


```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly.express as px
import seaborn as sns
import json
import requests

from sklearn.linear_model import LinearRegression
```

```
df = pd.read_csv('/content/dados_esgoto.csv')
df_limpo = df[df["populacao_urbana_atendida_agua_ibge"].notna()]
df
```



	ano	id_municipio	sigla_uf	populacao_atendida_agua	populacao_atendida_esgoto	populacao_urbana	populacao_urbana_residente_agua	populacao_urbana_atendida_agua	populacao_urbana_atendida_agua_ibge	populacao_urbana_residente_esgoto	...	investimento_recurso_onei
0	2018	2705606	AL	5276.0	3000.0	7665.0	7665.0	5276.0	12615.0	7665.0	...	
1	2018	2707107	AL	18502.0	8292.0	14245.0	14245.0	14238.0	24891.0	14245.0	...	
2	2021	2709152	AL	39000.0	28000.0	37674.0	37674.0	37500.0	44570.0	37674.0	...	
3	2017	2709301	AL	51589.0	45617.0	50799.0	50799.0	50740.0	66477.0	50799.0	...	
4	2021	2902104	BA	35540.0	20700.0	20874.0	20874.0	20820.0	54903.0	20874.0	...	
...	
119251	2022	1300805	AM	30805.0	NaN	NaN	NaN	NaN	33080.0	NaN	...	
119252	2022	2102077	MA	4678.0	NaN	NaN	NaN	NaN	12212.0	NaN	...	
119253	2022	1506104	PA	10000.0	NaN	NaN	NaN	NaN	10851.0	NaN	...	
119254	2022	2500577	PB	0.0	NaN	NaN	NaN	NaN	2953.0	NaN	...	
119255	2022	4311734	RS	2800.0	NaN	NaN	NaN	NaN	3131.0	NaN	...	

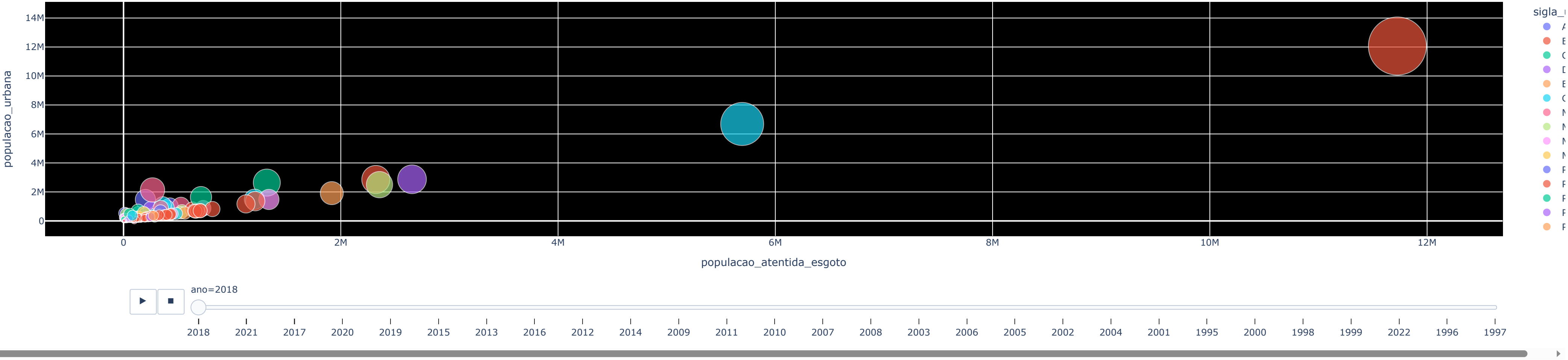
119256 rows × 133 columns

```
fig = px.scatter(
    df_limpo,
    x="populacao_atendida_esgoto", # Changed 'populacao_atendida_esgoto' to 'populacao_atendida_esgoto'
    y="populacao_urbana",
    size="populacao_urbana_atendida_agua_ibge",
    color="sigla_uf",
    hover_data=["populacao_atendida_agua", "populacao_urbana_atendida_agua", "populacao_urbana_residente_esgoto"],
    animation_frame="ano",
    size_max=55, # você pode ajustar esse valor,

)

fig.update_layout(
    plot_bgcolor = "Black",
)

fig.show()
```



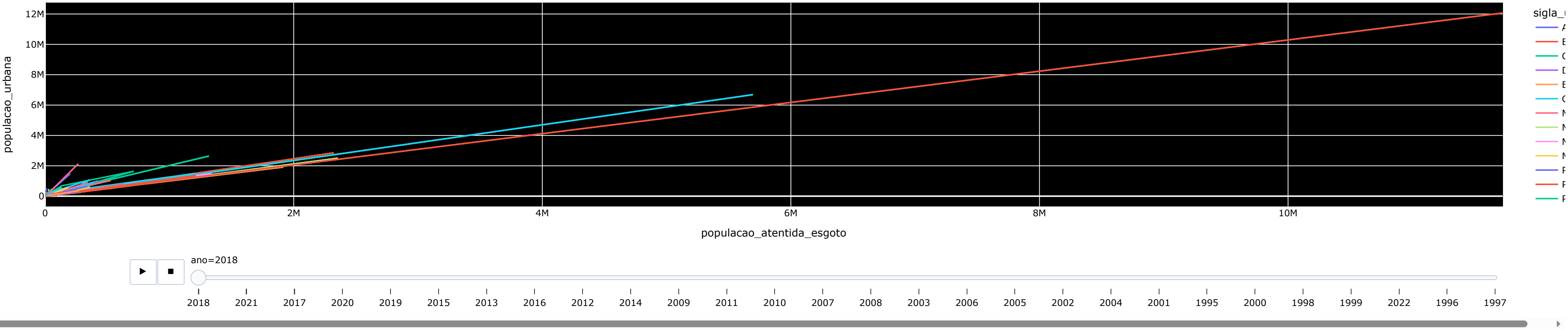
```
fig = px.line(  
    df_limpo,  
    x="populacao_atendida_esgoto", # Changed 'populacao_atendida_esgoto' to 'populacao_atendida_esgoto'  
    y="populacao_urbana",  
    color="sigla_uf",  
    animation_frame="ano",  
    title = "Gráfico de progressão: População Urbana X População Com Saneamento Básico"  
)
```

```
fig.update_layout(  
    plot_bgcolor = "Black",  
    font_color = "Black"  
)
```

```
fig.show()
```



Gráfico de progressão: População Urbana X População Com Saneamento Básico




Usaremos o próximo gráfico de regressão linear para estreitar uma relação entre o crescimento da população e o acesso a água.

```
df_limpo = df[["populacao_urbana", "populacao_urbana_atendida_agua"]].dropna()
```

```
x_1 = pd.DataFrame(df_limpo["populacao_urbana"])
y_1 = pd.DataFrame(df_limpo["populacao_urbana_atendida_agua"])
```

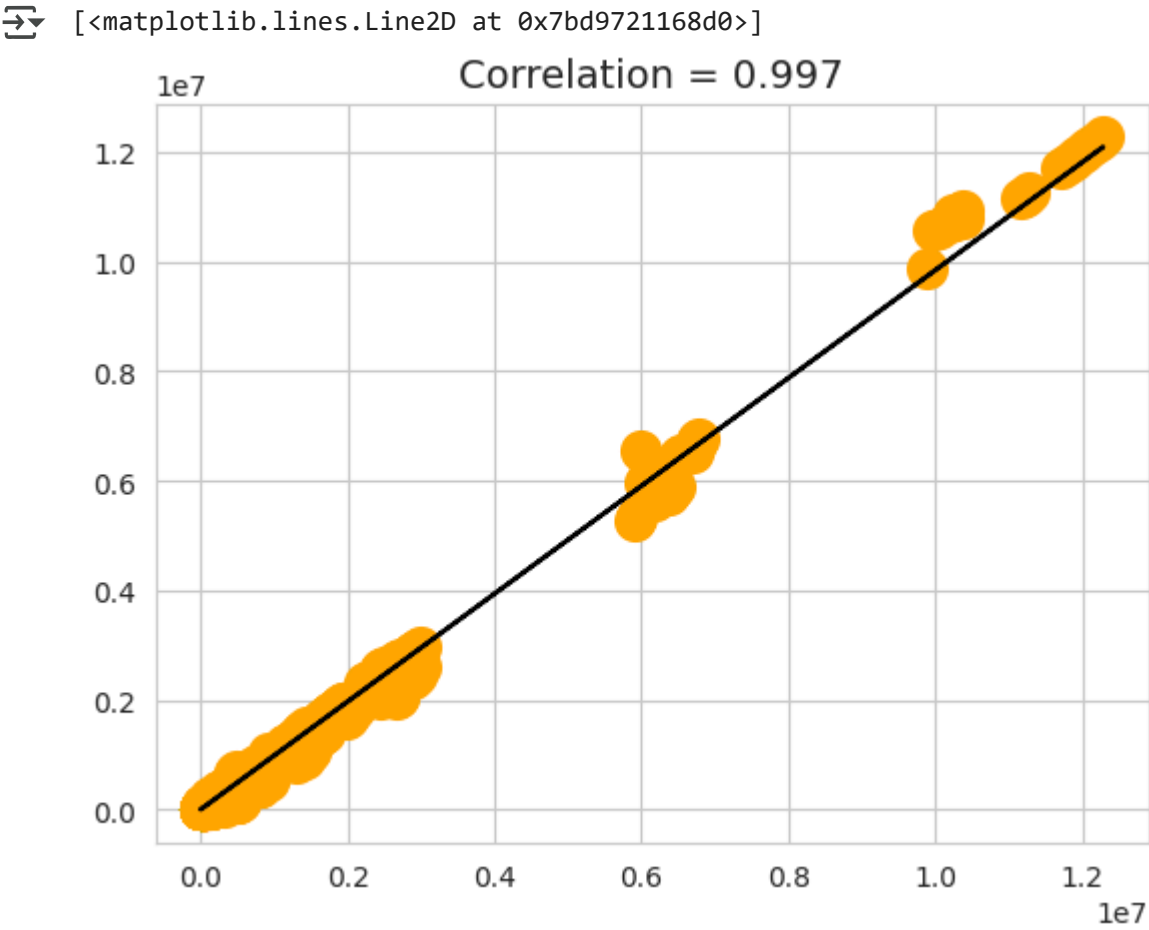
```
rgr_1 = LinearRegression()
rgr_1.fit(x_1, y_1)
```



▼ LinearRegression ⓘ ?

LinearRegression()

```
COLOR_1 = "orange"
COLOR_2 = "black"
sns.set_style("whitegrid")
plt.figure()
# Accessing the 'populacao_urbana' column from x_1 to ensure consistent shapes
plt.title(f'Correlation = {round(x_1["populacao_urbana"].corr(df_limpo["populacao_urbana_atendida_agua"]),3)}', fontsize = 14)
# Using x_1 and the corresponding predictions for plotting
plt.scatter(x_1["populacao_urbana"], df_limpo["populacao_urbana_atendida_agua"], color=COLOR_1,s=200)
plt.plot(x_1["populacao_urbana"], rgr_1.predict(x_1), color=COLOR_2)
```



Temos pela regressão que a correlação GERAL dos periodos analisados pela planilha é que a quantidade de pessoas com acesso a água é equivalente a **99,7%** da população urbana., devido ao preenchimento, interpolação ou outros critérios de limpeza, a porcentagem sobe um pouco, sendo considerado uma média de **98,75** de pessoas com acesso a água em relação a população total.

Faremos um gráfico para representação do crescimento da população com acesso á esgoto. Iremos usar uma aplicação demonstrando a ampliação das redes de esgoto nos estados do Brasil.


```
url = 'https://raw.githubusercontent.com/codeforamerica/click_that_hood/master/public/data/brazil-states.geojson'
geojson = requests.get(url).json()
```

```
print(geojson['features'][0]['properties'])
```

```
df_limpo = df[['sigla_uf', 'populacao_urbana_atendida_esgoto', 'ano',
              'populacao_urbana', 'populacao_urbana_atendida_agua']].dropna()
```

```
fig = px.choropleth(
    df_limpo,
    geojson=geojson,
    locations="sigla_uf",
    featureidkey="properties.sigla",
    color="populacao_urbana_atendida_esgoto",
```

```
        animation_frame="ano",
        color_continuous_scale="Viridis",
        scope="south america",
        range_color=(0, 800_000)
    )
fig.show()
```

 {'id': 1, 'name': 'Acre', 'sigla': 'AC', 'regiao_id': '3', 'codigo_ibg': '12', 'cartodb_id': 1, 'created_at': '2015-02-09T16:46:01Z', 'updated_at': '2015-02-09T16:46:01Z'}



E com esse gráfico é possível observar a progressão da população urbana atendida pelas redes de esgoto ao decorrer do tempo