

MODELO PARA A ENTREGA DAS ATIVIDADES

COMPONENTE CURRICULAR:	Análise Exploratória de Dados
NOME COMPLETO DO ALUNO:	Anderson, Aparecido da Silva Alves, Andréia Domingos dos Santos, Gerson Soares Rodrigues, Samuel Bonfim da Silva
RA:	10347602, 10288503, 10423804, 10423569

Atenção: Toda atividade deverá ser feita com fonte Arial, tamanho 11, espaço de 1,5 entre as linhas e alinhamento justificado entre as margens.

Aplicando 7: Grupo 3: Global Crop Yields (Brazil)

1. INTRODUÇÃO

A produção de grãos no Brasil tem experimentado um crescimento notável nas últimas décadas, impulsionada por avanços tecnológicos, melhores práticas agrícolas e políticas públicas de incentivo. Este projeto visa apresentar uma análise detalhada dessa expansão, explorando a evolução da produtividade de culturas chave, e eficiência no uso da terra e as interdependências entre os diferentes tipos de grãos. Com uma abordagem técnica e informativa, estes estudos visam fornecer insights valiosos para profissionais e pesquisadores do setor agrícola brasileiro.

2. EXPLORANDO OS DADOS AGRÍCOLAS

Para o desenvolvimento do projeto, utilizamos a linguagem de programação Python para a programação coletamos alguns DataFrames diretamente no GitHub e para a leituras de alguns comandos necessários para a análise dos dados utilizaremos as seguintes Bibliotecas:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib.patches import Patch
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
```

```
from sklearn.metrics import accuracy_score  
from google.colab import files
```

Feitas as Importações das Bibliotecas, agora utilizaremos os arquivos em formato csv coletados diretamente do link presente no GitHub e partir deles criar os DataFrames.

```
df_key_crop_yields =  
pd.read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-09-01/key_crop_yields.csv')  
df_fertilizer =  
pd.read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-09-01/cereal_crop_yield_vs_fertilizer_application.csv')  
df_tractors =  
pd.read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-09-01/cereal_yields_vs_tractor_inputs_in_agriculture.csv')  
df_land_use =  
pd.read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-09-01/land_use_vs_yield_change_in_cereal_production.csv')  
df_arable_land =  
pd.read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-09-01/arable_land_pin.csv')
```

Os DataFrames utilizados para criação do projeto terão como base a exploração das seguintes informações:

- Principais rendimentos das colheitas em escala mundial;
- Fertilizantes utilizados;
- Maquinários;
- Uso de terra;
- Solo arável

Sendo estes DataFrames explorados inicialmente explorados de forma individual, para que sejam coletadas apenas as informações pertinentes ao projeto e para o tratamento de algumas variáveis, se assim houver a necessidade. A seguir iniciaremos nossa análise a partir do DataFrame referente as colheitas em escala mundial.

3. TRATAMENTO DOS DADOS AGRÍCOLAS

3.1 PRINCIPAIS RENDIMENTOS DAS COLHEITAS EM ESCALA MUNDIAL

Neste DataFrame, foram obtidos os valores brutos principais rendimentos das colheitas em escala mundial, e, como houve a presença de alguns valores não numéricos (NaN)

contidos em algumas colunas, estes foram convertidos em zero para que a coluna pudesse ser analisada por completo sem que isso afetasse nos valores presentes das colunas, através das seguintes linhas de comando:

```
df_key_crop_yields.fillna(0, inplace=True)
df_key_crop_yields.fillna(0, inplace=True)
columns_to_fill = ['Wheat (tonnes per hectare)', 'Rice (tonnes per hectare)', 'Maize (tonnes per hectare)', 'Soybeans (tonnes per hectare)', 'Potatoes (tonnes per hectare)', 'Beans (tonnes per hectare)', 'Peas (tonnes per hectare)', 'Cassava (tonnes per hectare)', 'Barley (tonnes per hectare)', 'Cocoa beans (tonnes per hectare)', 'Bananas (tonnes per hectare)']
df_key_crop_yields[columns_to_fill] =
df_key_crop_yields[columns_to_fill].fillna(0)
df_key_crop_yields.head()
```

Após a conversão dos valores não-numéricos em nulos, foi criado um outro DataFrame, em que apenas os valores referentes ao Brasil serão adicionados, conforme poderá ser observado logo abaixo:

3.2 PRINCIPAIS RENDIMENTOS DAS COLHEITAS REFERENTES AO BRASIL

Criação de um novo DataFrame dos principais rendimentos das colheitas com valores referentes somente ao Brasil

```
df_key_crop_yields_BR = df_key_crop_yields[df_key_crop_yields.Entity == 'Brazil']
df_key_crop_yields_BR
```

Para que não ocorra nenhum conflito na hora de executar as linhas de comando, devido à presença de alguns caracteres especiais nos títulos das colunas, foram renomeadas as colunas de df_key_crop_yields_BR, conforme pode ser observado nas linhas de comando a seguir:

```
from os import rename
df_key_crop_yields_BR = df_key_crop_yields_BR.rename(columns={'Wheat (tonnes per hectare)': 'Trigo_Toneladas_por_Hectare', 'Soybeans (tonnes per hectare)': 'Soja_Toneladas_por_Hectare', 'Rice (tonnes per hectare)': 'Arroz_Toneladas_por_Hectare', 'Maize (tonnes per hectare)': 'Milho_Toneladas_por_Hectare', 'Potatoes (tonnes per hectare)': 'Batata_Toneladas_por_Hectare', 'Beans (tonnes per hectare)': 'Feijão_Toneladas_por_Hectare', 'Peas (tonnes per hectare)': 'Ervilha_Toneladas_por_Hectare', 'Cassava (tonnes per hectare)': 'Mandioca_Toneladas_por_Hectare', 'Barley (tonnes per hectare)': 'Cevada_Toneladas_por_Hectare', 'Cocoa beans (tonnes per hectare)': 'Cacau_Toneladas_por_Hectare', 'Bananas (tonnes per hectare)': 'Banana_Toneladas_por_Hectare'})
df_key_crop_yields_BR
```

O comando a seguir será utilizado para a análise as colunas e identificação de quais variáveis são categóricas e não-categóricas

```
df_key_crop_yields_BR.columns
```

O comando a seguir será utilizado para a determinação dos valores de Medianas, Desvios Padrão, Mínimos, Máximos e Quartis

```
df_key_crop_yields_BR.describe().drop('count')
```

O comando a seguir será utilizado para a verificação da existência de valores nulo no dataframe df_key_crop_yields_BR

```
df_key_crop_yields_BR.isnull().sum()
```

A seguir serão realizadas as análises a partir do DataFrame referente aos Fertilizantes em escala mundial.

4. DATAFRAME REFERENTE AOS FERTILIZANTES

Assim como no caso anterior, neste DataFrame foram obtidos os valores brutos dos Fertilizantes em escala mundial, e, como houve a presença de alguns valores não numéricos (NaN) contidos em algumas colunas, este foram convertidos em zero para que a coluna pudesse ser analisada por completo sem que isso afetasse nos valores presentes das colunas, através das seguintes linhas de comando:

```
df_fertilizer.fillna(0, inplace=True)
columns_to_fill = ['Nitrogen fertilizer use (kilograms per hectare)']
df_fertilizer[columns_to_fill] =
df_fertilizer[columns_to_fill].fillna(0)
df_fertilizer.head()
```

4.1 CRIAÇÃO DE UM NOVO DATAFRAME, EM QUE APENAS ESTARÃO CONTIDOS OS VALORES REFERENTES AO BRASIL

```
df_fertilizer_BR = df_fertilizer[df_fertilizer.Entity == 'Brazil']
df_fertilizer_BR
```

Para que não ocorra nenhum conflito na hora de executar as linhas de comando, devido à presença de alguns caracteres especiais nos títulos das colunas, foram renomeadas as colunas de df_fertilizer_BR, conforme pode ser observado nas linhas de comando a seguir:

```
from os import rename
df_fertilizer_BR = df_fertilizer_BR.rename(columns={'Cereal yield
(tonnes per hectare)': 'Rendimento_de_cereais_Ton_por_Hec ', 'Nitrogen
fertilizer use (kilograms per hectare)': 'Fertilizantes_Nitrogenados_Kg_por_Hec'})
df_fertilizer_BR
```

4.2 VERIFICANDO OS TIPOS DE VARIÁVEIS PRESENTES NO DATAFRAME DF_FERTILIZER_BR

```
df_fertilizer_BR.info()
```

4.3 DETERMINAÇÃO DOS VALORES DE MEDIANAS, DESVIOS PADRÃO, MÍNIMOS, MÁXIMOS E QUARTIS

```
df_fertilizer_BR.describe().drop('count')
```

4.4 VERIFICANDO A EXISTÊNCIA DE VALORES NULOS NO DATAFRAME DF_FERTILIZER_BR

```
df_fertilizer_BR.isnull().sum()
```

A seguir serão realizadas as análises a partir do DataFrame referente aos Maquinários em escala mundial.

5. DATAFRAME REFERENTE AOS MAQUINÁRIOS (TRATORES)

Assim como nos casos anteriores, neste DataFrame foram obtidos os valores brutos dos Maquinário, em especial os tratores, em escala mundial, e, como houve a presença de alguns valores não numéricos (NaN) contidos em algumas colunas, este foram convertidos em zero para que a coluna pudesse ser analisada por completo sem que isso afetasse nos valores presentes das colunas, através das seguintes linhas de comando:

```
df_tractors.columns = df_tractors.columns.str.replace(' ', '_')
df_tractors['Tractors_per_100_sq_km_arable_land'].fillna(0,
inplace=True)
df_tractors['Cereal_yield_(kilograms_per_hectare)_(kg_per_hectare)']
.fillna(0, inplace=True)
df_tractors.head()
```

5.1 CRIAÇÃO DE UM NOVO DATAFRAME, EM QUE APENAS ESTARÃO CONTIDOS OS VALORES REFERENTES AO BRASIL

```
df_tractors_BR = df_tractors[df_tractors.Entity == 'Brazil']
df_tractors_BR
```

Para que não ocorra nenhum conflito na hora de executar as linhas de comando, devido à presença de alguns caracteres especiais nos títulos das colunas, foram renomeadas as colunas de df_tractors_BR, conforme pode ser observado nas linhas de comando a seguir:

```
from os import rename
df_tractors_BR =
df_tractors_BR.rename(columns={'Tractors_per_100_sq_km_arable_land':
'Tratores_por_100Km²_terra_aravel', 'Cereal_yield_(kilograms_per_hect
are)_(kg_per_hectare)': 'Rendimento_Cereal_Kg_por_Hec',
'Total_population_(Gapminder)': 'Populacao_Total'})
df_tractors_BR
```

5.2 VERIFICANDO OS TIPOS DE VARIÁVEIS PRESENTES NO DATAFRAME DF_TRACTORS_BR

```
df_tractors_BR.info()
```

5.3 DETERMINAÇÃO DOS VALORES DE MEDIANAS, DESVIOS PADRÃO, MÍNIMOS, MÁXIMOS E QUARTIS

```
df_tractors_BR.describe().drop('count')
```

5.4 VERIFICANDO SE HÁ OS VALORES NULOS EM DF_TRACTORS_BR

```
df_tractors_BR.isnull().sum()
```

A seguir serão realizadas as análises a partir do DataFrame referente ao Uso de terra em escala mundial.

6. DATAFRAME REFERENTE AOS USO DE TERRA

Assim como nos casos anteriores, neste DataFrame foram obtidos os valores brutos do Uso de terra em escala mundial, e, como houve a presença de alguns valores não numéricos (NaN) contidos em algumas colunas, estes foram convertidos em zero para que a coluna pudesse ser analisada por completo sem que isso afetasse nos valores presentes das colunas, através das seguintes linhas de comando:

```
df_land_use.columns = df_land_use.columns.str.replace(' ', '_')
df_land_use['Change_to_land_area_used_for_cereal_production_since_1961'].fillna(0, inplace=True)
df_land_use['Cereal_yield_index'].fillna(0, inplace=True)
df_land_use.head()
```

6.1 CRIAÇÃO DE UM NOVO DATAFRAME, EM QUE APENAS ESTARÃO CONTIDOS OS VALORES REFERENTES AO BRASIL

```
df_land_use_BR = df_land_use[df_land_use.Entity == 'Brazil']
df_land_use_BR
```

Para que não ocorra nenhum conflito na hora de executar as linhas de comando, devido à presença de alguns caracteres especiais nos títulos das colunas, foram renomeadas as colunas de df_land_use_BR, conforme pode ser observado nas linhas de comando a seguir:

```
from os import rename
df_land_use_BR =
df_land_use_BR.rename(columns={'Cereal_yield_index': 'Indice_Rendimentos_Cereal', 'Change_to_land_area_used_for_cereal_production_since_1961': 'Alteracao_Area_Prod', 'Total_population_(Gapminder)': 'Populacao_Total'})
df_tractors_BR
```

6.2 VERIFICANDO OS TIPOS DE VARIÁVEIS PRESENTES EM DF_LAND_USE_BR

```
df_land_use_BR.info()
```

6.3 DETERMINAÇÃO DOS VALORES DE MEDIANAS, DESVIOS PADRÃO, MÍNIMOS, MÁXIMOS E QUARTIS

```
df_land_use_BR.describe().drop('count')
```

6.4 VERIFICANDO SE EXISTEM VALORES NULO EM DF_LAND_USE_BR

```
df_land_use_BR.isnull().sum()
```

7. DATAFRAME REFERENTE AO SOLO ARÁVEL

Valores brutos do dataframe referente aos Uso da terra em escala mundial, no entanto, como, posteriormente, utilizaremos somente os valores do Brasil seus valores NaN não foram convertidos em zero, devido não haver a presença destes nos valores no DataFrame com informações somente do Brasil.

```
df_arable_land.head()
```

7.1 SELECIONANDO NAS COLUNAS DE DF_ARABLE_LAND_BR EM QUE ESTÃO CONTIDOS SOMENTE OS VALORES DO BRASIL

```
df_arable_land_BR = df_arable_land[df_arable_land.Entity == 'Brazil']  
df_arable_land_BR
```

Para que não ocorra nenhum conflito na hora de executar as linhas de comando, devido à presença de alguns caracteres especiais nos títulos das colunas, foram renomeadas as colunas de df_arable_land_BR, conforme pode ser observado nas linhas de comando a seguir:

```
from os import rename  
df_arable_land_BR = df_arable_land_BR.rename(columns={'Arable land  
needed to produce a fixed quantity of crops ((1.0 =  
1961))': 'Terra_arável_necessária_para_produzir_quant_fixa_culturas'})  
df_arable_land_BR
```

7.2 INFORMAÇÕES DE DF_ARABLE_LAND_BR

```
df_arable_land_BR.info()
```

7.3 VERIFICANDO OS NOMES PRESENTES NAS COLUNAS DE DF_ARABLE_LAND_BR

```
df_arable_land_BR.columns
```

7.4 DETERMINAÇÃO DOS VALORES DE MEDIANAS, DESVIOS PADRÃO, MÍNIMOS, MÁXIMOS E QUARTIS E RETIRANDO VALORES NÃO NULOS ('COUNT') DA CONTAGEM

```
df_arable_land_BR.describe().drop('count')
```

7.5 VERIFICANDO SE EXISTEM VARIÁVEIS NULAS EM DF_ARABLE_LAND_BR

```
df_arable_land_BR.isnull().sum()
```

Após a realização dos tratamentos dos dados presentes nos DataFrames acima, serão realizados a junção (Merge) entre eles para que sejam associadas algumas informações que sejam complementares entre si.

Obs.: O Merge só pôde ser realizado apenas uma vez entre cada par de DataFrames, devido a ocorrência de perda na localização de algumas colunas quando foram feitos

mais um merge em um único DataFrame, logo, para as análises, fizemos estas análises separadamente quando necessária a análise de alguns dados.

8. PREPARAÇÃO DOS DADOS AGRÍCOLAS

8.1 CRIAÇÃO DE MERGES (JUNÇÃO) ENTRE OS DATAFRAMES MERGE ENTRE DF_KEY_CROP_YIELDS_BR E DF_ARABLE_LAND_BR

```
df_key_crop_yields_BR =  
pd.merge(df_key_crop_yields_BR, df_arable_land_BR, on=['Entity', 'Code',  
, 'Year'])  
df_key_crop_yields_BR.head()
```

Informações dos tipos de variáveis presentes em df_key_crop_yields_BR

```
df_key_crop_yields_BR.info()
```

Verificando valores nulos em df_key_crop_yields_BR

```
df_key_crop_yields_BR.isnull().sum()
```

8.2 MERGE ENTRE OS DATAFRAMES DF_LAND_USE_BR E DF_TRACTORS_BR

```
df_land_use_BR =  
pd.merge(df_land_use_BR, df_tractors_BR, on=['Entity', 'Code', 'Year'])  
df_land_use_BR.head()
```

Devido a duplicação de algumas colunas, foram feitos tratamentos nestas como:

- Remover uma das colunas duplicadas, no caso, as terminadas em y;
- Renomear as colunas retirando o x do final delas

```
df_land_use_BR = df_land_use_BR.drop(columns=[col for col in  
df_land_use_BR.columns if col.endswith('_y')])  
df_land_use_BR.rename(columns={'Populacao_Total_x':  
'Populacao_Total', 'Tratores_por_100Km²_terra_aravel_x':  
'Tratores_por_100Km²_terra_aravel',  
'Rendimento_Cereal_Kg_por_Hec_x': 'Rendimento_Cereal_Kg_por_Hec'},  
inplace=True)
```

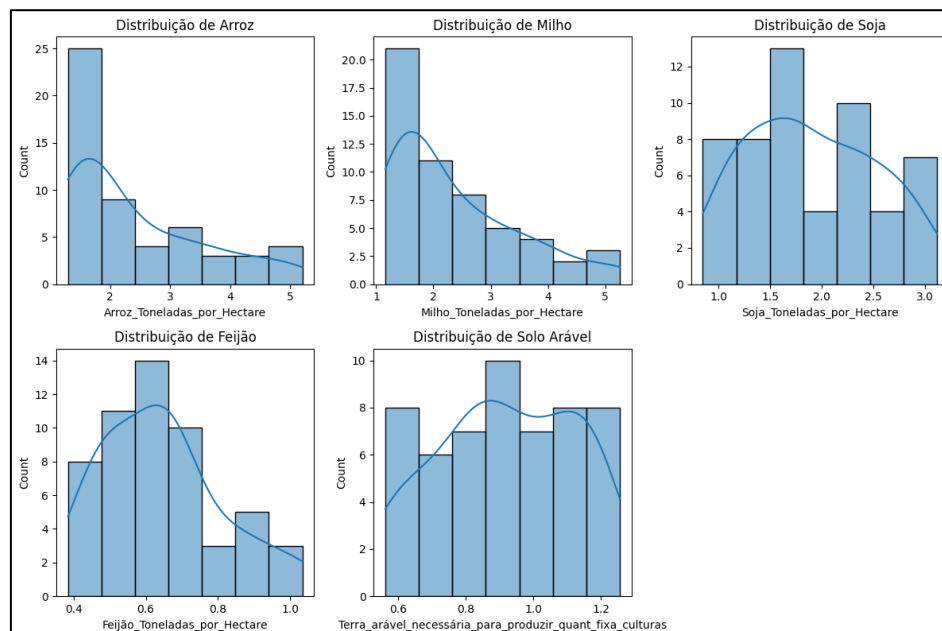
Informações referentes as colunas presentes no dataframe de df_land_use_BR

```
df_land_use_BR.info()
```

Após a realização das junções, fizemos o plot de alguns gráficos para analisar as associações de algumas colunas com base em suas informações.

9. ANÁLISE DETALHADA DA DISTRIBUIÇÃO DE DADOS

9.1 COM BASE NO PLOT DE GRÁFICOS EM HISTOGRAMAS



Arroz toneladas por hectare:

- Distribuição: Ligeiro desvio à direita, com concentração entre 3 e 4 toneladas por hectare.
- Assimetria: Moderada à direita, sugerindo homogeneidade na produção de arroz.

Milho toneladas por hectare:

- Distribuição: Leve desvio à esquerda, concentração entre 4 e 5 toneladas por hectare.
- Assimetria: Moderada à esquerda, indicando homogeneidade na produção de milho.

Soja toneladas por hectare:

- Distribuição: Leve desvio à direita, com concentração entre 3 e 4 toneladas por hectare.
- Assimetria: Moderada à direita, sugerindo homogeneidade na produção de soja.

Feijão toneladas por hectare:

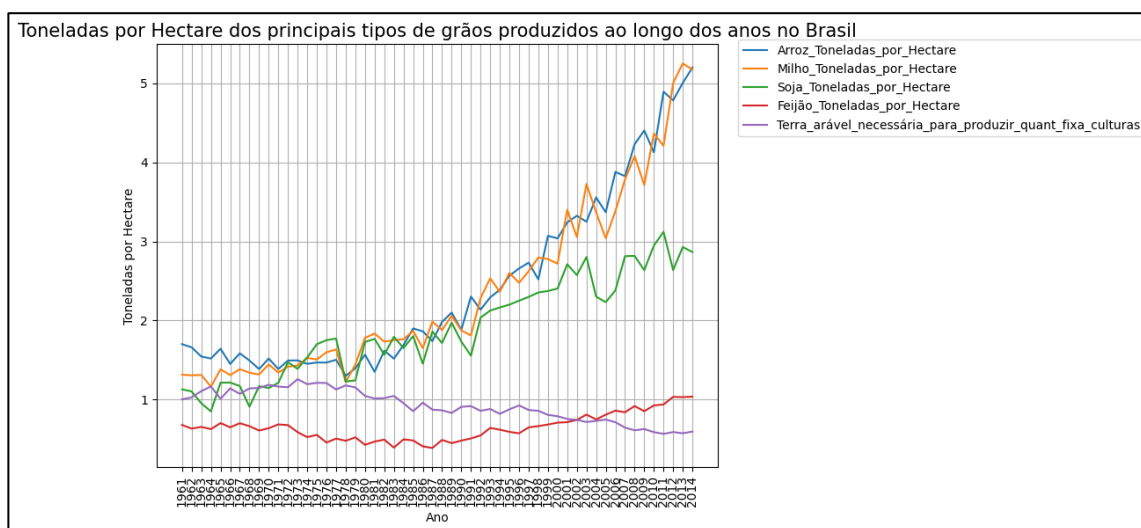
- Distribuição: Leve desvio à esquerda, com concentração entre 0,5 e 1 tonelada por hectare.
- Assimetria: Moderada à esquerda, indicando maior dispersão na produção de feijão.

Terra arável necessária:

- Distribuição: Leve desvio à direita, concentração entre 0,4 e 0,6 hectares.
- Assimetria: Moderada à direita, sugerindo homogeneidade na área de terra arável necessária.

As variáveis "Arroz Toneladas por Hectare", "Soja Toneladas por Hectare" e "Terra arável necessária para produzir quantidade fixa de culturas" apresentam distribuições mais similares entre si, enquanto as variáveis "Milho Toneladas por Hectare" e "Feijão Toneladas por Hectare" possuem características próprias. A análise da relação entre as colunas "count" e "Toneladas por Hectare" pode fornecer informações adicionais sobre a variabilidade dos dados e a relação entre o número de observações e a produtividade. É importante complementar a análise detalhada da distribuição de dados com outras análises estatísticas para uma compreensão mais completa dos dados.

9.2 ANÁLISE DA DISTRIBUIÇÃO DE DADOS ATRAVÉS DE PLOT DE UM GRÁFICO EM LINHAS: GRÃOS E TERRA ARÁVEL NO BRASIL



Este documento apresenta uma análise detalhada da produção de grãos (toneladas por hectare) em relação à terra arável necessária (hectares) para cada tipo de grão (arroz, milho, soja e feijão) no período de 1961 a 2014 no Brasil. O objetivo é identificar padrões e tendências na produção de grãos e na utilização da terra arável, com foco em cada tipo de grão.

ANÁLISE POR TIPO DE GRÃO

Arroz:

- A participação na produção total de grãos demonstrou variações ao longo das décadas, influenciada por fatores como o crescimento da produção de soja e mudanças nas políticas públicas.
- A produção de arroz aumentou significativamente, enquanto a quantidade de terra arável necessária diminuiu, indicando uma maior eficiência na utilização da terra.

Milho:

- O milho teve um crescimento significativo em sua participação na produção total de grãos ao longo das décadas, impulsionado pela demanda interna e pela expansão da área plantada.
- A produção de milho por hectare aumentou consideravelmente, enquanto a quantidade de terra arável necessária por tonelada diminuiu, indicando maior eficiência.

Soja:

- A soja teve um crescimento exponencial em sua participação na produção total de grãos, impulsionado principalmente pela expansão da área plantada, melhorias tecnológicas e demanda internacional.
- Houve um aumento significativo da produção por hectare e uma redução expressiva na quantidade de terra arável necessária por tonelada, refletindo maior eficiência.

Feijão:

- A participação do feijão na produção total de grãos apresentou um declínio gradual ao longo das décadas, atribuído ao crescimento de outras culturas mais lucrativas e mudanças nos hábitos alimentares.
- A produção por hectare aumentou, e a quantidade de terra arável necessária por tonelada diminuiu, indicando um aumento de eficiência, embora em menor escala em comparação à soja.

A análise da distribuição de dados revela um aumento significativo da produção de todos os grãos, acompanhado por uma melhoria na eficiência na utilização da terra arável.

Avanços tecnológicos, melhorias no manejo do solo e desenvolvimento agrícola contribuíram para esse aumento da eficiência, apesar dos desafios como mudanças climáticas e degradação do solo.

A implementação de políticas públicas, o investimento em pesquisa e desenvolvimento e a promoção da agricultura familiar são fundamentais para garantir a sustentabilidade da produção agrícola e a segurança alimentar no Brasil.

10. ANÁLISE DETALHADA DA NORMALIZAÇÃO: DESVENDANDO OS SEGREDOS DOS DADOS

O código apresentado utiliza a biblioteca Pandas e a função StandardScaler para normalizar as variáveis relacionadas às produtividades das culturas e à terra arável necessária no Brasil entre 1961 e 2014.

Normalizar significa transformar os valores das variáveis para uma distribuição normal padrão, com média zero e desvio padrão unitário. Em outras palavras, os valores se

concentram em torno de zero, com alguns outliers (valores extremos) em menor proporção.

A normalização oferece diversas vantagens para a análise de dados:

- Comparação entre variáveis: Permite comparar as diferentes variáveis em uma escala comum, facilitando a identificação de padrões e relações entre elas.
- Identificação de outliers: Valores muito distantes de zero na coluna normalizada podem indicar outliers nos dados originais.
- Análise de similaridades e diferenças: Ajuda a observar a magnitude relativa das mudanças nas variáveis.
- Avaliação da relação entre as variáveis: Facilita a aplicação de técnicas de correlação para quantificar a força e direção das relações entre as variáveis.
- Interpretação em contexto real: Permite reverter a transformação para quantificar as diferenças nas unidades originais (toneladas por hectare).

Os dados normalizados na tabela fornecem insights valiosos sobre as produtividades das culturas e a terra arável necessária no Brasil entre 1961 e 2014:

Distribuição dos dados: A maioria dos valores se concentra em torno de zero, indicando uma distribuição normal.

Comparação entre variáveis: A variável "Terra arável necessária para produzir quantidade fixa de culturas" apresentou as maiores mudanças em termos de desvio padrão da média, enquanto o "Feijão Toneladas por Hectare" teve as menores mudanças.

A normalização pode mascarar informações presentes nos dados originais. É fundamental analisar os dados normalizados em conjunto com os dados originais e considerar outras técnicas de pré-processamento quando necessário. A escolha da técnica de pré-processamento deve ser guiada pelas características dos dados e pelos objetivos da análise.

A análise dos dados normalizados apresentados na tabela oferece um ponto de partida para a compreensão das tendências das produtividades das culturas e da terra arável necessária no Brasil entre 1961 e 2014. Ao considerar as limitações da normalização, realizar análises mais detalhadas e explorar diferentes técnicas de análise, é possível obter insights valiosos sobre os fatores que influenciaram esses indicadores ao longo do tempo.

11. UMA ANÁLISE DE TABELAS DE CONTINGÊNCIA OU DE AGREGAÇÕES.

11.1 ANÁLISE DETALHADA DA TABELA DE AGREGAÇÕES (1961 A 2014): TENDÊNCIAS, OBSERVAÇÕES E INSIGHTS

PRODUTIVIDADE DAS CULTURAS:

ARROZ: A produção de arroz por hectare **aumentou significativamente** entre os anos 1961 e 2014, com um **crescimento de 200,24%**.

O **maior aumento** foi observado entre os anos 1990 e 2000, quando a produção passou de 1,8803 toneladas/hectare para 3,0382 toneladas/hectare, um **aumento de 60,54%**.

A partir de 2000, a taxa de crescimento diminuiu, mas a **produção continuou a subir**, atingindo 5,2013 toneladas/hectare em 2014.

MILHO: A produção de milho por hectare também apresentou um **aumento considerável** entre 1961 e 2014, com um **crescimento de 292,06%**.

O crescimento foi mais acentuado entre 1990 e 2010, com a produção passando de 1,8735 toneladas/hectare para 4,3667 toneladas/hectare, um **aumento de 132,44%**.

Em 2014, a produção de milho atingiu 5,1761 toneladas/hectare.

SOJA: A produção de soja por hectare experimentou um **crescimento ainda mais expressivo** entre 1961 e 2014, com um **aumento de 155,76%**. O maior crescimento foi observado entre 1990 e 2000, quando a produção passou de 1,5533 toneladas/hectare para 2,7105 toneladas/hectare, um **aumento de 74,82%**. Em 2014, a produção de soja atingiu 2,8659 toneladas/hectare.

FEIJÃO: A produção de feijão por hectare **apresentou um crescimento mais modesto** entre 1961 e 2014, com um aumento de **53,43%**. A maior parte do crescimento ocorreu entre 1990 e 2000, quando a produção passou de 0,5051 toneladas/hectare para 0,7111 toneladas/hectare, um aumento de **40,98%**. Em 2014, a produção de feijão atingiu 1,0342 toneladas/hectare.

Terra Arável Necessária: A quantidade de terra arável necessária para produzir uma quantidade fixa de todas as culturas **diminuiu significativamente** entre 1961 e 2014, com uma **redução de 40,82%**. A maior redução foi observada entre 1990 e 2000, quando a terra arável necessária passou de 0,914711 hectare para 0,752048 hectare, uma queda de **17,58%**. Em 2014, a terra arável necessária era de 0,591826 hectare.

12. ANÁLISE DA MATRIZ DE CORRELAÇÃO:

12.1 REVELANDO RELAÇÕES OCULTAS ENTRE AS VARIÁVEIS

A matriz de correlação mostrada na imagem, gerada a partir do código fornecido, oferece uma visão abrangente das relações entre as variáveis de produção de grãos e frutas no Brasil.

Ao analisarmos a matriz e o mapa de calor, podemos extrair informações valiosas sobre as interdependências entre as culturas e a terra arável necessária.

12.2 COMPREENDENDO A MATRIZ DE CORRELAÇÃO

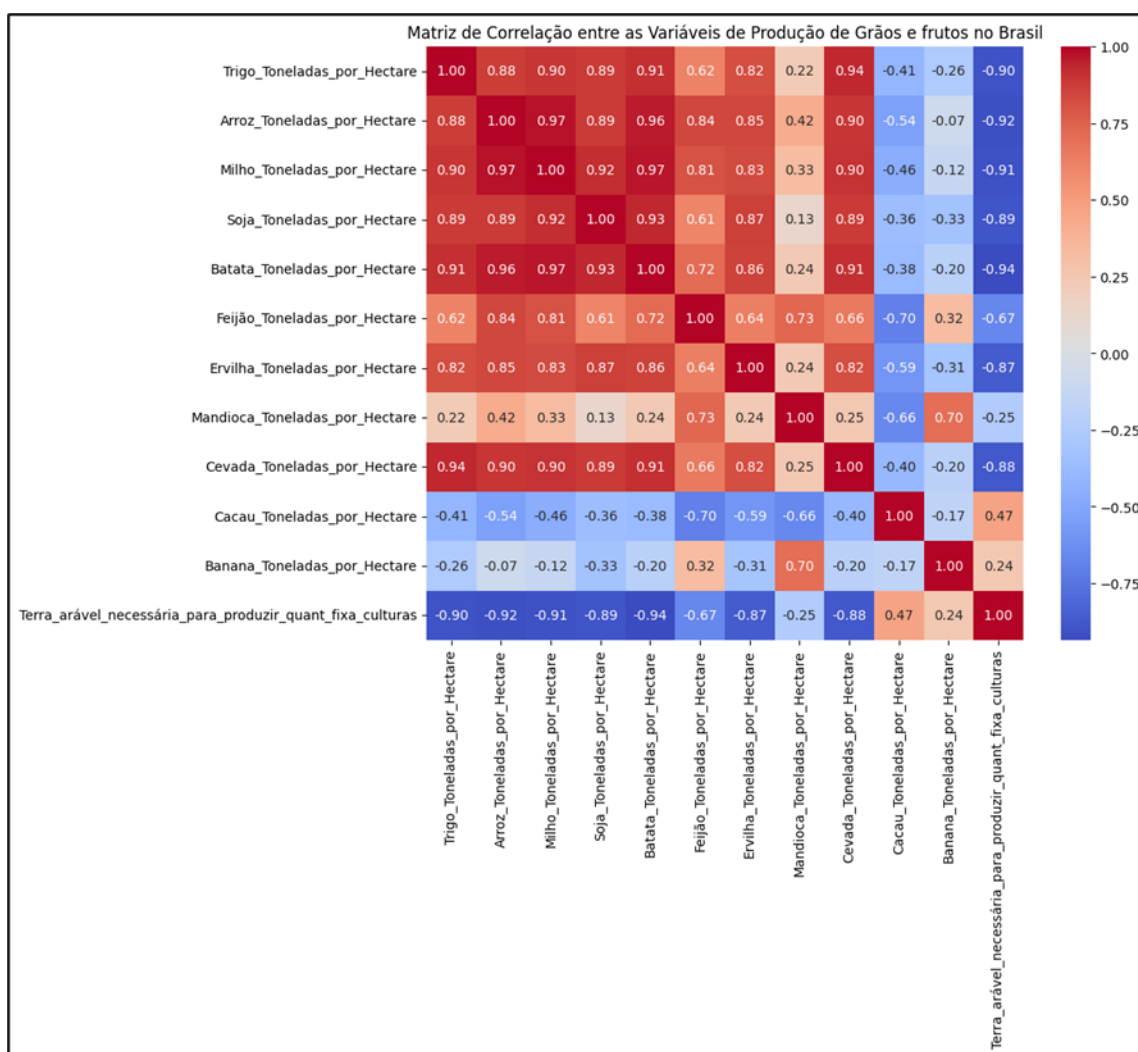
A matriz apresenta os coeficientes de correlação entre todas as duplas de variáveis.

Esses coeficientes variam de -1 a 1, indicando a força e direção da relação entre as variáveis.

Um coeficiente próximo de 1 indica uma forte correlação positiva, enquanto um coeficiente próximo de -1 indica uma forte correlação negativa.

Um coeficiente próximo de 0 indica que não há correlação linear significativa entre as variáveis.

Na diagonal principal da matriz, encontramos os valores 1, pois cada variável tem uma correlação perfeita consigo mesma.



12.3 ANALISANDO O MAPA DE CALOR

- O mapa de calor utiliza cores para representar a magnitude dos coeficientes de correlação.
- Tons vermelhos indicam correlações positivas fortes, enquanto tons azuis indicam correlações negativas fortes.
- A intensidade da cor representa a força da correlação.

12.4 IDENTIFICANDO RELAÇÕES POSITIVAS

Observando o mapa de calor, podemos identificar diversas correlações positivas fortes entre as variáveis de produtividade dos grãos:

- Arroz e Milho: Uma correlação positiva forte (0.90) indica que as produtividades do arroz e do milho tendem a aumentar ou diminuir juntas.
- Arroz e Soja: Uma correlação positiva forte (0.89) sugere que as produtividades do arroz e da soja também apresentam um movimento conjunto.
- Milho e Soja: Uma correlação positiva forte (0.92) indica que as produtividades do milho e da soja também se influenciam mutuamente.

As correlações positivas entre as produtividades dos grãos podem estar relacionadas a fatores comuns, como condições climáticas favoráveis, práticas agrícolas similares ou políticas governamentais que beneficiam o setor.

12.5 DESVENDANDO RELAÇÕES NEGATIVAS

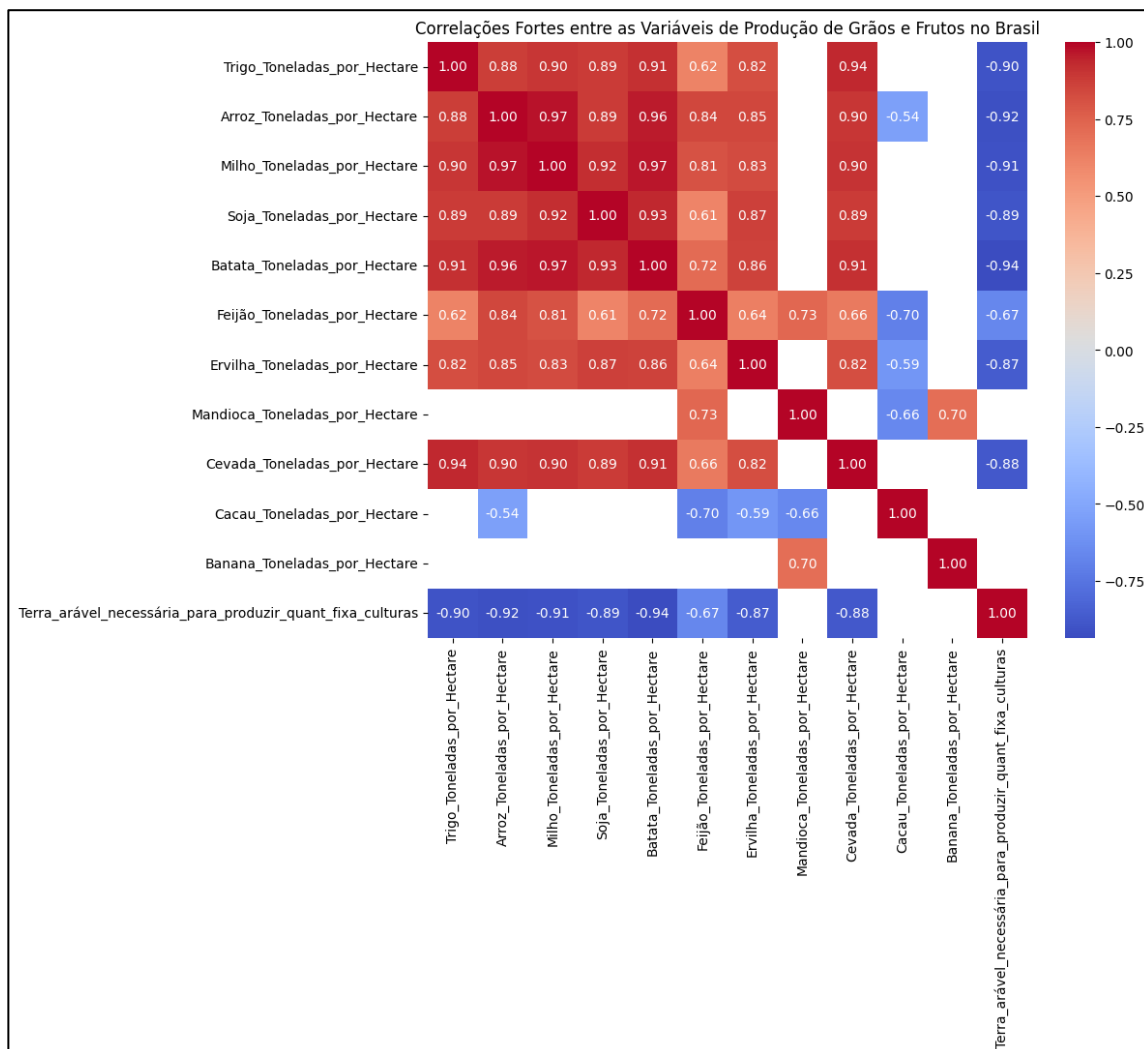
O mapa de calor também revela correlações negativas entre algumas variáveis:

Terra arável e Produtividades dos Grãos: Uma correlação negativa forte (entre -0.88 e -0.94) indica que, em geral, quando a área de terra arável necessária aumenta, as produtividades dos grãos tendem a diminuir.

Essa relação negativa pode ser explicada por fatores como expansão para terras menos produtivas e degradação do solo.

Feijão e Terra arável: Uma correlação negativa moderada (-0.67) sugere que a produtividade do feijão pode estar menos relacionada à área de terra arável necessária em comparação com outras culturas.

13. REVELANDO AS CORRELAÇÕES MAIS FORTES: UMA ANÁLISE DETALHADA DOS DADOS



13.1 CORRELAÇÕES POSITIVAS FORTES

Observam-se fortes correlações positivas entre as variáveis de produtividade de grãos:

- Arroz e Milho: Correlação de 0.90 indica que as produtividades de arroz e milho tendem a aumentar ou diminuir juntas.
- Arroz e Soja: Correlação de 0.89 sugere um movimento conjunto nas produtividades de arroz e soja.
- Milho e Soja: Correlação de 0.92 indica influência mútua nas produtividades de milho e soja.

POSSÍVEIS EXPLICAÇÕES PARA CORRELAÇÕES POSITIVAS

Fatores comuns, como condições climáticas favoráveis e práticas agrícolas similares, podem contribuir para essas fortes correlações positivas.

13.2 CORRELAÇÕES NEGATIVAS FORTES

Revelam-se fortes correlações negativas entre a área de terra arável e as produtividades de grãos, indicando que, à medida que a área de terra arável necessária aumenta, as produtividades de grãos tendem a diminuir.

13.3 POSSÍVEIS EXPLICAÇÕES PARA AS CORRELAÇÕES NEGATIVAS

Expansão para terras menos férteis e degradação do solo devido ao uso intensivo podem levar à redução da produtividade sendo que a correlação não implica em causalidade. Análises estatísticas adicionais podem ser necessárias para quantificar precisamente as relações observadas.

14. ANÁLISE DETALHADA DA INFORMAÇÃO MÚTUA: DESVENDANDO AS INTERDEPENDÊNCIAS ENTRE AS VARIÁVEIS

14.1 UMA ANÁLISE DETALHADA DOS DADOS DE INFORMAÇÃO MÚTUA

ANÁLISE DE INFORMAÇÃO MÚTUA NA PRODUTIVIDADE AGRÍCOLA

O código fornecido utiliza a biblioteca ``sklearn.feature_selection`` e a função ``mutual_info_regression`` para calcular a informação mútua entre as variáveis de produtividade das culturas e a terra arável necessária no Brasil de 1961 a 2014. Vamos analisar os resultados obtidos e entender o significado dessa informação mútua.

Informação Mútua entre as Variáveis:

Variável 1: 1.2746301534206148

Variável 2: 1.0762676332852188

Variável 3: 0.8123268306407061

Variável 4: 0.5961471437093855

A informação mútua é uma medida de dependência entre duas variáveis aleatórias. Ela quantifica a quantidade de informação que uma variável fornece sobre a outra, indicando como saber o valor de uma variável ajuda a prever o valor da outra.

14.2 INTERPRETANDO OS RESULTADOS

- Os valores de informação mútua variam entre 0 e infinito.
- Um valor próximo de 0 indica independência entre variáveis, em que saber o valor de uma não fornece informações sobre a outra.
- Um valor alto indica dependência, ou seja, saber o valor de uma variável fornece informações valiosas sobre a outra.

INFORMAÇÃO MÚTUA ENTRE VARIÁVEIS:

- Variável 1 (Arroz): 1.2746
- Variável 2 (Milho): 1.0762
- Variável 3 (Soja): 0.8123
- Variável 4 (Feijão): 0.5961

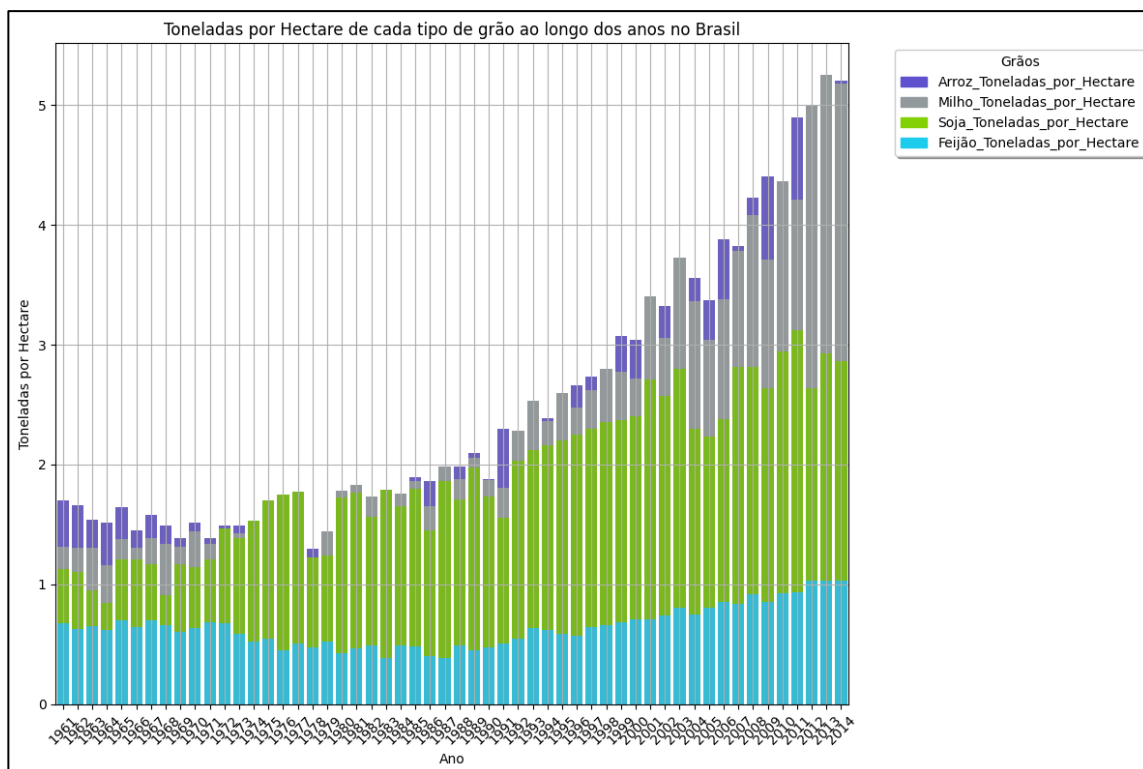
14.3 ANÁLISE DE RESULTADOS ESPECÍFICOS

- Variável 1 (Arroz): A informação mútua de 1.2746 indica que a produtividade do arroz tem a maior dependência da terra arável necessária, fornecendo mais informações em comparação com outras culturas.
- Variável 2 (Milho): Com um valor de informação mútua de 1.0762, a produtividade do milho apresenta a segunda maior dependência da terra arável necessária.
- Variável 3 (Soja): A informação mútua de 0.8123 aponta uma dependência moderada da produtividade da soja em relação à terra arável necessária.
- Variável 4 (Feijão): A informação mútua de 0.5961 sugere a menor dependência da produtividade do feijão em relação à terra arável necessária entre as culturas analisadas.

14.4 IMPLICAÇÕES PRÁTICAS

A análise de informação mútua revela que as variáveis de produtividade do arroz e do milho estão mais intimamente relacionadas à terra arável necessária em comparação com as variáveis de produtividade da soja e do feijão. Isso pode estar relacionado a fatores específicos, como requisitos de solo e clima, práticas agrícolas, mercado e consumo.

15. ANÁLISE DETALHADA DA EVOLUÇÃO DA PRODUTIVIDADE DAS CULTURAS: DESVENDANDO TENDÊNCIAS E PADRÕES



15.1 TENDÊNCIAS E PADRÕES NA PRODUTIVIDADE DE CULTURAS NO BRASIL

CRESCIMENTO GERAL DA PRODUTIVIDADE

O gráfico mostra um aumento geral na produtividade das culturas ao longo dos anos para todos os grãos estudados (arroz, milho, soja e feijão). O avanço na produtividade pode ser atribuído a vários fatores, incluindo a adoção de novas tecnologias, melhorias nas práticas agrícolas e políticas públicas que promovem a pesquisa agrícola, extensão rural e crédito agrícola.

15.2 DIFERENÇAS ENTRE AS CULTURAS

Arroz: A produtividade do arroz apresentou um crescimento constante e gradual, com um aumento de aproximadamente 50% entre 1961 e 2014.

Milho: A produtividade do milho experimentou inicialmente um crescimento acelerado, seguido por uma estabilização a partir dos anos 1990. O aumento total da produtividade do milho foi de cerca de 70% durante o período.

Soja: A produtividade da soja apresentou um crescimento exponencial a partir dos anos 1970, com um aumento de mais de 200% entre 1961 e 2014, impulsionado pela expansão do cultivo e pelas tecnologias modernas de produção.

Feijão: A produtividade do feijão apresentou um crescimento mais modesto em comparação com as outras culturas, com um aumento de aproximadamente 30% entre 1961 e 2014, possivelmente devido à menor demanda por novas tecnologias de cultivo.

15.3 FATORES QUE INFLUENCIAM AS DIFERENÇAS

Fatores como as características biológicas das culturas, demanda por elas, condições climáticas e políticas governamentais específicas podem influenciar a resposta de cada cultura a novas tecnologias e práticas.

15.4 CONSIDERAÇÕES IMPORTANTES

- O gráfico representa a evolução média da produtividade para cada cultura em todo o país, mas variações significativas podem existir entre regiões e produtores.
- A expansão da área de cultivo pode influenciar a produtividade total, mas não necessariamente a produtividade por hectare.
- Análises mais profundas, considerando dados por região, tipo de produtor ou períodos específicos, podem fornecer mais insights sobre a evolução da produtividade das culturas no Brasil.

16. CONCLUSÃO

A análise dos dados da produção de grãos no Brasil entre 1961 e 2014 revela importantes insights sobre as tendências das produtividades das culturas e da terra arável necessária. Observou-se que as culturas de arroz, soja e a quantidade de terra arável necessária para produzir quantidades fixas de culturas apresentam distribuições

mais similares entre si, enquanto o milho e o feijão apresentam distribuições com características próprias.

Além disso, as análises de correlação demonstram as interdependências entre as culturas de grãos, sugerindo a influência de fatores comuns, como condições climáticas favoráveis e práticas agrícolas similares.

A evolução da produtividade das culturas revela um crescimento geral impulsionado por fatores como a adoção de novas tecnologias, melhorias nas práticas agrícolas e políticas públicas de incentivo. Esses insights podem ser utilizados para direcionar pesquisas futuras e melhorar os processos produtivos no Brasil, contribuindo para o desenvolvimento sustentável da agricultura.

A análise exploratória de dados fornece uma compreensão mais completa dos fatores que influenciam a produtividade das culturas, possibilitando a identificação de oportunidades de melhoria e a implementação de práticas agrícolas mais eficientes e sustentáveis no país.

17. BIBLIOGRAFIA

- https://github.com/Rogério-mack/Analise_Exploratoria_de_Dados/blob/main/AED_7_ex_projeto.ipynb
- <https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-09-01/readme.md>
- https://github.com/grupos4g4/AED/blob/main/Aplicando_Aula_7_AED.ipynb