

Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM



Carga horária total: 70 horas



**VAMOS
COMEÇAR!**

1

FASE 1

Python para Análise de Dados

- Variáveis & Tipos de Dados
- Estruturas de Dados
- Fluxo Condicional & Repetição
- Arquivos & Funções
- Programação Funcional
- Programação Orientada a Objetos
- Módulos e Pacotes
- Tratamento de Erros
- Scripting

Desenvolvimento em Ciência de Dados e Metodologia Crisp-DM

- Perspectivas de Carreira
- Metodologia
- Elementos básicos de Python e Numpy
- Introdução ao Pandas e carga de dados
- Limpeza e preparação de dados
- Descritiva I
- Árvores I
- Git/Github - Controle de versionamento

FASE 2

2

3

FASE 3

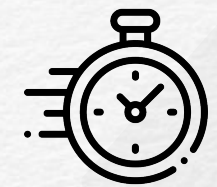
Desenvolvimento de aplicações com Python

- Agregações e operações em grupos
- Descritiva II
- Árvores de regressão
- Regressão I e II
- Scripting
- Streamlit I

BY
UIC

Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM



Carga horária total: 70 horas

Desenvolvimento de Modelos com Pandas e Python

- Métodos de análise
- Árvores II
- Regressão Logística I
- Streamlit II
- Markdown/ Terminal

FASE 4

4

5

FASE 5

Gestão avançada de Dados e Desenvolvimento de aplicações

- PostgreSQL
- Descritiva III
- Combinação de modelos
- Cálculo
- Algelin
- PCA
- Streamlit III

Agrupamento e segmentação de dados

- K-means
- Hierárquicos/aglomerativos
- Streamlit IV

FASE 6

6

BY
1100

Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM



Carga horária total: 70 horas

7

FASE 7

Modelagem estatística e Regressão com Python

- Probabilidade
- Inferência
- Regressão III e IV

Modelagem de dados categorizados

- Descritiva IV
- Regressão Logística II
- Streamlit V e Pycaret

FASE 8

8

9

FASE 9

SQL para análise de dados

- Base de dados e Linguagem SQL
- Tabelas em SQL
- Seleção e ordenação em SQL
- Filtragem e seleção condicional
- Agregações
- Múltiplas tabelas
- Técnicas avançadas



BY
UIC

Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM

Competência 1

Desenvolver habilidades em Python, abrangendo desde conceitos fundamentais até Programação Funcional e Orientada a Objeto. Formular módulos, pacotes e scripts para manutenção e reusabilidade de projetos. Avaliação e correção de erros em códigos Python, com expertise em manipulação de dados em diferentes formatos e criação de relatórios analíticos.



Habilidade 1

Empregar listas, conjuntos e dicionários para estruturar dados. Desenvolver funções organizadoras com parâmetros para tarefas específicas.

Habilidade 2

Implementar funções para tarefas repetitivas. Utilizar classes para criar e reutilizar objetos. Aplicar herança entre classes para reutilizar estruturas.

Habilidade 3

Corrigir erros de sintaxe e tempo de execução para performance. Validar códigos para garantir estruturação conforme padrões.

Habilidade 4

Selecionar dados relevantes para storytelling. Valorizar a apresentação gráfica dos dados usando bibliotecas.

FASE 1



Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM

Competência 2

Desenvolver uma visão geral em Ciência de Dados, aplicando metodologia Crisp DM, programação Python com Pandas para manipulação e análise eficaz de dados, expertise em limpeza, visualização e modelagem, aplicando árvores de decisão e gerenciamento eficiente usando Git/GitHub.



Habilidade 1

Identificar as habilidades e diferentes funções do Cientista de Dados, avaliando o mercado de trabalho e oportunidades.

Habilidade 2

Preparar dados para modelagem, analisar e identificar fontes de dados e aplicar a metodologia Crisp DM.

Habilidade 3

Compreender a utilização do Pandas para manipular dados, Matplotlib e Seaborn e manipular arrays usando Numpy.

Habilidade 4

Tratar dados ausentes, renomear e categorizar variáveis e integrar dados de diferentes fontes.

FASE 2



Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM

Competência 3

Dominar a manipulação eficiente de dados em Python, aplicando técnicas de análise descritiva, utilizando árvores de regressão e modelos estatísticos. Habilidade em regressão, automação de scripts e desenvolvimento de aplicativos web interativos com Streamlit.



Habilidade 1

Aplicar técnicas de análise descritiva, interpretar medidas de associação, identificar e tratar outliers e avaliar estabilidade de modelos.

Habilidade 2

Compreender, construir e avaliar árvores de regressão, realizar regressão com statsmodels e avaliar a qualidade do modelo de regressão.

Habilidade 3

Interpretar resultados de testes de hipóteses e selecionar modelos com métodos tradicionais e regularização.

Habilidade 4

Aplicar técnicas de scripting, executar scripts Python na linha de comando e utilizar editor de texto para facilitar a visualização de código.

FASE 3



Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM

Competência 4

Proficiência em análise avançada de dados com Pandas em Python, otimizando códigos e personalizando projetos. Desenvolvimento de modelos de classificação multinomial com árvores de decisão e validação cruzada. Aplicar regressão logística para análise e previsão. Criar relatórios interativos no Streamlit, integrando manipulação de arquivos e filtros. Manipular arquivos no Linux e criar documentos profissionais com Markdown e LaTeX no Jupyter Notebook.



Habilidade 1

Otimizar códigos reduzindo variáveis intermediárias. Utilizar for, list comprehension, .apply(), e tqdm para análise.

Habilidade 2

Aplicar classificação multinomial. Implementar cross-validation. Resolver problemas de classificação com dados de sensores.

Habilidade 3

Interpretar e aplicar regressão logística e utilizá-la para prever probabilidades. Avaliar a qualidade de modelos de regressão logística.

Habilidade 4

Utilizar Streamlit para otimizar leitura de arquivos. Manipular funções para gerenciar arquivos.

Habilidade 5

Aplicar comandos Linux para manipular arquivos. Criar documentos com Markdown e LaTeX no Jupyter Notebook.

FASE 4



Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM

Competência 5

Gerenciamento eficiente de bancos de dados (SQL, Python), análise de séries temporais financeiras e técnicas avançadas em modelos ensemble e Random Forest. Expert em algoritmos de boosting, cálculo diferencial e integral, álgebra linear e implementação de Análise de Componentes Principais (PCA). Destaque no desenvolvimento e gestão de aplicações Streamlit, incluindo deploy online com GitHub, Heroku e Render.



Habilidade 1

Executar consultas SQL para visualização e manipulação de dados. Criar funções Python para otimização de códigos em interação com bancos de dados.

Habilidade 2

Aplicar técnicas de análise descritiva em séries temporais. Utilizar pacote Plotly para criar gráficos interativos. Suavização de séries temporais.

Habilidade 3

Aplicar técnicas de ensemble. Otimizar hiperparâmetros em modelos. Identificar e ajustar overfitting em modelos de Random Forest.

Habilidade 4

Aplicar algoritmos de Boosting como AdaBoost e XGBoost. Identificar e otimizar parâmetros em modelos de Boosting.

FASE 5



BY
UIC

Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM

Competência 6

Domínio de técnicas avançadas de agrupamento, desde o domínio do algoritmo K-means até a implementação e interpretação de algoritmos de agrupamento hierárquico e aglomerativo. Destaca-se na utilização de diferentes tipos de ligação e medidas de distância para análise de dados mistos. Expertise na implementação personalizada da técnica de segmentação de clientes RFV no Streamlit para estratégias de marketing e análise do comportamento do cliente.



Habilidade 1

Aplicar o algoritmo K-means em conjuntos de dados variados. Habilidade na avaliação da qualidade dos grupos formados pelo algoritmo, compreendendo profundamente o processo.

Habilidade 2

Utilizar dendrogramas como ferramenta visual para compreender agrupamentos. Distinguir os tipos de ligação em algoritmos de agrupamento. Aplicar medidas de distância para dados mistos.

Habilidade 3

Aplicar a técnica RFV, Implementar o código RFV no Streamlit e personalizá-lo para diferentes ações, com abordagem prática e orientada para resultados.

FASE 6



BY
UIC

Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM

Competência 7

Capacitação em probabilidade, estatística e modelagem, utilizando modelos probabilísticos e funções de densidade de probabilidade em situações práticas. Aplicar técnicas de inferência estatística com Python para análise de dados. Expertise na construção, avaliação e aprimoramento de modelos de regressão avançados, incluindo testes de hipóteses, intervalos de confiança, previsões e correção de limitações como outliers, heterocedasticidade e multicolinearidade.



Habilidade 1

Aplicar modelos probabilísticos em situações práticas.

Habilidade 2

Aplicar inferência estatística na ciência de dados. Análise de dados, teste de hipóteses e avaliação da significância estatística com Python.

Habilidade 3

Aplicar testes e construir intervalos de confiança em parâmetros de regressão. Implementar técnicas para acomodar relações não lineares em modelos lineares.

Habilidade 4

Analisar as limitações e identificar de outliers em modelos de regressão. Corrigir heterocedasticidade e tratar multicolinearidade através de variáveis ortogonais.

FASE 7



BY
UIC

Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM

Competência 8

Habilidades avançadas em análise e modelagem de dados categorizados, com foco na estabilidade ao longo do tempo. Especializar-se na construção de modelos de risco de crédito, destacando-se na seleção de variáveis, regressão logística e avaliação de impacto. Expertise na implementação eficiente de soluções de aprendizado de máquina com Pycaret e técnicas de pré-processamento, aplicadas de maneira prática em contextos específicos.



Habilidade 1

Calcular WOE e IV para as categorias de variáveis. Aplicar análise longitudinal para verificar a estabilidade das variáveis ao longo do tempo.

Habilidade 2

Executar regressão logística e analisar os resultados. Avaliar o impacto de cada variável no modelo e ajustar conforme necessário.

Habilidade 3

Aplicar o pacote Pycaret para treinar e testar algoritmos de aprendizado de máquina.

Habilidade 4

Construir pipelines de processamento de dados para otimizar a sequência de etapas de análise. Realizar transformações de dados para melhorar a eficácia do modelo.

FASE 8



BY
IEC

Profissão: Cientista de Dados

JORNADA DE APRENDIZAGEM

Competência 9

Domínio em SQL para gerenciamento de bancos de dados. Habilidade destacada na manipulação de dados, aplicando filtros, funções e condicionais. Combinar recursos de múltiplas tabelas com subqueries, agregações e visões. Proficiência em técnicas de EDA para análise de conjuntos de dados complexos, com a capacidade de criar projetos impactantes usando Storytelling.



Habilidade 1

Aplicar conceitos básicos de SQL. Criar, configurar e gerenciar serviços de armazenamento de dados em nuvem.

Habilidade 2

Criar e gerenciar tabelas. Dominar comandos, operadores lógicos e funções para selecionar e ordenar dados em SQL.

Habilidade 3

Aplicar filtros para criar condições precisas. Usar agregação para calcular valores e obter informações estatísticas úteis.

Habilidade 4

Aplicar diferentes tipos de junções para combinar dados de múltiplas tabelas em consultas SQL.

Habilidade 5

Escrever subqueries para refinar e filtrar resultados. Fazer agregação por particionamento dos dados para obter resultados precisos

FASE 9



BY
UIC