

SITUAÇÃO DE APRENDIZAGEM

Curso: Programação em Python para Data Science

Unidade curricular (UC): Programação em Python para Data Science

Carga horária da UC: 60h

Nº de aulas: 20

Carga horária prevista para o desenvolvimento da Situação de Aprendizagem:

Objetivo da UC:

O Curso de Aperfeiçoamento Profissional Programação em Python tem por objetivo capacitar profiss para desenvolver aplicações em linguagem Python, por meio de técnicas de programação, seguindo práticas, procedimentos e normas.

Capacidades a serem desenvolvidas: Básicas () Técnicas (X) Socioemocionais ()

1. Capacidade analítica e interpretativa
2. Tomada de decisão baseada em dados
3. Domínio de ferramentas de análise de dados
4. Comunicação de resultados
5. Resolução de problemas com raciocínio lógico e computacional
6. Trabalho com projetos e entregas organizadas

Conhecimentos

Média, mediana, moda, variância, desvio padrão, distribuição de frequências

Criação e interpretação de gráficos: histogramas, dispersão, boxplots, barras

Sintaxe, manipulação de dados com Pandas, arrays com NumPy, uso de bibliotecas externas

Modelagem com regressão linear, avaliação com métricas (MSE, RMSE, R^2)

Limpeza de dados, tratamento de nulos, conversão de tipos, criação de colunas derivadas

Construção de narrativas visuais e textuais baseadas em dados reais

Google Colab, bibliotecas de visualização e modelagem

Estratégia de aprendizagem desafiadora

Situação-problema () Estudo de caso () Pesquisa Aplicada () Projeto (X) Integrador ()

Contextualização:

Você foi contratada(o) como cientista de dados júnior por uma empresa de análise de performance de streaming chamada **MovieScope**. Sua missão é analisar dados de filmes disponíveis em plataformas digitais, com o objetivo de gerar **insights sobre padrões de sucesso, características que influenciam a nota dos filmes e criar uma previsão de avaliação com base em dados históricos**.

Dataset Utilizado

🔗 **Dataset:** [IMDb Movies Dataset - TMDb 5000 Movie Dataset](#)

Este conjunto de dados contém informações sobre milhares de filmes, incluindo:

- Nome, diretor, elenco
Receita, orçamento, gênero
- Avaliações (nota média)
- Idioma original
Data de lançamento, popularidade
- e mais.

Desafio:

Exploração Inicial

- Importar e visualizar o dataset com **pandas**.
- Corrigir problemas de codificação, valores ausentes e inconsistências.
Analisar variáveis categóricas e numéricas.

Estatística Descritiva

- Calcular média, mediana, desvio padrão e variância de atributos como **orçamento, receita, nota média, popularidade**.
- Criar uma análise interpretativa desses valores.

Visualização de Dados

- Criar visualizações com **matplotlib** e **seaborn**:
-

-
- Histogramas (ex: distribuição das notas)
 - Gráficos de dispersão (ex: orçamento vs. receita)
 - Boxplots (ex: avaliação por gênero)
 - Gráficos de barras (ex: número de filmes por idioma)

Modelagem Preditiva

- Criar um modelo de **regressão linear** para prever a **nota média (vote_average)** com base em variáveis como:
 - orçamento
 - receita
 - popularidade
 - número de votos
- Avaliar o modelo com métricas apropriadas: MSE, RMSE, R^2 .

Apresentação Profissional

Elaborar uma apresentação final que conte a história dos dados:

Cenário → Perguntas → Análises → Modelos → Conclusões → Sugestões de negócio

Resultados esperados:

1. Notebook no Google Colab com análise, gráficos e modelos.
2. Apresentação final com storytelling.

Anexos (Figuras, esquemas, desenhos, leiaute, formulários, etc):

ESTRATÉGIAS DE ENSINO E INSTRUMENTOS DE AVALIAÇÃO E RECURSOS

Nº horas / aulas	Capacidades a serem trabalhadas	Conhecimentos relacionados	Estratégias de ensino e instrumentos de avaliação	Recursos e ambientes pedagógicos
Aula 1 (3h)	Aplicações em Data Science: Introdução à Data Science Compreender conceitos fundamentais de estatística descritiva. Tipos de Gráfico.	O que é Data Science e sua área de atuação. Data Science vs. Inteligência Artificial O que é estatísticas? Conceito de dados, informação e conhecimento. Tipos de gráficos.	Aula expositiva dialogada, exercícios práticos com dados fictícios. Avaliação diagnóstica.	Quadro, projetor, Google Colab
Aula 2 (3h)	Medidas de tendência central Analisar a variabilidade dos dados com medidas de dispersão.	Média, moda, mediana Amplitude, variância, desvio padrão	Aula invertida	Quadro, projetor, laboratório
Aula 3 (3h)	Introdução ao Colab e ferramentas que usaremos no curso. Como o Colab funciona? Calculando no Colab. Tipos de dados e estrutura de dados (dict, lists)	Média, moda, mediana Amplitude, variância, desvio padrão Variáveis, tipos de dados, dict, lists	Uso da lib statistics para cálculos Lista de atividades Uso de dados tratados e fictícios	Google Colab, statistics
Aula 4 (3h)	Representar dados graficamente e interpretar padrões no Colab 3.3.1. Plotagem 3.3.2. Exportação de Gráficos 3.3.3. KPIs 3.3.4. Gráficos	Histogramas, gráficos de dispersão, correlação e regressão linear	Exploração de gráficos reais e criação em Python com matplotlib Uso de dados tratados e fictícios	Google Colab, matplotlib
Aula 5 (3h)	Programação em Python: Tipos de dados e Manipulação de arquivos usando pandas Pandas 3.2.1. Comandos 3.2.2. Dataframes Conhecendo os dados	Tipos de dados, variáveis, dataframes, Arredondamento de valores	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Google Colab, Pandas
Aula 6 (3h)	Pandas Tratamento de Dados Condiçionais e Loops	Manipulação de arquivos Estruturas condicionais (if/else), laços de repetição (for/while), manipulação lógica de dados Agrupamentos com <code>groupby</code>	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Google Colab
Aula 7 (3h)	Analisando os dados com Pandas Pandas 3.2.3. Filtros 3.2.4. Gráficos	Filtros com condições, gráficos com Pandas	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Google Colab, Pandas
Aula 8 (3h)	Refinando os dados visuais Seaborn	Gráficos com Seaborn (boxplot, countplot, pairplot), análise de padrões visuais	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Google Colab

ESTRATÉGIAS DE ENSINO E INSTRUMENTOS DE AVALIAÇÃO E RECURSOS

Nº horas / aulas	Capacidades a serem trabalhadas	Conhecimentos relacionados	Estratégias de ensino e instrumentos de avaliação	Recursos e ambientes pedagógicos
	3.4.1.Comandos 3.4.2.Dados 3.4.3.Importação Pandas	Introdução a correlação		
Aula 9 (3h)	Seaborn - Atividade de Fixação Programação em Python: Relembrando estrutura de dados NumPy 3.1.1.Comandos; 3.1.2.Funções 3.1.3.Matrizes	Arrays, criação e manipulação de matrizes, funções básicas do NumPy	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Google Colab, Numpy
Aula 10 (3h)	Atividade de fixação e revisão da base da análise de dados	Biblioteca pandas, numpy e plotagem de gráficos	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Quadro, projetor
Aula 11 (3h)	Entender estruturas de modelos preditivos Scikit-learn 3.5.1.Comandos 3.5.2.Dados 3.5.3.Importação Utilizar bibliotecas de machine learning	Introdução à regressão linear, classificação, clustering e redução de dimensionalidade Scikit-learn: importação de dados, criação de modelos, métricas de avaliação	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Google Colab, Scikit-learn
Aula 12 (3h)	Reforçar testes e otimizações de modelos	Métricas de performance (R^2 , RMSE), tunagem de parâmetros	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Google Colab, Scikit-learn
Aula 13 (3h)	Reforçar testes e otimizações de modelos Atividades de Fixação de modelos preditivos	Ajuste de modelo, validação cruzada, interpretação de resultados Introdução à regressão linear, classificação, clustering e redução de dimensionalidade Scikit-learn: importação de dados, criação de modelos, métricas de avaliação Ajuste de modelo, validação cruzada, interpretação de resultados Métricas de performance (R^2 , RMSE), tunagem de parâmetros	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Google Colab, Scikit-learn
Aula 14 (3h)	Atividades de Fixação de modelos preditivos Testar hipóteses com modelos preditivos	Ajuste de modelo, validação cruzada, interpretação de resultados	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Google Colab

ESTRATÉGIAS DE ENSINO E INSTRUMENTOS DE AVALIAÇÃO E RECURSOS

Nº horas / aulas	Capacidades a serem trabalhadas	Conhecimentos relacionados	Estratégias de ensino e instrumentos de avaliação	Recursos e ambientes pedagógicos
		Introdução à regressão linear, classificação, clustering e redução de dimensionalidade Scikit-learn: importação de dados, criação de modelos, métricas de avaliação Ajuste de modelo, validação cruzada, interpretação de resultados Métricas de performance (R^2 , RMSE), tunagem de parâmetros		
Aula 15 (3h)	Extração de dados web com Python Scrapy 3.6.1.Comandos 3.6.2.Importação Análise de dados de sites • coleta de dados • limpeza de dados • organização de dados • visualização de dados Automação de extração	Web scraping com Scrapy: crawling, parsing, exportação de dados Coleta, limpeza, visualização e organização de dados coletados com Scrapy	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Google Colab, Scrapy
Aula 16 (3h)	Atividade: Scrapy 3.6.1.Comandos 3.6.2.Importação Comparar bancos relacionais e não relacionais	Coleta, limpeza, visualização e organização de dados coletados com Scrapy SQL vs. NoSQL, HDFS, estruturas de armazenamento em Data Science	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Quadro, projetor
Aula 17 (3h)	Atividade: Scrapy 3.6.1.Comandos 3.6.2.Importação	Coleta, limpeza, visualização e organização de dados coletados com Scrapy	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Quadro, projetor
Aula 18 (3h)	Planejar e construir projeto final	Integração dos conhecimentos: exploração, visualização, modelagem, apresentação	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Quadro, projetor
Aula 19 (3h)	Comparar bancos relacionais e não relacionais Criar dashboards interativos com Python Comunicar dados de forma eficaz	SQL vs. NoSQL, HDFS, estruturas de armazenamento em Data Science Interface visual com Streamlit, filtros dinâmicos, exibição de métricas e gráficos Princípios de visualização eficaz, estrutura de apresentação, storytelling com dados	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Quadro, projetor
Aula 20 (3h)	Finalizar e apresentar o projeto	Consolidação do trabalho, revisão de código, avaliação dos resultados	Código demonstrativo e exercícios guiados. Uso de dados de banco de ML	Quadro, projetor



ESTRATÉGIAS DE ENSINO E INSTRUMENTOS DE AVALIAÇÃO E RECURÇOS

Nº horas / aulas	Capacidades a serem trabalhadas	Conhecimentos relacionados	Estratégias de ensino e instrumentos de avaliação	Recursos e ambientes pedagógicos

Referências (livros, apostilas, sites, blog etc.):

Vale lembrar que:

- as estratégias e recursos aqui definidos poderão ser modificados pelo docente, desde que cumpra o desenvolvimento dos conteúdos formativos previstos;
- a intervenção mediadora deverá ser planejada pelo docente no momento da seleção das estratégias, considerando sua intencionalidade na aplicação;
- a avaliação formativa é feita durante o desenvolvimento de cada estratégia, cabendo ao docente definir os momentos em que os resultados do processo de ensino e de aprendizagem serão discutidos com os alunos.

INSTRUMENTO DE REGISTRO

Nome do aluno:		Turma:	
Capacidades básicas ou técnicas e socioemocionais	Critérios de Avaliação	Resultado	
		Aluno	Docente
<ul style="list-style-type: none"> Utilizar modelos estatísticos para coleta e tratamento de dados 	Aplica conceitos de média, mediana e desvio padrão para descrever um conjunto de dados.		
<ul style="list-style-type: none"> Elaborar programas em linguagem Python para coleta, limpeza, organização e visualização de dados 	Constrói scripts em Python utilizando estruturas de decisão (if/else) e laços (for/while) de forma lógica e funcional.		
<ul style="list-style-type: none"> Utilizar bibliotecas em python para data science 	Utiliza as bibliotecas Numpy e Pandas para manipular arrays e DataFrames.		
<ul style="list-style-type: none"> Realizar o tratamento, análise e visualização dos dados (refatoração) 	Realiza a limpeza de dados (tratamento de valores ausentes e duplicados) em um DataFrame.		
<ul style="list-style-type: none"> Demonstrar pensamento analítico 	Cria gráficos com Matplotlib/Seaborn para explorar e apresentar os dados de forma clara.		
<ul style="list-style-type: none"> Demonstrar autonomia 	Realiza as atividades propostas buscando soluções de forma independente e proativa.		
<ul style="list-style-type: none"> Demonstrar capacidade de organização 	Organiza o código e os arquivos do projeto de maneira lógica e limpa.		

Legenda:

- C= realizou
- R= não realizou



TABELA DE NÍVEIS DE DESEMPENHO

CrITÉRIOS de Avaliação	NÍvel de desempenho	Conversão em notas
Atingiu todos os objetivos com excelência	Nível 4	100 a 85
Atingiu os objetivos com segurança	Nível 3	84,9 a 70
Atingiu parcialmente os objetivos	Nível 2	69,9 a 50
Não atingiu os objetivos esperados	Nível 1	Abaixo de 50