

Data Science			
Nanodegree	Machine Learning e Inteligência Artificial	Semestre	2025.1
Dia da semana	Segunda, terça e quarta-feira	Data de entrega	23/04/2025
Professor	M.Sc. Fernando P. Pinheiro	Email	fernando.pinheiro@atitus.edu.br

Objetivo

Esse trabalho, que vale **100% da nota desta etapa da disciplina**, permitirá desenvolver habilidades práticas essenciais para a carreira em ciência de dados, especialmente no que diz respeito a compreender o processo de limpeza e tratamento de dados, identificar tipos de variáveis qualitativas e quantitativas, aplicar técnicas de tratamento e comunicar insights de dados de forma clara e objetiva.

Descrição do Trabalho

Grupo: de 3 até 5 pessoas.

Identificação do grupo: um líder do grupo deve listar os membros da equipe para serem adicionados na planilha de controle.

Cada equipe receberá um dataset referente a uma pesquisa real desenvolvida por alunos do primeiro semestre do curso de Ciência da Computação do nano Data Science. Esses datasets foram construídos por estudantes com baixa ou nenhuma experiência em pesquisa e análise de dados, portanto, possuem diversos problemas comuns a um dataset, tais como:

- Nomes de colunas não padronizados ou pouco descritivos
- Dados duplicados
- Informações faltantes
- Erros de digitação
- Inconsistência de formatos
- Dados desatualizados ou irrelevantes
- Entre outros.

Com os arquivos em mãos, o grupo deverá compreender o conteúdo da pesquisa, seu contexto e objetivo. Utilizando os dados coletados na pesquisa, irá realizar a limpeza do dataset, corrigindo erros, padronizando informações e removendo ou tratando dados inválidos.

Após o tratamento dos dados, deverá classificar os dados, identificando quais variáveis são qualitativas e quais são quantitativas (e suas subcategorias).

Após realizar as mudanças finais no dataset, deverá explicar o processo de limpeza:



Passo Fundo
Rua Senador Pinheiro, 304
Vila Rodrigues - 99070-220



Porto Alegre
Rua Dona Laura, 1020
Mont' Serrat - 90430-090



Ijuí
Treze de Maio, 67
Centro - 98700-000

- Quais dados estavam disponíveis, indisponíveis e duplicados no dataset?
- Quais dificuldades foram encontradas e como foram solucionadas?
- Quais dados foram removidos ou alterados e por que?
- Havia alguma informação irrelevante? Quais?
- Quais métodos e decisões foram utilizados?

Após esta limpeza, deverão realizar um resumo visual analítico dos dados, o que deve incluir gráficos e tabelas, comentando os principais achados (ou não) do projeto.

Apresentação Final

Cada grupo terá **até 10 minutos** para apresentar seu trabalho para a turma.

Critérios de Avaliação:

- Clareza e transparência na explicação do dataset
- Classificação das variáveis
- Precisão e eficácia na limpeza dos dados
- Qualidade da apresentação e capacidade de comunicação do grupo
- Entrega técnica
 - Material de apresentação
 - Dataset original
 - Dataset tratado
 - Arquivo ipynb documentado de forma detalhada com o tratamento realizado e com os gráficos gerados

Entregas

Um integrante do grupo deve submeter o material na atividade do AVA na seção "Avaliações". O grupo deve submeter o arquivo .pdf da apresentação utilizada, o arquivo csv original, o arquivo csv final e o arquivo .ipynb utilizado no tratamento e análise parcial dos dados.

Lembrem-se de deixar a identificação dos integrantes do grupo na apresentação em .pdf em ordem alfabética.



Passo Fundo

Rua Senador Pinheiro, 304
Vila Rodrigues - 99070-220



Porto Alegre

Rua Dona Laura, 1020
Mont' Serrat - 90430-090



Ijuí

Treze de Maio, 67
Centro - 98700-000