



UNIVERSIDADE FEDERAL DE OURO PRETO
CAMPUS MORRO DO CRUZEIRO
ESCOLA DE MINAS
DEPRO - DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
CURSO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

PEDRO FERNANDES DE SOUZA NETO

**Implementação de *Business Intelligence* através da
metodologia CRISP-DM para apoiar tomada de decisão
estratégica em uma corretora de imóveis.**

Ouro Preto
2022

PEDRO FERNANDES DE SOUZA NETO

Implementação de *Business Intelligence* através da metodologia CRISP-DM para apoiar tomada de decisão estratégica em uma corretora de imóveis.

Trabalho Final de Graduação apresentado como parte dos requisitos necessários à obtenção do título de Engenheiro de Produção na Universidade Federal de Ouro Preto.

Orientador: Prof. Dr. Helton Cristiano Gomes

Ouro Preto
2022

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

N469i Neto, Pedro Fernandes de Souza.
Implementação de Business Intelligence através da metodologia
CRISP-DM para apoiar tomada de decisão estratégica em uma corretora
de imóveis. [manuscrito] / Pedro Fernandes de Souza Neto. - 2022.
74 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. Helton Cristiano Gomes.
Monografia (Bacharelado). Universidade Federal de Ouro Preto. Escola
de Minas. Graduação em Engenharia de Produção .

1. Business Intelligence. 2. CRISP-DM. 3. Power BI. 4. Tecnologia da
Informação. I. Gomes, Helton Cristiano. II. Universidade Federal de Ouro
Preto. III. Título.

CDU 658.5

Bibliotecário(a) Responsável: Maristela Sanches Lima Mesquita - CRB-1716



FOLHA DE APROVAÇÃO

Pedro Fernandes de Souza Neto

Implementação de *Business Intelligence* através da metodologia CRISP-DM para apoiar a tomada de decisão estratégica em uma corretora de imóveis

Monografia apresentada ao Curso de Engenharia de Produção da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Engenheiro de Produção

Aprovada em 13 de abril de 2022

Membros da banca

Doutor - Helton Cristiano Gomes - Orientador - Universidade Federal de Ouro Preto
Mestre - Davi das Chagas Neves - Universidade Federal de Ouro Preto
Doutora - Irce Fernandes Gomes Guimarães - Universidade Federal de Ouro Preto

Helton Cristiano Gomes, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 18/04/2022



Documento assinado eletronicamente por **Helton Cristiano Gomes, PROFESSOR DE MAGISTERIO SUPERIOR**, em 13/04/2022, às 20:59, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0310926** e o código CRC **20611D50**.

Agradecimentos

Sou eternamente grato primeiramente a Deus, que me guiou, deu forças e me protegeu durante toda minha vida e durante esse caminho trilhado.

Agradeço da mesma forma aos meus pais, Jaci e Tânia, que são meus maiores exemplos de vida. Aqueles que nunca me deixaram faltar nada, sempre me deram condições de aprender e me desenvolver pessoal e profissionalmente, me apoiaram e guiaram em minhas escolhas e decisões.

Agradeço aos meus irmãos, Luciana e Fabiano estarem sempre ao meu lado nos momentos difíceis enfrentados nesse trajeto. E por fim, sou muito grato a todos os meus familiares e parentes, principalmente minha madrinha tia Vânia, que contribuíram para que eu chegasse nesse marco tão importante em minha vida.

Sou grato da mesma forma aos meus amigos, que me apoiaram e incentivarão em todos os momentos que precisei e também a toda pessoa que de certa forma contribuiu para que esse momento fosse alcançado com sucesso.

Agradeço ao professor Helton Gomes que me ajudou muito no início desse trabalho, auxiliando a organizar melhor minhas ideias e os processos da empresa, o que foi de grande valia para o trabalho realizado. Agradeço ao DEPRO por toda a luz do conhecimento na qual me conduziu até este momento.

Resumo

O presente trabalho propõe o projeto para implementação da técnica de Business Intelligence através da ferramenta Power BI da desenvolvedora Microsoft para o apoio a tomada de decisão estratégica em uma microempresa do setor imobiliário. O objetivo é introduzir a corretora de imóveis no cenário contemporâneo da utilização da tecnologia da informação para a extração de informações estratégicas baseadas em dados de operações anteriores da empresa. O planejamento das atividades de implementação bem como a manutenção e melhoria dos processos em análise de dados de inteligência de negócio são fundamentais para o sucesso do projeto, motivos nos quais motivou-se a escolha da metodologia CRISP-DM. Tal metodologia permite a execução do projeto dividida em seis etapas interdependentes, isto é, o sucesso da etapa subsequente depende da etapa anterior, essencial para respectivas validações sucessivas até a entrega do modelo em funcionamento e, especialmente, com performance garantida. Outra importante característica da metodologia escolhida para o trabalho é sua ciclicidade, permitindo o retorno ao início do projeto em ciclos para a implementação de novas melhorias principalmente na estrutura de TI, com implementação de novas ferramentas e técnicas para coleta e armazenamento de dados, permitindo-se a criação de um novo plano de ação a cada ciclo do projeto. Considerando-se um empreendimento de pequeno porte, o mesmo possui estrutura de TI em fase rudimentar, frente às novas empresas do setor lastreadas em tecnologias de alto nível. Portanto, a metodologia apresenta-se como a ideal para a solução da deficiência tecnológica do negócio tanto a curto prazo com a implementação do projeto quanto à longo prazo com a introdução de novas práticas, ferramentas e monitoramento do modelo ciclicamente.

Palavras-chave: *Business Intelligence*, CRISP-DM, Power BI, Tecnologia da Informação.

Abstract

The present work proposes the project for the implementation of the Business Intelligence technique through the Microsoft developer's Power BI tool to support strategic decision-making in a microenterprise in the real estate sector. The aim is to introduce the realtor in the contemporary scenario of the use of information technology to extract strategic information based on data from the company's previous operations. The planning of the implementation activities as well as the maintenance and improvement of the processes in analysis of business intelligence data are fundamental for the success of the project, reasons that motivated the choice of the CRISP-DM methodology. Such methodology allows the execution of the project divided into six interdependent stages, that is, the success of the later stage depends on the previous stage, essential for the respective successive validations until the delivery of the working model and, especially, with guaranteed performance. Another important characteristic of the methodology chosen for the work is its cyclicity, allowing the return to the beginning of the project in cycles for the implementation of new improvements, mainly in the IT structure, with the implementation of new tools and techniques for data collection and storage, allowing the creation of a new action plan at each project cycle. Considering a small enterprise, it has a rudimentary IT structure, compared to new companies in the sector backed by high-level technologies. Therefore, the methodology presents itself as ideal for the solution of the technological deficiency of the business both in the short term with the implementation of the project and in the long term with the introduction of new practices, tools and monitoring of the model.

Keywords: *Business Intelligence*, CRISP-DM, Power BI, Information Technology.

Lista de ilustrações

Figura 1 – Variação real dos preços dos imóveis no Brasil (% a.a.) no período 2004-2014 ..	15
Figura 2 – Estrutura da arquitetura de BI	17
Figura 3 – Ciclo da metodologia CRISP-DM.....	22
Figura 4 – Mind map: fenômeno, agentes e atributos do projeto.....	24
Figura 5 – Conjunto de medidas.....	29
Figura 6 – Passos para a transformação da base em Power Query	32
Figura 7 – Teste hipótese auxiliar 6.1	44
Figura 8 – Função <i>Performance Analyser</i> do <i>Power BI</i>	54
Figura 9 – Indicadores de desempenho satisfatórios.....	60
Figura 10 – Relação receitas real e ociosa: desempenho satisfatório.....	61
Figura 11 – Gráficos tipo rosca: desempenho satisfatório	61
Figura 12 – Relação tipos de receitas: desempenho satisfatório.....	62
Figura 13 – Relação receitas real e ociosa: desempenho insatisfatório.....	64
Figura 14 – Gráficos tipo rosca: desempenho insatisfatório	64
Figura 15 – Relação tipos de receitas: desempenho insatisfatório.....	65
Figura 16 – Anexo 2: Dashboard final do projeto modo inicialização	73
Figura 17 – Anexo 3: Dashboard imóvel com desempenho satisfatório.....	74
Figura 18 – Anexo 4: Dashboard imóvel com desempenho insatisfatório.....	74

Lista de gráficos

Gráfico 1 – Gráfico dados compilados do CRECI-SP.....	16
Gráfico 2 – Correlação de Pearson - Análise Multivariada.....	36
Gráfico 3 – Correlação de <i>Pearson</i> para medidas de desempenho	37
Gráfico 4 – Percentual.....	38
Gráfico 5 – Gráfico de barras com percentuais de tempo ocupado por cada tipo de presença de garagens	39
Gráfico 6 – Teste hipótese 2.....	40
Gráfico 7 – Teste hipótese 3.....	41
Gráfico 8 – Teste hipótese 4: receitas para a imobiliária.....	41
Gráfico 9 – Teste Hipótese 4: receitas para o locador.....	42
Gráfico 10 – Teste hipótese 5.....	42
Gráfico 11 – Teste hipótese 6.....	43
Gráfico 12 – Teste hipótese 7.....	45
Gráfico 13 – Teste hipótese 8.....	46
Gráfico 14 – Teste hipótese 9.....	47
Gráfico 15 – Teste hipótese 10: imóveis analisados por valores únicos	48
Gráfico 16 – Teste hipótese 10: imóveis analisados por contratos	49
Gráfico 17 – Indicadores: KPI's e cartões	50
Gráfico 18 – Relação receitas real e ocisa por contratos	51
Gráfico 19 – Gráficos tipo rosca.....	51
Gráfico 20 – Relação tipos de receitas.....	52
Gráfico 21 – Indicadores de desempenho insatisfatórios.....	63

Lista de tabelas

Tabela 1 – Descrição dos dados provenientes da plataforma InGaia.....	27
Tabela 2 – Descrição das variáveis (Features) derivadas das originais.....	28
Tabela 3 – Descrição de medidas.....	30

Lista de abreviaturas e siglas

AED - Análise Exploratória de Dados

BC - Banco Central

BI - Business Intelligence

BPM - Business Process Management

CRECISP - Conselho Regional de Corretores de Imóveis do Estado de São Paulo

CRISP-DM - Cross-Industry Standard Process for Data Mining

CRM - Customer Relationship Management

DAX - Data Analysis Expressions

DW - Data Warehouse

ERP - Enterprise Resource Planning

ETL - Extract Transform Load

IBGE - Instituto Brasileiro de Geografia e Estatística

IPEA - Instituto de Pesquisa Econômica Aplicada

KPI - Key Performance Indicator (Indicador Chave de Desempenho)

LGPD - Lei Geral de Proteção de Dados Pessoais

OLAP - Online Analytical Processing

POP - Procedimento Operacional Padrão

TACE - Teoria Austríaca dos Ciclos Econômicos

TI - Tecnologia da Informação

Sumário

1	Introdução.....	11
1.1	Objetivos.....	12
1.1.1	Objetivos Gerais	12
1.1.2	Objetivos Específicos	12
1.2	Justificativas.....	12
2	Revisão Bibliográfica	14
2.1	Mercado Imobiliário.....	14
2.2	O que é Business Intelligence	16
2.2.1	<i>Power BI.....</i>	<i>18</i>
2.3	Vantagens Competitivas e Desafios para a Implementação do Projeto	18
3	Metodologia	21
3.1	Compreensão do Negócio	22
3.1.1	Problemas do Negócio.....	22
3.1.2	Anamnese: Diagnóstico do Negócio na Perspectiva dos Gestores.....	24
3.1.3	Levantamento de Hipóteses.....	25
3.2	Compreensão dos Dados	26
3.2.1	Fontes dos Dados.....	26
3.2.2	Construção de <i>Features</i>	27
3.2.3	Medidas.....	28
3.3	Preparação dos Dados.....	30
3.3.1	Processo <i>Extraction, Transformation and Load</i> (ETL)	30
3.4	Modelagem	33
3.4.1	Análise Exploratória dos Dados (AED).....	33
3.4.2	Correlações	34
3.4.3	Insights.....	34
3.4.4	Testes e Resultados.....	35
3.4.5	Seleção dos Gráficos.....	49
3.5	Avaliação.....	52
3.5.1	Atendimento às necessidades.....	52
3.5.2	Desempenho	53
3.6	Implementação	55
4	Estudo de Caso.....	56
4.1	A Empresa	56
4.2	Processos de Tomada de Decisão	56
4.3	Problemas e Deficiências do Negócio	57

5	Resultados e Discussão	58
5.1	<i>Insights</i> e Informações de Negócio em <i>Power BI</i>	58
5.2	Dashboard	59
5.3	Plano de Ação.....	65
6	Conclusão e Considerações Finais.....	67
7	Referências Bibliográficas.....	69

ANEXOS

71

1 Introdução

De acordo com dados da Agência Paulista de Promoção de Investimentos e Competitividade - INVESTSP do Governo do Estado de São Paulo, entre os anos de 2008 e 2012, houve no estado um crescimento nos investimentos na indústria da construção civil de R\$34,9 bilhões para R\$60,5 bilhões respectivamente.

Tratando-se de um setor no qual está relacionado a outros setores de serviços de atividades imobiliárias, estes acompanharam o crescimento do setor imobiliário no estado de São Paulo com um crescimento de R\$4,2 bilhões em 2007 para R\$10,6 bilhões em 2011. Neste último ano, a participação paulista fechou em uma taxa de 41% de participação na receita total do Brasil. Apesar da desaceleração deste crescimento, imóveis construídos ainda neste período, em conjunto com novos projetos da construção civil, fazem com que a economia neste setor continue em movimento através dos serviços ligados ao setor, com destaque para a corretagem. É neste contexto que, imobiliárias de grande porte surgem para atender ao dinamismo da população principalmente na capital. Dentre estas empresas surgem startups de serviços imobiliários.

As altas demandas por moradias e espaços laborais com serviços de compra e venda de imóveis, a popularização da internet e de tecnologias como Smartphones e Tablets favoreceram o crescimento desta modalidade de empreendimentos no setor da corretagem imobiliária com tecnologias de alto nível, essenciais tanto à experiência do cliente quanto à própria empresa com grande potencial para a tomada de decisão em tempo real.

Devido ao surgimento deste novo conceito de empresas do ramo imobiliário pautadas na tecnologia e inovação como competidoras de empresas tradicionalmente deficiente em tecnologias da informação, como a empresa do presente trabalho, surgiu-se a motivação da implementação do projeto de BI através de uma metodologia robusta que abrange um conjunto de técnicas para a coleta, transformação e análise de dados brutos com o objetivo de extrair-se informações significativas a fim de se obter conhecimento tanto para novas oportunidades de mercado quanto monitoramento de métricas conhecidas a priori.

Diante desta necessidade, este trabalho terá como foco, o desenvolvimento de um modelo de BI através da ferramenta *Power BI* à partir da metodologia *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Esta metodologia permite tanto a implementação do projeto em curto prazo quanto seu acompanhamento, monitoramento e melhoria contínua em longo prazo por meio de seis etapas dependentes entre si, onde a etapa subsequente depende do sucesso da etapa anterior.

Por conseguinte, a proposta do trabalho é promover um modelo de inteligência de

negócios através dos dados disponíveis da corretora bem como abordar a importância de outras fontes de dados estratégicas disponíveis no negócio porém desestruturadas ou em fontes não utilizadas para a composição de um modelo em *Power BI* para o conhecimento de novas oportunidades de mercado e monitoramento de métricas através de um dashboard. Em primeira instância demonstrar a eficiência da ferramenta apenas com dados estruturados disponíveis no primeiro ciclo da metodologia CRISP-DM como resultado do trabalho e iniciar a cultura da companhia a uma cultura orientada a dados.

1.1 Objetivos

1.1.1 Objetivos Gerais

Implementar a técnica de BI em uma corretora de imóveis através da ferramenta *Power BI* pela metodologia CRISP-DM de modo a introduzir a empresa no cenário tecnológico para a tomada de decisões estratégicas, frente à concorrência.

1.1.2 Objetivos Específicos

Referente aos Objetivos Gerais, têm-se os seguintes objetivos específicos:

- Realizar uma revisão bibliográfica a respeito do mercado imobiliário, utilização de dados e suas relações;
- Levantar a preparação dos dados para a construção do modelo de BI para a empresa;
- Modelar o problema de negócio e definir modelo ideal para o negócio;
- Analisar e avaliar performance do modelo;
- Levantar todas as melhorias necessárias para o aprimoramento constante do modelo;
- Destacar a importância do projeto para a empresa.

1.2 Justificativas

Segundo Sharda *et al.* (2019) os ambientes gerenciais estão evoluindo cada vez mais em termos tecnológicos e, com isso, tornando-se mais competitivos, buscando-se decisões cada vez mais ágeis, assertivas e estratégicas, contando com uma crescente quantidade de dados processados e analisados em tempo real, exigindo-se assim suporte computadorizado cada vez mais complexo. Ainda, segundo os autores, ferramentas de análise de dados e BI, como armazenamento, mineração, processamento analítico online ou *Online Analytical Processing* (OLAP), *dashboards* e uso de sistemas baseados na nuvem são os pilares da gestão moderna.

Companhias tradicionais competem com empresas de alto nível tecnológico dentro de um contexto em que, a cada dia, tais empresas baseadas na inovação e sua globalização derrubam barreiras tradicionais de comércio e investimento (LYNN; MYTELKA; 1999). É neste ambiente competitivo que, segundo os autores, o setor imobiliário, apesar de compor um conjunto de necessidades essenciais básicas à sobrevivência e qualidade de vida humana com os serviços de moradia, aliado à outras necessidades básicas como alimentação, transporte, vestuário, higiene e saneamento, pode-se observar a crescente demanda tecnológica na qual empresas dotadas de grande capacidade tecnológica posicionam-se à frente de corretoras ortodoxas, tanto dispõem-se ferramentas gerenciais para tomadas de decisão complexas quanto oferecer uma melhor experiência ao cliente.

Dentre as inovações no ambiente gerencial destaca-se o BI, que consiste em uma combinação entre conhecimento de negócios, arquiteturas de dados e ferramentas analíticas empregadas em análises de desempenho gerencial. O principal objetivo do BI é possibilitar acesso interativo, às vezes em tempo real, a dados, permitir a manipulação de dados e oferecer aos gestores e analistas a capacidade de conduzir análises apropriadas. Ao analisarem dados, situações e desempenhos históricos e atuais, os tomadores de decisões obtêm vislumbres valiosos que lhes permitem tomar decisões mais embasadas. O processo de BI baseia-se na transformação de dados em informações, depois em decisões e por fim em ações (SHARDA *et al.*, 2019).

Perante o estudo desenvolvido e uma profunda análise através dos conceitos das bibliografias e ferramentas aplicadas, será possível propor melhorias nos processos do empreendimento por meio de visualizações gráficas nas quais trarão à luz informações outrora desconhecidas ou meramente hipotéticas por parte da gerência. Portanto, tais informações gerenciais permitirão a tomada de decisões estratégicas mais assertivas e contribuirão para uma possível reestruturação no modelo de negócios da empresa.

2 Revisão Bibliográfica

2.1 Mercado Imobiliário

Dos empreendimentos que impulsionam o mercado imobiliário destacam-se as empresas de construção civil, imobiliárias, empreiteiras de serviços gerais como empresas de acabamento e limpeza e agências de marketing e publicidade. Diante desta dinâmica, segundo (BALL; 1996), a construção habitacional representa um importante indicador de crescimento econômico. Segundo o autor, apesar de não ser o único indicador, as demandas deste setor refletem na mudança da população bem como na demanda por outros serviços.

A consolidação do mercado imobiliário, principalmente de corretagem de imóveis, é uma consequência de um aquecimento no mercado da construção civil, tanto para imóveis residenciais quanto comerciais. Tal impulso na construção civil no Brasil deve-se, segundo o relatório do Instituto de Pesquisa Econômica Aplicada (IPEA), ao país apresentar um crescimento significativo no setor de construção civil denominado na literatura como “Bolha” especulativa do setor imobiliário. Segundo (MENDONÇA; SASCHIDA; 2012), tal fenômeno ocorreu devido às políticas monetárias implementadas pela equipe econômica do governo federal da mesma época, ocasionando altos índices nos preços dos imóveis e a consequente expansão imobiliária. Ainda, segundo o relatório, podem-se destacar como responsáveis pelo aquecimento no setor o controle da inflação, queda na taxa de juros e consequentemente menores custos de financiamento, programas de obras públicas, aquecimento do mercado de trabalho e aumento da renda do trabalhador e, logo, maior acesso deste ao mercado imobiliário, principalmente com o incentivo do governo federal para aquisição de imóveis, inclusive, por famílias de baixa renda através de programas sociais na área.

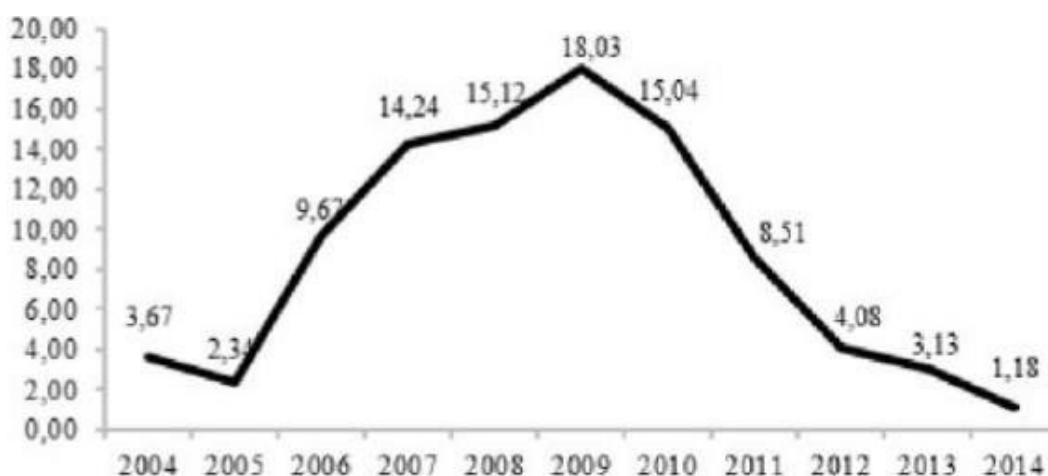
Uma bolha especulativa, de acordo com a perspectiva da Teoria Austríaca dos Ciclos Econômicos (TACE), não surge por força de um comportamento irracional, mas por consequência de um sinal exogenamente criado que faz com que os agentes direcionam erroneamente os investimentos em um determinado segmento da economia, fazendo-se com que os preços deste mercado sigam uma trajetória fortemente ascendente. Porém, tal elevação não se sustenta indefinidamente. Apesar da insustentabilidade do efeito de uma bolha, seja ela em qualquer setor, no caso do setor imobiliário na cidade de São Paulo - SP, a bolha limitou-se apenas à lucratividade sobre as construções e primeiras transações à elas relacionadas, isso ocorreu por dois principais motivos: o primeiro pelo fato de um imóvel ser um ativo permanente, ou seja, uma vez construído pode gerar dividendos ao proprietário permanentemente.

O segundo motivo pelo dinamismo da capital paulista, de acordo com dados do último censo do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2010, São Paulo está em 1º lugar no ranking como a cidade mais populosa do Brasil com 11.253.503 habitantes e ainda, de

acordo com a mesma fonte com dados de 2018 da mesma instituição, a cidade encontra-se em 1º lugar no ranking dos municípios brasileiros de maior PIB per capita do país, com R\$58.691,90. Tais concentrações de pessoas e capitais, ambos relacionados às concentrações industriais e do mercado financeiro, são favoráveis ao dinamismo do mercado imobiliário de corretagem na capital, mesmo após o fim da bolha imobiliária no país.

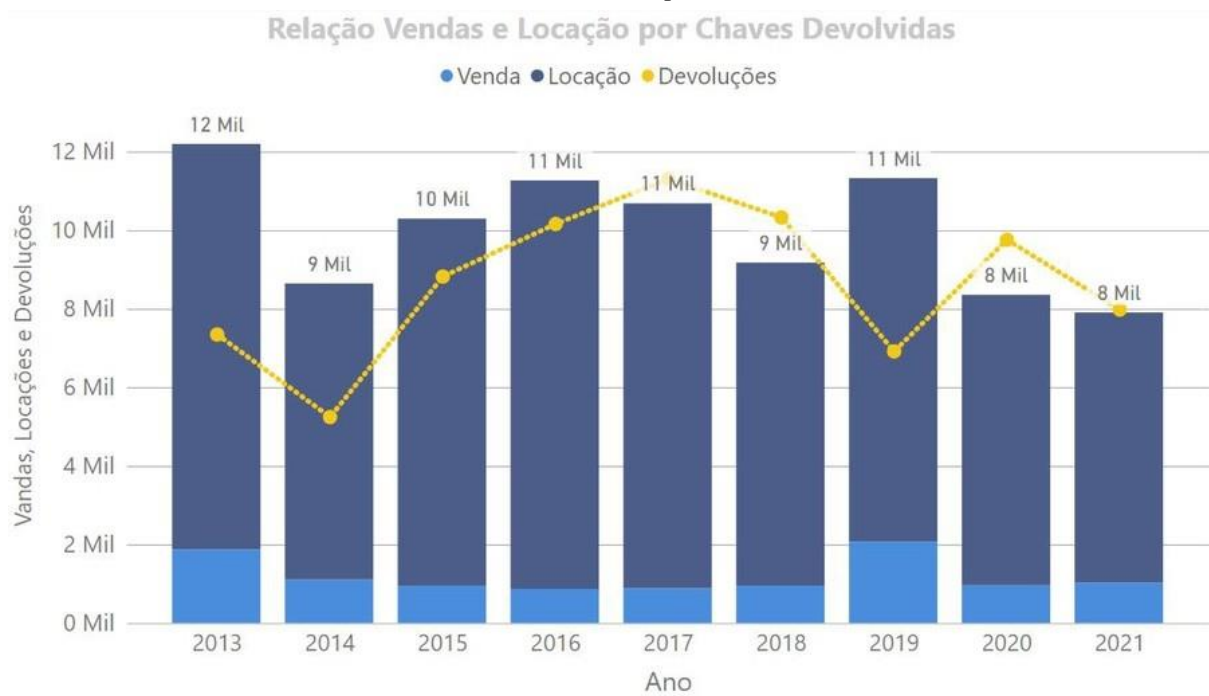
Conforme dados compilados dos relatórios do Conselho Regional de Corretores de Imóveis do Estado de São Paulo (CRECISP) entre os anos de 2013 e 2021, mesmo após o fim da bolha imobiliária, data marcada pelo declínio na variação de preços dos imóveis no Brasil à partir de 2009, segundo estudos dos autores Araujo *et al.* (2016) e baseado no levantamento da variação real dos preços pelo Banco Central de Brasil (BC), o mercado imobiliário na capital paulista manteve uma tendência de alta nas contagens de entregas de chaves, entregues por faixa de preços, tanto na categoria vendas quanto em locações superiores a quantidade de devoluções de chaves (motivos de inadimplência ou outros motivos). Com exceção apenas dos anos de 2017, 2018 e 2020 nos quais o número de chaves devolvidas superou o número de chaves entregues em vendas e locações.

Figura 1 – Variação real dos preços dos imóveis no Brasil (% a.a.) no período 2004-2014



Fonte: Banco Central do Brasil (2015)

Gráfico 1 – Gráfico dados compilados do CRECI-SP



Fonte: Elaborado pelo autor (2021)

Tais informações apresentadas mostram, segundo os autores citados até o momento, que o dinamismo e a competitividade no mercado imobiliário da cidade de São Paulo ainda prevalecem. Motivos suficientes e necessários para a busca de conhecimento do próprio negócio dentro das definições de inteligência de negócios, tratado no presente trabalho como BI.

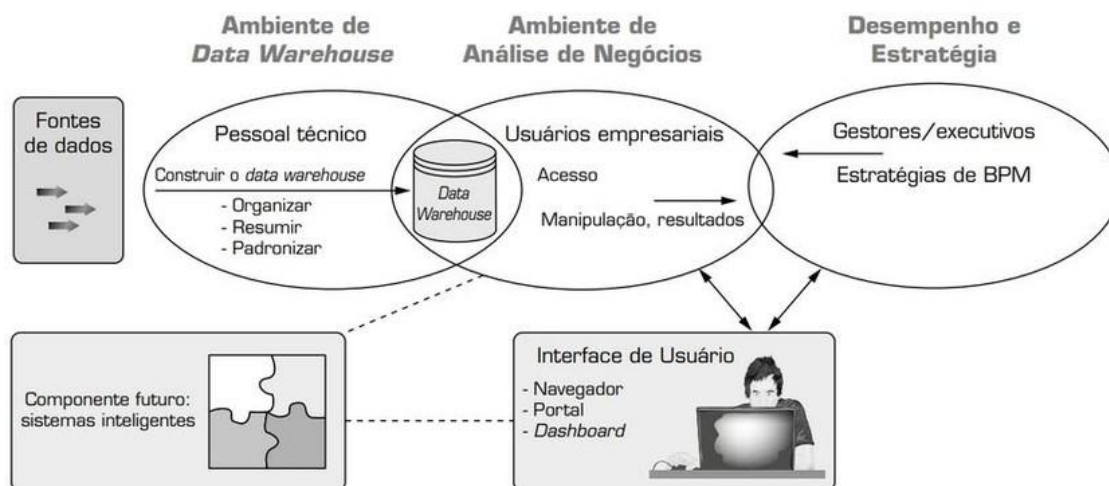
2.2 O que é Business Intelligence

Os autores Sharda *et al.* (2019) se referem ao BI, quanto à técnica, como um termo guarda-chuva que combina arquiteturas, ferramentas, bases de dados, ferramentas analíticas, aplicativos e metodologias. Conforme os autores, especificam-se quanto à técnica pois os mesmos elucidam na bibliografia que há diferenças entre o sistema ou a técnica BI e a tecnologia, esta última relacionada à softwares e seus respectivos fornecedores.

De acordo com Sharda *et al.* (2019), a arquitetura de um sistema BI é composta de *Data Warehouse* (DW) onde são armazenados os dados extraídos da fonte; *Análise de Negócios*, um conjunto de ferramentas para mineração e análise dos dados do DW; o *Business Process Management* (BPM) para o monitoramento de desempenho, ainda por parte dos desenvolvedores e o painel Interativo ou *dashboard*, a interface gráfica para o usuário final. O principal objetivo deste conjunto é fornecer aos gestores e analistas a capacidade de executarem análises de alto nível para a tomada de decisões com base em padrões históricos e atuais através de informações visuais por meio dos *dashboards*, em tempo real (SHARDA *et al.*, 2019).

Todo o processo é executado por meio da importação e transformação dos dados em informações que, subsequentemente, são utilizadas nas decisões estratégicas e ações.

Figura 2 – Estrutura da arquitetura de BI



Fonte: SHARDA; DELEN; TURBAN, 2019, p. 15

A principal motivação que leva a implementação do BI parte da crescente importância e diversificação das áreas de Tecnologia da Informação (TI) dentro das organizações independentemente do porte das mesmas, pois, todas as áreas de uma empresa geram grandes volumes de informações onde há a necessidade de suporte para a coleta e organização dos dados gerados em todas as áreas gerenciais (MICHAELSEN; GOMES, 2017).

No Brasil, o setor da TI nas instituições receberam uma importância ainda maior com o vigor da Lei de N° 3709/2018, a Lei Geral de Proteção de Dados Pessoais (LGPD), cuja disposição é, segundo o Art. 1° da lei:

“... o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.”

A LGPD tem por objetivo não apenas conferir às pessoas um maior controle sobre seus dados, mas também fomentar um ambiente de desenvolvimento econômico e tecnológico, mediante regras flexíveis e adequadas para lidar com os mais inovadores modelos de negócio baseados no uso de dados pessoais. Isso inclui modelos de negócio que se valem de algoritmos para auxiliar na tomada de decisões automatizadas (MONTEIRO, 2018). Tal disposição da lei trouxe uma

maior importância ao TI dentro da organização por demandar uma maior segurança dos dados da empresa, principalmente tratando-se de setores que dependem diretamente de dados pessoais de terceiros, destacando-se as áreas comerciais, principais portas de entrada de dados pessoais de clientes.

A TI nas organizações foi um fator motivador responsável por revolucionar as práticas das empresas principalmente na forma de gerenciamento dos projetos, principalmente pelo advento dos grandes volumes de dados que levou a uma nova abordagem conhecida como *Big Data* (FOX; HENDLER, 2011). Tal abordagem torna necessária a aplicação do BI nas organizações com o objetivo de extrair informações estratégicas de negócio.

2.2.1 Power BI

A ferramenta escolhida para o projeto é o *Power BI*, da desenvolvedora e distribuidora *Microsoft*. Dentre tantos outros distribuidores tanto pagos como *Tableau*, *Oracle*, gratuitos como *Google Studio* e opções *open source* como *Zoho Reports* e *Jaspersoft BI Tools*, apesar de ser um software pago, a ferramenta foi escolhida devido tanto à sua compatibilidade com sistema operacional *Windows* quanto à facilidade de manutenção de seu modelo no qual, o gestor poderá contar tanto com o suporte da *Microsoft Community* quanto se desenvolver para, futuramente, ser auto suficiente na manutenção do modelo sem a ajuda e acompanhamento de profissional qualificado, após a construção do modelo e sua publicação e uso pelo usuário final denominado *Deploy*.

O módulo pago da ferramenta é direcionado à trabalhos colaborativos onde o valor investido permite o compartilhamento do modelo entre membros da equipe. No caso do projeto do trabalho, as opções gratuitas da ferramenta atendem ao projeto, apenas com o *Power BI Desktop* e a publicação do *dashboard* no aplicativo online da própria ferramenta, onde é permitido a exibição no módulo tela inteira, suficientes para atender aos requisitos financeiros de uma microempresa.

2.3 Vantagens Competitivas e Desafios para a Implementação do Projeto

Pequenas empresas como a do presente trabalho, apesar de muitas vezes não contarem com uma complexa estrutura de DW, possuem grandes capacidades de receber e armazenar dados, sejam em planilhas ou diversas fichas preenchidas por meio de formulários, o suficiente para um bom aproveitamento na extração de informações através desses dados.

O processo de mudança tecnológica é o resultado do esforço das empresas em investir em atividades de pesquisa e desenvolvimento e na incorporação posterior dos resultados em novos processos, produtos e formas organizacionais (OLIVEIRA; BORSHIVER, 2013).

Segundo Sharda *et al.* (2019) os motivos fundamentais para investir em BI precisam estar alinhados com a estratégia de negócios da empresa. A ferramenta de BI não pode ser um mero exercício técnico para o departamento de TI. Ela precisa ser um agente de mudança do modo com que a empresa conduz seus negócios, aprimorando-se os processos comerciais e embasando os processos decisórios com dados mais concretos.

Muitos consultores e praticantes de BI envolvidos em iniciativas bem sucedidas na área sugerem que um quadro referencial para planejamento é uma pré-condição necessária. Ambos os autores elucidam a complexidade de um projeto de mudança tecnológica como um desafio para qualquer empreendimento, isso inclui a implementação da ferramenta abordada no presente trabalho, pois, uma mudança tecnológica abrange não apenas a complexidade em se adotar uma nova ferramenta mas também toda uma mudança na cultura da organização e principalmente engajamento de toda equipe envolvida no projeto, de tal maneira que os dados da empresa sejam coletados, armazenados e processados de maneira fidedigna ao processo.

Apesar do poder da metodologia para o negócio, a falta de atenção quanto aos requisitos citados podem representar não uma solução de longo prazo por meio de informações estratégicas de negócio, mas sim um fator de risco para a organização com o mau emprego de recursos como tempo, pessoal e capital em um projeto mal sucedido.

Pequenas empresas como a do trabalho desenvolvido, apesar de não contarem com uma complexa estrutura de DW, possuem grande capacidade de coleta e armazenamento de dados, sejam em planilhas ou formulários elaborados tanto de forma digitais e integradas à planilhas ou simplesmente formulários manuais que podem ser transcritos para planilhas à posteriori. É importante apenas que os gestores se atentem ao planejamento elaborado para a garantia da implementação da ferramenta da maneira correta, principalmente no que tange a gestão de custos e ferramentas de hardware e software adequados bem como segurança e proteção de privacidade e integração com sistemas e aplicativos (SHARDA *et al.*, 2019), requisitos essenciais para nortear o trabalho da equipe envolvida no projeto.

Apesar dos desafios envolvidos, a implementação do BI ainda representa uma vantagem estratégica e competitiva para a organização, seguindo-se os passos do desenvolvimento do projeto e respeitando-se o orçamento financeiro previsto pode-se alcançar benefícios diretos e indiretos. Os autores (FERREIRA *et al.*, 2020) enumeram os principais objetivos que podem ser alcançados com a implantação das ferramentas de BI, destacando-se como benefícios:

- Melhor tomada de decisões;
- Melhoria no planejamento estratégico por parte dos executivos;
- Redução de custos e aumento da receita em diversos setores da organização (maior eficiência na gestão de custos);
- Melhor compreensão de tendências e exploração de novas oportunidades de mercado;
- Elaboração de indicadores de desempenho como parâmetros para o estabelecimento de metas.

Diante do conhecimento apresentado, para micro e pequenas empresas, o desenvolvimento de um projeto de BI é possível e pode, a priori, trazer como resultados conhecimentos acerca do negócio explicado por meio dos dados. Tais resultados podem ser melhorados, a posteriori, em um processo de melhoria contínua e cíclica, com novos estudos e incremento do modelo, utilizando-se de uma metodologia cíclica adequada para projetos de mineração de dados.

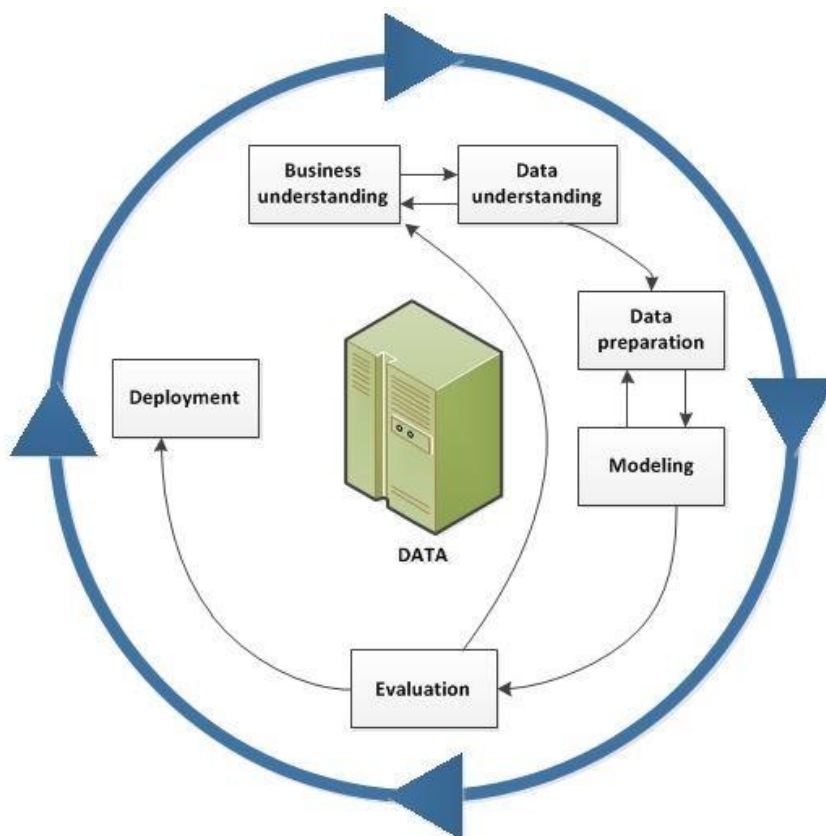
3 Metodologia

Com a evolução das empresas, tanto prestadoras quanto consumidoras de serviços de Mineração de Dados, por meio de um consórcio entre essas empresas, foi desenvolvido um modelo de gerenciamento voltado para projetos de mineração de dados, originalmente denominado CRISP-DM, cuja a abordagem é usada para resolver problemas de negócio desta natureza (SHEARER, 2000).

Independente da área de negócios aplicada, a metodologia consiste em uma sequência de fases de desenvolvimento para projetos de mineração de dados de uma forma estruturada cujos objetivos são, por meio dos dados, recursos humanos e computacionais disponíveis, transformar as necessidades de negócio em rotinas de mineração de dados. Executar transformações nos dados, estabelecer técnicas e ferramentas a serem utilizadas e avaliar os resultados obtidos com o modelo (SILVA *et al.*, 2007).

O CRISP-DM fornece um modelo de processo que entrega uma visão geral do ciclo de vida da Mineração de Dados. Este ciclo consiste em um fluxograma com seis fases onde o projeto é desenvolvido, seguindo a sequência respectiva: A Compreensão ou Entendimento do Negócio (*Business Understanding*), Compreensão ou Entendimento dos Dados (*Data Understanding*), Preparação dos Dados (*Data Preparation*), Modelagem (*Modeling*), Avaliação (*Evaluation*) e Implementação (*Deployment*).

Figura 3 – Ciclo da metodologia CRISP-DM



Fonte: IBM

O trabalho desenvolvido seguiu as seis etapas da metodologia a fim de se obter um direcionamento para o projeto da implementação da ferramenta *Power BI* na organização, compreendendo o primeiro ciclo do CRISP-DM com o objetivo de estabelecer as melhores práticas para o conhecimento do negócio e elaboração de insights que atendam às necessidades do empreendimento.

3.1 Compreensão do Negócio

3.1.1 Problemas do Negócio

A imobiliária apresentada no trabalho foi fundada em 2016 inicialmente como um escritório para uma corretora de imóveis independente que, no caso, é a atual proprietária. Com base em registros, a princípio escritos manualmente e posteriormente em planilhas digitais, o maior problema enfrentado no negócio é a falta de informações analíticas que poderiam ser extraídas de dados contratuais registrados para a tomada de decisões estratégicas.

Em relação aos recursos tecnológicos disponíveis, a empresa dispõe de quatro computadores tipo *desktop*, um para cada colaborador, onde são executados trabalhos de rotina como redação

de contratos, análises documentais e registros de transações de locação, compra e venda dos imóveis. Para a performance do trabalho diário da equipe, inúmeras ferramentas tecnológicas poderiam ser implementadas a partir de cada estação, tanto para melhorar processos operacionais trazendo melhores experiências para o cliente quanto extrair informações para a tomada de decisões estratégicas pelos gestores.

Apesar da variedade de problemas a serem solucionados por meio da tecnologia, os quais podem vir a calhar em uma gama de diferentes projetos, a proposta deste projeto é a apresentação de soluções para a tomada de decisões estratégicas por meio de insights pela ferramenta *Power BI*. Devido à dimensão da aplicabilidade da ferramenta e quantidade de possíveis projetos de visualizações, para o início do processo de implementação da ferramenta foi escolhido o problema mais significativo e de maior urgência segundo a proprietária da imobiliária: apoio a tomada de decisões na gestão contratual através da manutenção ou exclusão de imóveis da carteira para otimização do emprego de pessoal do escritório, para suporte ao inquilino em potencial e o locador. Este problema parte da observação da proprietária, demais colaboradores e parceiros nos quais observam-se imóveis ociosos que mesmo não gerando receita, ainda assim, demandam atenção da mão de obra da empresa, um recurso escasso dado à dimensão da instituição. Semelhantemente a um imóvel de alta procura, que gera um faturamento significativo tanto para os locadores quanto à imobiliária. Portanto, a proposta do projeto é baseada no fenômeno da entrada de receitas tanto para a imobiliária quanto para o locador do imóvel.

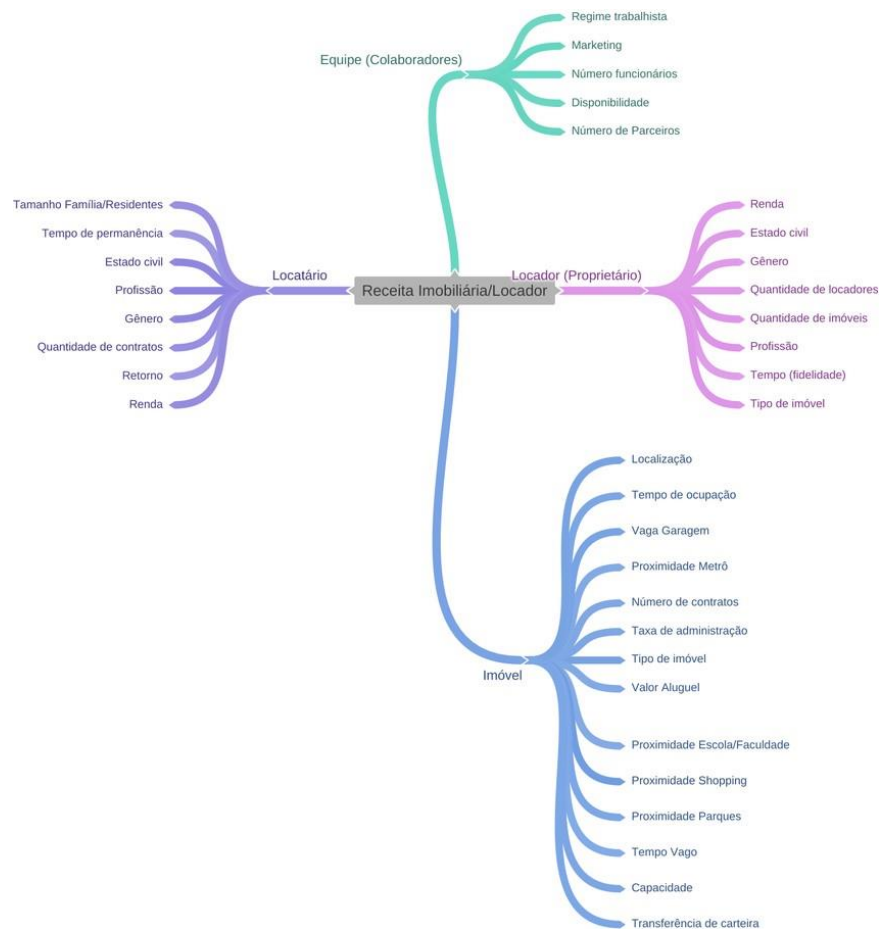
Para melhor vislumbre do problema de negócio cujo trabalho propõe a implementação de BI, o problema foi estruturado graficamente por meio de um *Mind Map*. A complexidade envolvida neste tipo de gráfico refere-se a potencialidade destes mapas em representar o pensamento humano e, a partir daí, organizá-lo de modo não linear, porém lógico e compreensível (ADORNO; TEIXEIRA, 2019).

O *Mind Map* parte de uma representação bidimensional e permite-se com que se organize e sistematize praticamente qualquer coisa a partir da visualização sistêmica como um todo. Partindo-se desta estrutura, dividiu-se o problema de negócio hierarquicamente em três diferentes componentes:

- **Fenômeno:** Este componente é o fenômeno a ser estudado e modelado, ocupa o cerne, a parte central do *Mind Map*. No caso, representa o problema de negócio a ser modelado, principal motivação do projeto, o elemento Receita Imobiliária/Locador.
- **Agente:** São elementos que constituem e impactam o fenômeno estudado. São representados pelas ramificações ou branches que partem do problema estudado. Como representantes neste mapa temos Locador (proprietário do imóvel), Locatário (inquilino), Equipe (Colaboradores da imobiliária) e Imóvel.

- **Atributos dos Agentes:** São as características que podem influenciar no comportamento dos agentes, representados no mapa pelas extremidades das ramificações. Para o agente representado pelo *branch* “Imóvel” temos como atributos o Endereço, Tempo de Ocupação, Tempo de Vacância, Vaga Garagem, Proximidade Metrô, Proximidade faculdades/universidades, Proximidade Parques, Proximidade Shoppings, Valor do Aluguel, Taxa de Administração, Tipo de Imóvel e Quantidade de Contratos.

Figura 4 – Mind map: fenômeno, agentes e atributos do projeto



Fonte: Elaborado pelo autor (2021)

Devido a falta de dados para incluir os demais agentes no projeto, estes serão tratados a partir da segunda etapa do ciclo do CRISP-DM.

3.1.2 Anamnese: Diagnóstico do Negócio na Perspectiva dos Gestores

Realizou-se uma entrevista mediante um questionário com a proprietária da imobiliária. O questionário teve como objetivo conhecer o modelo de negócio, os processos, levantar o conhecimento

da proprietária acerca do próprio empreendimento e identificar um possível problema de negócio passível de solução com aplicação do BI bem como identificar a deficiência do negócio com relação a tomada de decisão através do uso de dados e sua disposição e interesse em soluções inovadoras.

A conclusão a respeito da entrevista é que apesar da empresa funcionar com decisões baseadas em experiências anteriores e contar com uma pequena equipe que realiza tarefas com base em Procedimento Operacional Padrão (POP) a empresa ainda assim precisa de um planejamento mais estratégico no qual permitiria à empresa distribuir melhor as funções entre funcionários.

Em termos de tecnologias de TI a empresa conta apenas com suportes terceirizados que estão ao seu alcance como plataforma Ingaia Locação, uma plataforma destinada justamente a empresas do setor imobiliário, onde são armazenados dados de contrato e auxilia na gestão do mesmo bem como dados financeiros.

A plataforma conta com uma outra interface que é para o gerenciamento de clientes desde a fase de aquisição até o término do contrato, renovação ou até mesmo evasão, porém esta interface praticamente não é utilizada, o que demonstra uma grande subutilização da plataforma. A ferramenta também gera alguns insights que são muito simples para a tomada de decisões estratégicas ou tomadas de decisões de longo prazo como gráficos tipo pizza para percentuais de locação, séries temporais para crescimento ou decréscimo de receita e etc.

Referindo-se à concorrência destacam-se outras imobiliárias de mesmo porte que contam com uma mesma estrutura média em termos de tecnologia, diferenciando-se apenas no tamanho e qualificação de equipes. Com exceção à apenas duas empresas de grande porte, denominadas “unicórnios”, nas quais contam com tecnologias de DW e inteligência artificial que, além de favorecer melhor tomada de decisão contribuem significativamente para a experiência do cliente.

3.1.3 Levantamento de Hipóteses

Com o objetivo de realizar-se uma Análise Exploratória de Dados (AED), segundo (BAILEY; GATRELL, 1995) o levantamento de hipóteses tem o objetivo de ajudar o analista a desenvolver alguns questionamentos sobre o assunto e modelos apropriados para tais dados, principalmente na identificação de padrões, comportamentos, relações e dependências.

Levantou-se entre os membros da equipe, presente na reunião, hipóteses acerca dos atributos referente aos impactos causados pelos mesmos no fenômeno estudado para cada agente respectivamente com o objetivo de validar ou refutar hipóteses através da AED para a criação de novas variáveis necessárias tanto na AED quanto na construção do modelo de BI.

Apesar de levantar-se hipóteses dos atributos de todos os agentes, no primeiro ciclo do CRISP-DM, o projeto contempla a modelagem referente apenas ao agente Imóvel, devido aos dados tabulares disponíveis no sistema da empresa, as hipóteses dos demais agentes foram levantadas apenas à título de planejamento para os próximos ciclos do projeto. Dados referentes aos demais agentes se encontram dispersos em formulários e contratos impressos ou em recursos subutilizados distribuídos entre a plataforma de gestão como por exemplo o *Customer Relationship Management* (CRM) e demais recursos de TI da empresa que exigem ferramentas de preparação de dados com maior complexidade ficando, portanto, à cargo do segundo ciclo do CRISP-DM em diante.

As seguintes hipóteses foram levantadas pela equipe divididas por agentes, sendo o primeiro agente contemplado no primeiro ciclo do projeto:

- 1) Contratos de imóveis com vagas em garagem possuem os menores tempos de vacância (maior parte do tempo ocupados), logo, maior receita para a imobiliária.
- 2) Imóveis com garagem oferecem maior receita para a imobiliária.
- 3) Imóveis com maior tempo de ocupação tendem a gerar maior receita à empresa.
- 4) Imóveis de transferência de carteira têm uma viabilidade econômica melhor para empresa devido à ausência do tempo ocioso (maior receita para a imobiliária).
- 5) Imóveis com menor valor de aluguel tendem a ter uma maior ocupação.
- 6) Imóveis com menor valor de aluguel tendem a oferecer maior receita para a imobiliária.
- 7) Quanto maior o valor do aluguel maior a receita ociosa.
- 8) Imóveis com menores valores de aluguel oferecem menores receitas para a imobiliária.
- 9) Imóveis com mais contratos tendem a ter um tempo de ocupação maior.
- 10) Imóveis comerciais oferecem maior receita à imobiliária.

3.2 Compreensão dos Dados

3.2.1 Fontes dos Dados

Até o momento, levando-se em consideração que o projeto se encontra no primeiro ciclo da metodologia CRISP-DM, os dados para o estudo e implementação do modelo se limitam apenas

aos dados referentes aos imóveis e coordenadas geográficas de pontos estratégicos que podem estar diretamente ligados tanto a preferência dos clientes da empresa quanto ao próprio valor dos imóveis.

Os dados referentes aos imóveis da carteira da empresa foram coletados diretamente da plataforma Ingaia Locação, a plataforma terceirizada de cadastro e controle para a gestão de contratos da empresa, posteriormente adicionadas colunas de dados financeiros referentes à manutenção dos ativos por parte dos proprietários.

Para o início do projeto foram disponibilizados pela empresa e levantados no domínio público os seguintes dados:

Tabela 1 – Descrição dos dados provenientes da plataforma InGaia

VARIÁVEL	TIPO	SIGNIFICADO	INPUT
Contrato	String	Nome do contrato de locação (Abreviação + Código do contrato + Referência local)	não
Endereço	String	Lugadouro com Rua, número e apartamento	não
Cidade	String	Nome da cidade por extenso	não
UF	String	Abreviação da Unidade da Federação (UF)	não
Início	Data	Data de início do contrato de locação na imobiliária	não
Término	Data	Data de término do contrato de locação na imobiliária	não
Tipo	String	Tipo mercadológico do imóvel: Comercial ou Residencial	não
Taxa de Administração (%)	Float	Percentual arrecadado mensalmente pela imobiliária no aluguel do proprietário	não
Taxa de Administração (R\$)	Float	Percentual arrecadado mensalmente pela imobiliária no aluguel do proprietário em reais	não
Aluguel (R\$)	Float	Valor bruto do aluguel (sem desconto da Taxa de admin.)	não
Vaga Garagem	Int	Número de vagas disponível na garagem do imóvel	não
Entrada Contrato	Data	Data a qual o contrato iniciou-se na carteira da imobiliária	sim

Fonte: Elaborado pelo autor (2021)

3.2.2 Construção de *Features*

Para elucidar esta etapa da compreensão dos dados do projeto, deve-se levar em consideração dois principais conceitos da mineração de dados: *Feature* e *Feature Engineering*. Segundo (CHAPMAN; HALL, 2018) uma *Feature* para projetos tanto para o desenvolvimento de algoritmos de *Machine Learning* quanto para projetos de mineração de dados e *Data Analytics*, uma *Feature* representa um atributo ou variável para descrever aspectos individuais dos dados de uma base de dados, ou seja, todas as características que um objeto ou atributo da base carrega consigo (nome, tipo, dimensão e unidade). Ainda, segundo o mesmo autor, *Feature Engineering* representa uma definição mais abrangente.

O autor divide a definição da *Feature Engineering* em seis diferentes tópicos: *Feature Transformation*, *Feature Generation*, *Feature Selection*, *Feature Analysis and Evaluation*, *General Automatic Feature Engineering* e *Feature Engineering Application*. Para esta etapa do projeto

utilizou-se apenas da definição do primeiro tópico *Feature Transformation*, que significa a construção de uma nova Feature à partir de uma previamente existente na base de dados. A forma de construção depende da ferramenta utilizada no projeto, no caso da ferramenta *Power BI* da desenvolvedora Microsoft utilizada no projeto, novas Features foram criadas a partir das Features a priori através da construção de fórmulas, cuja sintaxe pertence ao *Power BI*, conhecida como *Data Analysis Expressions* (DAX) e no ambiente de preparação de dados da ferramenta *Power Query*.

Com base nas Features a priori da base de dados referente aos imóveis extraídos da plataforma Ingaia Locação e com o processo de transformação de Features descrito pelo autor têm-se as seguintes Features construídas:

Tabela 2 – Descrição das variáveis (Features) derivadas das originais

VARIÁVEL	TIPO	SIGNIFICADO	OPERAÇÃO	FÓRMULA DAX
Lugar	String	O prefixo do nome do contrato da base original, abreviação do imóvel	Transformação tipo Segregação em Power Query	Power Query
Número_Contrato	Int	Sufixo do nome do contrato original, qual ordem de contrato o imóvel se encontra	Transformação tipo Segregação em Power Query	Power Query
Tempo_ocupado (dias)	Int	Número de dias em que o imóvel permaneceu ocupado	Transformação no Power Query com Python*	Power Query
Tempo_vago (dias)	Int	Número de dias em que o imóvel permaneceu vago	Transformação no Power Query com Python*	Power Query
Tempo_total (dias)	Int	Número de dias totais em que o imóvel encontra-se na imobiliária	Coluna DAX	Tempo_total_dias = lista_imoveis[Tempo_ocupado] + lista_imoveis[Tempo_vago]
Portabilidade	String	Imóveis com números em dias Negativos em que o imóvel se encontra na imobiliária (Entrada na imobiliária posterior a data de início)	Transformação no Power Query com Python*	Power Query
Receita_imobiliária (R\$)	Float	Receita obtida pela imobiliária com imóvel ocupado	Coluna DAX	Receita_imobiliaria = lista_imoveis[Taxa adm]*(lista_imoveis[Tempo_ocupado]/30)
Receita_imobiliária_ociosa (R\$)	Float	Receita não obtida pela imobiliária devido ao tempo ocioso	Coluna DAX	Receita_imobiliaria = lista_imoveis[Taxa adm]*(lista_imoveis[Tempo_ocupado]/30)
Receita_locador (R\$)	Float	Receita obtida pelo locador com o imóvel ocupado	Coluna DAX	Receita_locador = (lista_imoveis[Aluguel] - lista_imoveis[Taxa adm])*(lista_imoveis[Tempo_ocupado]/30)
Receita_locador_ociosa (R\$)	Float	Receita não obtida pelo locador devido ao tempo ocioso	Coluna DAX	Receita_locador_ociosa = (lista_imoveis[Aluguel] - lista_imoveis[Taxa adm])*(lista_imoveis[Tempo_vago]/30)
Trim_inicio_contrato	String	Trimestre do início de cada contrato (Trimestre-Ano)	Coluna DAX	Trim_inicio_contrato = FORMAT(lista_imoveis[Início], "TQ-YYYY")
Ano_contrato	String	Ano no qual o contrato foi realizado	Coluna DAX	Ano_contrato = YEAR(lista_imoveis[Início])

Fonte: Elaborado pelo autor (2021)

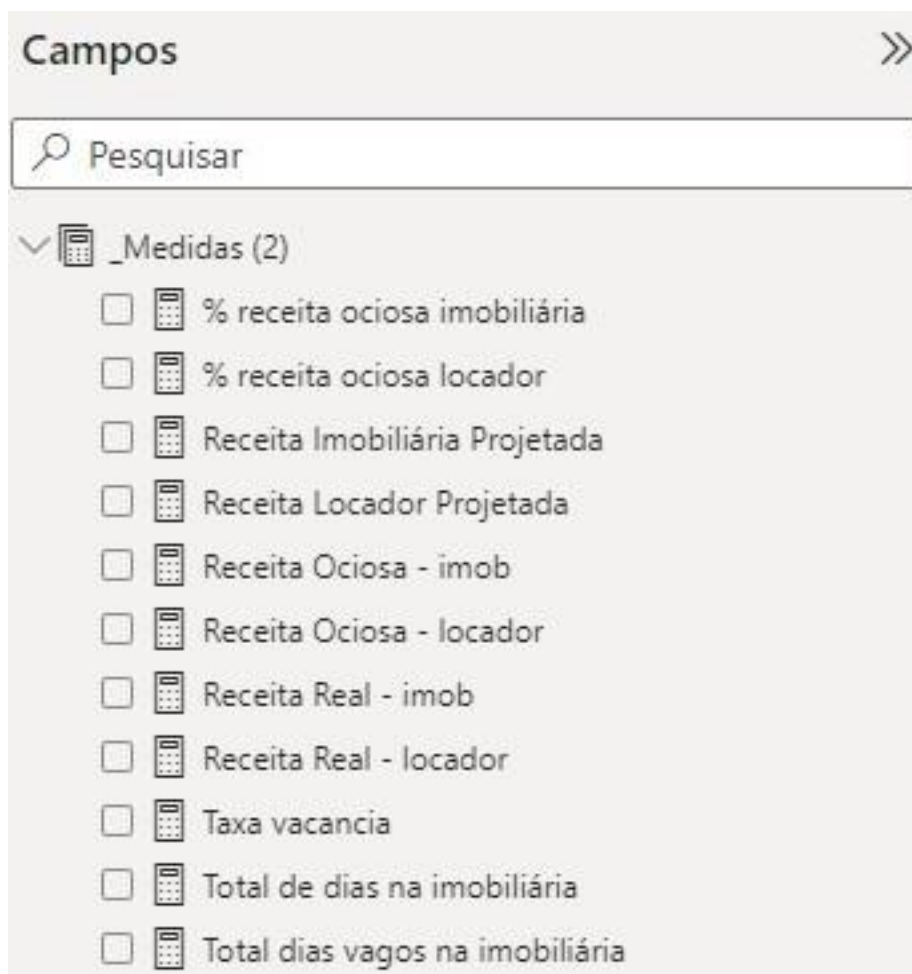
3.2.3 Medidas

Para uma análise *Ad Hoc*, isto é, análise personalizada especificamente para o negócio do cliente, é disponibilizado como um dos recursos do *Power BI* a ferramenta de cálculo de forma

sumarizada conhecido como Medidas. Utiliza-se este recurso em operações resumidas para cálculo responsivo a fim de potencializar o desempenho do modelo em que, as interações entre usuário e dashboard faça-se mais dinâmica e interativa, no qual se oferece ao usuário uma melhor usabilidade e resposta do modelo e melhor otimização dos recursos computacionais.

Nesta etapa do trabalho, visando-se a interação entre alguns visuais nos quais permitem o uso de métricas sumarizadas em funções da linguagem DAX, criou-se as seguintes medidas:

Figura 5 – Conjunto de medidas



Fonte: Elaborado pelo autor (2021)

Um resumo de cada medida aplicada é mostrado na tabela a seguir, na qual todos respectivos significados justificados de acordo com regras do negócio verificadas pela equipe, com sua respectiva finalidade e respectivas fórmulas DAX implementadas.

Tabela 3 – Descrição de medidas

MEDIDA	SIGNIFICADO	FINALIDADE	FÓRMULA DAX
% receita ociosa imobiliária	Percentual da receita na qual a imobiliária deixou de arrecadar com tempo vago	Cálculo de KPI	% receita ociosa imobiliária = SUM(lista_imoveis[Receita_imobiliaria_ociosa]) / (SUM(lista_imoveis[Receita_imobiliaria]) + SUM(lista_imoveis[Receita_imobiliaria_ociosa]))
% receita ociosa locador	Percentual da receita na qual o locador deixou de arrecadar com tempo vago	Cálculo de KPI	% receita ociosa locador = SUM(lista_imoveis[Receita_locador_ociosa]) / (SUM(lista_imoveis[Receita_locador]) + SUM(lista_imoveis[Receita_locador_ociosa]))
Receita Ociosa - locador	Receita na qual o locador deixou de arrecadar com tempo vago	Cálculo de cartões indicadores	Receita Ociosa - locador = SUM(lista_imoveis[Receita_locador_ociosa])
Receita Real - locador	Receita arrecadada pelo locador	Cálculo de cartões indicadores	Receita Real - locador = SUM(lista_imoveis[Receita_locador])
Receita Ociosa - Imob.	Receita na qual a imobiliária deixou de arrecadar com tempo vago	Cálculo de cartões indicadores e desempenho de contrato	Receita Ociosa - imob = SUM(lista_imoveis[Receita_imobiliaria_ociosa])
Receita Real - Imob	Receita arrecadada pela imobiliária	Cálculo de cartões indicadores e desempenho de contrato	Receita Real - locador = SUM(lista_imoveis[Receita_locador])
Receita Imobiliária Projetada	Receita da imobiliária caso o imóvel possuísse 100% do tempo de ocupação	Cálculo de cartões indicadores e receita por tipo de imóvel	Receita Imobiliária Projetada = [Receita Real - imob] + [Receita Ociosa - imob]
Receita Locador Projetada	Receita do locador caso o imóvel possuísse 100% do tempo de ocupação	Cálculo de cartões indicadores	Receita Locador Projetada = [Receita Real - locador] + [Receita Ociosa - locador]
Taxa vacância	Tempo vago em relação ao tempo total do imóvel na imobiliária	Cálculo de KPI	Taxa vacância = [Total dias vagos na imobiliária]/[Total de dias na imobiliária]
Total de dias na imobiliária	Somatório dos dias em que o imóvel se encontra na carteira da imobiliária	Cálculo de cartões indicadores	Total de dias na imobiliária = SUM(lista_imoveis[Tempo_total_dias])
Total de dias vagos na imobiliária	Somatório de todos os dias no qual o imóvel permaneceu vago durante todo o tempo na imobiliária	Cálculo de cartões indicadores	Total dias vagos na imobiliária = SUM(lista_imoveis[Tempo_vago])

Fonte: Elaborado pelo Autor (2021)

3.3 Preparação dos Dados

3.3.1 Processo *Extraction, Transformation and Load* (ETL)

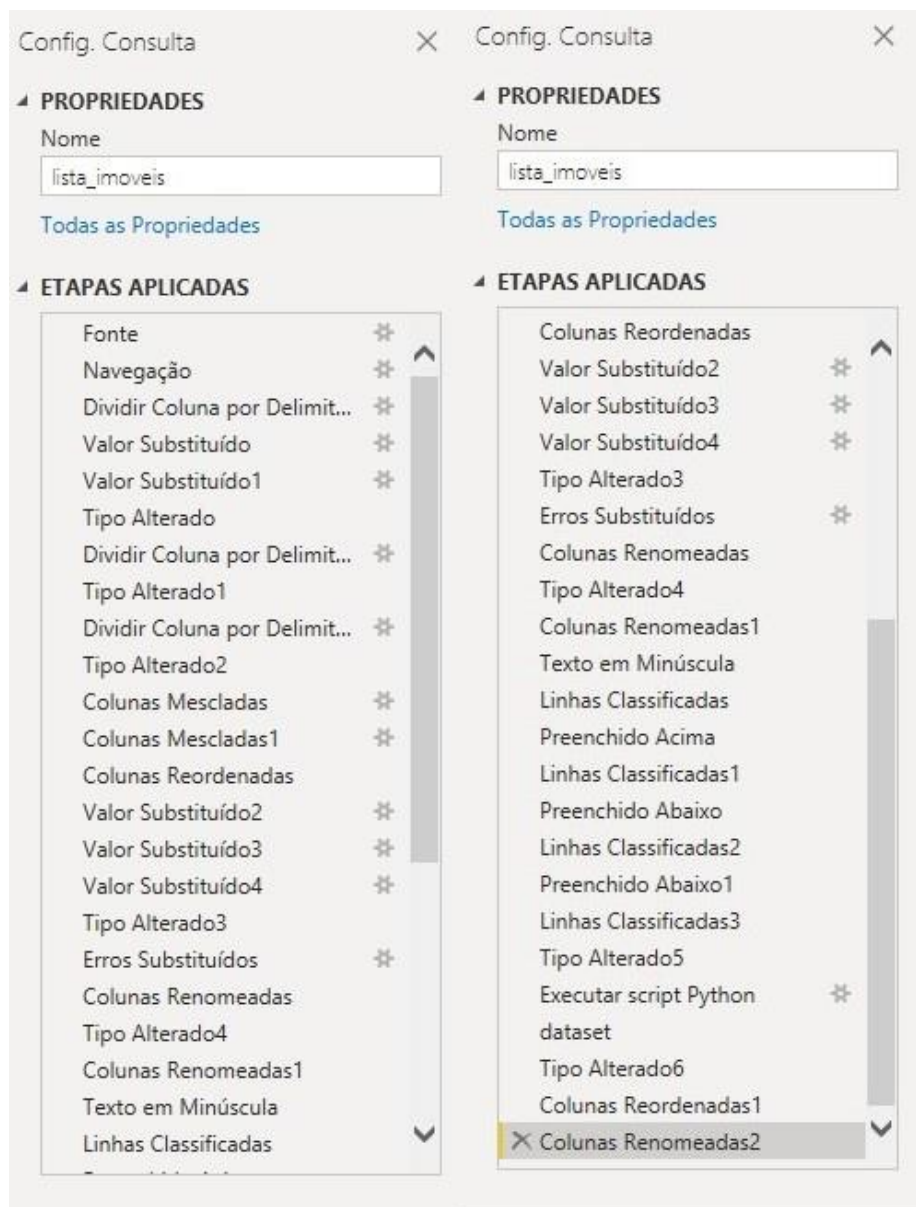
Em um projeto de BI, assim como qualquer projeto de dados em seus diversos níveis de complexidade, para a execução das análises os dados devem estar integrados e consolidados em um data warehouse. Esse processo é essencial pois, dados relacionados à diversas áreas do negócio provém de diversas fontes, incluindo-se plataformas de gestão, ferramentas gerenciais como CRM, *Enterprise Resource Planning* (ERP) e dados provenientes de domínios externos como a internet através de *Web Scraping*, técnica na qual consiste na rotina da obtenção de dados de forma automatizada por meio de algoritmos.

A prática para a ETL responsabiliza-se pela alimentação do DW com dados compatíveis ao projeto e desenvolvimento de modelos BI. Segundo Sharda *et al.* (2019), o processo consiste na extração, isto é, leitura de dados de uma ou mais base de dados, transformação que é a conversão dos dados

de uma ou mais base de dados extraídos de seu formato prévio para o formato ideal do ponto de vista do negócio para inserção no DW e finalmente, a carga, o processo de inserção dos dados transformados para dentro do DW. Uma importante observação é com relação a dados numéricos, que necessitam estar em conformidade com caracteres, unidades de medidas e principalmente com as configurações regionais dos respectivos interpretadores da ferramenta. Inconsistências desses requisitos podem levar ao enviesamento de resultados ou mesmo mau funcionamento do modelo. Ainda segundo o mesmo autor, este processo representa o maior desafio para as equipes de TI, pois é um processo que consome cerca de 70% do tempo de um projeto de mineração de dados. Tamanha complexidade se deve a necessidade de se obterem dados de qualidade para o projeto, pois o sucesso do resultado final do projeto pode ser prejudicado por mal conformidades e inconsistências na base de dados.

Neste trabalho, devido ao fato de não haver uma base de dados consolidada como uma estrutura de DW, todo o processo de ETL e consolidação da base de dados para o estudo foi construída na versão desktop da ferramenta *Power BI*, nos ambientes de transformação de dados *Power Query* e em tabelas por meio de fórmulas DAX. Os processos de *feature engineering* descritos anteriormente também fazem parte desta etapa do processo conforme as descrições na tabela 2 da etapa de construção de *features*. A figura a seguir representa os passos para a transformação da base de dados preparada para análise e construção do modelo.

Figura 6 – Passos para a transformação da base em Power Query



Fonte: Elaborado pelo autor (2021)

Apesar da versatilidade do recurso do *Power Query*, para a transformação e adequação da base de dados da fonte no conjunto de dados ideal para o desenvolvimento do projeto, a interface pode-se apresentar com algumas limitações nas quais o set de ferramentas para extração e transformação de dados não seja suficientemente eficiente, necessitando-se de recursos sobressalentes mais flexíveis e adaptáveis às necessidades da proposta. A fim de suprir-se as deficiências do conjunto de métodos de transformação oferecidos pela ferramenta, no *Power BI* conta-se, tanto no ambiente *Power Query* quanto na construção de visuais, interfaces internas para linguagens de programação dedicadas à exploração de dados e análises científicas: *R* e *Python*.

Conforme destacado tabela 2 da etapa de construção de features e o motivo ao qual esta etapa

está diretamente relacionada à transformação de tabelas, para criação de colunas de manipulação de variáveis do tipo data para análise de inteligência do tempo, utilizou-se a linguagem *Python* cálculo e criação das colunas cujas unidades mensuram-se em dias e os respectivos resultados dependem da interação da diferenças de datas entre diferentes contratos de um mesmo ativo: tempo ocupado, tempo vago e portabilidade para imóveis provenientes da transferência de carteira de outras corretoras. Para este tipo de transformação, cujo conjunto de ferramentas do *Power Query* apresentou-se limitações em termos de grau de liberdade para tal execução. Contudo, conforme o Anexo A, elaborou-se dentro do ambiente *Power BI*, o algoritmo na linguagem *Python* para a série de transformações.

3.4 Modelagem

3.4.1 Análise Exploratória dos Dados (AED)

Para uma maior compreensão a respeito do negócio, isto é, além do presente conhecimento do estado da técnica das métricas de negócio, o objetivo da AED é proporcionar a equipe conhecimentos tácitos a respeito do empreendimento e elucidar novas estratégias de negócio que são trazidas à tona através de insights gerados nesta etapa. De acordo com Lopes *et al.* (2019) a AED busca aumentar o conhecimento do pesquisador sobre a população a partir de uma amostra. Desta forma, pode-se descrever a AED como um conjunto de métodos adequados para a coleta, a exploração, descrição e interpretação de conjunto de dados numéricos. Ainda segundo os autores, estes métodos permitem a exploração dos dados com o intuito de identificar padrões de interesse e a representação dos dados caracterizados por estes padrões.

De acordo com (PAULINO; SINGER, 2006) a análise de dados categóricos é uma parte da análise multivariada, que interpreta a informação que está contida em dados discretos resultantes de contagens de eventos ou de unidades (pessoas, lugares, objetos etc.) possuindo certas características ou atributos definidos pela combinação das categorias de duas ou mais variáveis de interesse ou apenas categorias de uma variável. Com base nesta última argumentação e ampliando-se o escopo argumentativo de Lopes *et al.* (2019), a AED desenvolvida nesta etapa do trabalho conta com dados categóricos para a extração de *insights* relevantes ao negócio.

O levantamento de hipóteses junto a equipe de colaboradores da empresa, conforme descrito na etapa de conhecimento do negócio, além do direcionamento na extração de variáveis derivadas das variáveis originais também direciona o processo de AED no qual métricas para monitoramento do negócio e novos insights são levantados em seus respectivos testes.

A ferramenta escolhida para a AED no projeto é o *Python*, uma linguagem de programação de propósito geral, tanto para projetos de análise de dados, ciência de dados, desenvolvimento *web*

e de software. A escolha desta linguagem se deve à facilidade em produzir análises robustas com bibliotecas próprias específicas para análise de dados como *Pandas*, *Matplotlib* e *Seaborn* através da *Anaconda*, uma distribuição de código aberto que permite a execução de aplicações científicas da linguagem com o ambiente *Jupyter Notebook*.

3.4.2 Correlações

Para o conhecimento do negócio, é de suma importância saber como as variáveis relacionam-se entre si e, principalmente, com a variável resposta. Realizou-se no trabalho a análise de correlação de *Pearson*, método dedicado para análise de correlação de variáveis numéricas.

Para (JOHNSON; WICHERN, 2007) conforme citado por Lordelo *et al.* (2018) a matriz de correlação é utilizada para iniciar a análise estatística dos dados históricos dos diversos universos estudados, identificando-se visualmente as variáveis envolvidas no estudo que se relacionam-se entre si. A determinação do grau de relação entre duas variáveis é dada pelo coeficiente de *Pearson*, também chamado de coeficiente de correlação, ou simplesmente correlação para os pares de variáveis. Esse coeficiente é um número entre -1 e 1 que expressa o grau de dependência linear entre duas variáveis quantitativas, que quando negativa indica que uma variável diminui com o aumento da outra, e positiva quando uma variável aumenta com o aumento da outra.

Para duas variáveis quaisquer (x, y) sendo testadas, calcula-se o coeficiente de correlação de *Pearson*, como segue (HAIR *et al.*, 2005; JOHNSON; WICHERN, 2007). :

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (1)$$

3.4.3 Insights

Com origens no campo de estudo da psicanálise, a palavra *Insight* é amplamente empregada no campo da estatística e análise de dados como referência ao reconhecimento de um novo padrão encontrado em meio aos dados outrora desconhecido.

“a capacidade de entender verdades escondidas etc., especialmente de caráter ou situação” (ALLEN, 1990).

“o ato ou o resultado de alcançar a íntima ou oculta natureza das coisas ou perceber de maneira intuitiva” (MIFFLIN, 1994).

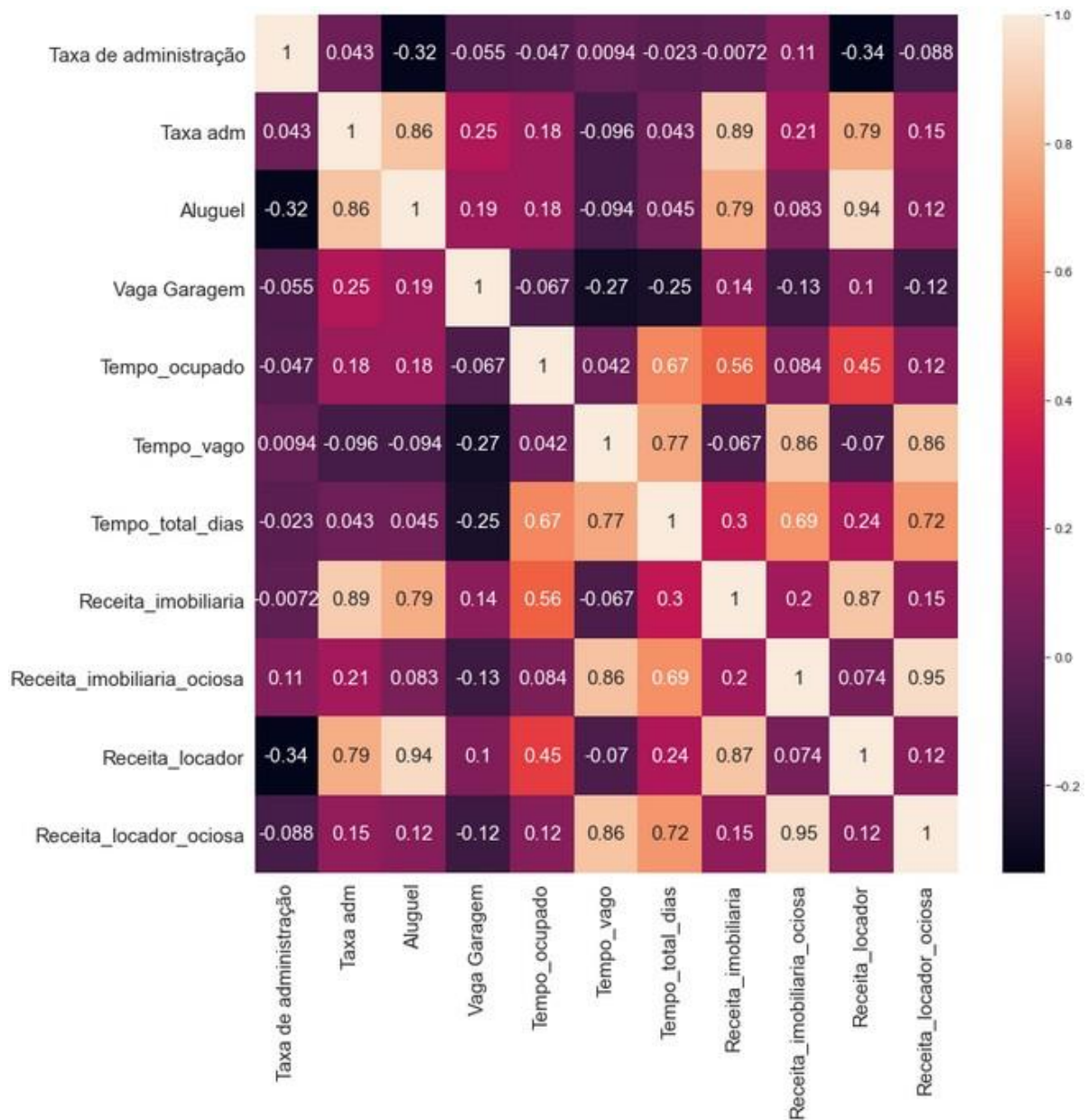
Em referência às citações de ambos os autores, mesmo egressos do campo da psicanálise, o significado de Insight no presente trabalho mostra-se pertinente ao fato de trazer à mostra informações de importância estratégica para o empreendimento. Por meio do teste das hipóteses levantadas na etapa de compreensão do negócio com análise bivariada pôde-se levantar informações significativas, a priori desconhecidas, para a composição do conjunto de visualizações do painel onde novas métricas do negócio serão monitoradas para a tomada de decisão na gestão de contratos dos imóveis da corretora.

3.4.4 Testes e Resultados

Aplicando-se os conhecimentos mencionados nas etapas 3.4.2. e 3.4.3. e colocando-se em prática as bibliotecas científicas disponíveis no *Python*, conforme mencionado na etapa 3.4.1., realizou-se uma sequência de testes referentes às dez hipóteses levantadas a priori da fase de problemas do negócio.

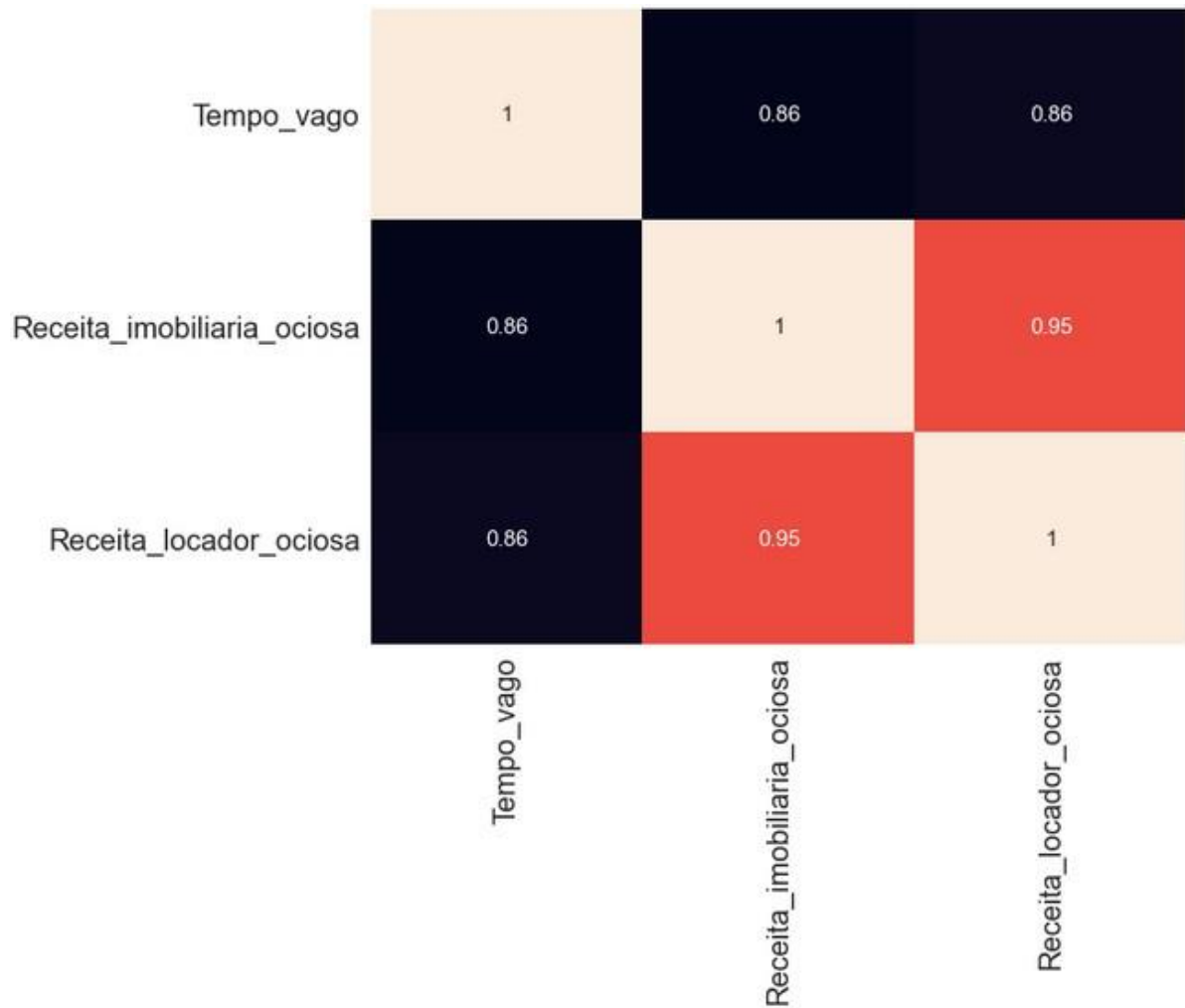
Através da análise multivariada e utilizando-se dos recursos do *Python*, realizou-se uma análise de correlação entre as variáveis numéricas para a identificação das relações as quais as variáveis possuem entre si e, principalmente, a relação da variável resposta receita imobiliária com as demais buscando encontrar atributos que possuem correlações próximas do valor 1.

Gráfico 2 – Correlação de Pearson - Análise Multivariada



Fonte: Elaborado pelo autor (2021)

As correlações levantadas no teste têm o objetivo de identificar variáveis que, além de estarem relacionadas à variável resposta de forma retórica, apresentam altos graus de dependência direta com a mesma confirmada pelo teste como mais próximas do valor 1, para a identificação e elaboração de KPI's em potencial. Feito isso, identificou-se duas variáveis com respectivos graus mais próximos do valor máximo, logo estabelecidas como indicador de desempenho do imóvel na gestão de contratos através do dashboard em *Power BI*.

Gráfico 3 – Correlação de *Pearson* para medidas de desempenho

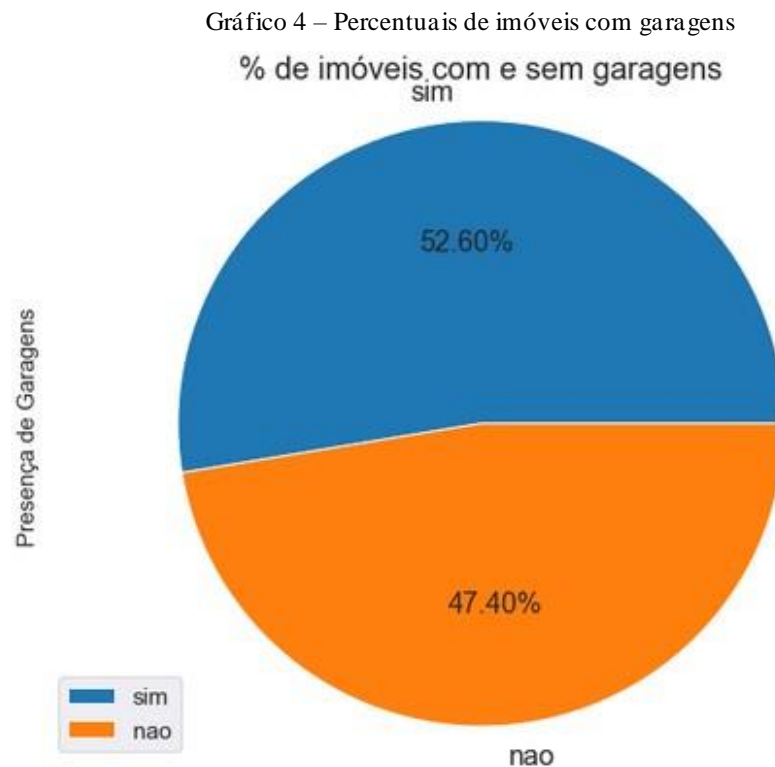
Fonte: Elaborado pelo autor (2021)

A correlação não faz qualquer pressuposição a priori sobre uma variável ser ou não dependente de outra(s) e não está preocupada com o tipo de relação entre variáveis; na verdade, ela oferece uma mera estimativa sobre o grau de associação entre elas. Por sua vez, a regressão busca descrever a dependência de uma variável de resposta em relação a uma ou mais variáveis explanatórias, assumindo implicitamente que há uma relação causal apontando na direção das variáveis explanatórias para a variável de resposta, quer a trajetória do efeito seja direta ou indireta (SHARDA *et al.*, 2019). Com base na argumentação do autor e o teste realizado selecionou-se os KPI's.

Escolheu-se para os KPI's, por fim, o visual do tipo gráfico de velocímetro para a composição do dashboard. A escolha deu-se, em consenso da equipe, pela sua facilidade de visualização e interpretação.

Para os testes de hipóteses, realizou-se um conjunto de análises bivariadas com o *Python* e suas bibliotecas para análise de dados anteriormente citadas serviram tanto para afirmar algumas hipóteses de métricas anteriormente pressupostas quanto negar outras.

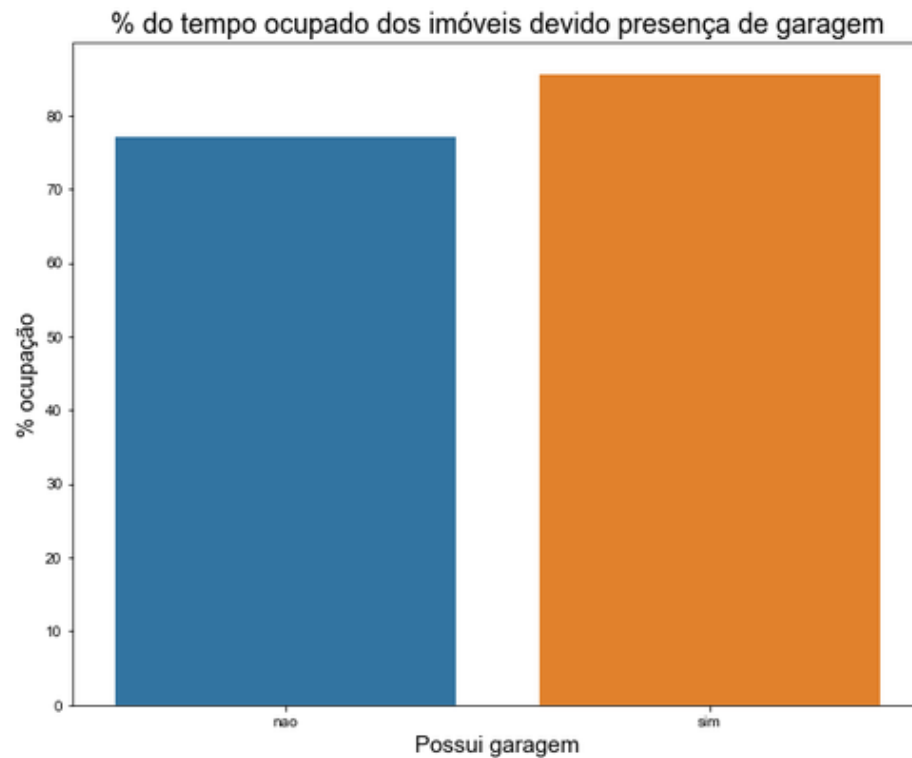
- 1) Contratos de imóveis com vagas em garagem possuem os menores tempos de vacância (maior parte do tempo ocupados), logo, maior receita para a imobiliária.



Fonte: Elaborado pelo autor (2021)

Neste gráfico tipo pizza demonstra-se que a quantidade de imóveis com garagens superam-se a quantidade de imóveis sem garagens a uma diferença de 5,2%.

Gráfico 5 – Gráfico de barras com percentuais de tempo ocupado por cada tipo de presença de garagens

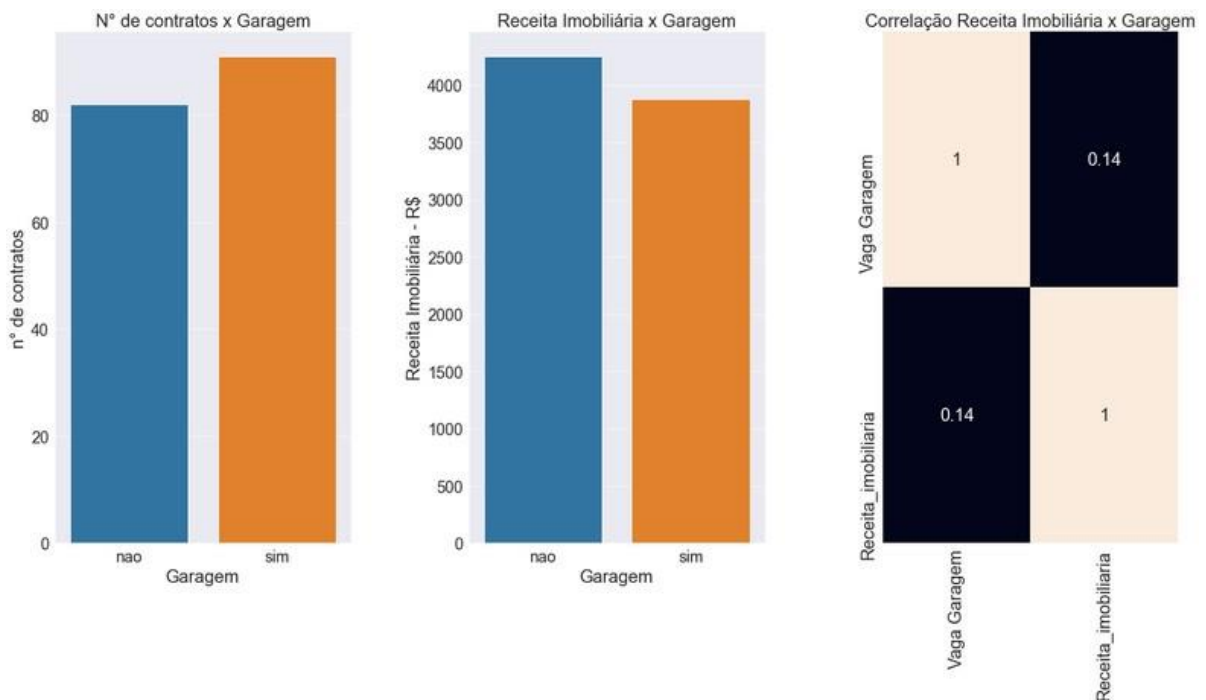


Fonte: Elaborado pelo autor (2021)

O resultado para este teste mostrou-se afirmativo. Analisando-se em percentuais, imóveis com garagem têm um percentual de ocupação maior em relação ao tempo total que o imóvel se encontra na carteira da imobiliária.

2) Imóveis com garagem oferecem maior receita para a imobiliária.

Gráfico 6 – Teste hipótese 2



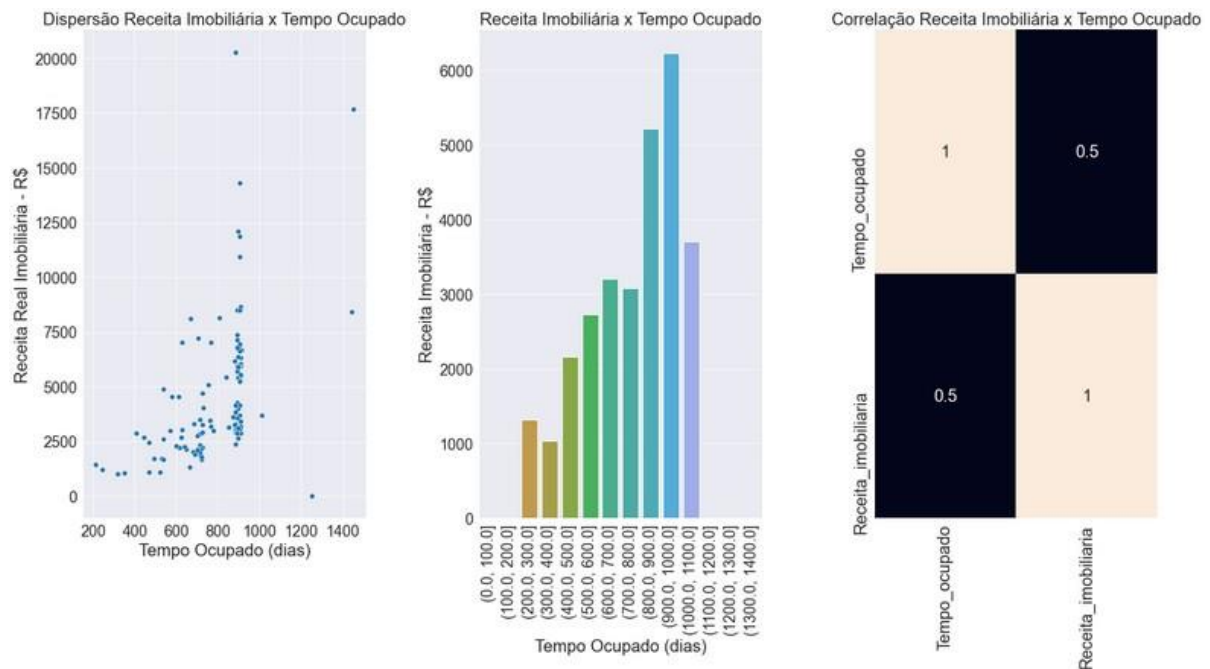
Fonte: Elaborado pelo autor (2021)

O resultado para este teste mostrou-se negativo. Apesar de na hipótese anterior confirmar-se que imóveis com garagem permanecem por mais tempo ocupados e haver correlação significativa entre receita e tempo ocupado, neste teste pode-se observar que mesmo havendo um número maior de contratos de imóveis com garagem, a maior receita da imobiliária é garantida por imóveis que não possuem garagens. Adiante, observa-se correlação significativamente baixa entre a quantidade de garagem e a receita da imobiliária.

Este comportamento pode ser explicado por outros fatores inerentes ao imóvel que não necessariamente seja a presença de garagem. Fatores como localização ou estrutura interna dos imóveis podem ser mais importantes que a presença de garagem.

3) Imóveis com maior tempo de ocupação tendem a gerar maior receita à empresa.

Gráfico 7 – Teste hipótese 3

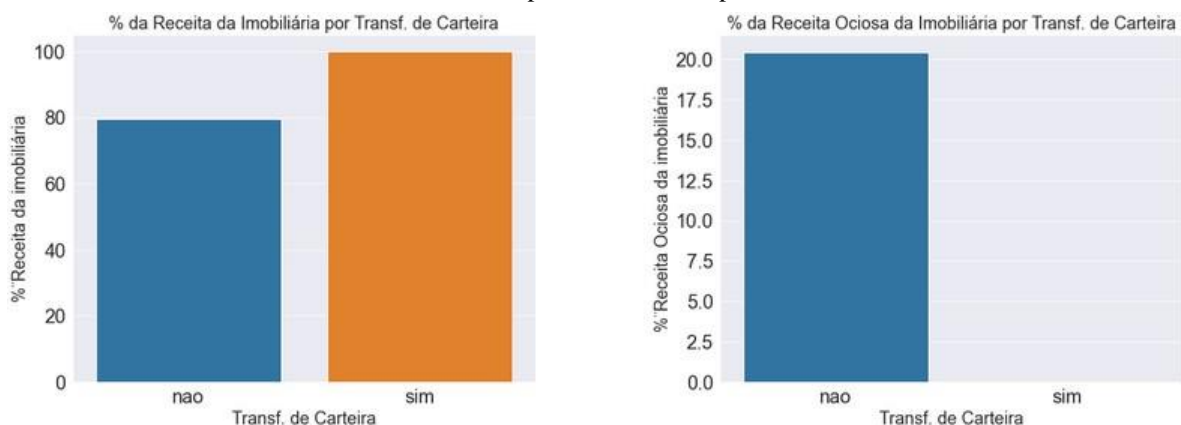


Fonte: Elaborado pelo Autor (2021)

O resultado para este teste mostrou-se afirmativo. Apesar de uma correlação intermediária, percebe-se visualmente que um maior tempo de ocupação do imóvel oferece sim uma maior receita para a empresa.

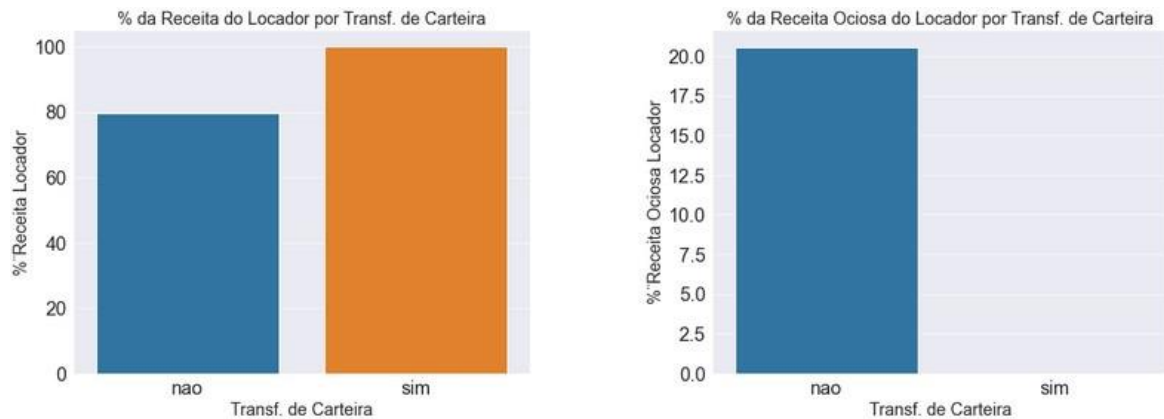
4) Imóveis de transferência de carteira têm uma viabilidade econômica melhor para empresa devido à ausência do tempo ocioso (maior receita para a imobiliária).

Gráfico 8 – Teste hipótese 4: receitas para a imobiliária



Fonte: Elaborado pelo autor (2021)

Gráfico 9 – Teste Hipótese 4: receitas para o locador

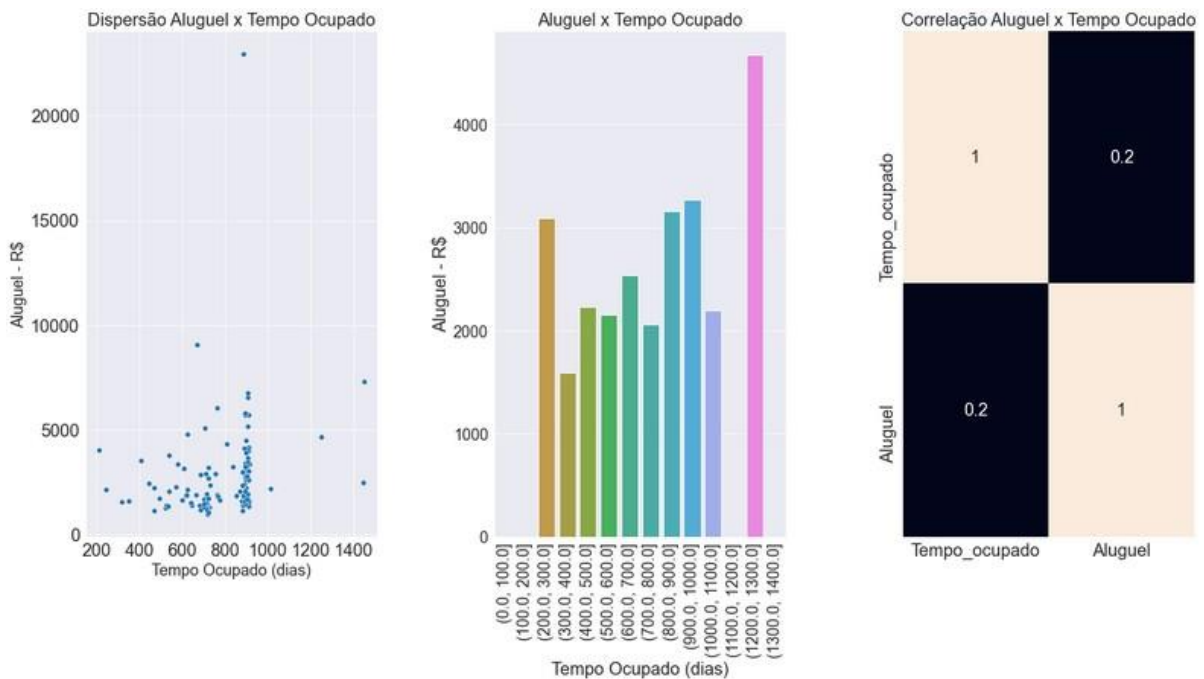


Fonte: Elaborado pelo autor (2021)

O resultado para este teste mostrou-se afirmativo. O percentual das receitas arrecadadas tanto por parte da imobiliária quanto para os locadores atingem 100% para os imóveis provenientes de transferência de carteira e, análogamente, em receita ociosa para ambas as partes os mesmos imóveis têm 0% de receita ociosa.

5) Imóveis com menor valor de aluguel tendem a ter uma maior ocupação.

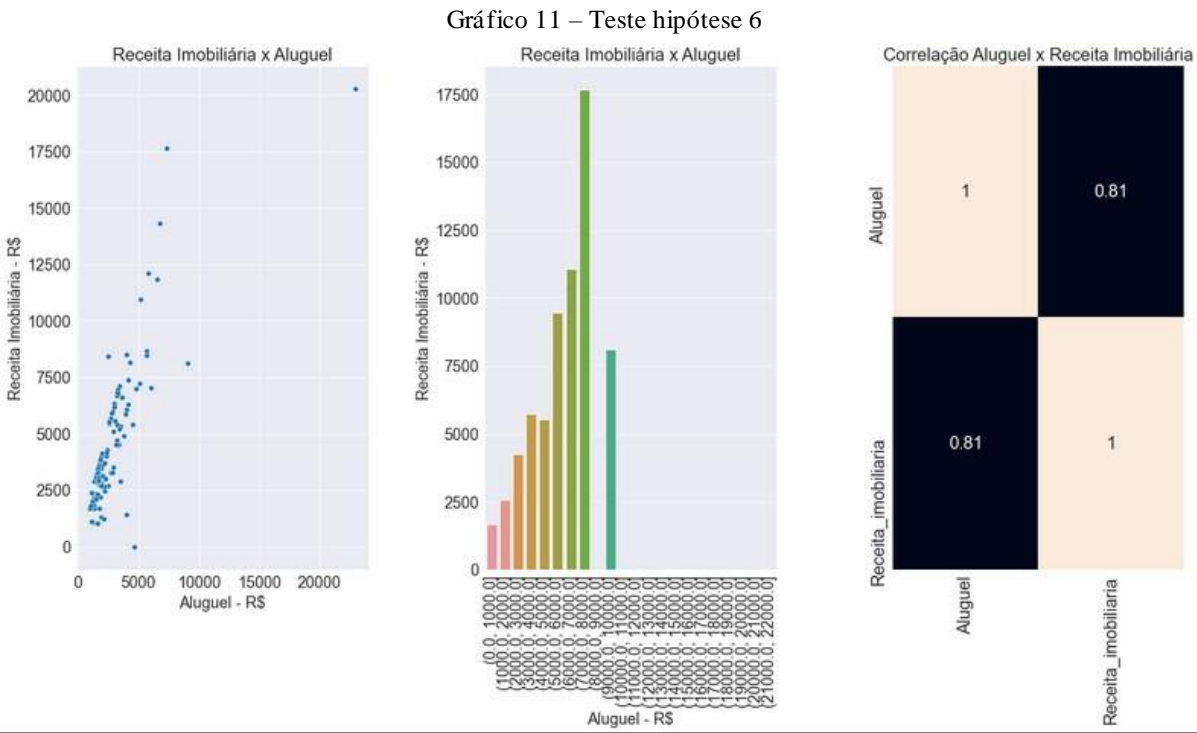
Gráfico 10 – Teste hipótese 5



Fonte: Elaborado pelo autor (2021)

O resultado para este teste mostrou-se negativo. Além de apresentar-se com baixa correlação, há uma diferença pouco significativa nas barras porém com uma leve tendência de maior tempo ocupado com maior valor de aluguel. Tais tendências do teste mostram que o preço dos imóveis podem não representar um único fator decisivo na preferência do cliente.

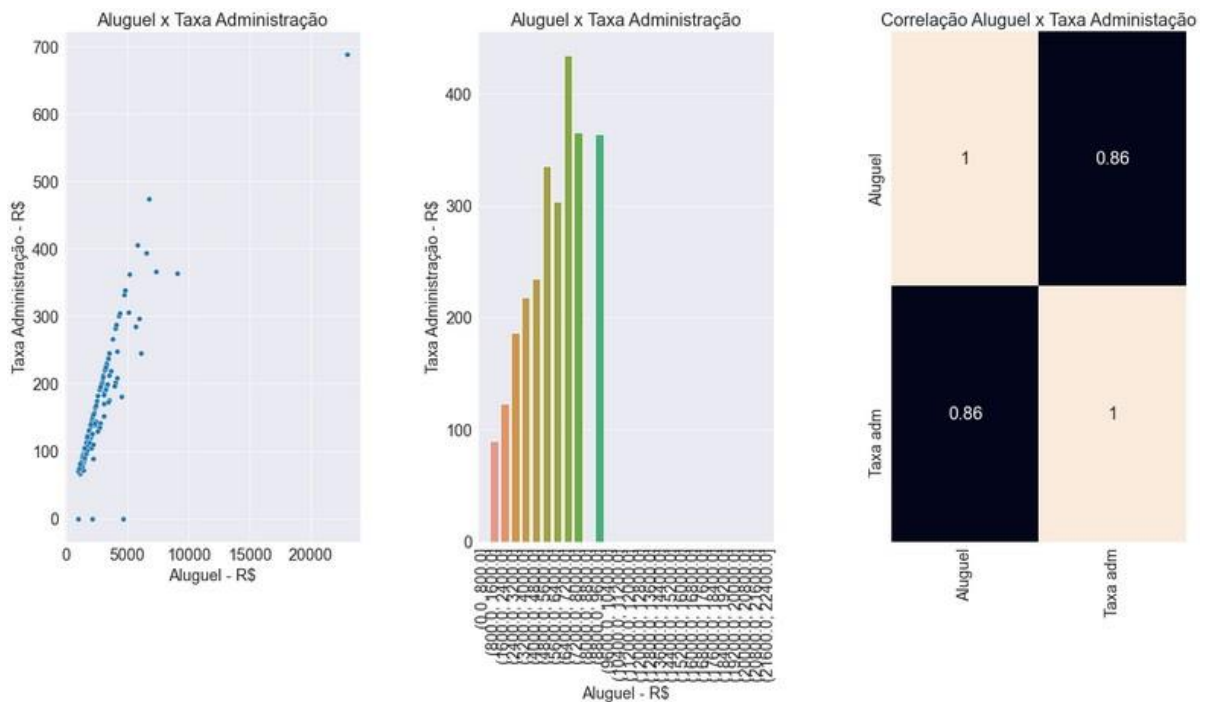
6) Imóveis com menor valor de aluguel tendem a oferecer maior receita para a imobiliária.



Fonte: Elaborado pelo autor (2021)

6.1. Imóveis com maior valor de aluguel possuem maiores taxas de administração

Figura 7 – Teste hipótese auxiliar 6.1



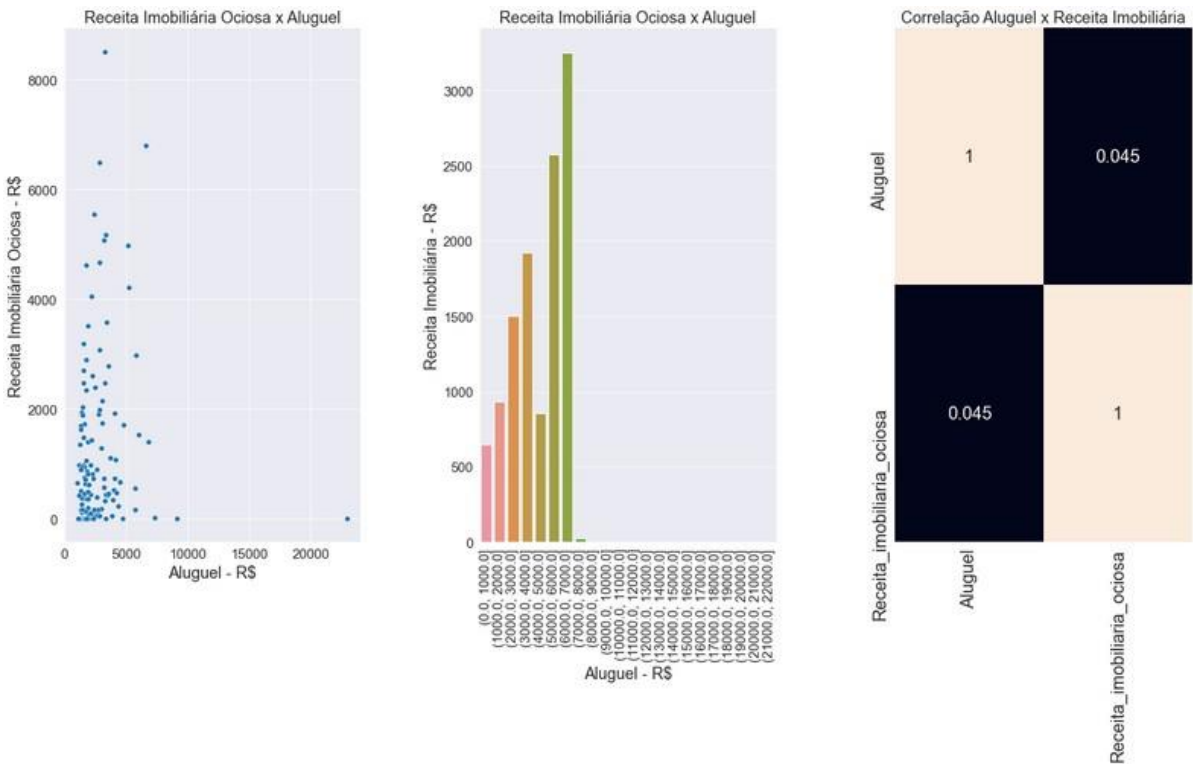
O resultado para este teste mostrou-se negativo: Imóveis com menores valores de aluguel, diferentemente do que supunha-se devido ao fato de, conforme a hipótese 5, acreditar que imóveis de menor valor permanecem por mais tempo ocupados devido a uma maior procura.

Tendo-se em vista que há uma correlação significativamente grande entre receita e tempo de ocupação, foi observado ainda na hipótese 5 que a ocupação de imóveis com maior valor de aluguel são maiores, incluindo-se o fato, provado por meio da análise de dados que, apesar de óbvio, o valor do aluguel é uma grandeza diretamente relacionada à taxa de administração arrecadada pela imobiliária.

Em resumo, imóveis com maior valor de aluguel além de possuírem maior ocupação e maior procura, oferecem uma arrecadação proporcionalmente maior em termos de taxa de administração conforme hipótese auxiliar 6.1. e, conseqüentemente, maior receita à imobiliária.

7) Quanto maior o valor do aluguel maior a receita ociosa.

Gráfico 12 – Teste hipótese 7

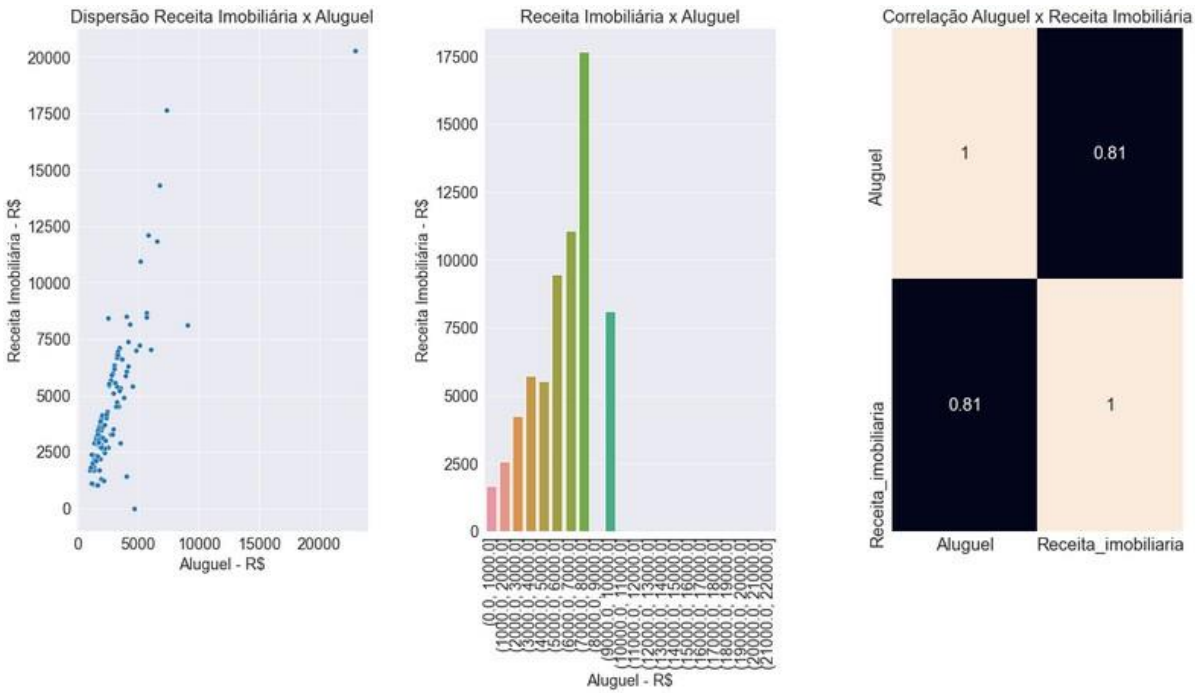


Fonte: Elaborado pelo autor(2021)

O resultado para este teste mostrou-se negativo: Análogamente à hipótese de nº 5 na qual constatou-se que imóveis com menor valor de aluguel tendem a oferecer maior receita para a imobiliária. Neste teste, de forma oposta, pôde-se concluir que imóveis com maior valor de aluguel tendem a ter mais procura conforme já avaliado anteriormente, e consequentemente menor receita ociosa.

8) Imóveis com menores valores de aluguel oferecem menores receitas para a imobiliária.

Gráfico 13 – Teste hipótese 8

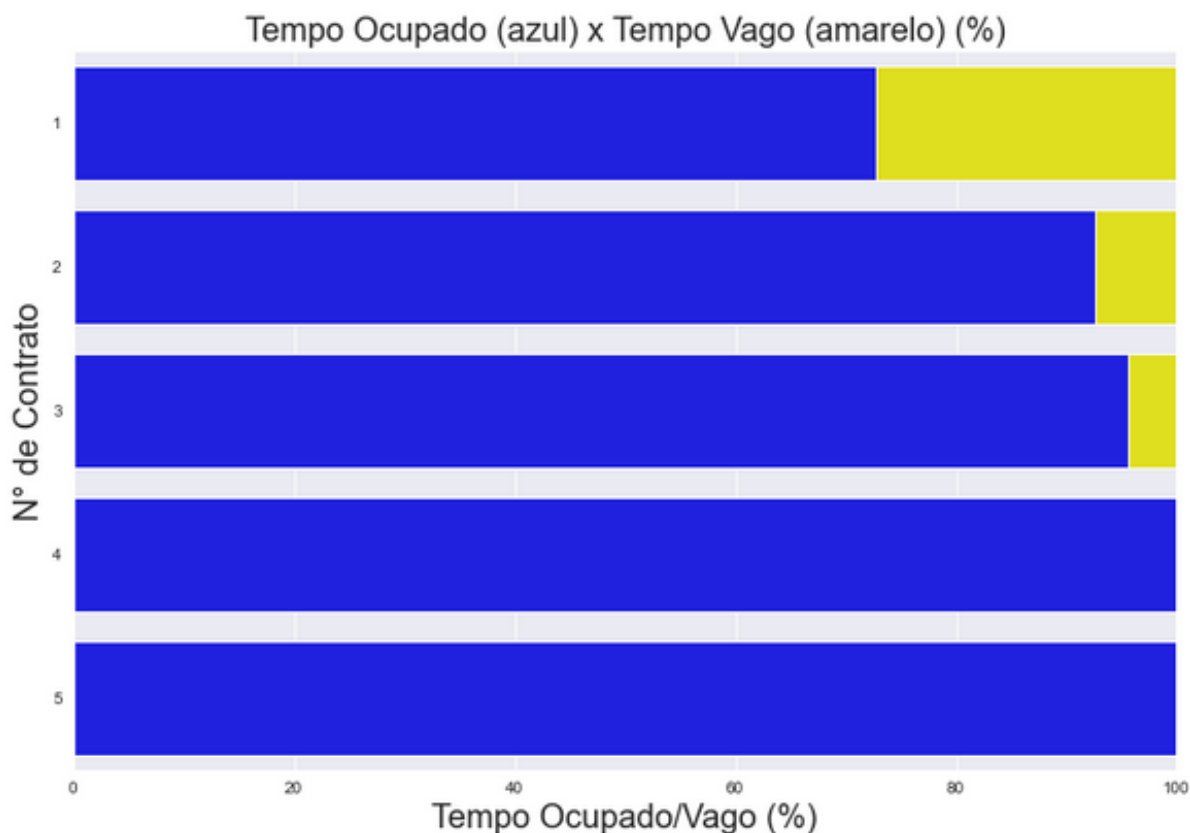


Fonte: Elaborado pelo autor (2021)

O resultado para este teste mostrou-se afirmativo. Análogamente à hipótese de nº 4 que foi constatado negativo para hipótese de que menores valores de aluguel oferecem maiores receitas à imobiliária, neste pode-se observar um comportamento de crescimento proporcional na relação entre aluguel e receita, quanto maior o valor do aluguel maior a receita para a imobiliária.

9) Imóveis com mais contratos tendem a ter um tempo de ocupação maior.

Gráfico 14 – Teste hipótese 9

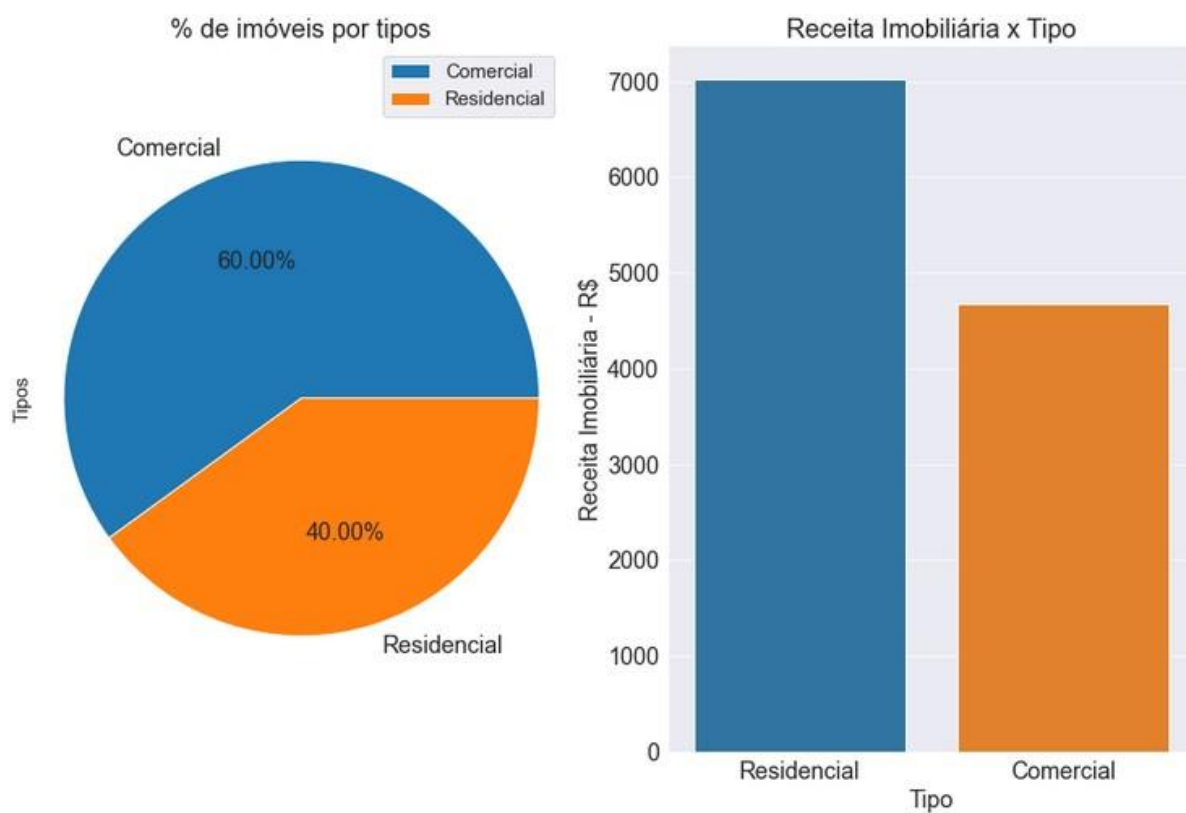


Fonte: Elaborado pelo autor (2021)

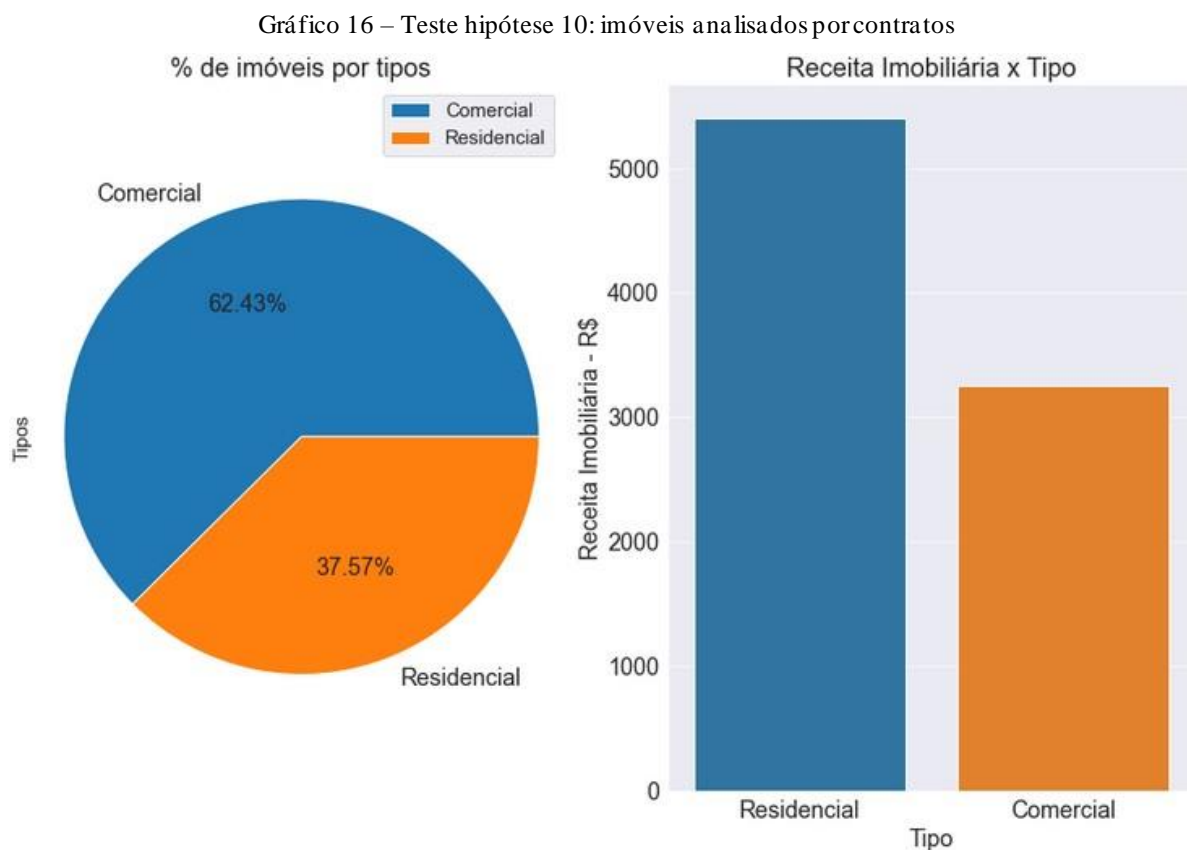
O resultado para este teste mostrou-se afirmativo. Os primeiros contratos quando um imóvel passa a integrar a carteira da imobiliária tendem a ter tempos vagos maiores e, com o passar do tempo, passa a dar lugar ao tempo ocupado. Este comportamento pode-se explicar por dois motivos: o primeiro a existência de imóveis com transferência de carteira que possuem 0% de tempo vago e que podem haver mais de um contrato desta modalidade ao longo do tempo. O segundo motivo pode ser uma possível melhoria nas estruturas dos imóveis nos quais podem representar maior aceitação social por parte dos clientes para com esses ativos, fazendo-se com que o tempo vago diminua ao em função do tempo.

10) Imóveis comerciais oferecem maior receita à imobiliária.

Gráfico 15 – Teste hipótese 10: imóveis analisados por valores únicos



Fonte: Elaborado pelo autor (2021)



Fonte: Elaborado pelo autor (2021)

O resultado para este teste mostrou-se negativo. Apesar da significativa superioridade, tanto em números de imóveis desse tipo quanto em contratos, esta categoria de imóvel não é responsável pela maior parte da receita da imobiliária e sim os imóveis do tipo residenciais.

Os imóveis residenciais apesar de corresponderem a 40% dos imóveis da carteira da corretora e 37,57% de todos os contratos, ativos e inativos, esta categoria representa 50% da receita total da imobiliária. Logo, pode-se afirmar que os imóveis do tipo residenciais oferecem uma maior receita à empresa, apesar de aparentemente a proporção de imóveis dizer o oposto.

De maneira geral, observou-se nos testes de hipóteses através da análise bivariada que não há relação entre causa e efeito entre ambos os fatores, abrindo-se precedentes para testes com outros fatores e possivelmente outros *insights* à partir dos testes realizados nos próximos ciclos do CRISP-DM.

3.4.5 Seleção dos Gráficos

Os testes de hipóteses realizados na AED e o teste de correlação de *Pearson* apoiaram a criação de modelos de visualização de dados validados com uma interface de interação simples, fácil

visibilidade e interpretação de resultados. Conforme citado na etapa anterior e com base nas visualizações criadas na etapa de AED foram determinadas as seguintes visualizações para o dashboard:

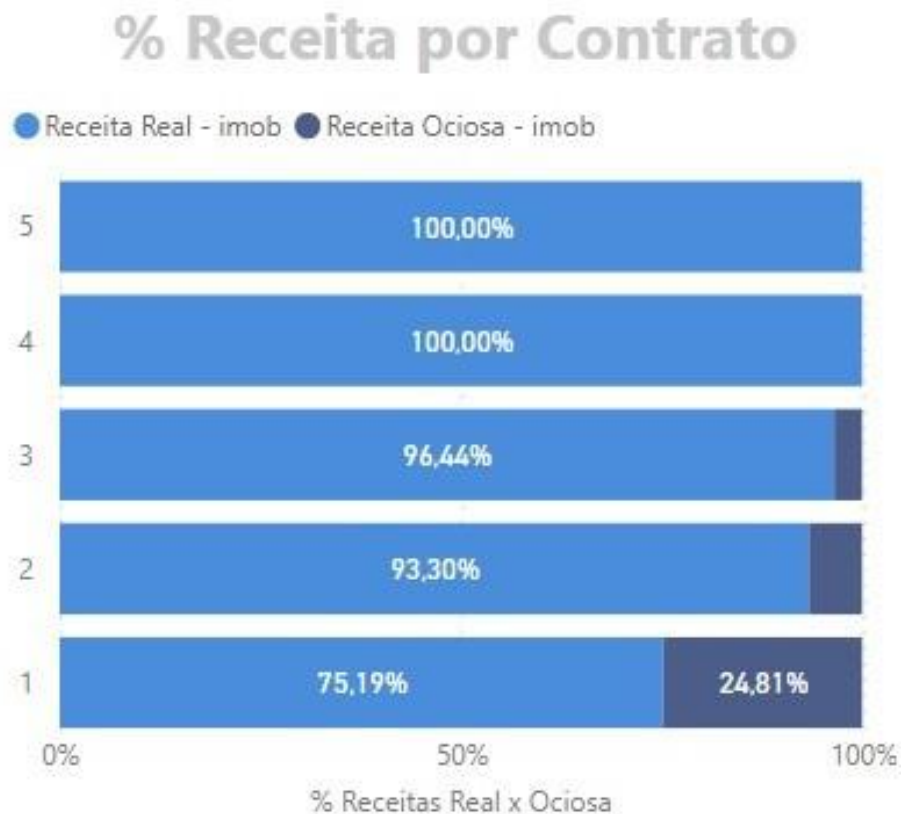
Gráfico 17 – Indicadores: KPI's e cartões



Fonte: Elaborado pelo autor (2021)

Com base no gráfico 3 de correlação de para medidas de desempenho, escolheu-se tais medidas devido ao alto grau de correlação de tais medidas com a variável resposta receita da imobiliária. O desempenho pode ser observado na configuração de inicialização do dashbord no qual exibe o desempenho não de algum imóvel específico mas sim da empresa em relação a todos os imóveis presentes na carteira.

Gráfico 18 – Relação receitas real e ociosa por contratos



Fonte: Elaborado pelo autor (2021)

Com base no teste de hipótese de nº 10, escolheu-se a visualização do teste a integrar o dashboard devido a importância de acompanhar-se o desempenho de imóveis com mais de um contrato e seu comportamento ao longo do tempo.

Gráfico 19 – Gráficos tipo rosca



Fonte: Elaborado pelo autor (2021)

Com base na avaliação das métricas da hipótese de nº 4, constatou-se a importância de avaliar-se a quantidade de imóveis provenientes de transferência de carteira, isto é, cujos proprietários optaram por introsuzi-los na imobiliária em detrimento da anterior pois, oferece uma rentabilidade de 100% a empresa. Constatou-se também que, com os insights extraídos da hipótese de nº 10 observou-se monitorar-se a proporção dos imóveis por tipo, cujo o intuito concentra-se em manter imóveis com maior rentabilidade a carteira.

Gráfico 20 – Relação tipos de receitas



Fonte: Elaborado pelo autor (2021)

De acordo com todas as métricas e insights extraídos na etapa de AED, observou-se a importância do monitoramento das relações entre receitas reais, isto é, arrecadação real da corretora e a receita ociosa cujo o objetivo representa a visualização de uma possível projeção de onde deveria-se estar o ponto ideal de arrecadação, demonstrado pela linha pontilhada em amarelo.

3.5 Avaliação

3.5.1 Atendimento às necessidades

As etapas de levantamento e testes de hipóteses a partir da AED, além da criação de novos insights, mostraram-se fundamentais para a verificação e validação das necessidades da companhia. O conjunto de práticas de tratamento dos dados na etapa de ETL mostrou-se útil na alimentação do modelo com dados compatíveis ao projeto, permitindo-se a execução de análises científicas para a concepção de novas informações para o negócio de modo robusto, isto é, a garantia de não haver problemas com informações resultantes do projeto como por exemplo problemas de bias.

3.5.2 Desempenho

Com a robustez comprovada das informações conforme as etapas anteriores, resta para a avaliação do modelo analisar o desempenho computacional do conjunto de visualizações a serem entregues através do *dashboard*.

No *Power BI Desktop*, permite-se avaliar a performance de cada um dos elementos de relatório, como visuais e fórmulas DAX. Utilizando-se o *Performance Analyzer*, é possível verificar e registrar parâmetros de medida de cada um de seus elementos de relatório quando submetem-se às interações do usuário e quais aspectos de seu desempenho consomem mais ou menos recursos.

O *Performance Analyzer* inspeciona e exibe a duração necessária para atualizar todos os aspectos visuais iniciados pelas interações do usuário e apresenta as informações para que o mesmo possa exibir, fazer *drill down* ou exportar os resultados. O *Performance Analyzer* pode ajudar a identificar aspectos visuais que estão afetando o desempenho dos relatórios e o motivo do impacto.

O recurso mede o tempo de processamento incluindo-se o tempo para criação ou atualização de um visual necessário para atualizar os elementos de relatório iniciados como resultado de qualquer interação do usuário que resulte na execução de uma consulta. Por exemplo, ajustar uma segmentação requer que o visual dela seja modificado e que uma consulta seja enviada para o modelo de dados, além de requerer os visuais afetados que devem ser atualizados como resultado das novas configurações.

Figura 8 – Função *Performance Analyser* do *Power BI*

Performance Analyzer	
▶ Iniciar gravação	↺ Atualizar visuais ⏹ Parar
◊ Limpar 📄 Exportar	
Nome	Duração (ms) ↓
↺ Visual atualizado	-
⊕ Caixa de texto	80
⊕ Imóvel	171
⊕ Relação Receitas (Trimestre): Real x Ociosa	521
⊕ % Receita por Contrato	578
⊕ Cartão de linha múltipla	774
⊕ Receita Locador	370
⊕ Receita Imobiliária	731
⊕ % Tempo Vago	626
⊕ % Receita Ociosa do Locador	330
⊕ % Receita Ociosa Imobiliária	413
⊕ Receita por Tipo de imóvel	660
⊕ Transferência de Carteira	392
⊕ Receita Projetada Locador	691
⊕ Receita Projetada Imobiliária	789

Saiba mais sobre como otimizar o desempenho do relatório em nosso [site de suporte](#). Localize a ajuda para ajustar seu relatório de especialistas parceiros de Power BI em [AppSource](#).

Fonte: Elaborado pelo autor (2021)

O quadro acima refere-se a análise de performance do modelo final pelo *Performance Analyser* do *Power BI*, a coluna da direita exibe a duração em milisegundos (ms) necessária para a exibição de cada elemento presente no painel bem como o tempo de execução de fórmulas DAX para exibir-se algumas visualizações. Ao longo do projeto, pode-se otimizar o desempenho através da preparação de dados ao passo que novos conjuntos de dados integram-se aos dados originais, normalizando-se o banco de dados a fim de evitar-se a redundância dos mesmos, o que pode afetar significativamente o desempenho da exibição dos *dashboards* em *Power BI*.

3.6 Implementação

Apesar de o modelo apresentar-se com desempenho satisfatório este pode ser incrementado nas próximas etapas da metodologia do projeto, reiterando-se o fato de que nas próximas fases do ciclo do CRISP-DM o modelo pode passar por aperfeiçoamentos e incorporação de novas visualizações onde, de acordo com a necessidade da corretora, objetiva alcançar um conjunto de dashboards dedicados a cada área da empresa.

Inicialmente, referente às primeiras informações do negócio disponíveis no primeiro painel do projeto para a gestão de contratos da imobiliária, este primeiro contato com a ferramenta ficar-se-á sob responsabilidade da proprietária da corretora. Uma vez que, para o compartilhamento destas informações faz-se necessário o engajamento de toda a equipe tanto para o manuseio da ferramenta quanto para a interpretação dos resultados que, ao longo dos próximos ciclos do projeto será apresentado à equipe pela proprietária para, por fim, incluir o restante da equipe tanto atual quanto vindouras a acessar e interpretar a ferramenta individualmente.

O grau de compartilhamento do painel com a equipe dependerá do nível de maturidade do projeto nos próximos ciclos e da integração da empresa à nova cultura *Data Driven*. Logo, a disponibilidade do modelo realizar-se-á por meio do próprio *Power BI Desktop* para, posteriormente de acordo com o gradual engajamento da equipe ao longo dos próximos ciclos do projeto, a empresa aderir-se a versão *Power BI pro*, com recursos colaborativos de compartilhamento e versão mobile dos dashboards construídos.

O principal objetivo do modelo é a obtenção, a médio e longo prazo, de imóveis que atendam aos requisitos necessários para a maximização da receita da imobiliária nos quais serão ajustados nos próximos ciclos e cuja a base de dados será retroalimentada com novos dados de imóveis mais fidedignos aos objetivos do modelo, tornando-se mais eficiente. O monitoramento da evolução deste objetivo fará-se através do *dashboard* na configuração de *overview*, isto é, na interface de inicialização do painel na qual, sem a seleção de um imóvel específico, é exibido o comportamento geral do negócio através dos indicadores chave de desempenho, os KPI's, e os cartões com medidas conforme o gráfico 17. Nesta configuração, quanto menor as medidas de desempenho para tempo vago e receita ociosa, significa melhor desempenho para a empresa.

4 Estudo de Caso

4.1 A Empresa

Fundada em 2016 em São Paulo - SP, a corretora de imóveis *New Offices* é constituída apenas pela proprietária mais quatro colaboradores e cerca de 2 outros corretores parceiros, comissionados e não comissionados. A relação do segundo tipo de parceiro com a empresa é de troca de favores onde ambas as partes suprem-se de determinados serviços indisponíveis em seus respectivos portfólios ao passo que surgem-se novas necessidades.

No portfólio de serviços da imobiliária estão presentes, predominantemente, aluguéis de imóveis corporativos, representando 52,46% dos serviços, e residenciais com 47,50%. Apesar de a empresa trabalhar com vendas de imóveis, não consta nos dados do trabalho informações de imóveis desta categoria. O público alvo da empresa, segundo os gestores, se encontra nos grupos cuja renda varia entre R\$10.000,00 e R\$12.000,00 e adultos entre os 30 e 40 anos de idade. Estudantes universitários também se encontram em meio aos clientes da empresa, devido a grande concentração de universidades nas regiões de influência da imobiliária e, apesar de ainda não mensurado devido à deficiência de dados para tal, pode ser notado por parte da gestão uma sazonalidade variando entre contratações e evasões em épocas de início e recessos acadêmicos das universidades.

4.2 Processos de Tomada de Decisão

Os processos essenciais para a empresa, segundo (Gonçalves 2000), estão diretamente associados às regras básicas do negócio. O autor usa como exemplo uma seguradora, prestadora de serviços semelhante à uma imobiliária tradicional cujo ponto principal da organização consiste na relação direta com o cliente, se tratando de um serviço executado mediante à contratação ao invés de um produto manufaturado. Adaptando-se à realidade de uma imobiliária, pode-se citar como processos essenciais os seguintes passos, subsequentemente: captação do cliente, elaboração do contrato de compra e venda, entrega de chaves, manutenção de vínculo contratual com clientes e pós venda ou serviço. Toda a parte estratégica de tomadas de decisões e estabelecimento de processos da empresa precisam ser melhoradas, a empresa dispõe de pessoas porém não dispõe de processos complexos, limitando-se a rotinas para micro tarefas realizadas no âmbito do escritório.

A corretora conta com a plataforma Ingaia Locação para gestão de contratos, onde são cadastrados dados de entrada e saída de clientes, locadores e locatários, assim como dados dos imóveis envolvidos nas transações. Como suporte analítico, a plataforma disponibiliza informações insuficientes em *dashboards*, padrão para todos os clientes os quais não contam com nenhum tipo de customização aplicado às necessidades do negócio, limitando-se exclusivamente a

informações financeiras de entrada e saída de contratos. As decisões são tomadas diariamente com base observações, suposições e experiências acumuladas ao longo da existência da empresa. A plataforma conta um sistema CRM para apoio ao marketing da companhia, auxiliando no monitoramento tanto das preferências dos clientes através do modelo funil de vendas.

Diante da defasagem tecnológica para o apoio a tomada de decisão em um cenário de incertezas, faz-se necessário mesmo tratando-se de um micro empreendimento, a disponibilidade de informações de maior complexidade e customização para todas as áreas do negócio tais como gestão de contratos, financeiro, marketing e recursos humanos. A necessidade da construção de um modelo *Ad Hoc* para o apoio a tomada de decisão da corretora pode concretizar-se por meio da implementação de um projeto de BI.

4.3 Problemas e Deficiências do Negócio

Os principais problemas que afetam a performance nos processos e consequentemente na gestão dos negócios na empresa estão relacionados à falta ou mal aproveitamento de ferramentas tecnológicas, tanto para a agilidade em processos quanto para a tomada de decisões estratégicas. A falta de ferramentas analíticas para a tomada de decisões não impede que os gestores conheçam as principais métricas da empresa, bem como a tomada de decisões operacionais de curto prazo, mesmo decisões baseadas em suposições. No entanto, a médio e longo prazo é um fator impeditivo para o vislumbre de novas oportunidades de negócio e implementação de planos de ação e campanhas de valor estratégico importantes para a alavancagem do negócio.

Além das deficiências cujas soluções são propostas no presente trabalho, problemas como entrada e saída de contratos tanto de locadores quanto de inquilinos, demora em processos de análises de contratos e análises cadastrais, são problemas que podem impactar negativamente na experiência do cliente que poderiam ser resolvidos com o uso da tecnologia, com destaque para o uso de técnicas de automação de processos e inteligência artificial.

Concorrência dentro do mesmo setor cujo modelo de negócio se baseia na tecnologia de alto nível, está proporcionando experiências satisfatórias aos clientes, principalmente com relação à redução da morosidade de processos decorrentes do excesso de burocracia envolvidas nos serviços deste segmento.

Os insights apresentados no trabalho têm como objetivo proporcionar aos gestores do empreendimento informações de negócio, outrora desconhecidas ou no plano das suposições, para permitir aos mesmos um passo adiante na consolidação de ações inéditas e representativas para o crescimento da empresa.

5 Resultados e Discussão

5.1 *Insights e Informações de Negócio em Power BI*

De acordo com a metodologia CRISP-DM, pôde-se observar que a construção de um *dashboard* para o apoio a tomada de decisões estratégicas por parte de da alta gestão de uma companhia, não limita-se apenas a utilização da ferramenta arbitrariamente e simplificada. Esse tipo de prática limita o autor à construção de painéis somente por meio do *Self Service BI*, operação na qual seleciona-se variáveis desejadas no visual escolhido. Esta prática, além de fornecer aos tomadores de decisões análises de impactos pouco significativos para o negócio, permite aos mesmos apenas o monitoramento de métricas conhecidas a priori. Permite que métricas como por exemplo lucros, prejuízos e spreads tenham seus comportamentos monitorados sem, no entanto, oferecer uma explicação científica dos fenômenos.

A geração de insights, conforme elucidado na etapa da metodologia do trabalho permite ao tomador de decisões, além das definições das melhores visualizações para o dashboard, permite-o conhecer o comportamento dos dados e um melhor direcionamento do negócio por meio de informações implícitas descobertas apenas em um processo de análise exploratória dos dados. Ao longo da aplicação da metodologia de gestão de projetos para mineração de dados observou-se que gerações de insights são essenciais para o conhecimento do negócio e auxiliam na elaboração de planos de ação assertivamente mais significativos e de maior impacto ao projeto.

Conforme a terceira etapa denominada Compreensão dos Dados da metodologia CRISP-DM, o levantamento de hipóteses de negócio por parte da equipe de colaboradores da imobiliária mostrou-se essencial, tanto para a geração de insights que trouxeram à luz conhecimentos acerca do negócio quanto indícios de importantes métricas de negócio cujos desempenhos devem ser constantemente monitoradas ao passo que novos contratos são integrados a carteira do empreendimento, bem como medidas de desempenho com eficiência diretamente relacionada a variável de interesse para os gestores, através da análise de correlações de *Pearson*: a receita da imobiliária.

De acordo com o trabalho desenvolvido, constatou-se que, apesar da praticidade da ferramenta *Power BI*, uma abordagem mais profunda em relação aos problemas de negócio e alguns padrões estratégicos não podem ser identificados apenas com o uso desta como uma única ferramenta. No presente trabalho mostrou-se o potencial que a ferramenta pode atingir em conjunto com outras aplicações como por exemplo linguagens de programação com bibliotecas direcionadas em análise de dados.

Utilizou-se a linguagem *Python* conjuntamente com bibliotecas direcionadas para análise de dados *Pandas*, *Numpy*, *Matplotlib*, *Seaborn* e *Statistics*. Estas ferramentas permitem uma abor-

dagem muito mais analítica ao estudo e, principalmente, um conhecimento significativo do comportamento dos agentes envolvidos no negócio através da base de dados disponível. Uma segunda ferramenta que pode-se empregar neste tipo de estudo é a linguagem de programação *R*, uma linguagem direcionada à estatística computacional e gráficos complexos.

Tendo-se em vista que o maior problema enfrentado pela proprietária e demais colaboradores é a falta de conhecimento do próprio negócio, do mercado cujo a imobiliária se encontra inserida e a falta de ferramentas para tecnologia da informação na resolução de problemas complexos, o conjunto de ambas as ferramentas mostrou-se essencial para tal finalidade.

5.2 Dashboard

Os estudos realizados na terceira etapa da metodologia CRISP-DM direcionada ao conhecimento do negócio, com destaque para o processo de análise exploratória dos dados, estabeleceu-se, com base nos insights extraídos, importantes métricas e medidas de desempenho para o monitoramento e tomadas de decisões estratégicas em relação à manutenção dos imóveis da carteira da empresa.

O dashboard construído mostrou-se fundamental para a constante avaliação do portfólio da companhia principalmente por oferecer ao gestor apoio a tomada de decisão em tempo real, através da análise de informações por meio de gráficos e KPI's instantaneamente para cada produto da carteira de forma individualizada. Através deste painel, o gestor tem recursos disponíveis suficientes para decidir sobre a permanência de determinado imóvel na imobiliária quanto à sua receita oferecida a empresa em função de seu respectivo tempo de permanência ocupados, isto é, com a permanência de inquilinos mediante o pagamento de aluguéis e dividendos percentuais do aluguel, pago ao proprietário, denominado pela empresa como a taxa de administração.

A retirada do imóvel da carteira poderá ser analisada através de três indicadores de desempenho, os KPI's, quanto ao percentual do tempo vago no imóvel, o percentual da receita ociosa para o locador e proporcionalmente ao percentual da receita ociosa para a imobiliária. Como referido anteriormente no trabalho, imóveis com altas taxas de vacância, apresentam-se como passivos para a imobiliária cuja a principal causa, segundo o gestor, é a comunicação com o proprietário no qual superestima o valor de seu ativo desconsiderando o passivo que o mesmo pode representar tanto para a empresa quanto para si. Diante deste problema, um dos objetivos do dashboard é auxiliar a argumentação da corretora perante a comunicação com o proprietário. Juntamente com indicadores de desempenho cartões, dispostos hierarquicamente, desempenham a função de exibir informações de receita de cada imóvel a fim de facilitar o poder de argumentação do profissional responsável no objetivo de persuadir o proprietário a aceitar ofertas de menores valores de aluguel pelo imóvel e, em última instância, a supressão do mesmo do portfólio da empresa.

O dashboard no ato de sua inicialização, conforme os gráficos 17, 18, 19 e 20 da etapa de modelagem, apresenta ao usuário como uma primeira interface uma visualização geral dos dados de todos os imóveis da base de todos os contratos ativos e encerrados. Demonstrando-se o poder do painel desenvolvido, abaixo têm-se exemplos do funcionamento das visualizações para imóveis com bom desempenho, isto é, candidatos a manter-se na carteira:

Figura 9 – Indicadores de desempenho satisfatórios



Fonte: Elaborado pelo autor (2021)

Considerando-se, ainda com base em observações da equipe da corretora, considera-se os indicadores abaixo de 20% aceitáveis a permanência do imóvel na carteira.

Figura 10 – Relação receitas real e ociosa: desempenho satisfatório

Fonte: Elaborado pelo autor (2021)

Observa-se na relação entre receitas real e ociosa uma melhoria nas taxas de ocupação deste imóvel ao longo do tempo cuja receita ociosa se mantém restrita apenas ao primeiro contrato.

Figura 11 – Gráficos tipo rosca: desempenho satisfatório

Fonte: Elaborado pelo autor (2021)

Figura 12 – Relação tipos de receitas: desempenho satisfatório

Fonte: Elaborado pelo autor (2021)

Observa-se nesta última visualização do dashboard a evolução no desempenho do imóvel cujos contratos firmaram-se nos trimestres seguintes, onde pode-se constatar a ausência da receita ociosa e o alcance da receita projetada.

Como exemplo, por outro lado, têm-se visualizações referentes ao imóvel de desempenho insatisfatório, considerando-se ainda com base nas observações da equipe de negócio, um índice de desempenho acima de 20% insatisfatório, fazendo-se do imóvel candidato a retirada da carteira.

Gráfico 21 – Indicadores de desempenho insatisfatórios



Fonte: Elaborado pelo autor (2021)

Com base nas correlações levantadas na etapa de AED, onde tempos de ocupação e vacância estão diretamente relacionados a ambas as receitas reais e ociosas, taxas significativamente acima de 20% indicam um possível transtorno à imobiliária e candidato à retirada do mesmo da carteira.

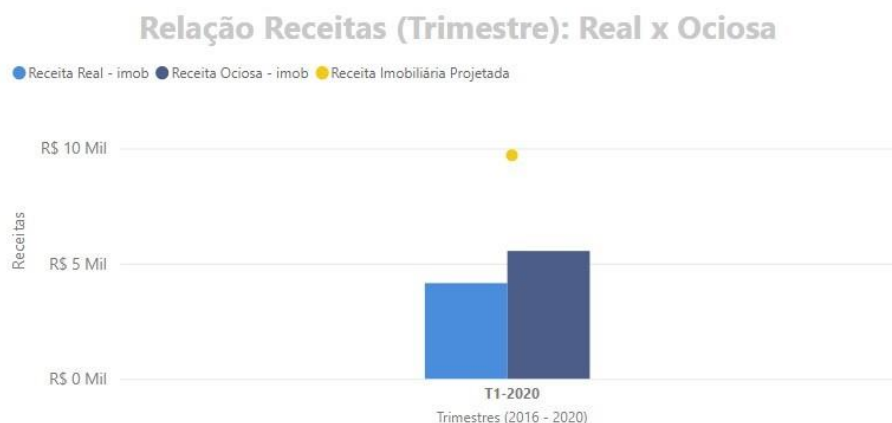
Figura 13 – Relação receitas real e ociosa: desempenho insatisfatório

Fonte: Elaborado pelo autor (2021)

Com apenas o primeiro contrato, observa-se uma receita ociosa de 57,20% sem apresentação de nenhuma evolução dentro de 2.063 dias na carteira da imobiliária, conforme o cartão indicador referente ao total de dias na imobiliária.

Figura 14 – Gráficos tipo rosca: desempenho insatisfatório

Fonte: Elaborado pelo autor (2021)

Figura 15 – Relação tipos de receitas: desempenho insatisfatório

Fonte: Elaborado pelo autor (2021)

O único trimestre cujo o contrato faz parte do último ano do conjunto de observações encontra-se com receita real inferior a receita ociosa.

Pode-se observar uma visão ampla das medidas de desempenho da empresa por meio de todos os contratos nos quais tange interesses de ambas as partes: a corretora e o locador.

5.3 Plano de Ação

O projeto desenvolvido e sustentado por todos os passos da metodologia escolhida trouxe uma nova perspectiva ao negócio no que se refere ao comportamento das métricas e ferramentas até o momento utilizadas. Padrões latentes em meio a poucos dados informacionais disponíveis nas bases da companhia, apesar da baixa capacidade de extração e armazenamento dos mesmos seja pela falta de recursos ou práticas irregulares descritas no presente trabalho, tornaram-se parte das deliberações rotineiras e das diretrizes da organização devido aos novos conhecimentos de negócio expostos por análise de insights. Desta forma prevê-se uma possível mudança no comportamento empreendedor da companhia no que se refere a orientação a dados. Segundo a nova perspectiva da corretora proporcionada através projeto, pode-se adotar para a segunda etapa do CRISP-DM em diante os seguintes planos de ação listados:

- Mudança nos fatores de apoio a decisão antes presumidos e refutados por novas descobertas em testes de hipóteses. Acreditava-se que imóveis de preços inferiores eram mais benéficos à empresa por oferecerem maior receita devido ao maior tempo de ocupação, ambas suposições foram refutadas pelos testes apontando-se preferência por imóveis de custos mais elevados. Outra suposição a priori submetida a ponderações está relacionada à presença de garagens. A ideia anterior que era priorização na carteira de imóveis com

uma ou mais vagas em garagens tornam-se obsoletas, sendo este fator irrelevante na procura deste tipo de propriedade por parte de potenciais inquilinos e impactante na receita da corretora;

- Coleta e consolidação de um banco de dados de todos os setores da empresa. Apesar da escassez de recursos humanos devido propriamente ao seu tamanho cujos colaboradores envolvidos caracterizam-se pela versatilidade de funções e setores, é possível a segregação e organização dos dados em diferentes entidades para cada setor. Tal ação envolver-se-á na criação de um DW consolidado com dados de diferentes setores de variáveis específicas segregados em entidades;
- Associada a ação anterior, com um conjunto de bases agregados em um DW, dentre as próximas ações encontra-se o emprego de novos dashboards concomitantes ao do presente trabalho. Esta ação ocorrer-se-á com base no levantamento e teste de novas hipóteses e a geração de novos insights a fim de expandir-se o conhecimento a respeito das outras áreas do negócio estudado e buscar o exercício de monitoramento de novas métricas que auxiliem na tomada de decisão e maximização das receitas do empreendimento;
- Adoção de novas práticas para utilização total de ferramentas disponíveis. Muito do potencial operacional do negócio é comprometido devido a subutilização de ferramentas devido a falta de conhecimento ou tempo por parte dos colaboradores. Conforme citado anteriormente no trabalho a empresa conta com um sistema de CRM subutilizado, não considerando-se que se trata de um sistema pago que pode representar a entrada de muitos dados com informações estratégicas para a imobiliária;
- Elaboração de campanhas e promoções para estímulo a transferência de carteira de imóveis de outras imobiliárias para a empresa em questão. Observou-se por meio aos estudos exploratórios que ativos de transferência de carteira apresentam-se com tempo de ocupação em 100%, em outras palavras, esta categoria oferece a totalidade da receita real para a corretora desde o início da parceria desta com o proprietário.

6 Conclusão e Considerações Finais

O principal objetivo do trabalho é a implementação do Business Intelligence em uma microempresa através do CRISP-DM, obedecendo-se a rigor, todos os respectivos passos da metodologia. Destacou-se como ferramenta principal utilizada o *Power BI* da desenvolvedora Microsoft, devido a sua facilidade em utilização cuja a interface conta com auxílio de algoritmos de inteligência artificial para construção de visualizações rápidas, além da disponibilidade de suporte e de documentação ambos cedidos pela desenvolvedora.

Com a metodologia escolhida CRISP-DM, pôde-se constatar no trabalho que a mesma proporciona um excelente direcionamento para o desenvolvimento e implementação do projeto do início ao fim, bem como seu devido acompanhamento e melhoria contínua.

Deve-se considerar que, para a adoção de uma cultura Data Driven por parte de uma empresa, dantes não orientada a dados, necessita-se não apenas a implementação de tecnologias, mas também a construção de todo um processo de inclusão digital e construção de *mindset* para o engajamento de todos os colaboradores envolvidos. Não apenas devido a presença de uma nova tecnologia mas, não menos importante, devido ao fato de o projeto ser devidamente conduzido por recursos humanos.

A metodologia permite a gradual adaptação da equipe a orientação a dados em razão de sua ciclicidade no seu desenvolvimento e implementação, provando-se robusta desde o levantamento de falhas e necessidades das áreas de TI da companhia até o engajamento ideal da equipe, mesmo com deficiências na base de dados inicial.

Reiterando-se a seção de Resultados e Discussão do trabalho, pode-se considerar que o projeto trouxe para empresa, de imediato, consideráveis vantagens. Em curto prazo, os *insights* construídos na etapa de análise exploratória, além de direcionar a construção do *dashboard*, trouxe à gestão a importância dos dados para a tomada de decisão em detrimento da pura e simples observação e intuição.

Novas diretrizes que objetivam o crescimento da receita e da empresa como a captação de imóveis com maiores valores de aluguel em oposição a imóveis com menor valor foram apontados em descobertas na AED, bem como a indicação de imóveis provenientes de transferência de carteira. Ambas diretrizes, orientadas a dados, cujas campanhas e planos de ação podem-se apresentar níveis de assertividade outrora não alcançados.

Por fim, de forma análoga porém no longo prazo, conforme mencionado nos parágrafos anteriores desta seção, prevê-se significativa importância do projeto na construção de um DW estruturado

e consolidado e recursos humanos capacitados cujo, ambos integrados concomitantemente ao desenvolvimento e mudança da nova cultura da companhia.

Como continuidade para melhoria do presente projeto, propõe-se o levantamento de novas hipóteses acerca dos outros agentes envolvidos no projeto e citados no trabalho. Como um segundo ciclo do CRISP-DM, far-se-á a realização de novos testes e novos *insights* inerentes a outros agentes e seus respectivos atributos que, possivelmente, impactem na receita da corretora.

7 Referências Bibliográficas

ABEL, Marcos Chedid. **Oinsight na psicanálise**. Psicologia: ciência e profissão, v. 23, p. 22-31, 2003

ADORNO, Luciano; TEIXEIRA, Júlio. **Mind Map 3D Interativo: recomendações para o desenvolvimento de uma ferramenta dinâmica para visualização sistêmica de dados**. 9ª CIDI Conferência Internacional de Design e Informação, 2019.

ALLEN, R. E. **The Concise Oxford English Dictionary**. 1990.

ARAÚJO, Assilio Luiz Zanella de; FERRARI FILHO, Fernando; BUENO, Eduardo. **Existe uma bolha imobiliária no Brasil?: uma análise teórica e empírica**. **Análise Econômica**. Porto Alegre. Vol. 34, n. 66 (set. 2016), p. 149-172, 2016.

BALL, Michael. **London and property markets: a long-term view**. Urban Studies, v. 33, n. 6, p. 859-876, 1996.

BAILEY, Trevor C. et al. **Interactive spatial data analysis**. Essex: Longman Scientific & Technical, 1995.

CRISP-DMHELP OVERVIEW. IBM, 2021. Disponível em <<https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=dm-crisp-help-overview>>. Acessado em 16 de dez. de 2021.

DONG, Guozhu; LIU, Huan (Ed.). **Feature engineering for machine learning and data analytics**. CRC Press, 2018.

FERREIRA, Antonio Augusto Pinto et al. **Business intelligence**. SITEFA-Simpósio de Tecnologia da Fatec Sertãozinho, v. 3, n. 1, p. 268-274, 2020.

FOX, P.; HENDLER, J. **Changing the Equation on Scientific Data Visualization**. **Science** 331, 705 (2011). Disponível em: <http://data2discovery.org/dev/wp-content/uploads/2013/05/Fox-and-Hendler_Visualization_Science-2011-Fox-705-8.pdf>. Acesso em 05 de dez. de 2021.

GONÇALVES, José Ernesto Lima. **As empresas são grandes coleções de processos**. Revista de administração de empresas, v. 40, n. 1, p. 6-9, 2000.

MERCADO IMOBILIÁRIO. **Investe-SP**, 2022. Disponível em <<https://www.investe.sp.gov.br/setores-de-negocios/mercado-imobiliario/>>. Acessado em 13 de mar. de 2022.

MYTELKA, Lynn et al. **Local clusters, innovation systems and sustained competitiveness**. UNU/INTECH Discussion Paper, n. 2005, 2000.

OLIVEIRA, T.; BORSHIVER, S. **Interação De Instituições Científicas e Tecnológicas Com Em-presas: Um Estudo do INT e Das Empresas Do Setor Químico**. GEINTEC Revista de Gestão, Inovação e Tecnologia, Aracaju, v. 3, n. 3, 2013.

PANORAMA POPULACIONAL CIDADE DE SÃO PAULO-SP. **IBGE**, 2021. Disponível em <<https://cidades.ibge.gov.br/brasil/sp/sao-paulo/panorama>>. Acesso em: 15 de nov. de 2021.

PESQUISA DE MERCADO - CAPITAL. **CRECISP**, 2021. Disponível em <<https://www.crecisp.gov.br/comunicacao/pesquisasmercado/capital>>. Acesso em: 16 de nov. de 2021.

POWERBIDOCUMENTATION. Documentação do Power BI, 2022. Disponível em <<https://docs.microsoft.com/pt-br/power-bi/>>. Acesso em: 15 de mar. de 2022.

R DOCUMENTATION. R-Project, 2022. Disponível em <<https://www.r-project.org/>>. Acessado em: 21 de jan. de 2022.

SHARDA, R.; DELEN, D.; TURBAN, E. **Business Intelligence e Análise de Dados para Gestão do Negócio**. 4ª Ed. Porto Alegre: Bookman, 2019.

SHEARER, C. **The CRISP-DM model: the new blueprint for data mining**. Journal of data warehousing, v. 5, n. 4, p. 13-22, 2000.

Anexos

Anexo 1 – Algoritmo em Python para cálculo e criação de colunas

```

1 import pandas as pd
2 import numpy as np
3
4 dataset['Entrada_imobiliaria'] = pd.to_datetime( dataset['Entrada_
    imobiliaria'], dayfirst=True)
5 dataset['Início'] = pd.to_datetime( dataset['Início'], dayfirst=
    True)
6 dataset['Término'] = pd.to_datetime( dataset['Término'], dayfirst=
    True)
7
8 for i in range(0, len( dataset) -1):
9     if (dataset['Codigo_Imovel'][i] == dataset['Codigo_Imovel'][i
        +1]) & ((dataset['Início'][i+1].year < dataset['Término'][
            i].year)):
10         dataset['Término'][i] = dataset['Início'][i+1]
11     elif ((dataset['Codigo_Imovel'][i] == dataset['Codigo_Imovel'
        ][i+1]) & ((dataset['Início'][i+1].year == dataset['Té
            rmino'][i].year) & (dataset['Início'][i+1].month < dataset
                ['Término'][i].month))):
12         dataset['Término'][i] = dataset['Início'][i+1]
13     else:
14         pass
15
16     #SUBSTITUINDODATASDEENTRADADOSIMOVEISNAIMOBILIARIANOS
        CONTRATOSSUBSEQUENTES
17     #DATADEENTRADAPRECEDEAPENASADATADEINICIODOPRIMEIRO
        CONTATO
18     if (dataset['Codigo_Imovel'][i] == dataset['Codigo_Imovel'][i
        +1]) & (dataset['Numero_contrato'][i+1] > 1):
19         dataset['Entrada_imobiliaria'][i+1] = dataset['Término'][
            i]
20
21 #CRIANDOCOLUNAS
22 dataset['Tempo_ocupado'] = dataset['Término'] - dataset['Início']
23 dataset['Tempo_vago'] = dataset['Início'] - dataset['Entrada_
    imobiliaria']
24 dataset['Tempo_ocupado'] = dataset['Tempo_ocupado'].dt.days.
    astype('int64')
25 dataset['Tempo_vago'] = dataset['Tempo_vago'].dt.days.astype('
    int64')
26
27 #DEFININDOCOLUNAPORTABILIDADE
28 dataset['Portabilidade'] = np.NaN
29
30 for k in range(0, len( dataset) -1):
31     if dataset['Tempo_vago'][k] < 0:
32         dataset['Portabilidade'][k] = 'sim'
33     else:
34         dataset['Portabilidade'][k] = 'nao'
35
36 #RESOLVENDOTEMPOOCUPADOPARATODASASLINHASSETEMPO
    DESOCUPADOPARAIMOVEISCOMAPENASUMOUPRIMEIROCONTRATO

```

```

37 for c in range(0, len( dataset) - 1):
38     if dataset['Tempo_vago'][ c] < 0:
39         dataset['Entrada_imobiliaria'][ c] = dataset['Início'][ c]
40
41 #RECALCULANDOCOLUNASTEMPOVAGO
42 dataset['Tempo_vago'] = dataset['Início'] - dataset['Entrada_
    imobiliaria']
43 dataset['Tempo_vago'] = dataset['Tempo_vago'].dt.days.astype('
    int64')

```

Figura 16 – Anexo 2: Dashboard final do projeto modo inicialização



Fonte: Elaborado pelo autor (2021)

Figura 17 – Anexo 3: Dashboard imóvel com desempenho satisfatório



Fonte: Elaborado pelo autor (2021)

Figura 18 – Anexo 4: Dashboard imóvel com desempenho insatisfatório



Fonte: Elaborado pelo autor (2021)