

Coleta e Disponibilização de Dados do IBGE Cidades

Kayque Avelar Oliveira
Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa
Florestal, Brasil
kayque.oliveira@ufv.br

Fabrcio Aguiar Silva
Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa
Florestal, Brasil
fabricio.asilva@ufv.br

RESUMO

O Instituto Brasileiro de Geografia e Estatísticas (IBGE) é o órgão responsável pela manutenção e disponibilização dos dados estatísticos de todo território nacional. No entanto a maneira como os dados são disponibilizados, pouco intuitiva, dificulta a extração para uma análise mais elaborada. Será apresentado nesse artigo um sistema que viabiliza a extração de dados de uma maneira mais clara. Usando conceitos de Data Science e outros recursos computacionais, o sistema desenvolvido neste trabalho é capaz de capturar os dados fornecidos pelo IBGE, formatá-los de maneira mais intuitiva e armazenar em uma base de dados local, com o objetivo de serem consultados para elaboração de gráficos e análises.

Os dados fornecidos pelo IBGE falam muito sobre o Brasil, como o brasileiro vive e como os gestores públicos têm gastado os recursos da nação. Olhando por essa perspectiva, o projeto assume características sociais, ao auxiliar a população no acesso às informações a fim de questionarem a situação em que o país se encontra. Este artigo aborda as estratégias usadas de fácil assimilação, atingindo o objetivo que é o entendimento dos dados disponibilizados pelo IBGE.

Palavras-Chave

Análise de Dados, IBGE, Cidades

1. INTRODUÇÃO

Com a promulgação do Decreto n° 218, em 26 de janeiro de 1938 (1), é criado o Instituto Brasileiro de Geografia e Estatística (IBGE), a partir do Instituto Nacional de Estatística (INE) (2). Com natureza federativa, o IBGE foi idealizado com finalidade de viabilizar e executar o levantamento organizado de todas estatísticas do território nacional. Na atualidade, os dados produzidos pelo IBGE são disponibilizados em seu site (3). Cada cidadão pode fazer uso do sistema oferecido no site do IBGE, a fim de realizar uma análise por conta própria.

Os dados coletados e disponibilizados pelo IBGE possuem relevância significativa para o entendimento de como os bra-

sileiros estão vivendo e como os gestores públicos estão aplicando os recursos. No entanto, as informações são apresentadas de uma maneira confusa, tornando inviável uma análise mais complexa, que traça correlações entre atributos de cada pesquisa, ou entre as cidades, deixando o cidadão limitado. Com isso, o potencial que os dados poderiam oferecer ficam restritos, já que poucas pessoas ou instituições enfrentam as dificuldades de coleta, preparação, organização e análise desses dados.

Para reduzir esse problema, neste trabalho foi desenvolvido um sistema com a finalidade de auxiliar tanto as pessoas que buscam fazer uma análise simples, quanto aquelas que procuram direcionar sua busca para uma perspectiva mais detalhada. Assim, todos os cidadãos se beneficiam do sistema, deixando os dados públicos mais fáceis de serem extraídos para uma análise posterior. Com isso, será mais fácil identificar os principais problemas enfrentados pelas cidades, e verificar possíveis formas de melhorar a vida dos cidadãos.

Este artigo está disposto da seguinte forma: Na Seção 2 serão apresentados os principais conceitos relacionados; na Seção 3 será apresentado o sistema desenvolvido, com a explicação de sua arquitetura e descrição do processo para coleta de dados e da API construída; na Seção 4 serão apresentadas algumas análises e resultados obtidos. Por fim, a Seção 5 apresenta as conclusões e alguns possíveis trabalhos futuros.

2. CONCEITOS RELACIONADOS

O IBGE é o órgão responsável por disponibilizar os dados e estatísticas do território nacional. Com administração federal e sede no Rio de Janeiro, o IBGE é encarregado por realizar censos e organizar os dados conseguidos, a fim de suprir outros órgãos do governo e demais organizações em geral. O interesse pelos dados estatísticos do Brasil tem crescido, não apenas em instituições governamentais, mas também da população. Percebendo isso, em 2011 foi decretado a lei n° 12.527/2011 - Lei de Acesso à Informação (LAI) (4), que define normas e regras para gastos públicos da União, Estados e Municípios. A lei afirma que os dados devem ser disponibilizados por meios eletrônicos (Portais de Transparência). Os Portais de Transparência são desenvolvidos pelas instituições governamentais e em alguns casos demonstram ser ineficientes pela maneira como os dados são disponibilizados, com formatação imprópria com grandes tabelas e até divergência de informações.

No Brasil alguns projetos fazem uso dos dados fornecidos pelo IBGE, temos o Observatório de Oncologia (5) que é uma plataforma online com monitoramento dinâmico de dados abertos, que visa compartilhar informações relevantes da área de oncologia no Brasil. Todos os dados usados pelo Ob-

servatório de Oncologia são de fontes governamentais. As bases usadas são fornecidas pelo Ministério da Saúde, Instituto Nacional do Câncer e também pelo IBGE.

Outro projeto interessante é o Repasse (6), se trata de uma ferramenta para visualização das transferências feitas pelo governo federal as cidades, com a finalidade de analisar o impacto do investimento por meio de uma visualização gráfica. Os usuários desse sistema podem refletir de maneira mais simples e eficiente diversos dados.

Os dados conseguidos pelo IBGE são disponibilizados por meio de uma API de dados agregados que hoje está em sua terceira versão. Essa API atualmente alimenta o SIDRA, (Sistema IBGE de Recuperação Automática) (7), ferramenta que disponibiliza os dados das pesquisas e censos realizados pelo IBGE. A API organiza os censos por pesquisas, e cada pesquisa possui seus indicadores. A figura 1 ilustra um retorno da API. A pesquisa referente ao censo escolar possui indicadores do ensino básico e ensino superior, que armazenam a quantidade de docentes, quantidade de matrículas, quantidade de matrículas em creches, quantidade de matrículas nas escolas públicas e privadas dentre outros.

```

{
  "id": 5902,
  "posicao": "1.1",
  "indicador": "Matrículas",
  "classe": "T",
  "children": [
    {
      "id": 77881,
      "posicao": "1.1.1",
      "indicador": "Ensino infantil",
      "classe": "I",
      "unidade": {
        "id": "matriculas",
        "classe": "N",
        "multiplicador": 1
      }
    },
    {
      "id": 5908,
      "posicao": "1.1.2",
      "indicador": "Ensino fundamental",
      "classe": "I",
      "unidade": {}
    }
  ]
}
  
```

Figura 1: Retorno de requisição a API do IBGE

O sistema desenvolvido nesse artigo faz uso dessa API e constrói uma relação entre as Pesquisas e Indicadores, armazenando os resultados a fim de construir um histórico para consulta dos usuários.

3. SOLUÇÃO

3.1 Arquitetura

A Figura 2 apresenta, de maneira geral, o funcionamento do sistema desenvolvido. Cada elemento ilustrado é responsável por um componente independente do sistema.

3.1.1 API IBGE

Possibilita o acesso aos dados e estatísticas produzidas pelo IBGE. Ao informar a pesquisa, indicador e localidade, a API retorna um conteúdo JSON com as informações referentes àquela consulta. Como é de responsabilidade do IBGE realizar o levantamento de todas estatísticas do território nacional, o retorno oferecido na consulta à API pode trazer informações que não são relevantes para uma determinada análise, tornando o uso confuso para usuários leigos.

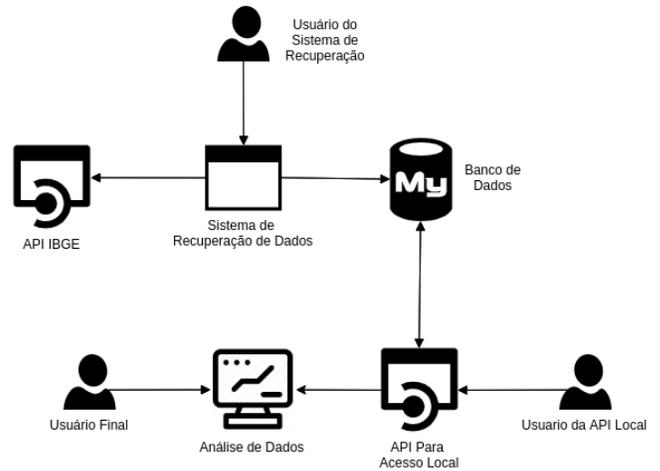


Figura 2: Arquitetura do sistema

3.1.2 Sistema de Recuperação de Dados

Responsável pela coleta e persistência dos dados. Este componente é um elemento fundamental para o sistema como um todo. Caso faça sentido o usuário poderá fazer uso apenas desse componente, construindo sua base personalizada, contendo apenas as pesquisas necessárias para a sua análise. De maneira mais abstrata, podemos subdividir o Sistema de Recuperação de Dados da forma apresentada na Figura 3, essa etapa acontece pelo menos 4 vezes em cada vez que o sistema é ativado, primeiro os dados das cidades são salvos, em seguida formatado e inserido no banco de dados local. Esse processo acontece novamente para os indicadores, pesquisas e por fim histórico..

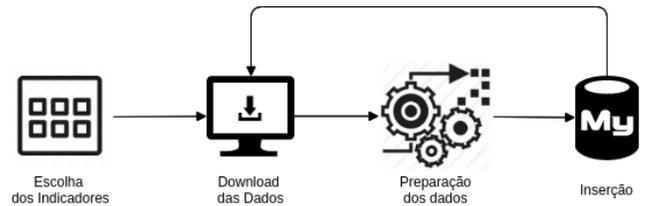


Figura 3: Arquitetura do Sistema de Recuperação de Dados

1. Escolha dos Indicadores: O IBGE atualmente disponibiliza 74 censos de pesquisas, e o conjunto de todas as pesquisas possuem 15001 indicadores. Para realizar uma análise específica, nem sempre é necessário obter todas as informações disponibilizadas. Por exemplo: caso o objetivo do usuário seja fazer um estudo de caso relacionando o Censo Escolar, com Desigualdades Sociais por Cor ou Raça no Brasil, não faria sentido obter informações da Produção Agrícola. Nesta etapa, o usuário irá escolher quais os indicadores serão recuperados, para que o tempo de execução e memória gasta não sejam acima do necessário.
2. Download dos Dados: Nesta etapa, é realizado o down-

load das informações para a máquina local em arquivos JSON, que posteriormente serão inseridos no banco de dados relacional. É aqui onde as consultas à API do IBGE são realizadas, para todas as cidades do Brasil, sempre levando em consideração os indicadores escolhidos. Por ser uma aplicação que faz uso da internet para obter os dados, o sistema fica suscetível a falhas de conexão. Contudo, a aplicação foi desenvolvida contornando esse contratempo, sendo que caso ocorra um problema de conexão, o sistema não segue adiante até que o download seja feito por completo.

3. Preparação dos Dados: A API retorna um JSON com diversas informações adicionais que devem ser ignoradas ou substituídas. Nessa etapa ocorre a leitura dos arquivos salvos no disco local, e em seguida as informações são extraídas dos arquivos JSON e formatadas, para posteriormente serem inseridas no banco de dados relacional.
4. Inserção: Os dados são inseridos no banco de dados relacional, seguindo o modelo que será apresentado a seguir.

3.1.3 Banco de Dados

Após coletados e preparados, os dados são armazenados em um banco de dados relacional para serem consultados posteriormente. A Figura 4 representa como o banco de dados foi modelado. Mesmo sendo um modelo simples, ele é capaz de armazenar todas as pesquisas feitas pelo IBGE.

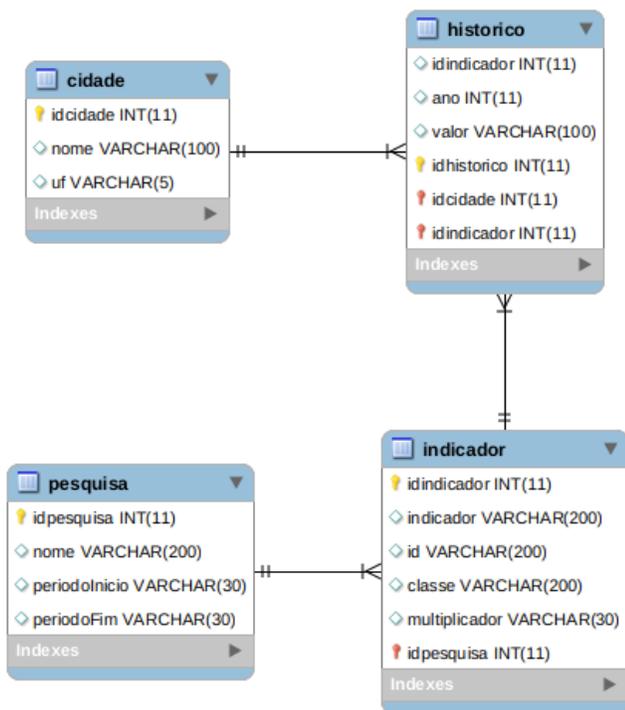


Figura 4: Modelo Relacional do Banco de Dados utilizado.

A seguir cada entidade do banco e seus relacionamentos são descritos:

1. Cidade: Cidades contidas no território nacional que tenham estatísticas levantadas pelo IBGE. Atualmente são 5570 cidades registradas.
2. Pesquisa: Conjunto de pesquisas realizadas pelo IBGE. Atualmente são 74 pesquisas, como por exemplo: Produção Agrícola - Lavoura Temporária, Serviços de saúde e Finanças públicas.
3. Indicadores: Conjunto de indicadores referentes às pesquisas realizadas pelo IBGE. Atualmente são 15001, como por exemplo: Atendimento de emergência, que informa a quantidade de atendimentos de emergencia realizados em uma região, Clínica médica, que informa a quantidade de estabelecimentos em uma localidade, Investimentos que armazena a quantidade de dinheiro investido em uma região.
4. Histórico: Armazena o valor de um determinado indicador em um período de tempo para uma cidade. Com isso é possível analisar como os valores das cidades tem evoluído ao longo do tempo.

3.1.4 API Para Acesso Local

Este componente foi construído com o propósito de facilitar o acesso aos elementos extraídos pelo Sistema de Recuperação de Dados. Caso um determinado grupo de pessoas tenha como objetivo realizar um estudo específico, este componente possibilita que todos tenham acesso aos dados extraídos pelo Sistema de Recuperação de dados. Desta maneira os dados serão salvos uma única vez e a API de acesso local irá intermediar o acesso a informação extraída.

Com este componente o usuário é capaz de buscar todas as cidades, pesquisas e indicadores. O usuário poderá fazer uma consulta de uma pesquisa específica para uma cidade específica, passando apenas como parâmetro, o id da cidade e o id da pesquisa, também é possível buscar uma determinada pesquisa para todas as cidades do Brasil, passando como parâmetro apenas o identificador da pesquisa e por fim também é possível fazer buscas de um ano específico, incluindo o ano nos parâmetros. A API Para Acesso Local, a cada requisição retorna um JSON com o valor correspondente a pesquisa, desta maneira uma aplicação pode consumir o retorno fornecido a fim de realizar análises.

3.1.5 Análise de Dados

Este é o componente encarregado de analisar os dados armazenados e, posteriormente, promover um entendimento do conteúdo extraído. No momento presente, essa etapa faz uso da API Rest de acesso local, para realizar consultas na base com um foco específico de análise (como por exemplo, ranking de cidades pelo IDH), e podem ser disponibilizados de forma gráfica. O principal foco aqui é selecionar parte dos dados já preparados e armazenados e buscar correlações para análise.

3.1.6 Usuário do Sistema de Recuperação

Este usuário está interessado em usufruir apenas do Sistema de Recuperação de Dados, assumindo a responsabilidade de escolher quais indicadores serão usados na busca e o tipo de análise a ser aplicada.

3.1.7 Usuário da API Local

Este usuário está interessado em fazer buscas aos dados extraídos pelo Sistema de Recuperação de Dados e armaze-

nados na base de dados já com as informações preparadas e organizadas, a fim de realizar sua análise individual.

3.1.8 Usuário Final

Este usuário tem acesso às análises produzidas por meio de gráficos e relatórios, fazendo uso indiretamente de todos os componentes do sistema.

4. RESULTADOS

A base construída neste trabalho contém 5.570 cidades, 73 pesquisas e 15.001 indicadores. Para popular o histórico foram selecionados 30 indicadores, o que levou cerca de 11 horas para construir a base para análise. Com o intuito de obter uma visão geral de cada cidade, os dados obtidos trazem informações abrangentes da saúde, educação, economia, território, população dentre outros. Para cada cidade foram realizadas 30 pesquisas totalizando 167.100 pesquisas efetuadas. O objetivo da tabela histórico é guardar as informações das pesquisas para todos os anos em que ocorreram. Por isso, neste caso de uso a tabela histórico possui 1.039.966 registros. Apesar da seleção de apenas 30 indicadores para este projeto, o sistema está preparado para consultar todos os indicadores, o que não foi feito ainda por restrições computacionais.

4.1 Gráficos

A análise feita neste estudo teve como maior objetivo encontrar correlações entre alguns dos indicadores escolhidos. A seguir serão apresentados os gráficos que possuem correlações mais satisfatórias. Nos gráficos, cada ponto representa uma cidade, e os eixos X e Y contemplam os valores dos indicadores analisados.

4.1.1 IDH X Esgotamento Sanitário

Como pode-se ver na Figura 5, existe uma correlação positiva entre o *Esgotamento Sanitário* e o *IDH*. A cidade tende a ter um *IDH* melhor quando o índice de esgotamento é maior.

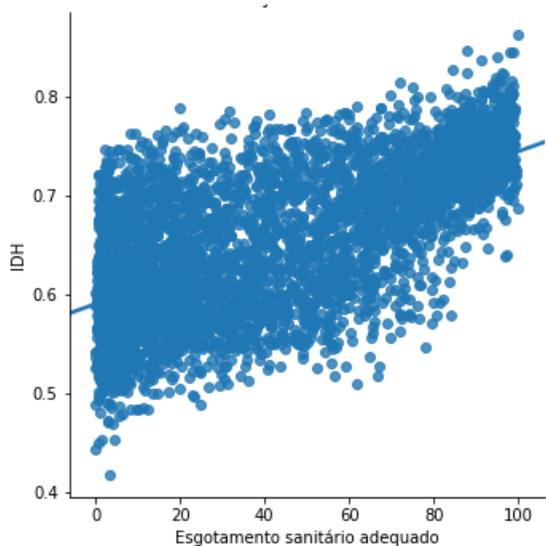


Figura 5: Correlação entre Esgotamento Sanitário e IDH.

4.1.2 População Ocupada x IDH

O indicador *População ocupada* representa a porcentagem da população empregada em um município. Este indicador é calculado da seguinte maneira: (pessoal ocupado no município/população total do município) x 100. Como podemos ver na Figura 6, existe uma certa correlação positiva entre a *população ocupada* e o *IDH*. Isso se justifica pois a população desocupada e desempregada pode passar por situações de miséria e conseqüentemente reduzir o IDH.

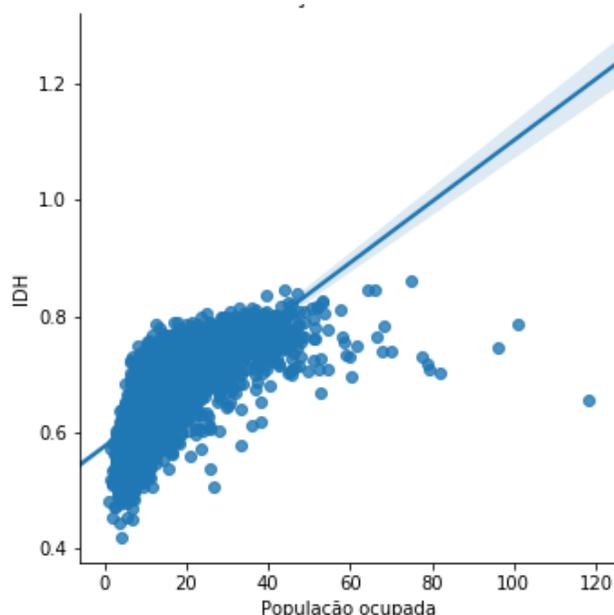


Figura 6: Correlação entre População ocupada e IDH

4.1.3 Esgotamento Sanitário x Salário médio mensal

A Figura 7 apresenta a correlação encontrada entre o *Esgotamento Sanitário* e o *Salário médio mensal*. A correlação apresentada é uma correlação positiva, ou seja quando o salário médio mensal da cidade é maior o esgotamento sanitário tende a ser maior. O esgotamento sanitário é uma necessidade básica, cidades que possuem o salário médio alto tendem a ser cidades mais desenvolvidas, que tem um sistema de esgoto apropriado.

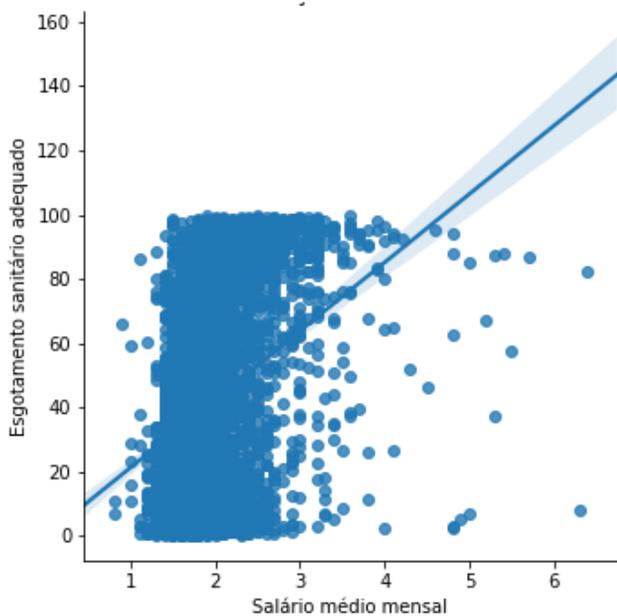


Figura 7: Correlação entre Esgotamento Sanitário e Salário médio mensal

4.1.4 Salário Médio Mensal x IDH

A Figura 8 apresenta a correlação encontrada entre o *salário médio mensal* e o *IDH*, como pode-se ver, existe uma correlação positiva entre o *salário médio mensal* e o *IDH*. A cidade tende a ter um *IDH* maior quando o *salário médio mensal* é maior. Isso é coerente supondo que a qualidade de vida é melhor quando o salário é maior.

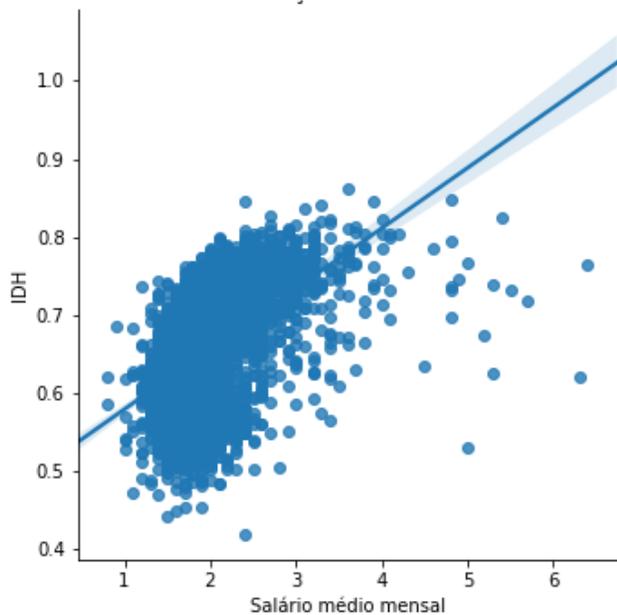


Figura 8: Correlação entre Salário Médio Mensal e IDH

4.1.5 Salário Médio Mensal x População Ocupada

A Figura 9 apresenta a correlação encontrada entre o *Salário médio mensal* e a *população ocupada*. Como mostrado existe uma correlação positiva entre os indicadores, isso quer dizer que as cidades que possuem um salário médio melhor tendem a ter mais pessoas trabalhando.

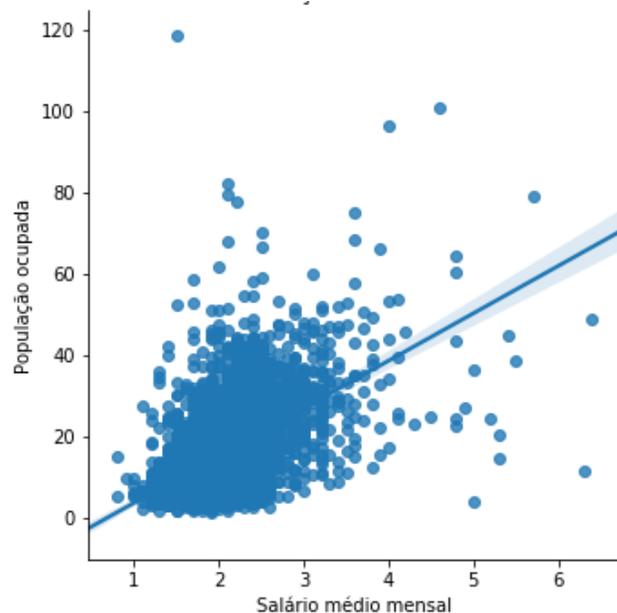


Figura 9: Correlação entre Salário Médio Mensal e População Ocupada

4.1.6 Mortalidade Infantil x Esgotamento Sanitário

A Figura 10 apresenta a correlação encontrada entre a *taxa de mortalidade infantil* e o *esgotamento sanitário*.

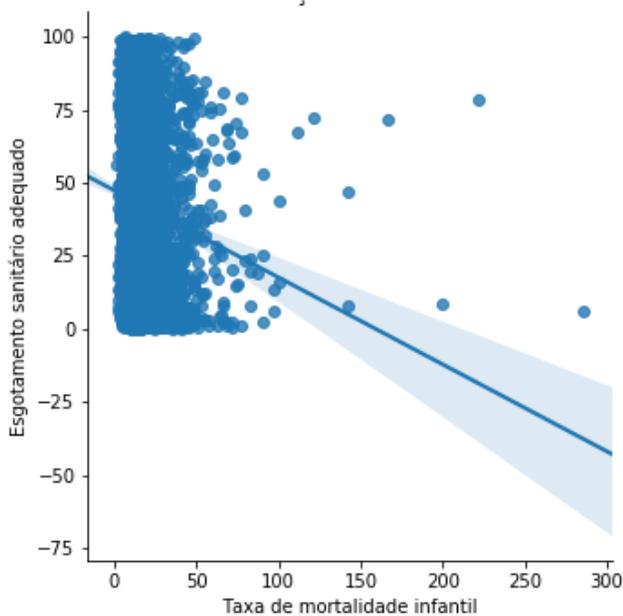


Figura 10: Correlação entre Mortalidade Infantil e Esgotamento Sanitário

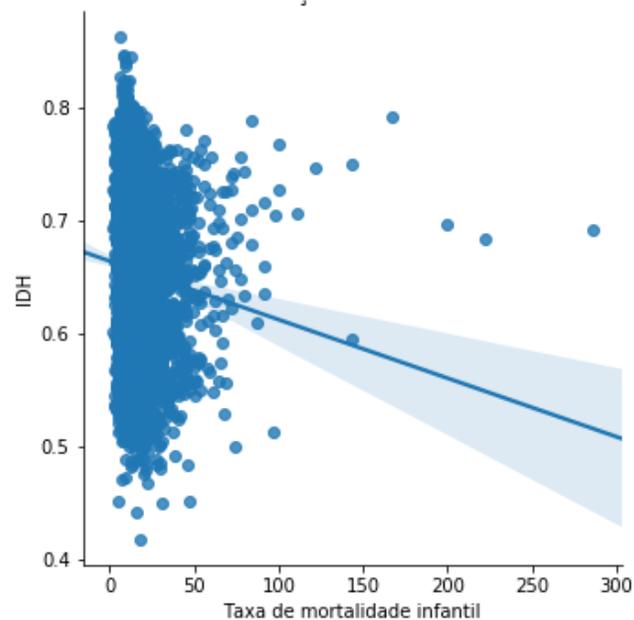


Figura 11: Correlação entre Mortalidade Infantil e IDH

4.1.7 Mortalidade Infantil x IDH

A Figura 11 apresenta a correlação encontrada entre a *taxa de mortalidade infantil* e o *IDH*. O Índice de Desenvolvimento Humano é calculado com base em três aspectos essenciais da população: Renda, educação e saúde, tendo isso em vista quando um desses aspectos apresentam melhorias o IDH tende a apresentar uma melhoria. A principal causa da mortalidade infantil predomina as causas evitáveis (8), sendo as causas evitáveis aquelas que podem ser parcialmente ou totalmente prevenidas por intervenção do serviço de saúde disponível. Sabendo disso é esperado que quanto maior a taxa de mortalidade infantil, menor é o IDH.

4.1.8 Salário médio mensal x Mortalidade infantil

A Figura 12 apresenta a correlação encontrada entre o *salário médio mensal* e a *mortalidade infantil*. É esperado que quanto menor o *salário médio mensal* do município maior a taxa de *mortalidade infantil*.

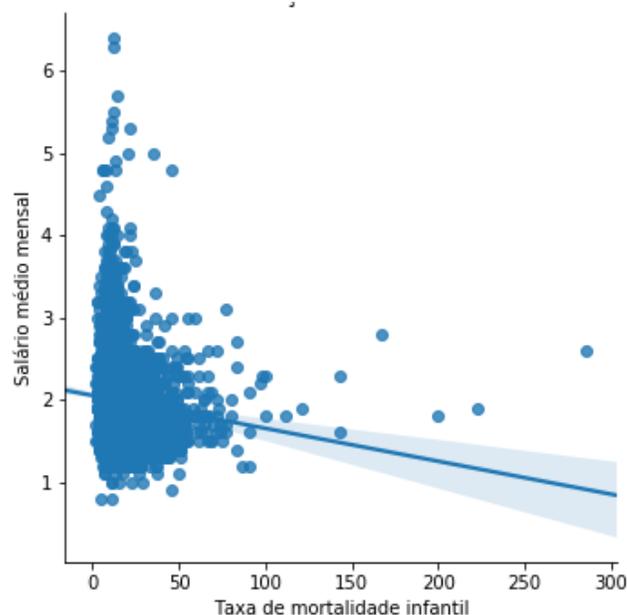


Figura 12: Correlação entre salário médio mensal e Mortalidade infantil

4.1.9 População Ocupada X Densidade Demográfica

A Figura 13 apresenta a correlação encontrada entre a *população ocupada* e a *densidade demográfica*. Como podemos observar, foi encontrada uma correlação positiva entre a população que está desempenhando algum tipo de trabalho e a densidade demográfica dos municípios. Ou seja, municípios maiores tendem a ter mais pessoas empregadas.

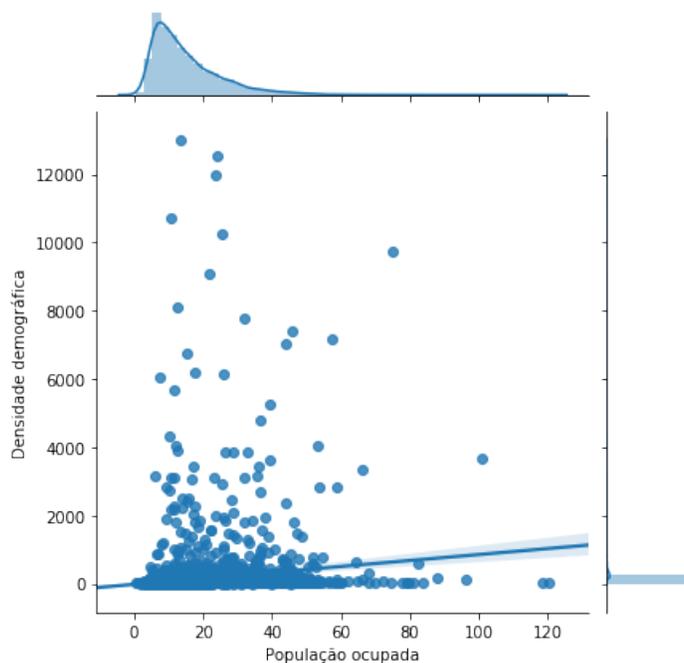


Figura 13: Correlação entre população ocupada e densidade demográfica

Uma outra maneira de realizar uma análise é criando grupos de cidades que possuem características semelhantes. Para isso, foi utilizado o k-means (9). O agrupamento por meio do k-means é um dos mais usados e mais rápidos métodos de clustering, que agrupa os dados em K grupos, sendo esse K pré-definido. Dizer quantos clusters queremos separar nossos dados é sempre um desafio, especialmente quando estamos trabalhando com agrupamento e aprendizado não-supervisionado. Para solucionar esse desafio, foi aplicado o *Elbow Method* (10), que visa aplicar o k-means em várias quantidades diferentes de clusters. Ao aumentar a quantidade de clusters, a diferença entre eles serão muito pequenas, e o objetivo é encontrar o meio termo em que os dados que formam cada agrupamento sejam o mais parecido possível e que os agrupamentos formados sejam o mais diferente um do outro. Na Figura 14, podemos ver que o melhor K seria 5, já que a partir daí os grupos passam a se tornarem menores e com maior semelhança.

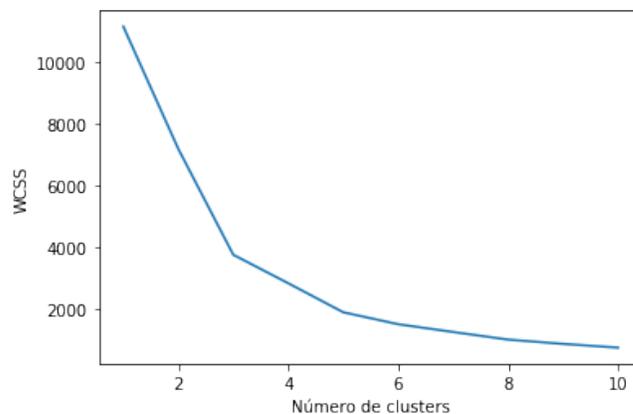


Figura 14: Definindo K com Elbow method

Para realizar o agrupamento das cidades, foram utilizados apenas os indicadores da *população ocupada* e *densidade demográfica*. Para este estudo de caso apresentar os dados em duas dimensões seria mais didático, mas nada impede do usuário buscar agrupamento com n indicadores. A Figura 15 mostra o k-means aplicado com K igual a 5.

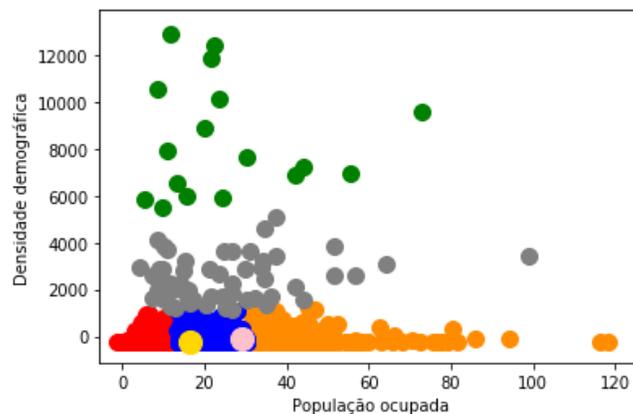


Figura 15: Clusters formados com K-means

Podemos observar os cinco grupos, cada grupo com uma cor diferente. É esperado que cada grupo tenha semelhança entre si. Por exemplo: o grupo vermelho e azul, são aquelas cidades que possuem uma baixa densidade demográfica, e pouca população ocupada. Já o grupo verde são aquelas cidades que provavelmente possuem várias pessoas desempregadas, tendo em vista que são cidades com grande densidade demográfica e pouca população ocupada. Em contrapartida, o grupo amarelo é formado por cidades com boa parte da população ocupada e baixa densidade demográfica. A nível de curiosidade foi plotado o ponto amarelo que representa a cidade de Florestal-MG e o ponto rosa que representa a cidade de Pará de Minas-MG.

4.2 Discussões

Analisando os gráficos apresentados podemos observar que vários indicadores possuem correlação com o IDH. Obser-

vando a Figura 5 percebemos que investir em saneamento básico pode ser fundamental para a melhora do Índice de Desenvolvimento Humano, as cidades que possuem o *esgotamento sanitário* adequado tende a ter o *IDH* mais alto. Além disso investir em saneamento básico promove queda no índice de mortalidade infantil que por consequência também aumenta o *IDH*.

Com o objetivo de aumentar a qualidade de vida da população e o *IDH*, os gestores devem investir em empregos que geram uma renda maior, quanto maior o salário médio de uma cidade menor é a taxa de mortalidade infantil e maior é o Índice de Desenvolvimento Humano. Além disso é possível encontrar perfis de cidades, e quais devem ter prioridade no investimento.

5. CONCLUSÕES E TRABALHOS FUTUROS

Há uma necessidade de análise dos dados estatísticos produzidos pelo Instituto Brasileiro de Geografia e Estatística. Por se tratar de grandes volumes de dados históricos, um usuário leigo poderá encontrar dificuldades em fazer uso dos sistemas apresentados pelo IBGE. Com o trabalho existente, é possível facilitar o processo de recuperação e análise dos dados, proporcionando ao usuário o acesso a informação sem dificuldade e de maneira intuitiva.

Com os recursos gráficos e correlações de atributos, é possível obter-se um entendimento de grande parte dos dados. É importante que o usuário tenha em mente a pergunta que ele deseja responder, dessa maneira o uso do sistema será voltado para seu objetivo, desde as escolhas dos indicadores. Por exemplo: caso o usuário queira classificar cidades com base na sua produção agrícola e o PIB, os indicadores e pesquisas escolhidos deverão ser voltados para este fim. Dessa maneira a coleta de dados será mais rápida e a análise mais coerente com o objetivo proposto. Usando este sistema o usuário terá em sua máquina apenas os dados necessários para análises em seu contexto, ignorando todas outras informações que não são interessantes. Isso economiza espaço no disco e processamento de dados indesejados. Além de facilmente conseguir atualizar os dados caso novas pesquisas sejam realizadas pelo IBGE.

Como trabalhos futuros, pretende-se desenvolver um modelo automatizado para a análise dos dados, relacionando os dados obtidos pelo IBGE com informações de outras bases a fim de encontrar padrões com o uso de técnicas de aprendizado de máquina. Também é esperado que a base coletada seja integrada a bases de dados coletadas de portais de transparência municipal, para que seja possível correlacionar as características das cidades com os detalhes dos gastos e investimentos públicos nas mesmas. Também seria interessante a construção de um ambiente colaborativo em que usuários disponibilizem scripts de análises feitas, dessa maneira em um só lugar poderiam ser encontrados vários tipos de análises diferentes, cada uma criando correlação com itens e bases diferentes.

6. REFERÊNCIAS

- [1]BRASIL. *DECRETO-LEI Nº 218*. Disponível em: <<https://www2.camara.leg.br/legin/fed/declei/1930-1939/decreto-lei-218-26-janeiro-1938-350934-publicacaooriginal-1-pe.html>>. Acessado em 24/11/2019.
- [2]SÍNTESE HISTORICA DO IBGE. Disponível em: <<https://memoria.ibge.gov.br/images/memoria/linha-do-tempo/LinhaDoTempoSemImagem.pdf>>.

Acessado em 24/11/2019.

[3]SITE IBGE CIDADES. <https://cidades.ibge.gov.br/>.

[4]BRASIL. *LEI Nº 12.527*. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm>. Acessado em 24/11/2019.

[5]OBSERVATÓRIO DE ONCOLOGIA. Disponível em: <<https://observatoriodeoncologia.com.br/>>. Acessado em 7/12/2019.

[6]REPASSE. Disponível em: <<http://repass.ufabc.edu.br/>>. Acessado em 7/12/2019.

[7]API IBGE. Disponível em: <<https://servicodados.ibge.gov.br/api/docs/agregados?versoa=3>>. Acessado em 24/11/2019.

[8]MORTALIDADE INFANTIL PRINCIPAIS CAUSAS. Disponível em: <<http://www.imb.go.gov.br/files/docs/publicacoes/informes-tecnicos/2018/12-mortes-evitaveis-na-infancia-201810.pdf>>. Acessado em 24/11/2019.

[9]KMEANS. Disponível em: <https://docs.aws.amazon.com/pt_br/sagemaker/latest/dg/algo-kmeans-tech-notes.html>. Acessado em 24/11/2019.

[10]ELBOW METHOD. Disponível em: <<https://www.scikit-yb.org/en/latest/api/cluster/elbow.html>>. Acessado em 24/11/2019.