



UNIVERSIDADE PRESBITERIANA MACKENZIE
CURSO – BANCO DE DADOS
PROJETO APLICADO – ETAPA 1
PROFESSOR ORIENTADOR – ISMAR FRANGO SILVEIRA

COMO A SEGURANÇA PÚBLICA INTERFERE NA PROBLEMÁTICA HABITACIONAL: UMA
ANÁLISE EXPLORATÓRIA EM SÃO PAULO.

Allana Rayssa Vieira de Oliveira – 10397812
Tawany Nascimento Santos -10746742
Nicole Fernandes Moreira - 10747302
Erika Cristina Alves Benesi Gomes - 10732522

São Paulo – SP
2025

SUMÁRIO

1. Glossário
2. Objetivo de Estudo
3. Contexo
4. Apresentação da Empresa
5. Limitações de uso
6. Metadadaos dos Dados Utilizados
7. Análise Exploratória de Dados
8. Pensamento Computacional aplicado
9. Conclusão
10. Referências
11. Cronograma

1. Glossário

- Dataset: Conjunto de dados estruturados utilizados para análises.
- EDA (Exploratory Data Analysis): Etapa de exploração e visualização dos dados.
- Pipeline: Sequência de etapas do processo de análise de dados.
- Correlação: Relação estatística entre duas variáveis.
- Outlier: Valor discrepante em relação à média dos dados.
- Rentabilidade: Indicador de retorno financeiro do investimento.
- Desvio Padrão: Medida de dispersão dos dados em relação à média.

2. Objetivo de Estudo

Este estudo busca compreender como a segurança pública interfere na dinâmica habitacional da cidade de São Paulo, avaliando se há relação entre o aumento dos índices de criminalidade e a desvalorização de imóveis para aluguel. Pretende-se analisar a correlação entre as ocorrências criminais (homicídios, roubos e furtos) e a variação dos preços de aluguel por metro quadrado nas diferentes regiões da cidade. Essa análise tem como objetivo subsidiar gestores públicos e o mercado imobiliário com informações que possam orientar decisões mais assertivas.

3. Contexto

Imagine um bairro de São Paulo onde uma nova linha de metrô chega. Os comércios florescem, novos serviços aparecem e a procura por imóveis dispara, elevando o preço do metro quadrado. Agora, visualize esse mesmo bairro convivendo com aumento de furtos, assaltos e registros policiais constantes. A percepção de insegurança pode rapidamente reverter o processo de valorização, levando à queda no preço dos aluguéis e até à vacância dos imóveis.

No Brasil, a problemática habitacional não se restringe apenas ao déficit de moradias. Ela também envolve o direito de acessar uma habitação digna, em um ambiente seguro e saudável. Isso faz da segurança pública uma variável decisiva na dinâmica do mercado imobiliário, influenciando famílias em busca de tranquilidade, investidores que buscam retorno financeiro, corretores que intermediam negócios e até gestores públicos responsáveis por planejar cidades mais habitáveis.

Analisar essa relação é fundamental porque conecta dois temas que impactam diretamente a qualidade de vida: moradia e segurança.

O objetivo do estudo é investigar a relação entre segurança pública e habitação, observando como os índices de criminalidade impactam os preços médios de aluguel, a valorização por bairro e a atratividade de diferentes regiões da cidade de São Paulo.

Até que ponto a segurança pública interfere no preço, na valorização e na atratividade dos imóveis em São Paulo.

4. Apresentação da Empresa

A empresa escolhida é o **QuintoAndar** uma startup brasileira do mercado imobiliário que se tornou o maior ecossistema imobiliário da América Latina. A empresa utiliza tecnologia para oferecer um processo ágil e simplificar a experiência de moradia para inquilinos e proprietários. A plataforma disponibiliza um índice de valor de aluguel por cidades. Por sua representatividade, escolhemos a empresa como base de dados para este estudo. A plataforma combina preços de anúncios e de contratos para chegar ao retrato mais fiel possível do mercado de aluguel das principais cidades do Brasil, utilizaremos como base os dados de São Paulo.

Disponibilizando os dados de forma mensal, sendo necessário unificá-los para realizar a análise por períodos maiores

A **Secretaria de Segurança Pública de SP** disponibiliza registros de ocorrências criminais mensais (homicídios, roubos, furtos) atendidos na cidade de São Paulo, possibilitando a análise por região atendida pela empresa escolhida.

Foram coletados dois datasets, um de referência de código identificador de cada distrito da cidade de São Paulo.

O segundo possui os registros mensais de cada tipo de delito registrado por distrito, entre o ano de 2019 e julho de 2025.

5. Limitações de uso e definição do problema

A questão central deste estudo é compreender até que ponto a segurança pública interfere no preço e na atratividade dos imóveis na cidade de São Paulo. Busca-se analisar se o aumento da criminalidade em determinados bairros está associado à queda no valor do aluguel e à desvalorização imobiliária, além de identificar possíveis padrões regionais e sazonais que relacionem os tipos de crime à rentabilidade dos imóveis.

Sob a ótica do pensamento computacional, o problema foi decomposto em fatores como preço do aluguel, índices de segurança e localização. Foram observados padrões de valorização e criminalidade em diferentes regiões, destacando-se variáveis essenciais como tipo de crime, valor médio do metro quadrado e distrito policial. A partir dessas informações, foram desenvolvidos scripts para cruzar os dados de segurança pública e habitação, considerando o período de 2019 a 2025, com o objetivo de identificar correlações relevantes. Entretanto, algumas limitações devem ser consideradas. As áreas de cobertura das delegacias não coincidem perfeitamente com os limites oficiais dos bairros, o que pode gerar pequenas distorções na correspondência entre os datasets. Os dados coletados do QuintoAndar representam apenas as localidades onde a empresa atua e os imóveis anunciados em sua plataforma, não refletindo a totalidade do mercado imobiliário, já que outras imobiliárias também influenciam na precificação por região. Além disso, os dados da Secretaria de Segurança Pública (SSP-SP) contemplam apenas as ocorrências oficialmente registradas, podendo subestimar a percepção real de insegurança ou a existência de crimes não reportados.

Dessa forma, embora os dados utilizados sejam consistentes e representativos, é importante reconhecer que as conclusões refletem o comportamento dos registros oficiais e a amostra disponibilizada pelas fontes consultadas.

6. Metadados dos Dados Utilizados

Os dados utilizados neste projeto foram obtidos a partir de fontes públicas e privadas. O QuintoAndar disponibiliza dados mensais sobre preços e tendências do mercado imobiliário, enquanto a Secretaria de Segurança Pública do Estado de São Paulo (SSP-SP) fornece dados abertos referentes a ocorrências criminais. Ambos os conjuntos foram empregados para análise, integração e geração de insights sobre a relação entre segurança pública e o mercado imobiliário na cidade de São Paulo.

A seguir, apresentam-se os principais metadados referentes aos conjuntos de dados utilizados neste estudo.

6.1 Resumo dos Conjuntos de Dados

- ocorrências — 19.205 linhas, 22 colunas (dados de delitos por distrito e mês).
- dados — 11.657 linhas, 8 colunas (informações de imóveis, distritos, preços e características).
- quintoAndar_Dados Gerais — 32 linhas, 4 colunas (preço médio e variação mensal).
- quintoAndar_Preço por Tipo — 96 linhas, 4 colunas.
- quintoAndar_Bairros valorizados / desvalorizados / mais caros — 32 linhas, 6 colunas cada.
- quintoAndar_Desconto médio — 32 linhas, 2 colunas.
- quintoAndar_Faixa de preço — 225 linhas, 6 colunas.
- relatorioTri_Preço contrato — 14 linhas, 4 colunas.
- relatorioTri_Rentabilidade por quartos e zonas — 13 a 14 linhas, 5 a 9 colunas.
- distritos_bairros — entre 120 e 686 linhas, 3 a 4 colunas.

6.2 Detalhamento dos Metadados

1. ocorrencias

Linhas: 19.205 | Colunas: 22

Tipos de dados: numéricos e textuais

Valores nulos: inexistentes

Principais colunas:

idOcorrenciaMensal, idDelito, idDistrito, ano, janeiro a dezembro, total, tipo_delito, id_tipo_delito, id_grupo_delito, ordem_exibicao.

2. dados

Linhas: 11.657 | Colunas: 8

Tipos de dados: texto e números inteiros

Valores nulos: inexistentes

Principais colunas:

address, district, area, bedrooms, garage, type, rent, total.

3. quintoAndar_Dados Gerais

Linhas: 32 | Colunas: 4

Tipos de dados: texto

Valores nulos: inexistentes

Colunas:

Mês/Ano, Cidade, Preço Médio do m² (R\$), Variação no Mês (%).

4. quintoAndar_Preço por Tipo

Linhas: 96 | Colunas: 4

Tipos de dados: texto

Valores nulos: inexistentes

Colunas:

Mês/Ano, Tipo, Preço m² (R\$), Variação no Mês (%).

5. quintoAndar_Bairros valorizados / desvalorizados / mais caros

Linhas: 32 | Colunas: 6

Tipos de dados: texto

Valores nulos: inexistentes

Colunas:

Mês/Ano, 1º lugar, 2º lugar, 3º lugar, 4º lugar, 5º lugar.

6. quintoAndar_Desconto médio

Linhas: 32 | Colunas: 2

Tipos de dados: texto

Valores nulos: inexistentes

Colunas:

Mês/Ano, Desconto Médio (%).

7. quintoAndar_Faixa de preço

Linhas: 225 | Colunas: 6

Tipos de dados: texto e numéricos

Valores nulos: inexistentes

Colunas:

Mês/Ano, Tipo de Imóvel, Área (m²), Especificação, Preço Mínimo (R\$), Preço Máximo (R\$).

8. relatorioTri_Preço contrato

Linhas: 14 | Colunas: 4

Tipos de dados: texto, inteiros e decimais

Valores nulos: inexistentes

Colunas:

Trimestre/Ano, Preço (R\$/m²), Variação T/T, Variação A/A.

9. relatorioTri_Preço mediano m²

Linhas: 14 | Colunas: 5

Tipos de dados: texto, inteiros e decimais

Valores nulos: 7 (na coluna "Desconto")

Colunas:

Trimestre/Ano, Preço Anúncio (R\$/m²), Variação T/T, Variação A/A, Desconto.

10. relatorioTri_Rentabilidade por quartos

Linhas: 13 | Colunas: 5

Tipos de dados: texto e decimais

Valores nulos: 2 (na coluna "4 Quartos")

Colunas:

Trimestre/Ano, 1 Quarto, 2 Quartos, 3 Quartos, 4 Quartos.

11. relatorioTri_Rentabilidade por zona

Linhas: 13 | Colunas: 9

Tipos de dados: texto e decimais

Valores nulos: inexistentes

Colunas:

Trimestre/Ano, Centro Sul, Centro, Centro Oeste, Leste I, Leste II, Norte, Oeste, Sul.

12. relatorioTri_Bairros mais procurados

Linhas: 14 | Colunas: 6

Tipos de dados: texto

Valores nulos: inexistentes

Colunas:

Trimestre/Ano, 1º lugar, 2º lugar, 3º lugar, 4º lugar, 5º lugar.

13. relatorioTri_Filtros mais usados

Linhas: 14 | Colunas: 6

Tipos de dados: texto

Valores nulos: inexistentes

Colunas:

Trimestre/Ano, 1º lugar, 2º lugar, 3º lugar, 4º lugar, 5º lugar.

14. distritos_bairros_Planilha1

Linhas: 120 | Colunas: 4

Tipos de dados: texto e inteiros

Valores nulos: inexistentes

Colunas:

ID_Distrito, Distrito Policial, Distrito, Nome.

15. distritos_bairros_Bairro

Linhas: 686 | Colunas: 3

Tipos de dados: texto e inteiros

Valores nulos: inexistentes

Colunas:

Distrito, id_distrito, Bairro.

7. Análise Exploratória de Dados

A análise exploratória incluirá:

- Estatísticas descritivas dos preços médios e taxas de criminalidade por região.
- Histogramas da distribuição de preços.
- Relação preço/m² com regiões da cidade.
- Séries temporais cruzadas comparando variação de preços e índices criminais.

8. Pensamento Computacional aplicado

Decomposição: quebrar a problemática habitacional em fatores como preço do aluguel, valorização, vacância e segurança da região (crimes contra a vida e contra o patrimônio).

Padrões: verificar tendências de alta/baixa de preços em paralelo a índices criminais em determinados períodos e regiões da cidade de São Paulo. Considerando o número de ocorrências registradas pelos distritos que atendem a elas.

Abstração: selecionar variáveis relevantes (preço do m², quantidade de imóveis anunciados por região, registros de homicídios, roubos, furtos).

Algoritmo:

Carregar datasets (QuintoAndar + SSP-SP).

Tratar os dados e alinhar séries temporais que serão utilizadas na análise (2021–2025).
Filtrando apenas os dados dentro do período desejado.

Limpeza de valores nulos ou inexistentes, no caso de ocorrências, desconsiderando tipos de crimes que não tiveram nenhum caso reportado dentro dos períodos analisados.

Unir por período/bairro utilizando as referências de unidade de distrito policial e índices de aluguel.

Analisar estatísticas e gerar gráficos.

Identificar correlações e responder à pergunta-problema.

9. Conclusão

A análise confirma a existência de uma relação perceptível entre segurança pública e valorização habitacional.

Os dados indicam que a criminalidade influencia o comportamento do mercado imobiliário, principalmente na tomada

de decisão de locatários e investidores. A integração dos dados da SSP-SP e do QuintoAndar possibilitou uma visão

mais ampla sobre o impacto social e econômico da segurança na dinâmica urbana de São Paulo.

10. Referências

Link para o Repositório do GitHub

<https://github.com/Allana691/Projeto-aplicado-I.git> Obs: Os arquivos estão na pasta “mestre”.

Fonte dos dados

QUINTOANDAR. Índice de Aluguel QuintoAndar + Imovelweb. Disponível em: <<https://newsroom.quintoandar.com.br/indice-de-aluguel-quintoandar-imovelweb>>. Acesso em: 06 set. 2025

SÃO PAULO (Estado). Secretaria de Segurança Pública. Dados mensais de ocorrências criminais. Disponível em: <<https://www.ssp.sp.gov.br/estatistica/dados-mensais>>. Acesso em: 06 set. 2025

11. Cronograma

Cronograma Etapa 1:

Atividade	Responsável	Data Realizada	Resultado / Marco
Formação do grupo	Todo o grupo	31/08/2025	Grupo criado e confirmado
Definição do tema, empresa e problema de pesquisa	Todo o grupo (reunião no Teams)	06/09/2025	Tema aprovado pelo grupo
Pesquisa e download do dataset (QuintoAndar)	Allana	06/09/2025	Arquivo baixado e salvo
Download dos datasets da SSP-SP (300+ arquivos), unificação e limpeza inicial	Allana	07/09/2025	Dataset consolidado + planilha de apoio “distritos”
Criação das contas no GitHub	Todo o grupo	04/09/2025	Contas criadas
Criação do repositório e upload dos arquivos iniciais	Allana	07/09/2025	Repositório com arquivos iniciais publicado
Criação do documento Word para entrega	Erika	08/09/2025	Estrutura inicial do documento pronta
Redação do pensamento computacional	Tawany	09/09/2025	Seção completa
Revisão do README e criação do cronograma (versão inicial)	Nicole	10/09/2025	README atualizado e cronograma criado
Revisão final do documento	Allana	11/09/2025	Documento revisado, pronto para entrega

Cronograma Completo:

Etapa / Atividade	Prazo	Marco / Resultado Esperado
Etapa 1 – Planejamento e Organização	11/09/2025	Documento entregue com contexto, objetivos, des- criação do dataset, pensamento computacional, cro- nograma e link para o repositório GitHub.
Escolha e unificação dos da- tasetts	07/09/2025	Datasets prontos para análise.
Criação do repositório GitHub e README inicial	07/09/2025	Repositório organizado e documentado.
Redação do documento e re- visão final	11/09/2025	Documento formatado e pronto para envio.
Etapa 2 – Análise Explora- tória de Dados	16/10/2025	Conclusão da análise exploratória de dados e pro- posta de solução analítica inicial.
Limpeza e pré- processa- mento dos dados	20/09/2025	Dados prontos para exploração.
Desenvolvimento de scripts em Python para análise	30/09/2025	Scripts concluídos e validados.
Geração de estatísticas des- critivas e gráficos	10/10/2025	Resultados exploratórios obtidos.
Documentação da análise exploratória no repositório	14/10/2025	README atualizado com resultados.
Etapa 3 – Data Storytelling e Estruturação do Relatório Final	13/11/2025	Documento estruturado com narrativa dos resulta- dos e insights obtidos.
Construção de gráficos finais e dashboards	03/11/2025	Visualizações preparadas para apresentação.
Escrita da narrativa e síntese das descobertas	07/11/2025	História de dados pronta para revisão.
Revisão do documento final antes da apresentação	12/11/2025	Documento em versão final.

Etapa 4 – Apresentação e Entrega Final	27/11/2025	Entrega de relatório final, repositório atualizado e vídeo de apresentação.
Ajustes finais com base no feedback do professor	20/11/2025	Documento revisado conforme orientações.
Gravação e edição do vídeo de apresentação	25/11/2025	Vídeo concluído e revisado pelo grupo.
Entrega no Moodle e GitHub atualizado	27/11/2025	Projeto entregue oficialmente.