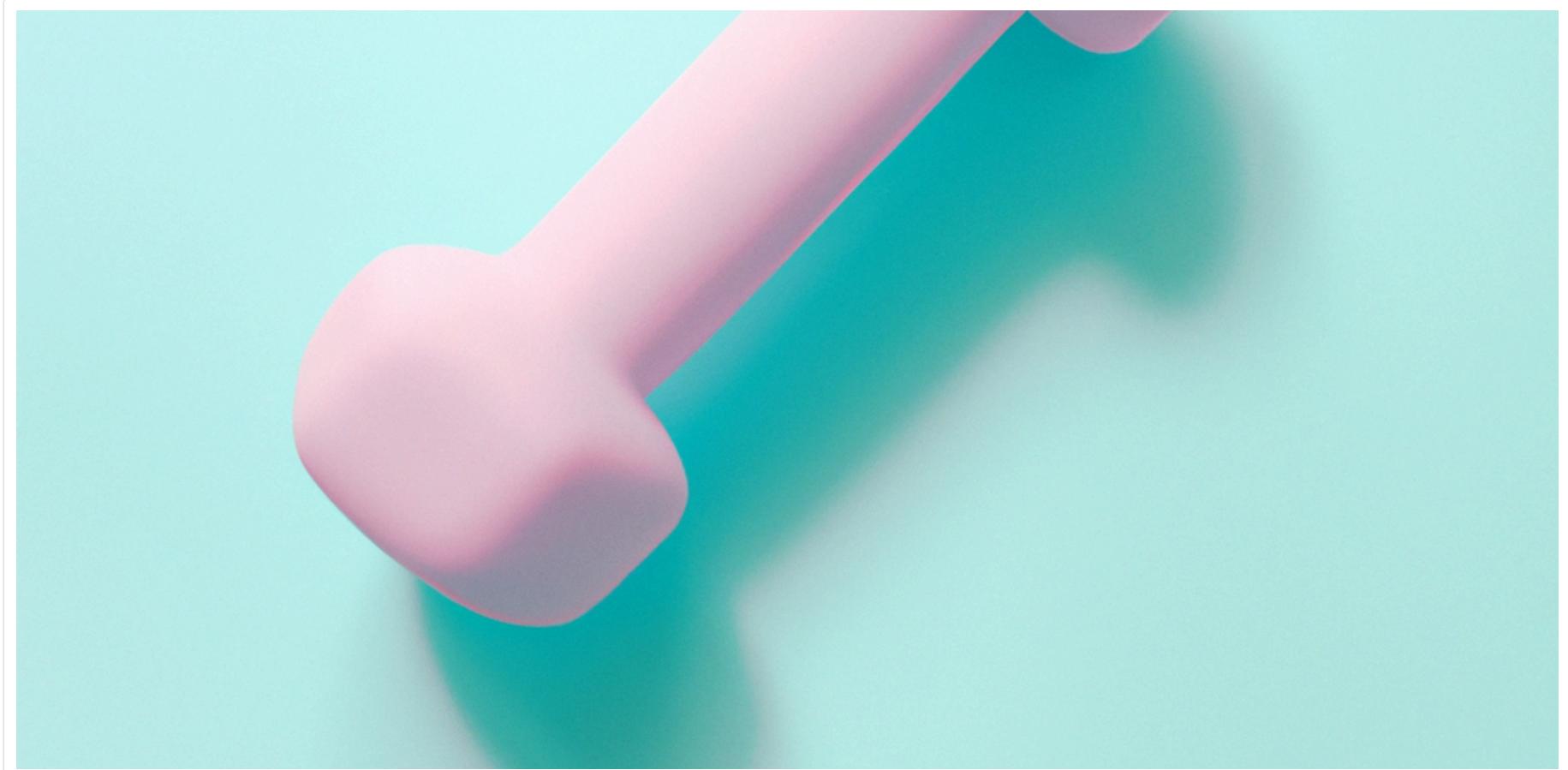


PROJECT: DATA-DRIVEN PRODUCT MANAGEMENT: CONDUCTING A MARKET ANALYSIS



You are a product manager for a fitness studio and are interested in understanding the current demand for digital fitness classes. You plan to conduct a market analysis in Python to gauge demand and identify potential areas for growth of digital products and services.

The Data

You are provided with a number of CSV files in the "Files/data" folder, which offer international and national-level data on Google Trends keyword searches related to fitness and related products.

workout.csv

Column	Description
'month'	Month when the data was measured.
'workout_worldwide'	Index representing the popularity of the keyword 'workout', on a scale of 0 to 100.

three_keywords.csv

Column	Description
'month'	Month when the data was measured.
'home_workout_worldwide'	Index representing the popularity of the keyword 'home workout', on a scale of 0 to 100.
'gym_workout_worldwide'	Index representing the popularity of the keyword 'gym workout', on a scale of 0 to 100.
'home_gym_worldwide'	Index representing the popularity of the keyword 'home gym', on a scale of 0 to 100.

workout_geo.csv

Column	Description
'country'	Country where the data was measured.
'workout_2018_2023'	Index representing the popularity of the keyword 'workout' during the 5 year period.

three_keywords_geo.csv

Column	Description
'country'	Country where the data was measured.
'home_workout_2018_2023'	Index representing the popularity of the keyword 'home workout' during the 5 year period.
'gym_workout_2018_2023'	Index representing the popularity of the keyword 'gym workout' during the 5 year period.
'home_gym_2018_2023'	Index representing the popularity of the keyword 'home gym' during the 5 year period.

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the workout trends data
workout = pd.read_csv('data/workout.csv')

workout['month'] = pd.to_datetime(workout['month'])
peak_row = workout.loc[workout['workout_worldwide'].idxmax()]
year_str = str(peak_row['month'].year)
print("Peak year for 'workout' searches:", year_str)

# Load and format
three_keywords = pd.read_csv('data/three_keywords.csv')
three_keywords['month'] = pd.to_datetime(three_keywords['month'])

# Filter for 2020 (COVID peak year)
covid_data = three_keywords[three_keywords['month'].dt.year == 2020]

# Calculate the column with highest average interest during 2020
peak_covid = covid_data[['home_workout_worldwide', 'gym_workout_worldwide', 'home_gym_worldwide']].mean().idxmax()

print("Most popular keyword during COVID:", peak_covid)

# Get the latest date in the data
latest_month = three_keywords['month'].max()

# Filter for the most recent 6 months
recent_data = three_keywords[three_keywords['month'] >= (latest_month - pd.DateOffset(months=6))]

# Get most popular keyword currently
current = recent_data[['home_workout_worldwide', 'gym_workout_worldwide', 'home_gym_worldwide']].mean().idxmax()

print("Most popular keyword now:", current)
# Set plot size
plt.figure(figsize=(12, 6))

# Plot each keyword
plt.plot(three_keywords['month'], three_keywords['home_workout_worldwide'], label='Home Workout')
plt.plot(three_keywords['month'], three_keywords['gym_workout_worldwide'], label='Gym Workout')
plt.plot(three_keywords['month'], three_keywords['home_gym_worldwide'], label='Home Gym')

# Customize the chart
plt.title('Global Search Interest Over Time')
plt.xlabel('Month')
plt.ylabel('Search Interest')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()

# Load data
workout_geo = pd.read_csv('data/workout_geo.csv')

# Filter for the 3 countries
```

```
filtered = workout_geo[workout_geo['country'].isin(['United States', 'Australia', 'Japan'])]

# Compute the average interest for each
top_country = filtered.groupby('country')[['workout_2018_2023']].mean().idxmax()

print("Country with highest interest:", top_country)

# Load the dataset
three_keywords_geo = pd.read_csv('data/three_keywords_geo.csv')

# Filter for Philippines and Malaysia
geo = three_keywords_geo[three_keywords_geo['Country'].isin(['Philippines', 'Malaysia'])]

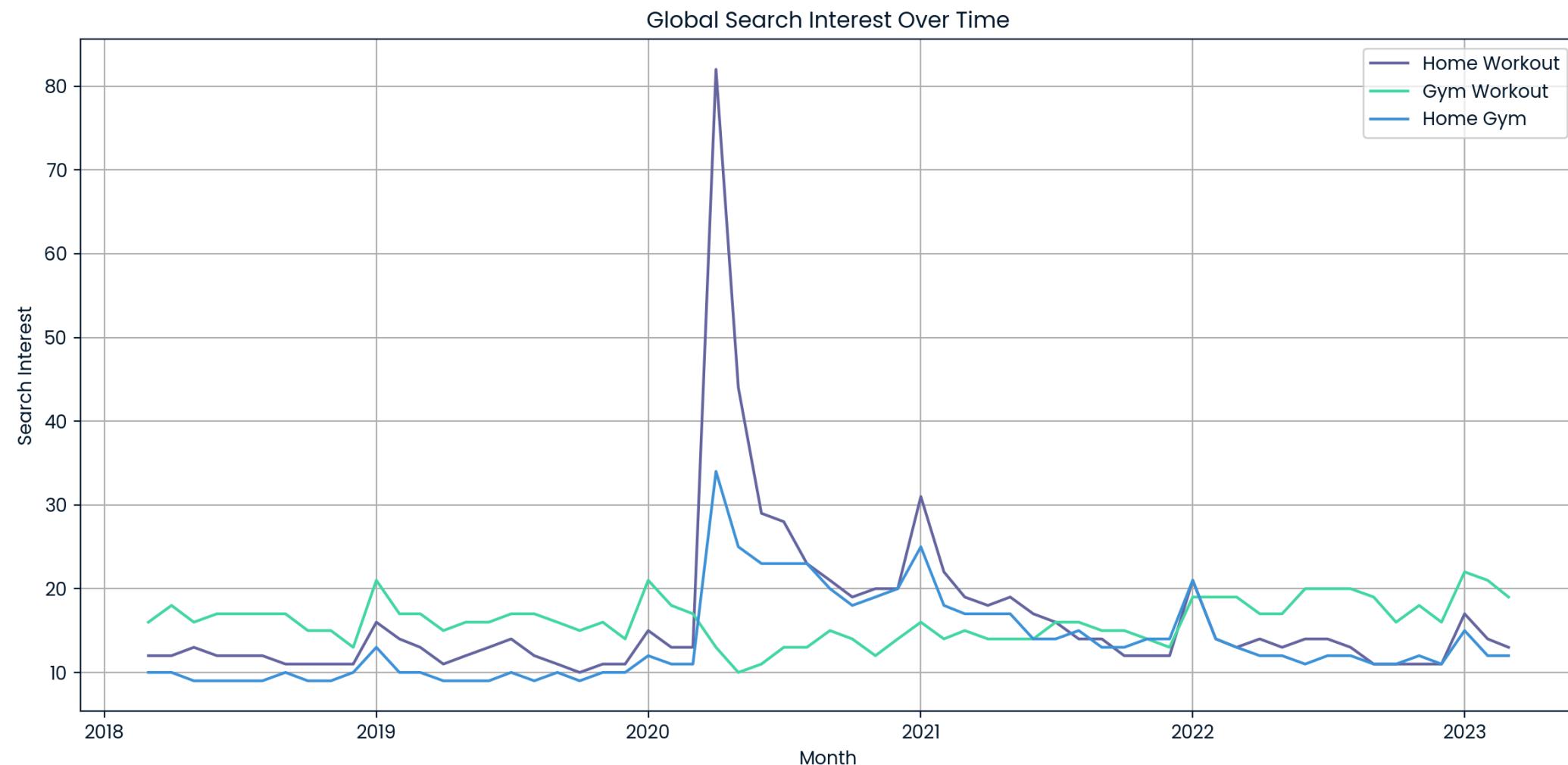
# Calculate the average home workout interest per country
home_workout_geo = geo.groupby('Country')[['home_workout_2018_2023']].mean().idxmax()

# Output the result
print("Country with highest home workout interest:", home_workout_geo)
```

Peak year for 'workout' searches: 2020

Most popular keyword during COVID: home_workout_worldwide

Most popular keyword now: gym_workout_worldwide



Country with highest interest: United States

Country with highest home workout interest: Philippines