

# PLANO DE ENSINO

## Programa de Formação em Análise de Dados

Formação completa em Dados com foco em Data Science

- **Aulas:**

1 aula (ao vivo) de 3.5h por semana 13h de conteúdo assíncrono, com 2 workshops em sextas feiras.

Aula das 19h às 22h30 - Terça-Feira

28 semanas - 100 horas de aula síncrono e total estimado 450 horas estimadas total  
O Programa Formação em Análise de Dados é um programa de formação com foco em ciência de dados aplicado em análise. Ao final da formação o profissional estará capacitado a entender as várias frentes de trabalho na área de análise de dados, desenvolver análises exploratórias sobre os dados, exibição e apresentação de informações obtidas pelas análises, além de manutenção de armazenagem desses dados em bancos de dados.

Componentes curriculares

	<b>Onboarding - 1 Semana(3,5 horas)</b>
<b>Bloco 01: Introdução à Área de Dados</b>	<ul style="list-style-type: none"><li>● Introdução à computação</li><li>● Conceitos de dados, informação, conhecimento</li><li>● Introdução a análise de dados</li><li>● Subáreas da área de dados</li><li>● Carreiras na área de dados</li></ul>
	<b>Introdução a programação - Python I - 4 Semanas(14 horas)</b>
	<ul style="list-style-type: none"><li>● IDEs de desenvolvimento</li><li>● Introdução a linguagem Python<ul style="list-style-type: none"><li>○ Operadores</li><li>○ Variáveis</li><li>○ Tipos de Dados</li><li>○ Condicionais</li><li>○ Laços de repetição</li><li>○ Funções</li><li>○ Arquivos</li><li>○ Exceções</li><li>○ Bibliotecas</li></ul></li></ul>
	<b>Programação aplicada - Python II - 6 Semanas (21 horas)</b>
	<ul style="list-style-type: none"><li>● Introdução a manipulação de dados com Python e bibliotecas<ul style="list-style-type: none"><li>○ Numpy<ul style="list-style-type: none"><li>■ Array</li><li>■ Métodos x Atributos</li><li>■ Inicialização de Arrays Numpy</li><li>■ Manipulação das dimensões de um array</li></ul></li></ul></li></ul>

- Slicing em Arrays
- Indexação
- Manipulação de Arrays
  - vstack/hstack
  - concatenate
  - append
- Funções Matemáticas
- Operações entre arrays
- Pandas
  - Series
    - Criação de uma series
    - Operações básicas com series
    - Filtros
    - Outras operações
    - Concatenação de series
  - Dataframes
    - Criação de um dataframe
    - Leitura de dados de csv e excel
    - Atributos de um dataframe
    - Acesso aos dados
    - Ordenação e descrição
    - Transformação de dados
    - Filtros
    - Mudança condicional
    - Export de dataframes
  - Limpeza e Manipulação de dados
    - Investigação e definição de tipos de dados
    - Análise de inconsistências
    - Dados faltantes
    - Análise de duplicatas
    - Remoção de colunas constantes
- Visualização de Dados
  - Matplotlib
  - Seaborn

### **Banco de Dados Relacional - 4 Semanas (14 Horas)**

- Conceitos
  - Bancos de dados.
  - SGBD.
  - Banco de dados relacional
  - Propriedades ACID (Atomicidade, Consistência, Isolamento, Durabilidade).
- Linguagem SQL
  - SQL ANSI
  - DDL
    - Create
    - Alter
    - Drop
  - DML e DQL
    - Select

	<ul style="list-style-type: none"> <li>■ Update</li> <li>■ Insert</li> <li>■ Delete</li> <li>○ Joins</li> <li>○ Comandos avançadas <ul style="list-style-type: none"> <li>■ CASE</li> </ul> </li> <li>○ Views</li> <li>○ Índices</li> </ul>
<b>Bloco 02 - Análise de Dados</b>	<b>Exploração de dados na internet e API - 2 Semanas (7 Horas)</b>
	<ul style="list-style-type: none"> <li>● Conceito, e entendimento de plataformas online <ul style="list-style-type: none"> <li>○ Considerações éticas na coleta e uso de dados online.</li> <li>○ Regulamentações de privacidade.</li> <li>○ Fontes de dados online.</li> <li>○ Coleta e análise de dados de redes sociais.</li> <li>○ Integração com APIs de plataformas sociais.</li> <li>○ JSON.</li> <li>○ XML.</li> <li>○ API <ul style="list-style-type: none"> <li>■ Conceitos básicos.</li> <li>■ Uso de APIs para coleta de dados.</li> <li>■ Exemplos práticos de integração com APIs populares.</li> </ul> </li> </ul> </li> </ul>
	<b>Projeto Prático I - 2 Semanas - (7 horas)</b>
	<p>O projeto 1 é sobre a extração, limpeza e manipulação de dados no INEP sobre o Censo da Educação Superior.</p> <p>Vamos disponibilizar os dados e os alunos precisarão imputar os dados de arquivos csv para um banco de dados sql. Após isso, vão precisar realizar diversas queries de SQL para responder questões ligadas ao negócio</p>
	<b>Workshop I - Uma introdução a IA - Foco em Machine Learning</b>
	<p>Conteúdo extra, para trazer um overview sobre um assunto relevante da área.</p>
	<b>Matemática e Estatística Aplicadas - 4 Semanas (14 horas)</b>
<ul style="list-style-type: none"> <li>● Estatística Descritiva <ul style="list-style-type: none"> <li>○ Tipos de Variáveis</li> <li>○ Medidas "Resumo" <ul style="list-style-type: none"> <li>■ Medidas de Tendência Central</li> <li>■ Medidas de Dispersão</li> <li>■ Quantis, quartis e percentis</li> </ul> </li> <li>○ Análise Bidimensional <ul style="list-style-type: none"> <li>■ Associação entre variáveis qualitativas</li> <li>■ Associação entre variáveis quantitativas</li> <li>■ Correlação <ul style="list-style-type: none"> <li>● Correlação de Spearman</li> <li>● Correlação de Pearson</li> </ul> </li> <li>■ Associação entre variáveis quali e quantitativas <ul style="list-style-type: none"> <li>● Correlação Point-Biserial</li> <li>● R2</li> </ul> </li> </ul> </li> </ul> </li> </ul>	

- Distribuição de Probabilidade
  - Distribuições Discretas
    - Distribuição de Bernoulli
    - Distribuição Binomial
  - Distribuições Contínuas
    - Distribuição Normal
    - Distribuição Normal Padrão
- Estatística Inferencial
  - O que é inferência
  - Amostragem
  - Lei dos Grandes Números
  - Teorema do Limite Central
  - Intervalo de Confiança
  - Teste de Hipótese
    - O que é um teste de hipótese
    - Exemplo simples
    - Metodologia Genérica
    - P-valor
    - Exemplos Práticos
      - Teste para a média
      - Teste de normalidade
- Análise Exploratória de Dados (EDA)
- Regressão Linear

### **Visualização de Dados e Storytelling - 1 Semana ( 3,5 horas)**

- Técnicas de apresentação de dados para análises explanatórias.
- Comunicação: história através dos dados, visualizações e escolha para cada demanda.
- Detalhes e os erros mais comuns na visualização de dados.
- Noção básica sobre acessibilidade na visualização de dados (pessoas daltônicas).

### **Workshop II - Uso de bibliotecas python para plotagem de dash**

Foco em plotly

### **Introdução a visualização de dados com uso de ferramentas de BI - 2 Semanas (14 horas)**

- Princípios de Design de Visualização:
  - Escolha de gráficos apropriados para diferentes tipos de dados.
- Business Intelligence (BI):
  - Definição e objetivos do BI.
  - Como o BI suporta a tomada de decisões nas organizações.
- Power BI: Introdução e Demonstração Prática:
  - Visão geral do Power BI e sua integração com o BI.
  - Demonstração prática: Criação de relatórios simples no Power BI.
  - Conectando e Importando Dados no Power BI:
- Importação de dados de diferentes fontes.
  - Transformação e limpeza de dados no Power BI.
  - Criando Dashboards Interativos:

### **Projeto Prático - 2 Semanas (7 horas)**

Projeto onde o aluno irá usar as mesmas bases de dados do projeto 1, porém aqui o foco vai ser em visualização de dados e geração de insights.

	Parte 1: Construir um dashboard em Power BI para se fazer uma análise do ensino superior no Brasil
--	--

	Parte 2: Usar o python para gerar insights e estatísticas e apresentar (preferencialmente em um ppt) as conclusões/insights para a turma.
--	---