PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA
Pós-graduação *Lato Sensu* em Ciência de Dados e Big Data

DIEISON GULARTE BASTOS

ANÁLISE DE DADOS DO BOLSA FAMÍLIA PARA DETERMINAR O TAMANHO DA FILA DE ESPERA DO PROGRAMA.

Belo Horizonte
2020

DIEISON GULARTE BASTOS

ANÁLISE DE DADOS DO BOLSA FAMÍLIA PARA DETERMINAR O TAMANHO DA FILA DE ESPERA DO PROGRAMA.

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Ciência de Dados e Big Data como requisito parcial à obtenção do título de especialista.

Belo Horizonte
2020

SUMÁRIO

1. Introdução

1.1. Contextualização

No contexto atual de políticas públicas da sociedade brasileira identifica-se a existência de diversos programas assistenciais que procuram minimizar as diferenças sociais e buscam contribuir para o desenvolvimento de uma sociedade mais igualitária. É o caso do Programa Bolsa Família, que foi criado pelo governo federal direcionado às famílias em situação de pobreza e de extrema pobreza em todo o País, de modo que consigam superar a situação de vulnerabilidade socioeconômica e pobreza.

O principal objetivo do Bolsa Família é combater a fome, a pobreza e promover a segurança alimentar e nutricional, por meio da transferência de renda às famílias cadastradas que se encontram dentro dos critérios de inclusão; além disso, através das condicionalidades, reforçar o acesso aos direitos básicos e aos serviços de saúde, segurança alimentar e assistência social.

O Programa Bolsa Família vive, sob o governo Jair Bolsonaro, o que pode ser um de seus momentos com maior fila de espera para ingresso da história, impedindo milhares de pessoas de obter acesso ao benefício por todo o Brasil.

1.2. O problema proposto

Para entender um pouco mais sobre o programa, vamos responder algumas perguntas e utilizar a técnica dos 5WS que nos dará uma melhor visão do problema e da solução.

Quem pode participar do programa?

Para se candidatar ao programa é necessário que a família esteja inscrita no Cadastro Único para programas sociais do Governo Federal, o cadastramento é um pré-requisito, mas não implica na entrada imediata das famílias no programa, nem no recebimento do benefício.

A população alvo é constituída por famílias que vivem em situação de pobreza e de extrema pobreza. Foi utilizado um limite de renda para definir esses dois patamares. Assim, podem fazer parte do programa: todas as famílias extremamente pobres, com renda por pessoa de até R$ 89,00 mensais; e famílias pobres, com renda por pessoa entre R$ 89,01 e R$ 178,00 mensais, desde que tenham em sua composição gestantes e crianças ou adolescentes entre 0 e 17 anos.

O que é a Fila de Espera?

A fila de espera do Bolsa Família é formada por famílias que estão registradas no Cadastro Único, se encaixam nos critérios do programa, mas que de fato não estão recebendo a transferência de renda. O governo federal não revela os números exatos dessa fila, sem disponibilizar os números absolutos mês a mês de famílias que preenchem os requisitos, mas ainda sem a bolsa.

Vamos a técnica dos 5WS:

(Why?) Por que esse problema é importante?

A fila de espera do programa pode indicar um crescimento da pobreza e extrema pobreza no país, bem como redução de novas concessões. Neste cenário e de extrema importância que a sociedade acompanhe esse indicador, com o intuito de reivindicar que o governo tome medidas para diminuir esse índice, melhorando a condição de vida dessas famílias que vivem com tão pouco, provendo o mínimo para uma vida mais digna.

(Who?) De quem são os dados analisados?

Os dados analisados são fornecidos pelo Ministério do Desenvolvimento Social (MDS) atual Ministério da Cidadania, é o órgão do governo federal responsável por realizar a articulação entre gestores federais, estaduais, municipais e a sociedade civil para manter um sistema de proteção social no país e promover políticas de assistência social, e é responsável pelo Programa Bolsa Família.

(What?): Quais os objetivos com essa análise?

Neste trabalho foi analisado os dados disponíveis do Bolsa Família e Cadastro Único, o objetivo geral do trabalho é determinar o tamanho da fila de espera do programa, composta por famílias que preenchem os requisitos, mas não recebem de fato a transferência de renda, ou seja, estão aguardando novas concessões do governo.

No portal de dados abertos, não foi possível encontrar informações atuais da composição das famílias pobres, inscritas no Cadastro Único, assim não podemos determinar se essas famílias preenchem todos os requisitos para essa faixa. Por esse motivo para definir a fila de espera do programa foi levado em consideração apenas as famílias extremamente pobres.

Como objetivos específicos tivemos:

- Analisar a distribuição de famílias beneficiárias no Brasil;
- Analisar a distribuição de famílias em situação de extrema pobreza no Brasil;
- Analisar novas concessões no programa.

(Where?): A análise levou em consideração todos os estados do Brasil.

(When?): Qual o período está sendo analisado?

Os dados analisados do Bolsa Família e Cadastro Único compreende o ano de 2019. Para o conjunto de dados de novas concessões, a análise engloba toda a série histórica.

2. Coleta de Dados

DataSet: Bolsa Família

Descrição: Esse conjunto de dados informa a quantidade de famílias beneficiárias e o valor repassado através da folha de pagamento do Bolsa Família, assim como o código IBGE do município e ano/mês de referência.

Fonte: Portal Brasileiro De Dados Abertos / Ministério do Desenvolvimento Social (MDS).

Formato: CSV

Data Coleta: 24/03/2020

Link: http://www.dados.gov.br/dataset/bolsa-familia-misocial

Nome da coluna/campo	Descrição	Tipo
Ibge	Código IBGE do município.	int64
anomes	Ano e mês de referência.	int64
qtd_familias_beneficiarias_bolsa_familia	Quantidade de famílias beneficiárias do programa bolsa família.	int64

valor_repassado_bolsa_familia	Valor repassado pelo programa bolsa família.	float64

DataSet: Cadastro Único

Descrição: Esse conjunto de dados apresenta a quantidade de famílias em situação de pobreza e extrema pobreza, o total de famílias e pessoas cadastradas no Cadastro Único, por faixas de renda per capita, assim como o código IBGE do município e o ano/mês de referência.

Fonte: Portal Brasileiro De Dados Abertos / Ministério do Desenvolvimento Social (MDS).

Formato: CSV

Data Coleta: 24/03/2020

Link: http://www.dados.gov.br/dataset/cadastro-unico-familias-pessoas-cadastradas-por-faixas-de-renda

Nome da coluna/campo	Descrição	Tipo
Ibge	Código IBGE do município.	int64
anomes	Ano/mês de referência.	int64
cadunico_tot_fam	Total de famílias cadastradas.	int64
cadunico_tot_pes	Total de pessoas cadastradas.	int64
cadunico_tot_fam_rpc_ate_meio_sm	Total de famílias com renda per capita até meio salário mínimo.	int64
cadunico_tot_pes_rpc_ate_meio_sm	Total de pessoas com renda per capita até meio salário mínimo.	int64
cadunico_tot_fam_pob	Total de famílias em situação de pobreza.	int64

cadunico_tot_pes_pob	Total de pessoas em situação de pobreza.	int64
cadunico_tot_fam_ext_pob	Total de famílias em situação de extrema pobreza.	int64
cadunico_tot_pes_ext_pob	Total de pessoas em situação de extrema pobreza.	int64
cadunico_tot_fam_pob_e_ext_pob	Total de famílias em situação de pobreza e de extrema pobreza.	int64
cadunico_tot_pes_pob_e_ext_pob	Total de pessoas em situação de pobreza e de extrema pobreza.	int64

DataSet: Concessões.

Descrição: Esse conjunto de dados apresenta a quantidade de novas concessões para o programa Bolsa Família, assim como o ano/mês de referência.

Fonte: CECAD – Consulta, Seleção e Extração de Informações do CadÚnico

Formato: CSV

Data Coleta: 28/03/2020

Link:https://aplicacoes.mds.gov.br/sagi/cecad20/agregado/resumovariavelCecad.php?id=352

Nome da coluna/campo	Descrição	Tipo
Periodo	Ano/mês de referência.	int64
Quantidade de novas concessões	Total de novas concessões.	float64

DataSet: Municípios latitude/longitude.

Descrição: Dataset utilizado para relacionamento com os datasets Bolsa Família e Cadastro Único, a fim de identificar o estado e a região, a partir do código IBGE do município.

Fonte: Disciplina - Data Discovery, OLAP e Visualização De Dados.

Formato: xlsx

Data Coleta: 11/03/2020

Nome da coluna/campo	Descrição	Tipo
Cod UF	Código IBGE do Estado.	int64
UF	Descrição do Estado.	String
Sgl UF	Sigla do Estado.	String
Cod IBGE	Código IBGE do Município.	int64
Nome Município	Descrição do Município.	String
cod_latitude	Código latitude do Município.	float64
cod_longitude	Código longitude do Município.	float64
Sgl Região	Sigla da região.	String
Região	Descrição da região.	String

DataSet: Estados

Descrição: Dataset utilizado para relacionamento com o dataset Concessões, a fim de identificar o estado e a região, a partir do código IBGE do estado.

Fonte: Kaggle

Formato: CSV

Data Coleta: 14/03/2020

Link: https://www.kaggle.com/joaopauloribsantos/dts-estados-municipios-ibge#estados.csv

Nome da coluna/campo	Descrição	Tipo
id_regiao	Identificador da região.	int64
nome_regiao	Descrição região.	string
sigla_regiao	Sigla da região.	string
id_estado	Código IBGE do Estado.	int64
sigla_estado	Sigla do Estado.	string
nome_estado	Descrição do Estado.	String

3. Processamento/Tratamento de Dados

Ferramenta

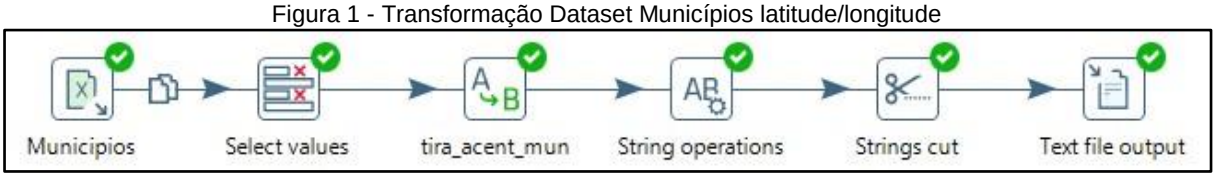
Para a escolha da ferramenta foi levado em consideração, o domínio da aplicação, a facilidade na manipulação da mesma e ser *Open Source*. Seguindo esses critérios a ferramenta escolhida para o processo de tratamento dos dados foi a plataforma *Pentaho*, conhecida como *Pentaho Open BI* é um software de código aberto, desenvolvida em Java e utilizado na área de *Business Intelligence*. A solução proporciona soluções para o processo de ETL, desenvolvimento de relatórios, análises OLAP, etc.

Especificamente foi utilizado o *Pentaho Data Integration* (PDI), conhecido como *Kettle*, é uma aplicação que auxilia no processo de integração e transformação de dados no processo de ETL. O *Kettle* oferece o *Spoon* uma aplicação que trabalha com componentes (*steps*) para projetar o fluxo de dados, desde as entradas de dados (*input*), passando pela transformação e gerando a saída (*output*).

DataSet: Municípios latitude/longitude.

A Figura 1 ilustra a transformação com as *steps* usadas para manipular esse dataset.

Figura 1 - Transformação Dataset Municípios latitude/longitude



Fonte: Adaptado de Spoon

A primeira *Step* a ser utilizada foi o *Microsoft Excel input*, com ele podemos selecionar um ou mais arquivos, escolher o *Sheet*, as *Fields* desejadas e carregar as informações.

Arquivo: municipios_latitude_longitude.xlsx.

Sheet: Dados Geo.

A próxima *Step* utilizada foi a *Select values*, com ela podemos selecionar as *fields* que desejamos trabalhar, renomear e até alterar o *Type* do dado, em alguns casos. A Figura 2 ilustra os *Fields* selecionados e a novas descrições atribuídas.

Figura 2 - *Fields* selecionados e renomeados

Fields :

#	Fieldname	Rename to	Length
1	UF	ESTADO	
2	Sgl UF	UF	
3	Cod IBGE	IBGE	
4	Nome Município	MUNICIPIO	
5	cod_latitude	LATITUDE	
6	cod_longitude	LONGITUDE	
7	Região	REGIAO	

Fonte: Adaptado de Spoon

A *field* IBGE teve seu *Type* alterado para String, para realizar as operações que podem ser feitas nesse formato.

Na sequência a *Step Replace in string*, que foi utilizada para substituir letras com acentuação para letras sem acentuação, a fim de evitar problemas de codificação. *Fields* afetados pela *step*: MUNICIPIO e ESTADO. A Figura 3 ilustra a configuração utilizada nessa *step*.

Figura 3 - Step Replace in string

#	In stream field	Out stream field	use RegEx	Search	Replace with
1	MUNICIPIO		S	[áàâãÁÀÂÃ]	a
2	MUNICIPIO		S	[éèêÉÈÊ]	e
3	MUNICIPIO		S	[íìîÍÌÎ]	i
4	MUNICIPIO		S	[óòôõÓÒÔÕ]	o
5	MUNICIPIO		S	[úùûÚÙÛ]	u
6	MUNICIPIO		S	[çÇ]	c
7	ESTADO		S	[áàâãÁÀÂÃ]	a
8	ESTADO		S	[éèêÉÈÊ]	e
9	ESTADO		S	[íìîÍÌÎ]	i
10	ESTADO		S	[óòôõÓÒÔÕ]	o
11	ESTADO		S	[úùûÚÙÛ]	u

Fonte: Adaptado de Spoon

Na sequência a *Step String operations*, que foi utilizada para retirar possíveis espaços nas *fields* tipo string e deixá-las como maiúscula. A Figura 4 ilustra essa *step*.

Figura 4 - *String operations*

#	In stream field	Out stream field	Trim type	Lower/Upper
1	IBGE		both	none
2	ESTADO		both	upper
3	MUNICIPIO		both	upper
4	REGIAO		both	upper

Fonte: Adaptado de Spoon

A próxima *Step Strings cut*, foi utilizada para selecionar parte da *field* IBGE, pelo fato de conter 7 dígitos neste dataset e 6 dígitos nos datasets (Bolsa Família e Cadastro Único), sem o dígito verificador. A figura 5 ilustra a *step*.

Figura 5 - Strings cut

#	In stream field	Out stream field	Cut from	Cut to
1	IBGE		0	6

Fonte: Adaptado de Spoon

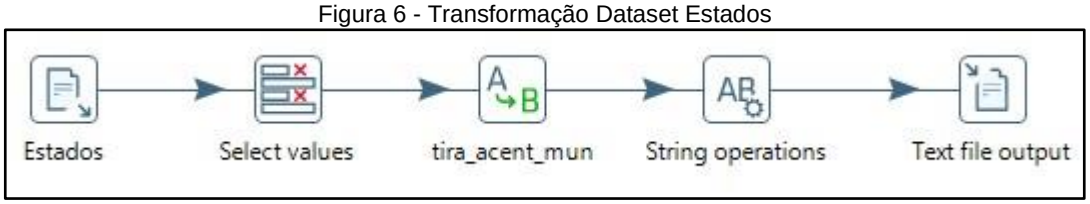
Para finalizar essa transformação a *Step Text file output*, foi utilizada para gerar o arquivo de saída no formato CSV com delimitador ";", incluindo as *fields* selecionadas e tratadas no processo. A opção pelo formato CSV foi para padronizar os datasets e por preferência de trabalhar com esse formato.

Arquivo gerado: municipios_lat_long_tratado.csv.

DataSet: Estados

Para este dataset o tratamento foi muito semelhante ao anterior, a Figura 6 ilustra a transformação com as *steps* utilizadas.

Figura 6 - Transformação Dataset Estados



Fonte: Adaptado de Spoon

CSV file input (Estados): importar o arquivo já no formato CSV com delimitador ","; **Arquivo:** estados.csv.

Select values: selecionar os campos: id_estado, nome_estado, sigla_estado e nome_regiao.

Replace in string (tira_acent_mun): substituir letras com acentuação por letras sem acentuação, campo afetado: nome_estado.

String operations: retirar possíveis espaços nas *fields* tipo string e deixá-las como maiúscula, campos afetados: nome_estado, sigla_estado e nome_regiao.

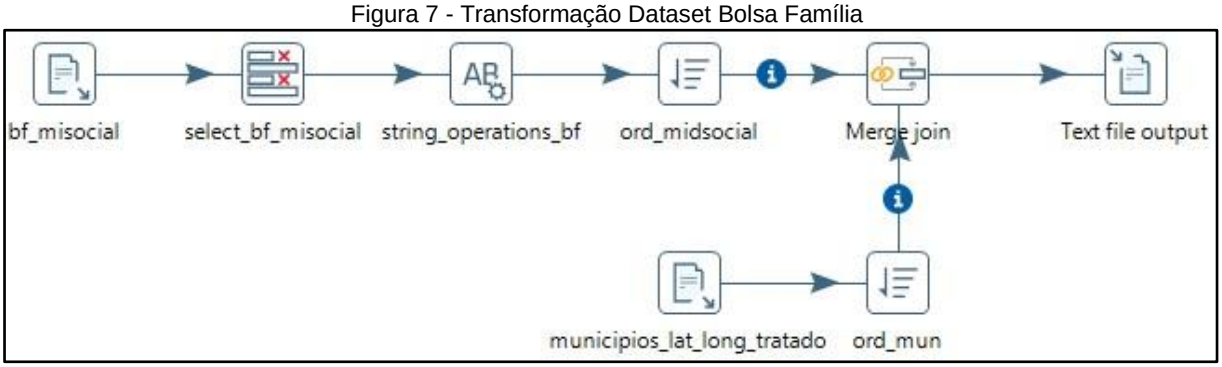
Text file output: gerar o arquivo de saída no formato CSV, com as *fields* selecionadas e tratadas no processo usando o delimitador ";".

Arquivo gerado: estados_tratado.csv.

DataSet: Bolsa Família

A Figura 7 ilustra a transformação com as *steps* utilizados para este dataset.

Figura 7 - Transformação Dataset Bolsa Família



Fonte: Adaptado de Spoon

Text file input (bf_misocial): importar o arquivo: misocial2019.csv com delimitador “,”. Foi utilizado o ano de 2019 contendo os meses de janeiro a dezembro. A Figura 8 ilustra os *Fields* presentes no arquivo.

Figura 8 - *Fields* do Dataset

#	Name	Type
1	ibge	Integer
2	anomes	Integer
3	qtd_familias_beneficiarias_bolsa_familia	Integer
4	valor_repassado_bolsa_familia	Number

Fonte: Adaptado de Spoon

Select Values (select_bf_misocial): utilizado para alterar o *type* das *fields* ibge e anomes para string, o formato da *field* valor_repassado_bolsa_familia para #,##0.00 adaptando o decimal para “,” e o grupo para “.” seguindo o formato do Brasil. A Figura 9 representa as alterações.

Figura 9 - Alterações Step Select values

#	Fieldname	Rename to	Type	Length	Precision	Binary to Normal?	Format
1	ibge		String			N	
2	anomes		String			N	
3	valor_repassado_bolsa_familia		Number	15	0	N	#,##0.00

Fonte: Adaptado de Spoon

String operations (string_operations_bf): retirar possíveis espaços nas *fields* tipo string, campos afetados ibge e anomes.

CSV file input (municipios_lat_long_tratado): importar o dataset municipios_lat_long_tratado.csv, para realizar a junção com este dataset, delimitador “;”.

Sort rows (ord_midsocial): para realizar a junção dos datasets é necessário ordena-los pela chave que será usada na junção, no caso a *field* ibge. A *step Sort rows* vai realizar essa tarefa. No dataset Bolsa Família as *fields* ordenadas foram ibge e anomes. Já o dataset municipios_lat_long_tratado foi ordenado pelo campo ibge, ambos em ordem crescente. A Figura 10 ilustra a *step*.

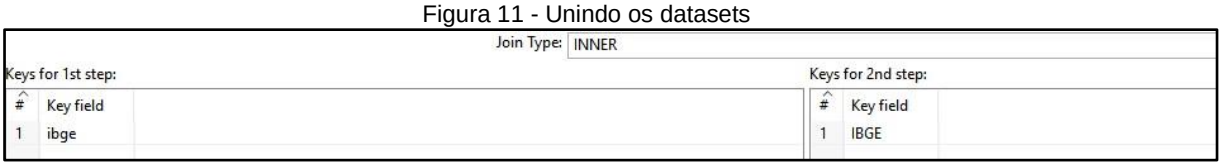
Figura 10 - Ordenar campos

#	Fieldname	Ascending
1	ibge	S
2	anomes	S

Fonte: Adaptado de Spoon

Merge join: essa *step* vai realizar a associação dos datasets pela chave ibge. O tipo de junção utilizado foi *INNER* que retorna os dados quando houver pelo menos uma correspondência em ambos os datasets. A Figura 11 ilustra a step.

Figura 11 - Unindo os datasets



Fonte: Adaptado de Spoon

Text file output: gerar o arquivo de saída no formato CSV, com as *fields* selecionadas e tratadas no processo, usando o delimitador “;”. **Arquivo gerado:** bolsa_familia_midsocial.csv. A Figura 12 ilustra essa *step*.

Figura 12 - Campos gerados no arquivo de saída

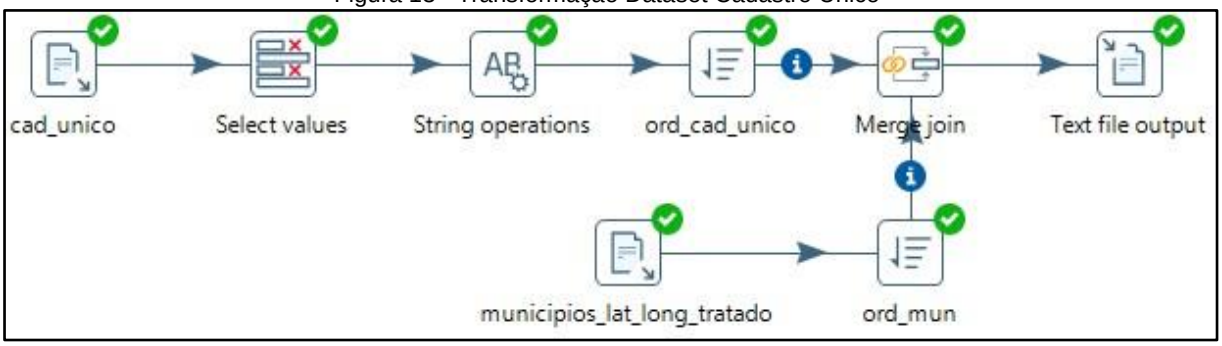
#	Name	Type	Format
1	ESTADO	String	
2	UF	String	
3	REGIAO	String	
4	MUNICIPIO	String	
5	LATITUDE	String	
6	LONGITUDE	String	
7	ibge	String	#
8	anomes	String	#
9	qtd_familias_beneficiarias_bolsa_familia	Integer	#
10	valor_repassado_bolsa_familia	Number	#,##0.00

Fonte: Adaptado de Spoon

DataSet: Cadastro Único

Para este dataset o tratamento é semelhante ao anterior, a Figura 13 ilustra a transformação com as *steps* utilizadas.

Figura 13 - Transformação Dataset Cadastro Único



Fonte: Adaptado de Spoon

Text file input (cad_unico): importar o arquivo: cadunico2019.csv com delimitador “,”. Vamos utilizar o ano de 2019 que contém os meses de janeiro a dezembro.

Select Values: utilizado para alterar o *type* das *fields:* ibge e anomes para string. A Figura 14 ilustra os *fileds* selecionados.

Figura 14 - Campos selecionados

#	Fieldname
1	ibge
2	anomes
3	cadunico_tot_fam_pob
4	cadunico_tot_pes_pob
5	cadunico_tot_fam_ext_pob
6	cadunico_tot_pes_ext_pob
7	cadunico_tot_fam_pob_e_ext_pob
8	cadunico_tot_pes_pob_e_ext_pob

Fonte: Adaptado de Spoon

String operations: retirar possíveis espaços nas *fields* tipo string, campos afetados: ibge e anomes.

CSV file input (municipios_lat_long_tratado): importar o dataset municipios_lat_long_tratado.csv, para realizar a junção com este dataset, delimitador “;”.

Sort rows (ord_cad_unico): no dataset Cadastro Único as *fields* ordenadas foram: ibge e anomes. Já o dataset municipios_lat_long_tratado foi ordenado pela *field* ibge, ambos em ordem crescente.

Merge join: essa step vai realizar a associação dos datasets pela chave ibge. O tipo de junção utilizado foi *INNER*.

Text file output: gerar o arquivo de saída no formato CSV, com as *fields* selecionadas e tratadas no processo, usando o delimitador ";". **Arquivo gerado:** cad_unico.csv. A Figura 15 ilustra essa *step*.

Figura 15 - Campos gerados no arquivo de saída

#	Name	Type
1	ESTADO	String
2	UF	String
3	REGIAO	String
4	MUNICIPIO	String
5	ibge	String
6	LATITUDE	String
7	LONGITUDE	String
8	anomes	String
9	cadunico_tot_fam_pob	Integer
10	cadunico_tot_pes_pob	Integer
11	cadunico_tot_fam_ext_pob	Integer
12	cadunico_tot_pes_ext_pob	Integer
13	cadunico_tot_fam_pob_e_ext_pob	Integer
14	cadunico_tot_pes_pob_e_ext_pob	Integer

Fonte: Adaptado de Spoon

DataSet: Concessões.

Para este dataset os dados foram disponibilizados com duas colunas "Período e Quantidade de novas concessões", no site é necessário selecionar um estado por vez e realizar a consulta. Após a consulta os dados ficam disponíveis na aba Série Histórica de 01/2014 a 12/2019, com a opção de baixar em formato CSV. Ao baixar e abrir o arquivo, identifiquei o primeiro problema, A Figura 16 ilustra.

Figura 16 - Dados em formato Html

```
<table id="cria_tabela_historico8396" class="table table-striped  table-hover"
style="width:80%">      <thead><tr><th>Período</th><th>Quantidade de novas
concessões</th><th></th></tr></thead><tbody><tr><td>02/2020</td><td>163</td><td></t
d></tr><tr><td>01/2020</td><td>11.293</td><td></td></tr><tr><td>12/2019</td><td>286
</td><td></td></tr><tr><td>11/2019</td><td>332</td><td></td></tr><tr><td>10/2019</t
d><td>277</td><td></td></tr><tr><td>09/2019</td><td>229</td><td></td></tr><tr><td>0
8/2019</td><td>367</td><td></td></tr><tr><td>07/2019</td><td>155</td><td></td></tr>
<tr><td>06/2019</td><td>138</td><td></td></tr><tr><td>05/2019</td><td>11.238</td><t
d></td></tr><tr><td>04/2019</td><td>9.600</td><td></td></tr><tr><td>03/2019</td><td
>8.844</td><td></td></tr><tr><td>02/2019</td><td>6.774</td><td></td></tr><tr><td>01
/2019</td><td>9.030</td><td></td></tr><tr><td>12/2018</td><td>23.845</td><td></td><
/tr><tr><td>11/2018</td><td>12.278</td><td></td></tr><tr><td>10/2018</td><td>9.747<
/td><td></td></tr><tr><td>09/2018</td><td>11.075</td><td></td></tr><tr><td>08/2018<
/td><td>9.022</td><td></td></tr><tr><td>07/2018</td><td>10.876</td><td></td></tr><t
r><td>06/2018</td><td>12.347</td><td></td></tr><tr><td>05/2018</td><td>10.603</td><
td></td></tr><tr><td>04/2018</td><td>11.238</td><td></td></tr><tr><td>03/2018</td><
td>7.797</td><td></td></tr><tr><td>02/2018</td><td>8.751</td><td></td></tr><tr><td>
01/2018</td><td>9.528</td><td></td></tr><tr><td>12/2017</td><td>8.057</td><td></td>
```

Fonte: Adaptado de Jupyter Notebook

Os dados estavam em formato de tabela Html, possivelmente algum problema na geração do CSV do site. A segunda questão levantada é que não está disponível nenhuma informação do estado pesquisado. Para resolver essas duas questões foi utilizado um pouco de *Web Scraping* com Python. A Figura 17 ilustra a primeira etapa.

Figura 17 - 1° Etapa *Web Scraping*

```
In [1]: import pandas as pd
        #biblioteca usada para consultar uma URL
        import urllib.request
        #funções BeautifulSoup para analisar os dados retornados do site
        from bs4 import BeautifulSoup

In [3]: #Consulte o site e retorne o html para a variável page
        page = urllib.request.urlopen(url)

In [4]: #armazene-o no formato BeautifulSoup
        soup = BeautifulSoup(page, 'html5lib')

In [5]: #Para identificar a tabela correta usaremos o atributo class.
        table = soup.find('table', attrs={'class': 'table table-striped table-hover'})
```

Fonte: Adaptado de Jupyter Notebook

A url utilizada na primeira etapa, foi a seguinte: https://aplicacoes.mds.gov.br/sagi/cecad20/agregado/resumovariavelCecad.php?uf_ibge=11&nome_estado=&p_ibge=&nome_municipio=Selecione+um+munic%C3%ADpio&id=352#

A Figura 18 ilustra a segunda etapa.

Figura 18 - 2° Etapa *Web Scraping*

```
In [7]: #Criar duas listas, uma para o periodo e outra para a quantidade
        #de novas concessoes
        A=[]
        B=[]

In [8]: #Aqui, precisamos iterar através de cada linha (tr) e, em seguida,
        #atribuir cada elemento de tr a uma variável e anexá-la a uma lista
        #para acessar o valor de cada elemento, usaremos o método find(text=True) para cada elemento
        #Na nossa tabela temos 3 celulas, porem so as duas primeiras nos interessa.
        for row in table.findAll("tr"):
            cells = row.findAll('td')
            if len(cells) == 3:
                A.append(cells[0].find(text=True))
                B.append(cells[1].find(text=True))

In [9]: colunas = ['uf_ibge', 'Periodo', 'Quantidade de novas concessoes']
        df = pd.DataFrame(columns=colunas)

In [11]: #A cada requisicao ao site informar o codigo ibge do estado
         #com isso teremos a informacao de cada estado no dataframe
         uf_ibge = 11
         df['uf_ibge']=uf_ibge
         df['Periodo']=A
         df['Quantidade de novas concessoes']=B
```

Fonte: Adaptado de Jupyter Notebook

Na sequência, basta exportar o DataFrame gerado em formato CSV com delimitador “;”, repetir o mesmo processo para todos os estados do Brasil, informando o código de cada estado no parâmetro uf_ibge na url utilizada. A Figura 19 ilustra o resultado.

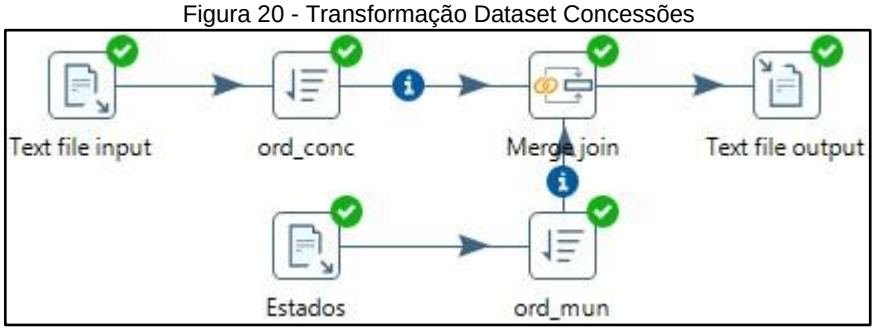
Figura 19 - Resultado *Web Scraping*

	uf_ibge	Periodo	Quantidade de novas concessoes
0	53	02/2020	25
1	53	01/2020	41
2	53	12/2019	43
3	53	11/2019	56
4	53	10/2019	60

Fonte: Adaptado de Jupyter Notebook

Por fim, vamos juntar o dataset resultante do *Web Scraping* com o dataset Estados, assim podemos complementar a análise com dados do estado e região. A Figura 20 ilustra a transformação com as *steps* utilizadas.

Figura 20 - Transformação Dataset Concessões



Fonte: Adaptado de Spoon

Text file input: importar todos os 27 arquivos gerados no *Web Scraping*, um para cada estado, no formato CSV com delimitador “;”. Ao fim da transformação será gerado apenas um arquivo. Cada arquivo foi nomeado com o respectivo código do IBGE.

CSV file input (Estados): importar o dataset: estados_tratado.csv para realizar a junção com este dataset, delimitador “;”.

Sort rows (ord_conc): neste dataset a *field* ordenada foi uf_ibge. No dataset estados_tratado o campo ordenado foi o id_estado, ambos em ordem crescente.

Merge join: essa *step* vai realizar a associação dos datasets pelo código ibge do estado, uf_ibge com id_estado. O tipo de junção utilizado foi *INNER*.

Text file output: gerar o arquivo de saída no formato CSV, usando o delimitador “;”.**Arquivo gerado:** concessoes.csv. A Figura 21 ilustra essa *step*.

Figura 21 - Campos gerados no arquivo de saída

#	Name	Type	Format
1	uf_ibge	Integer	#
2	sigla_estado	String	
3	nome_estado	String	
4	nome_regiao	String	
5	Periodo	String	
6	Quantidade_de_novas_concessoes	Number	#,##0.###

Fonte: Adaptado de Spoon

4. Análise e Exploração dos Dados

Para realizar a análise e exploração dos Dataframes foi utilizado o Pandas Profiling. O pandas profiling é uma ferramenta que gera um relatório html com informações estatísticas e as principais características do dataframe em apenas uma

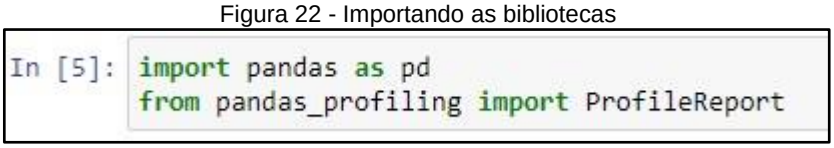
linha de código. Algumas das principais informações que o pandas profiling nos fornece:

- Qual o tamanho do dataset (MB, GB);
- Quantidade de linhas duplicadas;
- Quantidade de linhas do dataframe;
- O tipo de dado de cada coluna;
- Quais colunas contém missi values.

DataSet: Bolsa Família

Vamos analisar o dataset Bolsa Família com os dados brutos, ou seja, sem nenhum tratamento realizado. A Figura 22 ilustra as bibliotecas necessárias.

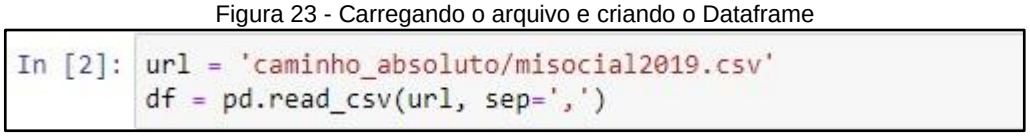
Figura 22 - Importando as bibliotecas



Fonte: Adaptado de Jupyter Notebook

O próximo passo e carregar o arquivo e criar o dataframe para ser utilizado no pandas profiling. A Figura 23 ilustra essa ação.

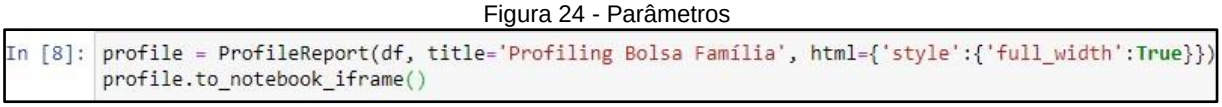
Figura 23 - Carregando o arquivo e criando o Dataframe



Fonte: Adaptado de Jupyter Notebook

Na sequência é criado uma variável com as configurações necessárias para gerar o relatório. Informar o dataframe: df, um título e o estilo do relatório, com isso já é possível gerá-lo. A Figura 24 ilustra a configuração utilizada.

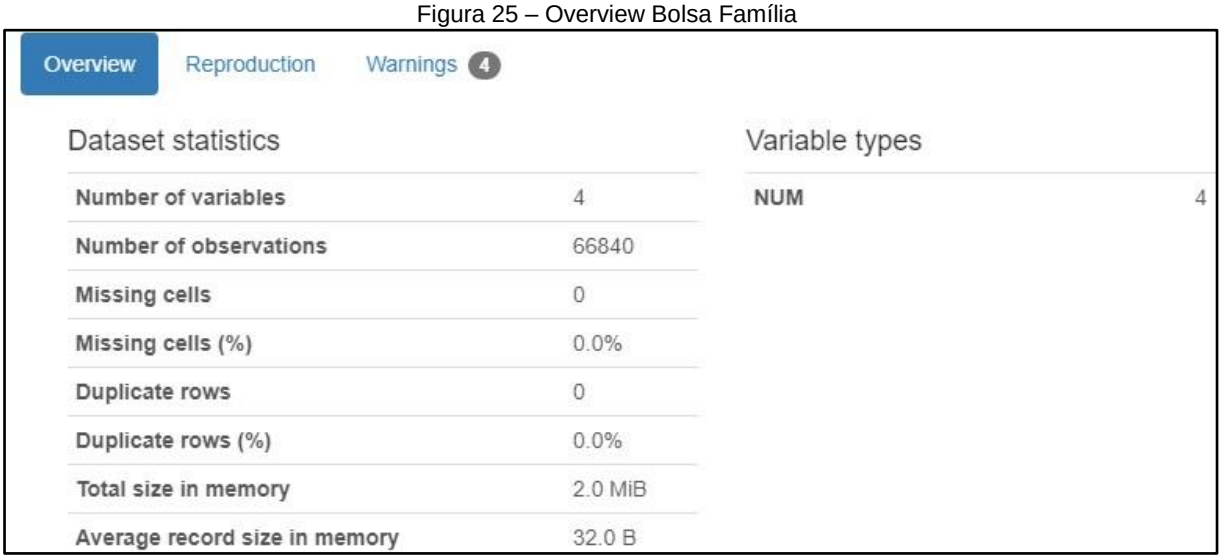
Figura 24 - Parâmetros



Fonte: Adaptado de Jupyter Notebook

A Figura 25 demonstra a principal seção do relatório, podemos ver que o dataset tem 4 variáveis, todas do tipo numérico, com 66.840 mil observações. Não temos valores ausentes nem linhas duplicadas.

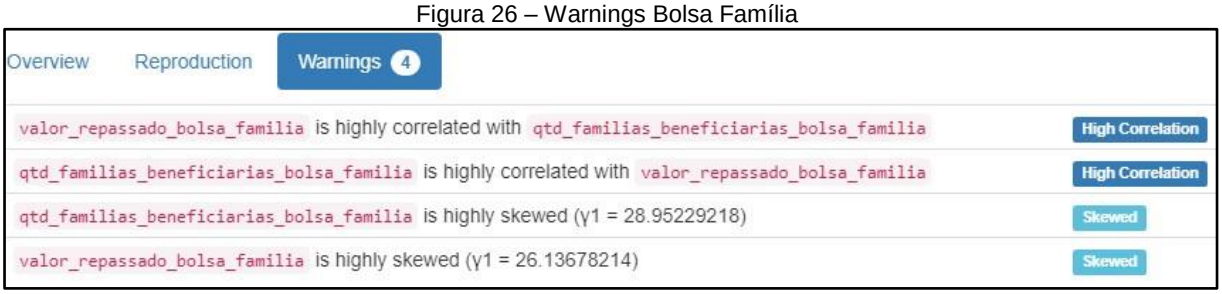
Figura 25 – Overview Bolsa Família



Fonte: Adaptado de Jupyter Notebook

A Figura 26 apresenta a seção de avisos, é possível notar que a variável qtd_familias_beneficiarias_bolsa_familia é altamente correlacionada com a variável valor_repassado_bolsa_familia. Ou seja, se a quantidade de famílias beneficiárias aumentar, o valor a ser repassado provavelmente subirá e vice-versa.

Figura 26 – Warnings Bolsa Família

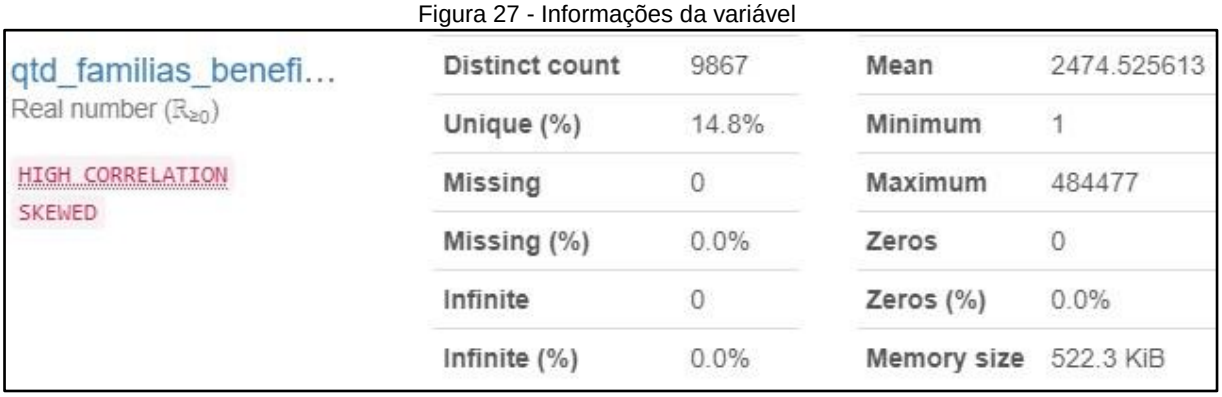


Fonte: Adaptado de Jupyter Notebook

Na Figura 26 temos o *Skewness* com valor positivo indicando uma distribuição altamente inclinada. Skewness é uma medida da assimetria. A curva inclinada negativamente tem uma longa cauda esquerda e vice-versa.

Levando em consideração toda a base de dados, temos uma média de 2474,52 mil famílias beneficiárias. O valor mínimo de 1 indica que em algum momento uma cidade teve apenas uma família beneficiária, já o valor máximo de 484.477 mil, possivelmente uma grande cidade ou capital. A Figura 27 ilustra as principais informações da variável (qtd_familias_beneficiarias_bolsa_familia).

Figura 27 - Informações da variável



qtd_familias_benefi…
Real number ($\mathbb{R}_{\geq 0}$)

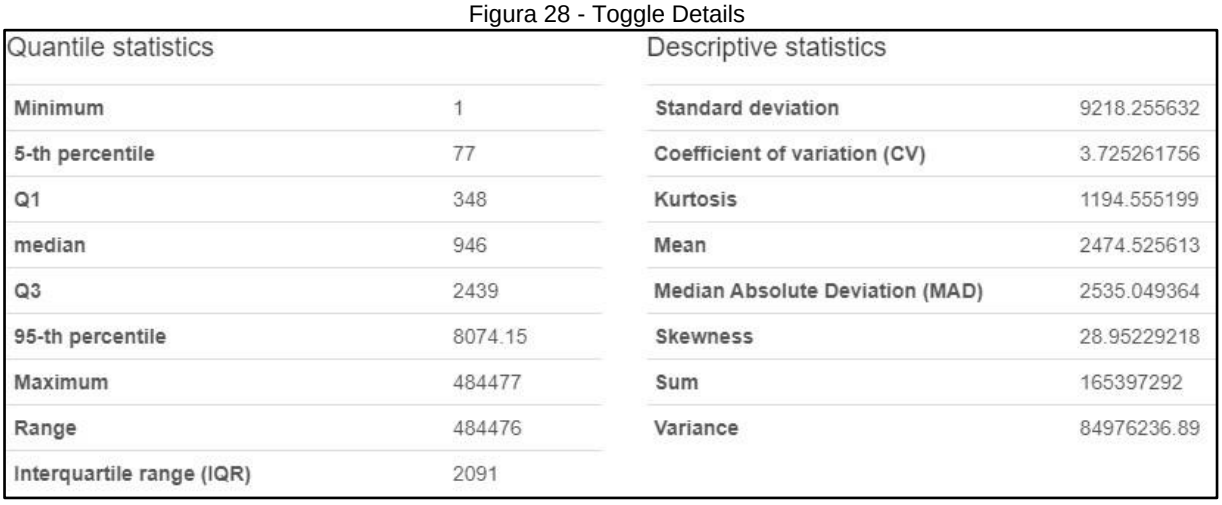
HIGH CORRELATION
SKEWED

Distinct count	9867	Mean	2474.525613
Unique (%)	14.8%	Minimum	1
Missing	0	Maximum	484477
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	522.3 KiB

Fonte: Adaptado de Jupyter Notebook

Em *Toggle details* temos um quadro com a estatística descritiva, como o desvio padrão de 9218,25 mil, indicando uma dispersão dos dados. A soma foi de 165.397.292 milhões de famílias beneficiárias, este não é o total em um único mês, e sim a soma de todos os meses do ano de 2019. A Figura 28 ilustra as informações.

Figura 28 - Toggle Details



Quantile statistics

Minimum	1
5-th percentile	77
Q1	348
median	946
Q3	2439
95-th percentile	8074.15
Maximum	484477
Range	484476
Interquartile range (IQR)	2091

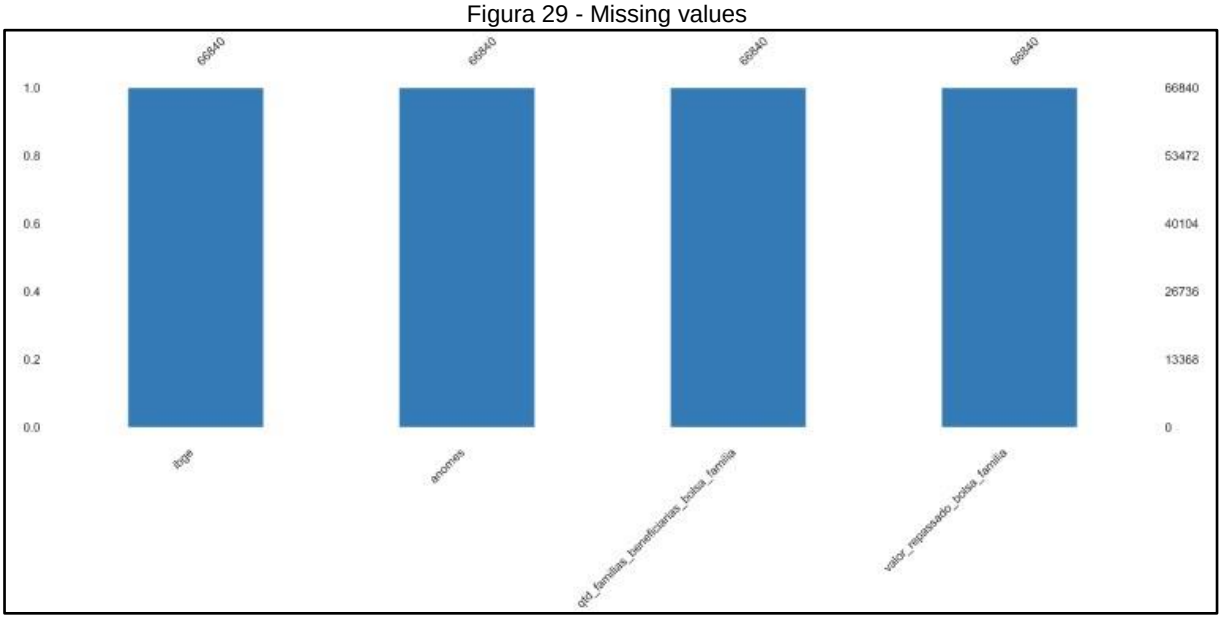
Descriptive statistics

Standard deviation	9218.255632
Coefficient of variation (CV)	3.725261756
Kurtosis	1194.555199
Mean	2474.525613
Median Absolute Deviation (MAD)	2535.049364
Skewness	28.95229218
Sum	165397292
Variance	84976236.89

Fonte: Adaptado de Jupyter Notebook

Na seção *missing values*, temos a quantidade de valores faltantes para cada uma das colunas. Como não temos valores faltantes nesse conjunto de dados, todas as colunas estão preenchidas, como mostra a Figura 29.

Figura 29 - Missing values



Fonte: Adaptado de Jupyter Notebook

A Figura 30 ilustra as 10 primeiras linhas do conjunto de dados.

Figura 30 - First rows

ibge	anomes	qtd_familias_beneficiarias_bolsa_familia	valor_repassado_bolsa_familia
110001	201901	1460	229913.0
110002	201901	3824	569381.0
110003	201901	173	25001.0
110004	201901	2608	396779.0
110005	201901	663	101141.0
110006	201901	362	48975.0
110007	201901	414	60428.0
110008	201901	1637	301156.0
110009	201901	1349	211191.0
110010	201901	2990	544379.0

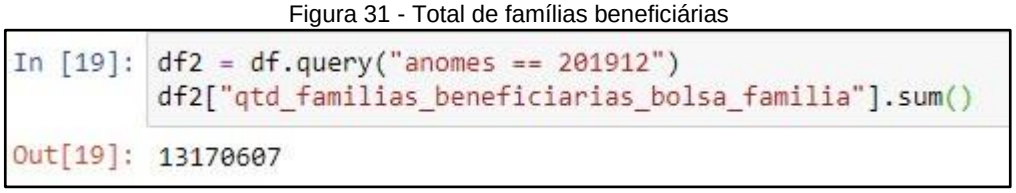
Fonte: Adaptado de Jupyter Notebook

Vamos fazer algumas perguntas para os nossos dados e verificar alguns *Insights*.

Qual o total de famílias beneficiárias no País?

Analisando o mês de dezembro de 2019, o programa fechou o ano com 13.170.607 milhões de famílias beneficiárias. A Figura 31 ilustra o resultado.

Figura 31 - Total de famílias beneficiárias



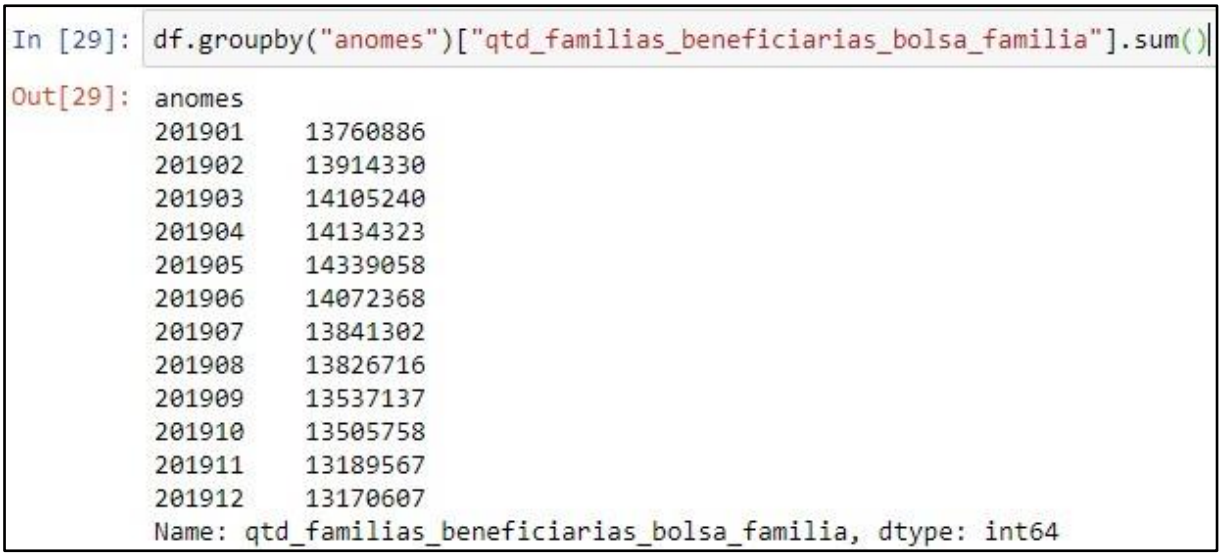
Fonte: Adaptado de Jupyter Notebook

O total de famílias beneficiárias cresceu ou reduziu ao longo do ano?

Na Figura 32 podemos observar que o número de famílias beneficiárias teve um crescimento de janeiro até maio de 2019. Em maio de 2019 o governo anunciava um recorde de 14,3 milhões de famílias contempladas com o auxílio, afirmando que a fila de espera do programa estava zerada.

Esse crescimento pode estar associado a melhorias no mecanismo de acesso ao programa e ao crescimento da extrema pobreza e pobreza no país, reflexo da crise econômica de anos anteriores.

Figura 32 - Total de famílias beneficiárias ao longo de 2019

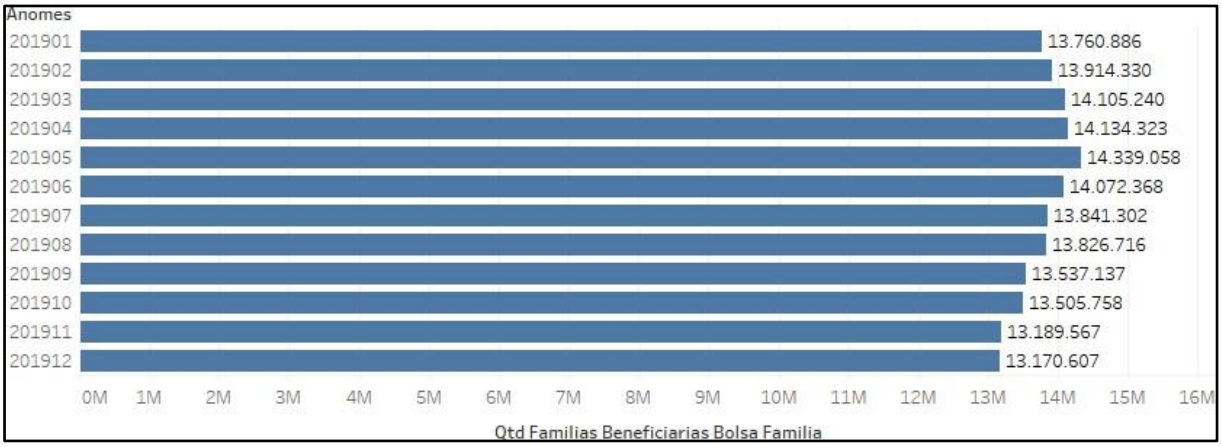


Fonte: Adaptado de Jupyter Notebook

O que aconteceu nos meses de junho a dezembro? Queda, o total de famílias reduziu mês a mês, fechando o ano com **590.279 mil famílias a menos** comparado a janeiro de 2019.

Na Figura 33 vamos visualizar essa mesma análise de forma gráfica.

Figura 33 - Total de famílias beneficiárias de forma gráfica



Fonte: Adaptado de Tableau

Ficou claro que o governo fez uma limpa no programa, mas é a fila de espera, continua zerada? Bom, essa informação não é disponibilizada nos dados abertos. Mas se o Cadastro Único é utilizado para determinar as famílias que ingressam no programa, vamos analisar e buscar a resposta desta pergunta.

DataSet: Cadastro Único

Para gerar o relatório é necessário carregar as bibliotecas, o arquivo e criar o dataframe, os mesmos passos do dataset anterior.

Na Figura 34 podemos observar que o dataset tem 12 variáveis todas do tipo numérico, com 66.840 mil observações. Não temos valores ausentes nem linhas duplicadas.

Figura 34 – Overview Cadastro Único

Dataset statistics

Number of variables	12
Number of observations	66840
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	6.1 MiB
Average record size in memory	96.0 B

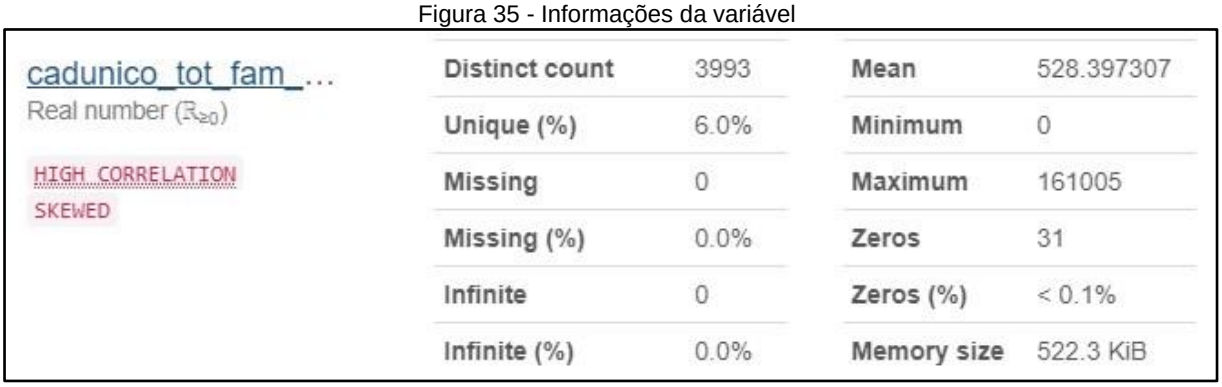
Variable types

NUM	12

Fonte: Adaptado de Jupyter Notebook

Na variável cadunico_tot_fam_pob podemos notar a presença de 31 zeros, que não significa falta de valores, possivelmente cidades com nenhuma família cadastrada na faixa de pobreza em um determinado período. Levando em consideração toda a base de dados, temos uma média de 528,39 famílias na faixa de pobreza. A Figura 35 ilustra as principais informações.

Figura 35 - Informações da variável



cadunico_tot_fam_...
Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
SKEWED

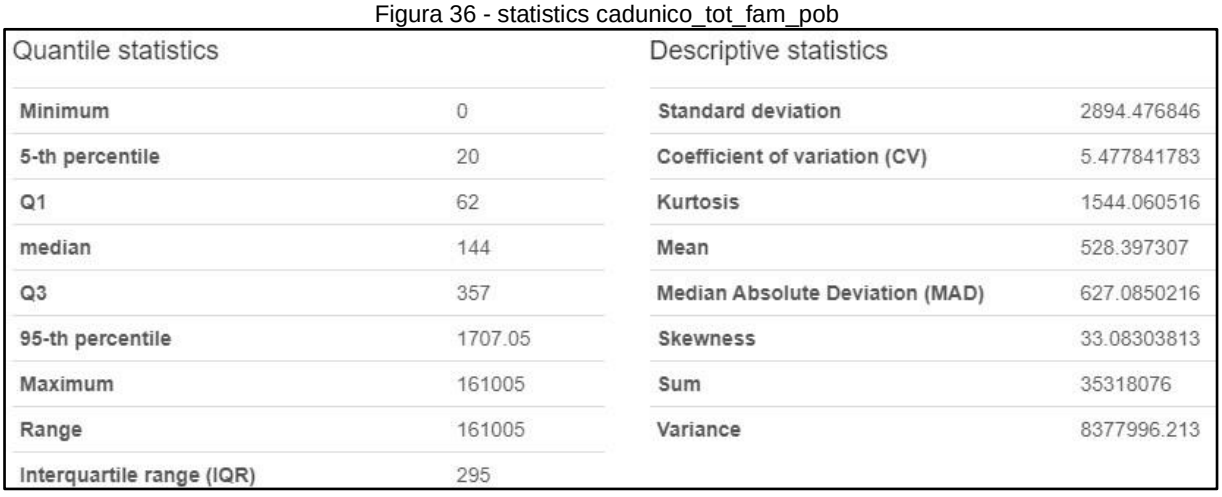
Distinct count	3993	Mean	528.397307
Unique (%)	6.0%	Minimum	0
Missing	0	Maximum	161005
Missing (%)	0.0%	Zeros	31
Infinite	0	Zeros (%)	< 0.1%
Infinite (%)	0.0%	Memory size	522.3 KiB

Fonte: Adaptado de Jupyter Notebook

Na Figura 36 as principais estatísticas da variável cadunico_tot_fam_pob, com o desvio padrão de 2894,47 mil, indicando uma dispersão dos dados. A soma foi de 35.318.076 milhões de famílias na faixa de pobreza, este não é o total de famílias em um único mês e sim a soma de todos os meses do ano de 2019.

Todas as variáveis deste dataset, contém o *Skewness* com valor positivo, indicando uma distribuição inclinada.

Figura 36 - statistics cadunico_tot_fam_pob



Quantile statistics

Minimum	0
5-th percentile	20
Q1	62
median	144
Q3	357
95-th percentile	1707.05
Maximum	161005
Range	161005
Interquartile range (IQR)	295

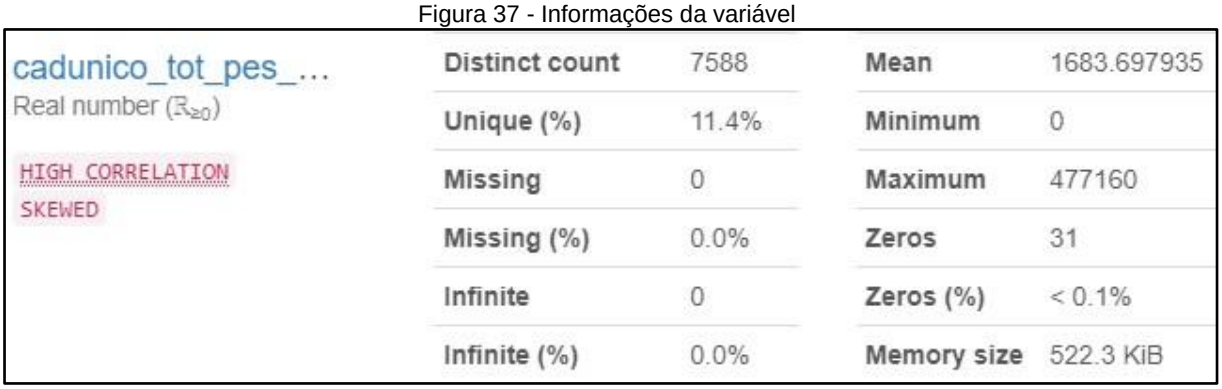
Descriptive statistics

Standard deviation	2894.476846
Coefficient of variation (CV)	5.477841783
Kurtosis	1544.060516
Mean	528.397307
Median Absolute Deviation (MAD)	627.0850216
Skewness	33.08303813
Sum	35318076
Variance	8377996.213

Fonte: Adaptado de Jupyter Notebook

Levando em consideração toda a base de dados, a variável cadunico_tot_pes_pob tem uma média de 1683,69 mil pessoas na faixa de pobreza. A Figura 37 ilustra as principais informações.

Figura 37 - Informações da variável

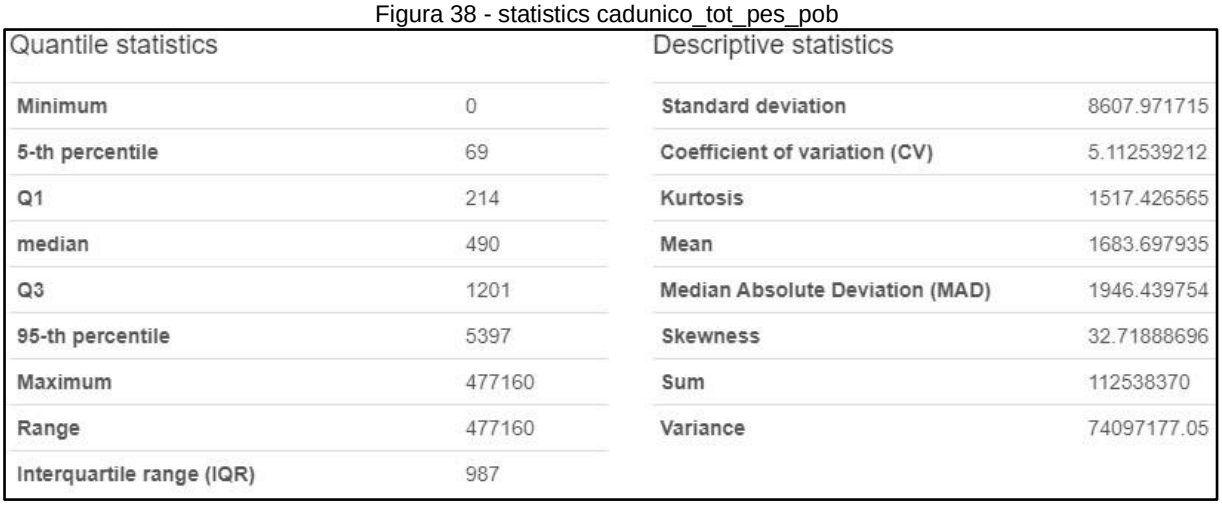


cadunico_tot_pes_…			
Real number ($\mathbb{R}_{\geq 0}$)			
HIGH CORRELATION			
SKEWED			
Distinct count	7588	**Mean**	1683.697935
Unique (%)	11.4%	**Minimum**	0
Missing	0	**Maximum**	477160
Missing (%)	0.0%	**Zeros**	31
Infinite	0	**Zeros (%)**	< 0.1%
Infinite (%)	0.0%	**Memory size**	522.3 KiB

Fonte: Adaptado de Jupyter Notebook

Na Figura 38 as principais estatísticas da variável cadunico_tot_pes_pob, com o desvio padrão de 8607,97 mil, indicando uma dispersão dos dados. A soma foi de 112.538.370 milhões de pessoas na faixa de pobreza, este não é o total de pessoas em um único mês e sim a soma de todos os meses do ano de 2019.

Figura 38 - statistics cadunico_tot_pes_pob

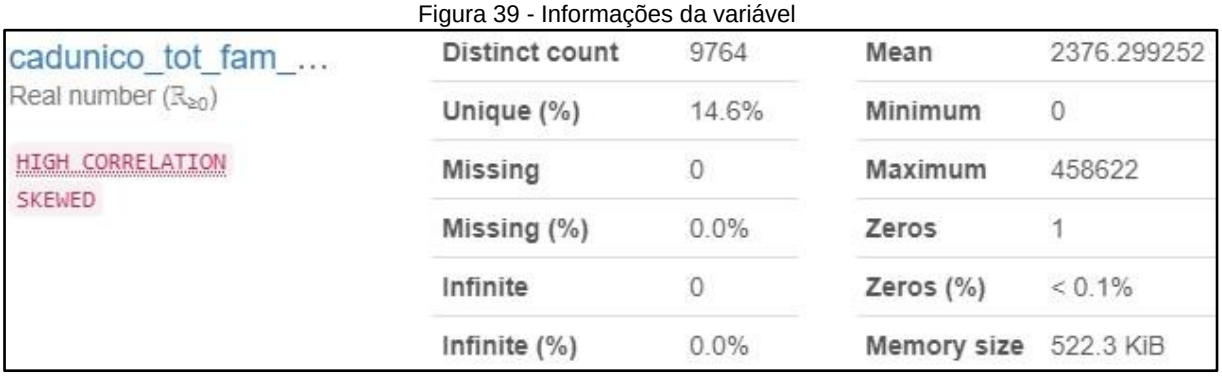


Quantile statistics		Descriptive statistics	
Minimum	0	**Standard deviation**	8607.971715
5-th percentile	69	**Coefficient of variation (CV)**	5.112539212
Q1	214	**Kurtosis**	1517.426565
median	490	**Mean**	1683.697935
Q3	1201	**Median Absolute Deviation (MAD)**	1946.439754
95-th percentile	5397	**Skewness**	32.71888696
Maximum	477160	**Sum**	112538370
Range	477160	**Variance**	74097177.05
Interquartile range (IQR)	987		

Fonte: Adaptado de Jupyter Notebook

Na variável cadunico_tot_fam_ext_pob podemos notar a presença de 1 Zero, neste caso possivelmente uma cidade com nenhuma família cadastrada na faixa de extrema pobreza em um determinado período. Temos uma média de 2376,29 mil famílias na faixa da extrema pobreza. A Figura 39 ilustra as principais informações.

Figura 39 - Informações da variável

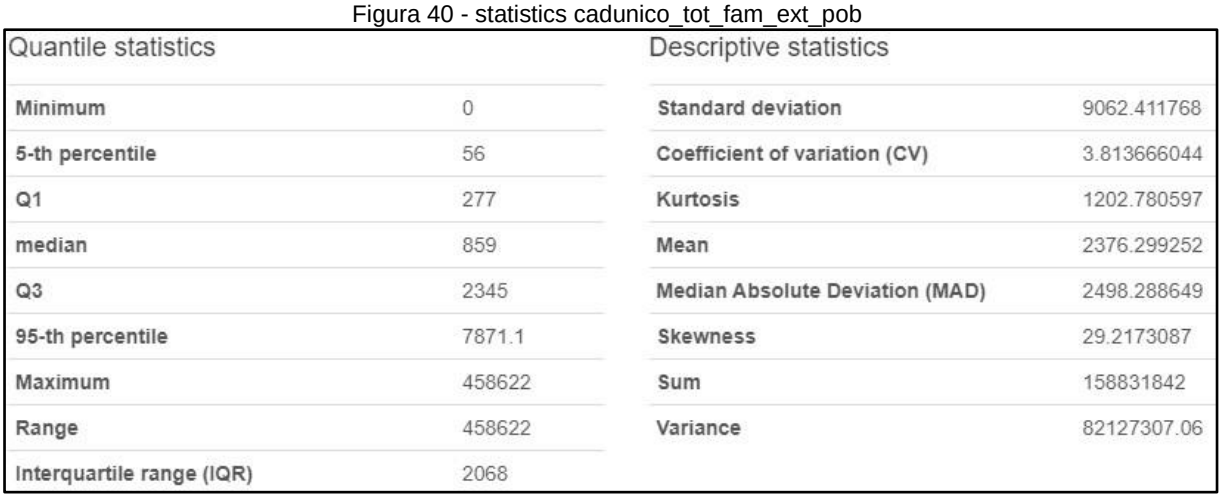


cadunico_tot_fam_…					
Real number ($\mathbb{R}_{\geq 0}$)		Distinct count	9764	Mean	2376.299252
HIGH CORRELATION		Unique (%)	14.6%	Minimum	0
SKEWED		Missing	0	Maximum	458622
		Missing (%)	0.0%	Zeros	1
		Infinite	0	Zeros (%)	< 0.1%
		Infinite (%)	0.0%	Memory size	522.3 KiB

Fonte: Adaptado de Jupyter Notebook

Na Figura 40 as principais estatísticas da variável cadunico_tot_fam_ext_pob, como o desvio padrão de 9062,41 mil, indicando uma dispersão dos dados. A soma foi de 158.831.842 milhões de famílias na faixa da extrema pobreza, este não é o total de famílias em um único mês e sim a soma de todos os meses do ano de 2019.

Figura 40 - statistics cadunico_tot_fam_ext_pob

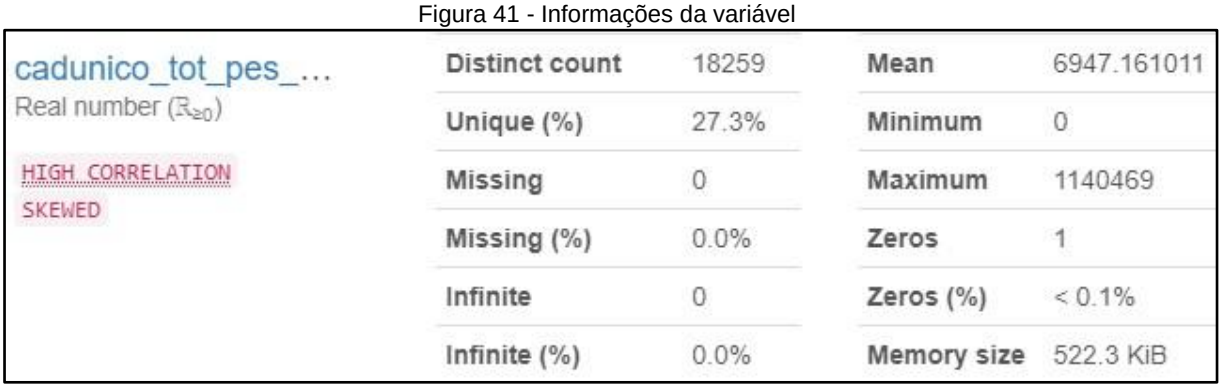


Quantile statistics		Descriptive statistics	
Minimum	0	Standard deviation	9062.411768
5-th percentile	56	Coefficient of variation (CV)	3.813666044
Q1	277	Kurtosis	1202.780597
median	859	Mean	2376.299252
Q3	2345	Median Absolute Deviation (MAD)	2498.288649
95-th percentile	7871.1	Skewness	29.2173087
Maximum	458622	Sum	158831842
Range	458622	Variance	82127307.06
Interquartile range (IQR)	2068		

Fonte: Adaptado de Jupyter Notebook

Levando em consideração toda a base de dados a variável cadunico_tot_pes_ext_pob tem uma média de 6947,16 mil pessoas na faixa da extrema pobreza. A Figura 41 ilustra as principais informações.

Figura 41 - Informações da variável

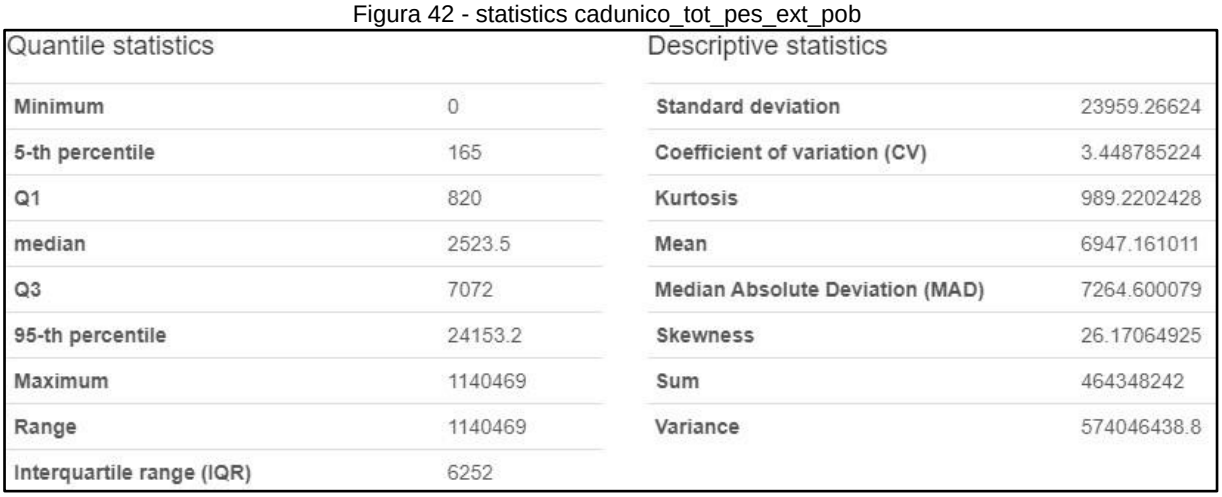


cadunico_tot_pes_…					
Real number ($\mathbb{R}_{\geq 0}$)			Distinct count	18259	Mean	6947.161011
HIGH CORRELATION			Unique (%)	27.3%	Minimum	0
SKEWED			Missing	0	Maximum	1140469
			Missing (%)	0.0%	Zeros	1
			Infinite	0	Zeros (%)	< 0.1%
			Infinite (%)	0.0%	Memory size	522.3 KiB

Fonte: Adaptado de Jupyter Notebook

Na Figura 42 as principais estatísticas da variável cadunico_tot_pes_ext_pob, como o desvio padrão de 23959,26 mil, indicando uma dispersão dos dados. A soma foi de 464.348.242 milhões de pessoas na faixa da extrema pobreza, este não é o total de pessoas em um único mês e sim a soma de todos os meses do ano de 2019.

Figura 42 - statistics cadunico_tot_pes_ext_pob

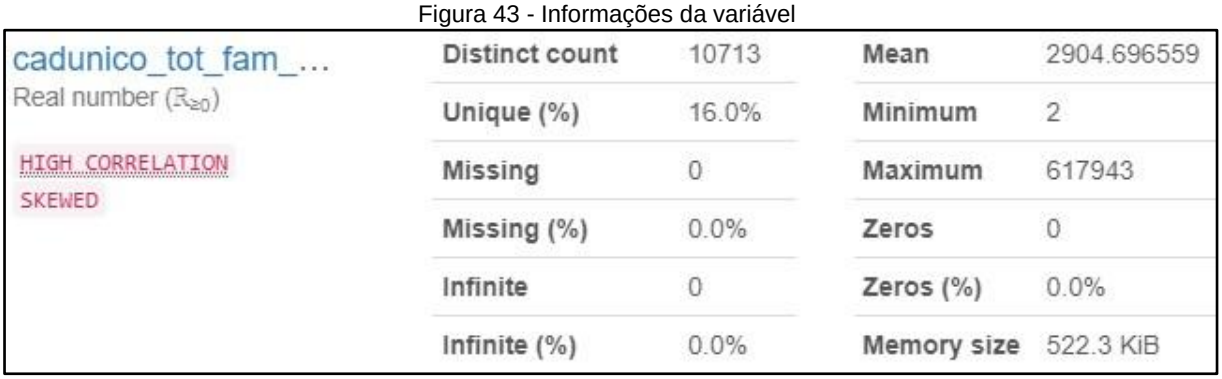


Quantile statistics		Descriptive statistics	
Minimum	0	Standard deviation	23959.26624
5-th percentile	165	Coefficient of variation (CV)	3.448785224
Q1	820	Kurtosis	989.2202428
median	2523.5	Mean	6947.161011
Q3	7072	Median Absolute Deviation (MAD)	7264.600079
95-th percentile	24153.2	Skewness	26.17064925
Maximum	1140469	Sum	464348242
Range	1140469	Variance	574046438.8
Interquartile range (IQR)	6252		

Fonte: Adaptado de Jupyter Notebook

A média da variável cadunico_tot_fam_pob_e_ext_pob foi de 2904,69 mil famílias na faixa da extrema pobreza e pobreza. A Figura 43 ilustra as principais informações.

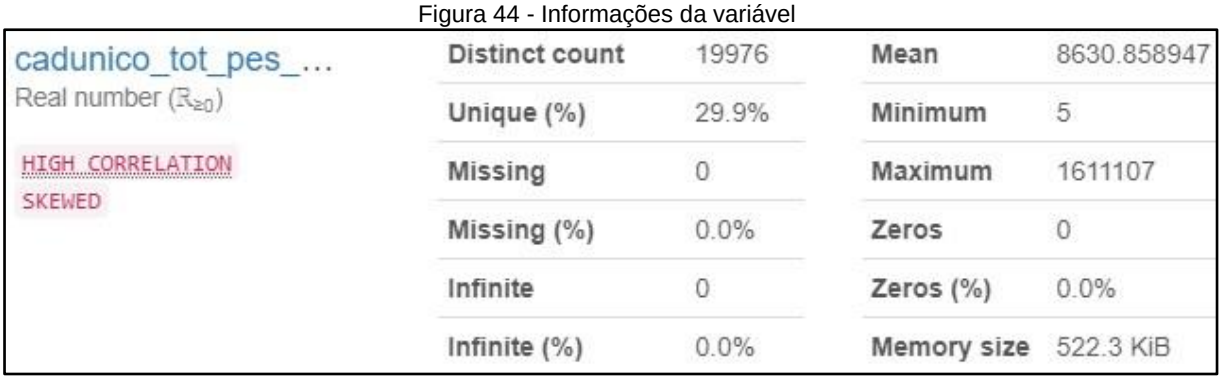
Figura 43 - Informações da variável



Fonte: Adaptado de Jupyter Notebook

A média da variável cadunico_tot_pes_pob_e_ext_pob foi de 8630,85 mil pessoas na faixa da extrema pobreza e pobreza. A Figura 44 ilustra as principais informações.

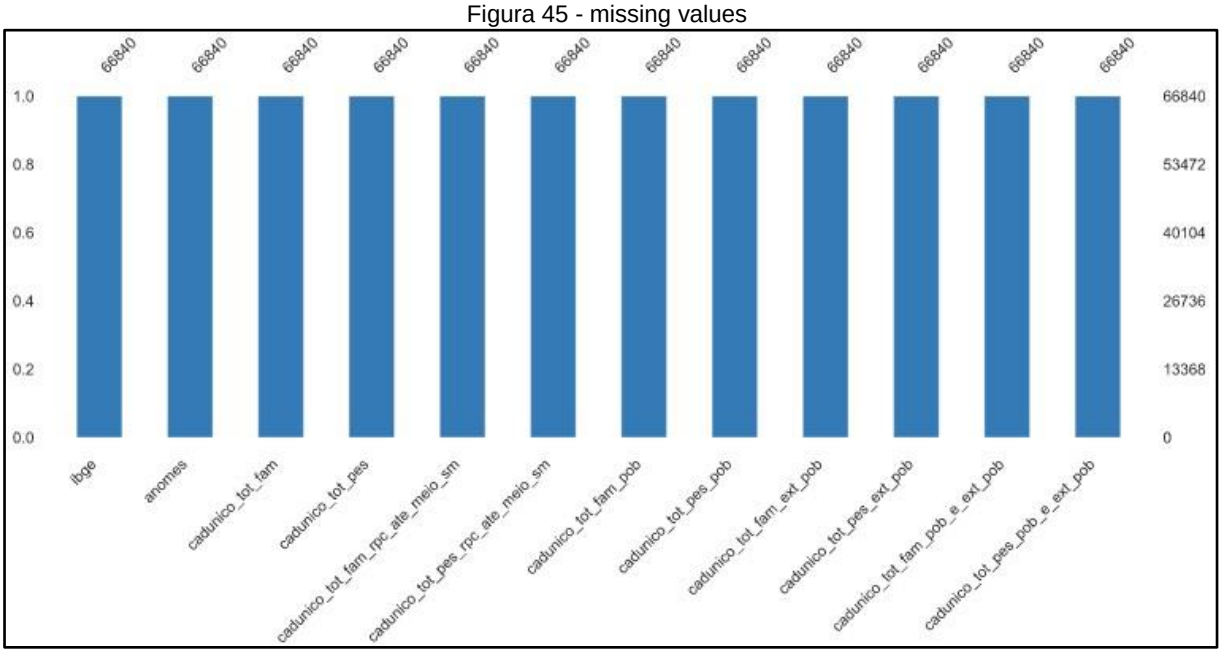
Figura 44 - Informações da variável



Fonte: Adaptado de Jupyter Notebook

Como não temos valores faltantes nesse conjunto de dados, todas as colunas da aba *Count* estão preenchidas, como mostra a Figura 45.

Figura 45 - missing values



Fonte: Adaptado de Jupyter Notebook

A Figura 46 ilustra as 10 primeiras linhas do conjunto de dados.

Figura 46 - First Rows Cadastro Único

	ibge	anomes	cadunico_tot_fam_pob	cadunico_tot_pes_pob	cadunico_tot_fam_ext_pob	cadunico_tot_pes_ext_pob
0	110001	201901	669	2401	968	3226
1	110002	201901	2312	7626	2492	7763
2	110003	201901	122	441	89	289
3	110004	201901	1636	5553	1675	5630
4	110005	201901	360	1222	428	1470
5	110006	201901	245	899	168	536
6	110007	201901	288	967	230	844
7	110008	201901	277	915	1513	5348
8	110009	201901	726	2597	874	3160
9	110010	201901	1167	4353	2270	9727

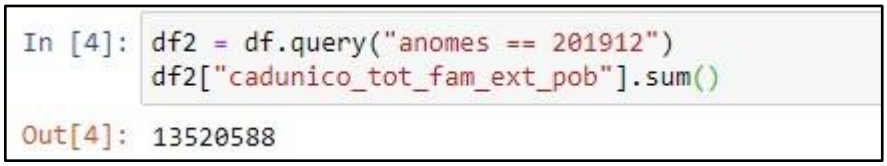
Fonte: Adaptado de Jupyter Notebook

Vamos fazer algumas perguntas para os nossos dados e verificar alguns *Insights*.

Qual o total de famílias em situação de extrema pobreza?

Analisando o mês de dezembro de 2019, o ano fechou com 13.520.588 milhões de famílias em situação de extrema pobreza. Número superior ao de famílias beneficiárias no mesmo período, a Figura 47 ilustra o resultado.

Figura 47 - Total de famílias em situação de extrema pobreza



Fonte: Adaptado de Jupyter Notebook

Qual o total de famílias em situação de pobreza?

São 2.853.527 milhões de famílias em situação de pobreza em dezembro de 2019. A Figura 48 ilustra o resultado.

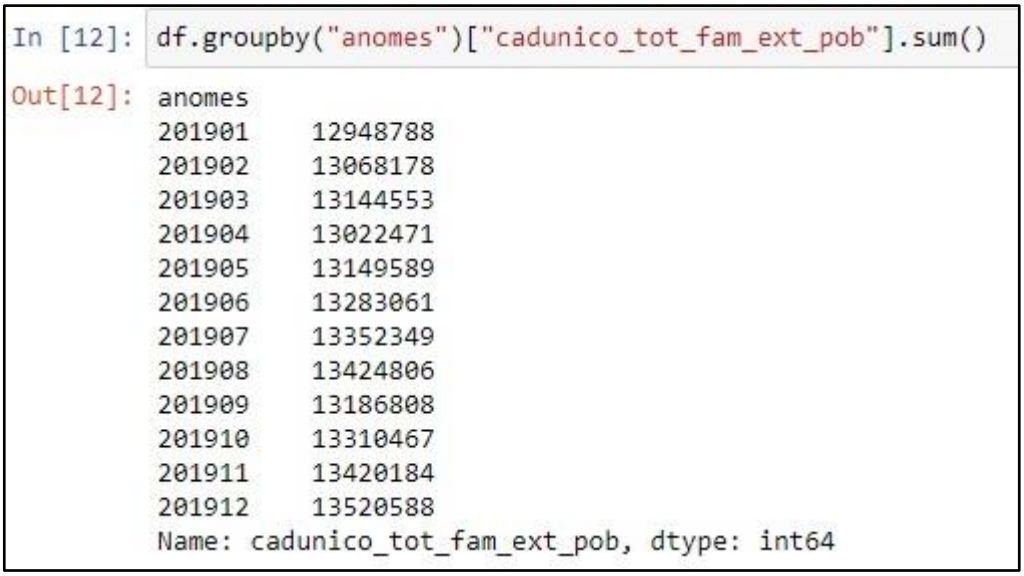
Figura 48 - Total de famílias em situação de pobreza

Fonte: Adaptado de Jupyter Notebook

O total de famílias em situação de extrema pobreza cresceu ou reduziu ao longo do ano?

Podemos observar na Figura 49, que o número de famílias em situação de extrema pobreza oscilou durante o ano 2019, se mantendo em mais de 13 milhões. O ano fechou com crescimento de 571.800 mil famílias, comparado a janeiro do mesmo ano.

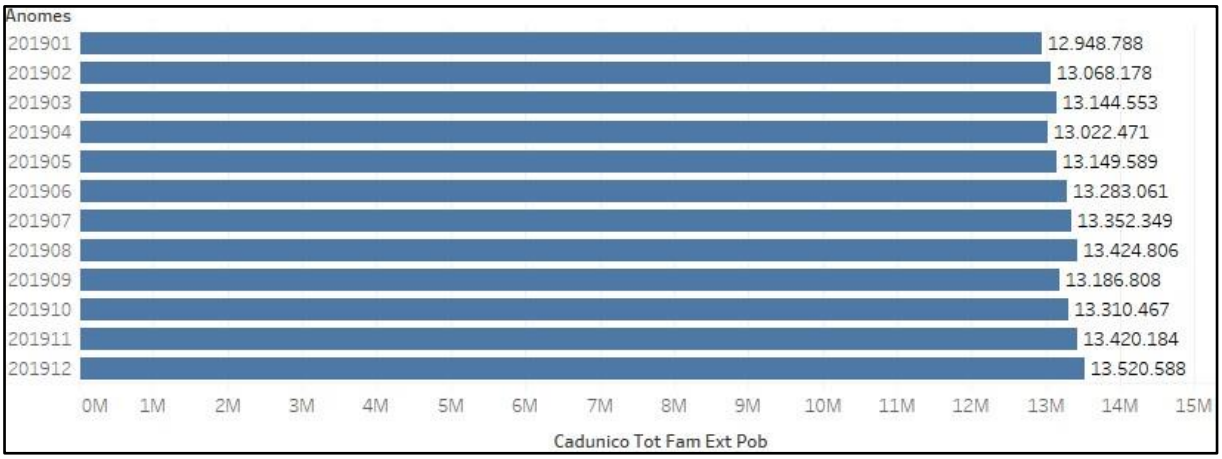
Figura 49 - Total de famílias em situação de extrema pobreza ao longo de 2019



Fonte: Adaptado de Jupyter Notebook

Na Figura 50 vamos visualizar essa mesma análise de forma gráfica.

Figura 50 - Total de famílias em situação de extrema pobreza de forma gráfica



Fonte: Adaptado de Tableau

O total de famílias em situação de pobreza cresceu ou reduziu ao longo do ano?

Podemos observar na Figura 51, que o número de famílias em situação de pobreza teve uma redução ao longo de 2019. O ano fechou com 167.408 mil famílias a menos, comparado a janeiro do mesmo ano. Essas famílias podem ter deixado a pobreza como também entrado na faixa de extrema pobreza.

Figura 51 - Total de famílias em situação de pobreza ao longo de 2019

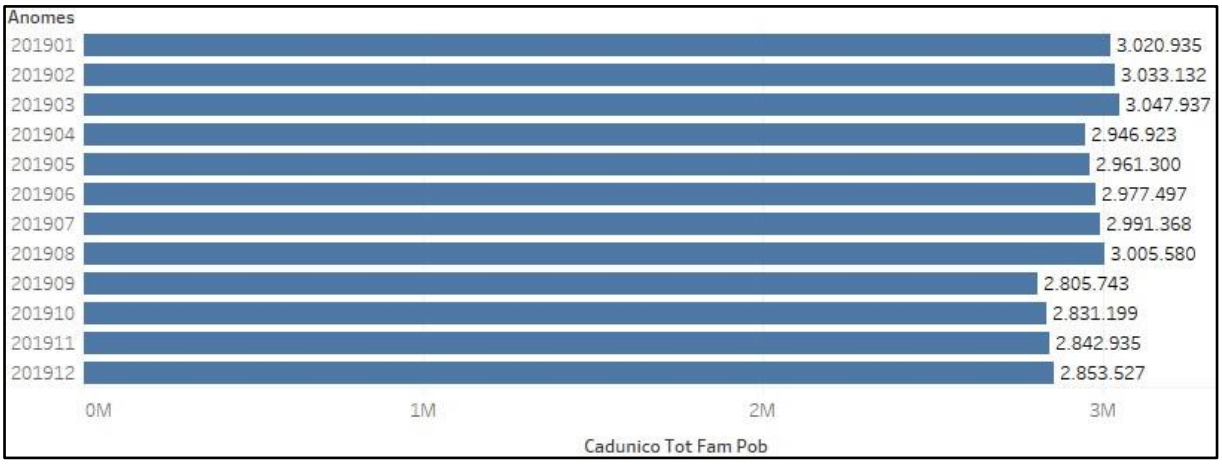
```
In [7]: df.groupby("anomes")["cadunico_tot_fam_pob"].sum()

Out[7]: anomes
        201901    3020935
        201902    3033132
        201903    3047937
        201904    2946923
        201905    2961300
        201906    2977497
        201907    2991368
        201908    3005580
        201909    2805743
        201910    2831199
        201911    2842935
        201912    2853527
        Name: cadunico_tot_fam_pob, dtype: int64
```

Fonte: Adaptado de Jupyter Notebook

Na Figura 52 vamos visualizar essa mesma análise de forma gráfica.

Figura 52 - Total de famílias em situação de pobreza de forma gráfica



Fonte: Adaptado de Tableau

O total de famílias em situação de pobreza e extrema pobreza.

O total de famílias em situação de pobreza e extrema pobreza em dezembro de 2019 foi de 16.374.115 milhões, vivendo com o valor mensal de até R$ 178 por pessoa. A Figura 53 ilustra o resultado.

Figura 53 - Total de famílias em situação de pobreza e extrema pobreza

```
In [12]: df2["cadunico_tot_fam_pob_e_ext_pob"].sum()
Out[12]: 16374115
```

Fonte: Adaptado de Jupyter Notebook

O total de pessoas em situação de pobreza e extrema pobreza.

O total de pessoas em situação de pobreza e extrema pobreza em dezembro de 2019 foi de 48.068.312 milhões, uma parcela considerável da população Brasileira. A Figura 54 ilustra o resultado.

Figura 54 - Total de pessoas em situação de pobreza e extrema pobreza

```
In [13]: df2["cadunico_tot_pes_pob_e_ext_pob"].sum()
Out[13]: 48068312
```

Fonte: Adaptado de Jupyter Notebook

DataSet: Concessões.

A análise será feita da forma como o arquivo foi adquirido, um arquivo para cada estado, antes de realizar a junção de todos em um único arquivo. O estado selecionado para demonstrar a análise foi o estado com código IBGE 11, pelo fato de ter sido o primeiro dataset criado.

Na Figura 55 podemos notar que o dataset tem 3 variáveis, 2 categóricas e uma do tipo numérico, com 74 observações. Não temos valores ausentes nem linhas duplicadas.

Figura 55 – Overview Concessões

Dataset statistics		Variable types	
Number of variables	3	CAT	2
Number of observations	74	NUM	1
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	5.9 KiB		
Average record size in memory	81.7 B		

Fonte: Adaptado de Jupyter Notebook

A média de novas concessões ao longo dos anos é de 161,37 para este estado, com no mínimo uma concessão e no máximo 964 concessões. A Figura 56 ilustra as principais informações.

Figura 56 - Informações da variável

Quantidade de novas concessoes Real number ($\mathbb{R}_{\geq 0}$)				
	Distinct count	71	Mean	161.374473
	Unique (%)	95.9%	Minimum	1
	Missing	0	Maximum	964
	Missing (%)	0.0%	Zeros	0
	Infinite	0	Zeros (%)	0.0%
	Infinite (%)	0.0%	Memory size	720.0 B

Fonte: Adaptado de Jupyter Notebook

Na Figura 57 as principais estatísticas da variável Quantidade de novas concessões, como o desvio padrão de 299,06, indicando uma dispersão dos dados.

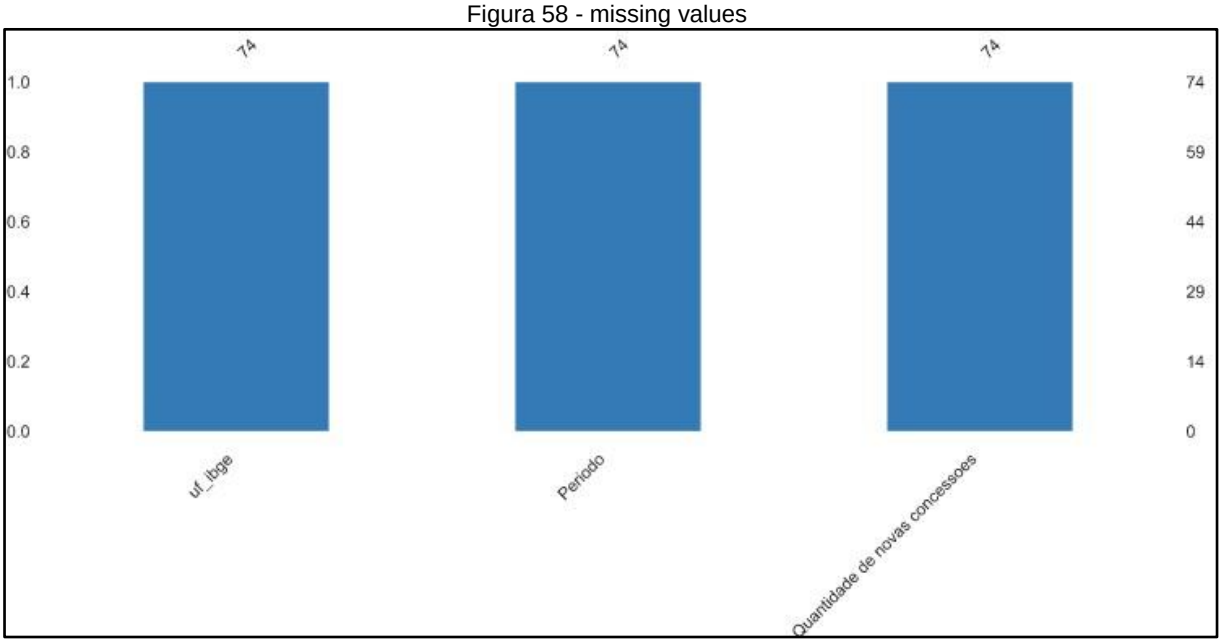
Figura 57 - statistics Quantidade de novas concessões

Quantile statistics		Descriptive statistics	
Minimum	1	Standard deviation	299.0602043
5-th percentile	1.1983	Coefficient of variation (CV)	1.853206389
Q1	1.648	Kurtosis	1.591700602
median	4.7875	Mean	161.374473
Q3	59	Median Absolute Deviation (MAD)	228.1792962
95-th percentile	917.3	Skewness	1.762636029
Maximum	964	Sum	11941.711
Range	963	Variance	89437.00578
Interquartile range (IQR)	57.352		

Fonte: Adaptado de Jupyter Notebook

Como não temos valores faltantes nesse conjunto de dados, todas as colunas estão preenchidas na seção *missing values*, como mostra a Figura 58.

Figura 58 - missing values



Fonte: Adaptado de Jupyter Notebook

A Figura 59 ilustra as 10 últimas linhas do conjunto de dados.

Figura 59 - Last Rows Concessões

	uf_ibge	Periodo	Quantidade de novas concessoes
64	11	10/2014	1
65	11	09/2014	35
66	11	08/2014	59
67	11	07/2014	4.790
68	11	06/2014	1.092
69	11	05/2014	1.330
70	11	04/2014	1.800
71	11	03/2014	2.404
72	11	02/2014	950
73	11	01/2014	622

Fonte: Adaptado de Jupyter Notebook

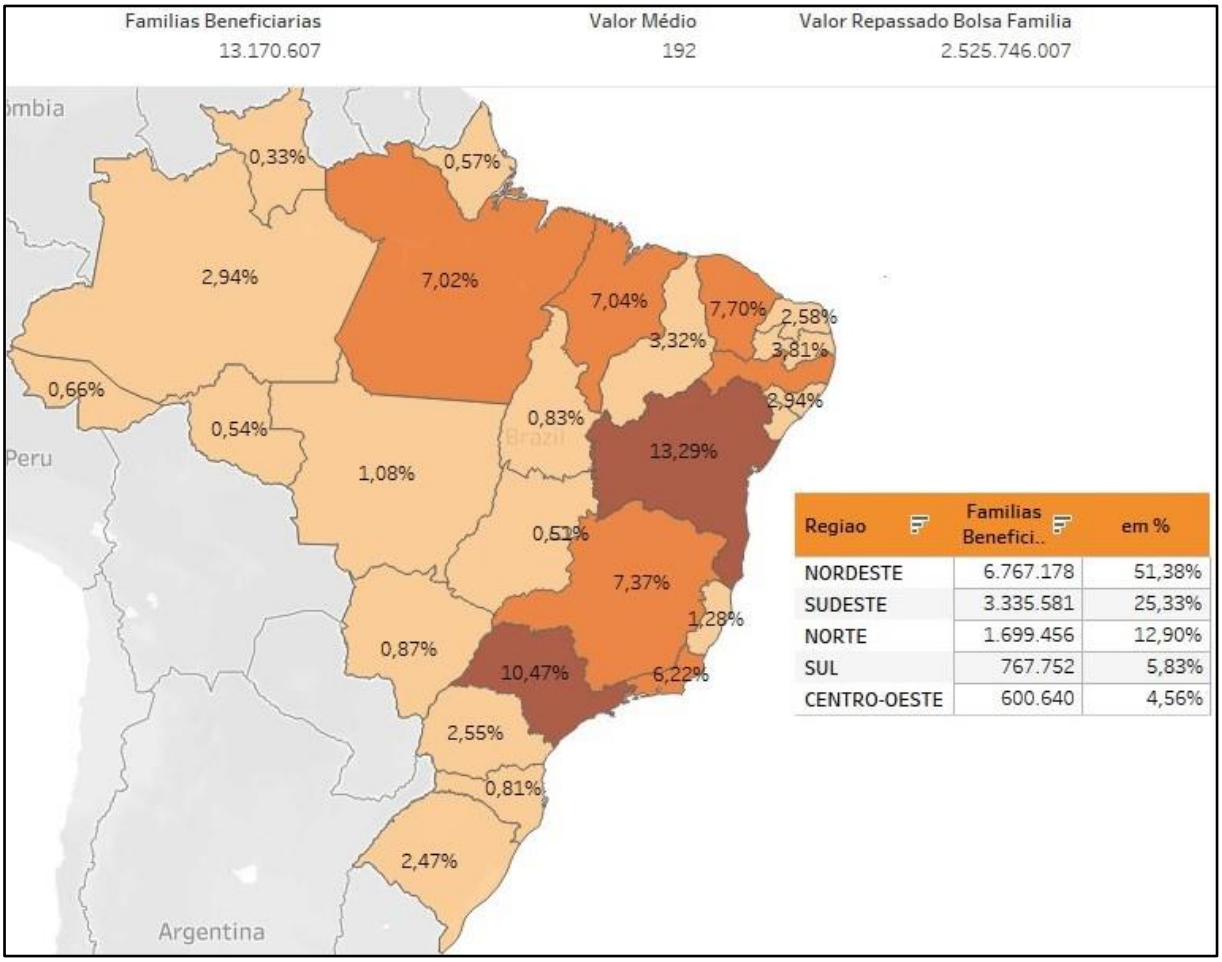
5. Apresentação dos Resultados

Após as etapas de Processamento/Tratamento e Análise Exploratória de Dados, vamos aos principais resultados do trabalho. Vamos analisar a distribuição de famílias beneficiárias por estado e região, para isso foi utilizado a ferramenta *Tableau* pela facilidade de uso e licença estudantil disponibilizada pela PUC.

O mapa na Figura 60 mostra que a distribuição das famílias beneficiárias no Brasil, proporcionalmente, está extremamente concentrada nas regiões do Nordeste e Sudeste. Os três estados com maior percentual, estão nessas regiões: Bahia 13,29%, São Paulo 10,47% e Pernambuco 8,56%.

O Nordeste aparece com mais da metade das famílias beneficiárias; 51,38% das 13.170.607 milhões de famílias, seguido por Sudeste 25,33%, Norte 12,90%, Sul e Centro-Oeste entre 4% e 6%. Com isso, percebemos que o programa Bolsa Família alcança todos os estados do Brasil, concentrando a maior parte nas regiões mais populosas.

Figura 60 - Distribuição das famílias beneficiárias no Brasil



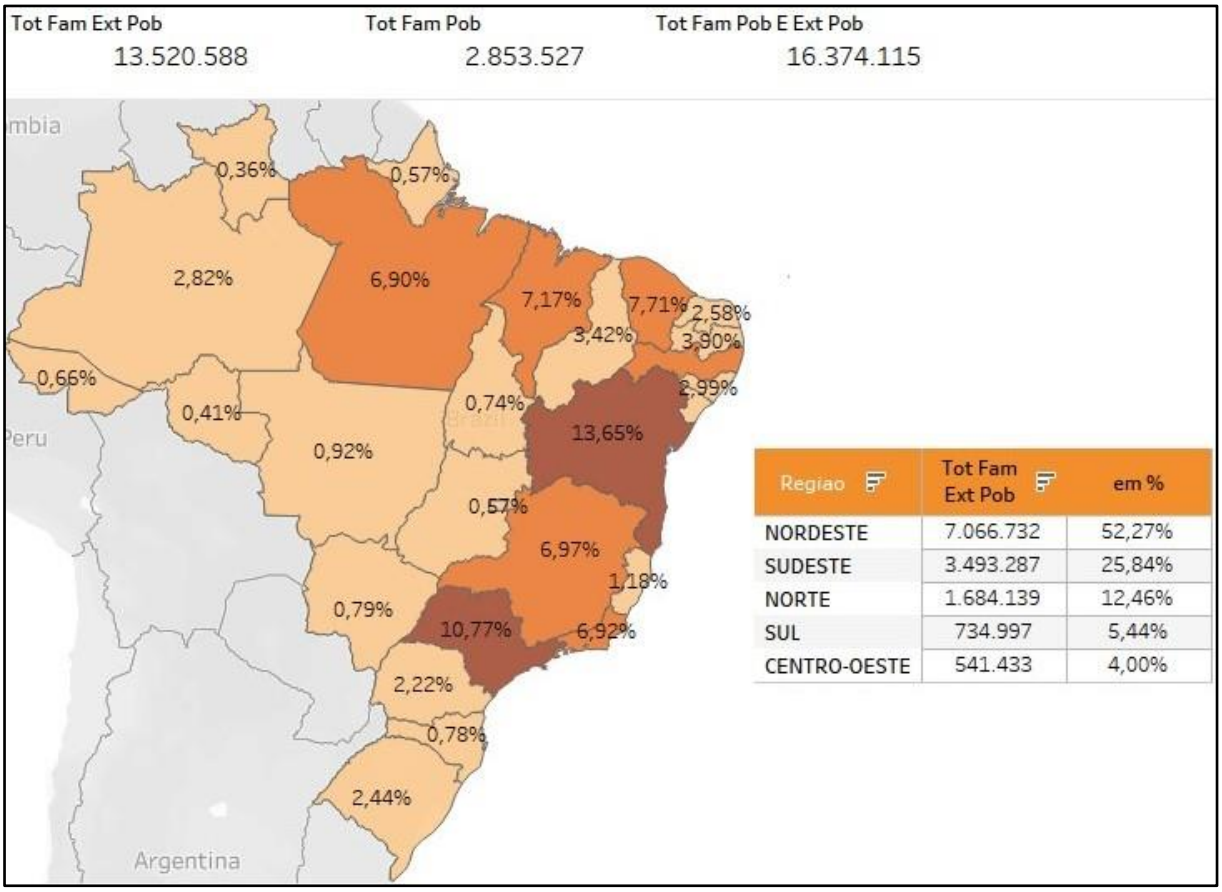
Fonte: Adaptado de Tableau

O valor médio pago as famílias foi de R$ 192, totalizando R$ 2.525.746.007 bilhões, repassados pelo programa no mês de dezembro de 2019.

Analisando a distribuição das famílias em extrema pobreza, o mapa na Figura 61 nos mostra percentuais muito próximos da Figura 60, os três estados com maior percentual são: Bahia 13,65%, São Paulo 10,77% e Pernambuco 8,72%. Olhando o mapa como um todo, inferimos que distribuição dos beneficiários tem atingindo as regiões com maior concentração populacional e de extrema pobreza.

O Nordeste aparece com 52,27% das 13.520.588 milhões de famílias em extrema pobreza. Esse é mais um indicador que o Nordeste é a região mais pobre do Brasil, situação que pode ser atribuída ao desenvolvimento histórico, político e social desfavorável. A região Sudeste aparece com 25,84%, seguido por Norte 12,46%, Sul e Centro-Oeste entre 4% e 6%.

Figura 61 - Distribuição das famílias em situação de extrema pobreza no Brasil



Regiao	Tot Fam Ext Pob	em %
NORDESTE	7.066.732	52,27%
SUDESTE	3.493.287	25,84%
NORTE	1.684.139	12,46%
SUL	734.997	5,44%
CENTRO-OESTE	541.433	4,00%

Fonte: Adaptado de Tableau

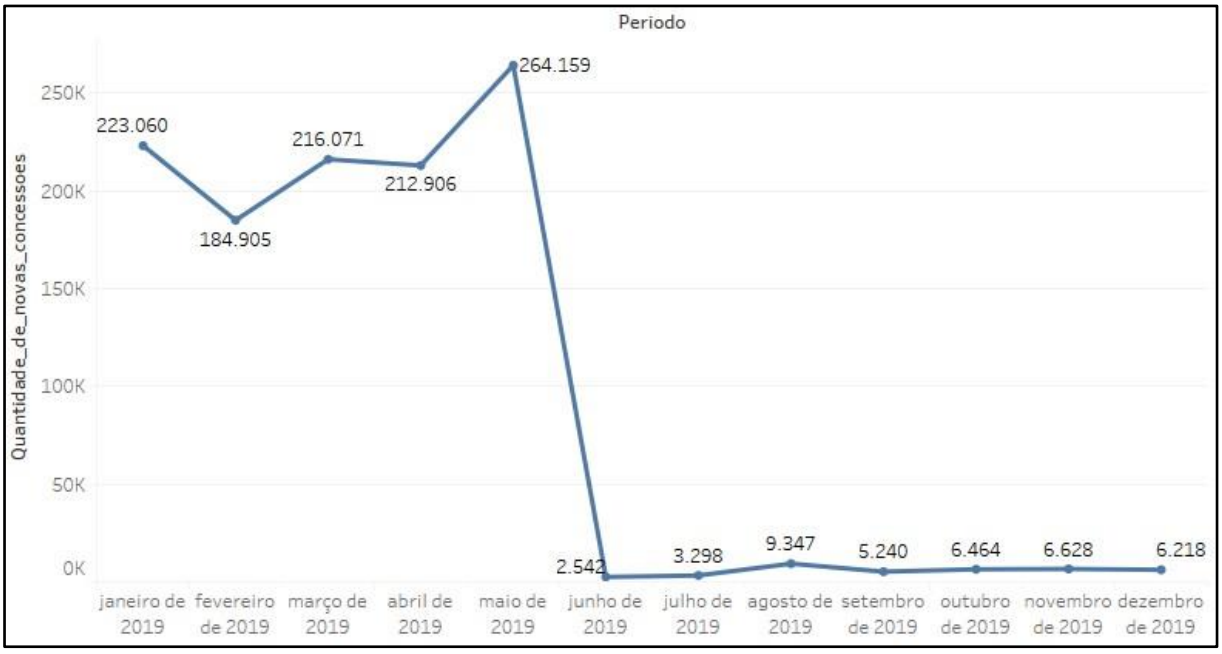
Olhando de forma isolada para as regiões Nordeste e Sudeste, temos um percentual maior comparado com o mapa de famílias beneficiárias. Isso nos permitir concluir, que existem famílias no cadastro único em situação de extrema pobreza que ainda não são beneficiárias, estão aguardando novas concessões, formando assim a fila de espera do programa.

Para enriquecer os dados apresentados, vamos analisar a quantidade de novas concessões ao longo de 2019 e de toda a série histórica disponível.

Podemos observar na Figura 62, que o total de novas concessões oscilou de janeiro até maio de 2019, tendo uma média de 220.220,2 mil concessões por mês. No mês de junho o governo concedeu 2.542 concessões em todo o País, uma queda drástica comparado aos meses anteriores.

Em maio de 2019 o governo anunciava que a fila de espera do programa estava zerada, o que pode estar relacionado a essa queda no número de novas concessões no mês de junho do mesmo ano.

Figura 62 - Quantidade de novas concessões no ano de 2019

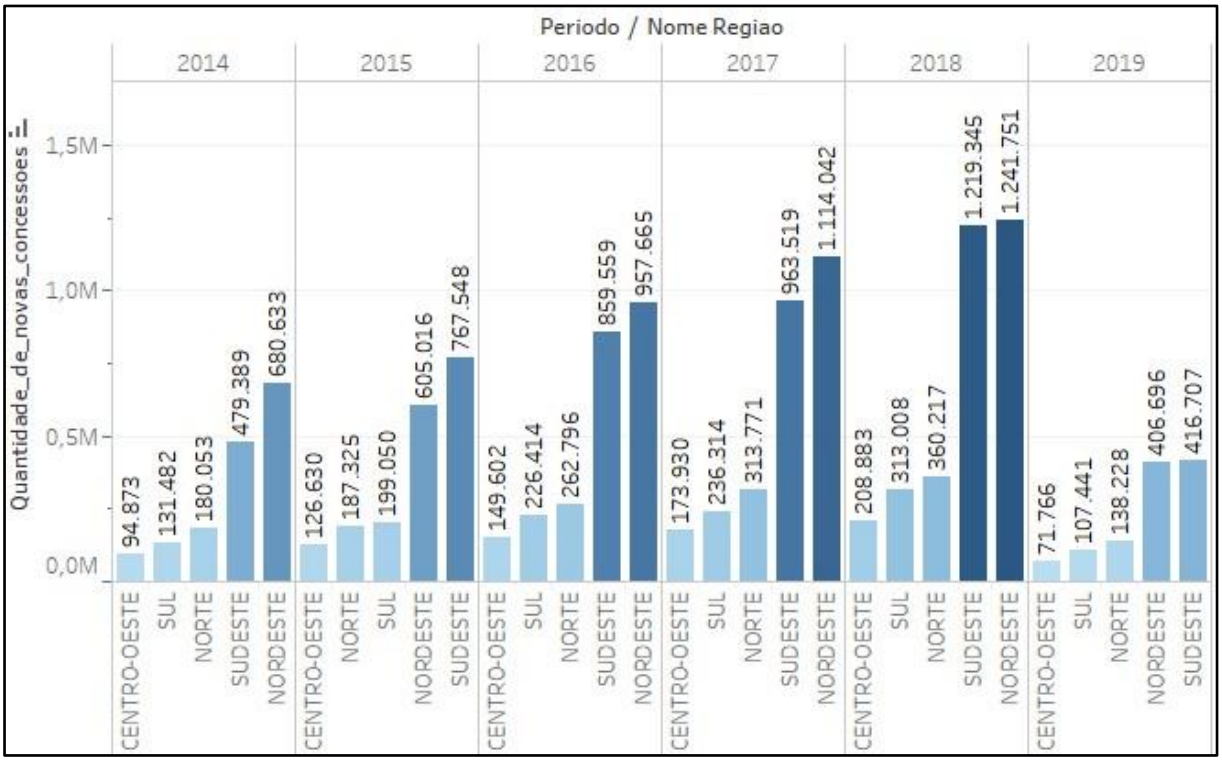


Fonte: Adaptado de Tableau

Entre junho e dezembro a quantidade de novas famílias que entraram no programa despencou, ficando com uma média de 5.676,71 concessões por mês, muito abaixo comparado a média dos meses de janeiro a maio de 2019.

Na Figura 63 é possível perceber um aumento, crescente, no número de novas concessões em todas as regiões entre 2014 e 2018. Nordeste e Sudeste com mais da metade de todas as concessões nos últimos cinco anos, fato que pode estar relacionado a grande concentração de população nessas regiões, bem como o aumento da extrema pobreza e pobreza no País.

Figura 63 - Série histórica por período e região



Fonte: Adaptado de Tableau

Na contramão dos anos anteriores, 2019 apresenta o menor número de concessões dos últimos cinco anos, em todas as regiões. O quadro é inédito na história do programa, pelo menos entre 2014, primeiro ano de dados disponíveis, e 2018.

Dessa forma, a rotatividade do programa acaba sendo desigual, por um lado, como observado ao longo da análise, há famílias que continuam saindo do programa, por outro, a entrada está praticamente emperrada, sendo que há demanda. Isso leva à redução no número total de famílias beneficiárias e ao aumento na fila de espera para entrar no programa.

Afinal qual o tamanho da fila de espera do Bolsa Família?

Ao longo da análise percebemos que o total de famílias em situação de extrema pobreza é maior que o total de famílias beneficiárias, essa diferença é a fila de espera do Bolsa Família.

Um ponto importante a ser considerado neste trabalho, com os dados disponíveis não é possível determinar em que faixa as famílias beneficiárias permaneciam, antes de receber de fato o benefício, com isso não temos a proporção exata de famílias em extrema pobreza que já são beneficiárias do programa, ou seja,

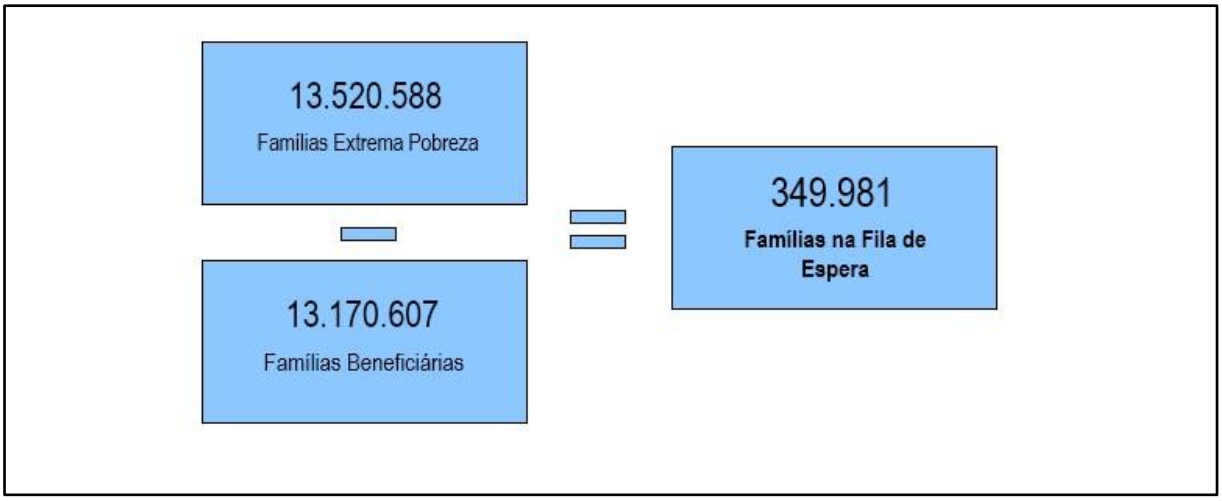
o tamanho da fila não reflete com exatidão as famílias em extrema pobreza aguardando novas concessões.

Assim, vamos utilizar uma fórmula que considera todas as famílias beneficiárias sendo oriundas da extrema pobreza, logo temos a seguinte fórmula:

Fila de espera = Total de famílias em extrema pobreza – Total de famílias beneficiárias.

A Figura 64 apresenta o resultado da aplicação desta fórmula.

Figura 64 - Total de famílias na fila de espera



Fonte: Autor

Fila de espera: 349.981 mil famílias, essa é a fila de espera ao final de dezembro de 2019. Transformando em porcentagem, temos aproximadamente 2,59% do total de famílias em situação de extrema pobreza.

Multiplicando o total de pessoas em extrema pobreza, 39.090.704 milhões x 2,59%, temos aproximadamente 1.012.449 milhões de pessoas na fila de espera do programa. Ou seja, menos famílias/pessoas registradas no Cadastro Único, com renda comprovadamente baixa, estão conseguindo ter acesso ao benefício. Assim, elas formam a fila de espera do programa.

O número total de famílias/pessoas na fila de espera pode ser ainda maior, caso sejam consideradas as famílias/pessoas em situação de pobreza com renda mensal entre R$ 89,01 e R$ 178 por pessoa.

A situação de filas no Bolsa Família não é novidade, o que foge da normalidade são os cortes sucessivos e a redução drástica de novas concessões, isso em um

cenário em que a extrema pobreza está aumentando no país. Isso acaba por diminuir a efetividade do programa e aumentar a desigualdade no país.

Além de ser um importante instrumento para atacar a desigualdade de renda no Brasil, o Bolsa Família cumpre papel relevante na atividade econômica do país. A transferência de renda para famílias mais pobres aumenta o consumo que, por sua vez, reflete pela economia. Sem isso, sofrem a economia local e as pessoas mais carentes.

6. Links

Repositório: https://github.com/dieisongularte/PUCMG_TCC
Vídeo: https://youtu.be/9lSXne7063k