

+CiênciaSJC

Exploração e Análise de Dados usando Python

Jonas Lucas Durão

Introdução

Este cronograma foi desenvolvido para o projeto **+CiênciaSJC**, com o tema *Exploração e Análise de Dados usando Python*, com foco em estudantes do Ensino Médio. O objetivo é introduzir conceitos de ciência de dados, modelagem computacional e aprendizado de máquina, utilizando a linguagem Python e a biblioteca Pandas desde o início do processo de aprendizagem.

Módulo 1: Estatística e Visualização de Dados

Objetivo: Explorar conceitos estatísticos e aprender a visualizar dados de forma clara e intuitiva.

Tópicos:

- Estatística descritiva com Pandas: média, mediana, moda, variância e desvio padrão.
- Introdução à visualização de dados:
 - Gráficos com `pandas.plot()`.
 - Personalização de gráficos com `matplotlib` e `seaborn`.
- Projeto prático: visualização e análise de gastos diários durante o ano.

Módulo 2: Limpeza e Preparação de Dados

Objetivo: Ensinar técnicas fundamentais para preparar dados para análise e modelagem.

Tópicos:

- Tratamento de valores ausentes: `dropna()`, `fillna()`.
- Correção de tipos de dados e normalização.
- Manipulação de DataFrames: filtragem, ordenação, criação de novas colunas.
- Projeto prático: preparação de um dataset sobre empregos e salários para análise posterior.

Módulo 3: Introdução à Modelagem Computacional e Simulação

Objetivo: Explorar a criação de modelos computacionais simples para simular fenômenos e realizar previsões.

Tópicos:

- Conceito de modelagem computacional.
- Modelos simples com Pandas: projeções de crescimento populacional e simulação de consumo energético.
- Introdução à probabilidade aplicada em simulações.
- Projeto prático: simulação de investimentos previsão de gastos energéticos.

Módulo 4: Introdução ao Aprendizado de Máquina (Machine Learning)

Objetivo: Apresentar noções básicas de aprendizado de máquina de forma prática, utilizando ferramentas acessíveis.

Tópicos:

- Conceitos iniciais de Machine Learning.
- Separação de dados e avaliação dos dados.
- Modelos iniciais com `scikit-learn`: regressão linear, árvore de decisão e KNN.
- Avaliação de modelos: erro médio absoluto, R^2 , matriz de confusão.
- Projeto prático: previsão de gastos mensais ou de valorização de investimentos.

Módulo 5: Projeto Final Integrador

Objetivo: Consolidar conhecimentos através de um projeto de análise de dados e modelagem computacional.

Sugestões de projetos:

- Análise e previsão de comportamento de consumo entre adolescentes.
- Simulação de crescimento e evolução de salários em diferentes setores.
- Análise comparativa de valorização de diferentes criptomoedas e opções de investimento.

Atividades:

- Apresentação dos resultados.
- Discussão sobre os impactos sociais, éticos e científicos da análise e modelagem de dados.

Ferramentas e Recursos

- **Linguagem:** Python 3.
- **Bibliotecas:** pandas, numpy, matplotlib, seaborn, scikit-learn.
- **Ambiente:** Google Colab ou Jupyter Notebook.

Competências Desenvolvidas (BNCC)

- **EM13CO11:** Criar e explorar modelos computacionais simples para simular e fazer previsões.
- **EM13CO12:** Produzir, analisar, gerir e compartilhar informações a partir de dados, utilizando princípios de ciência de dados.