**Componente:** Análise Exploratória de Dados

**Aula: AULA 7**

**TRILHA DE APRENDIZAGEM**

1. **TÍTULO DA AULA (RECURSO OBRIGATÓRIO)**

|  |
| --- |
| Casos de Análise Exploratória com R e Python |

1. **OBJETIVO DA AULA (RECURSO OBRIGATÓRIO)**

|  |
| --- |
| Nesta aula você entenderá:   * Como criar Análises Exploratórias de Dados com R e Python * Como empregar Seleções e Visualizações dos dados para responder perguntas e verificar hipóteses sobre os dados * O que é e como verificar correlações entre variáveis numéricas e explorar relações entre variáveis categóricas com tabelas de contingência |

1. **ORDEM DOS RECURSOS NA TRILHA**

Na primeira coluna, constam os recursos que você poderá indicar no formulário. Preencha na segunda coluna a ordem que você deseja que os recursos fiquem na trilha de aprendizagem, numerando-os a partir de 1.

Considere como quantidade ideal de recursos na Trilha entre 5 e 6 itens.

|  |  |
| --- | --- |
| **Recursos** | **Ordem** |
| Ponto de Partida | Este recurso não pode ter a ordem alterada |
| Texto de Apoio (obrigatório) | 1 |
| Professor Resolve 1 (opcional) | 2 |
| Praticando 1 (opcional) |  |
| Professor Resolve 2 (opcional) |  |
| Praticando 2 (opcional) |  |
| Momento com o Professor (opcional) | 4 |
| Explorando Conceitos (opcional) | 3 |
| Aplicando Conhecimento | Este recurso não pode ter a ordem alterada |

1. **PONTO DE PARTIDA (RECURSO OBRIGATÓRIO)**

Toda aula iniciará com uma videoaulachamada “Ponto de Partida”.  **Você poderá gravar diversos vídeos ao decorrer da trilha, desde que totalizem 20 minutos por aula**.

Em decorrência da pandemia de Covid-19, você poderá optar por gravar as videoaulas no estúdio, mediante prévio agendamento, ou no formato “vídeo caseiro”, utilizando seus próprios recursos. Indique, a seguir, o formato de gravação:

**☐ Gostaria de gravar a aula no estúdio.**

**X Gostaria de gravar a aula no formato homevídeo.**

Preencha a seguir as informações necessárias sobre a videoaula “Ponto de Partida”:

|  |
| --- |
| **PONTO DE PARTIDA** |
| Duração da videoaula: (Sugestão: de 5 a 8 minutos)  Título da videoaula: Um roteiro de EDA  Descrição da videoaula: Vou apresentar para você a filosofia do EDA. O EDA é mais uma abordagem ou ponto de vista sobre os dados do que propriamente um método. É uma exploração aberta dos dados onde, diferentemente da modelagem estatística, procuramos deixar que os dados revelem a sua estrutura.  Recursos que serão utilizados: Browser com acesso a Internet e Google Colaboratory. |

1. **TEXTO DE APOIO (RECURSO OBRIGATÓRIO)**

Por uma questão de praticidade, este item está posicionado no final deste documento.

1. **PROFESSOR RESOLVE + PRATICANDO (ESTA ATIVIDADE PODE GERAR NOTA DE PARTICIPAÇÃO)**

Neste recurso, o aluno terá acesso a um combinado de videoaula(s) e exercícios para praticar. Você poderá escolher como esses recursos serão apresentados na trilha.

Escolha uma das opções a seguir:

**FORMATO 1 (PROFESSOR RESOLVE + PRATICANDO)** – O aluno assistirá à(s) videoaula(s) e, em seguida, realizará uma atividade. Nesse caso, a atividade deve ser, preferencialmente, de autocorreção. Se não for de autocorreção, você deverá elaborar um gabarito para divulgação.

**FORMATO 2 (PRATICANDO + PROFESSOR RESOLVE)** – Neste caso, o aluno realiza a atividade e, em seguida, assiste à(s) videoaula(s) com a resolução dos exercícios.

**FORMATO 3 (PROFESSOR RESOLVE)** – Neste caso, o aluno assistirá à(s) videoaula(s) e não haverá exercício vinculado.

Quantidade de videoaulas que você pretende gravar **neste recurso**: Escolher um item.

Você poderá optar por gravar as videoaulas no estúdio, mediante prévio agendamento, ou no formato “vídeo caseiro”, utilizando seus próprios recursos. Indique, a seguir, o formato de gravação:

**Gostaria de gravar a aula no estúdio.**

**Gostaria de gravar a aula no formato homevídeo.**

Preencha o quadro abaixo para cada videoaula que você pretende gravar neste recurso.

|  |
| --- |
| **PROFESSOR RESOLVE** |
| Duração da videoaula: 5-8min  Título da videoaula: Explorando Cases em R e Python  Descrição da videoaula: Vamos acompanhar algumas explorações de dados juntos? Nesse vídeo eu vou mostrar para você como trabalhar alguns conjuntos de dados o que pode ser útil para você realizar o seu projeto de EDA.  Recursos que serão utilizados: Browser com acesso a Internet e Google Colaboratory. |

|  |
| --- |
| **PRATICANDO**  **Faça aqui a descrição da atividade** |
| Descrição da atividade: (Preencha aqui a atividade proposta. Caso você opte pelo Formato 1, poderá utilizar exercícios fechados, como questões de múltipla escolha, assim, o aluno receberá o feedback automaticamente. Se a atividade não for de autocorreção, lembre-se de incluir o gabarito com a resolução, o qual será divulgado pelo professor mediador). |

1. **MOMENTO COM O PROFESSOR**

Ao decorrer do componente, os alunos podem ter contato direto com o professor mediador, em três formatos diferentes:

* FORMATO 1 (BBB) – O professor conteudista prevê uma aula “ao vivo” na trilha. Não se trata de plantão de dúvidas. É um encontro com propósito: pode ser a explicação de um conceito, ou a resolução de um exercício, ou uma discussão a partir de um vídeo do YouTube a que todos assistiram. Metade das aulas de seu curso deve ter um encontro neste formato. Você indicará a temática para esse BBB, e o professor mediador será o responsável por planejar e executar essa aula.
* FORMATO 2 (FÓRUM DE DISCUSSÃO) – O professor conteudista apresenta uma questão problematizadora e propõe uma discussão com a turma. Você pode oferecer diversos recursos para essa discussão: um objeto de aprendizagem, um texto, um podcast, um link para site ou vídeo. O professor mediador, então, acompanhará o Fórum ao decorrer da disciplina, incentivando a interação entre os alunos e contribuindo com as postagens.
* FORMATO 3 (FÓRUM DE DÚVIDAS) – O professor conteudista pode reservar um espaço para que os alunos postem suas dúvidas. O professor mediador as responderá, de forma que todos possam ter acesso ao conteúdo.

**Lembre-se de que metade das aulas deve ter uma videoconferência, e você poderá escolher em quais aulas o BBB será realizado.**

Para a aula que está elaborando, escolha uma das opções a seguir:

**Gostaria de aplicar o BBB nesta aula.**

**Gostaria de aplicar um Fórum de Discussão nesta aula.**

**Gostaria de aplicar um Fórum de dúvidas nesta aula.**

**Não vou usar nenhum recurso “Momento com o professor nesta aula”.**

Agora, em função da escolha acima, preencha o quadro correspondente:

Se você optou por propor uma aula via BBB, indique a temática da aula e seu objetivo:

|  |
| --- |
| **MOMENTO COM O PROFESSOR – BBB** |
| Assunto que será tratado na aula: Projeto EDA  Orientações para o mediador: Aproveite esse encontro para discutir o projeto que é apresentado no Aplicando Conhecimento desta aula. Pontos importantes para enfatizar:   1. Trabalho obrigatoriamente em grupo 2-4 alunos para o desenvolvimento de *softskills* de colaboração entre os alunos 2. O projeto pode ser desenvolvido em R ou Python 3. Aproveite para enfatizar as análises obrigatórias no projeto 4. Apresente o modelo de Datafólio que é opcional para o projeto, mas destaque a importância da *comunicação dos dados* das análises nas empresas, negócios, pesquisas acadêmicas etc.   Numere os grupos atribuindo assim o conjunto de dados de cada grupo. A cada nova turma o professor pode fazer uma cópia da atividade e incluir ou excluir novos conjuntos de dados.  Duração: (Indique a duração prevista para a realização desse BBB. Exemplo: 60 minutos) |

Se você optou por propor um fórum de discussão, indique a proposta no bloco a seguir:

|  |
| --- |
| **MOMENTO COM O PROFESSOR – FÓRUM DE DISCUSSÃO** |
| Descrição do fórum: Achados e Perdidos de Recursos de R ou Python para Análise Exploratória de Dados  Aponte aqui um recurso de R ou Python que não apresentamos no curso mas você, nas suas explorações, acha um recurso útil e que pode ser empregado pelos colegas. Não deixe de acrescentar um exemplo e aponte um link de referência para o recurso.  Pontuação: 0.5  Recursos do fórum: - |

Se você optou por propor um fórum de dúvidas, não é necessário adicionar outras informações, pois já temos um formato específico para esse recurso.

1. **EXPLORANDO OS CONCEITOS**

Neste recurso, você terá a liberdade de propor materiais complementares para esta aula. Você poderá indicar animação, links externos, games, simuladores, exercícios resolvidos, podcast, entre outros.

**Gostaria de propor um material complementar nesta aula.**

**Não gostaria de propor um material complementar nesta aula.**

Se você optou por propor um material complementar, indique a proposta no bloco a seguir:

|  |
| --- |
| **EXPLORANDO OS CONCEITOS** |
| Descrição do recurso:  1. Acesse https://www.kaggle.com/ e pesquise por EDA. Você vai encontrar uma série de \*cases\* de exploração de dados que poderão interessar a você.  2. O próximo passo que você daria após essa análise dos dados talvez fosse a preparação dos dados tornando-os mais adequados para criação de modelos estatísticos ou de aprendizado de máquina. Muitas dessas transformações você verá adiante no curso, mas você pode já buscar conhecer algumas dessas técnicas. Para uma rápida introdução sobre isso em `Python` você pode acessar https://www.kaggle.com/learn/data-cleaning.  3. Quer ler mais sobre EDA? Acesse as referências online:  Hadley Wickham, Garrett Grolemund. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. Disponível em: https://r4ds.had.co.nz/.  \_\_\_. Engineering Statistics Handbook. Exploratory Data Analysis. Disponível em: https://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm  E leia mais sobre o processo de Análise Exploratória de Dados. |
|  |

1. **APLICANDO CONHECIMENTO (RECURSO OBRIGATÓRIO)**

Ao final de cada aula, o aluno realizará uma atividade de encerramento. Esta atividade é avaliativa, e você deverá indicar a pontuação dela no quadro de pontuação.

Indique a atividade no quadro a seguir:

|  |
| --- |
| **APLICANDO CONHECIMENTO** |
| Descrição da atividade:  Objetivo  Fazer uma Análise Exploratória de Dados (Exploratory Data Analysis) envolvendo as bases de dados definidas. Trabalho obrigatoriamente em grupo de 2-4 alunos.  Descrição do Projeto  O trabalho deve ser original e desenvolvido em R ou Python notebook com uma Análise e Exploração dos Dados, de um conjunto de dados definido. O R ou Python notebook deve ser 100% executável (empregue os dados a partir dos links originais).  A análise de incluir estatísticas descritivas, seleções, transformações, agrupamentos de dados e exploração gráfica dos dados respondendo ao menos 6 perguntas relevantes no contexto dos dados. Deve ser entregue também uma apresentação em \*\*vídeo (até 10min)\*\* dos resultados. Para a apresentação você pode querer empregar um datafólio, mas este é uma entrega opcional.  Garanta que seu projeto tenha ao menos:  1. Uma análise de distribuição de dados  2. Transformações dos dados envolvendo junção de dados, transformações de atributos e/ou normalização  3. Uma análise tabelas de contingência ou de agregações  4. Uma análise de correlação  5. Uma análise de informação mútua, emprego de atributos derivados ou teste de interdependência  Acesse  <https://github.com/Rogerio-mack/Analise_Exploratoria_de_Dados/blob/main/AED_7_ex_projeto.ipynb>  para mais detalhes da atividade, como as bases de dados elegíveis e os templates.  Pontuação: 2 |

1. **GABARITO – APLICANDO CONHECIMENTO (RECURSO OBRIGATÓRIO)**

Forneça o gabarito aqui ou, se for disponibilizar o gabarito de outra forma, converse com a Designer, para que ela lhe oriente.

|  |
| --- |
| **GABARITO** |
| Disponível em:  Não se aplica. |

**TABELA DE PONTUAÇÃO DA AULA**

Preencha a tabela abaixo com a pontuação aplicada nessa aula e a soma com as demais aulas.

**Lembrete:** a soma de todas as atividades do componente totaliza 10 pontos.

|  |  |
| --- | --- |
| Total de pontos aplicados nessa aula. | 2,5\_\_\_\_ /10,0 pontos |
| Total dos pontos que temos até o momento (soma das aulas anteriores). | 8.5\_\_\_\_ / 10,0 pontos |

1. **TEXTO DE APOIO (RECURSO OBRIGATÓRIO)**

Em todas as aulas, você deverá produzir um texto de apoio que apresente o conteúdo essencial à aula. O texto de apoio deve conter os aspectos mais importantes, como conceitos, ideias, teses, trechos etc., a respeito do assunto. **O texto deve ser de autoria própria, permitindo que o aluno aprofunde o conhecimento e assimile os conteúdos**.

Apresente consistência, ou seja, organize informações importantes ou sinalize caminhos relevantes para a construção do conhecimento.

**Orientações para a elaboração:**

- Para cada aula, você deve desenvolver cerca de **10 páginas (mínimo 8, máximo 12).** Utilize a seguinte formatação: fonte Arial, tamanho 12, espaçamento 1,5 entrelinhas e alinhamento justificado.

- Use sempre a linguagem dialógica, de forma que você passe para o aluno a ideia de que está conversando com ele. Exemplos de expressões: “Você reparou como...”, “Você já estudou XYZ, agora veja...”, “Você deve prestar atenção em XYZ...”, “Percebeu como fica tal coisa...”

- Indique imagens ilustrativas, assim como gráficos e esquemas. O CEDaD detém a assinatura do banco de imagens Getty Images, não gerando, assim, custo ao professor conteudista. Busque as imagens que deseja utilizar e indique o link exato no material: <https://www.gettyimages.com.br/>

- Ao longo do texto, utilize elementos extratextuais para interagir com o aluno. Exemplos: Saiba mais, Curiosidades, Atenção, entre outros. Para identificar esses elementos, utilize a seguinte formatação: (fonte em vermelho e entre parênteses). Exemplo:

*[...]*

*A classe Pilha estenderá a classe Vetor e, portanto, terá acesso aos métodos públicos dessa classe. É muito importante que você retome os conceitos de herança que você já estudou em componentes anteriores!*

- Veja um modelo desse texto no ambiente de entrega de materiais.

Escreva nas páginas a seguir o texto de apoio desta aula:

|  |
| --- |
| TÍTULO DO TEXTO DE APOIO: (Crie um título para este texto)  TEXTO: (Escreva nas páginas a seguir o texto de apoio desta aula.)  Documento postado em word à parte no Moodle. |