



Tópicos Especiais em Computação I

**Universidade Regional Integrada do Alto Uruguai e das Missões -
Campus Erechim**

Prof. Jackson Felipe Magnabosco

A descoberta de conhecimento em base de dados (KDD)



Introdução

- Definir descoberta de conhecimento em base de dados.
- Explicar as etapas do processo de descoberta de conhecimento em base de dados.
- Demonstrar o processo de descoberta de conhecimento em base de dados.

Etapas do Processo KDD

[illegible]

Extração de Conhecimento em Bases de Dados

Introdução ao KDD

Knowledge Discovery in Databases

Descoberta de conhecimento em base de dados.





O que é KDD

- O processo de extração de conhecimento de base de dados.
- Essencial para diversos setores da economia.
- Busca por informações úteis em grandes conjuntos de dados.



Etapas do KDD

- Entendimento do Domínio da Aplicação e Seleção de Dados
- Sugestão de Ferramentas Computacionais
- Pré-processamento e Transformação dos Dados
- Mineração de Dados e Interpretação de Resultados



Entendimento do Domínio

- Definir objetivos desejados.
- Monitorar variáveis importantes.
- Buscar fontes de dados complementares.



Ferramentas Computacionais

- Desenvolvimento de algoritmos.
- Ferramentas comuns:
 - Python
 - R
 - Matlab



Pré-processamento de Dados

- Integrar dados de diferentes bases.
- Lidar com dados faltantes.
- Evitar falhas de aquisição que comprometam o processo.



Mineração de Dados

- Verificar se os resultados são consistentes.
- Utilizar gráficos para interpretar e transmitir os resultados.
- Importância de transformar dados em informações sólidas.

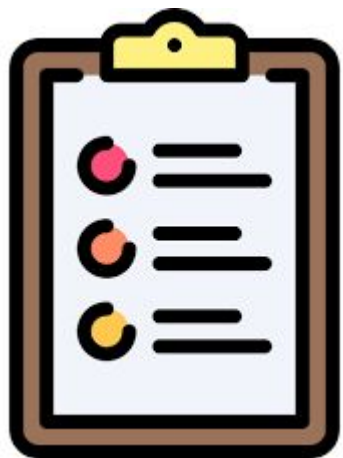


Conclusão

- O KDD é aplicado em diversas áreas e auxilia na tomada de decisões com base em dados.
- A extração de conhecimento pode ser transformadora quando bem executada.

KDD aplicado em dados de consumo de energia

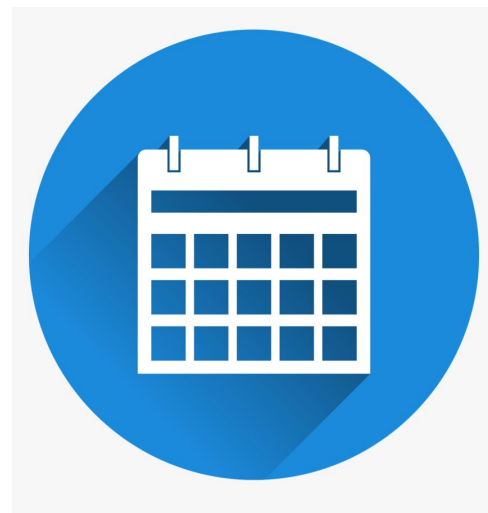
Dados



9 variáveis

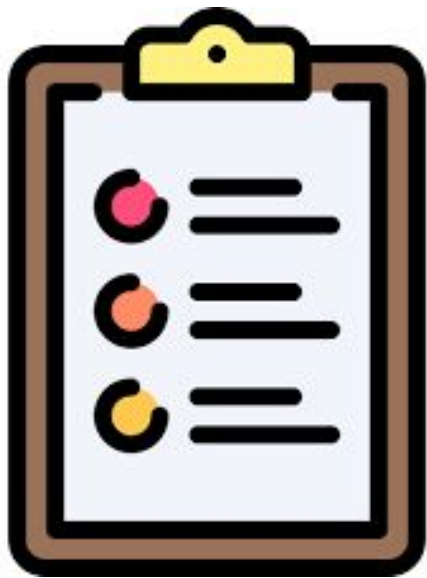


2 milhões de
observações



De 12/2006 a
11/2010

Variáveis



- Data.
- Horário.
- Potência ativa média por minuto (kW).
- Potência reativa média por minuto (kW).
- Tensão por minuto.
- Intensidade da corrente média por minuto.
- Média de energia.

Grupo 1

- Lavadora.
- Forno elétrico.
- Micro-ondas.



Grupo 2

- Lavadora.
- Secadora.
- Geladeira
- Luz




Grupo 3

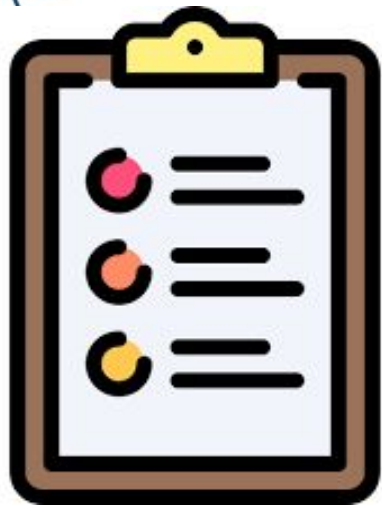
- Aquecedor
- Ar-condicionado





Objetivos desejados com o KDD

- Classificador do perfil de consumo energético.
 - Grupo 1
 - Grupo 2
 - Grupo 3
- 



Dados
coletados



Problema
formulado

<code>

Desenvolvimento
de algoritmos

Atividade



Gerar Gráficos de Predição e Distribuição

Objetivo: Carregar os dados do arquivo `household_power_consumption.txt`, treinar um modelo de regressão para prever `Global_active_power`, e gerar gráficos que comparem as previsões com os valores reais, além de gráficos de distribuição para várias variáveis.

Instruções:

1. Carregue os dados do arquivo.
2. Separe os dados em conjuntos de treino e teste.
3. Treine um modelo de regressão.
4. Gere um gráfico que compare os valores reais com as previsões.
5. Gere gráficos de distribuição para as variáveis `Global_active_power`, `Global_reactive_power`, `Voltage`, `Global_intensity`, `Sub_metering_1`, `Sub_metering_2` e `Sub_metering_3`.
6. Salve todos os gráficos em arquivos.

Ferramentas



Python





Python



- Um dos principais recursos.
- Linguagem mais utilizada na ciência de dados
- Infinitude de recursos.
- Gratuito




Como Baixar o Python

Acesse o site oficial do Python:

- Abra o navegador e vá para <https://www.python.org>.

Baixe o Python:

- Clique no botão "**Download Python**".
 - O site detecta automaticamente o sistema operacional do seu computador (Windows, Mac, ou Linux).
 - Confirme se a versão mais recente está sendo baixada.
- 

Python

[Donate](#)[GO](#)[Socialize](#)[About](#)[Downloads](#)[Documentation](#)[Community](#)[Success Stories](#)[News](#)[Events](#)

```
# Python 3: Fib
>>> def fib(n):
>>>     a, b =
>>>     while a
>>>         pri
>>>         a,
>>>     print()
>>> fib(1000)
0 1 1 2 3 5 8 1
```

[All releases](#)[Source code](#)[Windows](#)[macOS](#)[Other Platforms](#)[License](#)[Alternative Implementations](#)

Download for Windows

[Python 3.13.0](#)

Note that Python 3.9+ *cannot* be used on Windows 7 or earlier.

Not the OS you are looking for? Python can be used on many operating systems and environments.

[View the full list of downloads.](#)

Python is a programming language that lets you work quickly
and integrate systems more effectively. [>>> Learn More](#)



Como Instalar o Python

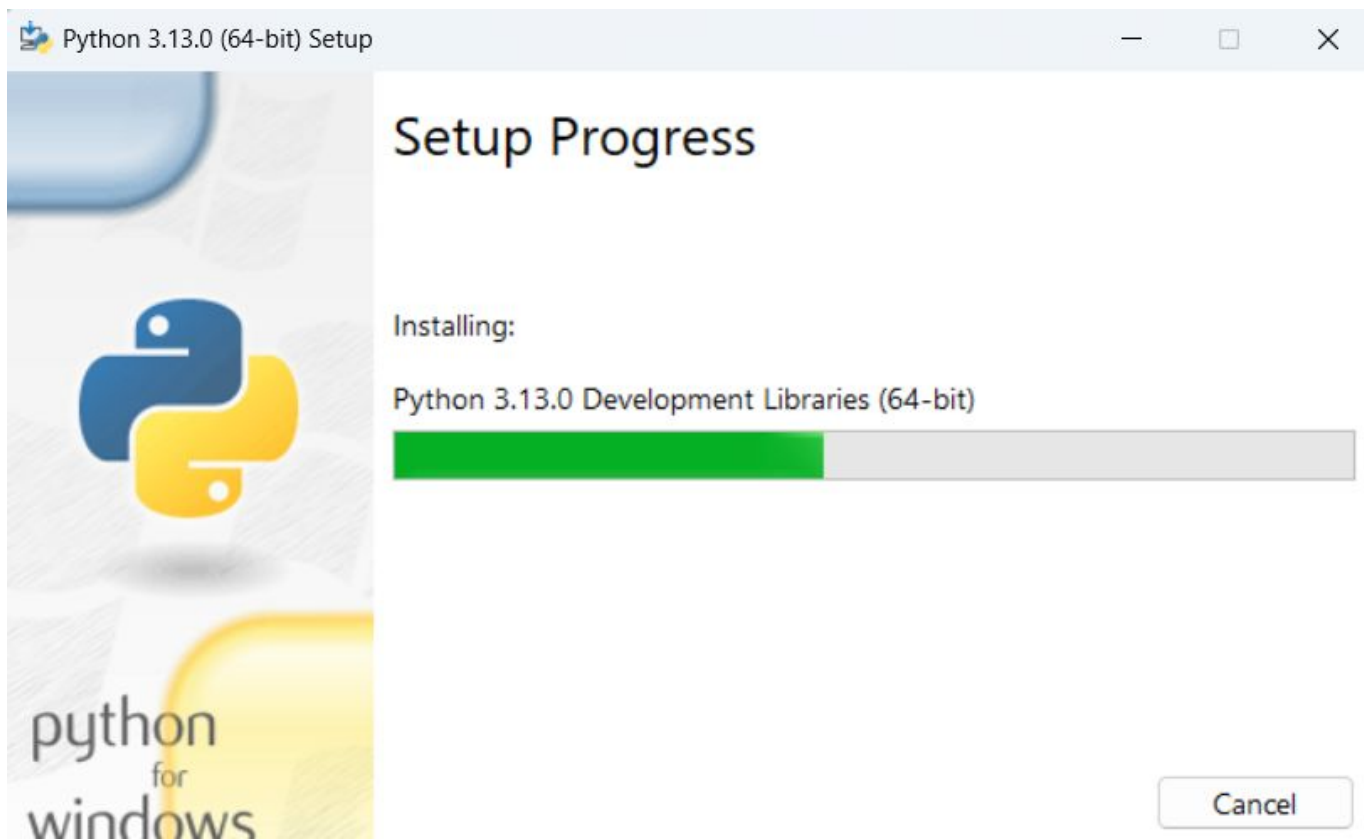
Instale o Python:

- Execute o arquivo baixado.
- **IMPORTANTE:** Marque a opção "**Add Python to PATH**" para que o Python funcione no terminal.
- Clique em "**Install Now**" e aguarde.

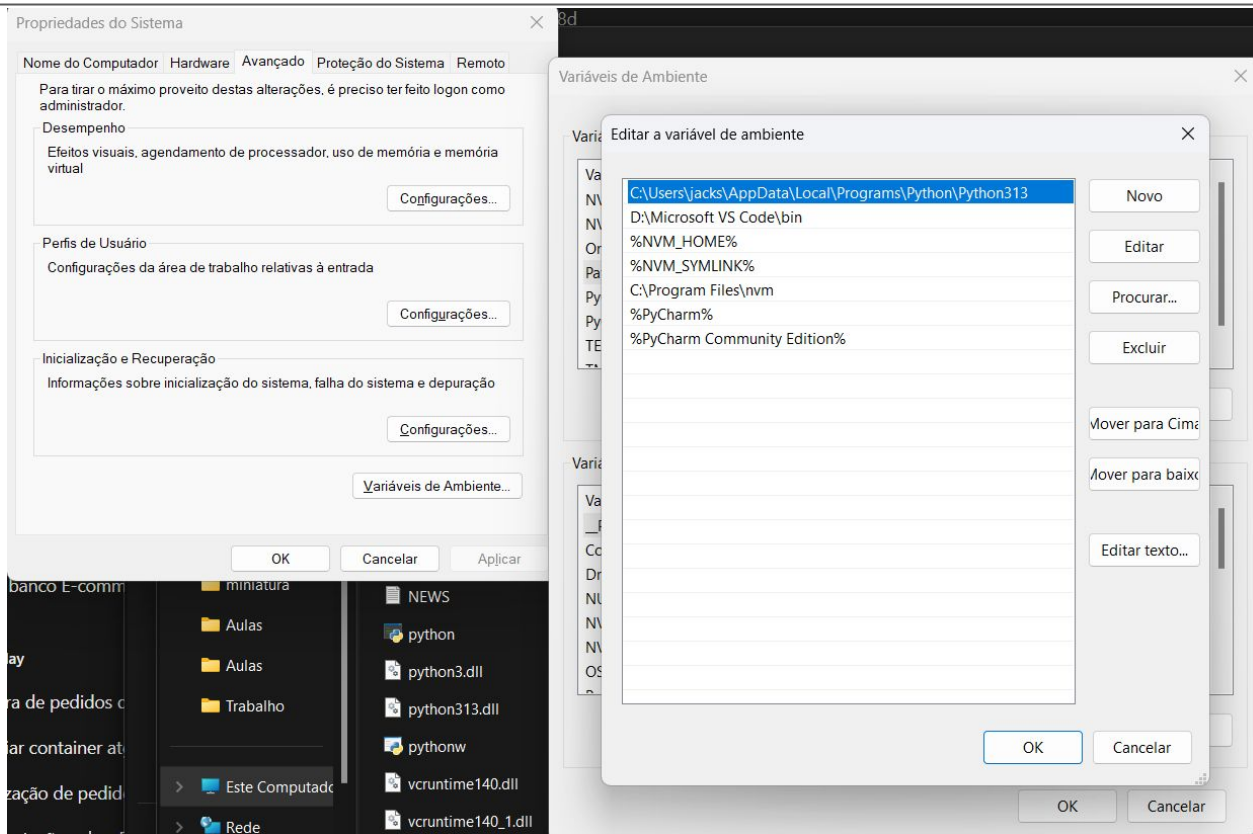
Verifique a instalação:

- Abra o **Prompt de Comando** ou terminal.
- Digite `python --version` ou `python3 --version` para verificar se o Python foi instalado corretamente.

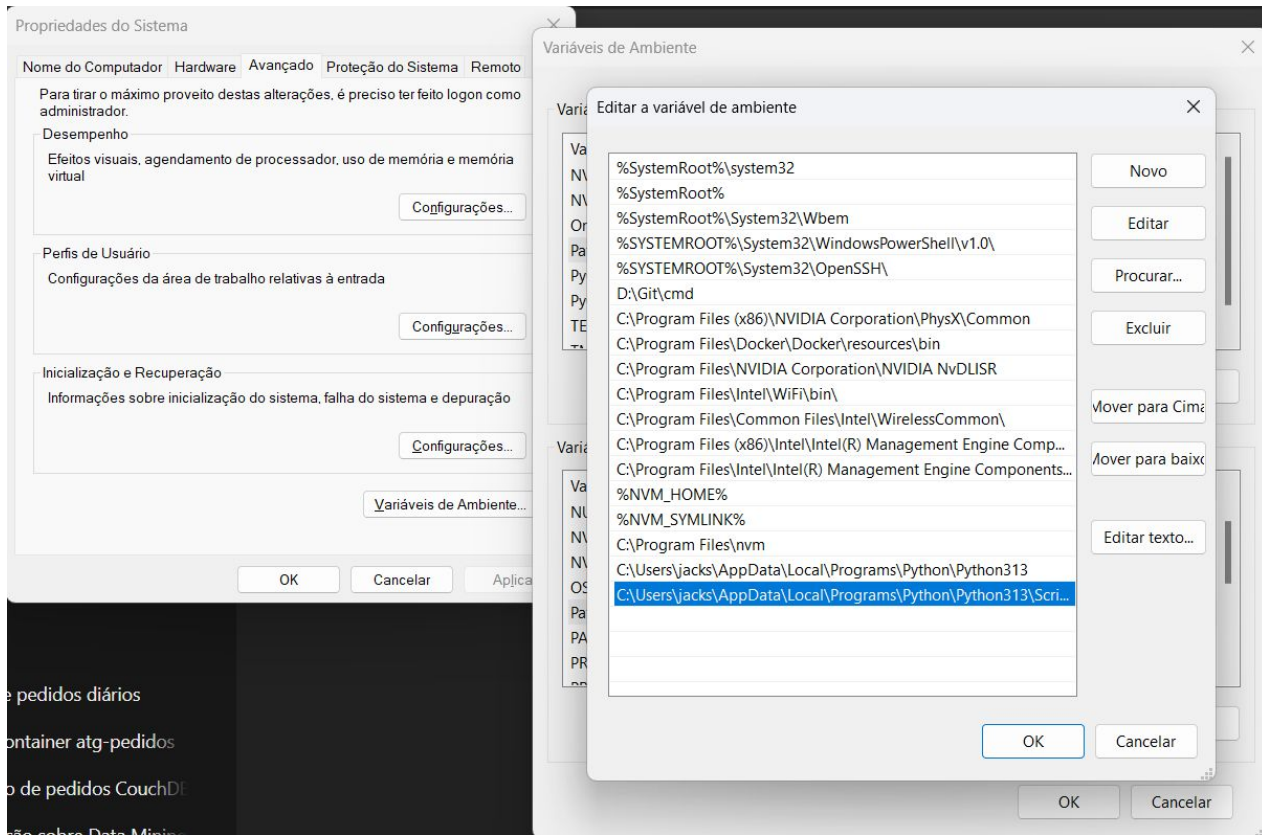
Como Instalar o Python



Como adicionar o path do Python



Como adicionar o path do Pip



Como validar



Git CMD



```
C:\Users\jacks>python --version
```

```
Python 3.13.0
```

```
C:\Users\jacks>pip --version
```

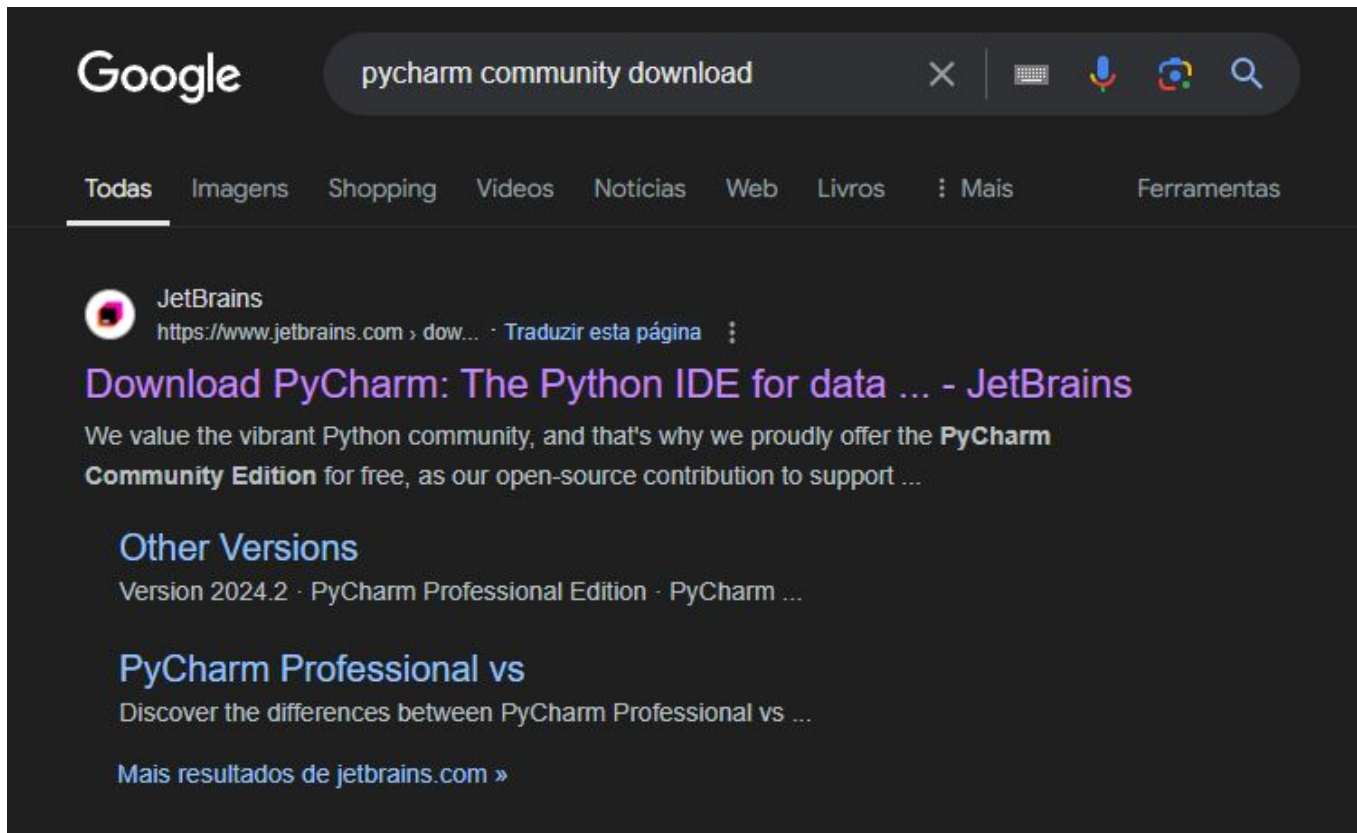
```
pip 24.2 from C:\Users\jacks\AppData\Local\Programs\Python\Python313\Lib\site-packages\pip (python 3.13)
```



Pycharm



Como baixar o Pycharm



Como baixar o Pycharm



PyCharm Community Edition

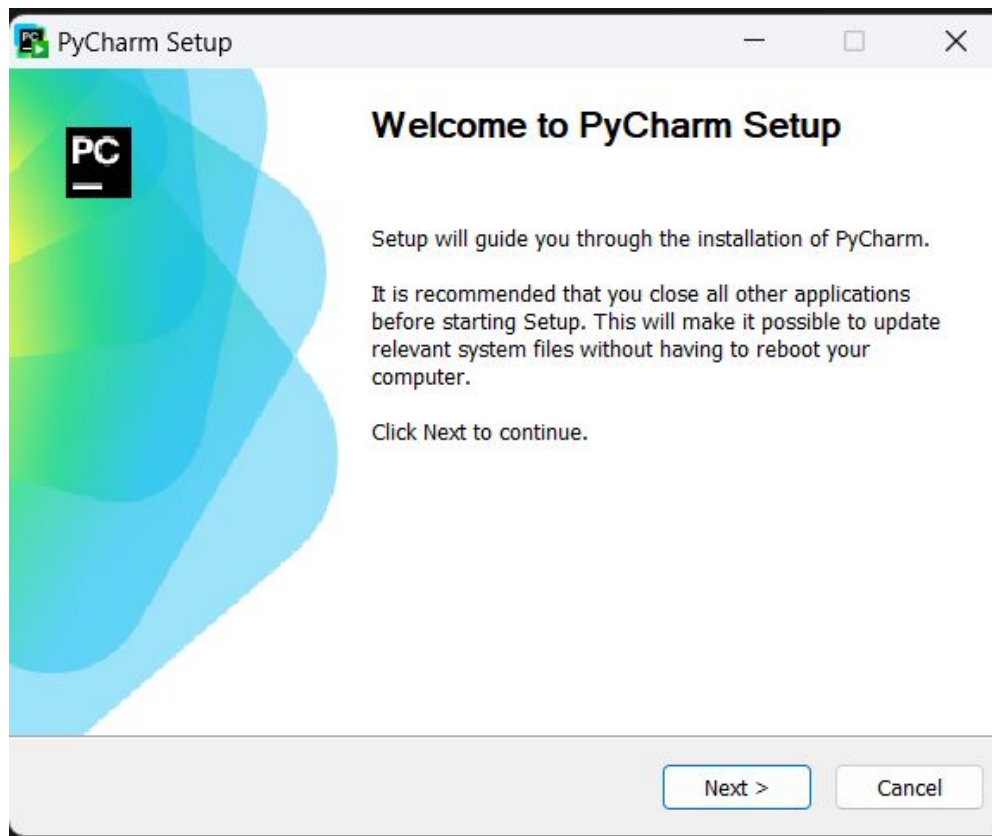
The IDE for Pure Python Development

Download

.exe (Windows) ▼

Free, built on open source

Como instalar o Pycharm



Baixar base de dados

[Datasets](#)[Contribute Dataset](#)[About Us](#)

Individual Household Electric Power Consumption

Donated on 8/29/2012

Measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering values are available.

Dataset Characteristics

Multivariate, Time-Series

Subject Area

Physics and Chemistry

Associated Tasks

Regression, Clustering

Feature Type

Real

Instances

2075259

Features

9

[DOWNLOAD \(19.7 MB\)](#)[IMPORT IN PYTHON](#)[CITE](#)

29 citations

75163 views

Keywords

power consumption



Pip install

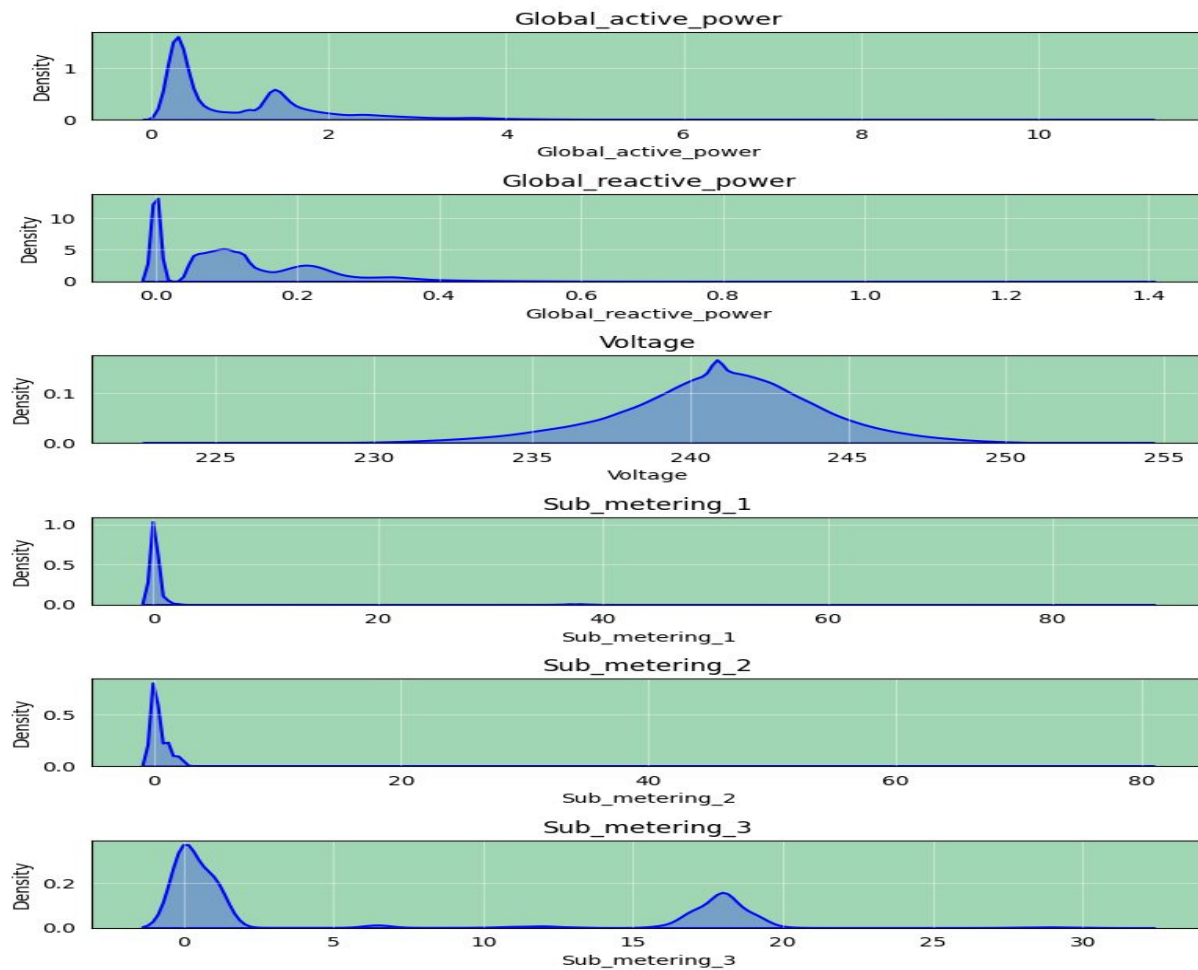
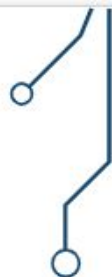


```
.venv\Scripts\activate
```

```
pip install pandas numpy matplotlib seaborn  
scikit-learn impute
```

```
python kdd.py
```

Resultados



Hora de codificar!
Vamos mergulhar no Python e
construir nossas análises.

Hands on





Disponível em:

- **GitHub:**
`https://github.com/jacksonn455/kdd-topicos-especiais-computacao/tree/main`

Dúvidas ou sugestões ?





Tópicos Especiais em Computação I

**Universidade Regional Integrada do Alto Uruguai e das Missões -
Campus Erechim**

Prof. Jackson Felipe Magnabosco