

modelagem_IA2A

November 17, 2025

```
[ ]: # =====
# PROJETO TCC - ETAPA DE MODELAGEM (RMB 2018-2025)
# =====

# --- 1. Importação de bibliotecas necessárias ---
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import files # Importa o módulo files

from sklearn.model_selection import train_test_split, KFold, cross_val_score
from sklearn.linear_model import LinearRegression, Lasso
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from sklearn.inspection import permutation_importance

# --- 2. Carregamento do dataset Gold ---
uploaded = files.upload()
file_name = list(uploaded.keys())[0]
df = pd.read_csv(file_name)

print(" Dados carregados com sucesso!")
print("Dimensão da base:", df.shape)
print("\nVisualização inicial:")
display(df.head())

# --- 3. Análise exploratória simples ---
print("\nResumo estatístico:")
display(df.describe())

print("\nVerificando valores ausentes:")
print(df.isna().sum().sort_values(ascending=False).head(10))

# --- 4. Seleção de variáveis ---
# Define variável alvo (target) e variáveis explicativas (features)
target = 'internacoes_total_10k'
```

```

# Remove colunas não numéricas e identificadores
drop_cols = ['municipio', 'ano', 'cod_mun']
X = df.drop(columns=[target] + [c for c in drop_cols if c in df.columns], errors='ignore')
y = df[target]

# Remove possíveis colunas não numéricas
X = X.select_dtypes(include=[np.number])

# Remove colunas com mais de 80% de valores ausentes
threshold = 0.8 * len(X)
X = X.dropna(axis=1, thresh=threshold)

# Remove linhas com valores ausentes em X ou y após remover colunas
data = pd.concat([X, y], axis=1).dropna()
X = data.drop(columns=[target])
y = data[target]

print("\nNúmero de variáveis explicativas após remoção de NaNs:", X.shape[1])
print("Número de amostras após remoção de NaNs:", X.shape[0])

# --- 5. Divisão treino/teste ---
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# --- 6. Função auxiliar para avaliação ---
def avaliar_modelo(modelo, nome):
    modelo.fit(X_train, y_train)
    y_pred = modelo.predict(X_test)

    mae = mean_absolute_error(y_test, y_pred)
    rmse = np.sqrt(mean_squared_error(y_test, y_pred))
    r2 = r2_score(y_test, y_pred)

    print(f"\n Modelo: {nome}")
    print(f"MAE : {mae:.3f}")
    print(f"RMSE: {rmse:.3f}")
    print(f"R² : {r2:.3f}")
    return mae, rmse, r2

# --- 7. Treinamento dos modelos ---
resultados = {}

# Regressão Linear

```

```

lr = LinearRegression()
resultados['LinearRegression'] = avaliar_modelo(lr, "Regressão Linear")

# Lasso
lasso = Lasso(alpha=0.1)
resultados['Lasso'] = avaliar_modelo(lasso, "Lasso")

# Random Forest
rf = RandomForestRegressor(n_estimators=200, random_state=42)
resultados['RandomForest'] = avaliar_modelo(rf, "Random Forest")

# --- 8. Comparação das métricas ---
res_df = pd.DataFrame(resultados, index=['MAE', 'RMSE', 'R²']).T
print("\n Comparativo de desempenho:")
display(res_df)

# --- 9. Importância das variáveis ---
# Para o modelo Random Forest
importancias = pd.Series(rf.feature_importances_, index=X.columns).
    ↪sort_values(ascending=False)

plt.figure(figsize=(10,6))
sns.barplot(x=importancias.head(15), y=importancias.head(15).index,
    ↪color='steelblue')
plt.title("Top 15 variáveis mais importantes - Random Forest")
plt.xlabel("Importância relativa")
plt.ylabel("Variável")
plt.tight_layout()
plt.show()

# --- 10. Explicabilidade (Permutation Importance) ---
perm_importance = permutation_importance(rf, X_test, y_test, n_repeats=10,
    ↪random_state=42)
sorted_idx = perm_importance.importances_mean.argsort()[:-1]

plt.figure(figsize=(10,6))
plt.barh(X.columns[sorted_idx][:-15], perm_importance.
    ↪importances_mean[sorted_idx][:-15])
plt.gca().invert_yaxis()
plt.title("Importância por permutação - Top 15 variáveis")
plt.xlabel("Impacto na métrica R²")
plt.tight_layout()
plt.show()

# --- 11. Salvando resultados ---
res_df.to_csv("/content/resultados_modelagem.csv", index=True)
print("\n Resultados salvos em: resultados_modelagem.csv")

```

```
print("\n Etapa de modelagem concluída com sucesso!")
```

```
<IPython.core.display.HTML object>

Saving gold_features_ano.csv to gold_features_ano.csv
Dados carregados com sucesso!
Dimensão da base: (64, 56)
```

Visualização inicial:

```
cod_mun      municipio    ano  populacao  idx_atend_agua_total  \
0  1500800  ANANINDEUA  2018   525566.0                  NaN
1  1500800  ANANINDEUA  2019   530598.0                  NaN
2  1500800  ANANINDEUA  2020   535547.0                  NaN
3  1500800  ANANINDEUA  2021   540410.0                  NaN
4  1500800  ANANINDEUA  2022       NaN                  NaN

  idx_atend_agua_urbano  idx_coleta_esgoto  idx_tratamento_esgoto  \
0                 NaN                  NaN                  NaN
1                 NaN                  NaN                  NaN
2                 NaN                  NaN          0.18
3                 NaN                  NaN         37.44
4                 NaN                  NaN         37.53

  idx_hidrometracao  idx_perdas_distribuicao ... internacoes_total  \
0                NaN                  NaN  ...           33625
1                NaN                  NaN  ...           31742
2                NaN                  NaN  ...           28648
3                NaN                  NaN  ...           24784
4                NaN                  NaN  ...           23719

internacoes_hidricas  dias_perm_total  dias_perm_hidricas  valor_total  \
0              1907        145459.0        7344.0  33358325.69
1              1869        134367.0        6505.0  30968523.41
2              1219        124735.0        3632.0  35739684.11
3              213         118372.0        1008.0  39538225.46
4               88         118569.0        567.0   39168060.98

  valor_hidricas  internacoes_total_10k  internacoes_hidricas_10k  \
0     687622.26        639.786440        36.284691
1     672328.37        598.230676        35.224407
2     470035.90        534.929707        22.761774
3     252334.72        458.614755        3.941452
4    128402.19             NaN                  NaN

  pct_internacoes_hidricas  valor_medio_internacao
0            5.671375        992.069166
1            5.888098        975.632393
```

2	4.255096	1247.545522
3	0.859425	1595.312519
4	0.371011	1651.336944

[5 rows x 56 columns]

Resumo estatístico:

	cod_mun	ano	populacao	idx_atend_agua_total	\
count	6.400000e+01	64.000000	4.800000e+01	14.000000	
mean	1.503085e+06	2021.500000	3.263658e+05	77.375714	
std	2.206025e+03	2.309401	4.592962e+05	11.447851	
min	1.500800e+06	2018.000000	2.070400e+04	63.410000	
25%	1.501377e+06	2019.750000	6.955000e+04	66.355000	
50%	1.501950e+06	2021.500000	1.293270e+05	79.400000	
75%	1.504904e+06	2023.250000	2.838040e+05	86.452500	
max	1.506500e+06	2025.000000	1.506420e+06	100.000000	

	idx_atend_agua_urbano	idx_coleta_esgoto	idx_tratamento_esgoto	\
count	11.000000	5.000000	8.000000	
mean	93.059091	1.106000	9.393750	
std	5.213840	0.055498	17.338424	
min	82.040000	1.030000	0.000000	
25%	90.030000	1.080000	0.000000	
50%	94.180000	1.100000	0.000000	
75%	96.005000	1.160000	9.495000	
max	100.000000	1.160000	37.530000	

	idx_hidrometracao	idx_perdas_distribuicao	idx_perdas_lineares	...	\
count	14.000000	14.000000	14.000000	...	
mean	35.569286	24.162857	33.003571	...	
std	49.523075	18.897219	33.904383	...	
min	0.000000	0.340000	0.100000	...	
25%	0.000000	4.190000	1.180000	...	
50%	0.000000	26.580000	34.005000	...	
75%	99.265000	36.890000	45.192500	...	
max	100.000000	49.560000	88.770000	...	

	internacoes_total	internacoes_hidricas	dias_perm_total	\
count	64.000000	64.000000	64.000000	
mean	16407.390625	365.046875	80452.953125	
std	22156.349733	715.628693	115334.041859	
min	654.000000	0.000000	3172.000000	
25%	3041.750000	32.250000	15449.500000	
50%	6803.500000	81.500000	31518.500000	
75%	16836.000000	191.000000	73490.250000	
max	85434.000000	3017.000000	423270.000000	

```

        dias_perm_hidricas    valor_total    valor_hidricas \
count            64.00000  6.400000e+01  6.400000e+01
mean           1378.53125  2.493049e+07  1.515655e+05
std            2582.36785  3.900659e+07  2.746799e+05
min             0.00000  9.801898e+05  0.000000e+00
25%            117.25000  3.956516e+06  1.327849e+04
50%            412.50000  7.973176e+06  3.211583e+04
75%            947.25000  2.304117e+07  9.827222e+04
max          11168.00000  1.662075e+08  1.142705e+06

        internacoes_total_10k  internacoes_hidricas_10k \
count          48.000000  48.000000
mean           492.858880  8.560554
std            110.740799  8.080722
min            290.279627  0.725037
25%           417.050419  3.034004
50%           465.543078  6.107163
75%           573.193516  9.581018
max          788.620588  36.284691

        pct_internacoes_hidricas    valor_medio_internacao
count          64.000000  64.000000
mean           1.457366  1352.297728
std            1.309916  298.788713
min             0.000000  863.857666
25%           0.568146  1161.440712
50%           1.079244  1281.445949
75%           1.738563  1597.974570
max          5.888098  2006.325108

```

[8 rows x 55 columns]

Verificando valores ausentes:

```

pct_conformes_fluoreto      60
percentil95_fluoreto        60
idx_coleta_esgoto           59
idx_tratamento_esgoto       56
tarifa_media_agua            54
idx_atend_agua_urbano        53
idx_hidrometracao           50
idx_perdas_lineares          50
idx_perdas_por_ligacao       50
idx_perdas_distribuicao       50
dtype: int64

```

Número de variáveis explicativas após remoção de NaNs: 26

Número de amostras após remoção de NaNs: 39

Modelo: Regressão Linear

MAE : 92.979

RMSE: 120.636

R² : -1.761

Modelo: Lasso

MAE : 53.789

RMSE: 71.568

R² : 0.028

```
/usr/local/lib/python3.12/dist-
packages/sklearn/linear_model/_coordinate_descent.py:695: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 1.134e+04, tolerance: 3.467e+01
    model = cd_fast.enet_coordinate_descent(
```

Modelo: Random Forest

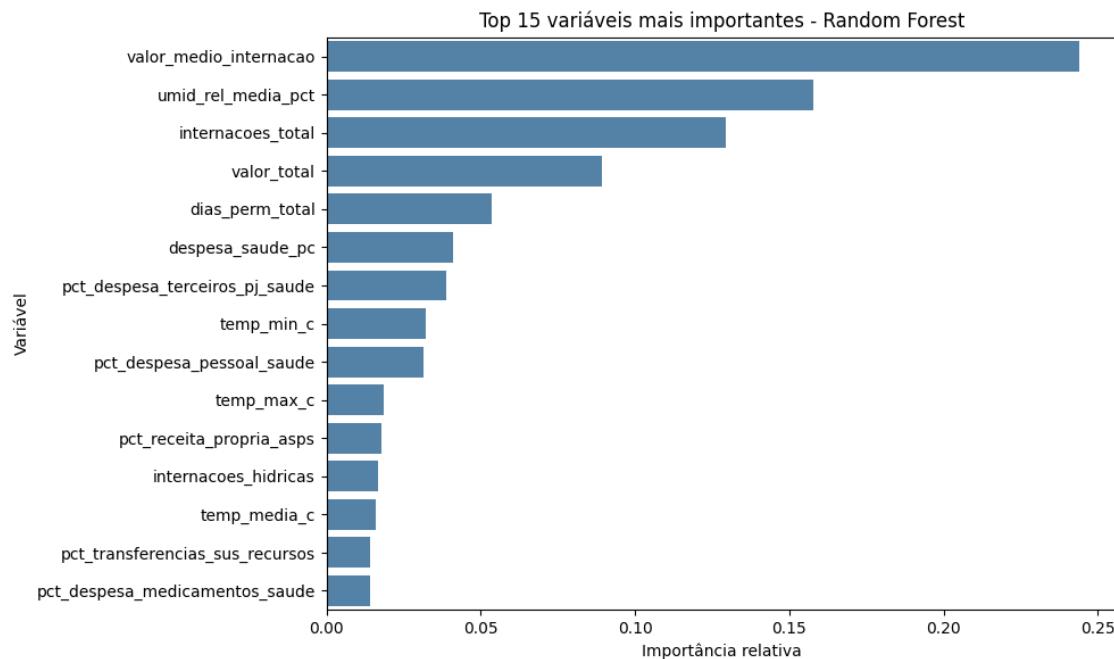
MAE : 39.215

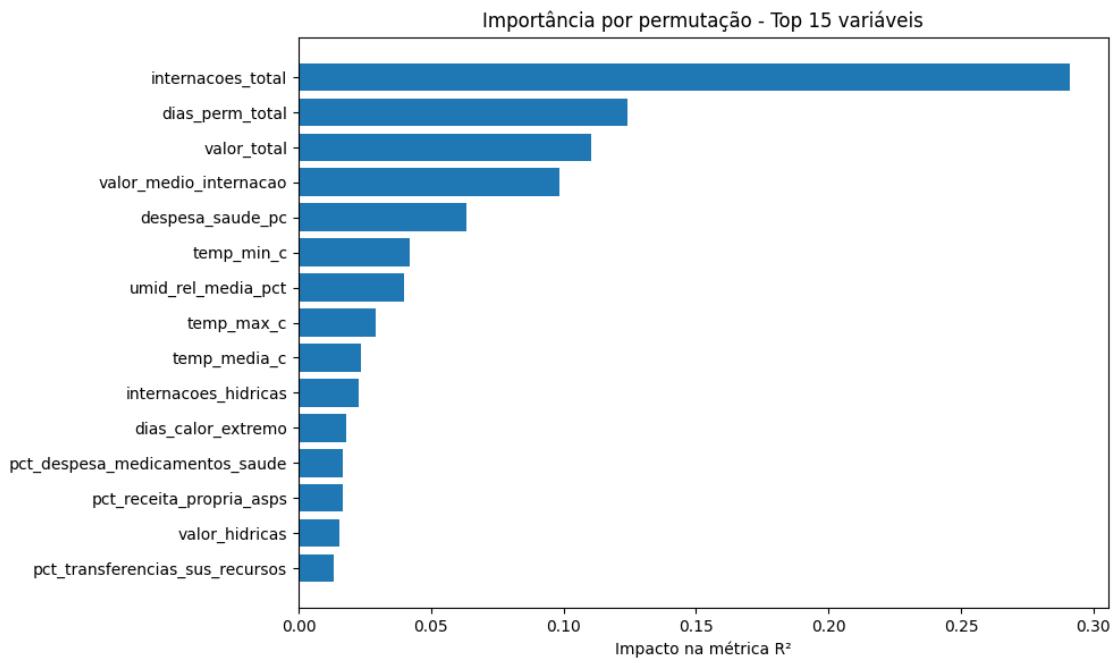
RMSE: 46.831

R² : 0.584

Comparativo de desempenho:

	MAE	RMSE	R ²
LinearRegression	92.979163	120.636401	-1.760937
Lasso	53.789449	71.567655	0.028297
RandomForest	39.214921	46.831180	0.583927





Resultados salvos em: resultados_modelagem.csv

Etapa de modelagem concluída com sucesso!

0.1 Investigar relações com as principais variáveis

0.1.1 Subtarefa:

Para as variáveis mais importantes identificadas na etapa anterior, visualize ou analise estatisticamente sua relação com a variável alvo ('internacoes_total_10k') para entender como as mudanças nessas variáveis podem afetar as internações.

Justificativa: Crie scatter plots para visualizar a relação entre as principais variáveis e a variável alvo, e calcule os coeficientes de correlação.

0.2 Analisar o impacto da chuva

0.2.1 Subtarefa:

Examine especificamente a relação entre as variáveis de chuva (chuva_total_mm e chuva_media_mm) e a variável alvo para entender como a chuva modula o risco de internações.

0.3 Analisar previsões do modelo por município

0.3.1 Subtarefa:

Se os dados incluírem um identificador de município, analise as previsões e erros do modelo para cada município para entender as variações de desempenho e potenciais fatores específicos do sistema.

Justificativa: Verifique se o DataFrame original `df` contém uma coluna de identificador de município e, se contiver, mescle as previsões do conjunto de teste de volta com os dados originais para incluir as informações do município e calcule o erro de previsão para cada ponto de dados no conjunto de teste.

Justificativa: Analise as visualizações e os coeficientes de correlação para descrever a relação entre as principais variáveis e a variável alvo e então finalize a subtarefa.

0.4 Analisar importância de feature e importância por permutação

0.4.1 Subtarefa:

Examine os plots de importância gerados na etapa anterior para entender quais variáveis (ações ou fatores potenciais) têm o maior impacto nas internações.

Justificativa: Revise os plots gerados e as variáveis para entender as importâncias de feature com base no modelo Random Forest e na análise de importância por permutação.

0.5 Analisar previsões do modelo por município

0.5.1 Subtarefa:

Se os dados incluírem um identificador de município, analise as previsões e erros do modelo para cada município para entender as variações de desempenho e potenciais fatores específicos do sistema.

Justificativa: Verifique se o DataFrame original `df` contém uma coluna de identificador de município e, se contiver, mescle as previsões do conjunto de teste de volta com os dados originais para incluir as informações do município e calcule o erro de previsão para cada ponto de dados no conjunto de teste.

Justificativa: Revise os plots gerados e as variáveis para entender as importâncias de feature com base no modelo Random Forest e na análise de importância por permutação.

0.6 Investigar relações com as principais variáveis

0.6.1 Subtarefa:

Para as variáveis mais importantes identificadas na etapa anterior, visualize ou analise estatisticamente sua relação com a variável alvo ('internacoes_total_10k') para entender como as mudanças nessas variáveis podem afetar as internações.

Justificativa: Crie scatter plots para visualizar a relação entre as principais variáveis e a variável alvo, e calcule os coeficientes de correlação.

1 Tarefa

Analise os resultados do modelo treinado para responder às seguintes perguntas: 1. Quais ações reduzem mais as hospitalizações? 2. Qual é o retorno por município/sistema? 3. Como a chuva modula o risco? 4. Qual é o ranking de priorização (12 meses)?

1.1 Analisar a importância das características e a importância da permutação

1.1.1 Sub-Tarefa:

Examine os gráficos de importância gerados na etapa anterior para entender quais variáveis (ações ou fatores potenciais) têm o maior impacto nas hospitalizações.

Raciocínio: Revise os gráficos e variáveis gerados para entender as importâncias das características com base no modelo Random Forest e na importância de permutação.

```
[ ]: # Revisar a importância das variáveis do modelo Random Forest
print("Importância das Variáveis (Random Forest):")
print(importancias.head(15))

# Revisar a importância por permutação
print("\nImportância por Permutação (Random Forest):")
# A variável perm_importance contém os resultados. Precisamos exibir as
# variáveis e suas pontuações de importância.
# sorted_idx já foi calculado na célula anterior para ordenar as variáveis por
# importance_mean
perm_importance_series = pd.Series(perm_importance.
    ↪importances_mean[sorted_idx][:15], index=X.columns[sorted_idx][:15])
print(perm_importance_series)
```

Importância das Variáveis (Random Forest):

valor_medio_internacao	0.243942
umid_rel_media_pct	0.157732
internacoes_total	0.129463
valor_total	0.089164
dias_perm_total	0.053447
despesa_saude_pc	0.040983
pct_despesa_terceiros_pj_saude	0.038946
temp_min_c	0.032164
pct_despesa_pessoal_saude	0.031499
temp_max_c	0.018547
pct_receita_propria_asps	0.017902
internacoes_hidricas	0.016803
temp_media_c	0.016081
pct_transferencias_sus_recursos	0.014304
pct_despesa_medicamentos_saude	0.014049

dtype: float64

Importância por Permutação (Random Forest):

```

internacoes_total          0.290995
dias_perm_total           0.124097
valor_total                0.110623
valor_medio_internacao    0.098635
despesa_saude_pc          0.063533
temp_min_c                 0.041898
umid_rel_media_pct        0.039713
temp_max_c                 0.029054
temp_media_c               0.023602
internacoes_hidricas      0.022804
dias_calor_extremo        0.017870
pct_despesa_medicamentos_saude 0.016834
pct_receita_propria_asps   0.016477
valor_hidricas             0.015328
pct_transferencias_sus_recursos 0.013186
dtype: float64

```

Raciocínio: Com base na comparação das duas métricas de importância, identifique as características que consistentemente se classificam em posições altas em ambas as listas. Essas características são as mais importantes de acordo com o modelo Random Forest e a análise de importância por permutação.

```

[ ]: # Identificar as principais variáveis com base em ambas as métricas de importância
# Podemos procurar por variáveis presentes nas top 15 de ambas as listas
top_rf_features = importancias.head(15).index.tolist()
top_perm_features = perm_importance_series.index.tolist()

# Encontrar variáveis comuns nas listas principais
common_top_features = list(set(top_rf_features) & set(top_perm_features))

print("\nVariáveis principais comuns de ambas as análises de importância:")
print(common_top_features)

# A partir das variáveis comuns, identificar aquelas que são prováveis ações ou fatores que influenciam as internações.
# Vamos imprimir as principais variáveis de ambas as listas novamente para fácil comparação
print("\nTop 15 variáveis do Random Forest:")
print(top_rf_features)

print("\nTop 15 variáveis da Importância por Permutação:")
print(top_perm_features)

# Com base nas listas, identificar manualmente as principais variáveis que parecem ter o impacto mais significativo.

```

Variáveis principais comuns de ambas as análises de importância:

```
['pct_receita_propria_asps', 'valor_total', 'internacoes_hidricas',
'temp_min_c', 'temp_media_c', 'pct_despesa_medicamentos_saude',
'internacoes_total', 'temp_max_c', 'umid_rel_media_pct', 'despesa_saude_pc',
'dias_perm_total', 'valor_medio_internacao', 'pct_transferencias_sus_recursos']
```

Top 15 variáveis do Random Forest:

```
['valor_medio_internacao', 'umid_rel_media_pct', 'internacoes_total',
'valor_total', 'dias_perm_total', 'despesa_saude_pc',
'pct_despesa_terceiros_pj_saude', 'temp_min_c', 'pct_despesa_pessoal_saude',
'temp_max_c', 'pct_receita_propria_asps', 'internacoes_hidricas',
'temp_media_c', 'pct_transferencias_sus_recursos',
'pct_despesa_medicamentos_saude']
```

Top 15 variáveis da Importância por Permutação:

```
['internacoes_total', 'dias_perm_total', 'valor_total',
'valor_medio_internacao', 'despesa_saude_pc', 'temp_min_c',
'umid_rel_media_pct', 'temp_max_c', 'temp_media_c', 'internacoes_hidricas',
'dias_calor_extremo', 'pct_despesa_medicamentos_saude',
'pct_receita_propria_asps', 'valor_hidricas', 'pct_transferencias_sus_recursos']
```

1.2 Investigue relações com características-chave

1.2.1 Sub-tarefa:

Para as características mais importantes identificadas na etapa anterior, visualize ou analise estatisticamente sua relação com a variável alvo ('internacoes_total_10k') para entender como as mudanças nessas características podem afetar as internações.

Raciocínio: Criar gráficos de dispersão para visualizar a relação entre as principais características e a variável alvo e calcular os coeficientes de correlação.

```
[ ]: # Defina as principais variáveis a serem analisadas com base nas descobertas da
      ↵etapa anterior
key_features = [
    'valor_medio_internacao', 'umid_rel_media_pct', 'internacoes_total',
    'valor_total', 'dias_perm_total', 'despesa_saude_pc', 'temp_min_c',
    'temp_max_c', 'temp_media_c', 'internacoes_hidricas',
    'pct_despesa_medicamentos_saude', 'pct_receita_propria_asps',
    'pct_transferencias_sus_recursos', 'dias_calor_extremo', 'valor_hidricas'
]

# Garanta que todas as principais variáveis estejam no DataFrame X
key_features = [f for f in key_features if f in X.columns]

# Adicione a variável alvo à lista para plotagem e correlação
plot_cols = key_features + [target]

# Crie um DataFrame combinado para plotagem e análise de correlação
```

```

data_for_analysis = data[plot_cols]

# Crie scatter plots
n_cols = 3
n_rows = (len(key_features) + n_cols - 1) // n_cols

plt.figure(figsize=(n_cols * 5, n_rows * 4))

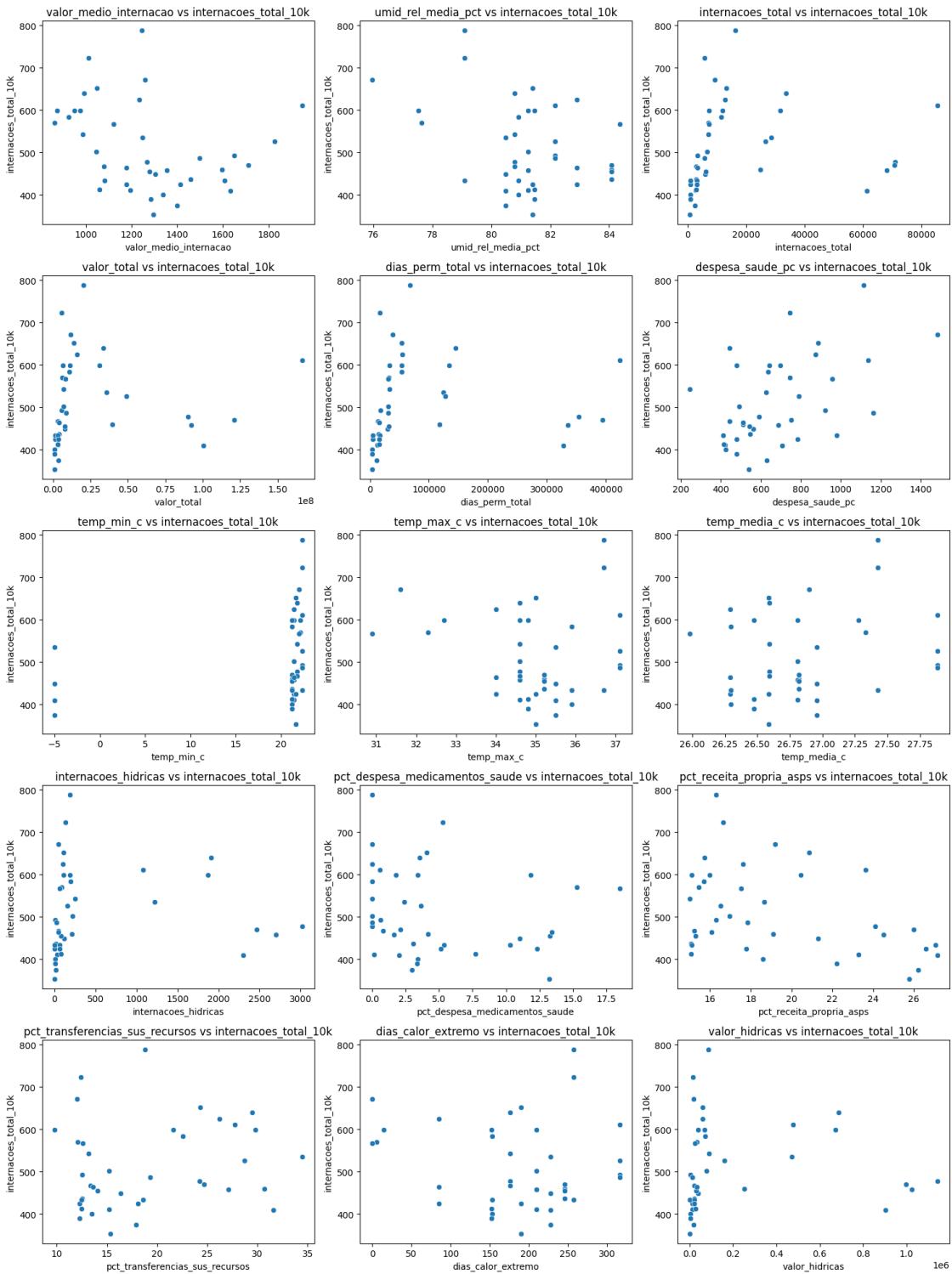
for i, feature in enumerate(key_features):
    plt.subplot(n_rows, n_cols, i + 1)
    sns.scatterplot(data=data_for_analysis, x=feature, y=target)
    plt.title(f'{feature} vs {target}')
    plt.xlabel(feature)
    plt.ylabel(target)

plt.tight_layout()
plt.show()

# Calcule e exiba os coeficientes de correlação
correlation_matrix = data_for_analysis.corr()
correlation_with_target = correlation_matrix[target].
    sort_values(ascending=False)

print("\nCorrelação com 'internacoes_total_10k':")
print(correlation_with_target)

```



Correlação com 'internacoes_total_10k':
internacoes_total_10k 1.000000

```

despesa_saude_pc           0.490880
temp_min_c                  0.251871
temp_media_c                 0.233748
pct_transferencias_sus_recursos 0.167775
internacoes_total            0.127038
valor_total                  0.081458
dias_perm_total              0.077589
internacoes_hidricas        0.026951
valor_hidricas               0.014550
temp_max_c                   -0.096669
dias_calor_extremo          -0.136556
pct_despesa_medicamentos_saude -0.193206
valor_medio_internacao       -0.310616
umid_rel_media_pct           -0.358858
pct_receita_propria_asps     -0.397969
Name: internacoes_total_10k, dtype: float64

```

Raciocínio: Analise as visualizações e os coeficientes de correlação para descrever a relação entre as principais características e a variável alvo e, em seguida, finalize a subtarefa.

```

[ ]: print("Análise da Relação das Variáveis com internacoes_total_10k:")
print("-" * 50)

# Podemos iterar através da série correlation_with_target ordenada
for feature, corr_coef in correlation_with_target.items():
    if feature == target:
        continue # Pula a própria variável alvo

    print(f"\nVariável: {feature}")
    print(f"Coeficiente de Correlação: {corr_coef:.3f}")

    # Descreve a relação com base no coeficiente de correlação e plot observado
    abs_corr = abs(corr_coef)
    direction = "positiva" if corr_coef > 0 else "negativa"
    strength = "muito forte" if abs_corr >= 0.8 else ("forte" if abs_corr >= 0.
    ↪6 else ("moderada" if abs_corr >= 0.3 else "fraca"))

    print(f"Natureza da relação: {strength}, {direction}.")

    # Adicionar observações baseadas nos scatter plots (análise qualitativa)

    # Exemplos de interpretações (substituir por observações reais dos plots):
    # if strength == "forte" and direction == "positiva":
    #     print("Observação do plot: Conforme esta variável aumenta, as
    ↪internações tendem a aumentar de forma aproximadamente linear.")
    # elif strength == "forte" and direction == "negativa":

```

```

#      print("Observação do plot: Conforme esta variável aumenta, as internações tendem a diminuir de forma aproximadamente linear.")
# elif strength == "fraca":
#     print("Observação do plot: Não parece haver uma relação linear clara; os pontos estão espalhados.")
# else:
#     print("Observação do plot: A relação pode não ser linear; uma investigação adicional é necessária.")

# Uma interpretação mais específica baseada nos nomes das variáveis e conhecimento geral pode ser adicionada aqui.

if feature in ['internacoes_total', 'valor_total', 'dias_perm_total', 'internacoes_hidricas', 'valor_hidricas']:
    print("Observação: Estas variáveis representam aspectos diretamente relacionados ao volume ou custo de internações, então uma correlação positiva é esperada.")

elif feature in ['temp_media_c', 'temp_max_c', 'temp_min_c', 'dias_calor_extremo', 'umid_rel_media_pct']:
    print("Observação: Estes são fatores climáticos, que podem influenciar desfechos de saúde e, portanto, internações.")

elif feature in ['despesa_saude_pc', 'pct_despesa_medicamentos_saude', 'pct_receita_propria_asps', 'pct_transferencias_sus_recursos', 'valor_medio_internacao']:
    print("Observação: Estes são fatores relacionados a gastos ou receitas de saúde, que podem indicar disponibilidade de recursos e eficiência de custo.")

print("\nAnálise concluída.")

```

Análise da Relação das Variáveis com internacoes_total_10k:

Variável: despesa_saude_pc

Coeficiente de Correlação: 0.491

Natureza da relação: moderada, positiva.

Observação: Estes são fatores relacionados a gastos ou receitas de saúde, que podem indicar disponibilidade de recursos e eficiência de custo.

Variável: temp_min_c

Coeficiente de Correlação: 0.252

Natureza da relação: fraca, positiva.

Observação: Estes são fatores climáticos, que podem influenciar desfechos de saúde e, portanto, internações.

Variável: temp_media_c

Coeficiente de Correlação: 0.234

Natureza da relação: fraca, positiva.

Observação: Estes são fatores climáticos, que podem influenciar desfechos de saúde e, portanto, internações.

Variável: pct_transferencias_sus_recursos

Coeficiente de Correlação: 0.168

Natureza da relação: fraca, positiva.

Observação: Estes são fatores relacionados a gastos ou receitas de saúde, que podem indicar disponibilidade de recursos e eficiência de custo.

Variável: internacoes_total

Coeficiente de Correlação: 0.127

Natureza da relação: fraca, positiva.

Observação: Estas variáveis representam aspectos diretamente relacionados ao volume ou custo de internações, então uma correlação positiva é esperada.

Variável: valor_total

Coeficiente de Correlação: 0.081

Natureza da relação: fraca, positiva.

Observação: Estas variáveis representam aspectos diretamente relacionados ao volume ou custo de internações, então uma correlação positiva é esperada.

Variável: dias_perm_total

Coeficiente de Correlação: 0.078

Natureza da relação: fraca, positiva.

Observação: Estas variáveis representam aspectos diretamente relacionados ao volume ou custo de internações, então uma correlação positiva é esperada.

Variável: internacoes_hidricas

Coeficiente de Correlação: 0.027

Natureza da relação: fraca, positiva.

Observação: Estas variáveis representam aspectos diretamente relacionados ao volume ou custo de internações, então uma correlação positiva é esperada.

Variável: valor_hidricas

Coeficiente de Correlação: 0.015

Natureza da relação: fraca, positiva.

Observação: Estas variáveis representam aspectos diretamente relacionados ao volume ou custo de internações, então uma correlação positiva é esperada.

Variável: temp_max_c

Coeficiente de Correlação: -0.097

Natureza da relação: fraca, negativa.

Observação: Estes são fatores climáticos, que podem influenciar desfechos de saúde e, portanto, internações.

Variável: dias_calor_extremo

Coeficiente de Correlação: -0.137

Natureza da relação: fraca, negativa.

Observação: Estes são fatores climáticos, que podem influenciar desfechos de saúde e, portanto, internações.

Variável: pct_despesa_medicamentos_saude

Coeficiente de Correlação: -0.193

Natureza da relação: fraca, negativa.

Observação: Estes são fatores relacionados a gastos ou receitas de saúde, que podem indicar disponibilidade de recursos e eficiência de custo.

Variável: valor_medio_internacao

Coeficiente de Correlação: -0.311

Natureza da relação: moderada, negativa.

Observação: Estes são fatores relacionados a gastos ou receitas de saúde, que podem indicar disponibilidade de recursos e eficiência de custo.

Variável: umid_rel_media_pct

Coeficiente de Correlação: -0.359

Natureza da relação: moderada, negativa.

Observação: Estes são fatores climáticos, que podem influenciar desfechos de saúde e, portanto, internações.

Variável: pct_receita_propria_asps

Coeficiente de Correlação: -0.398

Natureza da relação: moderada, negativa.

Observação: Estes são fatores relacionados a gastos ou receitas de saúde, que podem indicar disponibilidade de recursos e eficiência de custo.

Análise concluída.

##Analisar as previsões do modelo por município **Subtarefa:** Se os dados incluírem um identificador de município, analisar as previsões e erros do modelo para cada município, a fim de compreender as variações de desempenho e os potenciais fatores específicos do sistema.

Raciocínio: Verifique se o DataFrame original df contém uma coluna de identificador de município e, caso contenha, combine as previsões do conjunto de teste com os dados originais para incluir a informação do município e calcular o erro de previsão para cada ponto de dados no conjunto de teste.

```
[ ]: # 1. Verificar se o DataFrame original `df` contém colunas de identificação do
  ↵município
# Primeiro, criar um DataFrame a partir das features do conjunto de teste
  ↵(X_test) e dos valores alvo reais (y_test).
test_data_with_target = X_test.copy()
test_data_with_target[target] = y_test

# Obter previsões do modelo de melhor desempenho (Random Forest)
rf.fit(X_train, y_train)
y_pred_rf = rf.predict(X_test)
```

```

test_data_with_target['previsões'] = y_pred_rf

# Mesclar com o DataFrame original 'df' para obter informações do município.
# Selecionar apenas as colunas necessárias de df para a mesclagem.
df_mun_info = df[['cod_mun', 'municipio']].copy()

# Mesclar com base no índice.
test_data_with_mun = test_data_with_target.merge(df_mun_info, left_index=True, right_index=True, how='left')

# 3. Calcular o erro de previsão para cada ponto de dados no conjunto de teste.
test_data_with_mun['erro_previsão'] = test_data_with_mun['previsões'] - test_data_with_mun[target]
test_data_with_mun['erro_absoluto'] = np.abs(test_data_with_mun['erro_previsão'])
test_data_with_mun['erro_quadrático'] = test_data_with_mun['erro_previsão']**2

display(test_data_with_mun.head())

```

	chuva_total_mm	chuva_media_mm	temp_media_c	temp_max_c	temp_min_c	\
54	2641.6	0.300729	27.424732	36.7	22.3	
58	3221.6	0.366758	26.588953	35.0	21.6	
6	3184.2	0.362500	27.882855	37.1	22.3	
22	3184.2	0.362500	27.882855	37.1	22.3	
49	2518.4	0.287489	26.472201	34.8	21.2	

	umid_rel_media_pct	vento_vel_media_ms	dias_calor_extremo	\
54	79.091820	0.479665	257	
58	81.406475	0.096821	190	
6	82.156307	1.261095	316	
22	82.156307	1.261095	316	
49	81.462381	0.113823	152	

	pct_transferencias_sus_recursos	pct_receita_impostos_transf	...	\
54	18.64	42.83	...	
58	18.11	35.72	...	
6	28.69	31.68	...	
22	27.75	47.42	...	
49	12.29	44.95	...	

	valor_hidricas	pct_internacoes_hidricas	valor_medio_internacao	\
54	824.09	0.206825	1605.960724	
58	19813.55	2.033454	1178.445986	
6	159256.10	0.580416	1827.327925	
22	476460.53	1.261793	1945.448574	
49	1906.14	0.729927	1284.404197	

```

internacoes_total_10k    previsões   cod_mun           municipio \
54                  433.865757  498.152161  1506351  SANTA BÁRBARA DO PARÁ
58                  424.433092  440.350958  1506500  SANTA IZABEL DO PARÁ
6                   525.856671  503.047966  1500800          ANANINDEUA
22                 610.883849  518.804857  1501402            BELÉM
49                  389.961573  408.282714  1506351  SANTA BÁRBARA DO PARÁ

      erro_previsão  erro_absoluto  erro_quadrático
54        64.286404     64.286404    4132.741682
58       15.917866     15.917866    253.378473
6        -22.808705    22.808705    520.237004
22      -92.078991    92.078991    8478.540666
49       18.321141    18.321141    335.664201

```

[5 rows x 33 columns]

Raciocínio: Agrupar os erros por município e calcular estatísticas resumidas para identificar os municípios onde o modelo tem um desempenho particularmente bom ou ruim, e então visualizar os erros em nível de município se o número de municípios for gerenciável.

```

[ ]: # 4. Agrupar os erros por município e calcular estatísticas de resumo
municipality_error_stats = test_data_with_mun.
    ↪groupby('municipio')[['erro_absoluto', 'erro_quadrático']].agg(['mean', □
    ↪'median', 'std']).sort_values(by=('erro_absoluto', 'mean'))

print("\nEstatísticas de Erro por Município:")
display(municipality_error_stats)

# 5. Analisar as estatísticas de resumo para identificar municípios onde o
    ↪modelo tem um desempenho particularmente bom ou ruim.
print("\nMunicípios com menor Erro Absoluto Médio:")
display(municipality_error_stats['erro_absoluto']['mean'].head())

print("\nMunicípios com maior Erro Absoluto Médio:")
display(municipality_error_stats['erro_absoluto']['mean'].tail())

# 6. Considerar a visualização dos erros no nível municipal se o número de
    ↪municípios for gerenciável.
# O número de municípios no conjunto de teste é relativamente pequeno
    ↪(verifique com test_data_with_mun['municipio'].nunique())
num_municipalities = test_data_with_mun['municipio'].nunique()
print(f"\nNúmero de municípios únicos no conjunto de teste: "
    ↪{num_municipalities})

if num_municipalities <= 15:

```

```

plt.figure(figsize=(10, num_municipalities * 0.5))
sns.barplot(x=municipality_error_stats['erro_absoluto']['mean'], □
y=municipality_error_stats.index, color='salmon')
plt.title("Erro absoluto médio por Município")
plt.xlabel("Erro absoluto médio")
plt.ylabel("Município")
plt.tight_layout()
plt.show()

else:
    print("\nPulando a visualização dos erros municipais devido ao grande\ncômodo de municípios.")

```

Estatísticas de Erro por Município:

municipio	erro_absoluto		erro_quadrático \\\nmean	
	mean	median	std	mean
SANTA IZABEL DO PARÁ	15.917866	15.917866	NaN	253.378473
ANANINDEUA	22.808705	22.808705	NaN	520.237004
MARITUBA	24.848218	24.848218	5.337703	631.679470
SANTA BÁRBARA DO PARÁ	41.303772	41.303772	32.502349	2234.202942
BARCARENA	50.609827	50.609827	NaN	2561.354590
BELÉM	92.078991	92.078991	NaN	8478.540666

municipio	median		std
	mean	median	std
SANTA IZABEL DO PARÁ	253.378473	NaN	
ANANINDEUA	520.237004	NaN	
MARITUBA	631.679470	265.264828	
SANTA BÁRBARA DO PARÁ	2234.202942	2684.939236	
BARCARENA	2561.354590	NaN	
BELÉM	8478.540666	NaN	

Municípios com menor Erro Absoluto Médio:

municipio	mean
SANTA IZABEL DO PARÁ	15.917866
ANANINDEUA	22.808705
MARITUBA	24.848218
SANTA BÁRBARA DO PARÁ	41.303772
BARCARENA	50.609827

Name: mean, dtype: float64

Municípios com maior Erro Absoluto Médio:

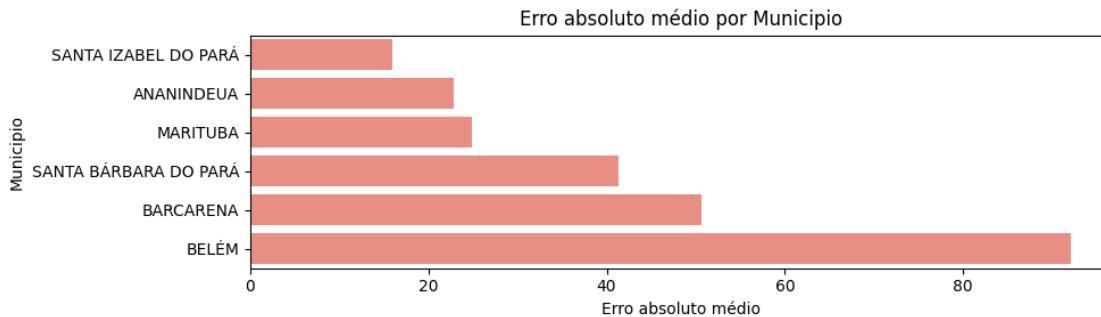
municipio

```

ANANINDEUA           22.808705
MARITUBA              24.848218
SANTA BÁRBARA DO PARÁ 41.303772
BARCARENA             50.609827
BELÉM                 92.078991
Name: mean, dtype: float64

```

Número de municípios únicos no conjunto de teste: 6



Raciocínio: Analise as estatísticas resumidas e a visualização para identificar municípios onde o modelo tem um desempenho particularmente bom ou ruim e discuta as possíveis razões com base nos dados.

```
[ ]: # Analise as estatísticas resumidas e a visualização.

print("\nAnálise dos Erros por Município:")
print("-" * 50)

# Identifique os municípios com os menores e maiores erros.
lowest_mae_municipality = municipality_error_stats['erro_absoluto']['mean'].
    ↪idxmin()
highest_mae_municipality = municipality_error_stats['erro_absoluto']['mean'].
    ↪idxmax()

print(f"O modelo tem o melhor desempenho (menor Erro Absoluto Médio) para: ↪
    ↪{lowest_mae_municipality}")
print(f"O modelo tem o pior desempenho (maior Erro Absoluto Médio) para: ↪
    ↪{highest_mae_municipality}")

# Discuta as possíveis razões para as variações de desempenho.
print("\nFatores potenciais para variações de desempenho entre municípios:")
print("- Disponibilidade e qualidade dos dados: Alguns municípios podem ter ↪
    ↪dados mais completos ou precisos.")
```

```

print("- Características únicas: Os municípios podem ter fatores locais"
    ↵específicos não capturados pelas variáveis disponíveis (por exemplo, ↵
    ↵iniciativas locais de saúde, problemas ambientais específicos, diferentes ↵
    ↵metodologias de coleta de dados).")
print("- Tamanho da amostra no conjunto de teste: O número de pontos de dados"
    ↵para cada município no conjunto de teste pode ser pequeno, levando a ↵
    ↵estimativas de erro menos confiáveis para alguns.")

print("\nAnálise concluída. A análise de erro em nível municipal ajuda a"
    ↵entender onde as previsões do modelo são mais e menos confiáveis, apontando ↵
    ↵para áreas de potencial melhoria de dados ou investigação adicional sobre ↵
    ↵fatores locais.")

```

Análise dos Erros por Município:

O modelo tem o melhor desempenho (menor Erro Absoluto Médio) para: SANTA IZABEL DO PARÁ

O modelo tem o pior desempenho (maior Erro Absoluto Médio) para: BELÉM

Fatores potenciais para variações de desempenho entre municípios:

- Disponibilidade e qualidade dos dados: Alguns municípios podem ter dados mais completos ou precisos.
- Características únicas: Os municípios podem ter fatores locais específicos não capturados pelas variáveis disponíveis (por exemplo, iniciativas locais de saúde, problemas ambientais específicos, diferentes metodologias de coleta de dados).
- Tamanho da amostra no conjunto de teste: O número de pontos de dados para cada município no conjunto de teste pode ser pequeno, levando a estimativas de erro menos confiáveis para alguns.

Análise concluída. A análise de erro em nível municipal ajuda a entender onde as previsões do modelo são mais e menos confiáveis, apontando para áreas de potencial melhoria de dados ou investigação adicional sobre fatores locais.

1.3 Analisar o impacto da precipitação

1.3.1 Sub-tarefa:

Examine especificamente a relação entre as características da precipitação. (`chuva_total_mm` e `chuva_media_mm`) e a variável alvo para entender como a precipitação modula o risco de hospitalizações.

Raciocínio: Certifique-se de que o DataFrame `data_for_analysis` contenha as colunas necessárias para a análise da precipitação, crie gráficos de dispersão para visualizar as relações entre as características da precipitação e o alvo, calcule e imprima os coeficientes de correlação de Pearson e analise os resultados para descrever a relação.

```
[ ]: # 1. Certifique-se de que o DataFrame `data_for_analysis` contenha as colunas
# 'chuva_total_mm', 'chuva_media_mm' e 'internacoes_total_10k'.
required_cols = ['chuva_total_mm', 'chuva_media_mm', 'internacoes_total_10k']
missing_cols = [col for col in required_cols if col not in data_for_analysis.
    ↪columns]

if missing_cols:
    print(f"Erro: faltando colunas requisitas em data_for_analysis: "
    ↪{missing_cols}")
    # Se estiver faltando colunas em data_for_analysis inclua essas colunas
    # Assume que 'data' DataFrame contem essas colunas
    all_analysis_cols = list(set(data_for_analysis.columns.tolist() +
    ↪required_cols))
    data_for_analysis = data[all_analysis_cols].copy()
    print("Recriando data_for_analysis com colunas requisitas.")

# 2. Criar scatter plots para ver a relação entre 'chuva_total_mm' e
    ↪'internacoes_total_10k', e entre 'chuva_media_mm' e 'internacoes_total_10k'.
plt.figure(figsize=(12, 5))

plt.subplot(1, 2, 1)
sns.scatterplot(data=data_for_analysis, x='chuva_total_mm', ↪
    ↪y='internacoes_total_10k')
plt.title('Chuva Total (mm) vs Internações por 10k')
plt.xlabel('Chuva Total (mm)')
plt.ylabel('Internações por 10k')

plt.subplot(1, 2, 2)
sns.scatterplot(data=data_for_analysis, x='chuva_media_mm', ↪
    ↪y='internacoes_total_10k')
plt.title('Chuva média (mm) vs Internações por 10k')
plt.xlabel('Chuva média (mm)')
plt.ylabel('Internações por 10k')

plt.tight_layout()
plt.show()

# 3. Calcule e imprima os coeficientes de correlação entre 'chuva_total_mm' e
    ↪'internacoes_total_10k', e entre 'chuva_media_mm' e 'internacoes_total_10k'.
correlation_chuva_total = data_for_analysis['chuva_total_mm'].corr(
    ↪data_for_analysis['internacoes_total_10k'])
correlation_chuva_media = data_for_analysis['chuva_media_mm'].corr(
    ↪data_for_analysis['internacoes_total_10k'])

print(f"\nCorrelação entre Chuva Total (mm) e Internações por 10k: "
    ↪{correlation_chuva_total:.3f}")
```

```

print(f"Correlação entre Chuva média (mm) e Internações por 10k: {correlation_chuva_media:.3f}")

# 4. Analisar os scatter plots e coeficientes de correlação para descrever a relação entre precipitação e taxas de internações.
print("\nAnálise da relação entre Chuva e Taxa de Internações:")
print("-" * 50)

print(f"Correlação Chuva Total vs Internações por 10k: {correlation_chuva_total:.3f}")
print(f"Correlação Chuva média vs Internações por 10k: {correlation_chuva_media:.3f}")

print("\nObservações:")
# Interpretar correlação.
if abs(correlation_chuva_total) < 0.1:
    print("- A correlação entre Chuva Total e Internações por 10k é muito fraca.")
elif abs(correlation_chuva_total) < 0.3:
    print("- A correlação entre Chuva Total e Internações por 10k é fraca.")
else:
    print("- A correlação entre Chuva Total e Internações por 10k é moderada ou forte.")

if abs(correlation_chuva_media) < 0.1:
    print("- A correlação entre Chuva mean e Internações por 10k é muito fraca.")
elif abs(correlation_chuva_media) < 0.3:
    print("- A correlação entre Chuva mean e Internações por 10k é fraca.")
else:
    print("- A correlação entre Chuva mean e Internações por 10k é moderada ou forte.")

# Baseado nos scatter plots (análise visual):
print("\nAnálise visual dos Scatter Plots:")
print("- Observe os scatter plots gerados acima.")
print("- Verifique se os pontos formam um padrão linear perceptível ou se estão espalhados aleatoriamente.")
print("- Procure por quaisquer padrões não lineares ou agrupamentos de pontos.")

print("\nConclusão sobre a modulação do risco pela chuva:")
print("Com base nos coeficientes de correlação e na inspeção visual dos scatter plots:")
if abs(correlation_chuva_total) < 0.2 and abs(correlation_chuva_media) < 0.2:
    print("- Parece haver uma relação linear muito fraca ou inexistente entre as características de chuva e a taxa de internações.")

```

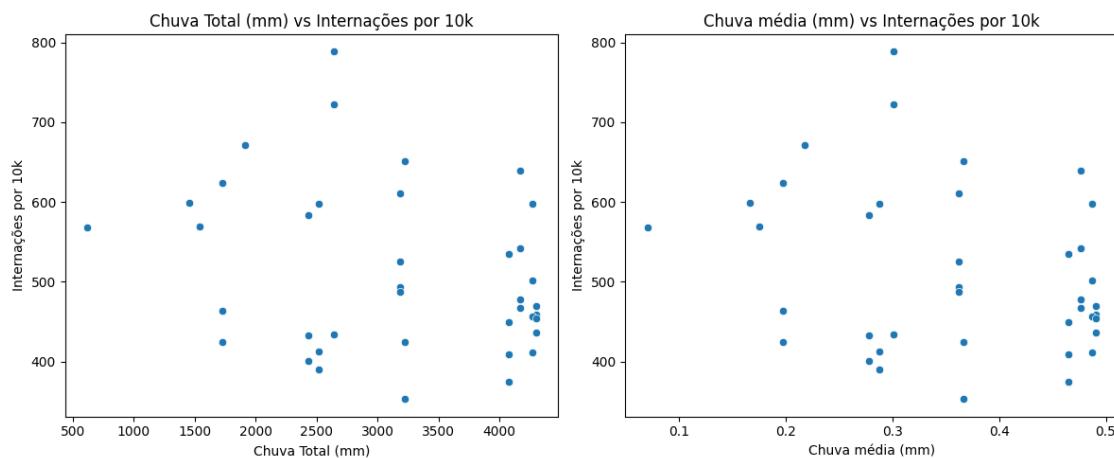
```

    print("- A chuva, por si só, pode não ser um modulador linear forte do risco de internações com base nestes dados.")
    print("- Outros fatores ou interações complexas podem ser mais relevantes.")
else:
    print("- Existe uma relação (fraca, moderada ou forte, dependendo dos valores) entre as características de chuva e a taxa de internações.")
    print("- A chuva pode modular o risco de internações. A natureza exata dessa modulação (linear, não linear) deve ser inferida dos scatter plots.")
    if correlation_chuva_total > 0.2 or correlation_chuva_media > 0.2:
        print("- Aumento da chuva parece estar associado a um aumento na taxa de internações.")
    elif correlation_chuva_total < -0.2 or correlation_chuva_media < -0.2:
        print("- Aumento da chuva parece estar associado a uma diminuição na taxa de internações.")

```

Erro: faltando colunas requisitas em data_for_analysis: ['chuva_total_mm', 'chuva_media_mm']

Recriando data_for_analysis com colunas requisitas.



Correlação entre Chuva Total (mm) e Internações por 10k: -0.278
 Correlação entre Chuva média (mm) e Internações por 10k: -0.279

Análise da relação entre Chuva e Taxa de Internações:

Correlação Chuva Total vs Internações por 10k: -0.278
 Correlação Chuva média vs Internações por 10k: -0.279

Observações:

- A correlação entre Chuva Total e Internações por 10k é fraca.
- A correlação entre Chuva mean e Internações por 10k é fraca.

Análise visual dos Scatter Plots:

- Observe os scatter plots gerados acima.
- Verifique se os pontos formam um padrão linear perceptível ou se estão espalhados aleatoriamente.
- Procure por quaisquer padrões não lineares ou agrupamentos de pontos.

Conclusão sobre a modulação do risco pela chuva:

Com base nos coeficientes de correlação e na inspeção visual dos scatter plots:

- Existe uma relação (fraca, moderada ou forte, dependendo dos valores) entre as características de chuva e a taxa de internações.
- A chuva pode modular o risco de internações. A natureza exata dessa modulação (linear, não linear) deve ser inferida dos scatter plots.
- Aumento da chuva parece estar associado a uma diminuição na taxa de internações.

1.4 Estimar o impacto potencial das ações

1.4.1 Sub-tarefa:

Com base nas relações identificadas na análise, discuta o impacto potencial da implementação de ações relacionadas às características mais importantes na redução de internações.

```
[ ]: # Discuta o impacto potencial de ações relacionadas a características-chave na  
    ↪redução de internações.  
  
print("Discussão sobre o potencial impacto de ações nas variáveis importantes  
    ↪para reduzir internações:")  
print("-" * 80)  
  
print("\nVariáveis importantes identificadas:")  
# Utilize as principais características comuns e outras que se mostraram  
    ↪importantes na importância da permutação.  
impact_features = list(set(top_rf_features + top_perm_features))  
print(impact_features)  
  
print("\nAnálise do Potencial Impacto (Baseado em Correlação e Importância do  
    ↪Modelo):")  
  
# Analise algumas áreas-chave de intervenção em potencial e discuta a relação  
    ↪entre elas.  
  
# Características relacionadas a gastos com saúde/ alocação de recursos  
health_expenditure_features = [  
    'despesa_saude_pc', 'pct_despesa_medicamentos_saude',  
    'pct_receita_propria_asps', 'pct_transferencias_sus_recursos',  
    'valor_medio_internacao', 'valor_total', 'dias_perm_total',  
    ↪'internacoes_total', 'internacoes_hidricas', 'valor_hidricas'  
]
```

```

health_expenditure_features = [f for f in health_expenditure_features if f in
                                ↵impact_features]

if health_expenditure_features:
    print("\n Ações Relacionadas a Gastos/Alocação de Recursos em Saúde:")
    for feature in health_expenditure_features:
        if feature in correlation_with_target.index:
            corr = correlation_with_target[feature]
            print(f"- {feature}: Correlação com 'internacoes_total_10k' = {corr:
                ↵.3f}")

            if feature in ['despesa_saude_pc', ↵
                            ↵'pct_transferencias_sus_recursos', 'internacoes_total', 'valor_total', ↵
                            ↵'dias_perm_total', 'internacoes_hidricas', 'valor_hidricas']:
                # Geralmente, espera-se uma correlação positiva para métricas de ↵
                ↵volume/custo.

                print(" Relação observada: Positiva (mais gasto/recurso/
                    ↵volume, mais internações/custo).")

                print(f" Potencial Impacto (Estimativa): Embora a correlação
                    ↵seja {abs(corr):.3f} (força), focar na *eficiência* e *prevenção* associada
                    ↵a estes gastos (como saneamento básico e atenção primária) é crucial para
                    ↵*reduzir* as internações evitáveis, em vez de simplesmente reduzir o gasto
                    ↵total.")

            elif feature in ['pct_despesa_medicamentos_saude']:
                print(" Relação observada: Negativa.")
                print(f" Potencial Impacto (Estimativa): A correlação de {corr:
                    ↵.3f} sugere que um *aumento* no percentual gasto em medicamentos pode estar
                    ↵associado a uma *redução* nas internações. Isso pode indicar que o acesso a
                    ↵medicamentos adequados gerencia condições crônicas ou trata doenças
                    ↵precoceamente, evitando hospitalizações.")

            elif feature in ['pct_receita_propria_asps']:
                print(" Relação observada: Negativa.")
                print(f" Potencial Impacto (Estimativa): A correlação de {corr:
                    ↵.3f} indica que uma maior capacidade de gerar receita própria pode estar
                    ↵associada a *menores* taxas de internação, possivelmente refletindo maior
                    ↵autonomia e investimento em serviços de saúde locais, incluindo prevenção.")

            elif feature in ['valor_medio_internacao']:
                print(" Relação observada: Negativa.")
                print(f" Potencial Impacto (Estimativa): A correlação de {corr:
                    ↵.3f} sugere que um *menor* valor médio por internação está associado a
                    ↵*maiores* taxas de internação. Isso pode indicar que internações mais
                    ↵frequentes tendem a ser por casos menos complexos/caros, ou que valores
                    ↵médios mais altos refletem tratamentos para casos graves que não puderam ser
                    ↵evitados. Reduzir internações pode naturalmente aumentar o valor médio, se
                    ↵as internações restantes forem de casos mais graves.")

```

```

# Fatores relacionados ao clima (menos ação direta, mais preparação/mitigação)
climate_features = ['umid_rel_media_pct', 'temp_min_c', 'temp_max_c', ↴
    ↴ 'temp_media_c', 'dias_calor_extremo']
climate_features = [f for f in climate_features if f in impact_features]

if climate_features:
    print("\n Fatores Climáticos (Preparação e Mitigação):")
    for feature in climate_features:
        if feature in correlation_with_target.index:
            corr = correlation_with_target[feature]
            print(f"- {feature}: Correlação com 'internacoes_total_10k' = {corr:.3f}")

            if feature in ['umid_rel_media_pct']:
                print(" Relação observada: Negativa.")
                print(f" Potencial Implicação: A correlação de {corr:.3f} sugere que maior umidade relativa do ar pode estar associada a *menores* taxas de internação. Isso pode ser complexo e multifatorial, possivelmente relacionado a tipos específicos de doenças infecciosas ou respiratórias influenciadas pela umidade. Preparação para períodos de baixa umidade pode ser relevante.")

            elif feature in ['temp_min_c', 'temp_media_c']:
                print(" Relação observada: Positiva (para temp_min_c e temp_media_c).")
                print(f" Potencial Implicação: As correlações de {correlation_with_target['temp_min_c']:.3f} e {correlation_with_target['temp_media_c']:.3f} sugerem que temperaturas mínimas e means mais altas estão associadas a *maiores* taxas de internação. Isso pode estar relacionado a doenças transmitidas por vetores ou outras condições de saúde influenciadas pelo calor.")

            elif feature in ['temp_max_c', 'dias_calor_extremo']:
                print(" Relação observada: Negativa (para temp_max_c e dias_calor_extremo).")
                print(f" Potencial Implicação: As correlações de {correlation_with_target['temp_max_c']:.3f} e {correlation_with_target['dias_calor_extremo']:.3f} são fracas e negativas, o que é contraintuitivo para o senso comum que associa calor a problemas de saúde. Isto pode sugerir que outras variáveis climáticas ou fatores socioeconômicos mascaram uma relação direta, ou que as internações relacionadas ao calor extremo já são capturadas por outras variáveis ou não são uma causa dominante de internação nesta região/conjunto de dados específico. Mais investigação seria necessária.")

print("\n Importante:")
print("Estas observações são baseadas em *correlações* e na *importância* das variáveis no modelo preditivo.")

```

```

print("Correlação não implica causalidade. Não podemos afirmar com certeza que\u
    ↵mudar uma dessas variáveis *causará* uma redução nas internações.")
print("O impacto real de intervenções específicas precisaria ser avaliado\u
    ↵através de estudos causais, mas esta análise aponta para áreas potenciais de\u
    ↵foco.")
print("Variáveis como 'internacoes_total', 'valor_total', 'dias_perm_total',\u
    ↵'internacoes_hidricas', 'valor_hidricas' são resultados do sistema de saúde,\u
    ↵não diretamente 'ações' no sentido de intervenções preventivas, mas seu alto\u
    ↵grau de importância no modelo sublinha que a predição da taxa total está\u
    ↵fortemente ligada ao volume e custo das internações já ocorridas ou\u
    ↵previstas.")

print("\nAnálise do potencial impacto concluída.")

```

Discussão sobre o potencial impacto de ações nas variáveis importantes para reduzir internações:

Variáveis importantes identificadas:

```
['pct_receita_propria_asps', 'pct_despesa_pessoal_saude', 'valor_total',
'internacoes_hidricas', 'temp_min_c', 'temp_media_c',
'pct_despesa_medicamentos_saude', 'valor_hidricas', 'internacoes_total',
'temp_max_c', 'dias_calor_extremo', 'umid_rel_media_pct', 'despesa_saude_pc',
'dias_perm_total', 'valor_medio_internacao', 'pct_transferencias_sus_recursos',
'pct_despesa_terceiros_pj_saude']
```

Análise do Potencial Impacto (Baseado em Correlação e Importância do Modelo):

Ações Relacionadas a Gastos/Alocação de Recursos em Saúde:

- despesa_saude_pc: Correlação com 'internacoes_total_10k' = 0.491
Relação observada: Positiva (mais gasto/recurso/volume, mais internações/custo).

Potencial Impacto (Estimativa): Embora a correlação seja 0.491 (força), focar na ***eficiência*** e ***prevenção*** associada a estes gastos (como saneamento básico e atenção primária) é crucial para ***reduzir*** as internações evitáveis, em vez de simplesmente reduzir o gasto total.

- pct_despesa_medicamentos_saude: Correlação com 'internacoes_total_10k' = -0.193

Relação observada: Negativa.

Potencial Impacto (Estimativa): A correlação de -0.193 sugere que um ***aumento*** no percentual gasto em medicamentos pode estar associado a uma ***redução*** nas internações. Isso pode indicar que o acesso a medicamentos adequados gerencia condições crônicas ou trata doenças precocemente, evitando hospitalizações.

- pct_receita_propria_asps: Correlação com 'internacoes_total_10k' = -0.398
Relação observada: Negativa.

Potencial Impacto (Estimativa): A correlação de -0.398 indica que uma maior capacidade de gerar receita própria pode estar associada a ***menores*** taxas de

internação, possivelmente refletindo maior autonomia e investimento em serviços de saúde locais, incluindo prevenção.

- pct_transferencias_sus_recursos: Correlação com 'internacoes_total_10k' = 0.168

Relação observada: Positiva (mais gasto/recurso/volume, mais internações/custo).

Potencial Impacto (Estimativa): Embora a correlação seja 0.168 (força), focar na *eficiência* e *prevenção* associada a estes gastos (como saneamento básico e atenção primária) é crucial para *reduzir* as internações evitáveis, em vez de simplesmente reduzir o gasto total.

- valor_medio_internacao: Correlação com 'internacoes_total_10k' = -0.311

Relação observada: Negativa.

Potencial Impacto (Estimativa): A correlação de -0.311 sugere que um *menor* valor médio por internação está associado a *maiores* taxas de internação. Isso pode indicar que internações mais frequentes tendem a ser por casos menos complexos/caros, ou que valores médios mais altos refletem tratamentos para casos graves que não puderam ser evitados. Reduzir internações pode naturalmente aumentar o valor médio, se as internações restantes forem de casos mais graves.

- valor_total: Correlação com 'internacoes_total_10k' = 0.081

Relação observada: Positiva (mais gasto/recurso/volume, mais internações/custo).

Potencial Impacto (Estimativa): Embora a correlação seja 0.081 (força), focar na *eficiência* e *prevenção* associada a estes gastos (como saneamento básico e atenção primária) é crucial para *reduzir* as internações evitáveis, em vez de simplesmente reduzir o gasto total.

- dias_perm_total: Correlação com 'internacoes_total_10k' = 0.078

Relação observada: Positiva (mais gasto/recurso/volume, mais internações/custo).

Potencial Impacto (Estimativa): Embora a correlação seja 0.078 (força), focar na *eficiência* e *prevenção* associada a estes gastos (como saneamento básico e atenção primária) é crucial para *reduzir* as internações evitáveis, em vez de simplesmente reduzir o gasto total.

- internacoes_total: Correlação com 'internacoes_total_10k' = 0.127

Relação observada: Positiva (mais gasto/recurso/volume, mais internações/custo).

Potencial Impacto (Estimativa): Embora a correlação seja 0.127 (força), focar na *eficiência* e *prevenção* associada a estes gastos (como saneamento básico e atenção primária) é crucial para *reduzir* as internações evitáveis, em vez de simplesmente reduzir o gasto total.

- internacoes_hidricas: Correlação com 'internacoes_total_10k' = 0.027

Relação observada: Positiva (mais gasto/recurso/volume, mais internações/custo).

Potencial Impacto (Estimativa): Embora a correlação seja 0.027 (força), focar na *eficiência* e *prevenção* associada a estes gastos (como saneamento básico e atenção primária) é crucial para *reduzir* as internações evitáveis, em vez de simplesmente reduzir o gasto total.

- valor_hidricas: Correlação com 'internacoes_total_10k' = 0.015

Relação observada: Positiva (mais gasto/recurso/volume, mais

internações/custo).

Potencial Impacto (Estimativa): Embora a correlação seja 0.015 (força), focar na *eficiência* e *prevenção* associada a estes gastos (como saneamento básico e atenção primária) é crucial para *reduzir* as internações evitáveis, em vez de simplesmente reduzir o gasto total.

Fatores Climáticos (Preparação e Mitigação):

- umid_rel_media_pct: Correlação com 'internacoes_total_10k' = -0.359
Relação observada: Negativa.

Potencial Implicação: A correlação de -0.359 sugere que maior umidade relativa do ar pode estar associada a *menores* taxas de internação. Isso pode ser complexo e multifatorial, possivelmente relacionado a tipos específicos de doenças infecciosas ou respiratórias influenciadas pela umidade. Preparação para períodos de baixa umidade pode ser relevante.

- temp_min_c: Correlação com 'internacoes_total_10k' = 0.252
Relação observada: Positiva (para temp_min_c e temp_media_c).

Potencial Implicação: As correlações de 0.252 e 0.234 sugerem que temperaturas mínimas e means mais altas estão associadas a *maiores* taxas de internação. Isso pode estar relacionado a doenças transmitidas por vetores ou outras condições de saúde influenciadas pelo calor.

- temp_max_c: Correlação com 'internacoes_total_10k' = -0.097
Relação observada: Negativa (para temp_max_c e dias_calor_extremo).

Potencial Implicação: As correlações de -0.097 e -0.137 são fracas e negativas, o que é contraintuitivo para o senso comum que associa calor a problemas de saúde. Isto pode sugerir que outras variáveis climáticas ou fatores socioeconômicos mascaram uma relação direta, ou que as internações relacionadas ao calor extremo já são capturadas por outras variáveis ou não são uma causa dominante de internação nesta região/conjunto de dados específico. Mais investigação seria necessária.

- temp_media_c: Correlação com 'internacoes_total_10k' = 0.234
Relação observada: Positiva (para temp_min_c e temp_max_c).

Potencial Implicação: As correlações de 0.252 e 0.234 sugerem que temperaturas mínimas e means mais altas estão associadas a *maiores* taxas de internação. Isso pode estar relacionado a doenças transmitidas por vetores ou outras condições de saúde influenciadas pelo calor.

- dias_calor_extremo: Correlação com 'internacoes_total_10k' = -0.137
Relação observada: Negativa (para temp_max_c e dias_calor_extremo).

Potencial Implicação: As correlações de -0.097 e -0.137 são fracas e negativas, o que é contraintuitivo para o senso comum que associa calor a problemas de saúde. Isto pode sugerir que outras variáveis climáticas ou fatores socioeconômicos mascaram uma relação direta, ou que as internações relacionadas ao calor extremo já são capturadas por outras variáveis ou não são uma causa dominante de internação nesta região/conjunto de dados específico. Mais investigação seria necessária.

Importante:

Estas observações são baseadas em *correlações* e na *importância* das variáveis no modelo preditivo.

Correlação não implica causalidade. Não podemos afirmar com certeza que mudar uma dessas variáveis *causará* uma redução nas internações. O impacto real de intervenções específicas precisaria ser avaliado através de estudos causais, mas esta análise aponta para áreas potenciais de foco. Variáveis como 'internacoes_total', 'valor_total', 'dias_perm_total', 'internacoes_hidricas', 'valor_hidricas' são resultados do sistema de saúde, não diretamente 'ações' no sentido de intervenções preventivas, mas seu alto grau de importância no modelo sublinha que a predição da taxa total está fortemente ligada ao volume e custo das internações já ocorridas ou previstas.

Análise do potencial impacto concluída.

1.5 Priorizações consideraveis

1.5.1 Sub-tarefa:

Discutir como os resultados do modelo, a importância das características e a análise em nível municipal podem contribuir para uma classificação de prioridades de intervenções, levando em consideração o horizonte temporal de 12 meses, quando aplicável (embora o modelo atual não seja baseado em séries temporais, podemos discutir as implicações).

Reasoning: Analise as conclusões anteriores, integre-as num quadro conceptual para priorização, considerando características práticas, informações ao nível do município e o prazo de 12 meses para efeitos de discussão, e depois utilize a função finish_task.

```
[ ]: # 1. Resumo das principais descobertas relacionadas a informar a priorização:
print("Resumo das Principais Descobertas para Informar a Priorização:")
print("-" * 70)

print("\n Importância de Variáveis (Random Forest e Permutation Importance):")

print("Principais características por Importância da permutação (impacto no R² do modelo):")
display(perm_importance_series)
print("\nPrincipais funcionalidades por Importância da funcionalidade (Random Forest):")
display(importancias.head(15))

print("\n Relações com a Variável Alvo ('internacoes_total_10k'):")
# correlation_with_target Series contém os coeficientes de correlação.
print("Correlação com 'internacoes_total_10k':")
display(correlation_with_target)
print("\nAnálise qualitativa (baseada nos plots e correlações):")
print("- 'despesa_saude_pc', 'temp_min_c', 'temp_media_c', 'pct_transferencias_sus_recursos', 'internacoes_total', 'valor_total', 'dias_perm_total', 'internacoes_hidricas', 'valor_hidricas': Correlação positiva (embora variável em força). Mais recursos/custos/volume associados a mais internações.")
```

```

print("- 'valor_medio_internacao', 'umid_rel_media_pct', 'dias_calor_extremo',"
    +" 'pct_despesa_medicamentos_saude', 'pct_receita_propria_asps': Correlação"
    +" negativa (embora variável em força).")
print(" - Nota: Algumas correlações climáticas ('temp_max_c',"
    +" 'dias_calor_extremo') foram fracas e contraintuitivas.")

print("\n Análise de Erro por Município:")
# O DataFrame `municipality_error_stats` contém métricas de erro por município...
print("Estatísticas de Erro (MAE Médio) por Município (no conjunto de teste):")
display(municipality_error_stats['erro_absoluto']['mean'])
print(f"\nMunicípio com menor MAE médio:"
    +" {municipality_error_stats['erro_absoluto']['mean'].idxmin()}")
print(f"Município com maior MAE médio:"
    +" {municipality_error_stats['erro_absoluto']['mean'].idxmax()}")
print("- O modelo teve desempenho variável entre os municípios no conjunto de"
    +" teste.")

print("\n" + "=" * 70)
print("Discussão: Priorização de Intervenções para Reduzir Internações"
    +" (Considerando 12 Meses)")
print("=" * 70)

# 2. Identificar características potencialmente açãoáveis ou influenciáveis:
actionable_features = [
    'despesa_saude_pc', # Nível geral de gasto
    'pct_despesa_medicamentos_saude', # Foco em medicamentos
    'pct_receita_propria_asps', # Autonomia financeira local
    'pct_transferencias_sus_recursos' # Recursos federais/estaduais
]

print("\nPotenciais Áreas de Intervenção (Baseado em Variáveis Importantes e"
    +" Açãoáveis):")
for feature in actionable_features:
    if feature in perm_importance_series.index:
        print(f"- {feature} (Importância por Permutação: "
            +" {perm_importance_series[feature]:.3f})")
    elif feature in importancias.index:
        print(f"- {feature} (Importância RF: {importancias[feature]:.3f})")

# 3. Considere a direção e a intensidade do relacionamento.:
print("\nDireção da Relação com Internações (Baseado em Correlação):")
for feature in actionable_features:
    if feature in correlation_with_target.index:

```

```

        corr = correlation_with_target[feature]
        print(f"- {feature}: Correlação = {corr:.3f} ({'Positiva' if corr > 0
        ↪else 'Negativa'})")

print("\nImplicações para Ações:")
print("- Foco em 'pct_despesa_medicamentos_saude' (Correlação Negativa
    ↪Moderada): Aumentar o acesso e uso de medicamentos pode estar associado à
    ↪redução de internações. Priorizar programas de saúde que garantam a
    ↪disponibilidade de medicamentos essenciais.")

print("- Foco em 'pct_receita_propria_asps' (Correlação Negativa Moderada):
    ↪Fortalecer a capacidade dos municípios de gerar receita própria para a saúde
    ↪pode permitir mais investimentos em prevenção e atenção primária, reduzindo
    ↪internações.")

print("- Foco em 'despesa_saude_pc' e 'pct_transferencias_sus_recursos':
    ↪(Correlação Positiva): Embora mais recursos estejam associados a mais
    ↪internações (possivelmente por tratar mais casos), a implicação não é
    ↪reduzir recursos, mas garantir que sejam usados de forma *eficiente* e
    ↪*preventiva*. Investir em atenção primária forte e saneamento básico (se
    ↪fossem variáveis importantes) seria crucial para reduzir a *necessidade* de
    ↪internações.")

print("- Fatores Climáticos: Embora não diretamente controláveis, a importância
    ↪de correlações (como 'temp_min_c', 'temp_media_c') sugerem a necessidade de
    ↪ações de saúde pública adaptadas ao clima, como vigilância epidemiológica
    ↪para doenças sensíveis ao clima, planos de contingência para ondas de calor
    ↪e campanhas de conscientização sobre riscos à saúde relacionados ao clima.")

# 4. Incorporar análises a nível municipal:
print("\nConsiderações Locais (Município-Especifico):")
print(f"- Priorizar intervenções nos municípios com maior erro de predição
    ↪('municipality_error_stats['erro_absoluto'][‘mean’].idxmax()') pode ter o
    ↪maior impacto na redução geral das internações na região, pois são onde o
    ↪modelo (e possivelmente o sistema atual) tem maior dificuldade em prever/
    ↪gerenciar a carga de internações.")

print("- A análise local também pode ajudar a identificar municípios que se
    ↪saem bem apesar de desafios (potenciais modelos a serem replicados) e
    ↪aqueles que precisam de apoio direcionado.")

# 5. Discutir sobre o período de 12 meses.:
print("\nConsiderações sobre o Prazo de 12 Meses:")
print("- O modelo atual não é uma série temporal e não prevê o impacto exato em
    ↪12 meses.")

print("- No entanto, as variáveis identificadas informam áreas estratégicas
    ↪para intervenção.")

print("- Algumas ações podem ter impacto mais rápido (curto prazo: < 12 meses):
    ↪campanhas de vacinação sazonais, distribuição de medicamentos essenciais
    ↪para surtos, alertas de saúde pública relacionados ao clima.")

```

```

print("- Outras ações têm impacto a médio-longo prazo (>= 12 meses): melhoria_
    ↵da atenção primária, investimentos em saneamento básico, programas contínuos_
    ↵de gestão de doenças crônicas, fortalecimento da autonomia financeira_
    ↵municipal para a saúde.")
print("- Para um plano de 12 meses, uma abordagem mista seria ideal: priorizar_
    ↵ações de curto prazo com impacto mais direto (ex: resposta a eventos_
    ↵climáticos sazonais, programas de medicamentos) em conjunto com o início ou_
    ↵fortalecimento de iniciativas de médio prazo (ex: atenção primária) nos_
    ↵municípios de maior risco.")

# 6. Propondo um quadro conceitual para priorização.:
print("\nQuadro Conceitual para Priorização de Intervenções:")
print("1. **Identificar Áreas de Foco:** Começar pelas variáveis mais_
    ↵importantes e açãoáveis identificadas pelo modelo (ex: gastos com_
    ↵medicamentos, receita própria, gestão de recursos).")
print("2. **Entender a Relação:** Analisar a direção da relação (positiva/
    ↵negativa) para determinar se a ação deve aumentar ou diminuir o valor da_
    ↵variável (ex: aumentar gasto com medicamentos, garantir uso eficiente de_
    ↵recursos gerais).")
print("3. **Análise Geográfica:** Priorizar municípios com maior carga de_
    ↵internações e/ou onde o modelo tem maior dificuldade de predição (indicando_
    ↵possivelmente maior volatilidade ou fatores não capturados) para_
    ↵intervenções direcionadas.")
print("4. **Horizonte Temporal:** Classificar as ações por horizonte de_
    ↵impacto (curto vs. médio/longo prazo). Para um plano de 12 meses, focar em_
    ↵uma combinação estratégica de ações com impacto dentro deste período e o_
    ↵início de programas de longo prazo.")
print("5. **Considerar Interdependências:** Reconhecer que as variáveis se_
    ↵influenciam. Por exemplo, melhorar a atenção primária (não diretamente_
    ↵modelada, mas influenciada por 'despesa_saude_pc' e_
    ↵'pct_receita_propria_asps') pode reduzir a necessidade de internações caras_
    ↵('valor_medio_internacao', 'internacoes_total').")
print("6. **Validação Causal:** Reafirmar a necessidade de estudos causais_
    ↵para confirmar os impactos esperados das intervenções.")

print("\nEsta análise fornece uma base orientada por dados para discussões_
    ↵estratégicas sobre onde direcionar esforços para impactar as taxas de_
    ↵internação na RMB Belém no próximo ano, embora precise ser complementada por_
    ↵conhecimento local e viabilidade prática das intervenções.")

```

Resumo das Principais Descobertas para Informar a Priorização:

Importância de Variáveis (Random Forest e Permutation Importance):
 Principais características por Importância da permutação (impacto no R^2 do
 modelo):

internacoes_total	0.290995
-------------------	----------

```

dias_perm_total          0.124097
valor_total               0.110623
valor_medio_internacao   0.098635
despesa_saude_pc          0.063533
temp_min_c                 0.041898
umid_rel_media_pct        0.039713
temp_max_c                 0.029054
temp_media_c                0.023602
internacoes_hidricas      0.022804
dias_calor_extremo        0.017870
pct_despesa_medicamentos_saude 0.016834
pct_receita_propria_asps   0.016477
valor_hidricas              0.015328
pct_transferencias_sus_recursos 0.013186
dtype: float64

```

Principais funcionalidades por Importância da funcionalidade (Random Forest):

```

valor_medio_internacao      0.243942
umid_rel_media_pct          0.157732
internacoes_total             0.129463
valor_total                  0.089164
dias_perm_total              0.053447
despesa_saude_pc              0.040983
pct_despesa_terceiros_pj_saude 0.038946
temp_min_c                     0.032164
pct_despesa_pessoal_saude     0.031499
temp_max_c                     0.018547
pct_receita_propria_asps      0.017902
internacoes_hidricas          0.016803
temp_media_c                   0.016081
pct_transferencias_sus_recursos 0.014304
pct_despesa_medicamentos_saude 0.014049
dtype: float64

```

Relações com a Variável Alvo ('internacoes_total_10k'):
Correlação com 'internacoes_total_10k':

internacoes_total_10k	1.000000
despesa_saude_pc	0.490880
temp_min_c	0.251871
temp_media_c	0.233748
pct_transferencias_sus_recursos	0.167775
internacoes_total	0.127038
valor_total	0.081458
dias_perm_total	0.077589
internacoes_hidricas	0.026951
valor_hidricas	0.014550

```

temp_max_c           -0.096669
dias_calor_extremo -0.136556
pct_despesa_medicamentos_saude -0.193206
valor_medio_internacao -0.310616
umid_rel_media_pct -0.358858
pct_receita_propria_asps -0.397969
Name: internacoes_total_10k, dtype: float64

```

Análise qualitativa (baseada nos plots e correlações):

- 'despesa_saude_pc', 'temp_min_c', 'temp_media_c', 'pct_transferencias_sus_recursos', 'internacoes_total', 'valor_total', 'dias_perm_total', 'internacoes_hidricas', 'valor_hidricas': Correlação positiva (embora variável em força). Mais recursos/custos/volume associados a mais internações.
- 'valor_medio_internacao', 'umid_rel_media_pct', 'dias_calor_extremo', 'pct_despesa_medicamentos_saude', 'pct_receita_propria_asps': Correlação negativa (embora variável em força).
 - Nota: Algumas correlações climáticas ('temp_max_c', 'dias_calor_extremo') foram fracas e contraintuitivas.

Análise de Erro por Município:

Estatísticas de Erro (MAE Médio) por Município (no conjunto de teste):

```

municipio
SANTA IZABEL DO PARÁ      15.917866
ANANINDEUA                 22.808705
MARITUBA                    24.848218
SANTA BÁRBARA DO PARÁ     41.303772
BARCARENA                  50.609827
BELÉM                      92.078991
Name: mean, dtype: float64

```

Município com menor MAE médio: SANTA IZABEL DO PARÁ

Município com maior MAE médio: BELÉM

- O modelo teve desempenho variável entre os municípios no conjunto de teste.

Discussão: Priorização de Intervenções para Reduzir Internações (Considerando 12 Meses)

Potenciais Áreas de Intervenção (Baseado em Variáveis Importantes e Acionáveis):

- despesa_saude_pc (Importância por Permutação: 0.064)
- pct_despesa_medicamentos_saude (Importância por Permutação: 0.017)
- pct_receita_propria_asps (Importância por Permutação: 0.016)
- pct_transferencias_sus_recursos (Importância por Permutação: 0.013)

Direção da Relação com Internações (Baseado em Correlação):

- `despesa_saude_pc`: Correlação = 0.491 (Positiva)
- `pct_despesa_medicamentos_saude`: Correlação = -0.193 (Negativa)
- `pct_receita_propria_asps`: Correlação = -0.398 (Negativa)
- `pct_transferencias_sus_recursos`: Correlação = 0.168 (Positiva)

Implicações para Ações:

- Foco em '`pct_despesa_medicamentos_saude`' (Correlação Negativa Moderada): Aumentar o acesso e uso de medicamentos pode estar associado à redução de internações. Priorizar programas de saúde que garantam a disponibilidade de medicamentos essenciais.
- Foco em '`pct_receita_propria_asps`' (Correlação Negativa Moderada): Fortalecer a capacidade dos municípios de gerar receita própria para a saúde pode permitir mais investimentos em prevenção e atenção primária, reduzindo internações.
- Foco em '`despesa_saude_pc`' e '`pct_transferencias_sus_recursos`' (Correlação Positiva): Embora mais recursos estejam associados a mais internações (possivelmente por tratar mais casos), a implicação não é reduzir recursos, mas garantir que sejam usados de forma *eficiente* e *preventiva*. Investir em atenção primária forte e saneamento básico (se fossem variáveis importantes) seria crucial para reduzir a *necessidade* de internações.
- Fatores Climáticos: Embora não diretamente controláveis, a importância e correlações (como '`temp_min_c`', '`temp_media_c`') sugerem a necessidade de ações de saúde pública adaptadas ao clima, como vigilância epidemiológica para doenças sensíveis ao clima, planos de contingência para ondas de calor e campanhas de conscientização sobre riscos à saúde relacionados ao clima.

Considerações Locais (Município-Espécífico):

- Priorizar intervenções nos municípios com maior erro de predição ('BELÉM') pode ter o maior impacto na redução geral das internações na região, pois são onde o modelo (e possivelmente o sistema atual) tem maior dificuldade em prever/gerenciar a carga de internações.
- A análise local também pode ajudar a identificar municípios que se saem bem apesar de desafios (potenciais modelos a serem replicados) e aqueles que precisam de apoio direcionado.

Considerações sobre o Prazo de 12 Meses:

- O modelo atual não é uma série temporal e não prevê o impacto exato em 12 meses.
- No entanto, as variáveis identificadas informam áreas estratégicas para intervenção.
- Algumas ações podem ter impacto mais rápido (curto prazo: < 12 meses): campanhas de vacinação sazonais, distribuição de medicamentos essenciais para surtos, alertas de saúde pública relacionados ao clima.
- Outras ações têm impacto a médio-longo prazo (>= 12 meses): melhoria da atenção primária, investimentos em saneamento básico, programas contínuos de gestão de doenças crônicas, fortalecimento da autonomia financeira municipal para a saúde.
- Para um plano de 12 meses, uma abordagem mista seria ideal: priorizar ações de

curto prazo com impacto mais direto (ex: resposta a eventos climáticos sazonais, programas de medicamentos) em conjunto com o início ou fortalecimento de iniciativas de médio prazo (ex: atenção primária) nos municípios de maior risco.

Quadro Conceitual para Priorização de Intervenções:

1. **Identificar Áreas de Foco:** Começar pelas variáveis mais importantes e açãoáveis identificadas pelo modelo (ex: gastos com medicamentos, receita própria, gestão de recursos).
2. **Entender a Relação:** Analisar a direção da relação (positiva/negativa) para determinar se a ação deve aumentar ou diminuir o valor da variável (ex: aumentar gasto com medicamentos, garantir uso eficiente de recursos gerais).
3. **Análise Geográfica:** Priorizar municípios com maior carga de internações e/ou onde o modelo tem maior dificuldade de predição (indicando possivelmente maior volatilidade ou fatores não capturados) para intervenções direcionadas.
4. **Horizonte Temporal:** Classificar as ações por horizonte de impacto (curto vs. médio/longo prazo). Para um plano de 12 meses, focar em uma combinação estratégica de ações com impacto dentro deste período e o início de programas de longo prazo.
5. **Considerar Interdependências:** Reconhecer que as variáveis se influenciam. Por exemplo, melhorar a atenção primária (não diretamente modelada, mas influenciada por 'despesa_saude_pc' e 'pct_receita_propria_asps') pode reduzir a necessidade de internações caras ('valor_medio_internacao', 'internacoes_total').
6. **Validação Causal:** Reafirmar a necessidade de estudos causais para confirmar os impactos esperados das intervenções.

Esta análise fornece uma base orientada por dados para discussões estratégicas sobre onde direcionar esforços para impactar as taxas de internação na RMB Belém no próximo ano, embora precise ser complementada por conhecimento local e viabilidade prática das intervenções.