
TERM PROJECT DOCUMENTATION |

ANALYSIS OF GYM USAGE AND DEMOGRAPHIC DATA

GROUP 3: Cansu Karabulut (2203973) - Mathilde Mayer (2408329) - Seneca Miller (2405932) - Anna Pavlova (2303217)

Prepared for: Mr. László Salló

CONTENTS

ABSTRACT	3
1. INTRODUCTION.....	3
2. DATA & METHODOLOGY.....	3
2.1 Datasets	3
2.1.1 Gym Data	3
2.1.2 Census Data.....	3
2.1.3. External Factors Data (API).....	4
2.2 Methodology.....	4
3. ANALYSIS & FINDINGS.....	4
3.1 Revenue analyses	4
3.2 Sport popularity analyses.....	4
3.3 Demographic features analyses	4
3.4 External factors analyses	4
4. CONCLUSIONS & RECOMMENDATIONS.....	5

ABSTRACT

This study focuses on how demographic and external factors influence gym usage patterns in the United States in 2023. The datasets include information such as user age, gender and location, alongside workout types, program category and calories burned. Findings reveal that men subscribing more to gym memberships, although the difference is small in some cities, and preferences for popular workout types vary among user demographics.

1. INTRODUCTION

The gym industry plays a crucial role in improving physical and mental health. Understanding user behavior is vital for optimizing the expansion of sports facilities.. This study addresses the following questions:

- How much revenue do gyms generate in the United States, depending on the type of programme (Pro, Basic or Student)?
- What are the most popular sports in different regions?
- Do user demographic features influence gym membership rates?
- What external factors can influence gym attendance?

The results provide actionable insights for gym businesses to implement targeted marketing strategies.

2. DATA & METHODOLOGY

2.1 Datasets

2.1.1 Gym Data

- Columns: user_id, gym_id, checkin_time, checkout_time, workout_type, calories_burned, location, gym_type, facilities, Month (Name), Day of Week (Name)
- Description: Contains user activity and usage frequency details.

2.1.2 Census Data

- Columns: user_id, first_name, last_name, age, gender, birthdate, sign_up_date, user_location, subscription_plan, price_per_month, features, State, Median Age, Male Population, Female Population, Total Population, Average Household Size, State Code
- Description: Includes demographic and regional data in the United States.

2.1.3. External Factors Data (API)

- Columns: moon_phases, avgtempCs, Values, Dates
- Description : Contains meteorological and astrological data in New York in 2023.

2.2 Methodology

- **Importating the data** on KNIME : The gym dataset was imported from MySQL with the node “MySQL Connector”; the census dataset was directly imported on KNIME with a “CSV reader” and the weather data comes from the import of an API.
- **Cleaning the data** by removing the missing values, applying the right format (“String to Date&Time”) and applying filters.
- **Manipulating the data** by joining tables together (“Joiner”, “Concatenate”) and creating new columns (“Math Formula”).
- **Visualyizing the data** by producing graphs (“Bar Chart”, “Scatter Plot”, “Line Plot”, “Pie Chart”) in order to answer the questions set out in the introduction.

3. ANALAYSIS & FINDINGS

3.1 Revenue analyses

- Most of the revenue earned by gyms in the US comes from ‘Pro’ programmes (around 50% of revenue).
- The ‘Student’ programmes are the least profitable, but are nonetheless an interesting way of making younger customers more reliable and encouraging them to sign up to other programmes at a later date.

3.2 Sport popularity analyses

- The popularity of sports varies from city to city in the United States.
- There is great interest in weighlfling in Houston and Philadelphia, cardio in Los Angeles and pilates in Chicago.

3.3 Demographic features analyses

- More men than women take out gym memberships.
- The difference is significant in Atlanta, Orlando, Miami and Seattle.
- In contrast, in the other cities, there is little difference between the number of men and women who take out a subscription.

3.4 External factors analyses

- The analysis focuses on a single city, New York, in order to vary only one external factor.

- The weather has a strong influence on gym attendance: attendance is high when the sky is overcast, as people are not encouraged to do sport outdoors.
- The lunar phases influence also the check-ins: when the moon is full, people go to the gym less.
- Applied K-Means Clustering to analyze gym attendance preferences in New York, focusing on factors such as the day of the week, attendance frequency, average calories burned, workout type preferences, and average exercise duration.

4. CONCLUSIONS & RECOMMENDATIONS

This study reveals that user behaviour varies according to their demographic characteristics and exogenous factors. In order to increase their income, gyms need to adapt their marketing strategy according to their location and the profile of their users, particularly those who subscribe to the 'Pro' programme. They can also adapt their marketing strategies to various external factors, such as the weather, to encourage users to come when these factors are unfavourable (fog).