

Plano Sessão

Data Science Aplicada Um Guia para Análise, Modelação e ComunicaçãoR

Formador: JOÃO SILVA

Objetivo Geral

Capacitar os formandos para atuar em equipas de ciência de dados, com domínio das ferramentas em R para coleta, análise, modelação e comunicação de dados.

Destinatários

Formandos com conhecimentos básicos em programação e estatística.

Carga Horária

60 horas, distribuídas em 7 módulos.

Introdução

Data Collection and Preparation: Assisting in the collection, cleaning, and preprocessing of large datasets from various sources, ensuring data quality and integrity.

Exploratory Data Analysis (EDA): Performing initial data exploration and visualization to identify patterns, trends, and anomalies, and communicating findings to the team.

Model Development Support: Collaborating with senior colleagues on the development and implementation of machine learning models, including feature engineering, model training, and evaluation.

Reporting and Documentation: Generating reports, dashboards, and presentations to effectively communicate data insights and model performance to technical and non-technical stakeholders.

Research and Learning: Staying up-to-date with the latest advancements in data science, machine learning, and artificial intelligence, and continuously improving your skills.

Ad-hoc Analysis: Conducting ad-hoc data analysis to answer specific business questions and provide data-driven recommendations.

Tooling and Infrastructure: Assisting in the maintenance and optimization of data science tools and infrastructure.

Introdução

Este curso foi concebido para preparar formandos que desempenham funções essenciais no ciclo de vida de projetos de Ciência de Dados, abrangendo desde a recolha e preparação de dados até à implementação e manutenção de modelos preditivos.

Recolha e Preparação de Dados: Apoiar na recolha, limpeza e pré-processamento de grandes conjuntos de dados provenientes de várias fontes, garantindo a qualidade e a integridade dos dados.

Análise Exploratória de Dados (EDA): Realizar a exploração inicial e a visualização de dados para identificar padrões, tendências e anomalias, comunicando as conclusões à equipa.

Apoio ao Desenvolvimento de Modelos: Colaborar com colegas séniores no desenvolvimento e implementação de modelos de aprendizagem automática, incluindo engenharia de variáveis, treino e avaliação de modelos.

Relatórios e Documentação: Produzir relatórios, *dashboards* e apresentações para comunicar de forma eficaz as perceções obtidas a partir dos dados e o desempenho dos modelos, tanto a públicos técnicos como não técnicos.

Investigação e Aprendizagem: Manter-se atualizado relativamente aos mais recentes avanços em ciência de dados, aprendizagem automática e inteligência artificial, melhorando continuamente as suas competências.

Análise *Ad-hoc*: Realizar análises de dados pontuais para responder a questões específicas de negócio e fornecer recomendações baseadas em dados.

Ferramentas e Infraestrutura: Apoiar na manutenção e otimização das ferramentas e da infraestrutura de ciência de dados.

Resumo

Este curso fornece uma visão prática e integrada do trabalho em Ciência de Dados com R. Os estudantes irão:

- Aprender a recolher, limpar e preparar dados de múltiplas fontes, garantindo qualidade e integridade.
- Explorar dados com técnicas de análise exploratória e visualização para identificar padrões e anomalias.
- Apoiar no desenvolvimento de modelos de *machine learning*, desde a engenharia de variáveis até à avaliação.
- Criar relatórios, dashboards e apresentações para comunicar resultados a públicos técnicos e não técnicos.
- Manter-se atualizados com as tendências e avanços na área.
- Realizar análises pontuais para responder a questões específicas de negócio.
- Utilizar e otimizar ferramentas e infraestruturas de suporte à Ciência de Dados.

Módulos de Formação

Módulo 1 – Data Collection and Preparation (10h)

Objetivos:

- Compreender a importância da recolha e preparação de dados no ciclo de vida de um projeto de Data Science.
- Importar dados de diversas fontes, incluindo ficheiros locais (CSV, Excel), APIs REST, e bases de dados relacionais (MySQL, PostgreSQL, SQLite).
- Aplicar técnicas de limpeza e transformação de dados utilizando pacotes do ecossistema *tidyverse* (*dplyr*, *tidyr*) e funções auxiliares do pacote *janitor*.
- Garantir a integridade e qualidade dos dados através de verificação de tipos, deteção de valores ausentes, duplicados e inconsistências.
- Documentar o processo de recolha e preparação para assegurar reprodutibilidade e transparência.

Conteúdos:

- Introdução ao processo de *Data Collection* e *Data Preparation*.
- Leitura de dados:
 - Ficheiros CSV e Excel com `readr` e `readxl`.
 - Acesso a APIs com `httr` e `jsonlite`.
 - Conexão a bases de dados com DBI e RSQLite/RMySQL.
- Limpeza de dados:
 - Normalização de nomes de colunas com `janitor::clean_names()`.
 - Tratamento de valores ausentes (NA) e duplicados.
 - Conversão de tipos de dados (datas, fatores, numéricos).
- Transformação de dados:
 - Filtragem, ordenação e seleção de variáveis.
 - Criação de variáveis derivadas.
 - *Reshaping* de dados com `pivot_longer()` e `pivot_wider()`.
- Boas práticas de organização e documentação do código.

Atividades Práticas:

- Importar e limpar um dataset real obtido do Kaggle ou de uma API pública (por exemplo, dados meteorológicos ou de vendas).
- Criar uma função personalizada em R para automatizar tarefas de limpeza recorrentes (remoção de duplicados, normalização de nomes, tratamento de NA).

- Validar a integridade dos dados através de verificações automáticas (contagem de registos, tipos de variáveis, intervalos de valores).
- Documentar o processo de recolha e preparação num ficheiro **RMarkdown**, incluindo código, explicações e exemplos de antes/depois da limpeza.

Recursos e Ferramentas:

- R e RStudio.
- Pacotes: `tidyverse`, `janitor`, `httr`, `jsonlite`, `DBI`.
- Fontes de dados: Kaggle, APIs públicas (ex.: OpenWeatherMap, World Bank), bases de dados de teste.

Módulo 2 – Exploratory Data Analysis (EDA) (10h)

Objetivos:

- Compreender o papel da Análise Exploratória de Dados no ciclo de vida de um projeto de Data Science.
- Explorar dados utilizando estatísticas descritivas e visualizações gráficas para obter uma compreensão inicial do conjunto de dados.
- Identificar padrões, tendências, outliers e relações entre variáveis.
- Formular hipóteses iniciais com base nas observações obtidas.
- Comunicar de forma clara os principais achados da análise exploratória.

Conteúdos:

- Introdução à Análise Exploratória de Dados (EDA) e sua importância.
- Estatísticas descritivas:
 - Medidas de tendência central: média, mediana, moda.
 - Medidas de dispersão: variância, desvio padrão, amplitude.
 - Distribuições de frequência e percentis.
- Visualização de dados:
 - Gráficos básicos com `ggplot2`: histogramas, boxplots, scatterplots, gráficos de barras.
 - Visualizações interativas com `plotly`.
 - Mapas de calor e matrizes de correlação com `corrplot`.
- Identificação de outliers e valores atípicos:
 - Métodos gráficos (boxplot, scatterplot).
 - Métodos estatísticos (IQR, Z-score).

- Análise de correlação:
 - Correlação de Pearson, Spearman e Kendall.
 - Interpretação de coeficientes de correlação.
- Boas práticas na apresentação de resultados da EDA.

Atividades Práticas:

- Calcular estatísticas descritivas para um dataset real (ex.: vendas, saúde pública, meteorologia).
- Criar visualizações com `ggplot2` para explorar distribuições e relações entre variáveis.
- Utilizar `plotly` para criar gráficos interativos que permitam explorar dados dinamicamente.
- Gerar e interpretar uma matriz de correlação com `corrplot`.
- Identificar e documentar padrões, outliers e correlações relevantes.

Recursos e Ferramentas:

- R e RStudio.
- Pacotes: `ggplot2`, `plotly`, `corrplot`, `dplyr`.
- Datasets de apoio: vendas online, dados de saúde pública, dados meteorológicos.

Módulo 3 – Model Development Support (12h)

Objetivos:

- Compreender o papel do desenvolvimento de modelos no ciclo de vida de um projeto de Data Science.
- Apoiar na criação de modelos preditivos supervisionados e não supervisionados.
- Realizar engenharia de variáveis (*feature engineering*) para melhorar a performance dos modelos.
- Selecionar e aplicar algoritmos adequados ao tipo de problema (classificação, regressão, agrupamento).
- Avaliar modelos utilizando métricas apropriadas e técnicas de validação.
- Documentar e comunicar o processo de modelagem e os resultados obtidos.

Conteúdos:

- Introdução ao fluxo de trabalho de modelagem preditiva.
- Preparação dos dados para modelagem:
 - Divisão em conjuntos de treino, validação e teste.

- Normalização e padronização de variáveis.
- Codificação de variáveis categóricas.
- Engenharia de variáveis:
 - Criação de novas variáveis a partir de dados existentes.
 - Seleção de variáveis relevantes (*feature selection*).
 - Redução de dimensionalidade (PCA).
- Treino de modelos:
 - Utilização do pacote `caret` para treino e avaliação.
 - Fluxo de trabalho com `tidymodels` e `recipes`.
 - Algoritmos comuns: regressão linear, regressão logística, árvores de decisão, *random forest*, k-NN.
- Avaliação de modelos:
 - Métricas para regressão: RMSE, MAE, R^2 .
 - Métricas para classificação: Acurácia, Precisão, Recall, F1-score, AUC.
 - Validação cruzada e *resampling*.
- Comparação e seleção de modelos.

Atividades Práticas:

- Criar um modelo de regressão para prever preços de imóveis utilizando `caret`.
- Desenvolver um modelo de classificação para prever churn de clientes com `tidymodels`.
- Implementar um `recipe` para normalizar dados, criar variáveis derivadas e codificar variáveis categóricas.
- Comparar o desempenho de pelo menos dois algoritmos diferentes para o mesmo problema.
- Documentar o processo de modelagem e apresentar os resultados com métricas e gráficos.

Recursos e Ferramentas:

- R e RStudio.
- Pacotes: `caret`, `tidymodels`, `recipes`, `ggplot2`, `dplyr`.
- Datasets de apoio: preços de imóveis, churn de clientes, datasets públicos do UCI Machine Learning Repository.

Módulo 4 – Reporting and Documentation (8h)

Objetivos:

- Compreender a importância da comunicação clara e estruturada dos resultados em projetos de Data Science.
- Criar relatórios técnicos e executivos que transmitam de forma eficaz as descobertas e conclusões.
- Desenvolver dashboards interativos para visualização e monitorização de métricas e indicadores.
- Adaptar a comunicação para diferentes públicos-alvo (técnico e não técnico).
- Documentar o código, processos e decisões para garantir reprodutibilidade e manutenção futura.

Conteúdos:

- Boas práticas de comunicação de resultados em Data Science.
- Relatórios com **RMarkdown**:
 - Estrutura de um documento técnico.
 - Inclusão de código, tabelas e gráficos.
 - Exportação para HTML, PDF e Word.
- Dashboards interativos:
 - Criação com **flexdashboard**.
 - Aplicações web com **shiny**.
 - Integração de visualizações dinâmicas (**plotly**, **leaflet**).
- Storytelling com dados:
 - Estrutura narrativa para apresentação de insights.
 - Uso de visualizações para reforçar mensagens-chave.
- Documentação de código e processos:
 - Comentários claros e consistentes.
 - Ficheiros **README** e guias de utilização.
 - Versionamento com Git/GitHub.

Atividades Práticas:

- Criar um relatório técnico em **RMarkdown** com análise exploratória e resultados de um modelo preditivo.
- Desenvolver um dashboard interativo com **flexdashboard** ou **shiny** para monitorizar indicadores-chave.

- Preparar uma apresentação executiva com gráficos e conclusões para um público não técnico.
- Documentar todo o processo de análise, incluindo código, decisões e fontes de dados.

Recursos e Ferramentas:

- R e RStudio.
- Pacotes: `rmarkdown`, `flexdashboard`, `shiny`, `plotly`, `leaflet`.
- Ferramentas de versionamento: Git e GitHub.
- Datasets de apoio: conjuntos de dados utilizados nos módulos anteriores.

Módulo 5 – Research and Learning (6h)

Objetivos:

- Desenvolver a capacidade de pesquisa autónoma e contínua em Data Science, Machine Learning e Inteligência Artificial.
- Identificar e avaliar novas bibliotecas, pacotes e técnicas relevantes para projetos em R.
- Manter-se atualizado com as tendências e avanços tecnológicos na área.
- Integrar novos conhecimentos e ferramentas em projetos práticos.
- Fomentar a aprendizagem colaborativa e a partilha de conhecimento dentro da equipa.

Conteúdos:

- Fontes de informação e atualização:
 - Repositórios oficiais (CRAN, Bioconductor).
 - Comunidades e fóruns (*Stack Overflow*, RStudio Community, R-bloggers).
 - Publicações científicas e técnicas (arXiv, IEEE, ACM).
- Avaliação de novas ferramentas:
 - Critérios de seleção (popularidade, manutenção, documentação, compatibilidade).
 - Testes de desempenho e integração.
- Aprendizagem contínua:
 - Cursos online (Coursera, edX, DataCamp).
 - Participação em conferências e *meetups* (useR!, R/Finance, EARL).
 - Leitura de *white papers* e estudos de caso.
- Partilha de conhecimento:

- Apresentações internas.
- Publicação de artigos técnicos ou tutoriais.
- Contribuição para projetos *open source*.

Atividades Práticas:

- Pesquisar e apresentar um pacote R recente, explicando a sua utilidade e aplicabilidade.
- Implementar um exemplo prático com uma técnica ou biblioteca recém-descoberta.
- Criar um breve relatório ou **RMarkdown** com a avaliação de uma nova ferramenta.
- Participar num fórum ou comunidade online, contribuindo com uma resposta ou tutorial.

Recursos e Ferramentas:

- R e RStudio.
- Acesso a internet para pesquisa e participação em comunidades.
- Plataformas de cursos online (Coursera, edX, DataCamp).
- Repositórios de código (GitHub, GitLab).

Módulo 6 – Ad-hoc Analysis (8h)

Objetivos:

- Compreender o conceito e a importância da análise *ad-hoc* para responder a questões específicas de negócio ou investigação.
- Desenvolver capacidade de formular hipóteses e estruturar análises rápidas e direcionadas.
- Utilizar R para realizar consultas, filtragens e cálculos específicos de forma eficiente.
- Produzir resultados claros e acionáveis a partir de análises pontuais.
- Comunicar conclusões de forma objetiva e adaptada ao público-alvo.

Conteúdos:

- Introdução à análise *ad-hoc*:
 - Diferença entre análise exploratória e análise *ad-hoc*.
 - Quando e por que utilizar este tipo de abordagem.
- Formulação de perguntas e hipóteses:
 - Identificação de necessidades de informação.
 - Definição de métricas e indicadores relevantes.

- Técnicas de análise rápida em R:
 - Filtragem e agregação de dados com `dplyr`.
 - Criação de resumos estatísticos direcionados.
 - Visualizações rápidas para suporte à decisão.
- Boas práticas na apresentação de resultados:
 - Clareza e objetividade.
 - Uso de visualizações simples e diretas.
 - Contextualização das conclusões.

Atividades Práticas:

- Receber uma questão de negócio simulada (ex.: “Quais os 5 produtos mais vendidos no último trimestre por região?”) e responder com análise em R.
- Criar um relatório breve com tabelas e gráficos que respondam a uma pergunta específica.
- Utilizar `ggplot2` para criar visualizações rápidas que suportem a resposta.
- Apresentar conclusões e recomendações baseadas na análise realizada.

Recursos e Ferramentas:

- R e RStudio.
- Pacotes: `dplyr`, `ggplot2`, `lubridate`.
- Datasets de apoio: vendas online, dados de saúde pública, dados meteorológicos.

Módulo 7 – Tooling and Infrastructure (6h)

Objetivos:

- Compreender a importância das ferramentas e da infraestrutura no suporte a projetos de Data Science.
- Configurar e manter ambientes de desenvolvimento reprodutíveis e eficientes em R.
- Integrar R com sistemas de controlo de versões e plataformas colaborativas.
- Automatizar tarefas e fluxos de trabalho para aumentar a produtividade.
- Garantir a segurança, organização e escalabilidade dos projetos.

Conteúdos:

- Organização de projetos:
 - Estrutura de pastas e convenções de nomenclatura.
 - Utilização do pacote `here` para caminhos relativos.

- Gestão de dependências e ambientes:
 - Criação de ambientes isolados com **renv**.
 - Documentação de pacotes utilizados.
- Controlo de versões:
 - Introdução ao Git e GitHub.
 - Fluxos de trabalho colaborativos (*branching*, *pull requests*).
- Automatização de tarefas:
 - Pipelines de análise com **targets** ou **drake**.
 - Agendamento de scripts com **cronR** ou tarefas agendadas no sistema operativo.
- Integração e interoperabilidade:
 - Comunicação entre R e outras linguagens (Python via **reticulate**).
 - Conexão com APIs e bases de dados.
- Boas práticas de segurança e backup.

Atividades Práticas:

- Criar a estrutura de um projeto em R com **here** e **renv**.
- Configurar um repositório GitHub e realizar operações básicas (commit, push, pull).
- Implementar um pipeline de análise com **targets** para automatizar a execução de scripts.
- Agendar a execução automática de um script de recolha e limpeza de dados.
- Documentar o ambiente e as dependências do projeto para partilha com outros utilizadores.

Recursos e Ferramentas:

- R e RStudio.
- Pacotes: **here**, **renv**, **targets**, **drake**, **reticulate**, **cronR**.
- Ferramentas de controlo de versões: Git e GitHub.
- Plataformas colaborativas: GitHub, GitLab.

Módulo 8 – Deployment e Manutenção de Modelos (6h)

Objetivos:

- Compreender o processo de disponibilização (*deployment*) de modelos de Machine Learning em ambientes de produção.
- Implementar soluções para servir modelos desenvolvidos em R de forma segura e escalável.
- Monitorizar o desempenho de modelos após a implementação.
- Aplicar técnicas de manutenção e atualização de modelos para garantir a sua relevância e eficácia ao longo do tempo.
- Documentar e comunicar o processo de *deployment* e manutenção a equipas técnicas e não técnicas.

Conteúdos:

- Conceitos de *deployment*:
 - Diferença entre ambiente de desenvolvimento, teste e produção.
 - Ciclo de vida de um modelo em produção.
- Ferramentas e abordagens para *deployment* em R:
 - APIs com **plumber**.
 - Aplicações web com **shiny**.
 - Integração com serviços externos (Docker, cloud computing).
- Monitorização de modelos:
 - Métricas de desempenho em produção.
 - Detecção de *data drift* e *concept drift*.
- Manutenção e atualização:
 - Re-treino periódico.
 - Gestão de versões de modelos.
 - Automação de pipelines de atualização.
- Boas práticas de segurança e escalabilidade.

Atividades Práticas:

- Criar uma API simples com **plumber** para servir previsões de um modelo treinado.
- Desenvolver uma aplicação **shiny** para visualização de resultados e interação com o modelo.
- Configurar um processo de monitorização de métricas de desempenho.
- Simular um cenário de *data drift* e aplicar um re-treino do modelo.

- Documentar o processo de *deployment* e manutenção num relatório técnico.

Recursos e Ferramentas:

- R e RStudio.
- Pacotes: `plumber`, `shiny`, `caret` ou `tidymodels`, `ggplot2`.
- Ferramentas externas: Docker, GitHub Actions, serviços de cloud (AWS, Azure, GCP).
- Datasets de apoio: modelos desenvolvidos nos módulos anteriores.

Módulo 9 – Avaliação (4h)

Objetivos:

- Definir critérios claros e objetivos para avaliar o desempenho dos formandos.
- Aplicar instrumentos de avaliação diagnóstica, formativa e sumativa.
- Garantir que a avaliação reflete as competências técnicas e comportamentais desenvolvidas.
- Fornecer feedback construtivo para promover a melhoria contínua.
- Documentar os resultados da avaliação de forma transparente e organizada.

Conteúdos:

- Tipos de avaliação:
 - Diagnóstica – levantamento de conhecimentos prévios.
 - Formativa – acompanhamento contínuo do progresso.
 - Sumativa – medição final das competências adquiridas.
- Critérios de avaliação:
 - Domínio técnico (execução correta de tarefas, uso adequado de ferramentas).
 - Resolução de problemas e pensamento crítico.
 - Clareza e qualidade da comunicação de resultados.
 - Participação, colaboração e gestão do tempo.
- Instrumentos de avaliação:
 - Fichas de trabalho e exercícios práticos.
 - Projetos individuais e em grupo.
 - Testes teóricos e práticos.
 - Apresentações orais e relatórios escritos.
- Feedback:

- Técnicas de feedback construtivo.
- Autoavaliação e avaliação por pares.

Atividades Práticas:

- Aplicar um questionário diagnóstico no início do curso.
- Avaliar uma ficha prática com base numa grelha de critérios predefinidos.
- Realizar uma sessão de feedback individual com cada formando.
- Conduzir a avaliação final através de um projeto integrador.

Recursos e Ferramentas:

- Grelhas de avaliação e rubricas.
- Plataformas de gestão de aprendizagem (LMS) para submissão e feedback.
- Ferramentas de colaboração online (Google Workspace, Microsoft 365).

Ponderação Sugerida:

Tipo de Avaliação	Peso
Diagnóstica	–
Formativa (fichas, participação, exercícios)	30%
Sumativa (projeto final, teste teórico-prático)	70%

Grelha de Avaliação – Projeto Final

Objetivo: Avaliar o desempenho dos formandos na execução do projeto final, considerando competências técnicas, metodológicas e comunicacionais.

Critério	Descritores de Desempenho	Peso
1. Planeamento e Organização	<ul style="list-style-type: none"> Excelente: Objetivos claros, plano de trabalho estruturado, gestão de tempo eficaz. Bom: Objetivos definidos, plano de trabalho adequado, gestão de tempo satisfatória. Insuficiente: Objetivos pouco claros, ausência de planeamento, má gestão de tempo. 	10%
2. Preparação e Qualidade dos Dados	<ul style="list-style-type: none"> Excelente: Dados limpos, consistentes e bem documentados; aplicação correta de técnicas de preparação. Bom: Dados preparados com pequenas inconsistências; aplicação adequada de técnicas. Insuficiente: Dados com erros significativos; preparação incompleta ou incorreta. 	15%
3. Análise Exploratória (EDA)	<ul style="list-style-type: none"> Excelente: Análise completa, identificação clara de padrões, outliers e correlações; visualizações eficazes. Bom: Análise adequada, identificação de alguns padrões e relações; visualizações satisfatórias. Insuficiente: Análise superficial, ausência de insights relevantes; visualizações pouco claras. 	15%
4. Desenvolvimento do Modelo	<ul style="list-style-type: none"> Excelente: Seleção apropriada do algoritmo, engenharia de variáveis eficaz, treino e validação corretos. Bom: Algoritmo adequado, engenharia de variáveis satisfatória, treino e validação aceitáveis. Insuficiente: Algoritmo inadequado, engenharia de variáveis insuficiente, treino/validação incorretos. 	20%
5. Avaliação e In		15%

Observações do Avaliador:

Grelha de Avaliação – Projeto Final (Escala 1 a 5)

Objetivo: Avaliar o desempenho dos formandos na execução do projeto final, atribuindo pontuações objetivas com base em critérios definidos.

Critério	Descritores de Desempenho	Peso	Pontuação (1-5)
1. Planeamento e Organização	1 = Sem planeamento; 3 = Planeamento básico; 5 = Planeamento completo, objetivos claros e gestão de tempo eficaz.	10%	
2. Preparação e Qualidade dos Dados	1 = Dados com erros graves; 3 = Dados preparados com pequenas falhas; 5 = Dados limpos, consistentes e bem documentados.	15%	
3. Análise Exploratória (EDA)	1 = Análise superficial; 3 = Análise adequada; 5 = Análise completa com insights relevantes e visualizações eficazes.	15%	
4. Desenvolvimento do Modelo	1 = Algoritmo inadequado; 3 = Algoritmo adequado mas otimização limitada; 5 = Algoritmo bem escolhido, otimizado e validado.	20%	
5. Avaliação e Interpretação dos Resultados	1 = Métricas incorretas ou mal interpretadas; 3 = Métricas corretas mas interpretação limitada; 5 = Métricas corretas, interpretação clara e fundamentada.	15%	
6. Comunicação e Apresentação	1 = Apresentação confusa; 3 = Apresentação adequada; 5 = Apresentação clara, estruturada e adaptada ao público.	15%	
7. Documentação e Reprodutibilidade	1 = Código sem comentários e não reprodutível; 3 = Documentação parcial; 5 = Código bem comentado, documentação completa e projeto reprodutível.	10%	
Total		100%	

Escala de Pontuação:

- 1 – Muito Insuficiente
- 2 – Insuficiente
- 3 – Satisfatório
- 4 – Bom
- 5 – Excelente

Cálculo da Nota Final:

$$NotaFinal = \frac{\sum(Pontuação \times Peso)}{SomadosPesos}$$

Observações do Avaliador:

Quiz de Revisão

1. Qual pacote é ideal para limpeza de dados em R?
a) ggplot2 b) janitor c) shiny d) lubridate
Resposta correta: b
2. Qual função é usada para dividir dados em treino e teste?
a) split() b) sample() c) initial_split() d) divide_data()
Resposta correta: c
3. Qual métrica avalia modelos de regressão?
a) AUC b) RMSE c) Accuracy d) Recall
Resposta correta: b
4. Qual ferramenta permite criar dashboards interativos em R?
a) rmarkdown b) shiny c) ggplot2 d) tidyr
Resposta correta: b
5. Qual pacote ajuda a gerir dependências e ambientes em R?
a) dplyr b) renv c) caret d) readr
Resposta correta: b