

Relatório - Análise de Série Temporal do Mercado Imobiliário

Eryka Carvalho

Fevereiro, 2025

Introdução

O mercado imobiliário brasileiro tem passado por transformações significativas nos últimos anos, com mudanças no perfil do consumidor e no comportamento das empresas do setor. Segundo a SCOD Brasil [?], cerca de 50% dos consumidores de imóveis estão na faixa etária entre 38 e 58 anos. Com base nessa informação, este relatório tem como objetivo analisar a série temporal da razão entre o número de empresas ativas e a população nessa faixa etária, por estado, no período de 2007 a 2022.

A análise inclui a estimativa dos dados para os anos de 2021 e 2022, utilizando modelos de séries temporais, e a identificação de estados com comportamentos similares. Além disso, serão identificados os estados que apresentam maior saturação no mercado e aqueles que oferecem oportunidades futuras.

Metodologia

Coleta de Dados

Para realizar a análise, foram coletados dados de duas fontes principais:

- **Dados das Empresas:** Utilizamos a Tabela 1757 do SIDRA, que contém informações sobre as empresas de construção civil, segundo as faixas de pessoal ocupado. Os dados foram obtidos diretamente da API do SIDRA, utilizando requisições HTTP, sem o uso de bibliotecas prontas como o *sidrapy*.
- **Dados Populacionais:** A projeção populacional por sexo e idade foi obtida do IBGE, na tabela "População por sexo e idade simples". Para ajustar a faixa etária de 38 a 58 anos, foi aplicada uma interpolação linear.

Processamento e Análise

- **Interpolação da Faixa Etária:** A faixa etária de 38 a 58 anos foi interpolada a partir dos dados disponíveis, utilizando o método de interpolação linear.
- **Integração dos Dados:** Os dados das empresas e da população foram integrados em um único dataframe, onde foi calculada a razão entre o número de empresas ativas e a população na faixa etária de interesse.
- **Análise de Séries Temporais:** Para estimar os valores de 2021 e 2022, foi utilizado o modelo ARIMA (AutoRegressive Integrated Moving Average), que é amplamente utilizado para previsão de séries temporais.
- **Agrupamento de Estados:** Utilizamos o algoritmo K-Means para agrupar os estados com base no comportamento temporal da razão calculada. O número ideal de clusters foi determinado pelo método do cotovelo.
- **Identificação de Saturados e Oportunidades:** Estados com razões altas foram considerados saturados, enquanto aqueles com razões baixas foram identificados como oportunidades futuras.

Desenvolvimento

Coleta e Preparação dos Dados

Os dados das empresas foram obtidos diretamente da API do SIDRA, utilizando requisições HTTP. A função `baixar_dados_empresas()` foi implementada para realizar o download dos dados e salvá-los em formato JSON. Da mesma forma, os dados populacionais foram baixados da tabela do IBGE e salvos em formato XLS.

Interpolação e Integração

A interpolação da faixa etária de 38 a 58 anos foi realizada utilizando a função `interpolar_faixa_etaria()`, que aplica uma interpolação linear para estimar os valores para os anos de 2021 e 2022. Os dados das empresas e da população foram então integrados em um único dataframe, onde foi calculada a razão entre o número de empresas ativas e a população na faixa etária de interesse.

Análise de Séries Temporais

A série temporal da razão foi analisada utilizando o modelo ARIMA. Os dados de 2007 a 2020 foram utilizados para treinar o modelo, que foi então utilizado para prever os valores de 2021 e 2022. A previsão foi realizada com a função `forecast()` do modelo ARIMA.

Agrupamento de Estados

Para agrupar os estados com base no comportamento temporal da razão, foi utilizado o algoritmo K-Means. O número ideal de clusters foi determinado pelo método do cotovelo, que consiste em plotar a soma dos quadrados das distâncias (WCSS) em função do número de clusters e identificar o ponto de "cotovelo".

Identificação de Saturados e Oportunidades

Os estados foram classificados como saturados ou com oportunidades futuras com base na média da razão calculada. Estados com razões altas foram considerados saturados, enquanto aqueles com razões baixas foram identificados como oportunidades.

Conclusão

A análise realizada permitiu identificar os estados com maior saturação no mercado imobiliário, bem como aqueles que apresentam oportunidades futuras. A utilização de modelos de séries temporais e técnicas de agrupamento proporcionou uma visão clara do comportamento do mercado ao longo dos anos. Os resultados obtidos podem ser utilizados para orientar estratégias de investimento e políticas públicas no setor imobiliário.

Referências

- SCOD Brasil. Tendências do mercado imobiliário para 2023. Disponível em: <https://scod.com.br/blog/post/tendencias-do-mercado-imobilirio-para-2023>.
- IBGE. Projeção da População. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html>.
- SIDRA. API SIDRA. Disponível em: <https://apisidra.ibge.gov.br/home/ajuda>.