Curso Ciência de Dados

Universidade Presbiteriana Mackenzie

2° semestre – ano letivo: 2025

CIÊNCIA DE DADOS

Ead

02

**Projeto**

**Aplicado**

CONTRACAPA

CIÊNCIA DE DADOS

Ead

Este Projeto tem como objetivo principal propor ......(preencher)..... é baseado no uso de técnicas e ferramentas desenvolvidas no curso de Ciência de Dados da Universidade Presbiteriana Mackenzie com a orientação do professor Felipe Albino dos Santos.

Sumário

[Apresentação 4](#_Toc207912871)

[Contexto do Estudo 5](#_Toc207912872)

[Referências de Aquisição do Dataset 6](#_Toc207912873)

[Origem dos Dados: 7](#_Toc207912874)

[Limitações de Uso: 7](#_Toc207912875)

[Período da Coleta: 7](#_Toc207912876)

[Descrição da Origem: 8](#_Toc207912877)

[Datasets em vigor: 8](#_Toc207912878)

[Considerações Finais 10](#_Toc207912879)

[Glossário 11](#_Toc207912880)

[Link Repositório ONLINE 11](#_Toc207912881)

[Referências Bibliográficas 12](#_Toc207912882)

[Figuras 12](#_Toc207912883)

[Tabelas 12](#_Toc207912884)



# Apresentação

Neste projeto a FinData Analytics uma empresa de tecnologia e análise de dados do setor financeiro irá auxiliar instituições financeiras, empresas de crédito e grandes redes de varejo nas tomadas de decisões financeiras a fim de reduzir riscos de inadimplência, utilizando ciência de dados para análise preditiva de crédito.

Utilizando dados de relatórios públicos em pdf, do mapa de inadimplência do Serasa, que contém textos explicativos, tabelas e gráficos de sobre a inadimplência no Brasil, e também sobre o comportamento do consumidor.

Dentre os métodos para a extração desses dados serão utilizados técnicas de extração de texto com a biblioteca pdfplumber do python e também visão computacional para extração de informações de imagens. Que futuramente serão convertidos em uma base única para realização de análises exploratórias e preditivas.

O objetivo geral do projeto é desenvolver uma solução capaz de analisar relatórios financeiros, consolidando informações de texto e visuais para gerar insights identificando grupos e/ou regiões com maior confiabilidade para concessão de crédito.

Metas

* Extrair e organizar informações textuais e gráficas de relatórios do Serasa.
* Realizar análise exploratória e tratamento da base consolidada.
* Definir e aplicar métodos de aprendizado de máquina para classificar perfis de risco.
* Medir a acurácia dos modelos aplicados.
* Construir um storytelling com os resultados obtidos, simulando a entrega de valor da empresa fictícia.
* Entregar relatório técnico, código no GitHub e apresentação final em vídeo

**Cronograma de Atividades**

|  |  |  |
| --- | --- | --- |
| **Etapa** | **Atividades Principais** | **Prazo de Entrega** |
| 1. Kick-off | Definir grupo | 11/agosto |
|  | Definir empresa e área de atuação | 04/setembro |
|  | Definir dados, objetivos e cronograma | 05/setembro |
|  | **Entrega da Etapa 1** | 05/setembro |
| 2. Exploração e preparação de dados | Definir bibliotecas Python e repositório GitHub | 10/setembro |
|  | Análise exploratória da base de dados | 17/setembro |
|  | Tratamento e preparação dos dados | 24/setembro |
|  | Definir bases teóricas dos métodos analíticos e cálculo de acurácia | 01/outubro |
|  | **Entrega da Etapa 2** | 03/outubro |
| 3. Desenvolvimento analítico | Aplicar métodos analíticos definidos à base de dados | 08/outubro |
|  | Calcular acurácia e comparar métodos | 12/outubro |
|  | Descrever resultados preliminares e gerar protótipo | 17/outubro |
|  | Definir modelo de negócio e elaborar storytelling inicial | 22/outubro |
|  | **Entrega da Etapa 3** | 24/outubro |
| 4. Conclusão e entrega final | Redação do relatório técnico final | 04/novembro |
|  | Finalização da apresentação storytelling em PPT | 10/novembro |
|  | Organização final do repositório GitHub | 14/novembro |
|  | Gravação e edição do vídeo de apresentação (YouTube) | 18/novembro |
|  | **Entrega da Etapa 4 (Final)** | 21/novembro |

# Contexto do Estudo

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci. Aenean nec lorem. In porttitor. Donec laoreet nonummy augue. Suspendisse dui purus, scelerisque at, vulputate vitae, pretium mattis, nunc. Mauris eget neque at sem venenatis eleifend. Ut nonummy.

Fusce aliquet pede non pede. Suspendisse dapibus lorem pellentesque magna. Integer nulla. Donec blandit feugiat ligula. Donec hendrerit, felis et imperdiet euismod, purus ipsum pretium metus, in lacinia nulla nisl eget sapien. Donec ut est in lectus consequat consequat. Etiam eget dui. Aliquam erat volutpat. Sed at lorem in nunc porta tristique. Proin nec augue. Quisque aliquam tempor magna. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Nunc ac magna. Maecenas odio dolor, vulputate vel, auctor ac, accumsan id, felis. Pellentesque cursus sagittis felis.

# Referências de Aquisição do Dataset

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci. Aenean nec lorem. In porttitor. Donec laoreet nonummy augue.

## Origem dos Dados:

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede.

Mauris et orci. Aenean nec lorem. In porttitor. Donec laoreet nonummy augue. Suspendisse dui purus, scelerisque at, vulputate vitae, pretium mattis, nunc. Mauris eget neque at sem venenatis eleifend. Ut nonummy. Fusce aliquet pede non pede.

## Limitações de Uso:

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci. Aenean nec lorem. In porttitor. Donec laoreet nonummy augue.

## Período da Coleta:

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci. Aenean nec lorem.

## Descrição da Origem:

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci. Aenean nec lorem. In porttitor. Donec laoreet nonummy augue. Suspendisse dui purus, scelerisque at, vulputate vitae, pretium mattis, nunc. Mauris eget neque at sem venenatis eleifend. Ut nonummy.

Fusce aliquet pede non pede. Suspendisse dapibus lorem pellentesque magna. Integer nulla. Donec blandit feugiat ligula. Donec hendrerit, felis et imperdiet euismod, purus ipsum pretium metus, in lacinia nulla nisl eget sapien. Donec ut est in lectus consequat consequat. Etiam eget dui. Aliquam erat volutpat. Sed at lorem in nunc porta tristique. Proin nec augue. Quisque aliquam tempor magna. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Nunc ac magna. Maecenas odio dolor, vulputate vel, auctor ac, accumsan id, felis. Pellentesque cursus sagittis felis.

## Datasets em vigor:

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci. Aenean nec lorem. In porttitor. Donec laoreet nonummy augue. Suspendisse dui purus, scelerisque at, vulputate vitae, pretium mattis, nunc. Mauris eget neque at sem venenatis eleifend. Ut nonummy.

Fusce aliquet pede non pede. Suspendisse dapibus lorem pellentesque magna. Integer nulla. Donec blandit feugiat ligula. Donec hendrerit, felis et imperdiet euismod, purus ipsum pretium metus, in lacinia nulla nisl eget sapien. Donec ut est in lectus consequat consequat. Etiam eget dui. Aliquam erat volutpat. Sed at lorem in nunc porta tristique. Proin nec augue. Quisque aliquam tempor magna. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Nunc ac magna. Maecenas odio dolor, vulputate vel, auctor ac, accumsan id, felis. Pellentesque cursus sagittis felis.

Pellentesque porttitor, velit lacinia egestas auctor, diam eros tempus arcu, nec vulputate augue magna vel risus. Cras non magna vel ante adipiscing rhoncus. Vivamus a mi. Morbi neque. Aliquam erat volutpat. Integer ultrices lobortis eros. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin semper, ante vitae sollicitudin posuere, metus quam iaculis nibh, vitae scelerisque nunc massa eget pede. Sed velit urna, interdum vel, ultricies vel, faucibus at, quam. Donec elit est, consectetuer eget, consequat quis, tempus quis, wisi. In in nunc. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Donec ullamcorper fringilla eros. Fusce in sapien eu purus dapibus commodo. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus.

Cras faucibus condimentum odio. Sed ac ligula. Aliquam at eros. Etiam at ligula et tellus ullamcorper ultrices. In fermentum, lorem non cursus porttitor, diam urna accumsan lacus, sed interdum wisi nibh nec nisl. Ut tincidunt volutpat urna. Mauris eleifend nulla eget mauris. Sed cursus quam id felis. Curabitur posuere quam vel nibh. Cras dapibus dapibus nisl. Vestibulum quis dolor a felis congue vehicula. Maecenas pede purus, tristique ac, tempus eget, egestas quis, mauris. Curabitur non eros. Nullam hendrerit bibendum justo. Fusce iaculis, est quis lacinia pretium, pede metus molestie lacus, at gravida wisi ante at libero.

# Considerações Finais

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci. Aenean nec lorem. In porttitor. Donec laoreet nonummy augue. Suspendisse dui purus, scelerisque at, vulputate vitae, pretium mattis, nunc. Mauris eget neque at sem venenatis eleifend. Ut nonummy.

Fusce aliquet pede non pede. Suspendisse dapibus lorem pellentesque magna. Integer nulla. Donec blandit feugiat ligula. Donec hendrerit, felis et imperdiet euismod, purus ipsum pretium metus, in lacinia nulla nisl eget sapien. Donec ut est in lectus consequat consequat. Etiam eget dui. Aliquam erat volutpat. Sed at lorem in nunc porta tristique. Proin nec augue. Quisque aliquam tempor magna. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Nunc ac magna. Maecenas odio dolor, vulputate vel, auctor ac, accumsan id, felis. Pellentesque cursus sagittis felis.

# Glossário

Granularidade: Nível de detalhe dos dados. Dados com alta granularidade são mais detalhados e específicos.

Dataset: Conjunto de dados estruturado para análise.

Software: Software é uma coleção de programas que dizem a um computador como executar tarefas específicas.

Repositório: Um repositório de software é um local de armazenamento de onde pacotes de software podem ser recuperados e instalados em um computador.

GitHub: Plataforma online para armazenar e compartilhar códigos, documentos e outros arquivos de projetos, facilitando a colaboração e a transparência na pesquisa.

Insights: Compreensões ou revelações importantes obtidas a partir da análise dos dados.

## Link Repositório ONLINE

<https://github.com/GrupoMackenzie/ProjetoAplicado02-DataScience-Mackenzie-2025>

# Referências Bibliográficas

## Figuras

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa.

Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim.

Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede.

Mauris et orci. Aenean nec lorem.

In porttitor. Donec laoreet nonummy augue.

## Tabelas

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa.

Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim.

Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede.

Mauris et orci. Aenean nec lorem.

In porttitor. Donec laoreet nonummy augue.