# Load libraries

import pandas as pd

import matplotlib.pyplot as plt

# Load the data

flights2022 = pd.read\_csv("flights2022.csv")

flights\_weather2022 = pd.read\_csv("flights\_weather2022.csv")

# Create route column

flights2022["route"] = flights2022["origin"] + "-" + flights2022["dest"]

# Calculate mean departure delay and number of canceled flights for each unique flight route

routes\_delays\_cancels = flights2022.groupby("route").agg(

mean\_dep\_delay=("dep\_delay", "mean"),

total\_cancellations=("dep\_time", lambda x: x.isna().sum())

).reset\_index()

# Identify routes with the highest mean departure delays

top\_routes\_by\_delay = routes\_delays\_cancels.sort\_values("mean\_dep\_delay", ascending=False).head(9)

# Identify routes with the highest number of cancellations

top\_routes\_by\_cancellations = routes\_delays\_cancels.sort\_values("total\_cancellations", ascending=False).head(9)

# Create a bar graph for highest number of cancellations

top9\_route\_cancels\_bar, ax = plt.subplots()

ax.bar(top\_routes\_by\_cancellations["route"], top\_routes\_by\_cancellations["total\_cancellations"])

ax.set\_xlabel("Route")

ax.set\_ylabel("Total Cancellations")

ax.set\_title("Routes with Highest Number of Cancellations")

ax.set\_xticklabels(top\_routes\_by\_cancellations["route"], rotation=90)

plt.show()

plt.close()

# Find mean departure delays and total cancellations by airline

airlines\_delays\_cancels = flights2022.groupby("airline").agg(

mean\_dep\_delay=("dep\_delay", "mean"),

total\_cancellations=("dep\_time", lambda x: x.isna().sum())

).reset\_index()

# Identify airlines with the highest mean departure delay

top\_airlines\_by\_delay = airlines\_delays\_cancels.sort\_values("mean\_dep\_delay", ascending=False).head(9)

# Identify airlines with the highest number of cancellations

top\_airlines\_by\_cancellations = airlines\_delays\_cancels.sort\_values("total\_cancellations", ascending=False).head(9)

# Create a bar graph for highest mean depearture delay

top9\_airline\_delays\_bar, ax = plt.subplots()

ax.bar(top\_airlines\_by\_delay["airline"], top\_airlines\_by\_delay["mean\_dep\_delay"])

ax.set\_xlabel("Airline")

ax.set\_ylabel("Mean Departure Delay")

ax.set\_title("Airlines with Highest Mean Departure Delays")

ax.set\_xticklabels(top\_airlines\_by\_delay["airline"], rotation=90)

plt.show()

# Are departure delays impacted by 10+ mph winds from each airport

flights\_weather2022["group"] = flights\_weather2022["wind\_gust"].apply(lambda x: ">= 10mph" if x >= 10 else "< 10 mph")

wind\_grouped\_data = flights\_weather2022.groupby(["group", "origin"]).agg(

mean\_dep\_delay=("dep\_delay", "mean")

)

print(wind\_grouped\_data)

wind\_response = True