**Módulo: Data Science I – Introdução I**

|  |
| --- |
| Curso: Programação Full Stack |
| Ênfase: Programação |
| Carga Horária: 24 horas |
| Professor responsável: Rafael Puyau |

### Apresentação:

O cientista de dados é o profissional responsável que coleta uma grande quantidade de dados não-estruturados para que deles se possa extrair informações relevantes. É um profissional deve, além de coletá-los, ser capaz gerenciá-los para depois, transformá-los em modelos utilizáveis. Também faz parte da sua função a interpretação dos dados e fazr que eles se comuniquem entre si, multiplicando a possibilidade de informações geradas, que agregam valor, tornando-se um produtos valiosos para uma empresa.

|  |
| --- |
| **Ementa:**  Capacitar o aluno para desempenhar as principais funções de um cientista de dados, para que ele consiga, através da coleta, do armazenamento e do tratamento de dados, gerar informações para a inteligência de negócio. Ele aprenderá como extarir e organizar um volume grande de informações para auxiliar na tomada de decisões. |
| **Objetivos:**  Através da introdução à biblioteca pandas e a visualização com matplotlib e seaborn, permitir que o aluno faça uma análise exploratória básica de dados num dataset sobre vinhos.  Neste módulo será utilizado o Google Colaboratory para a execução de todas as atividades programadas. Como pré-requisito: o aluno deve ter uma conta no Google (gmail).  O grau de complexidade desde módulo é composto por 90% de conteúdos básicos e 10% de conteúdos intermediários. |

|  |
| --- |
| **Conteúdo Programático**:   * Primeiros passos em Data Science e ao Google Colab; * Introdução à Biblioteca Pandas e suas estruturas de dados; * Expressões Lambda; * Principais métodos do Pandas - realizando os primeiros tratamentos; * Agrupamento, Seleção e Filtro; * Respondendo as questões do CEO - extraindo informações de um dataset; * Introdução à visualização com Pandas Plot, Matplotlib e Seaborn; * Visualização - respondendo as questões do CEO com storytelling. |

**Processo de avaliação:**

Será realizado um projeto de conclusão com a análise exploratória de dados no dataset de vinhos com visualização de alguns gráficos.

**Referências:**

(em construção)

**Planejamento do Curso**

|  |
| --- |
| **Aula 01 Data: Horário:** |
| **Docente:** |
| **Conteúdos:**   * Primeiros passos em Data Science e ao Google Colab; * O que é Data Science? * Como rodar e executar um código de Data Science? * Rodando localmente; * Ativando o ambiente virtual; * Desativando o ambiente virtual; * Instalando as bibliotecas; * Onde executar? * Vim / NeoVim; * Emacs; * Jupyter Notebook; * VS Code; * PyCharm; * Rodando na nuvem; * Google Colab; * Principais comandos; * Colocando a mão no código; * Importação da biblioteca; * Buscando ajuda para a função read\_csv; * Lendo o dataset; * Retornando as 5 primeiras linhas do dataset ; * Coletando informações sobre o dataset; * Pegando as colunas; * Alterando o tipo de dado de uma coluna numérica; * Método describe; * Hora de praticar! |

|  |
| --- |
| **Aula 02 Data: Horário:** |
| **Docente:** |
| **Conteúdos:**   * Introdução à Biblioteca Pandas e suas estruturas de dados; * Sobre a biblioteca Pandas; * Trabalhando com Pandas; * Função read\_csv( ); * Primeiros passos; * Mais aplicações do método describe( ); * Principais estruturas do Pandas * Séries; * Valores; * Índices; * Tipo; * Nome; * Dataframes; * Dados; * Índices; * Colunas; * Tipos; * Hora de praticar! |

|  |
| --- |
| **Aula 03 Data: Horário:** |
| **Docente:** |
| **Conteúdos:**   * Expressões Lambda; * Diferenças entre funções padrão e lambdas; * Outros nomes para as expressões lambdas; * Statements não permitidos; * Passagem de parâmetros; * Posicionais; * Nomeados; * Lista de tamanho variável (\*args); * Lista de tamanho variável (\*\*kwargs); * Utilizar expressão lambda é uma boa prática? * O real poder das expressões lambda; * map( ); * map( ) com dicionários; * filter( ); * reduce( ); * Hora de praticar! |

|  |
| --- |
| **Aula 04 Data: Horário:** |
| **Docente:** |
| **Conteúdos:**  Principais métodos do Pandas - realizando os primeiros tratamentos   * Rename( ); * Apply( ); * Map( ); * Agg( ); * Min( ); * Max( ); * Sum( ); * Mean( ); * Count( ); * Value\_counts( ); * Sort\_values( ); * Insert( ); * Drop( ); * Inserir novas colunas no dataframe; * Remover colunas no dataframe; * Formatando valores; * Hora de praticar! |

|  |
| --- |
| **Aula 05** Data: Horário: |
| Docente: |
| **Conteúdos:**   * Agrupamento, Seleção e Filtro; * Groupby( ); * Loc; * Iloc; * Query( ); * Set\_index( ); * Reset\_index( ); * Hora de praticar! |

|  |
| --- |
| **Aula 06 Data: Horário:** |
| **Docente:** |
| **Conteúdos:**   * Respondendo as questões do CEO - extraindo informações de um dataset * Quanto custa o vinho mais caro? * Quanto custa o vinho mais barato? * Qual o vinho mais caro? * Qual o vinho mais barato? * Quanto foi vendido em 2021? * Quantas garrafas no total? * Qual vinho teve a maior nota? * Qual vinho teve a menor nota? * Qual vinho teve a menor nota? * Qual a safra mais antiga? * Qual a safra mais recente? * Qual país teve mais garrafas vendidas? * Qual país teve menos garrafas vendidas? * Quantos vinhos abaixo de R$100,00 foram vendidos? * Quantos vinhos acima de R$100,00 foram vendidos? * Quantas garrafas de espumantes foram vendidas no mês de dezembro? * Quantos vinhos brancos e rosés foram vendidos nos meses de Janeiro à Março? * Qual foi o faturamento do primeiro trimestre de 2021? * Quantas garrafas de tinto foram vendidas no inverno de 2021? * Vale lembrar: Inverno de 21 de junho à 20 de setembro * Quais os 3 meses de maior faturamento? * Qual o ticket médio por mês?   Observação sobre a aula: esta aula é 100% prática - HANDS ON - para que o aluno sedimente todo o conhecimento passado nas aulas anteriores. |

|  |
| --- |
| **Aula 07 Data: xx/xx Horário:** |
| **Docente:** |
| **Conteúdos:**   * Introdução à visualização com Pandas Plot, Matplotlib e Seaborn; * Entendendo a visualização de dados; * Figure; * Plots; * Subplots; * Axes; * Axis; * Conhecendo o método Plot( ); * Histogramas; * Gráfico de Linhas; * Gráfico de Barras; * Mexendo em cada detalhe do seu gráfico com Matplotlib; * Introdução ao Seaborn; * Plotando gráficos mais bonitos; |

|  |
| --- |
| **Aula 08 Data: Horário:** |
| **Docente:** |
| **Conteúdos:**   * Visualização - respondendo as questões do CEO com storytelling * Quanto custa o vinho mais caro? * Quanto custa o vinho mais barato? * Qual o vinho mais caro? * Qual o vinho mais barato? * Quanto foi vendido em 2021? * Quantas garrafas no total? * Qual vinho teve a maior nota? * Qual vinho teve a menor nota? * Qual vinho teve a menor nota? * Qual a safra mais antiga? * Qual a safra mais recente? * Qual país teve mais garrafas vendidas? * Qual país teve menos garrafas vendidas? * Quantos vinhos abaixo de R$100,00 foram vendidos? * Quantos vinhos acima de R$100,00 foram vendidos? * Quantas garrafas de espumantes foram vendidas no mês de dezembro? * Quantos vinhos brancos e rosés foram vendidos nos meses de Janeiro à Março? * Qual foi o faturamento do primeiro trimestre de 2021? * Quantas garrafas de tinto foram vendidas no inverno de 2021? * Vale lembrar: Inverno de 21 de junho à 20 de setembro * Quais os 3 meses de maior faturamento? * Qual o ticket médio por mês?   Observação sobre a aula: esta aula é 100% prática - HANDS ON - para que o aluno sedimente todo o conhecimento passado na aula anterior. |