



Universidade do Minho

Unidade Curricular de Análise Inteligente em Sistemas de "Big Data"

**Investigar a relação da inflação com o custo de habitação,
alimentação e o crescimento populacional nos EUA**

Grupo 12

Rui Guilherme Monteiro (PG50739), Rodrigo Rodrigues (PG50726), José Gonçalves (PG50519)

Junho, 2023

Data de Recepção	
Responsável	
Avaliação	
Observações	

Investigar a relação da inflação com o custo de habitação, alimentação e o crescimento populacional nos EUA

Resumo

A habitação é uma necessidade básica, e o seu custo tem implicações significativas para os indivíduos, a economia e a sociedade como um todo. Compreender como a inflação e o crescimento populacional influenciam os preços das moradias é essencial para formular políticas econômicas e sociais eficazes. Este estudo investiga a interação complexa entre a inflação, o custo de habitação, o custo de alimentação e o crescimento populacional nos Estados Unidos da América, utilizando dados históricos abrangentes e métodos analíticos rigorosos, para examinar o período desde os anos 1970 até os dias atuais. Com este documento pretende-se sensibilizar o leitor para a compreensão mais abrangente das dinâmicas entre inflação, custo de habitação e de alimentação e crescimento populacional nos Estados Unidos, fornecendo *insights* valiosos para formuladores de políticas, economistas e investigadores interessados no tema. Com base nessas descobertas, espera-se que a tomada de decisões informada e estratégica possa ser facilitada, visando um mercado habitacional mais estável e inclusivo nos Estados Unidos da América.

Área de Aplicação: Economia, Políticas públicas, Mercado imobiliário, Planeamento urbano, Estratégias habitacionais

Palavras-Chave: *Big Data*, Inflação, Habitação, População, Estados Unidos da América, Python, Apache Spark, Power BI, MongoDB, *Dashboards*, *Big data tools*

Índice

<u>1. Introdução</u>	6	
<u>1.1. Contextualização</u>	6	
<u>1.2. Apresentação do Caso de Estudo</u>	6	
<u>1.3. Motivação e Objectivos</u>	7	
<u>1.4. Estrutura do Relatório</u>	7	
<u>2. Estado da Arte</u>	9	9
<u>2.1 Crescimento populacional e a sua relação com a inflação imobiliária</u>	10	
<u>2.2 Contribuições do presente estudo</u>	11	
<u>3. Arquitetura</u>	13	
<u>3.1 Análise dos Dados</u>	14	
<u>3.2 Tratamento de Dados e Merge dos Datasets</u>	15	
<u>3.3 Armazenamento dos dados</u>	17	
<u>3.4 Conexão com PowerBI</u>	20	
<u>3.5 Visualização dos dados</u>	20	
<u>4. Resultados e Discussão</u>	21	
<u>5. Conclusões e Trabalho Futuro</u>	31	
<u>Referências</u>	33	

1.Introdução

1.1.Contextualização

Nos últimos anos, o tema da inflação tem sido objeto de grande preocupação e discussão ao redor do mundo, mas especialmente nos Estados Unidos da América. O aumento constante nos preços dos bens e serviços tem levantado questões sobre o impacto que isso tem na vida cotidiana dos americanos. Além disso, a inflação pode vir a afetar outros fatores, como o acesso à habitação e o crescimento populacional.

Nesse contexto, o uso de análise inteligente de *Big Data* pode ser uma ferramenta poderosa para entender melhor essa relação. Analisando grandes conjuntos de dados, é possível identificar tendências e padrões que não seriam facilmente perceptíveis em uma análise mais limitada.

1.2.Apresentação do Caso de Estudo

Assim, no presente trabalho, será realizada uma investigação da relação da inflação com o custo de habitação e o crescimento populacional nos Estados Unidos, utilizando técnicas de *Big Data*. Serão analisados conjuntos de dados relevantes para essas áreas, como índices de preços, dados do mercado imobiliário e informações demográficas.

Para isso, foram utilizados conjuntos de dados de fontes confiáveis, como o *U.S. Bureau of Labor Statistics*. Foram escolhidos dados sobre índices de preços, como o Índice de Preços ao Consumidor (*Consumer Price Index*, CPI), dados do mercado imobiliário em diferentes zonas urbanas dos Estados Unidos da América (E.U.A.) (*House Price Index*, HPI) e dados relativos ao custo de alimentos, quer seja comida ou bebida, nos EUA (*Food and Beverage*).

1.3.Motivação e Objectivos

A inflação é um tema complexo que afeta a economia e a vida cotidiana das pessoas em todo o mundo. Com o aumento dos preços dos bens e serviços, muitos indivíduos e empresas enfrentam dificuldades financeiras, o que pode levar a um declínio na economia. Nesse contexto, a análise inteligente de *Big Data* pode ajudar a entender melhor essa relação e fornecer *insights* valiosos para tomadores de decisão, formuladores de políticas e empresas. Por essa razão, este relatório técnico visa explorar o uso de *Big Data* na análise da inflação relacionada ao custo da habitação e ao crescimento populacional nos Estados Unidos da América.

No que toca aos objetivos deste projeto, eles são os seguintes: analisar o conceito de inflação e seus efeitos na economia e na vida cotidiana dos americanos; explorar a importância da análise inteligente em sistemas de *Big Data* na compreensão da inflação e seus impactos; identificar e analisar conjuntos de dados relevantes que possam ser utilizados na análise da inflação relacionada com o acesso à habitação e o crescimento populacional nos Estados Unidos da América; identificar as principais tendências e padrões que relacionam a inflação, o custo habitacional e o crescimento populacional; explorar as implicações da inflação no custo habitacional e como isso pode afetar o acesso à moradia para diferentes segmentos da população.

1.4.Estrutura do Relatório

Capítulo 1: Introdução

Neste capítulo introdutório, são apresentados o contexto e o objetivo deste relatório. É feita uma breve descrição da importância da relação entre a inflação nos preços das habitações e o crescimento populacional nos Estados Unidos, no contexto de *Big Data*. Além disso, são apresentados os principais tópicos que serão abordados nos capítulos seguintes.

Capítulo 2: Estado da Arte

Este capítulo oferece uma visão geral dos trabalhos mais relevantes e recentes relacionados à relação entre a inflação nos preços das habitações e o crescimento populacional nos Estados Unidos, no contexto de *Big Data*. São abordados três tópicos

principais: inflação nos preços das habitações, crescimento populacional e sua relação com a inflação imobiliária, e as contribuições específicas do presente estudo.

Capítulo 3: Arquitetura

Neste capítulo, é apresentada a arquitetura utilizada no estudo. São detalhadas as etapas e processos envolvidos na análise dos dados, incluindo a análise dos dados, tratamento e *merge* do *dataset*, armazenamento do *dataset*, conexão com o Power BI e a visualização dos dados. Também são discutidos os resultados obtidos e suas respectivas discussões.

Capítulo 4: Resultados e Discussão

No último capítulo, é feito um resumo dos principais pontos abordados no relatório. São apresentadas as conclusões e os insights derivados da análise dos dados e discutidas as possíveis interpretações e limitações do estudo. Esse capítulo encerra o relatório, oferecendo uma visão geral dos resultados obtidos e sua relevância para o tema em questão.

2.Estado da Arte

Nesta secção culmina a pesquisa sobre a relação entre a inflação nos custos habitacionais, custos de alimentação e o crescimento populacional nos Estados Unidos, no contexto de Big Data. É fornecida uma visão geral dos trabalhos mais relevantes e recentes relacionados ao tema, posicionando o estudo atual dentro do contexto da pesquisa existente e estabelecendo a base para a contribuição do trabalho.

2.1. Crescimento populacional e sua relação com a inflação imobiliária

O crescimento populacional é um fator importante a ser considerado na análise da inflação nos preços das habitações. Diversos estudos têm explorado essa relação, utilizando técnicas de Big Data para obter insights mais detalhados.

Um estudo conduzido por Ye Jin Heo (2022) intitulado *“Population aging and house prices: Who are we calling old?”* [\[1\]](#) investigou a relação entre o envelhecimento populacional e os preços reais das habitações em 21 países da OCDE. Neste estudo, foi adotada uma abordagem empírica, redefinindo a razão de dependência de idosos utilizando a idade efetiva de aposentadoria e os anos restantes de expectativa de vida, a fim de explorar o efeito heterogêneo do envelhecimento nos preços reais das habitações. Os resultados revelaram que um aumento na razão de dependência com base nos anos restantes de expectativa de vida está associado a uma diminuição nos preços reais das habitações, enquanto a razão de dependência da população efetivamente aposentada não apresenta essa associação. Ao segmentar os grupos de "jovens-idosos" e "idosos-idosos", foi confirmado que a associação negativa com os preços reais das habitações é impulsionada pelo aumento da razão de dependência do grupo de "idosos-idosos". Esses achados sugerem, de forma geral, que **o envelhecimento populacional não implica necessariamente** uma diminuição contínua nos preços reais das habitações, uma vez que **o efeito negativo está relacionado à população muito idosa**, com uma expectativa de vida sendo curta.

O artigo "Demographics and the Housing Market: Japan's Disappearing Cities"

da autoria de Ms. Yuko Hashimoto et al. (2020) [\[2\]](#) investiga a relação entre o crescimento populacional e a inflação imobiliária no Japão. O estudo revela que as tendências demográficas desempenham um papel importante na disparidade regional. Enquanto as grandes cidades experimentam crescimento populacional devido à migração interna, as áreas rurais sofrem com o declínio populacional devido a baixas taxas de fertilidade e emigração. Os preços das habitações têm declinado persistentemente desde a década de 1990, mas de forma desigual, com as cidades maiores apresentando quedas menores. A pesquisa também mostra que **as mudanças na população afetam os preços das habitações de forma assimétrica**, com quedas mais acentuadas em áreas com **declínio populacional**. Essa dinâmica pode levar a disparidades regionais maiores e impactar a economia japonesa, afetando a riqueza das famílias e a saúde financeira dos bancos. O estudo destaca a **necessidade de políticas que promovam um crescimento mais equilibrado** entre as regiões e que evitem quedas excessivas nos preços das habitações nas áreas rurais.

Um terceiro e último estudo [\[3\]](#), intitulado "Housing Price and Population Growth across China: The Role of Housing Supply", foi conduzido por Hongbo Wang e Dan S. Rickman da Oklahoma State University - Stillwater. O estudo, realizado em março de 2019, emprega um modelo de crescimento de equilíbrio espacial para examinar empiricamente o papel do crescimento da oferta de habitação nas diferenças de preço da habitação e crescimento populacional nas províncias, regiões autónomas, municípios e principais cidades da China continental no período de 2002 a 2015. Os resultados revelaram que áreas do Leste, como Shanghai, Ningbo, Qingdao e Xiamen, apresentaram o menor crescimento na oferta de habitação, enquanto as regiões autónomas e áreas do Sudoeste e Nordeste apresentaram o maior crescimento. Essas diferenças no crescimento da oferta de habitação não apenas **influenciaram significativamente o crescimento** relativo dos preços da habitação, mas também tiveram um grande impacto no crescimento populacional relativo das regiões, o que sugere que **políticas de oferta de terras e habitação são componentes críticos do crescimento regional** na China.

2.2. Crescimento populacional e sua relação com a inflação na alimentação

O artigo "Does inflation kill? Exposure to food inflation and child mortality" da autoria de Daniel Kidane, Andinet Woldemichael (2020) [\[4\]](#) investiga o impacto da exposição à inflação dos preços dos alimentos na mortalidade infantil, durante os períodos críticos da primeira infância. Os pesquisadores utilizam dados da Etiópia, combinando informações do Inquérito Demográfico e de Saúde (DHS) com dados de preços de alimentos locais em alta frequência. Os resultados mostram que a exposição a uma alta inflação dos preços dos alimentos durante a gestação e a infância aumenta significativamente a mortalidade infantil. Esses impactos são não lineares e variam de acordo com o mês específico de exposição. Os resultados são consistentes com estudos anteriores e robustos a diferentes especificações empíricas. Esses resultados têm importantes implicações na formulação de intervenções eficazes para mitigar o impacto da exposição à alta inflação dos preços dos alimentos no bem-estar das crianças, que são os membros mais vulneráveis da sociedade. As políticas públicas existentes podem ser aprimoradas para proteger mulheres grávidas e crianças pequenas durante episódios de inflação. É crucial implementar ações políticas fortes para reduzir os efeitos na saúde e mortalidade infantil, especialmente diante da persistência da alta inflação dos preços dos alimentos ao longo da última década. Programas de segurança alimentar existentes podem ser ajustados para atender a famílias com mulheres grávidas e bebês, visando reduzir o impacto da inflação dos alimentos, que é mais pronunciado entre crianças expostas durante a gestação e os primeiros seis meses de vida. Embora o estudo se concentre na Etiópia, os resultados também são relevantes para outros países em desenvolvimento que enfrentaram aumentos nos preços dos alimentos nas últimas décadas.

2.3. Contribuições do presente estudo

Apesar dos avanços mencionados anteriormente, ainda há lacunas significativas no conhecimento sobre a relação entre a inflação nos preços das habitações e o crescimento populacional nos Estados Unidos. O presente estudo visa explorar essas

lacunas, utilizando técnicas de Big Data para obter insights precisos.

O presente estudo utilizará um conjunto de dados que integra índices de preços de imóveis que medem as variações de preços de imóveis residenciais nos Estados Unidos da América, dados de inflação e informações relevantes sobre crescimento populacional. Combinando métodos de análise de dados e uma abordagem multidisciplinar, espera-se que este estudo ofereça insights sobre a relação complexa entre a inflação nos preços das habitações e o crescimento populacional nos Estados Unidos, contribuindo para o avanço do campo e fornecendo orientações para políticas públicas relacionadas ao mercado imobiliário.

Cada uma das ferramentas possui, inerentemente, as suas características tornando-se assim necessário efetuar uma análise prévia das tecnologias que servirão de suporte para a arquitetura a ser implementada. A decisão da arquitetura a adotar teve em conta o peso de cada um dos artigos analisados, tendo sido comparadas as várias abordagens de modo a alcançar um entendimento das ferramentas adequadas para fornecer uma abordagem capaz de suportar o problema em mãos. Tendo escolhido as tecnologias surge naturalmente o processo de criação de um pipeline, cujos detalhes serão apresentados em maior detalhe na secção seguinte. Os artigos expandem o estudo prévio tanto do tema como das ferramentas de big data, materializando o mesmo numa solução de arquitetura para o caso de uso definido.

3. Arquitetura

Para responder ao caso de uso apresentado foi implementada uma arquitetura que é descrita em várias fases. Na primeira fase, os dados foram analisados e tratados, procurando eliminar erros e inconsistências. Após este tratamento, passou-se à junção dos dados (*merge*), onde se uniram os *datasets* utilizados no estudo e, de seguida, na terceira fase procedeu-se ao armazenamento do *dataset* resultante numa base de dados MongoDB. Na quarta e última fase realizou-se visualização de dados, respetivamente. A figura abaixo representa visualmente a estrutura e etapas do pipeline implementado.

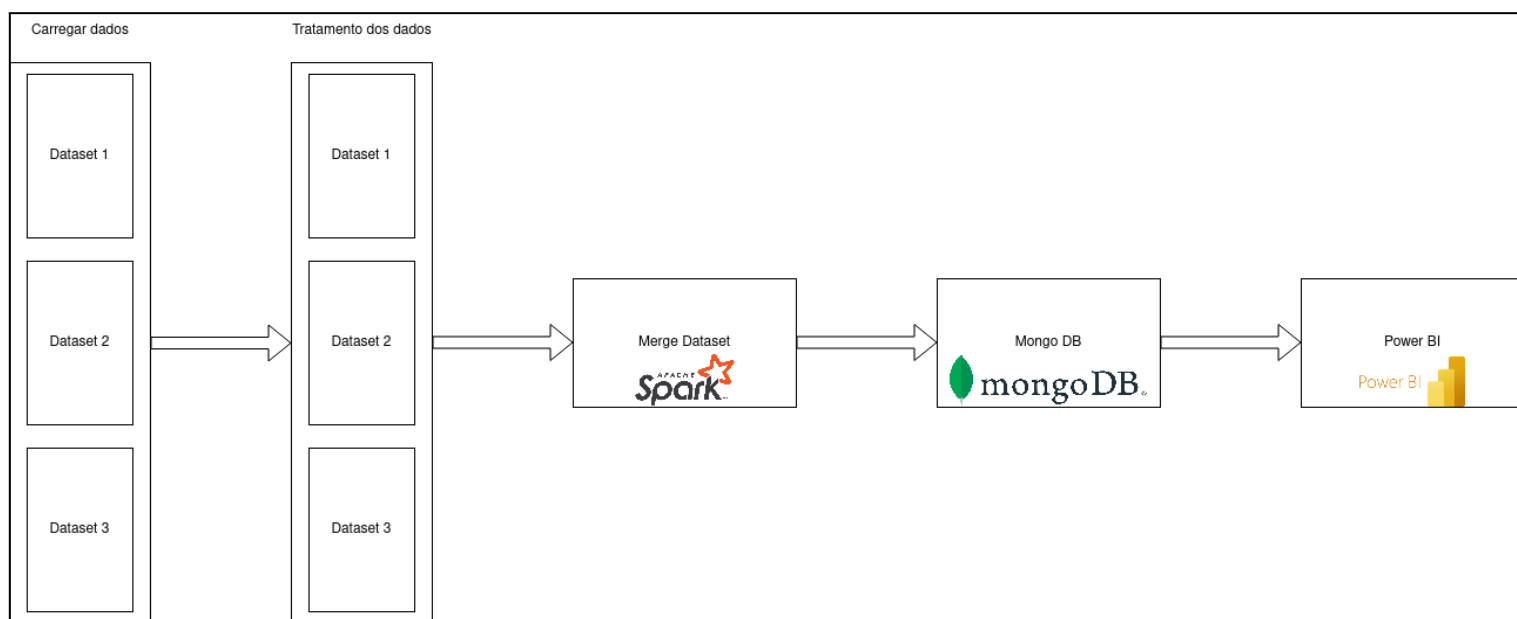


Figura 1 - Esquema de arquitetura do projeto

3.1. Análise dos dados

Neste projeto foram utilizadas diferentes fontes de dados. Para atingir os objetivos propostos, foram selecionados três conjuntos de dados. O primeiro é inerente a dados relativos à população nos E.U.A recolhidos da informação oficial disponibilizada pelo governo através do censo nacionais, transformada em formato CSV, possuindo dezenas de milhares de linhas com informações para cada estado. Seguidamente, o segundo é chamado de *House Price Index*, ou seja, contém um índice do preço das casas em vários locais urbanos dos E.U.A. em formato JSON e possui 121 mil linhas e 10 colunas. Por último, o *dataset* utilizado como base para a inflação é chamado de *Food & Beverage*, e possui um índice do preço da alimentação em geral nos E.U.A. em formato CSV com 119 mil linhas e 5 colunas. Nestes *datasets* em formato CSV, a informação é relativa ao *Consumer Price Index*.

3.1.1. Dataset 1 - US Population Size

Ao longo dos anos, os censos têm desempenhado um papel fundamental na coleta de informações demográficas precisas e abrangentes em todo o mundo. Os Estados Unidos são conhecidos pela sua ênfase na recolha de dados precisos e confiáveis sobre sua população. Não havendo disponibilidade de um *dataset* abrangente que cobrisse o período de 1975 a 2010, decidiu-se construir um *dataset* extenso juntando os dados fornecidos pelo governo dos EUA ao longo das décadas no site oficial do *Census Bureau dos Estados Unidos*, onde se pode encontrar uma fonte valiosa de dados, com informações detalhadas sobre a população para cada estado em diferentes momentos. Esses ficheiros incluem variáveis como o ano, o estado, a população total e outras informações relevantes.

No entanto, trabalhar com tantos dados foi um desafio. Primeiro, foi necessário extrair e analisar cuidadosamente os ficheiros CSV para garantir que os dados estivessem formatados corretamente e que todas as informações relevantes fossem incluídas. Em seguida, criou-se um script personalizado para limpar os dados e garantir a consistência em todo o conjunto de dados, bem como agrupar dados por ano, estado e *age group* (jovens, adultos, idosos). Após a limpeza dos dados, construiu-se um conjunto de dados unificado que continha informações sobre a população total de

cada estado dos EUA entre 1975 e 2010.

Esses tipos de dados estruturados permitem análises estatísticas e a aplicação de diferentes técnicas para explorar as relações entre o tamanho da população e outros fatores, como inflação e custo de habitação, conforme mencionado no resumo do estudo proposto.

3.1.2. Dataset 2 - Food & Beverages

O conjunto de dados sobre Food & Beverage (Alimentos e Bebidas) é baseado no Índice de Preços ao Consumidor (CPI) compilado pelo Bureau of Labor Statistics (BLS) dos Estados Unidos. O CPI é uma medida estatística que acompanha as mudanças nos preços de bens e serviços ao longo do tempo em grupos de gastos importantes, como alimentação, moradia, vestuário, transporte e cuidados médicos, geralmente adquiridos por consumidores urbanos.

O *dataset* apresenta informações sobre a inflação em produtos alimentares ao longo das décadas, em períodos mensais. Cada entrada no conjunto de dados inclui o valor do índice de preços para alimentos e bebidas naquele período específico. As informações fornecidas são representadas por ano, período (mês) e valor do índice. Esses dados podem ser usados para analisar as flutuações nos preços de alimentos e bebidas ao longo dos anos e identificar possíveis tendências ou padrões.

3.1.3. Dataset 3 - House Price Index

O FHFA House Price Index (FHFA HPI®) é uma coleção abrangente de índices públicos e de preços de residências que medem as mudanças nos valores de residências unifamiliares com base em dados de todos os 50 estados e mais de 400 cidades americanas, desde meados da década de 1970 até 2023.

3.2. Etapa 1 - Tratamento de Dados e Merge do Dataset

O pipeline começa por utilizar a biblioteca Pandas da linguagem Python para fazer realizar o tratamento dos conjuntos de dados provenientes dos censos do EUA. Algumas etapas de tratamento incluem renomear colunas, limpar e transformar valores em determinadas colunas, agrupar e somar dados do mesmo ano/estado. No final, um novo DataFrame é criado para armazenar os dados agregados. O objetivo

geral é preparar os dados para uma análise posterior, facilitando a compreensão e o uso das informações contidas no conjunto de dados. De seguida, pretende-se tratar os dados dos 3 datasets mencionados e realizar o *merge* dos *datasets* escolhidos para este estudo, utilizando PySpark. Esta biblioteca permite a representação de tabelas com vários tipos de dados através do uso de *dataframes* que fornecem ao utilizador maior facilidade na manipulação destes objetos. O *merge* dos datasets foi feito com base nas colunas “Year”, “State” e “State_id” para assim unir os novos atributos (colunas) de cada dataset juntando entradas relativas ao mesmo ano no mesmo estado.

3.2.1. Renomeação de colunas

Uma vez que os *datasets* CSV e JSON apresentam, à partida, colunas diferentes ou com nomes distintos, foi necessário fazer o devido tratamento. Neste caso, renomeou-se a coluna “yr” do dataset JSON para “Year” de forma a ficar igual à coluna do *dataset* CSV para efetuar o *merge* com sucesso.

Em ambos os *datasets* foi também necessário proceder ao tratamento de colunas, nomeadamente, transformar o tipo de dados de determinadas colunas como “index_sa” e “period” em valores numéricos, onde no segundo, se implementou um mapeamento de valores string para números (ex: 'M01': 1).

3.2.2. Registos anteriores a 1975

Visto que nem todos os *datasets* possuem informação para determinados anos, filtrou-se apenas as entradas de dados com ano igual ou superior a 1975, uma vez que a partir desse ano todos os *datasets* em análise contêm dados.

3.2.3. Redução da dimensionalidade e período considerado para análise

Uma vez que se está a trabalhar com conjuntos de dados com centenas de milhares de entradas e, especialmente ao fazer *merge*, ficou-se na presença de um elevado número de entradas no nosso *dataset* resultante, foi decidido tratar os dados com as médias anuais em vez de fazer uma análise tão minuciosa ao nível do mês. Assim sendo, o estudo da relação da inflação com o crescimento populacional foi feito ao nível do ano, ou seja, analisando os dados agregados de ano para ano ao longo de várias décadas.

3.2.4. Tratamentos de valores nulos, duplicados e remoção de colunas redundantes

De forma a obter um *dataset* resultante do *merge* com maior qualidade possível numa fase inicial, procedeu-se a tratamento de Nan Values com a devida substituição pelo valor da média na coluna “*index_sa*”. Adicionalmente, removeram-se valores nulos e duplicados, bem como colunas desnecessárias para a nossa análise, isto é, cuja informação que disponibilizam não acrescenta valor significativo à análise em mãos pois o seu objetivo inicial servia outros propósitos (exemplo: coluna “*footnote_codes*” que deveria conter códigos mas neste caso apresenta apenas valores nulos).

3.2.5. Conversão para formato Parquet e *inner join*

Tendo tratado e visualizado o resultado do tratamento dos *datasets* do caso de estudo, procedeu-se à conversão dos mesmos para formato *Parquet* e, de seguida, à realização da união dos mesmos com auxílio do *Apache Spark*, mais especificamente com a utilização da função “*join*” para efetuar um *inner join* pela coluna *year*.

3.3. Etapa 2 - Armazenamento do Dataset

De modo a armazenar a informação, permitindo não só garantir a integridade da mesma como a sua disponibilidade em qualquer dispositivo, recorreu-se ao Sistema de gestão de Bases de Dados MongoDB, uma ferramenta *noSQL* de formato documental.

A script designada para este efeito é escrita em Python e utiliza a biblioteca PyMongo para estabelecer conexão e importar os dados para o MongoDB. Uma vez

que este trabalha com um formato documental, o objetivo principal passa por importar dados agrupados por ano armazenando-os de forma eficiente e escalável.

```
_id: ObjectId('647b871d86c7c0e2249c5f53')
year: 1975
state_id: 1
state: "Alabama"
kids: 1961900
adults: 1333001
retired: 383915
populationSize: 3678816
index_housing: 73.39
index_food: 60.19
```

```
_id: ObjectId('647b871d86c7c0e2249c5f54')
year: 1976
state_id: 1
state: "Alabama"
kids: 1986831
adults: 1353185
retired: 395123
populationSize: 3735139
index_housing: 77.07
index_food: 62.08
```

...

```
_id: ObjectId('647b871d86c7c0e2249c5f77')
year: 1975
state_id: 2
state: "Alaska"
kids: 230269
adults: 137107
retired: 8797
populationSize: 376173
index_housing: 65.6
index_food: 60.19
```

```
_id: ObjectId('647b871d86c7c0e2249c5f78')
year: 1976
state_id: 2
state: "Alaska"
kids: 245001
adults: 146742
```

Figura 2 - Documentos MongoDB

Os documentos da coleção MongoDB contêm as seguintes informações relacionadas ao crescimento populacional, inflação habitacional e inflação nos alimentos: "year": Representa o ano a que os dados se referem; "state_id": É um identificador numérico para o estado. "state": Indica o nome do estado em que os dados foram recolhidos; "kids": Refere-se à população de crianças no estado no ano especificado. "adults": Indica a população de adultos no estado no ano especificado; "retired": Representa a população de aposentados no estado no ano especificado; "populationSize": Refere-se ao tamanho total da população no estado no ano especificado; "index_housing": É o índice de inflação habitacional para o estado no ano especificado; "index_food": Indica o índice de inflação nos alimentos para o estado no ano especificado.

Essas informações podem ser utilizadas para analisar a relação entre a inflação habitacional e a inflação nos alimentos com o crescimento populacional nos 50 diferentes estados. Com base nesses dados, é possível realizar análises estatísticas e visualizações para compreender melhor os padrões e tendências relacionados a esses fatores.

3.4. Etapa 3 - Conexão com o Power BI

Após ter armazenado os dados tratados na primeira etapa do pipeline na base de dados MongoDB, a próxima etapa é referente à conexão com a ferramenta de análise de dados Microsoft Power BI. Para estabelecer essa conexão, irá ser utilizado o ODBC (Open Database Connectivity). O ODBC é um padrão que permite a conexão com diversos tipos de bases de dados por meio de um driver específico. Neste caso, utilizaremos um driver ODBC adequado para o MongoDB, que permitirá aceder e visualizar os dados no Power BI.

3.5. Etapa 4 - Visualização dos dados

Na etapa 4, realizou-se a análise dos dados no Microsoft Power BI, utilizando as

coleções de dados do MongoDB. Foram criados diversos gráficos e dashboards com o objetivo de explorar a relação entre a inflação e o crescimento populacional nos Estados Unidos ao longo dos anos.

Para visualizar a evolução da inflação ao longo do tempo, criou-se um gráfico de linhas temporais que representa a evolução da inflação a nível de habitação e de alimentação (`index_housing` e `index_food`). Esse gráfico permitiu identificar as variações da inflação ao longo dos anos e observar possíveis tendências. Outro gráfico importante foi o gráfico representativo do crescimento populacional. Nesse gráfico, está representado o crescimento populacional (`populationSize`) para cada ano (1975-2010). Essa visualização ajudou a identificar os anos com maior ou menor crescimento populacional, permitindo uma análise mais precisa dessa variável.

Além disso, foi criado um gráfico de dispersão que relaciona o crescimento populacional e a inflação. Cada ponto nesse gráfico representa uma localidade específica (`place_name`), permitindo identificar possíveis correlações entre essas duas variáveis.

Também foi utilizado um mapa para visualizar as diferenças de inflação nos diferentes estados. Nesse mapa, as localidades foram marcadas geograficamente e cores ou tamanhos diferentes foram utilizados para representar o crescimento populacional ou a inflação em cada localidade. Essa visualização facilitou a identificação de padrões geográficos e a compreensão das diferenças entre regiões.

Todos esses gráficos e dashboards foram reunidos num dashboard comparativo, que permitiu filtrar os dados por ano, localidade ou outros critérios. Dessa forma, foi possível realizar comparações e explorar diferentes aspectos da inflação e do crescimento populacional nos Estados Unidos.

A análise no Power BI proporcionou uma visão abrangente da relação entre a inflação e o crescimento populacional ao longo do tempo. Os *insights* obtidos através dos gráficos e dashboards criados ajudaram a identificar tendências, padrões e possíveis correlações entre as variáveis. Essa compreensão mais profunda dos dados é essencial para o estudo da economia e da demografia nos Estados Unidos.

4.Resultados e Discussão

Nesta secção irão ser apresentadas as dashboards criadas no PowerBI para a visualização de dados para responder ao caso de estudo em questão.

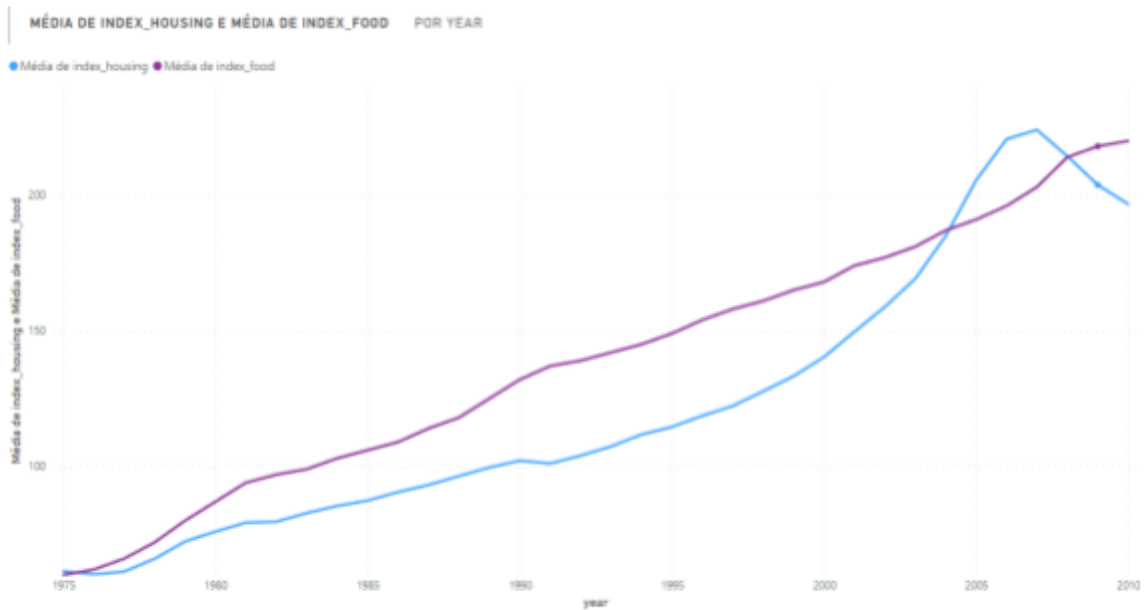
Evolução da Inflação na Habitação e do tamanho da População ao longo dos anos

O seguinte gráfico de linhas mostra a evolução da taxa de inflação ao longo dos anos (1975-2010), em comparação com o crescimento populacional nos Estados Unidos.

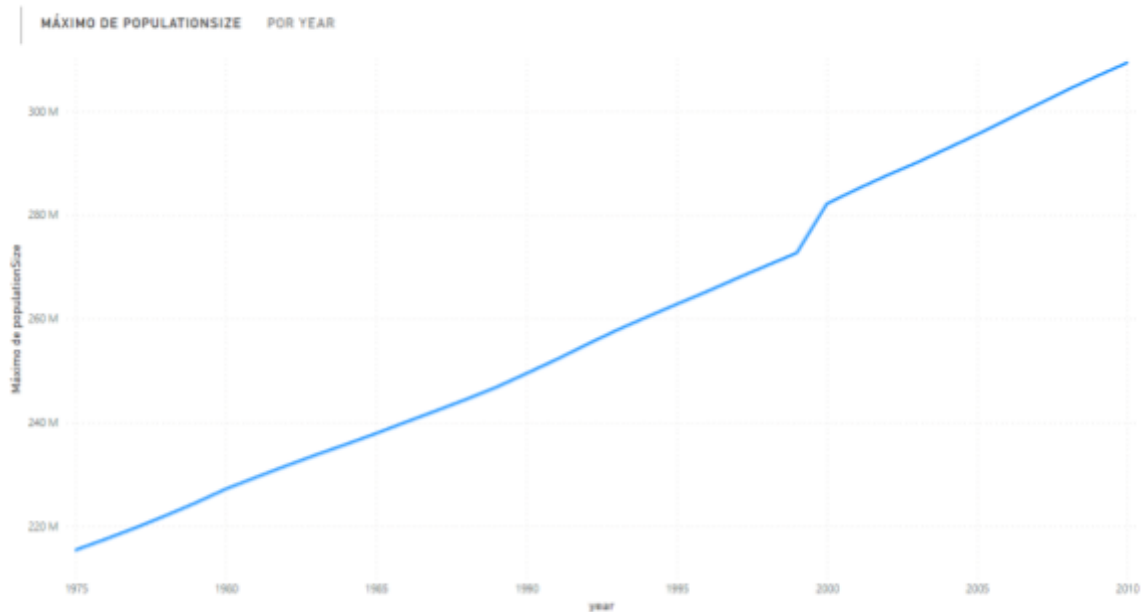
A análise do gráfico de linhas que mostra a evolução da taxa de inflação na habitação e do tamanho da população nos Estados Unidos ao longo dos anos (1975-2010) revela tendências interessantes. Observamos que ambas as variáveis apresentaram um crescimento contínuo ao longo do período analisado.

No que diz respeito à taxa de inflação na habitação, é possível notar um aumento gradual ao longo dos anos, indicando que os preços das habitações experimentaram um crescimento constante durante esse período. Esse padrão sugere a presença de pressões inflacionárias persistentes no setor imobiliário, possivelmente impulsionadas por uma demanda crescente em relação à oferta de habitação.

Por outro lado, a linha que representa o crescimento populacional mostra uma tendência ascendente semelhante. Isso indica que a população dos Estados Unidos também aumentou consistentemente durante o período considerado. Esse crescimento populacional pode ter sido impulsionado por diversos fatores, como migração interna, nascimentos e imigração.



Inflação na habitação e nos alimentos ao longo dos anos



Crescimento populacional ao longo dos anos

No entanto, ao considerarmos as conclusões dos estudos mencionados anteriormente, podemos adicionar algumas perspectivas adicionais à análise do gráfico. Por exemplo, o estudo de Ye Jin Heo (2022) [\[1\]](#) destacou a importância da

composição etária da população na dinâmica dos preços das habitações. Embora o gráfico não forneça dados específicos sobre a estrutura etária da população, é importante reconhecer que mudanças na composição da população, como o envelhecimento populacional, podem influenciar os preços das habitações.

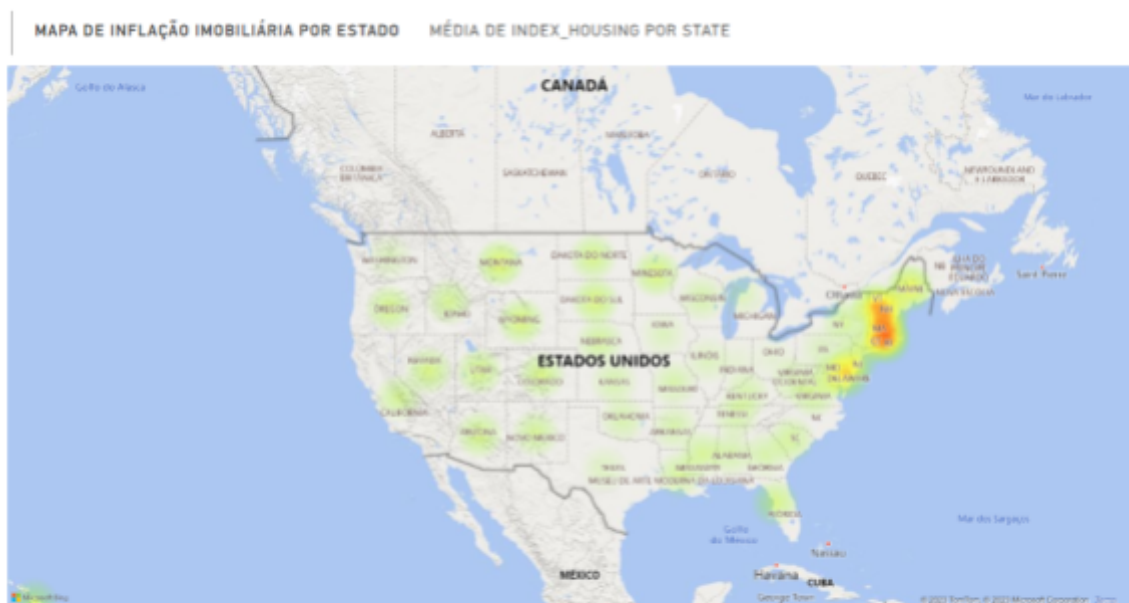
Além disso, o estudo realizado por Ms. Yuko Hashimoto et al. (2020) [\[2\]](#) revelou que as tendências demográficas podem desempenhar um papel significativo na disparidade regional do mercado imobiliário. No contexto dos Estados Unidos, isso sugere que diferentes áreas geográficas podem experimentar dinâmicas distintas nos preços das habitações, dependendo de fatores demográficos e migração interna.

Considerando o estudo de Hongbo Wang e Dan S. Rickman (2019) [\[3\]](#), que examinou a relação entre o crescimento da oferta de habitação e as diferenças nos preços das habitações e no crescimento populacional, é possível inferir que a oferta de habitação também pode desempenhar um papel importante na dinâmica do mercado imobiliário. Se houver desequilíbrios entre a demanda e a oferta de habitação, isso pode afetar os preços das habitações e potencialmente influenciar o crescimento populacional em determinadas regiões.

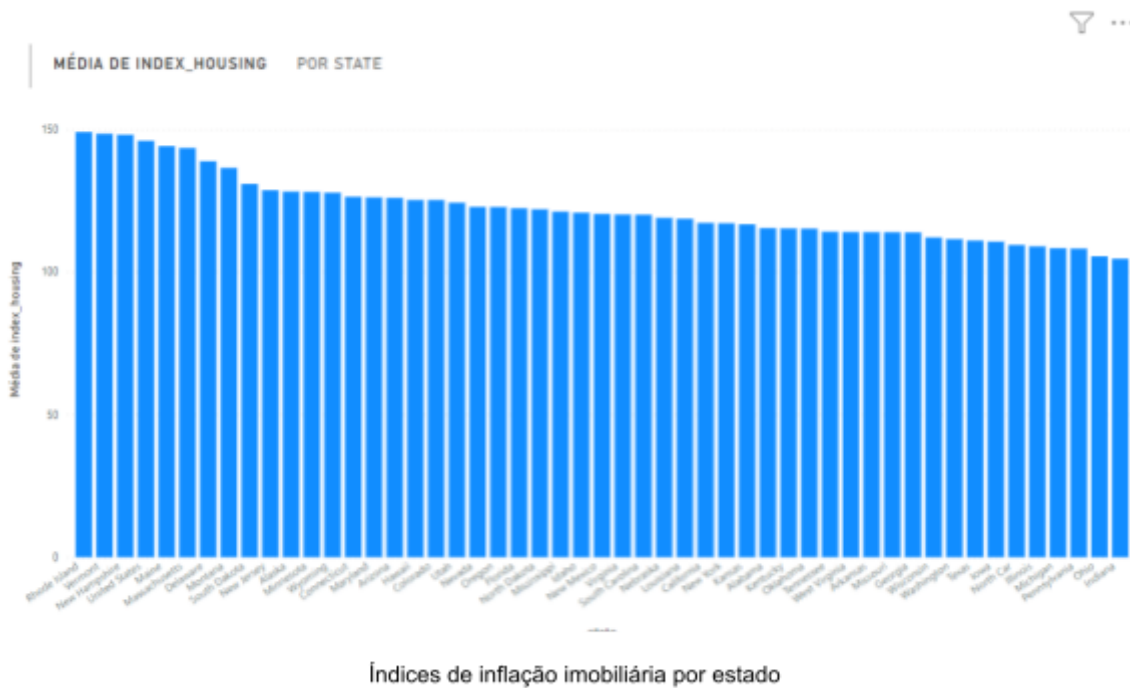
Portanto, ao analisar o gráfico, é importante considerar essas perspectivas adicionais e reconhecer que fatores demográficos, mudanças na oferta de habitação e disparidades regionais podem desempenhar um papel na evolução da inflação na habitação e no crescimento populacional nos Estados Unidos ao longo dos anos.

Mapa da Inflação na Habitação por Estado

Criou-se um mapa interativo que mostra a taxa de inflação em cada estado dos EUA. As cores diferentes representam diferentes faixas de inflação e permitem a exploração dos dados clicando em cada estado para obter informações detalhadas. Ao comparar a taxa de inflação por Estado, pode-se concluir quais os Estados apresentam maior taxa de inflação e selecionar um conjunto de Estados para posteriormente averiguar se essas taxas elevadas se traduzem em acréscimo ou decréscimo populacional.



Mapa de inflação imobiliária por estado



Estados com **inflação imobiliária mais alta**: Rhode Island, Vermont e New Hampshire apresentam as maiores médias de *index_housing*, indicando uma inflação imobiliária significativamente alta. Isso pode ser atribuído a fatores como procura elevada por imóveis, crescimento populacional, restrições na oferta de moradias e atratividade desses estados.

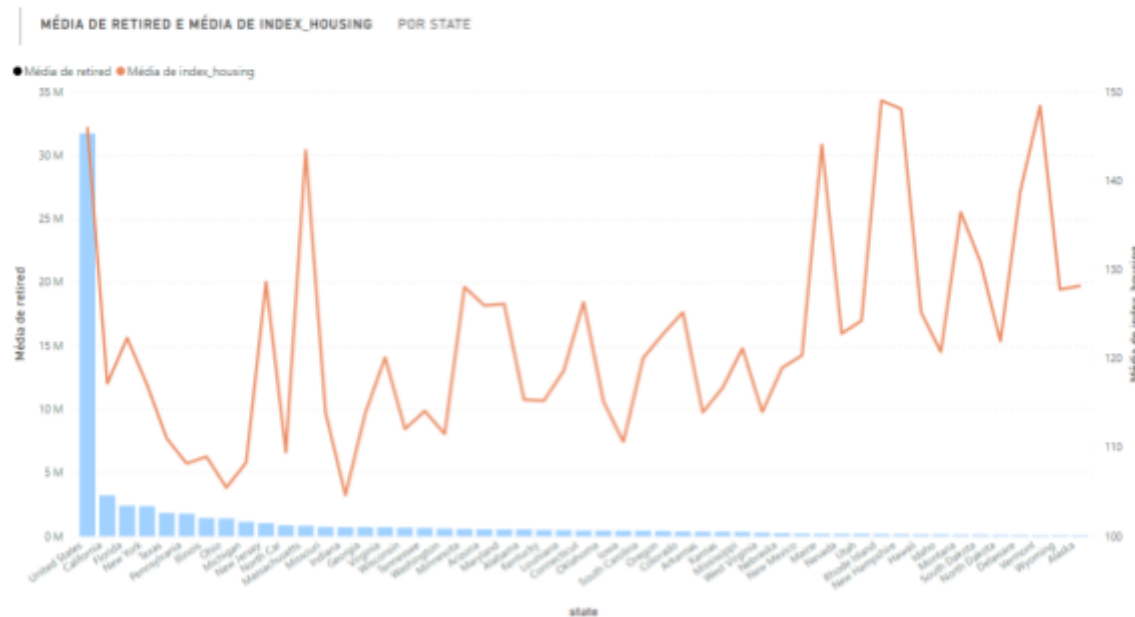
Estados com **inflação imobiliária abaixo da média nacional**: Alguns estados, como Illinois, Michigan, Pennsylvania, Ohio e Indiana, têm médias de *index_housing* abaixo da média nacional dos EUA, sugerindo que esses estados podem ter uma procura mais moderada por imóveis ou uma oferta relativamente maior, resultando numa inflação imobiliária mais baixa.

Variações regionais: Podemos observar algumas tendências regionais na inflação imobiliária. Por exemplo, estados do **Nordeste**, como Rhode Island, Vermont, New Hampshire e Maine, apresentam **altos índices de inflação imobiliária**, enquanto estados do **Meio-Oeste**, como Illinois, Michigan, Ohio e Indiana, têm **índices relativamente mais baixos**. Essas variações podem estar relacionadas a diferenças econômicas, demográficas e políticas entre as regiões.

Estados de **alta inflação imobiliária populares**: Alguns estados conhecidos pela sua atratividade e alta procura por imóveis, como Califórnia, Nova York e Massachusetts, têm médias de *index_housing* **acima da média nacional**. Isso reflete a influência de fatores como grandes centros urbanos, setores económicos desenvolvidos e qualidade de vida, que impulsionam os preços dos imóveis nessas regiões.

Estados com **inflação imobiliária moderada**: Alguns estados, como Texas, North Carolina, Washington e Wisconsin, apresentam médias de *index_housing* próximas à média nacional, o que sugere uma combinação de procura e oferta equilibradas, bem como influências económicas e demográficas específicas desses estados.

Relação entre número de reformados e inflação imobiliária



Relação entre inflação imobiliária e número de reformados

A partir do gráfico, parece não haver uma relação clara entre o índice de inflação imobiliária e o número de pessoas reformadas nos diferentes estados dos EUA. No entanto, é possível efetuar uma análise mais aprofundada calculando a correlação entre essas variáveis.

Calculando a correlação entre o índice de inflação imobiliária e o número de pessoas reformadas para ver se existe alguma relação significativa, através da fórmula da correlação de Pearson, obtem-se um valor de 0,126.

$$\Sigma X = 145.94722222222222 + 117.16555555555556 + \dots + 128.14527777777778$$

$$\Sigma Y = 31715647.805555556 + 3204382.888888889 + \dots + 27472.25$$

$$\Sigma(X * Y) = (145.94722222222222 * 31715647.805555556) + \dots$$

$$\Sigma X^2 = (145.94722222222222)^2 + (117.16555555555556)^2 + \dots$$

$$\Sigma Y^2 = (31715647.805555556)^2 + (3204382.888888889)^2 + \dots$$

$$N = 51 \text{ (número de estados)}$$

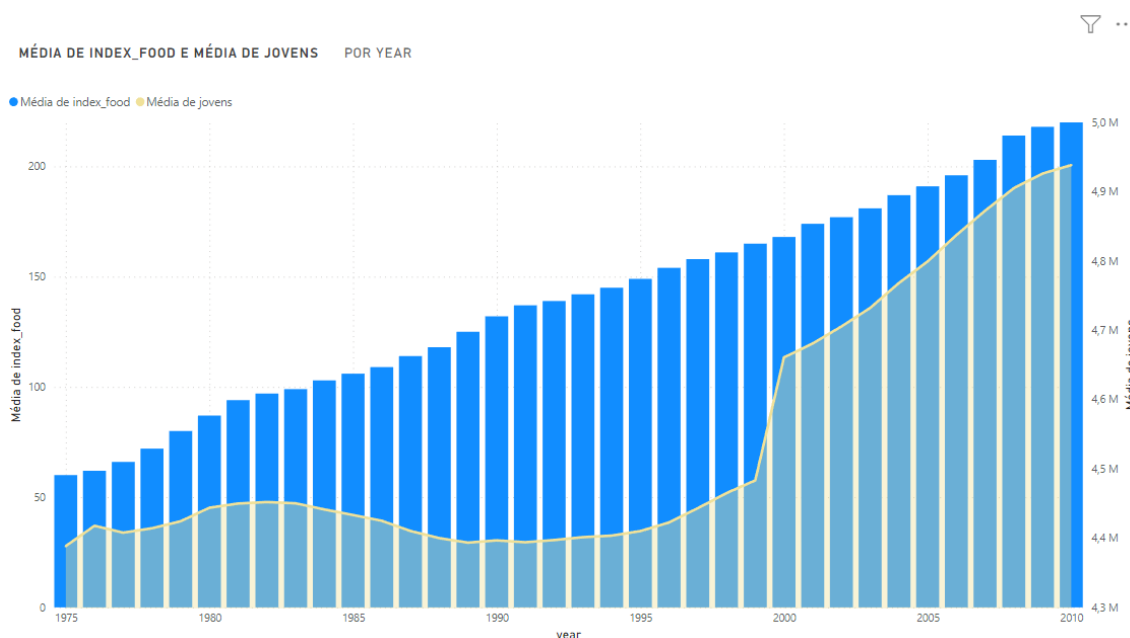
$$\text{correlation} = (N * \Sigma(X * Y) - \Sigma X * \Sigma Y) / \text{sqrt}((N * \Sigma X^2 - (\Sigma X)^2) * (N * \Sigma Y^2 - (\Sigma Y)^2))$$

Esse valor indica uma **correlação muito fraca** entre as duas variáveis.

A correlação fraca sugere que o índice de inflação imobiliária não é um fator determinante para o número de pessoas reformadas nos diferentes estados dos EUA. Outros fatores, como o custo de vida, a demografia, a disponibilidade de serviços e a qualidade de vida, podem ter uma influência mais significativa na decisão das pessoas de se aposentarem em determinados estados.

Portanto, com base nos dados fornecidos, não é possível encontrar uma relação clara entre o índice de inflação imobiliária e o número de pessoas reformadas nos diferentes estados dos EUA.

Relação entre número de jovens e inflação nos alimentos



Ao analisar os dados, observamos que tanto a média de *index_food* (inflação nos alimentos) quanto a média de jovens (número de pessoas entre 0 e 29 anos) tendem a

aumentar ao longo do tempo, indicando um aumento dos preços dos alimentos e um crescimento populacional na faixa etária de 0 a 29 anos.

A média de *index_food* apresenta um aumento mais rápido e consistente ao longo dos anos em comparação com a média de jovens, o que pode sugerir que os preços dos alimentos aumentam a uma taxa mais acelerada do que o crescimento populacional na faixa etária de 0 a 29 anos.

Observou-se algumas flutuações nas médias de *index_food* e jovens de um ano para outro. Por exemplo, a inflação na comida teve um aumento significativo entre 1975 e 1980, mas depois mostrou um crescimento mais moderado. Em relação ao número de pessoas novas, houve um aumento gradual ao longo dos anos, com algumas variações menores.

É possível observar que entre 1983 e 1995 houve um abrandamento no crescimento populacional em termos de população jovem, enquanto a inflação continuou a crescer nesse período, pelo que nesta década pode-se ter observado um impacto significativo da inflação no crescimento da população jovem.

A partir da análise, verifica-se um aumento expressivo no número de pessoas jovens a partir de 2000. É possível observar um aumento considerável no número de pessoas novas a partir de 2000, enquanto a inflação na comida também continua a aumentar. Esse período pode estar relacionado a fatores demográficos, como o aumento da taxa de natalidade.

5. Conclusões e Trabalho Futuro

O trabalho realizado teve como objetivo fazer uma análise da relação entre inflação nos preços das habitações e na alimentação e o crescimento populacional nos Estados Unidos, usando ferramentas de Big Data para fornecer uma visão geral dos principais tópicos e estudos relacionados ao assunto, oferecendo informações relevantes para o leitor.

Relativamente a pontos fracos, o relatório poderia se beneficiar de uma discussão mais aprofundada sobre as limitações e possíveis vieses dos dados utilizados. Por exemplo, é importante considerar outras variáveis que podem influenciar a relação entre inflação imobiliária e crescimento populacional, como políticas governamentais, taxas de juros e condições econômicas gerais.

Além disso, algumas conclusões são baseadas em análises exploratórias simples, como a correlação entre variáveis. Seria interessante realizar análises estatísticas mais robustas para confirmar ou refinar as conclusões apresentadas.

Futuramente, uma possível tarefa futura seria aprofundar a análise considerando outras variáveis relevantes, como taxas de juros, políticas habitacionais e indicadores econômicos, para obter uma compreensão mais completa dos fatores que afetam a relação entre inflação imobiliária e crescimento populacional.

Além disso, o trabalho poderia se beneficiar de uma análise espacial mais detalhada, considerando fatores regionais e locais que podem influenciar a dinâmica do mercado imobiliário.

Uma extensão interessante seria realizar análises de previsão ou modelagem para entender melhor as tendências futuras e as possíveis relações causais entre as variáveis estudadas.

Uma diferença potencial em relação a trabalhos já publicados é a utilização de Big Data na análise da relação entre inflação nos preços das habitações e crescimento populacional nos Estados Unidos. A abordagem de usar conjuntos de dados extensos e técnicas de visualização avançadas pode fornecer insights adicionais e uma compreensão mais abrangente do fenômeno.

Além disso, o trabalho também inclui informações sobre a inflação nos alimentos, adicionando uma dimensão adicional à análise, permitindo uma compreensão mais

abrangente dos fatores que afetam o crescimento populacional e a inflação imobiliária nos Estados Unidos. A inclusão da inflação nos alimentos permite explorar a relação entre o custo de vida, o comportamento de consumo e a dinâmica demográfica.

Ao analisar os dados sobre a inflação nos alimentos, o trabalho identifica tendências e flutuações ao longo do tempo. Essas informações são combinadas com os dados sobre o crescimento populacional, especialmente na faixa etária de 0 a 29 anos, para examinar possíveis correlações e padrões.

Essa abordagem amplia a perspectiva do trabalho, considerando não apenas o setor imobiliário, mas também o impacto da inflação nos alimentos na tomada de decisões sobre moradia e migração populacional. Isso pode fornecer insights adicionais sobre os fatores socioeconômicos e demográficos que influenciam a dinâmica do mercado imobiliário nos Estados Unidos.

Portanto, uma contribuição adicional deste trabalho é a inclusão da inflação nos alimentos como um fator importante a ser considerado na compreensão da relação entre inflação imobiliária, crescimento populacional e tomada de decisões relacionadas à moradia e migração. Isso diferencia o trabalho de estudos anteriores que podem ter se concentrado exclusivamente na relação entre inflação imobiliária e crescimento populacional, sem considerar o impacto da inflação nos alimentos.

Referências

- [1] Heo, Ye Jin. "Population aging and house prices: Who are we calling old?" - Science Direct, 19 October 2021,
<https://www.sciencedirect.com/science/article/pii/S2212828X22000494>
- [2] Hashimoto, Y., Hong, G. H., and Zhang, X. (2020). Demographics and the Housing Market: Japan's Disappearing Cities. *IMF Working Papers* 2020, 200, A001, available from: <https://doi.org/10.5089/9781513557700.001.A001>
- [3] Wang, Hongbo & Rickman, Dan. (2019). Housing Price and Population Growth across China: The Role of Housing Supply. *International Regional Science Review*. 43. 016001761983588. 10.1177/0160017619835885, available from:
https://www.researchgate.net/publication/331685408_Housing_Price_and_Population_Growth_across_China_The_Role_of_Housing_Supply
- [4] Kidane, D., & Woldemichael, A. (2020). Does inflation kill? Exposure to food inflation and child mortality. *Food Policy*, 92, 101838. ISSN 0306-9192. DOI: 10.1016/j.foodpol.2020.101838, available from:
https://www.sciencedirect.com/science/article/pii/S0306919220300221?casa_token=sP2jgNTevqEAAAAA:xq-Ec7UbQNpFT6Ulzcl_LjdLBmqu65pQ_ghs9zMQxoFoicmbH553gVibBN9cR6l5aJdU9BW2Guk