

Utilizando o dataset Individual Household Electric Power Consumption, disponível no repositório da UCI Machine Learning Repository, resolva os exercícios abaixo. O dataset contém medições de consumo elétrico em uma residência entre 2006 e 2010, com registros a cada minuto.

1. Carregue o dataset e exiba as 10 primeiras linhas.
2. Explique a diferença entre as variáveis `Global_active_power` e `Global_reactive_power`.
3. Verifique se existem valores ausentes no dataset. Quantifique-os.
4. Converta a coluna `Date` para o tipo `datetime` e crie uma nova coluna com o dia da semana correspondente.
5. Filtre os registros apenas do ano de 2007 e calcule a média de consumo diário de `Global_active_power`.
6. Gere um gráfico de linha mostrando a variação de `Global_active_power` em um único dia à sua escolha.
7. Crie um histograma da variável `Voltage`. O que pode ser observado sobre sua distribuição?
8. Calcule o consumo médio por mês em todo o período disponível no dataset.
9. Identifique o dia com maior consumo de energia ativa global (`Global_active_power`).
10. Compare o consumo médio de energia ativa global em dias de semana versus finais de semana.
11. Calcule a correlação entre as variáveis `Global_active_power`, `Global_reactive_power`, `Voltage` e `Global_intensity`.
12. Crie uma nova variável chamada `Total_Sub_metering` que some `Sub_metering_1`, `Sub_metering_2` e `Sub_metering_3`.
13. Verifique se há algum mês em que `Total_Sub_metering` ultrapassa a média de `Global_active_power`.
14. Faça um gráfico de série temporal do `Voltage` para o ano de 2008.
15. Compare o consumo entre os meses de verão e inverno (no hemisfério norte).
16. Aplique uma amostragem aleatória de 1% dos dados e verifique se a distribuição de `Global_active_power` é semelhante à da base completa.
17. Utilize uma técnica de normalização (Min-Max Scaling) para padronizar as variáveis numéricas principais.
18. Aplique K-means para segmentar os dias em 3 grupos distintos de consumo elétrico. Interprete os resultados.
19. Realize uma decomposição de série temporal (tendência, sazonalidade e resíduo) para `Global_active_power` em um período de 6 meses.
20. Treine um modelo de regressão linear simples para prever `Global_active_power` a partir de `Global_intensity`. Avalie o erro do modelo.