# Morocco Electricity Consumption Analysis



DATA ANALYTICS - COMP6886001 - LC01

ASSESSMENT OF LEARNING

Disusun Oleh:

JOSH NICHOLAS SUTANTO 2702234825

FERLIE HERNATA 2702231262

GIOVINCENT RICEL'S TANOTO 2702226786

NIKOLAUS MARVIN LIAYASA 2702233702

RENDY RIADY 2702234421

RASXELZ CORNELIUS KHU 2702237316

Bina Nusantara University - Computer Science

2025

# BAB I

**Introduction**

## 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.

## Rumusan Masalah

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

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

## Tujuan Projek

* 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.

* 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.

* 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

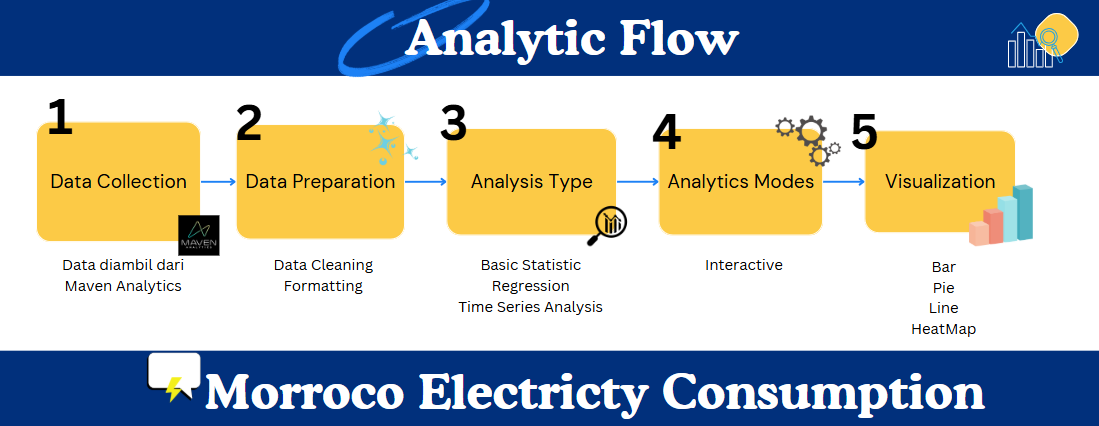
**Metodologi dan Alur Pekerjaan**

## Dataset

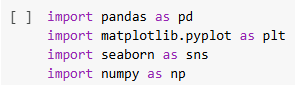
Dataset diambil dari Maven Analytics dengan sumber UCI Machine Learning Repository.Dataset berbentuk .csv dan memiliki 61.755 baris data dengan 9 atribut yaitu:

* Datetime: Tanggal dan waktu pencatatan (*datetime*)
* Temperature: Suhu lingkungan (*int*)
* Humidity: Kelembaban lingkungan (*int*)
* WindSpeed: Kecepatan angin (*int*)
* GeneralDiffuseFlows: Aliran difus umum (*int*)
* PowerConsumption\_Zone1: Konsumsi listrik di Zona 1 (*int*)
* PowerConsumption\_Zone2: Konsumsi listrik di Zona 2 (*int*)
* PowerConsumption\_Zone3: Konsumsi listrik di Zona 3 (*int*)

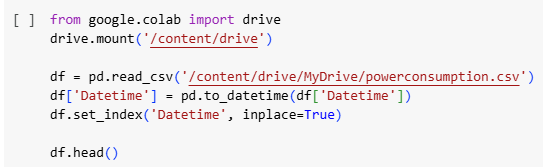
## Analysis Flow

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.

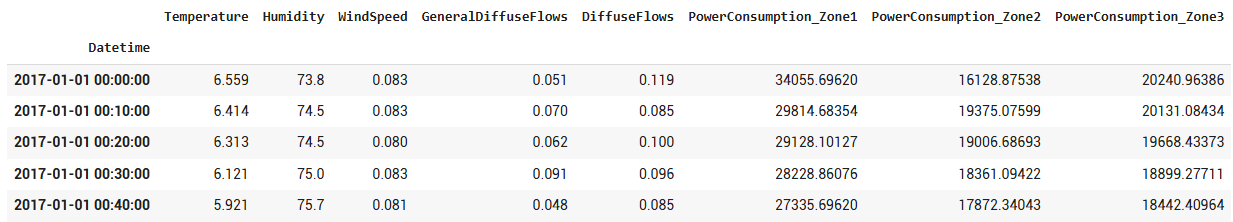
1. Import Library

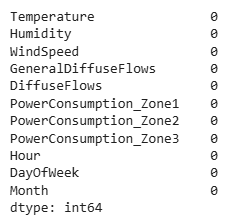
* pandas dan numpy: untuk manipulasi, transformasi, dan analisis data numerik serta statistik dasar.
* matplotlib dan seaborn: untuk membuat grafik visual seperti line plot, box plot, scatter plot, dan heatmap.

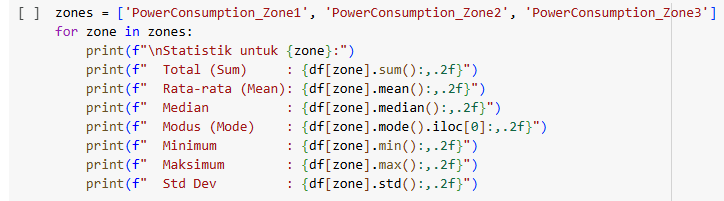
1. Data Loading & Preprocessing

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:

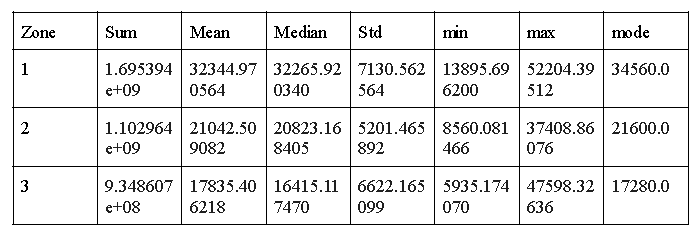
* Konversi Kolom Datetime: Kolom Datetime dikonversi ke dalam format waktu (datetime object) agar dapat digunakan dalam analisis deret waktu.
* Pembersihan Data: Menghapus missing values untuk memastikan integritas data.
* Outlier Handling: Outlier tetap dipertahankan, karena dalam konteks konsumsi listrik,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

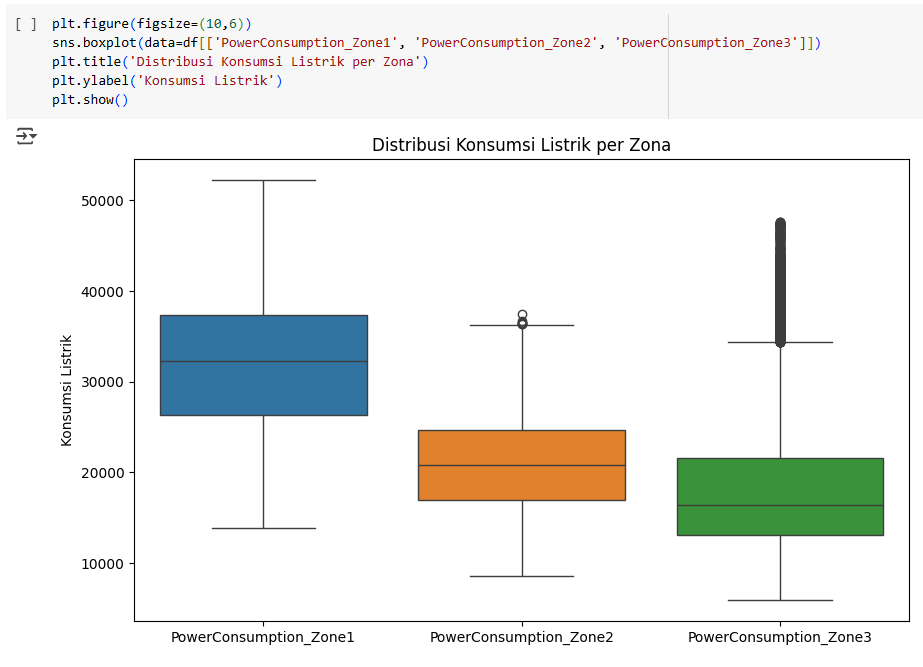
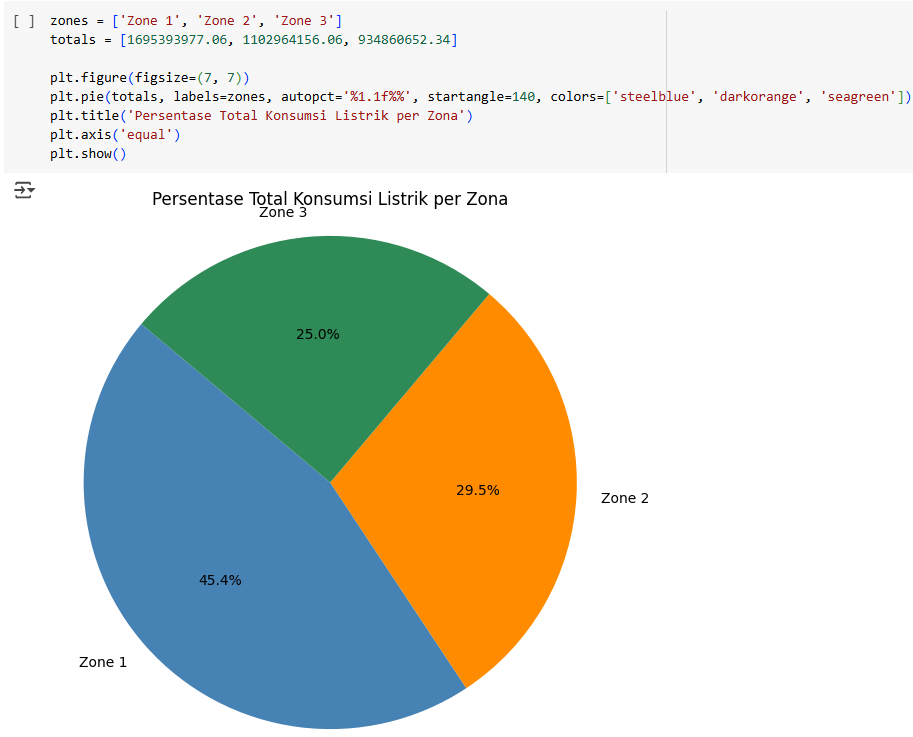
Setelah tahap ini, data siap untuk dianalisis lebih lanjut. Potongan sampel data ditampilkan sebagai verifikasi keberhasilan preprocessing.



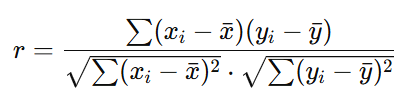
1. Basic Statistic Analysis

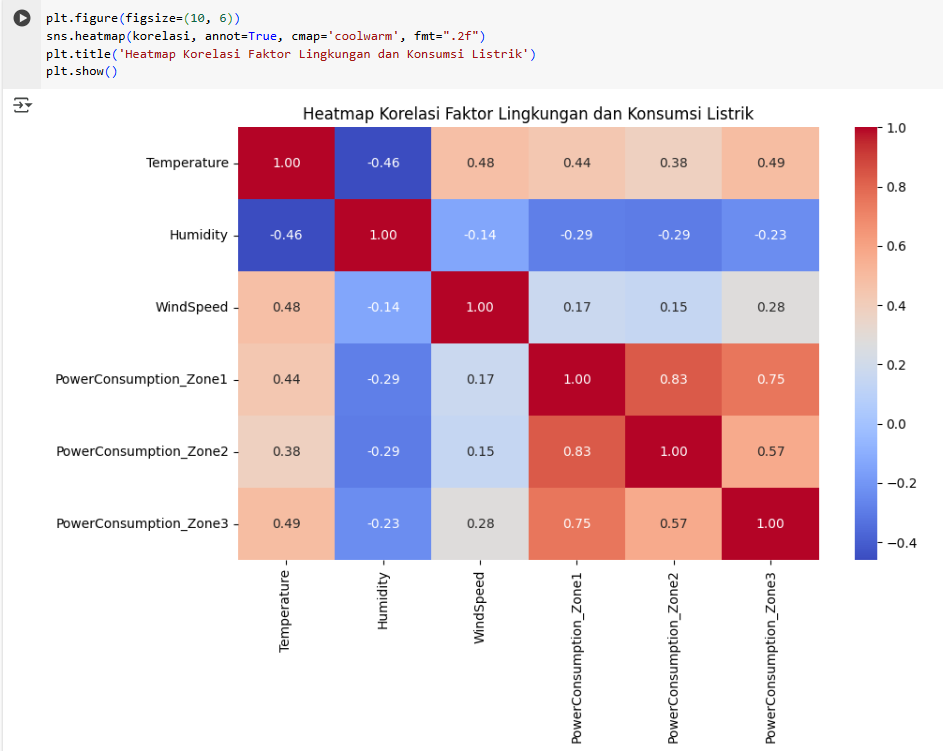
Sebagai analisis deskriptif mendasar, setiap zona akan dihitung total, rata-rata, median, min, max penggunaan listrik. Berikut hasilnya, agar data lebih menarik, digunakan visualisasi boxplot dan pie chart.

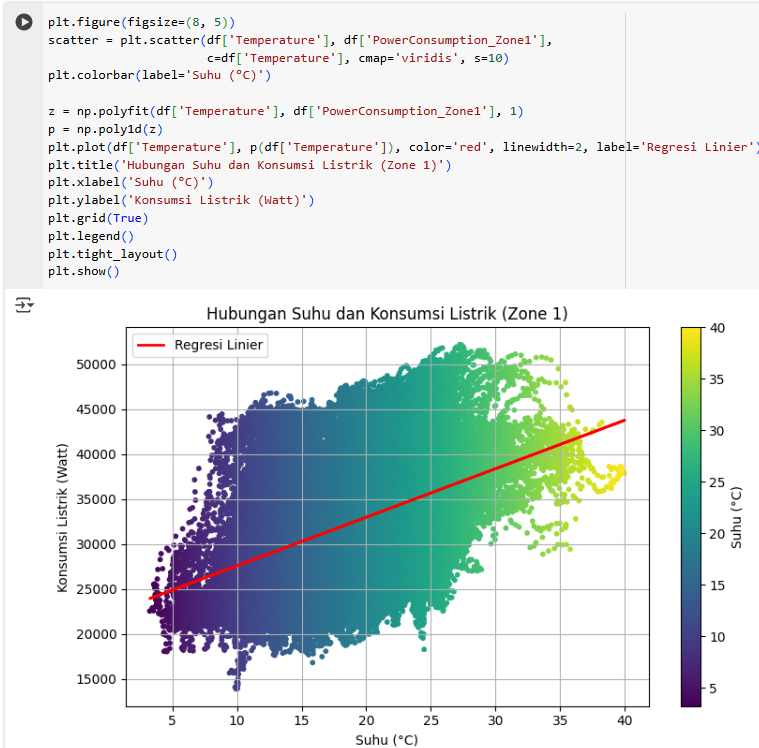




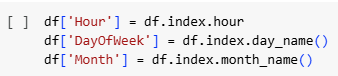
1. Analisis Korelasi Faktor Lingkungan

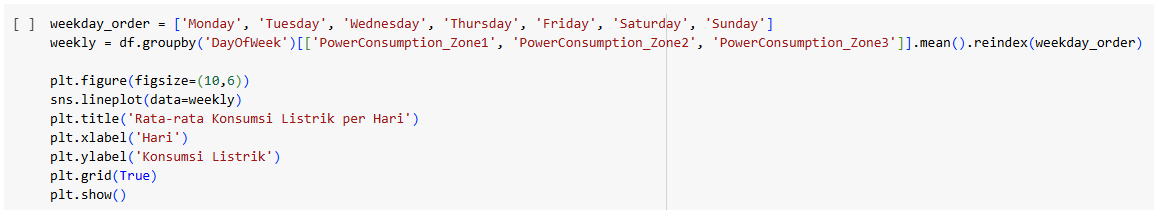
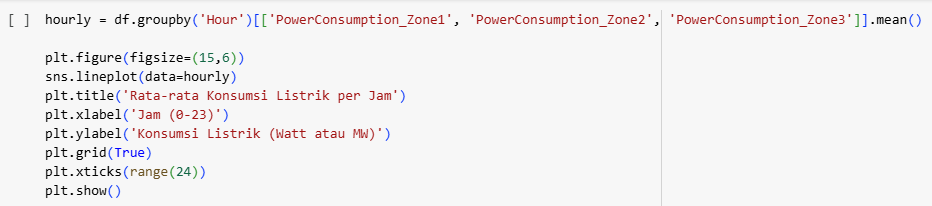
Dalam dataset disediakan 3 faktor lingkungan yaitu temperatur, kecepatan angin, dan kelembaban. Untuk mengetahui faktor apa yang paling mempengaruhi penggunaan listrik, maka digunakan perhitungan pearson. Teknik pearson adalah teknik untuk menghitung korelasi antar 2 variabel, dengan hasil koefisien. 

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.

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.

1. Time Series Analysis

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



# BAB III

**Diskusi dan Hasil Analisis**

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

## Basic Statistic

### 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.

### 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.

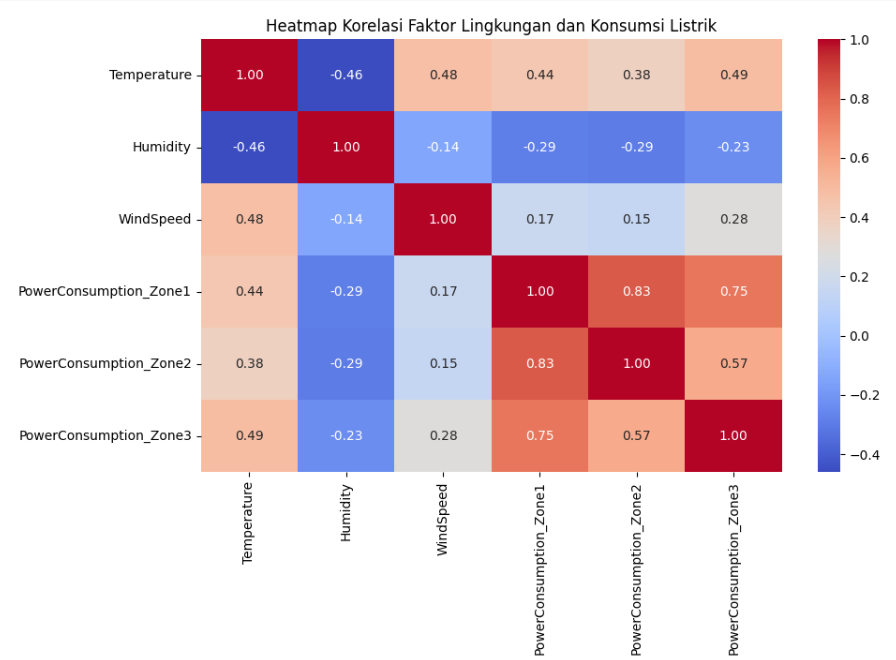
### Zone 3 (Hijau)

* 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.

## Analisis Korelasi

Analisis korelasi dilakukan untuk memahami hubungan linear antara variabel lingkungan (temperatur, kelembaban, dan kecepatan angin) terhadap konsumsi listrik di tiga zona berbeda. Korelasi diukur menggunakan metode Pearson, yang menghasilkan nilai antara -1 dan 1 yang berarti:

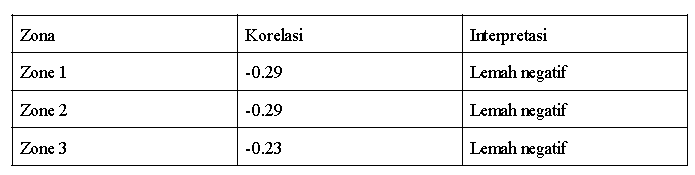
* Mendekati 1 → korelasi positif kuat (kedua variabel meningkat bersamaan).
* Mendekati -1 → korelasi negatif kuat (satu meningkat, yang lain menurun).
* Mendekati 0 → tidak ada hubungan linear signifikan.

Gambar diatas menunjukkan heatmap hasil perhitungan korelasi, dari heatmap ini, kita dapat menarik beberapa kesimpulan penting yang akan dibagi kedalam beberapa sub bab dibawah ini:

### Pengaruh Temperatur

Temperatur menunjukkan korelasi positif sedang dengan konsumsi listrik di semua zona. Hal ini berarti ketika suhu udara meningkat, konsumsi listrik juga cenderung meningkat. Efek ini paling kuat terlihat di Zona 3.

### Pengaruh Kelembaban



Kelembaban memiliki korelasi negatif lemah terhadap konsumsi listrik. Artinya, saat kelembaban meningkat, konsumsi listrik sedikit menurun. Hal ini bisa disebabkan karena udara yang lembab sering kali terasa lebih sejuk, sehingga orang mungkin mengurangi penggunaan alat pendingin seperti AC. Namun, pengaruh ini tidak terlalu signifikan karena korelasinya tergolong rendah.

### Pengaruh Kecepatan Angin

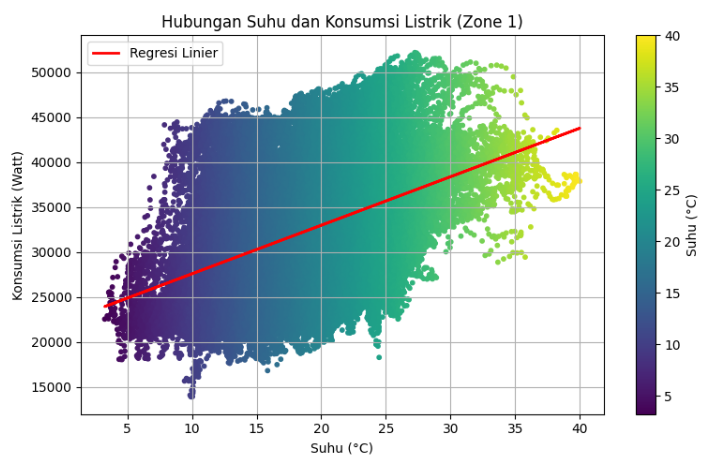
Kecepatan angin memiliki korelasi yang tergolong lemah terhadap konsumsi listrik. Hanya di Zona 3 korelasinya sedikit lebih tinggi, namun tetap di bawah ambang korelasi sedang. Angin mungkin hanya sedikit mempengaruhi perubahan suhu atau ventilasi alami, tapi dampaknya tidak cukup besar untuk memicu perubahan signifikan pada konsumsi listrik.

### Korelasi Antar Zona

Terdapat korelasi tinggi antara zona konsumsi listrik, terutama antara Zona 1 dan Zona 2. Zona-zona ini kemungkinan terpengaruh oleh faktor eksternal yang sama, seperti musim, hari kerja vs. akhir pekan, atau kebijakan nasional (misalnya jam kerja). Korelasi ini menunjukkan bahwa meskipun zona berbeda, pola konsumsi mereka bergerak searah, walaupun dalam skala dan intensitas yang berbeda. Dari hasil analisis:

* 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.

### Analisis Regresi Hubungan Temperatur Terhadap Listrik

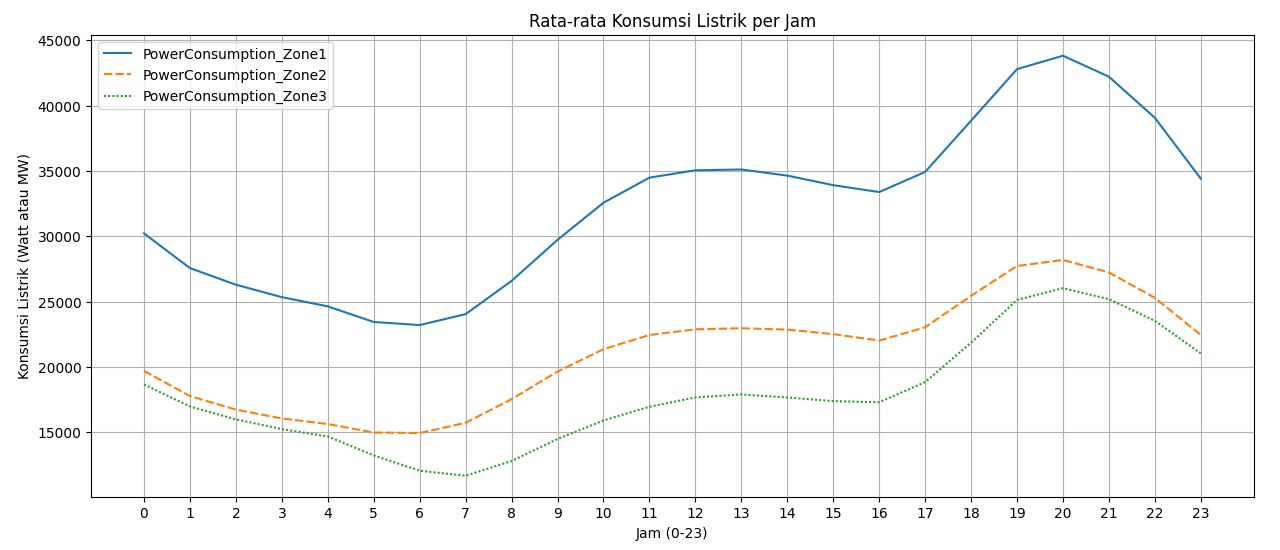
Untuk menguji seberapa kuat pengaruh suhu terhadap konsumsi listrik, dilakukan analisis regresi linier sederhana. Gambar scatter plot di atas memperlihatkan hubungan antara suhu (sumbu X) dan konsumsi listrik (sumbu Y) di Zona 1, dilengkapi dengan garis regresi berwarna merah yang menunjukkan tren hubungan linear antara keduanya.

* 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.

Terdapat hubungan positif yang cukup kuat dan konsisten. Analisis ini memperkuat hasil korelasi sebelumnya, di mana temperatur merupakan faktor lingkungan paling dominan yang mempengaruhi tingkat konsumsi listrik. Hal ini memperkuat dugaan bahwa penggunaan perangkat pendingin (seperti AC) meningkat seiring dengan naiknya suhu.

## Analisis Pola Musiman

### 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.

### Pola Mingguan

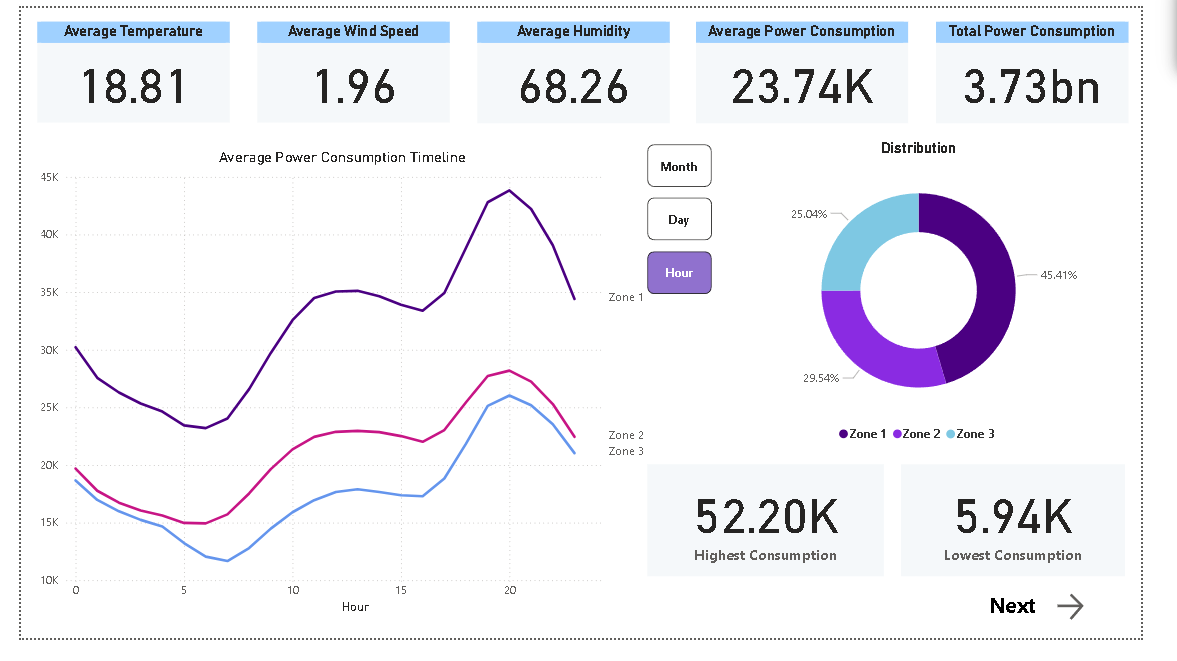
* Konsumsi listrik cenderung stabil sepanjang minggu, dengan sedikit fluktuasi.
* Terdapat penurunan konsumsi yang nyata pada hari Minggu, terutama di Zona 1 dan 2.
* Zona 3 menunjukkan pola yang relatif stabil sepanjang minggu.

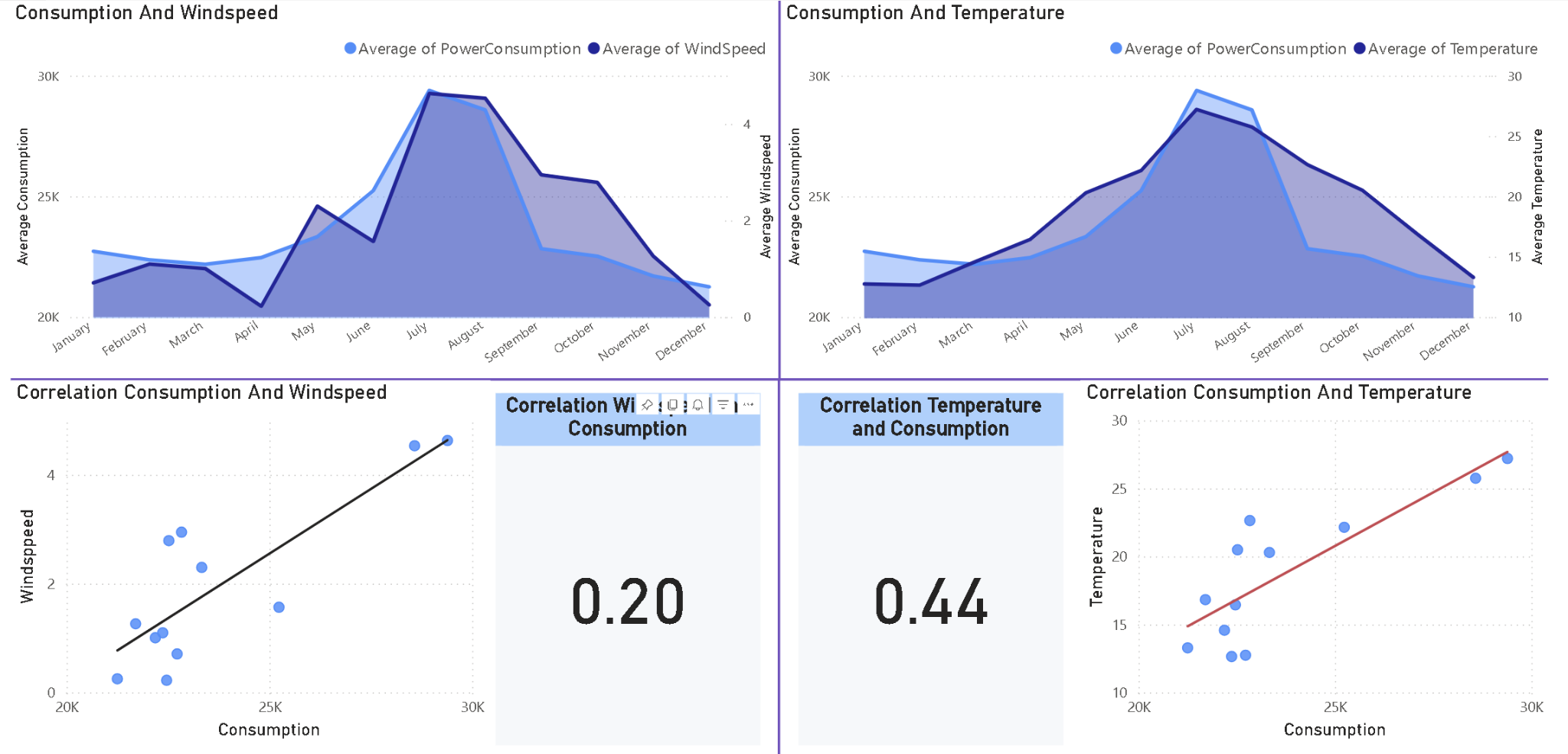
### Pola Bulanan

* 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.

Hasil yang didapat:  
 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).

## Visualisasi Data (Power BI Dashboard)





# BAB IV

**Kesimpulan**

Zona 1 tercatat sebagai area dengan konsumsi listrik tertinggi dan paling stabil sepanjang tahun, dengan median sekitar 32.265 watt dan hanya sedikit nilai outlier, menunjukkan pola penggunaan yang konsisten, mungkin karena karakteristiknya yang lebih industri atau komersial. Zona 2 berada di tingkat menengah, median sekitar 20.823 watt dengan fluktuasi moderat, indikasi zona semi-komersial atau campuran perumahan. Sementara itu, Zona 3 memiliki median terendah, sekitar 16.415 watt, namun sering mengalami outlier yang cukup banyak, menandakan konsumsi yang tidak stabil dan sangat dipengaruhi aktivitas harian di area residensial.

Dari sisi faktor lingkungan, suhu udara muncul sebagai pendorong utama perubahan konsumsi listrik di ketiga zona dimana semakin tinggi suhu, maka semakin meningkat penggunaan, terutama untuk perangkat pendingin di Zona 3. Kelembaban berpengaruh negatif meski lemah, karena udara yang lebih lembab terasa sedikit lebih sejuk sehingga kebutuhan pendingin menurun. Kecepatan angin nyaris tidak berdampak signifikan, hanya sedikit terasa pada Zona 3. Selain itu, pola konsumsi di masing-masing zona menunjukkan korelasi yang cukup kuat, mengindikasikan bahwa faktor eksternal seperti perubahan musiman atau hari libur nasional mempengaruhi penggunaan secara bersamaan di seluruh wilayah.

Untuk pola musiman, konsumsi harian memuncak dua kali tepatnya pada sekitar pukul 13.00 dan puncak lebih tinggi di pukul 20.00, sementara periode paling sepi terjadi pada dini hari (00.00–06.00). Secara mingguan, terlihat penurunan nyata pada hari Minggu di Zona 1 dan 2, mencerminkan aktivitas ekonomi yang lebih rendah di akhir pekan. Jika dilihat dari bulan, puncak konsumsi berada pada musim panas (Agustus untuk Zona 1 dan 2, Juli untuk Zona 3), sedangkan musim semi (April–Mei) mencatat penurunan. Pada musim dingin, Zona 2 mengalami kenaikan kecil di November, sementara Zona 1 dan 3 justru mencatat level terendah pada periode yang sama. Pemahaman tentang pola harian, mingguan, dan tahunan ini penting untuk membantu penyedia energi mengoptimalkan distribusi dan merancang kebijakan tarif atau program edukasi penggunaan listrik yang lebih bijak.

LINK GOOGLE COLAB: <https://colab.research.google.com/drive/1Nf9_4CRpYm85e04EdyCQNRXX_OUZJVsB?usp=sharing>

LINK POWER BI

<https://app.powerbi.com/links/CM34Hb-NOT?ctid=3485b963-82ba-4a6f-810f-b5cc226ff898&pbi_source=linkShare>