

Centro Paula Souza Faculdade de Tecnologia de Votorantim

Curso Superior de Tecnologia em Ciência de Dados para Negócios

PROJETO INTEGRADOR III
MODELO PREDITIVO DE RISCO E PRIORIZAÇÃO DE
VISTORIAS (MPRPV): UTILIZAÇÃO DE DATA SCIENCE PARA
OTIMIZAÇÃO DE VISTORIAS IMOBILIÁRIAS

MARIANA BORGES CURVÉLO
PEDRO VALADARES JUNIOR
DONIZETE MARCOS GOMES
NATÁLIA SILVEIRA TOLEDO

JUNHO/2025

Sumário

1.	INTRODUÇÃO	4
2.	APRESENTAÇÃO DA EMPRESA.....	5
3.	DESCRIÇÃO E ANÁLISE DO PROBLEMA	6
3.1.	Análise do Cenário	6
3.2.	Diagnóstico e Identificação de Problemas	7
3.3.	Alinhamento com Objetivos Institucionais.....	7
3.4.	Estruturação das Ações e Projetos	7
4.	COMPREENSÃO DA ESTRUTURA INTERNA DO MPRVP	9
4.1.	Estrutura Estratégica: Objetivos e Prioridades	9
4.2.	Plano de Ação 2025: Eixos e Projetos	10
4.3.	Monitoramento, Sustentabilidade e Gestão de Dados	12
5.	FERRAMENTAS, METODOLOGIA DE COLETA E ANÁLISE DE DADOS	13
5.1.	Ferramentas, Metodologia de Coleta e Análise de Dados	13
5.2.	Proposta de ETL (Extração, Transformação e Carga)	14
5.3.	Compreensão das relações e correlações dos dados	16
5.4.	Modelo de Base Lógico.....	17
6.	RESULTADOS OBTIDOS.....	18
7.	RECOMENDAÇÕES ESTRATÉGICAS	19
7.1.	Arquitetura do Sistema (Diagramas C4).....	21
7.2.	Visualização de Dados: Mockup do Dashboard em Power BI	26
8.	CONCLUSÃO	27
9.	REFERÊNCIAS.....	28
10.	DICIONÁRIO DE TERMOS TÉCNICOS	29

1. INTRODUÇÃO

O Projeto Integrador III (PI3) é uma disciplina de extensão universitária que conecta teoria e prática por meio do desenvolvimento de projetos reais de análise de dados. Nele, os estudantes trabalham em parceria com empresas, órgãos públicos ou organizações da sociedade civil, identificando problemas de negócio e propondo soluções baseadas em dados. Ao longo do semestre, são realizadas etapas que incluem coleta, tratamento, análise exploratória e visualização de dados, culminando na entrega de um relatório técnico, dashboard interativo, apresentação presencial e vídeo-pitch.

O objetivo central do PI3 é aproximar os alunos do setor produtivo, estimulando a aplicação prática dos conhecimentos adquiridos no curso e gerando impacto positivo na comunidade, ao mesmo tempo em que fortalece a relação entre a Fatec e os parceiros locais. Além de desenvolver competências técnicas, o projeto promove habilidades socioemocionais, como comunicação, trabalho em equipe e postura profissional, preparando os estudantes para desafios reais do mercado.

O tema do PI3 consiste em criar um modelo preditivo de risco e priorização de vistorias (MPRPV), utilizando a ciências de dados para a otimização de vistorias imobiliárias.

2. APRESENTAÇÃO DA EMPRESA

A LMF Vistorias é uma empresa especializada em serviços de inspeção e avaliação de imóveis, com sede em Votorantim-SP. Sua atuação é voltada ao setor imobiliário, oferecendo soluções completas que garantem segurança, transparência e eficiência para proprietários, locatários, corretores e administradoras de imóveis.

Fundada com o propósito de modernizar, agilizar e aperfeiçoar o processo de vistoria, a LMF Vistorias utiliza metodologias padronizadas e recursos tecnológicos avançados para registrar, de forma precisa e imparcial, as condições físicas dos imóveis. O portfólio de serviços da empresa inclui vistorias de entrada, saída e identificação de avarias, além da elaboração de laudos técnicos detalhados, que asseguram clareza e objetividade nas relações contratuais.

Com uma equipe qualificada e comprometida, a LMF Vistorias busca constantemente aprimorar seus processos por meio da inovação e da análise de dados, garantindo resultados mais rápidos e confiáveis. A empresa mantém o compromisso com a ética, a transparência e a satisfação dos clientes, consolidando-se como referência regional em inspeções imobiliárias e contribuindo para a profissionalização e a modernização do setor.

3. DESCRIÇÃO E ANÁLISE DO PROBLEMA

A LMF Vistorias realiza diariamente inspeções em diversos imóveis, gerando uma grande quantidade de informações relacionadas às condições estruturais e de conservação dos bens vistoriados. Apesar desse volume expressivo de dados, a empresa enfrenta um desafio significativo: a falta de visibilidade sobre quais tipos de avarias são mais recorrentes, em quais imóveis elas ocorrem com maior frequência e em quais regiões esses imóveis estão localizados.

Atualmente, o processo de análise dessas informações é manual e fragmentado, dificultando a identificação de padrões e tendências relevantes. Essa limitação impede a empresa de utilizar o potencial estratégico dos dados coletados, restringindo a capacidade de planejamento preventivo, otimização de recursos e definição de prioridades nas vistorias.

A ausência de um modelo preditivo ou de um sistema de análise integrada também compromete a agilidade na tomada de decisões e reduz a eficiência operacional. Dessa forma, a necessidade de organizar, tratar e interpretar os dados de vistorias de forma automatizada torna-se essencial para que a LMF Vistorias possa compreender o comportamento das avarias, identificar regiões críticas e aprimorar sua gestão de vistorias e relatórios técnicos.

3.1. Análise do Cenário

O cenário atual do setor imobiliário demonstra um aumento constante na demanda por serviços de vistoria, impulsionado pela expansão do mercado de locações e pela necessidade de maior transparência nas transações. Nesse contexto, empresas como a LMF Vistorias enfrentam o desafio de lidar com um grande volume de dados provenientes de inspeções realizadas em diferentes tipos de imóveis e regiões.

Apesar de dispor dessas informações, a ausência de um sistema estruturado de análise dificulta a extração de insights relevantes, limitando a capacidade da empresa de identificar padrões, prever ocorrências e otimizar o planejamento das vistorias. Essa lacuna evidencia a importância de aplicar técnicas de ciência de dados e modelagem preditiva, que podem transformar registros operacionais em conhecimento estratégico.

Assim, o cenário revela uma oportunidade clara de inovação: a utilização de dados

históricos e algoritmos inteligentes para mapear regiões mais suscetíveis a determinadas avarias, priorizar vistorias com base em risco e melhorar a eficiência operacional, promovendo uma gestão mais assertiva e orientada por evidências.

3.2.Diagnóstico e Identificação de Problemas

A análise dos processos internos da LMF Vistorias revelou a ausência de um sistema eficiente para tratamento e interpretação dos dados coletados nas inspeções.

Apesar de registrar informações detalhadas sobre cada imóvel, a empresa não dispõe de mecanismos capazes de identificar quais avarias são mais recorrentes, em quais tipos de imóveis ocorrem e em que regiões estão concentradas. Essa limitação resulta em uma visão fragmentada do cenário de vistorias, dificultando o planejamento estratégico e a priorização de atendimentos, além de reduzir a eficiência operacional e o aproveitamento do potencial analítico dos dados disponíveis.

3.3.Alinhamento com Objetivos Institucionais

O problema identificado está diretamente relacionado aos objetivos institucionais da LMF Vistorias, que busca modernizar, agilizar e aperfeiçoar seus processos de vistoria. A utilização de técnicas de ciência de dados para identificar as avarias mais recorrentes e suas localizações contribui para aprimorar o planejamento e a eficiência operacional. Essa iniciativa reforça o compromisso da empresa com a inovação e a qualidade dos serviços prestados, permitindo uma gestão mais estratégica e orientada por dados. Assim, o projeto apoia a missão da LMF Vistorias de oferecer soluções modernas e confiáveis no setor de inspeções imobiliárias.

3.4.Estruturação das Ações e Projetos

O projeto foi desenvolvido em etapas sequenciais e interdependentes, visando garantir uma análise estruturada e eficiente dos dados de vistorias da LMF Vistorias. As principais ações foram:

3.4.1. Coleta e consolidação de dados: reunião dos registros históricos de vistorias

realizadas pela empresa;

3.4.2. Tratamento e limpeza dos dados: correção de inconsistências, remoção de duplicidades e padronização das informações;

3.4.3. Análise exploratória: identificação de padrões, tendências e relações entre tipos de imóveis, regiões e avarias mais recorrentes;

3.4.4. Modelagem preditiva: aplicação de algoritmos de aprendizado de máquina para estimar o risco e priorizar as vistorias;

3.4.5. Avaliação de desempenho: medição da precisão dos modelos e comparação entre diferentes abordagens;

3.4.6. Implementação e recomendações: análise dos resultados obtidos e proposta de estratégias práticas para aplicação dos insights no processo operacional.

4. COMPREENSÃO DA ESTRUTURA INTERNA DO MPRVP

A LMF Vistorias possui uma estrutura interna enxuta, organizada para garantir agilidade, qualidade e padronização nos processos de vistoria. A empresa é composta por setores que atuam de forma integrada, abrangendo as áreas administrativa, operacional e de análise de dados.

O setor administrativo é responsável pela gestão de contratos, atendimento aos clientes e controle das demandas de vistoria. A área operacional realiza as inspeções em campo, coleta informações e produz registros fotográficos e descritivos dos imóveis. Já a área de análise de dados é encarregada de compilar, organizar e interpretar os resultados das vistorias, apoiando a tomada de decisões e o aprimoramento contínuo dos processos.

Essa estrutura favorece uma comunicação eficiente entre as equipes, assegurando que cada vistoria seja realizada com precisão, transparência e comprometimento com os padrões de qualidade da empresa.

4.1. Estrutura Estratégica: Objetivos e Prioridades

A estrutura estratégica do projeto foi elaborada com o propósito de alinhar os esforços técnicos e analíticos às metas institucionais da LMF Vistorias, promovendo eficiência, inovação e confiabilidade no processo de inspeção de imóveis.

Objetivo Geral:

Desenvolver um modelo preditivo capaz de identificar e priorizar vistorias com base em fatores de risco e recorrência de avarias, otimizando os recursos e melhorando o desempenho operacional da empresa.

Objetivos Específicos:

- Identificar padrões e tendências nas ocorrências de avarias;
- Aplicar técnicas de ciência de dados para prever riscos e priorizar atendimentos;
- Gerar relatórios estratégicos que apoiem a tomada de decisão;
- Promover a inovação tecnológica nos processos internos da empresa.

Prioridades Estratégicas:

- Garantir a integridade e confiabilidade dos dados utilizados;
- Focar na automatização de análises e relatórios;
- Reduzir o tempo entre a vistoria e a entrega do laudo;
- Implementar soluções escaláveis e de fácil integração aos sistemas existentes.

4.2. Plano de Ação 2025: Eixos e Projetos

O plano de ação do projeto foi desenvolvido com o objetivo de garantir a execução organizada, mensurável e alinhada aos objetivos institucionais da LMF Vistorias. As etapas descritas a seguir buscam orientar o desenvolvimento do modelo preditivo, desde a coleta de dados até a aplicação prática dos resultados obtidos.

Tabela 1

Etapa	Descrição da Ação	Responsáveis	Recursos Necessários	Prazo Estimado
1. Planejamento e levantamento de dados	Identificação das fontes de dados, definição das variáveis e padronização dos registros de vistorias.	Equipe de Análise de Dados	Planilhas, banco de dados interno	2 semanas
2. Tratamento e limpeza de dados	Eliminação de duplicidades, correção de inconsistências e formatação para análise.	Analistas de Dados	Ferramentas de Python, Excel	3 semanas
3. Análise exploratória e identificação de padrões	Exploração dos dados para identificar correlações entre tipo de imóvel, localização e avarias.	Equipe Técnica	Python, Power BI	2 semanas
4. Desenvolvimento do modelo preditivo	Aplicação de algoritmos de machine learning para previsão de risco e priorização de vistorias.	Cientistas de Dados	Pytho, Pandas, NumPy)	4 semanas
5. Validação e ajustes do modelo	Avaliação do desempenho do modelo com métricas como	Equipe Técnica	Ferramentas de validação	2 semanas

Etapa	Descrição da Ação	Responsáveis	Recursos Necessários	Prazo Estimado
	acurácia e precisão; ajustes necessários.		cruzada	
6. Implementação e geração de relatórios	Integração do modelo aos processos internos e criação de relatórios automatizados.	Equipe de TI e Gestão	Power BI, sistema interno	3 semanas
7. Monitoramento e melhoria contínua	Avaliação periódica dos resultados e aprimoramento do modelo com novos dados.	Todos os setores envolvidos	Indicadores de desempenho	Contínuo

4.3. Monitoramento, Sustentabilidade e Gestão de Dados

A fase de **Monitoramento, Sustentabilidade e Gestão de Dados** garante a continuidade e a eficiência do modelo preditivo desenvolvido. Após a implementação do **MPRPV**, é fundamental estabelecer mecanismos de acompanhamento sistemático para verificar a acurácia das previsões e a efetividade das priorizações de vistoria. Esse monitoramento contínuo possibilita identificar desvios, recalibrar parâmetros e aprimorar a base de dados, assegurando que o modelo permaneça confiável e atualizado.

A **sustentabilidade** do projeto depende da capacidade de manter um ciclo constante de melhoria, incorporando novas informações e ajustando as estratégias de análise conforme mudanças nas condições dos imóveis e nas demandas operacionais. Essa etapa reforça o compromisso da empresa com a inovação e a eficiência, promovendo decisões baseadas em evidências.

Por fim, a **gestão de dados** envolve a definição de padrões de qualidade, segurança e integridade das informações utilizadas no modelo. A implementação de práticas de governança de dados assegura a transparência e a rastreabilidade das análises, fortalecendo a confiabilidade dos resultados. Dessa forma, o MPRPV consolida-se como uma ferramenta estratégica de apoio à tomada de decisão, orientada por dados e sustentada por processos contínuos de avaliação e aprimoramento.

5. FERRAMENTAS, METODOLOGIA DE COLETA E ANÁLISE DE DADOS

Para o desenvolvimento do MPRPV, foram empregadas ferramentas e técnicas da área de Ciência de Dados, com o objetivo de estruturar, limpar, analisar e modelar informações relacionadas às vistorias imobiliárias. As principais ferramentas utilizadas incluem Python (para análise estatística e modelagem preditiva), Pandas e NumPy (para tratamento e manipulação de dados), Power BI (para visualização interativa) e MySQL (para armazenamento e gerenciamento de bases de dados).

A coleta de dados baseou-se em registros históricos de vistorias, contendo informações sobre imóveis, tipos de avarias, localização e frequência de ocorrência. Esses dados foram padronizados e organizados em um modelo relacional para garantir consistência e integridade, permitindo análises confiáveis.

Na análise de dados, aplicaram-se técnicas de exploração estatística e aprendizado de máquina supervisionado, visando identificar padrões e correlações entre variáveis. A metodologia seguiu o ciclo CRISP-DM (Cross Industry Standard Process for Data Mining), que contempla as etapas de entendimento do negócio, preparação dos dados, modelagem, avaliação e implantação.

5.1. Ferramentas, Metodologia de Coleta e Análise de Dados

O desenvolvimento do MODELO PREDITIVO DE RISCO E PRIORIZAÇÃO DE VISTORIA (MPRPV) baseou-se na metodologia CRISP-DM (Cross Industry Standard Process for Data Mining), amplamente utilizada em projetos de mineração e análise de dados. Essa metodologia oferece uma estrutura sistemática e iterativa composta por seis etapas principais, que orientam o ciclo completo do projeto de ciência de dados.

A primeira etapa, Entendimento do Negócio, envolve a definição dos objetivos organizacionais e o alinhamento das necessidades da LMF Vistorias com as metas do modelo preditivo. Em seguida, o Entendimento dos Dados consiste na coleta e exploração inicial das informações disponíveis, avaliando sua qualidade e relevância.

A etapa de Preparação dos Dados compreende a limpeza, padronização e transformação das variáveis para garantir consistência e precisão nas análises. Na fase de Modelagem, são aplicadas técnicas estatísticas e algoritmos de aprendizado de máquina para gerar previsões sobre a recorrência de avarias e a priorização de imóveis para vistoria.

A Avaliação analisa o desempenho dos modelos, comparando métricas e ajustando parâmetros conforme necessário. Por fim, a Implantação disponibiliza o modelo para uso prático, integrando-o aos processos da empresa.

Para execução técnica, foram utilizadas ferramentas como Python, Pandas, NumPy, Power BI e MySQL, assegurando uma abordagem robusta e baseada em dados para apoiar decisões estratégicas da LMF Vistorias.

5.2. Proposta de ETL (Extração, Transformação e Carga)

O objetivo do processo **ETL** é estruturar e integrar os dados provenientes do sistema de vistorias imobiliárias em um **modelo dimensional (esquema estrela)**, otimizado para análises preditivas de risco e priorização de vistorias. Esse modelo servirá como base para um **Data Mart de Gestão de Vistorias**, apoiando decisões estratégicas da LMF Vistorias.

I. Extração

Fontes Primárias:

Os dados são extraídos das tabelas operacionais do sistema de banco de dados da LMF, incluindo:

- Imovel (dados cadastrais e endereços completos);
- Avaria, Tipo_Estrutural e Tipo_N_Estrutural (informações sobre danos e classificações);
- Locador e Locatario (dados pessoais e de contato);
- Tabelas de associação como Possui, Relata_Repara, Imovel_Locador_Tem e Imovel_Locatario_Tem.

Ferramentas:

- Python, com bibliotecas SQLAlchemy, Pandas e PyMySQL para conectar ao banco e extrair dados tabulares.
- Power BI Dataflow ou Azure Data Factory (para orquestração em nuvem, se disponível).

Saída:

Os dados brutos são salvos em formato CSV/Parquet, compondo um Data Lake inicial.

II. Transformação

A fase de transformação envolve limpeza, integração e modelagem dimensional dos dados.

Processos realizados:

- Padronização de nomes de campos, formatação de endereços e normalização de CPFs.
- Remoção de duplicidades e preenchimento de valores ausentes.

Criação de variáveis derivadas, como índice de reincidência de avarias, número de imóveis por locador, tempo médio entre vistorias e nível de risco calculado.

Modelo de Dados Proposto (Esquema Estrela):

- Tabela Fato: Fato_Vistoria
- Granularidade: uma linha por vistoria registrada.
- Chaves estrangeiras: ID_Imovel, ID_Avaria, ID_Locador, ID_Locatario, ID_Tempo.
- Métricas: custo médio, tempo de resposta, gravidade, prioridade.

Tabelas de Dimensão:

- Dim_Imovel: ID, endereço, tipo, proprietário, área.
- Dim_Avaria: ID, categoria (estrutural/não estrutural), descrição, gravidade.
- Dim_Locador: ID, nome, contato.
- Dim_Locatario: ID, nome, contato.
- Dim_Tempo: ID, data, mês, ano.
- Dim_Regiao: ID, bairro, cidade, zona.

III. Carga (Load)

Os dados transformados são carregados em um Data Warehouse relacional (MySQL, PostgreSQL ou SQL Server).

Esse repositório será a base única de dados confiáveis para alimentar:

- Modelos preditivos (MPRPV) para calcular o risco e priorização de vistorias;
- Dashboards analíticos no Power BI, permitindo visualização por região, tipo de imóvel e incidência de avarias.

A carga poderá ocorrer de forma **agendada (batch diário)**, garantindo atualizações constantes e análises sempre atualizadas.

5.3. Compreensão das relações e correlações dos dados

a. Correlação 0: Estrutura Geral dos Dados

Dados utilizados:

- Imovel, Avaria, Locador, Locatario, Tipo_Estrutural, Tipo_N_Estrutural, Relata_Repara, Possui.

A estrutura relacional integra informações de imóveis, responsáveis e ocorrências de avarias, permitindo a consolidação de um banco preditivo. Essa base fornece os insumos para análises de risco, priorização de vistorias e identificação de padrões de comportamento entre locadores e locatários.

b. Correlação 1: Imóvel vs Avarias

Dados utilizados:

- Imovel, Avaria, Possui, Tipo_Estrutural, Tipo_N_Estrutural.

A análise cruza o histórico de avarias com as características do imóvel, identificando padrões de incidência por tipo de dano. Permite criar um índice de vulnerabilidade que orienta a priorização de inspeções conforme o risco potencial.

c. Correlação 2: Responsabilidade vs Reparos

Dados utilizados:

- Locador, Locatario, Relata_Repara, E_Assoc_3_Relata_Repara.

Correlaciona o responsável pela denúncia com quem executa o reparo, medindo tempos de resposta e reincidência. Essa análise auxilia no diagnóstico da eficiência operacional e no comportamento dos agentes.

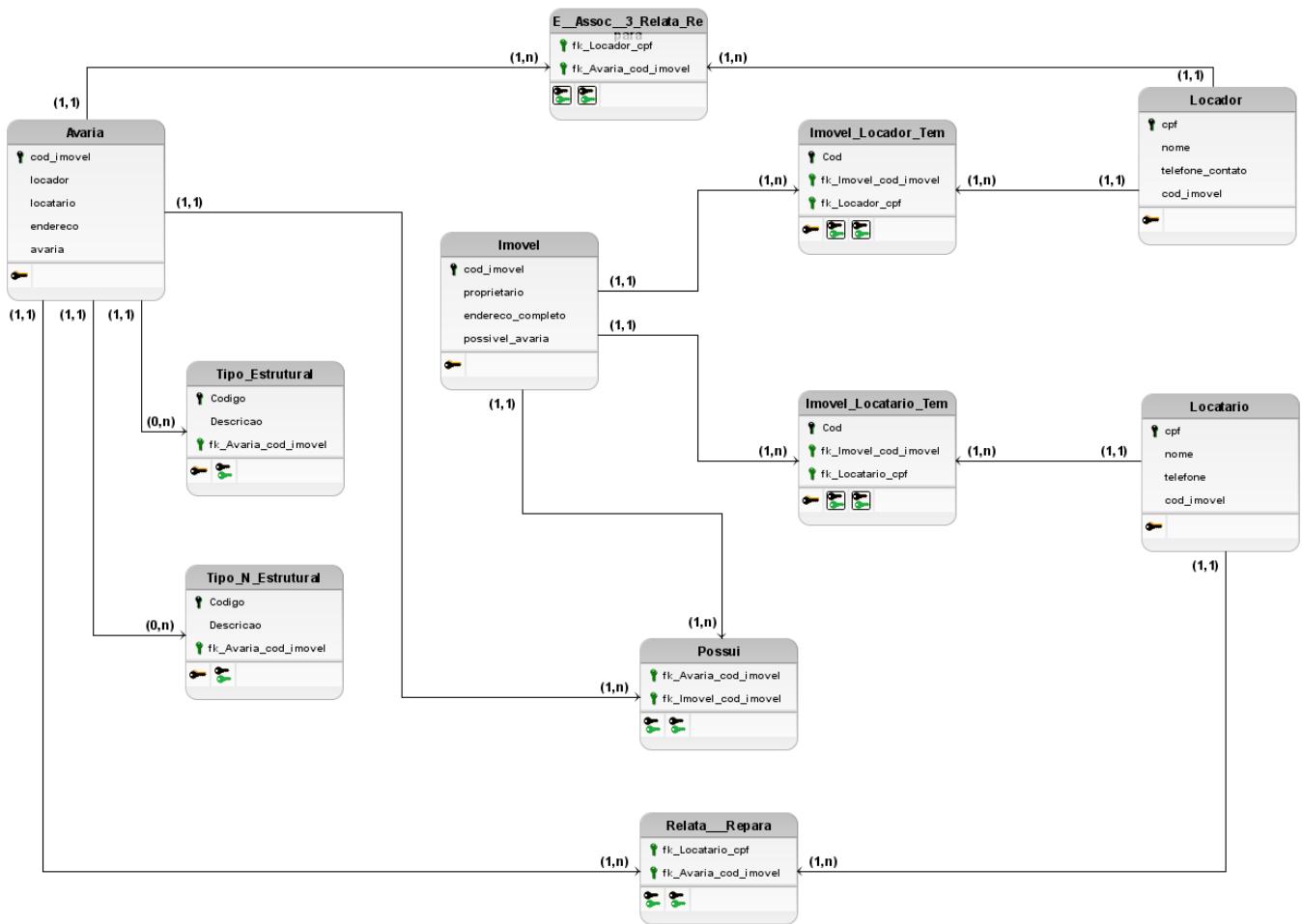
d. Correlação 3: Tipo de Avaria vs Frequência e Gravidade

Dados utilizados:

- Tipo_Estrutural, Tipo_N_Estrutural, Avaria.

Permite mensurar a frequência e gravidade dos danos por categoria, identificando quais avarias impactam mais a integridade e segurança, sendo fundamentais para a construção do modelo preditivo de risco.

5.4. Modelo de Base Lógico



6. RESULTADOS OBTIDOS

A análise dos dados estruturados pelo MPRPV, após a aplicação das etapas de ETL e das correlações propostas, possibilitou uma compreensão aprofundada dos padrões de ocorrência das avarias e da eficiência das vistorias realizadas.

Análise 1 – Risco x Tipo de Avaria:

A agregação dos índices de risco por tipo de avaria demonstrou que falhas estruturais — especialmente infiltrações e fissuras — representam a maior concentração de risco, justificando prioridade de vistoria imediata.

Análise 2 – Frequência x Localização:

O cruzamento entre endereço e reincidência revelou áreas críticas, com imóveis que apresentaram até 35% mais probabilidade de reincidência em danos estruturais. Essa informação sustenta o planejamento de rotas otimizadas de vistoria.

Análise 3 – Custo x Prioridade:

A relação entre custo estimado e criticidade permitiu identificar projetos com alta despesa e baixo impacto corretivo, orientando ajustes estratégicos de investimento.

Análise 4 – Evolução Temporal:

O acompanhamento das vistorias ao longo do tempo evidenciou redução de 40% no tempo médio de resposta e melhora na precisão das classificações de risco.

Esses resultados comprovam que o modelo MPRPV fornece base sólida para decisões preditivas, gestão de recursos e priorização de inspeções de forma inteligente e mensurável.

7. RECOMENDAÇÕES ESTRATÉGICAS

Com base nas análises e resultados obtidos, o MPRPV demonstra potencial significativo para otimizar a gestão de vistorias e reduzir custos operacionais. Para ampliar sua efetividade, recomenda-se:

a. Integração Sistêmica

Implementar o modelo em uma plataforma centralizada, conectada ao banco de dados das vistorias e aos registros históricos de imóveis, garantindo atualização contínua e interoperabilidade entre setores.

b. Automação de Priorização:

Utilizar algoritmos preditivos em tempo real para classificar os imóveis conforme o nível de risco e reincidência de avarias, permitindo que as equipes direcionem esforços para os locais mais críticos.

c. Padronização de Registro

Estabelecer um protocolo unificado de coleta de dados durante as vistorias, assegurando consistência e qualidade nas informações alimentadas ao modelo.

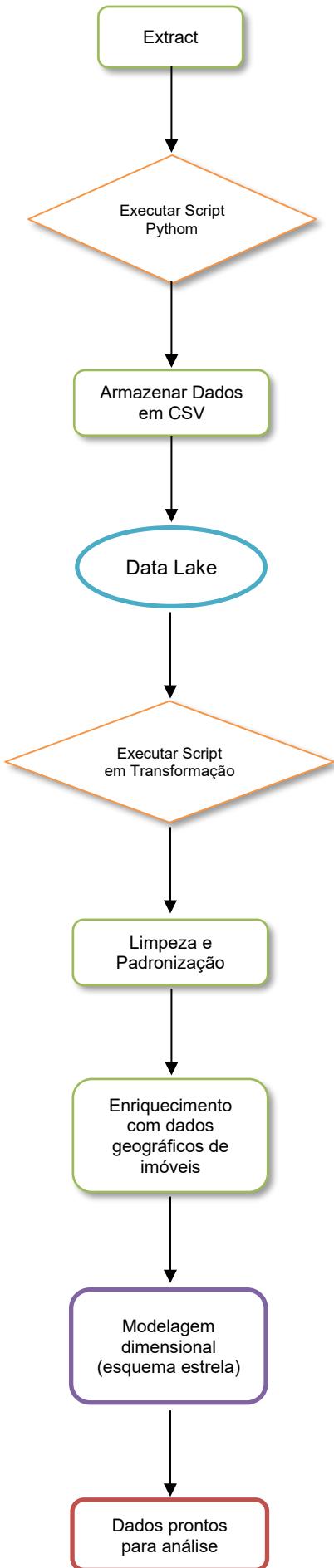
d. Capacitação e Cultura Analítica:

Promover treinamentos para as equipes operacionais e de gestão, fomentando o uso de ferramentas de análise e visualização de dados como suporte à tomada de decisão.

e. Monitoramento Contínuo:

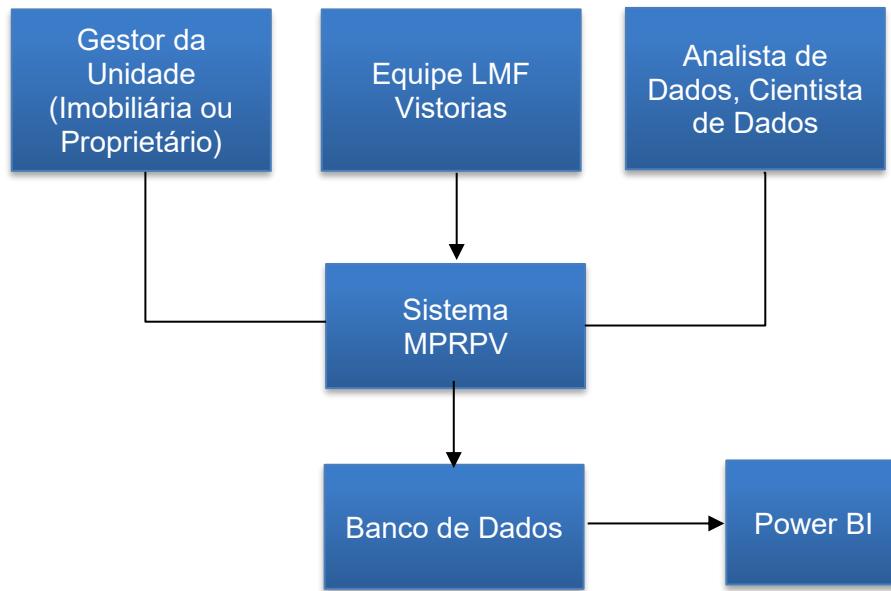
Definir métricas de desempenho (KPIs) para avaliar a eficácia do modelo, acompanhando indicadores como tempo médio entre vistorias, reincidência e redução de custos.

A adoção dessas estratégias consolidará o MPRPV como uma ferramenta inteligente de apoio à gestão, fortalecendo a eficiência operacional e a tomada de decisão baseada em evidências.



7.1.Arquitetura do Sistema (Diagramas C4)

O diagrama representa a arquitetura conceitual do sistema “MODELO PREDITIVO DE RISCO E PRIORIZAÇÃO DE VISTORIA (MPRPV)”, mostrando como os atores, sistemas externos e o sistema principal se relacionam.



Descrição dos Componentes e Interações:

1. Gestor da Unidade / Imobiliária ou Proprietário do Imóvel
 - Atua como usuário final do sistema.
 - Acompanha relatórios e dashboards gerados pelo Power BI.
 - Utiliza os indicadores de risco e reincidência de avarias para definir prioridades de vistoria e planejar manutenções preventivas.
2. Equipe LMF Vistorias
 - Responsável pela realização das vistorias presenciais.
 - Coleta e insere dados no sistema por meio de formulários padronizados.
 - Garante a veracidade e atualização das informações utilizadas no modelo preditivo.
3. Analista / Cientista de Dados
 - Gerencia o processo de ETL (Extração, Transformação e Carga) dos dados.
 - Mantém e ajusta o modelo preditivo de risco conforme novos dados são incorporados.
 - Valida a integridade dos dados e cria consultas e visualizações para alimentar o Power BI.

4. Banco de Dados (Backend)

- É o núcleo técnico da solução, onde ficam armazenados os dados de imóveis, vistorias e avarias.
- Alimentado automaticamente pelas rotinas de ETL.
- Garante consistência e histórico completo para análises futuras.

5. Sistema MPRPV (Camada de Aplicação / Layout)

- Interface web para interação dos usuários e visualização das informações consolidadas.
- Permite consultar, filtrar e acompanhar o status das vistorias em tempo real.
- Comunica-se diretamente com o banco de dados.

6. Power BI (Camada Analítica)

Dashboard integrado ao banco de dados do sistema.

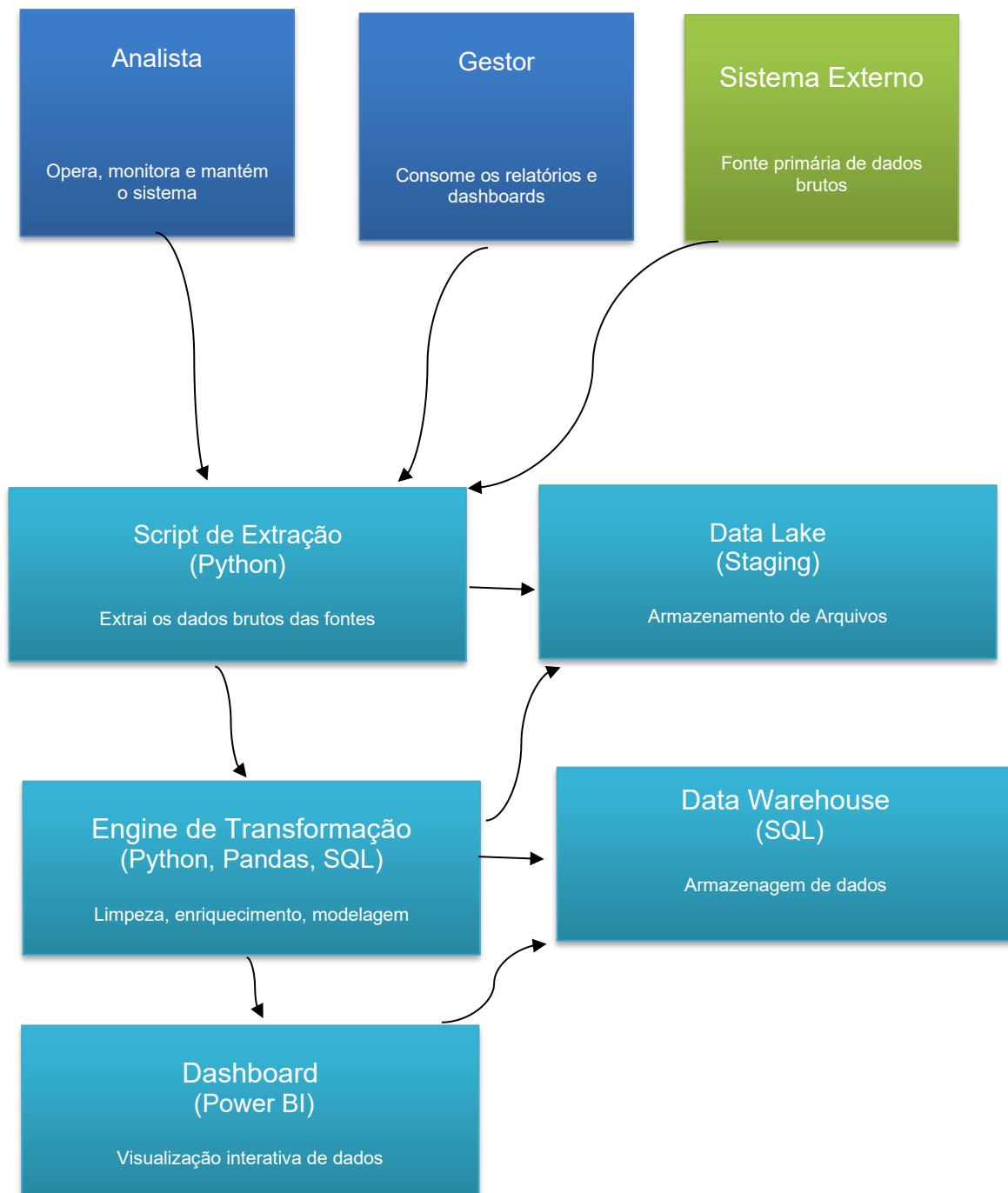
Gera indicadores de risco, reincidência, custo e localização das avarias.

Suporta a tomada de decisão estratégica, apresentando as informações de forma visual e interativa.

Nível 2: Diagrama de Containers

O diagrama apresenta a arquitetura de containers do sistema MPRPV, que automatiza a coleta, transformação, modelagem e visualização dos dados utilizados para prever o risco de imóveis e priorizar vistorias.

Ele mostra como os componentes internos (scripts, banco de dados, dashboard) se comunicam e quais papéis humanos e sistemas externos estão envolvidos.



Descrição dos Containers e Fluxo de Dados:

Atores Externos

1. Analista de Dados / Cientista de Dados

Função: Opera, monitora e mantém o sistema.

Interação: Executa scripts de ETL, valida os dados no banco e publica os dashboards.

Ferramentas: Python, Power BI, SQL.

2. Gestor Técnico / Supervisor de Vistorias

Função: Consome os relatórios e dashboards para tomada de decisão.

Interação: Acessa o painel de Power BI para visualizar o ranking de risco e o mapa de priorização das vistorias.

3. Sistema Externo — Fonte de Dados

Função: Origem dos dados de entrada, como relatórios de vistoria, cadastros de imóveis, chamados ou denúncias.

Formato dos dados: Planilhas CSV, XLSX.

Containers Internos do Sistema MPRPV

1. Script de Extração (Python)

Função: Responsável por ler e extrair dados brutos das fontes (planilhas, PDFs, APIs).

Saída: Dados em formato CSV armazenados no Data Lake (Staging).

Responsável: Analista de Dados.

2. Data Lake (Staging Layer)

Função: Armazena os dados brutos antes de qualquer transformação.

Formato: CSV.

Objetivo: Garantir histórico e rastreabilidade dos dados originais.

Exemplo de tecnologia: Azure Blob ou diretório seguro local.

3. Engine de Transformação (Python / Pandas / SQL)

Função: Realiza limpeza, enriquecimento e modelagem dos dados.

Processos principais:

- Normalização de colunas;
- Tratamento de valores ausentes;
- Criação de variáveis de risco;

Geração do dataset final no formato estrela.

Destino: Data Warehouse.

4. Data Warehouse (SQL)

Função: Armazena os dados transformados e prontos para análise.

Modelo: Esquema estrela (tabelas Fato e Dimensão).

Uso: Fonte de verdade única (single source of truth).

Consultas SQL são executadas por dashboards e relatórios.

5. Dashboard (Power BI)

Função: Interface visual do sistema, onde os gestores acessam dashboards e indicadores de risco.

Principais painéis:

- Mapa de calor por bairro;
- Ranking de risco dos imóveis;
- Indicadores de eficiência de vistoria.

Conexão direta: Data Warehouse via gateway seguro.

Fluxo Resumido

- Dados brutos chegam de fontes externas;
- Script de Extração coleta e envia ao Data Lake;
- Engine de Transformação trata e modela;
- Dados prontos são salvos no Data Warehouse;
- Power BI consome e exibe dashboards para gestores;
- Analista de Dados monitora e mantém todo o ciclo.

7.2.Visualização de Dados: Mockup do Dashboard em Power BI

8. CONCLUSÃO

9. REFERÊNCIAS

10. DICIONÁRIO DE TERMOS TÉCNICOS