```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# 1. Carregar os dados e transformar para formato longo
file_path =
'/content/drive/MyDrive/Python/TrensRJ/Dados_Supervia.xlsx'
df = pd.read_excel(file_path, sheet_name='Supervia2019_2024')

# Transformar para formato longo
df_long = df.melt(id_vars=['Estação', 'Ramal'], var_name='Data',
value_name='Valor')
df_long['Data'] = pd.to_datetime(df_long['Data'], errors='coerce')

# 2. Converter coluna 'Valor' para numérica
df_long['Valor'] = pd.to_numeric(df_long['Valor'], errors='coerce')

# 3. Remover valores ausentes
df_long = df_long.dropna(subset=['Valor'])

# 4. Filtrar período desejado
start_date = '2019-01-01'
end_date = '2024-03-31'
df_filtered = df_long[(df_long['Data'] >= start_date) &
(df_long['Data'] <= end_date)].copy()

# 5. Salvar os dados filtrados em Parquet
parquet_path =
'/content/drive/MyDrive/Python/TrensRJ/supervia_filtrado_2019_2024.
parquet'
df_filtered.to_parquet(parquet_path)
print(f"Dados filtrados salvos em: {parquet_path}")

# 6. Visualização com Boxplot por ramal
ramais = df_filtered['Ramal'].unique()
for ramal in ramais:
    plt.figure(figsize=(12, 6))

    # Filtrar dados do ramal
    data_ramal = df_filtered[df_filtered['Ramal'] == ramal]

    # Remover outliers (opcional)
    q1 = data_ramal['Valor'].quantile(0.25)
    q3 = data_ramal['Valor'].quantile(0.75)
    iqr = q3 - q1
    lower_bound = q1 - 1.5 * iqr
    upper_bound = q3 + 1.5 * iqr
    data_no_outliers = data_ramal[(data_ramal['Valor'] >=
lower_bound) & (data_ramal['Valor'] <= upper_bound)]
```

```python
# Boxplot
sns.boxplot(x='Data', y='Valor', data=data_no_outliers)
plt.title(f'Boxplot do Ramal: {ramal} (jan/19 - mar/24)')
plt.xlabel('Data')
plt.ylabel('Valores')
plt.xticks(rotation=45)
plt.grid(True)
plt.ylim(lower_bound, upper_bound)
plt.tight_layout()
plt.show()
```