import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

# Assume `paths\_counts` is your DataFrame containing journey data.

# Example DataFrame:

# paths\_counts = pd.DataFrame({

# 'journey\_name': ['Journey 1', 'Journey 1', 'Journey 2', 'Journey 2', 'Journey 3'],

# 'path': ['Path A', 'Path B', 'Path A', 'Path C', 'Path D'],

# 'count': [100, 200, 150, 50, 300]

# })

# Step 1: Filter the top N paths and journeys

top\_n\_paths = 10 # Change this value as needed

top\_n\_journeys = 10 # Change this value as needed

# Get the top N paths

top\_paths = paths\_counts.groupby('path')['count'].sum().nlargest(top\_n\_paths).index

# Get the top N journeys

top\_journeys = paths\_counts.groupby('journey\_name')['count'].sum().nlargest(top\_n\_journeys).index

# Filter the DataFrame

filtered\_data = paths\_counts[

(paths\_counts['path'].isin(top\_paths)) & (paths\_counts['journey\_name'].isin(top\_journeys))

]

# Step 2: Create the pivot table for the heatmap

heatmap\_data = filtered\_data.pivot(index='journey\_name', columns='path', values='count')

heatmap\_data.fillna(0, inplace=True) # Replace NaNs with 0 for missing counts

# Step 3: Plot the Heatmap

plt.figure(figsize=(12, 8)) # Adjust figure size for better readability

sns.heatmap(

heatmap\_data,

annot=True, # Annotate each cell with the count

fmt='.0f', # Display counts as integers

cmap='YlGnBu', # Use a color map

cbar\_kws={'label': 'Frequency'}

)

plt.title("Heatmap of Paths for Each Journey", fontsize=16, pad=20)

plt.xlabel("Paths", fontsize=12)

plt.ylabel("Journeys", fontsize=12)

plt.xticks(rotation=45, ha='right', fontsize=10)

plt.yticks(fontsize=10)

plt.tight\_layout()

plt.show()