memiliki pengaruh yang sangat terbatas terhadap konsumsi listrik, meskipun di Zona 3 terlihat sedikit lebih signifikan.

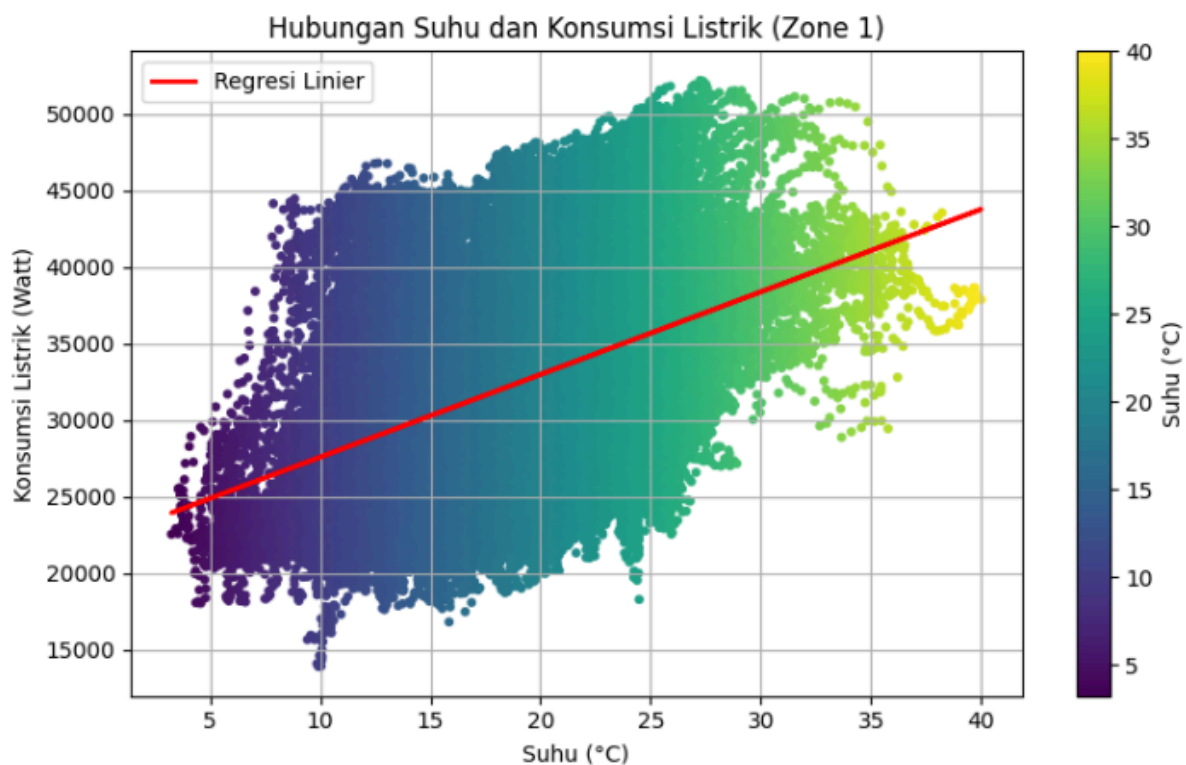
3.2.4 Korelasi Antar Zona

Zona	Korelasi	Interpretasi
Zone 1 vs 2	0.83	Korelasi sangat tinggi, menunjukkan pola konsumsi yang serupa.
Zone 1 vs 3	0.75	Berkorelasi tinggi.
Zone 2 vs 3	0.57	Korelasi sedang hingga kuat.

Kesimpulan Umum:

- Temperatur adalah faktor lingkungan dengan pengaruh paling kuat dan konsisten terhadap konsumsi listrik.
- Kelembaban memiliki pengaruh negatif yang lemah.
- Kecepatan angin memberikan pengaruh kecil, terutama di Zona 3.
- Konsumsi listrik antar zona menunjukkan korelasi yang tinggi, mengindikasikan bahwa faktor-faktor eksternal seperti musim atau hari besar nasional mempengaruhi pola konsumsi secara serentak di seluruh wilayah.

3.2.5 Analisis Regresi Hubungan Temperatur Terhadap Listrik



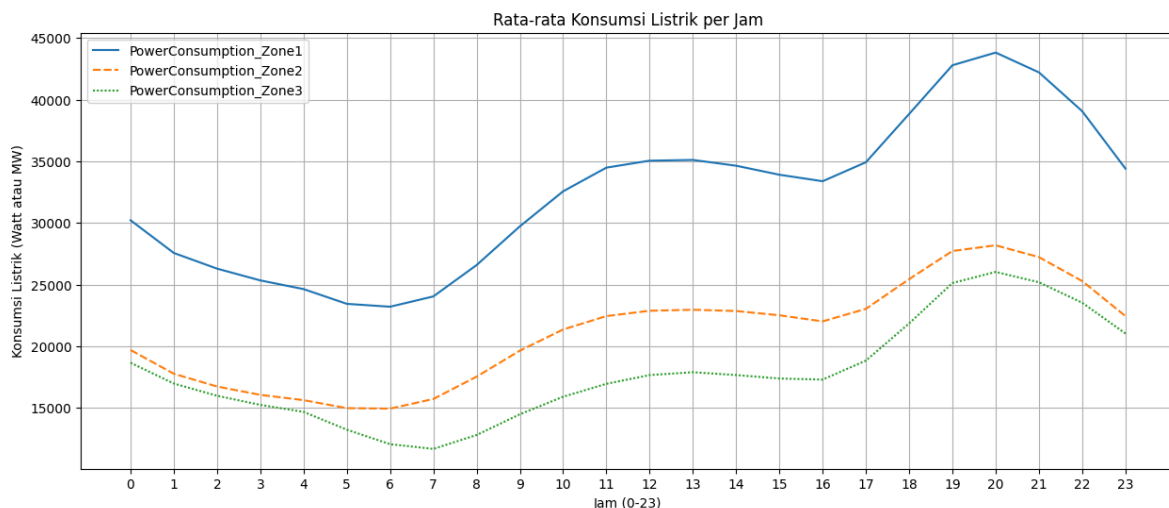
Visualisasi scatter plot antara suhu dan konsumsi listrik menunjukkan:

- Pola trend naik: semakin tinggi suhu (sumbu X), semakin tinggi konsumsi listrik (sumbu Y).
- Sebaran titik menunjukkan konsentrasi data terbesar pada rentang suhu 15°C hingga 30°C.
- Pada rentang tersebut, konsumsi listrik menunjukkan kecenderungan meningkat, terutama pada suhu tinggi.

Kesimpulan: Terdapat hubungan positif moderat antara suhu dan konsumsi listrik, khususnya di Zona 1. Hal ini memperkuat dugaan bahwa penggunaan perangkat pendingin (seperti AC) meningkat seiring dengan naiknya suhu.

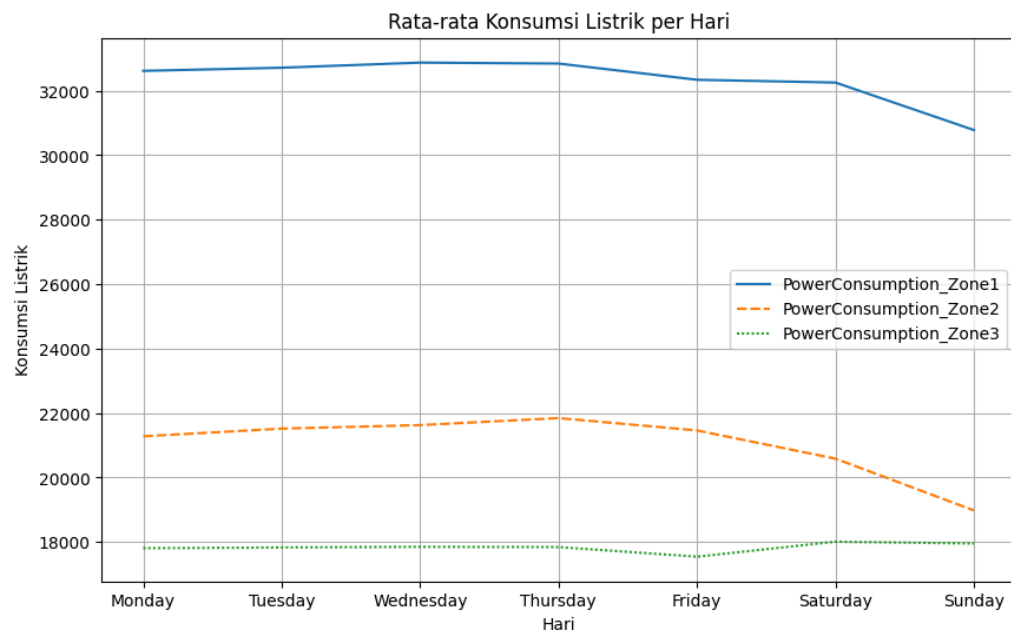
3.3 Analisis Pola Musiman

3.3.1 Pola Harian



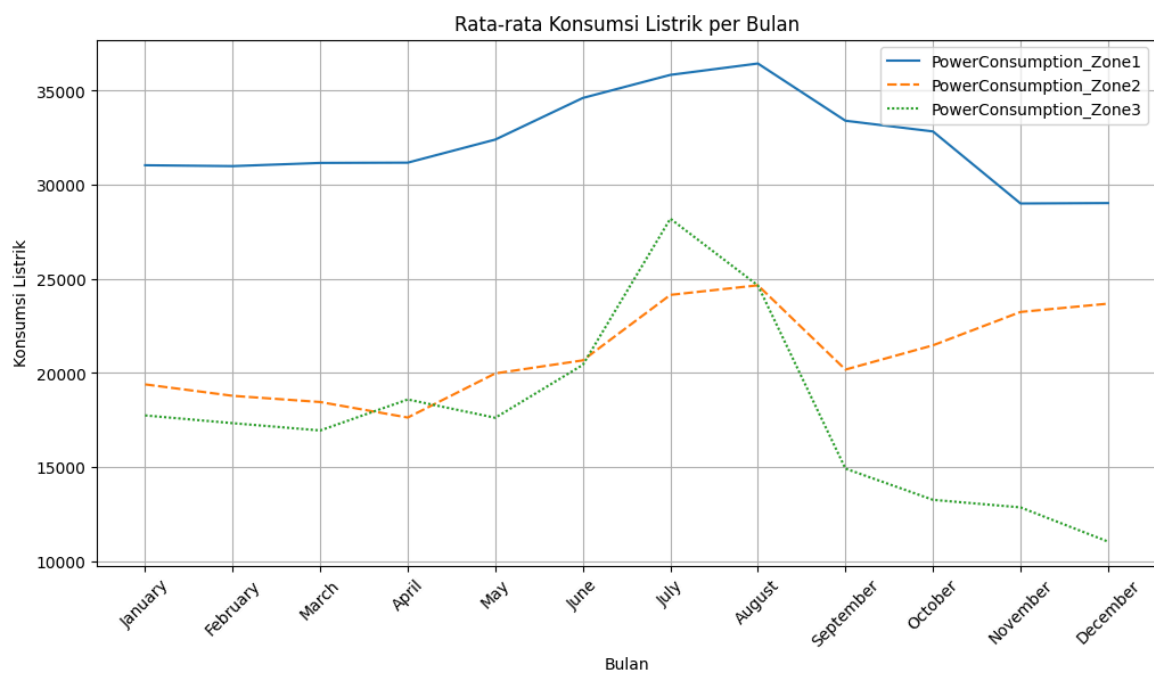
- Terdapat dua puncak konsumsi listrik di semua zona:
 - Puncak pertama sekitar pukul 13.00 (siang hari)
 - Puncak kedua, yang lebih signifikan, terjadi sekitar pukul 20.00 (malam hari)
- Konsumsi terendah tercatat pada rentang waktu 00:00–06:00 (larut malam hingga dini hari)
- Zona 1 secara konsisten menunjukkan konsumsi tertinggi sepanjang hari.

3.3.2 Pola Mingguan



- Konsumsi listrik cenderung stabil sepanjang minggu, dengan sedikit fluktuasi.
- Terdapat penurunan konsumsi yang nyata pada hari Minggu, terutama di Zona 1 dan 2.
- Zona 3 menunjukkan pola yang relatif stabil sepanjang minggu.

3.3.3 Pola Bulanan



- Puncak Musim Panas:
 - Terjadi pada bulan Agustus untuk Zona 1 dan 2.
 - Terjadi pada bulan Juli untuk Zona 3.
- Penurunan Konsumsi:
 - Terjadi pada bulan April dan Mei, menunjukkan penurunan konsumsi saat musim semi.
- Musim Dingin:
 - Zona 2 menunjukkan puncak kecil pada bulan November, menandakan kenaikan konsumsi pada musim dingin.
 - Sebaliknya, Zona 1 dan 3 menunjukkan konsumsi terendah pada periode musim dingin.

Kesimpulan:

Zona 1 menunjukkan konsumsi tertinggi dalam seluruh periode harian, mingguan, dan musiman. Pola konsumsi mengikuti aktivitas manusia (jam kerja dan malam hari), serta musim (musim panas dan hari besar nasional).

LINK GOOGLE DRIVE:

<https://drive.google.com/drive/folders/1oRDwUmqFC--8m4rVuNcEGNuso78GqCVa?usp=sharing>

LINK GOOGLE COLAB:

https://colab.research.google.com/drive/1Nf9_4CRpYm85e04EdyCQNRXX_OUZJVsB?usp=sharing