

Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação

## **Otávio de Meira Lima**

MONOGRAFIA EM SISTEMAS DE INFORMAÇÃO I - PESQUISA TECNOLÓGICA  
**Emprego de Ciência de Dados na Redução do Déficit  
Habitacional Brasileiro**

Professor Adriano Alonso Veloso  
Orientador

**Resumo:** Este documento se refere ao relatório final da disciplina Monografia em Sistemas de Informação I - Emprego de Ciência dos Dados na Redução do Déficit Habitacional Brasileiro. Contém análises exploratórias ao redor dos dados disponíveis pelo projeto federal Casa Verde Amarela, analisando atributos de imóveis e de seus compradores e referenciais teóricos já existentes no mercado.

**Abstract:** This document refers to the final report of the subject Monograph in Information Systems I - Employment of Data Science in Reducing the Brazilian Housing Deficit. Contains exploratory analyzes around the data available by the Casa Verde Amarela federal project, analyzing attributes of properties and their buyers and theoretical references already existing in the market.

## 1. Introdução

De acordo com um levantamento da EmCasa[1], empresa de intermediação imobiliária, quatro em cada dez brasileiros pretendem comprar imóveis até o ano de 2023. Os aspectos mais importantes considerados na hora de escolher um imóvel são a localização, região e valor. Porém, grande parte desse grupo que têm interesse em comprar uma casa ou imóvel não conseguem alcançar seu objetivo, principalmente devido à questões financeiras.

Pensando nisso, o governo brasileiro criou, em 2009, um programa federal chamado Minha Casa Minha Vida[2] com o objetivo de combater o déficit habitacional do país e proporcionar a realização da compra da casa própria para famílias de baixa renda. Apresentado com orçamento do Ministério das Cidades e gestão da Caixa Econômica Federal, o projeto social oferece subsídio de até R\$47,5 mil e taxa de juros abaixo do mercado para facilitar a aquisição de imóveis residenciais na cidade ou no campo até um determinado valor.

Apesar dos avanços significativos nos últimos anos, a questão da moradia ainda é um grande problema social em todo o país. Uma pesquisa feita pela Associação Brasileira de Incorporadoras Imobiliárias (Abrainc) e da Fundação Getúlio Vargas (FGV), com dados de 2017, o Brasil tem um déficit habitacional de pelo menos 7,8 milhões de domicílios. Pensando nos novos rumos que o programa precisa tomar, o atual governo de 2019-2022 substituiu o programa Minha Casa Minha Vida pelo programa Casa Verde Amarela, lançado em 2020 com o mesmo apelo social. A expectativa atual é de auxiliar pelo menos 1,6 milhões de famílias de baixa renda até o final de 2024.

O programa atualmente auxilia a população em 6 grupos diferentes, sendo 3 urbanos e 3 rurais. São eles:

- a. Urbano 1 – renda bruta familiar mensal até R\$ 2.400 mil
- b. Urbano 2 – renda entre R\$ 2.400,01 e R\$ 4.400 mil
- c. Urbano 3 – renda entre R\$ 4.400,01 e R\$ 8 mil
- d. Rural 1 – renda bruta familiar anual até R\$ 29 mil
- e. Rural 2 – renda anual entre R\$ 29.000,01 e R\$ 52.800,00 mil
- f. Rural 3 – renda bruta familiar anual entre R\$ 52.800,01 até R\$ 96 mil

As famílias que vivem no ambiente rural, em geral, não possuem uma taxa de entrada de renda fixa mensal. A maioria de sua renda é proveniente da venda de produtos agrícolas, que possuem suas vendas sazonais e podem afetar as famílias caso elas fossem classificadas por renda mensal.

O projeto, em parceria com a UFMG, disponibilizou os dados para a realização de pesquisas que podem auxiliar na redução do déficit habitacional brasileiro. Esses dados serão utilizados para analisar a atual situação da população e, futuramente, será possível aplicar um modelo para auxiliar um comprador antes de fechar uma compra definitiva de um imóvel.

## 2. Referencial

O mercado imobiliário está bem aquecido e o uso de tecnologias de aprendizado de máquina é inevitável. No mercado norte-americano, a maioria dos dados de imóveis e terras estão disponíveis ao público. Dentre os dados, estão histórico do lote, documentações passadas, preços de compra anteriores e até valores de hipoteca. Atualmente, os algoritmos de aprendizado de máquina são utilizados para percorrer milhões de documentos em segundos, armazenando valores do imóvel e até mesmo informações pessoais dos atuais donos.

A LoanSnap utiliza algoritmos em vários passos do processo de hipoteca com foco em auxiliar o futuro dono do imóvel a calcular seus gastos e valores de finanças a serem calculadas.

A Compass possui uma AI capaz de realizar triangulações com todas as informações disponíveis ao público e predizer quais imóveis estarão futuramente à venda. Dessa forma, um agente pode abordar o dono e oferecer seus serviços de corretagem com maior probabilidade de conversão.

Já o algoritmo Zestimate[3] é capaz de avaliar o preço de um imóvel com um modelo de rede neural que correlaciona localização, aspectos de moradia, preço de mercado e tendências de valor. Quanto mais informações sobre o imóvel, mais

preciso é o resultado.

As empresas Loft e QuintoAndar são as mais relevantes quando se fala de mercado imobiliário no Brasil. Elas intermediam e facilitam a compra de imóveis, disponibilizando dados relevantes e ajudando seus clientes na escolha do imóvel ideal. QuintoAndar ainda possui uma plataforma que auxilia o dono do lar a calcular o valor ideal para sua casa ou apartamento.

Foram citados diversos exemplos de emprego da tecnologia no mercado imobiliário, tanto americano quanto brasileiro. Porém, todas as soluções possuem foco na venda do imóvel, expansão de negócios e intermediação de compras e aluguéis. O problema a ser resolvido, proposto por esse trabalho, é encontrar uma maneira de auxiliar o consumidor a encontrar o imóvel ideal baseado nos seus dados pessoais, disponibilizados pelo projeto Casa Verde Amarela. O foco, então, é na análise dos dados pessoais da população em busca de um imóvel, ao contrário do referencial citado que foca na análise dos imóveis em si.

### 3. Contribuição

Os dados disponíveis pelo projeto Casa Verde Amarela contém diversos atributos, tanto do imóvel quanto do seu comprador. Dentre eles, estão:

- Avaliação do imóvel
- Renda Total Apurada
- Valor do Financiamento
- Valor Recursos Próprios
- Valor Reserva
- Valor Recursos Próprios Aportados
- Renda Comprovada
- Renda não Comprovada
- Idade
- Estado Civil
- Sexo
- Rating
- ...

No total, foram coletados 2914 dados sobre compradores e seus imóveis.

A idade média da população atendida no programa é de 36 anos, sendo que a grande maioria se encontra no estado civil Solteiro. O programa também recebe população que se encaixa em todos os estados civis, mas solteiros estão em maior quantidade, seguidos por casados e divorciados. Esses dados estão de acordo com

o esperado da população, já que, de acordo com um estudo feito pelo IBGE, a idade média dos brasileiros casados está cada vez crescendo, atingindo atualmente a marca de 30,8 anos para os homens e 28,2 anos para as mulheres.



Imagen 1: gráfico de barras para visualização da população por estado civil

Foi realizada uma análise mais profunda em relação à renda da população que adquiriu o imóvel. Como o projeto tem como público alvo uma população mais carente, é possível perceber que a Renda Comprovada média de R\$2627 reais não é o suficiente para adquirir uma casa/apartamento de alto padrão. O valor médio dos imóveis fica em R\$153.559,00. Segundo dados de 2018 divulgados pela FipeZap[4], é preciso desembolsar, em média, R\$ 458 mil para comprar a casa própria com um ou dois dormitórios em 60m<sup>2</sup>.

Os valores que o comprador possui em reserva, dados pelos atributos “Valor Reserva” e “Valor Recursos Próprios Aportados” possuem um peso significativo ao comprar um imóvel. Isso é explicado pelo fato de que o valor do apartamento vai além de apenas o seu valor, mas também inclui mudanças, decoração, reformas, mobílias, gastos com documentação, dentre outras.

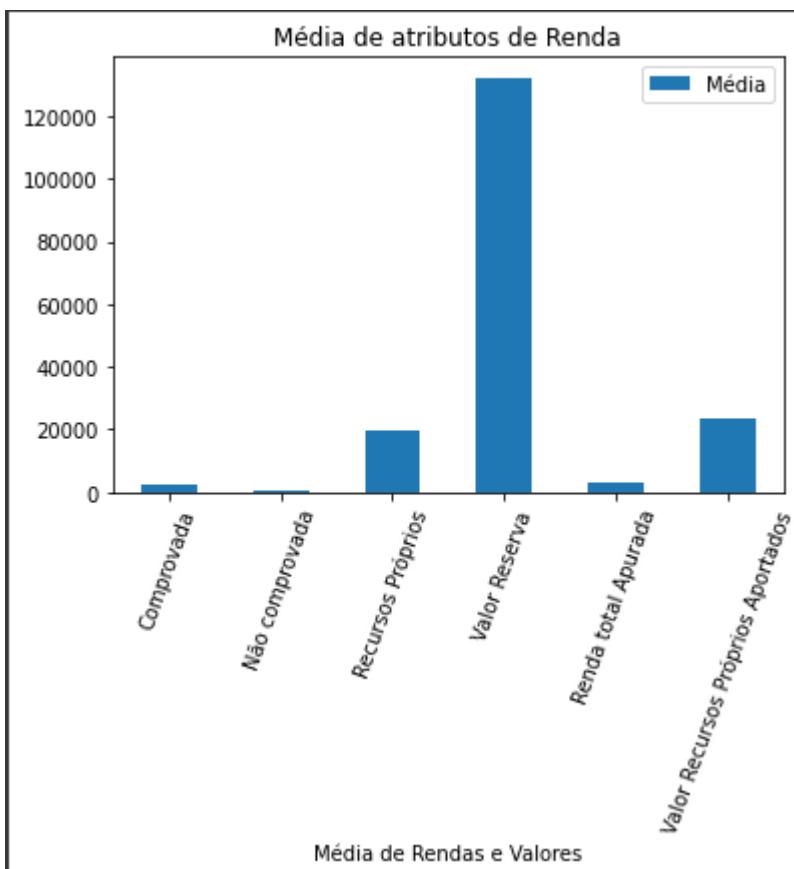


Imagen 2: gráfico de barras para visualização de atributos de renda média

As variáveis mais relevantes para essa análise circulam ao redor do valor do imóvel e da renda de seu comprador. É possível perceber uma forte correlação com a Renda Comprovada do cidadão com a Avaliação do Imóvel. É de se esperar que quanto maior a renda do comprador, mais caro o imóvel que ele pode comprar. Os valores em reserva do comprador são aqueles que mais se correlacionam proporcionalmente com o valor do imóvel. A variável “Rating” significa uma avaliação do comprador, que leva em consideração dívidas e compromissos cumpridos que possui junto com o Governo.

	Avaliação do Imóvel	Valor Recursos Próprios	Renda Comprovada	Renda não Comprovada	Valor Reserva	Valor Recursos Próprios Aportados	Rating
Avaliação do Imóvel	1.00000	0.132624	0.534850	-0.175708	0.836414	0.704879	0.250871
Valor Recursos Próprios	0.132624	1.00000	-0.105954	-0.042562	0.022419	0.211293	-0.046268
Renda Comprovada	0.534850	-0.105954	1.00000	-0.284365	0.440057	0.347801	0.173366
Renda não Comprovada	-0.175708	-0.042562	-0.284365	1.00000	-0.125076	-0.070123	-0.078889
Valor Reserva	0.836414	0.022419	0.440057	-0.125076	1.00000	0.384068	0.186143
Valor Recursos Próprios Aportados	0.704879	0.211293	0.347801	-0.070123	0.384068	1.00000	0.188746
Rating	0.250871	-0.046268	0.173366	-0.078889	0.186143	0.188746	1.00000

Imagen 3: correlação de Pearson entre as variáveis mais relevantes do imóvel e do comprador.

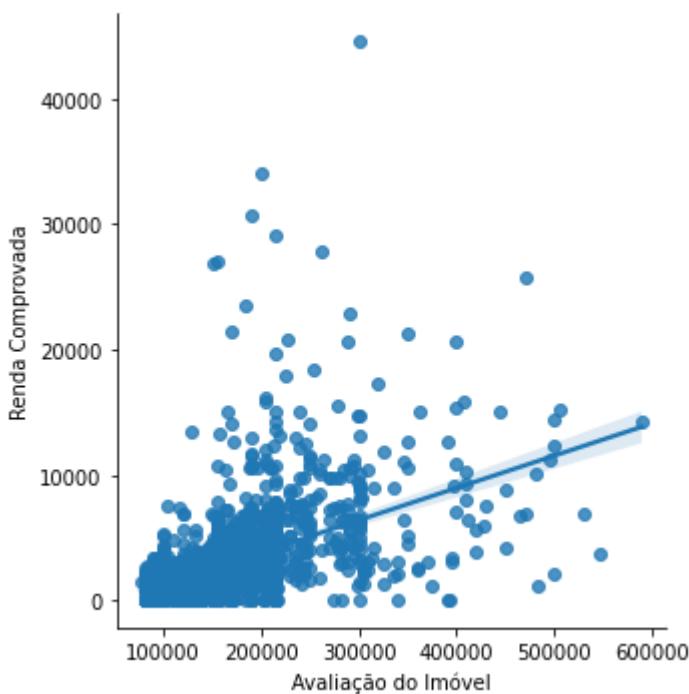


Imagen 4: correlação de Pearson entre “Renda Comprovada” e “Avaliação do Imóvel”.

O gráfico a seguir mostra a relevância do valor que o comprador possui em reserva para a compra de um imóvel pelo programa Casa Verde Amarela. A correlação de Pearson (0.836) entre esses dois atributos é a maior que se pode encontrar nessa base de dados. A visualização desse valor é dada pela inclinação acentuada da reta chamada “Best Fit”.

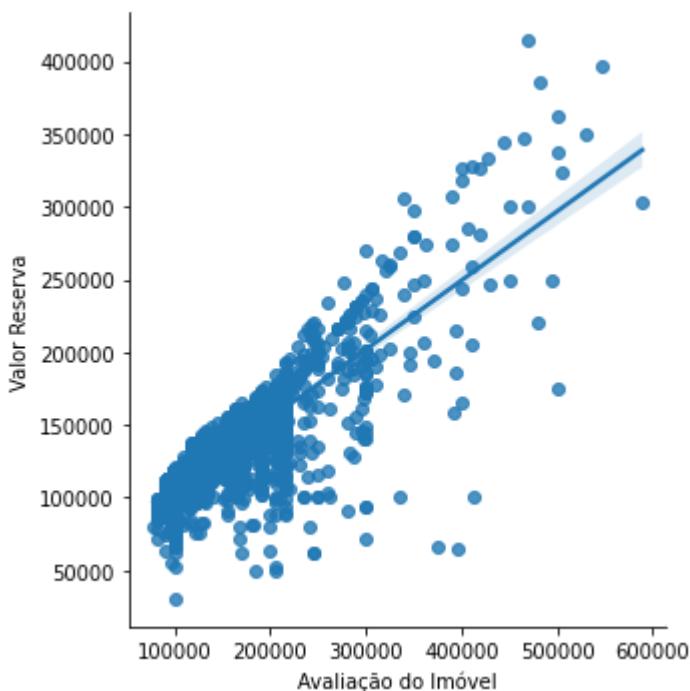


Imagen 5: correlação de Pearson entre “Valor Reserva” e “Avaliação do Imóvel”.

O valor de recursos próprios aportados pode ser interpretado como um valor que o comprador aplicou para render seu dinheiro e evoluir seu patrimônio. Não está especificado onde o dinheiro está aportado, mas sua correlação de Pearson (0.704) mostra sua relevância para a compra de um novo imóvel. Esse valor é conhecido pelo Governo como uma fonte de renda para a aquisição da casa/apartamento e por isso é tão relevante para a compra, assim como o valor reserva apontado pelo comprador.

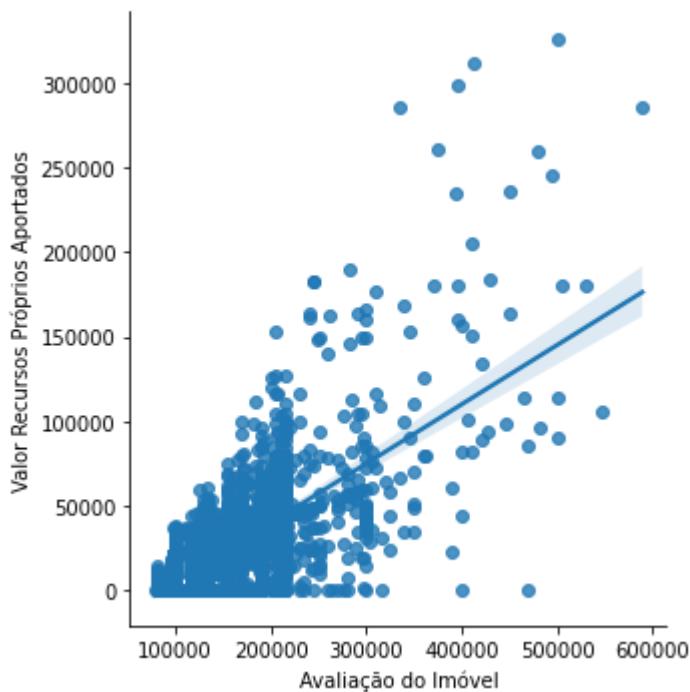


Imagen 6: correlação de Pearson entre “Valor Recursos Próprios Apurados” e “Avaliação do Imóvel”.

O atributo “Rating”, apesar de ser categórico, também possui correlação positiva com o valor do imóvel. É de se esperar que pessoas com melhor avaliação e menos dívidas com o governo irão conseguir maior financiamento para auxiliar na compra do seu imóvel.

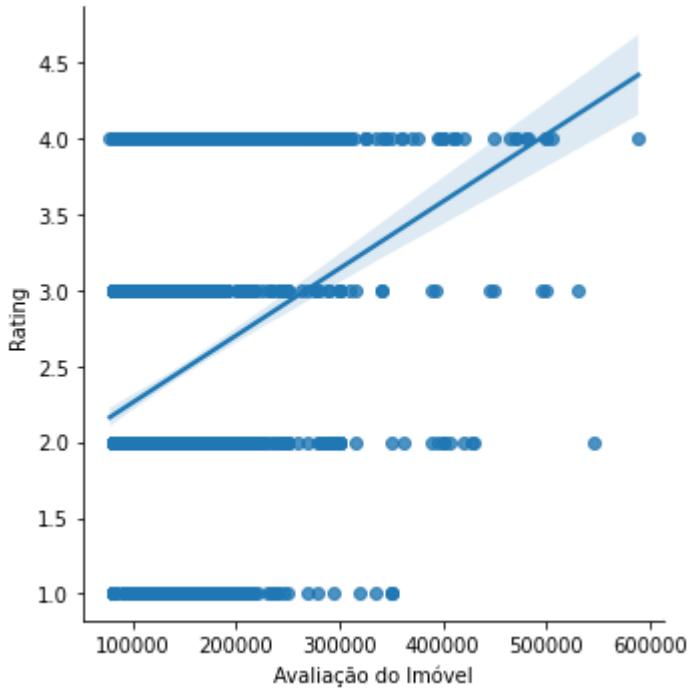


Imagen 7: correlação de Pearson entre “Rating” e “Avaliação do Imóvel”.

#### 4. Conclusões

A análise exploratória anterior deixou evidente quais atributos pessoais mais se relacionam com os atributos do imóvel, que nesse momento se caracteriza pelo seu valor. Médias foram calculadas para avaliar analisar quanto a população descrita nessa base de dados se aproxima da população real. Dessa forma, dados futuros poderão ser avaliados por um modelo preditivo capaz de auxiliar programas na diminuição do déficit habitacional brasileiro. O modelo deve ser capaz de predizer, dadas as informações dos indivíduos, que imóveis se adequam mais ao seu perfil econômico e social antes de finalizar uma compra definitiva.

Uma forma proposta de atingir o objetivo é utilizar modelos de aprendizado de máquina supervisionado, combinando uma porcentagem dessa base de dados analisada durante esse documento para treinar o modelo e o restante para testá-lo. Existem diversos modelos já existentes, como Random Forest, Knn, Decision Trees, entre outros, que podem ser utilizados para a predição.

## 5. Referências

[1] Estudo sobre pretenção de compra de imóveis:

<https://veja.abril.com.br/coluna/radar/quatro-em-cada-dez-brasileiros-pretendem-comprar-imoveis-ate-2023/>

[2] Programa Casa Verde e Amarela:

[https://emccamp.com.br/blog/programa-casa-verde-e-amarela/?gclid=Cj0KCQjwj7CZBhDHA\\_RlsAPPWv3dMBghmFz8ZXFzfQIQigiKXU0E3Jo4YVyGCgfyfH1LvfeoEYi4TUWsaAq-TEALw\\_wcB](https://emccamp.com.br/blog/programa-casa-verde-e-amarela/?gclid=Cj0KCQjwj7CZBhDHA_RlsAPPWv3dMBghmFz8ZXFzfQIQigiKXU0E3Jo4YVyGCgfyfH1LvfeoEYi4TUWsaAq-TEALw_wcB)

[3] Zillow Estimate, ou Zestimate:

<https://analyticsindiamag.com/zestimate-algorithm/>

<https://www.prnewswire.com/news-releases/zillow-launches-new-neural-zestimate-yielding-major-accuracy-gains-301312541.html>

[4] Custo médio de imóvel no Brasil:

<https://noticias.r7.com/economia/pesquisa-apartamento-padrao-custa-em-media-r-458-mil-no-pais-04012018>