Impact of Gym Accessibility on Training Preferences

Kassidy M, Freddy C, Sanmi L, Shriya Y

Introduction to our research topic

Project Description:

We aim to investigate how the relationship between gym accessibility factors may affect an individual's training preferences.
Decided training categories

- Gym/weight training
- At-home training
- Endurance/marathon training

Factors that got evaluated

- Cost of gym memberships
- Means of transportation
- Income levels by state

Training Preferences by Gym Membership Cost

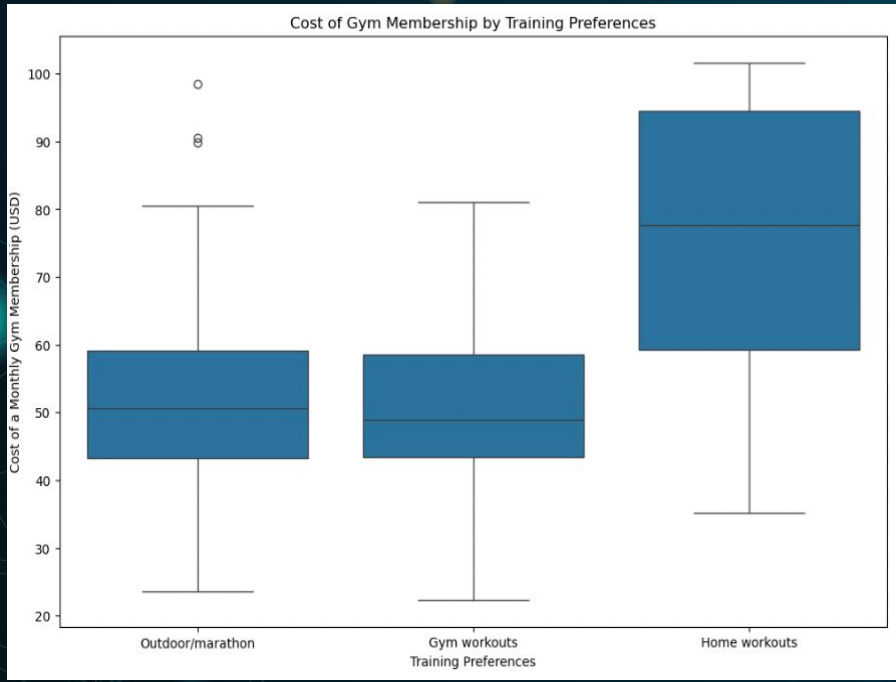
1. Loading and Preparing Data: CSV data file from Kaggle.
2. The data set, Healthy Lifestyle Cities 2021, had 10 total columns. For my part of the project I only used 2 columns: “Membership Cost” and “Outdoor Activities”
3. Cleaned the membership cost column
4. I created a training preferences column based on the number of hours spent outdoors.

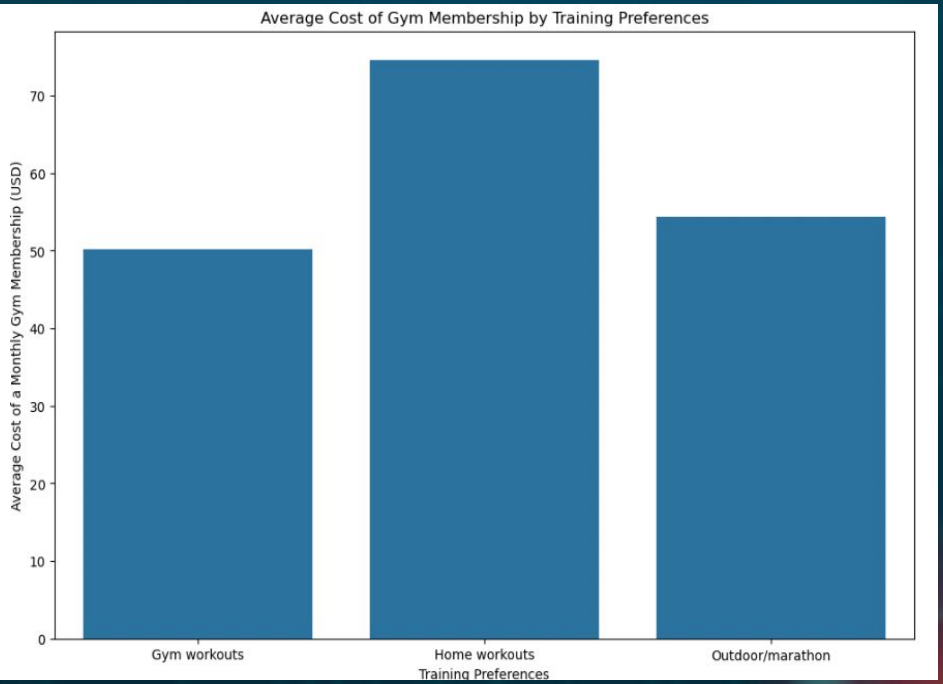
```
# Step 3: Create a new column 'Training Preferences' based on 'Outdoor activities(City)' and cost of membership

def classify_training_preferences(row):
    # Extract values
    outdoor_activities = row['Outdoor activities(City)']
    cost = row['Cost of a monthly gym membership(City)']

    # Define thresholds for classification based on outdoor activities
    if outdoor_activities > data['Outdoor activities(City)'].median():
        return 'Outdoor/marathon'
    elif outdoor_activities < data['Outdoor activities(City)'].median() / 2:
        return 'Home workouts'
    else:
        return 'Gym workouts'
```

Healthy lifestyle metrics of top 44 cities.

1. Sunshine hours(City)
2. Cost of a bottle of water(City)
3. Obesity levels(Country)
4. Life expectancy(years) (Country)
5. Pollution(Index score) (City)
6. Annual avg. hours worked
7. Happiness levels(Country)
8. Outdoor activities(City)
9. Number of take out places(City)
10. Cost of a monthly gym membership(City)

(Any graphs for Cost vs Training preferences)





ANOVA test

Interpretation

- F-statistic: 3.8765
- p-value: 0.0287

```
F-statistic: 3.8765
p-value: 0.0287
```

Conclusion

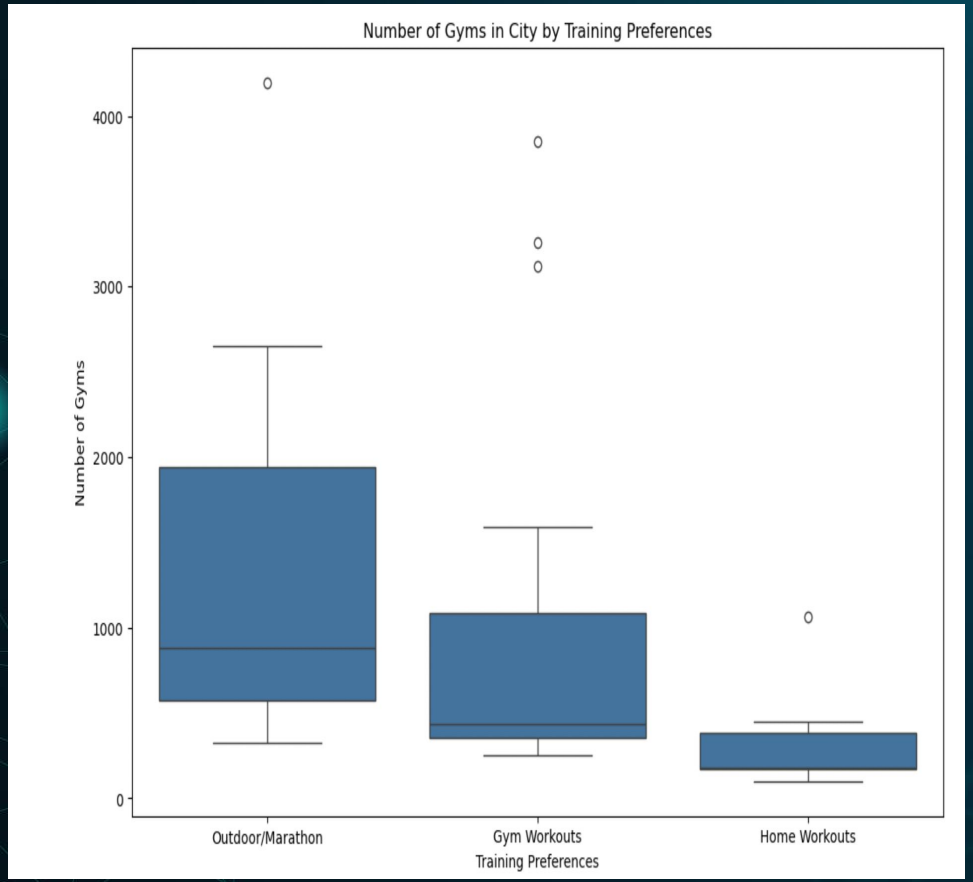
This indicates that you can reject the null hypothesis, suggesting that there is a statistically significant difference in gym membership costs among the different training preference groups (Gym workouts, Outdoor/marathon, Home workouts)

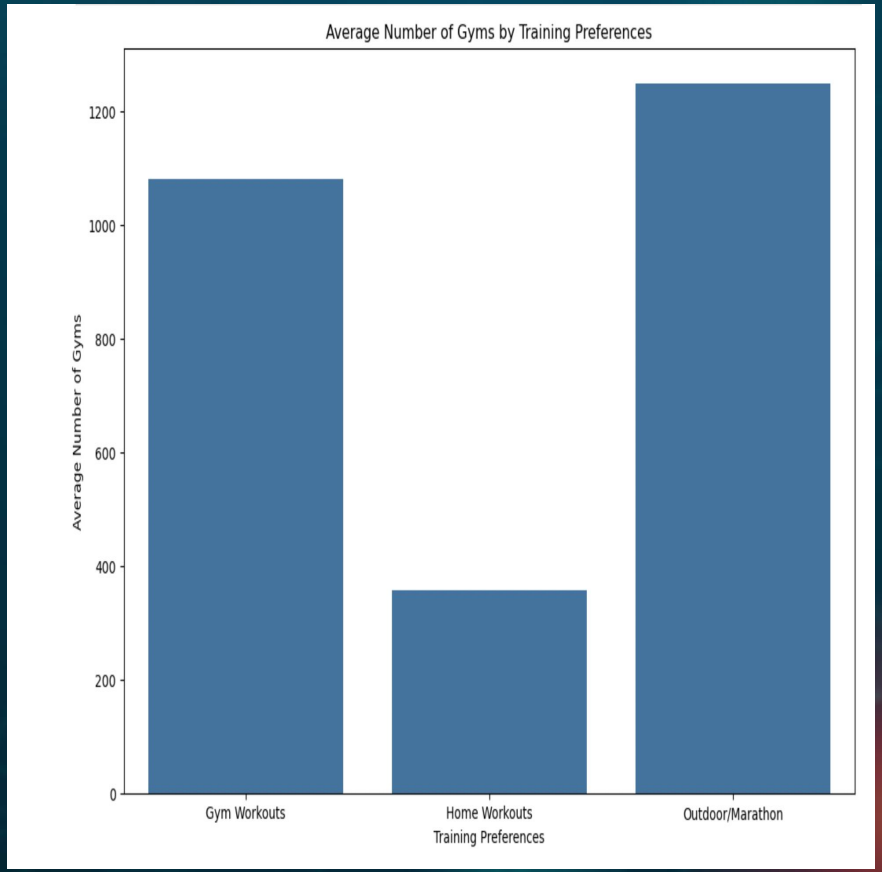
Location

```
data = [
    ['Amsterdam',478],
    ['Sydney',1633],
    ['Vienna',250],
    ['Stockholm',586],
    ['Copenhagen',340],
    ['Helsinki',371],
    ['Fukoka',415],
    ['Berlin',573],
    ['Barcelona',485],
    ['Vancouver',601],
    ['Melbourne',327],
    ['Beijing',2044],
    ['Bangkok',620],
    ['Buenos Aires',500],
    ['Toronto',433],
    ['Madrid',575],
    ['Jakarta',418],
    ['Seoul',456],
    ['Frankfurt',448],
    ['Geneva',100],
    ['Tel Aviv',419],
    ['Istanbul',2423],
    ['Cairo',881],
    ['Taipei',331],
    ['Los Angeles',2406],
    ['Mumbai',3255],
    ['Boston',1063],
    ['Dublin',255],
    ['Tokyo',886],
    ['Chicago',551],
    ['Hong Kong',1189],
```

1. Collected data for each city
2. Set up a data frame showing each city and number of gym locations per city.
3. Using prior code to calculate the training preferences
4. Merged data sets

VISUAL REPRESENTATION





ANOVA TEST

Interpretation :

F-Statistic: 1.731846

P-Value: 0.18996

```
outdoor = merged_df[merged_df['Training Preferences']== 'Outdoor/Marathon']['Number of Gyms']
gym = merged_df[merged_df['Training Preferences']== 'Gym Workouts']['Number of Gyms']
home = merged_df[merged_df['Training Preferences']== 'Home Workouts']['Number of Gyms']
stats.f_oneway(outdoor, gym, home)
```

```
F_onewayResult(statistic=1.7318462745669077, pvalue=0.18996230363946046)
```

Conclusion:

This indicates that you cannot reject the null hypothesis, suggesting that there is not a statistically significant difference in gym membership costs among the different training preference groups (Gym workouts, Outdoor/marathon, Home workouts)

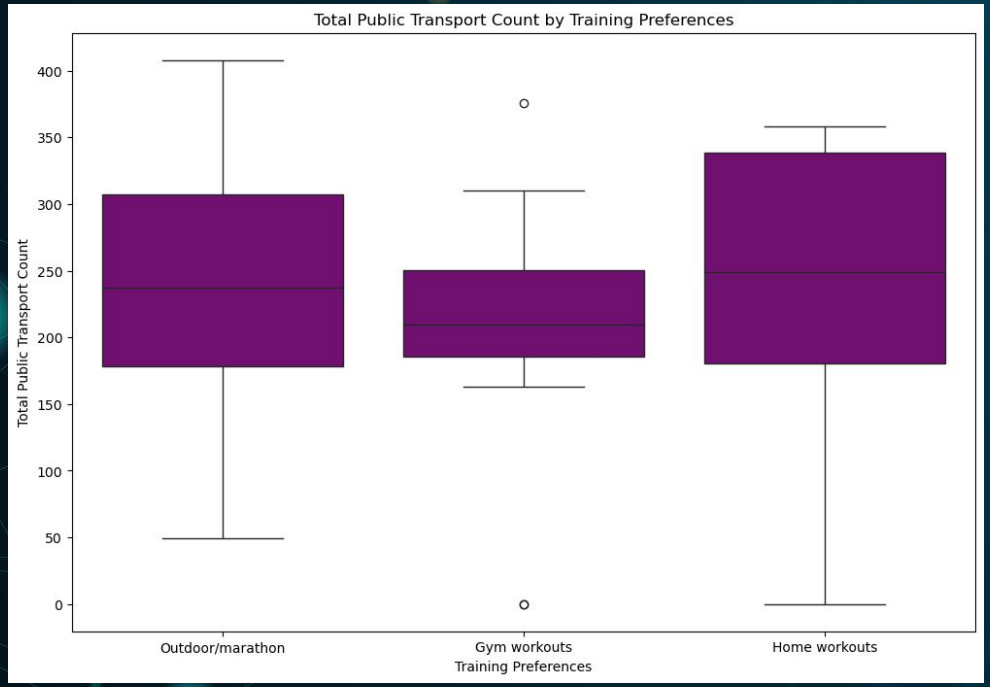
Transportation

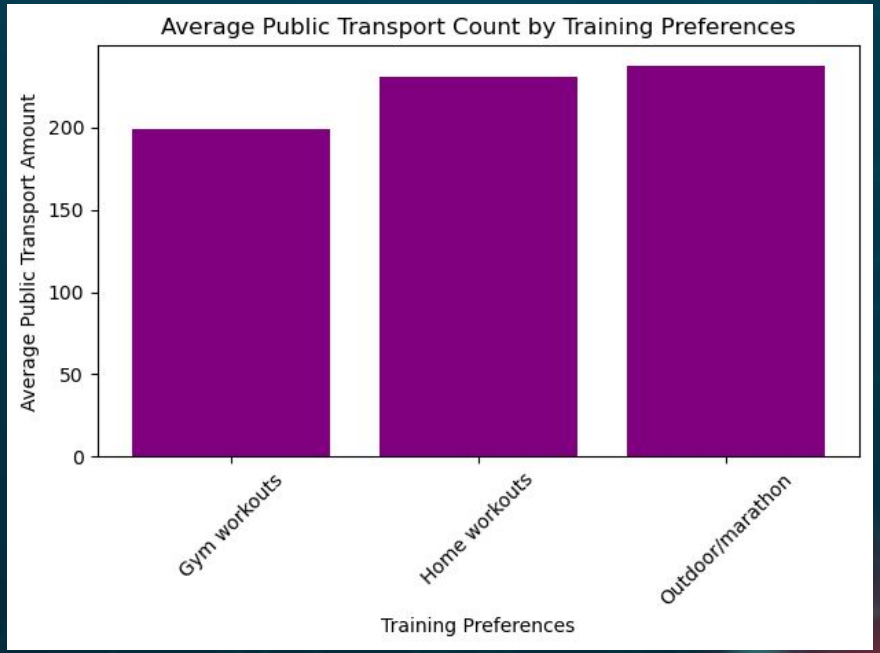
1. Get City Coordinates using Geoapify API Geocode
2. Calculate the number of public transportation locations using Geoapify API Places
3. Using prior code to calculate the training preferences

```python
# Create a function to get each of the city coordinates
def get_city_coordinates(city_name):
    # Set base url for geoapify_geocode
    geoapify_geocode_url = "https://api.geoapify.com/v1/geocode/search"

    # Set parameters dictionary
    params = {
        "text": city_name,
        "apiKey": geoapify_key,
        "limit": 1
    }
```

```python
# Create a function for the if/else statement to get number of public transportation
def num_of_public_transport(lat, lon, radius = 5000):

    # Set base url for geoapify_places
    geoapify_places_url = "https://api.geoapify.com/v2/places"

    # List the public transport categories we want to use
    transport_categories = [
        "public_transport.light_rail",
        "public_transport.monorail",
        "public_transport.subway",
        "public_transport.bus",
        "public_transport.tram"
    ]
    total_public_transport = 0
```

Visual representation(s)





Statistical Test

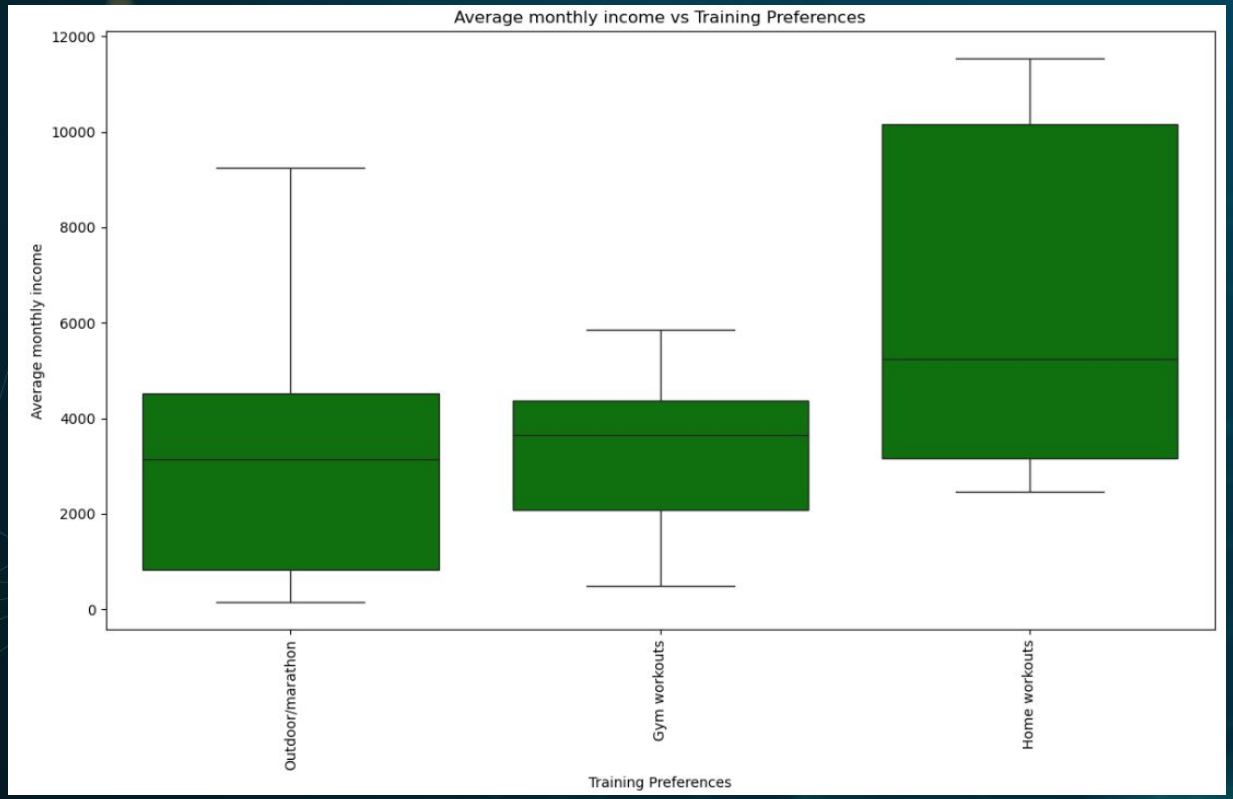
ANOVA Table
Why: We are testing against 3 categories: Gym, Home, and Outdoor Workouts

P-value: 0.6048

Conclusion: We cannot say that the amount of public transportation affects how people train because it is not statistically significant according to our P-value.

```
outdoor = combined_df[combined_df["Training Preferences"]== "Outdoor/marathon"]["Total Public Transport Count"]
gym = combined_df[combined_df["Training Preferences"]== "Gym workouts"]["Total Public Transport Count"]
home = combined_df[combined_df["Training Preferences"]== "Home workouts"]["Total Public Transport Count"]
stats.f_oneway(outdoor, gym, home)
```

Training Preferences by Income Level



Statistical Test

- F-statistic: 4.68
- P-value = 0.01535

```
F_onewayResult(statistic=4.685430627690373, pvalue=0.015353111161191245)
```

Conclusion

Analysis Tools:

- Visualization: Employed bar graphs and box-and-whisker plots to depict relationships between factors and training preferences.
- Statistical Test: Used an ANOVA test to determine the statistical significance of each factor.

Key Findings:

- Non-Significant Factors: Number of gym locations and available transportation showed no statistical significance (p-value > .05).
- Significant Factor: Membership cost (p-value = .0287) and income level (p-value = .01535) were statistically significant , influencing training method preference.

Conclusion: Membership cost significantly impacts training preferences in urban areas.

Credits

- Healthy Lifestyle Cities Report 2021 (kaggle.com) for Costs and City Names
- Geoapify API for Transportation
- Google for Locations and Income