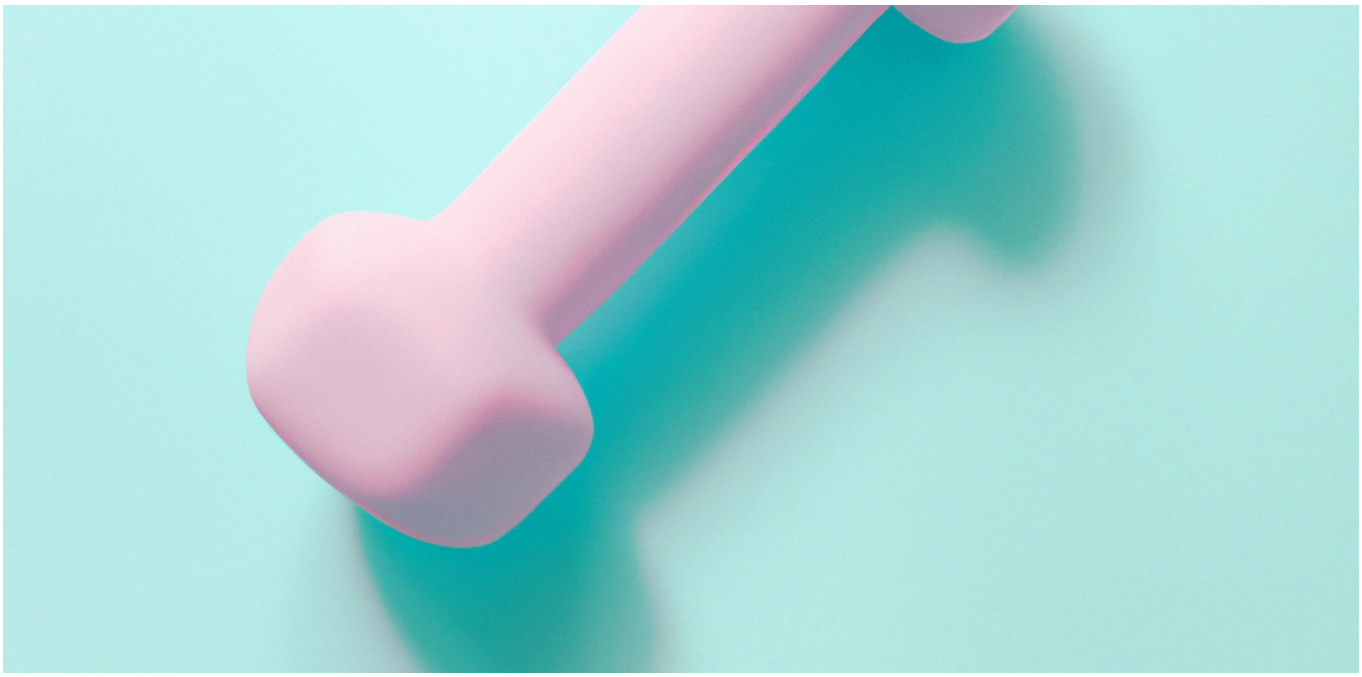
PROJECT: DATA-DRIVEN PRODUCT MANAGEMENT: CONDUCTING A MARKET ANALYSIS



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 the necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
```

```
import datetime as dt
import time as tm
```

```
# Load data on global interest in workouts
workouts_df = pd.read_csv('data/workout.csv', parse_dates = ['month'])
workouts_df.head()
```

index	month	workout_worldwide
0	2018-03-01T00:00:00.000	
1	2018-04-01T00:00:00.000	
2	2018-05-01T00:00:00.000	
3	2018-06-01T00:00:00.000	
4	2018-07-01T00:00:00.000	

Rows: 5 Expand

```
year = workouts_df.sort_values('workout_worldwide', ascending=False)
year.head()

print(year.month.dt.year)#year) #to_datetime(month))#.dt.year)

#print(year.month.to_datetime(parts))#.dt.year)
```

```
25    2020
26    2020
34    2021
27    2020
46    2022
      ...
45    2021
19    2019
20    2019
55    2022
57    2022
Name: month, Length: 61, dtype: int64
```

```
# Group by year and sum the 'workout_worldwide' values
workouts_per_year = workouts_df.groupby(workouts_df['month'].dt.year)['workout_worldwide'].sum()
workouts_per_year
```

	workout_worl...
2018	536
2019	646
2020	773
2021	685
2022	646
2023	174

Rows: 6 Expand

```
sorted = workouts_per_year.sort_values(ascending=False)
sorted.index[0]
```

```
2020
```

```
#store year where global search for 'workout' at peak'
year_str =str(sorted.index[0])

print(year_str)

"""

top_country =                #country w/ highest interest for workouts

#expanding virtual home workout service to either Philippines or Malaysia
home_workout_geo =             # higher interest in home workouts between the 2 countries
"""
```

```
2020
```

```
'\n\ntop_country =                #country w/ highest interest for workouts \n\n#expanding virtual home workout service to either Philippines or Malaysia\nhome_workout_geo =             # higher interest in home workouts between the 2 countries \n'
```

```
keywords3_df = pd.read_csv('data/three_keywords.csv', parse_dates = ['month'])
keywords3_df.head()
```

	month	home_workout_world...	gym_workout_world...	home_gym_wo...
0	2018-03-01T00:00:00.000	12	16	
1	2018-04-01T00:00:00.000	12	18	
2	2018-05-01T00:00:00.000	13	16	
3	2018-06-01T00:00:00.000	12	17	
4	2018-07-01T00:00:00.000	12	17	

Rows: 5 Expand

```
keywords3_years = keywords3_df.groupby(keywords3_df['month'].dt.year)
['gym_workout_worldwide','home_workout_worldwide', 'home_gym_worldwide'].sum()
keywords3_years
#most popular keywords of the available during covid &now
#peak_covid, current =
```

	gym_workout_world...	home_workout_world...	home_gym_wo...
2018	161	117	94
2019	197	148	118
2020	171	327	239
2021	176	206	191
2022	220	160	152
2023	62	44	39

Rows: 6 Expand

```
keywords3_years.index#
```

```
Int64Index([2018, 2019, 2020, 2021, 2022, 2023], dtype='int64', name='month')
```

```
keywords3_years.iloc[0:3]
```

	gym_workout_world...	home_workout_world...	home_gym_wo...
2018	161	117	94
2019	197	148	118
2020	171	327	239

Rows: 3 Expand

```
keywords3_years.loc[2020]
```

index	
gym_workout_worldwide	171
home_workout_worldwide	327
home_gym_worldwide	239

Rows: 3 Expand

```
peak_covid = keywords3_years.loc[2020].sort_values(ascending=False).index[0]
current = keywords3_years.loc[2023].sort_values(ascending=False).index[0]
current, peak_covid
```

```
('gym_workout_worldwide', 'home_workout_worldwide')
```

```
countries_df = pd.read_csv('data/workout_geo.csv', index_col=0)#parse_dates=['month'])
countries_df.head()
```

country	workout_2018...
Guam	
Falkland Islands (Islas Malvinas)	
Cook Islands	
Brunei	
Palau	

Rows: 5 Expand

```
country_x=['United States', 'Australia', 'Japan']
#countries_df.index[]
```

```
#countries_df.loc['United States']
i = countries_df.loc[country_x]
i
```

country	workout_2018...
United States	100
Australia	77
Japan	1

Rows: 3 Expand

```
top_country = i.sort_values(by='workout_2018_2023', ascending=False).index[0]
top_country
```

```
'United States'
```

```
geowords_df = pd.read_csv('data/three_keywords_geo.csv', index_col=0)#parse_dates=['month'])
geowords_df.head()
```

C	home_workout_2018_...	gym_workout_2018_...	home_gym_201...
Gibraltar			
Lesotho			
Guam			
Botswana			
Brunei			

Rows: 5 Expand

```
mesa = geowords_df.loc[['Philippines','Malaysia']]
mesa
```

Co...	home_workout_2018_...	gym_workout_2018_...	home_gym_201...
Philippines	52	38	10
Malaysia	47	38	15

Rows: 2 Expand

```
home_workout_geo = mesa[['home_workout_2018_2023']].sort_values(by='home_workout_2018_2023',
ascending=False).index[0]
home_workout_geo
```

```
'Philippines'
```