



# Final Report

PureAnalytics

April 2024



## Contents

### 1. Background

- 1.1 The Business Problem
- 1.2 The Objective

### 2. Product Analysis

- 2.1 Visit and Off-Peak Product Analysis

### 3. Class Popularity Analysis

- 3.1 Timetable Classes
- 3.2 Class Space Analysis
- 3.3 Class Inclusion Score (CIS)
- 3.4 Extra Classes

### 4. Recommendations

- 4.1 Off-Peak Product
- 4.2 Operational Adjustments
- 4.3 Class Adjustments
- 4.4 Conclusions
- 4.5 Future Considerations

### 5. Appendix

## 1. Background

In 2022-23, the UK health and fitness market reached £5.4bn, fuelled by strong demand and a focus on health and wellbeing amongst consumer discretionary spending. The COVID-19 pandemic had significantly disrupted the leisure industry, with health and fitness facilities still not recovering to pre-2019 club numbers and penetration rates. Some market share has been permanently lost to online alternatives developed during the pandemic.

Despite this, market value hit a record high due to increased average pricing to offset inflationary pressures. Private sector growth is driven by the low-cost segment, which has expanded to represent approximately 19% of the market. The Gym Group and Pure Gym make up 80% of this segment.

### 1.1 The Business Problem

*To maintain a competitive edge in the market, PureGym must strategically enhance its product portfolio to both attract new members and foster loyalty among existing ones. A pivotal aspect of this strategy lies in meticulously aligning membership packages, class offerings, and timetables with the evolving preferences and demands of the membership base.*



### 1.2 The Objective

To analyse member utilisation patterns across different gyms and product tiers, focusing on off-peak product optimisation and class popularity. To identify opportunities for improvement in these areas and thereby increase competitive advantage, enhance the member experience, and achieve the company's mission of promoting a healthier nation.

## 2. Product Analysis

### 2.1 Gym Visit Analysis

PureGym seeks to analyse gym visit patterns to pinpoint the busiest and least busy hours. This analysis will assess the effectiveness of the off-peak product and its designated hours at the specified gyms. The objective is to optimise resource allocation, enhance member experiences, and ensure the profitability of the gyms.

**Initial Wrangling Using Python:**

Process	Action
Import Libraries	Import libraries to support the handling of data – pandas & numpy and visualization's – matplotlib & seaborn
PEP8	All code was written to PEP8 guidelines & code was written to ensure no warnings arose
Import Data	Using Pandas read method 'visit_data.csv' & 'gym_data_modified.csv' were read into DataFrame objects. The gym_data_modified file was derived from gym_data to set out the off-peak timing data as detailed in the Python code file
Initial Review	Data was reviewed using the head, shape, describe & dtypes methods to get a sense of the data
Check for duplicates & missing values	Duplicates and missing values are explored using duplicated and isna methods
Convert date and time columns	All columns with date and time data are converted to datetime format for analysis
Covert binary data columns to boolean	Facility presence marked as true or false
Add calculations	Calculations added for time duration in gym duration and extraction of day of the week, month of the year and year from visit date information.
Remove invalid data	Remove data with < 15 mins spent in the gym, negative durations and excessively long durations
Remove unnecessary columns	Remove all columns not required for analysis

Merge DataFrames	Join the two DataFrame with a left join to enable data about the gyms to be added to the visits
Create an offpeak Boolean column	Use the merged data to add a Boolean column, flagging whether the visit took place during an off-peak time period or not
Adjust Off-Peak at 168	Adjusts the Off-Peak designation at gym 168, removing 6am-8:30am and 11:30am to 13:30pm based on visit data and Employer feedback
Export cleaned data	Full data set and a manipulated data set for tableau analysis (detailed below)

### Extract from Python to Tableau

A key step involved creating an extract from the “visit\_clean” dataset summarised in Section 2.1 for detailed analysis in Tableau.

- We defined 15-minute intervals throughout the day using pd.date\_range to categorise visits into specific time slots, then aggregated data by center, year, month, day, and membership category.
- By exporting the processed data to Tableau, we enable a more dynamic and interactive analysis environment.
- A breakdown of the code used to create the extract is summarised below.

Process	Action
Defining Time Intervals	Creates 15-minute time intervals throughout the day with pd.date_range to help categorize visits into specific time slots.
Aggregating Data	Aggregates data by center, year, month, day, and membership category within each time interval. Counts visits for each category and interval, using nested loops to iterate through centers and dates.
Calculating and Structuring the Final Output	Converts aggregated data into a DataFrame and calculates an Average Visits column, showing visits per day adjusted by how often each day occurs monthly. Structures results into a pivot table for clarity, categorised by center, date, and time interval.
Exporting to CSV	Exports the pivot table to a CSV file named final_average_visitor_counts.csv, providing a structured format for further analysis or reporting.

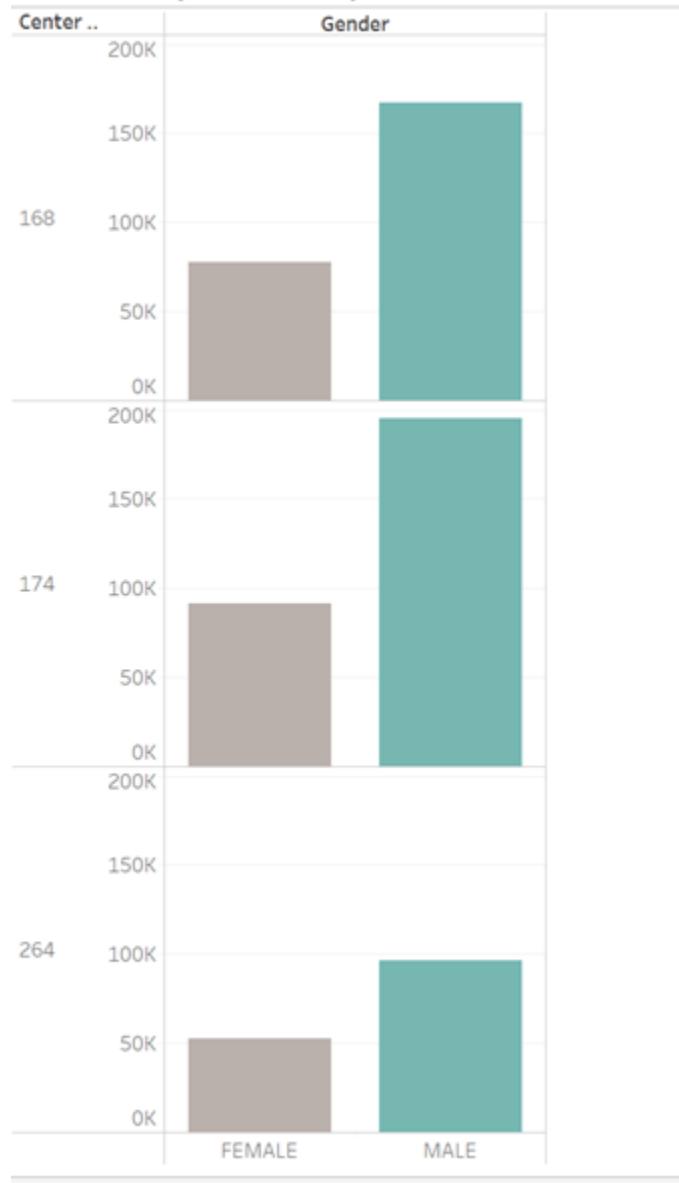
## Analytical Approach

The workbooks utilised for producing the analysis are listed in Appendix 5.3 of this report.

Process	Action	Visualization tool
<b>2.1.1 Demographic Analysis</b>	Examines demographic profiles (age, gender) to determine who uses the gym and when.	<ul style="list-style-type: none"> <li>Bar Plot: Displays the count of visits by gender and center.</li> </ul>
<b>2.1.2 Visit Duration Analysis</b>	Analyzes the duration of member stays to gauge overall engagement and facility utilization.	<ul style="list-style-type: none"> <li>Histogram: Shows the distribution of visit durations, segmented by gender.</li> </ul>
<b>2.1.3 Center Facilities Analysis</b>	Assesses the impact of specific facilities (e.g., spin classes, functional space) on member visitation trends.	<ul style="list-style-type: none"> <li>Bar Plot: Illustrates the average number of visitors in each gym, categorized based on the facilities available.</li> </ul>
<b>2.1.4 Temporal and Membership Analysis</b>	<p>Identifies peak usage times and attendance patterns across different times of the day, week, and month.</p> <p>Focuses on the usage of Off-Peak memberships across centers to evaluate their effectiveness and popularity.</p>	<ul style="list-style-type: none"> <li>Bar Plot: Represents the number of visits by month, divided by membership type.</li> <li>Bar Plot: Depicts the number of visits by day of the week, differentiated by membership type.</li> <li>Bar Plot: Visualizes the average number of visitors in the gym throughout the day, from Monday to Friday.</li> <li>Line Plots: Track the number of visitors throughout the day, plotted for each day of the week.</li> </ul>

### 2.1.1 Demographic Analysis Demographic Analysis:

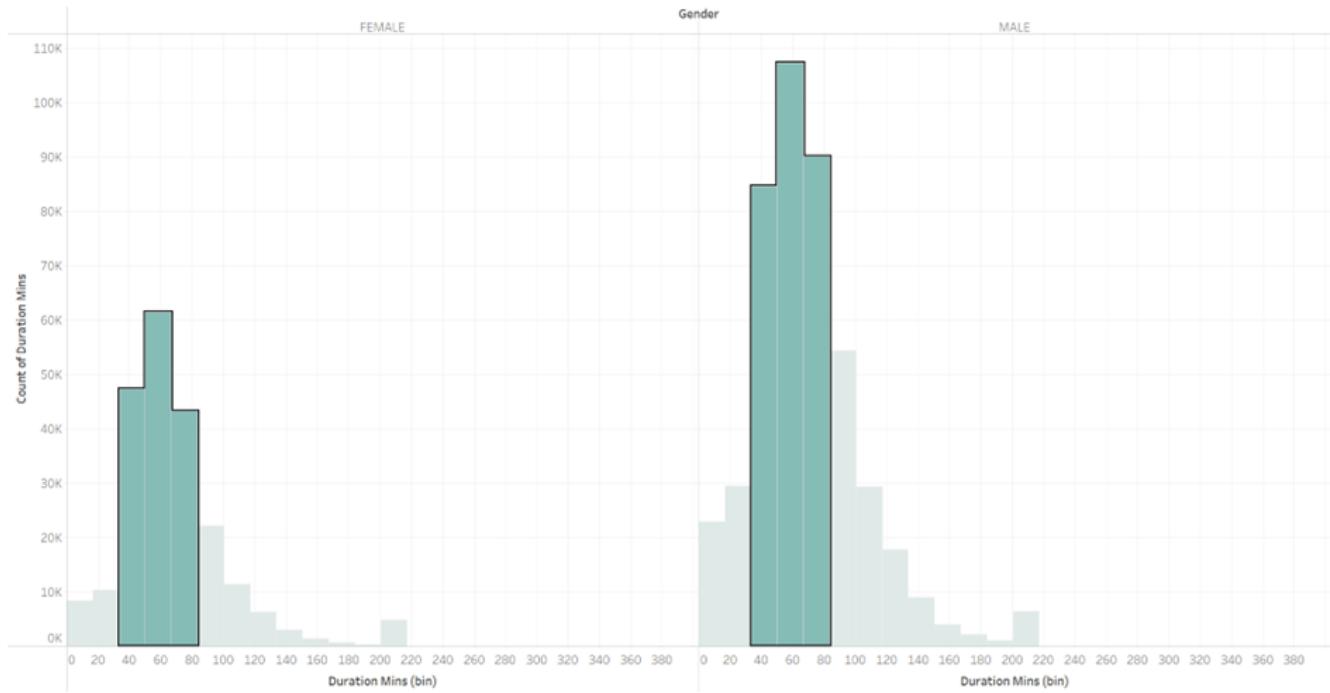
Count of Gym Visits by Gender and Center



Across all three gyms, visits by male members are approximately twice those of female members, indicating a significant gender disparity in gym attendance.

## 2.1.2 Visit Duration Analysis

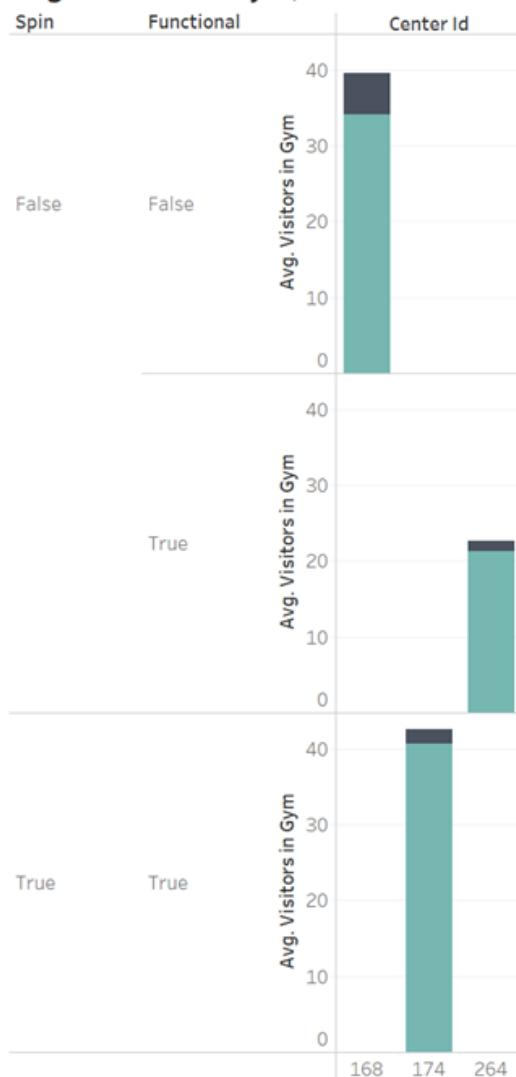
Distribution of Visit Duration



In the analysis, the average duration of gym visits is centred around the 60-minute mark, showing no significant differences between genders. This uniformity across male and female members indicates a consistent usage pattern irrespective of gender.

### 2.1.3 Facilities Analysis

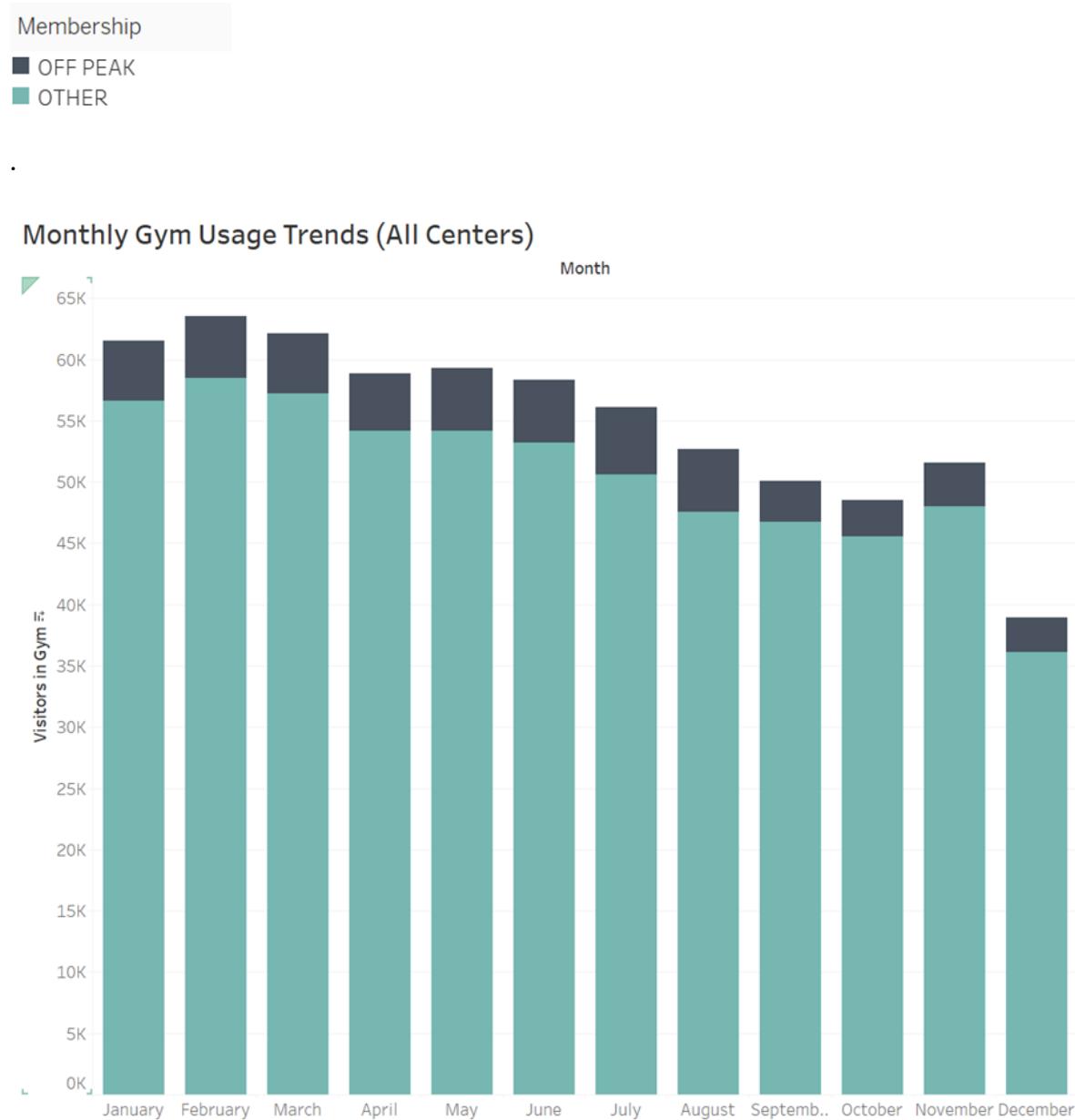
Avg Visitors in Gym, based on Facilities available, by Centre



Due to the limited dataset, which includes only three locations and lacks a consistent combination of spin and functional space facilities, it is challenging to draw insightful conclusions regarding how specific amenities influence gym visit patterns.

#### 2.1.4 Temporal and Membership Analysis

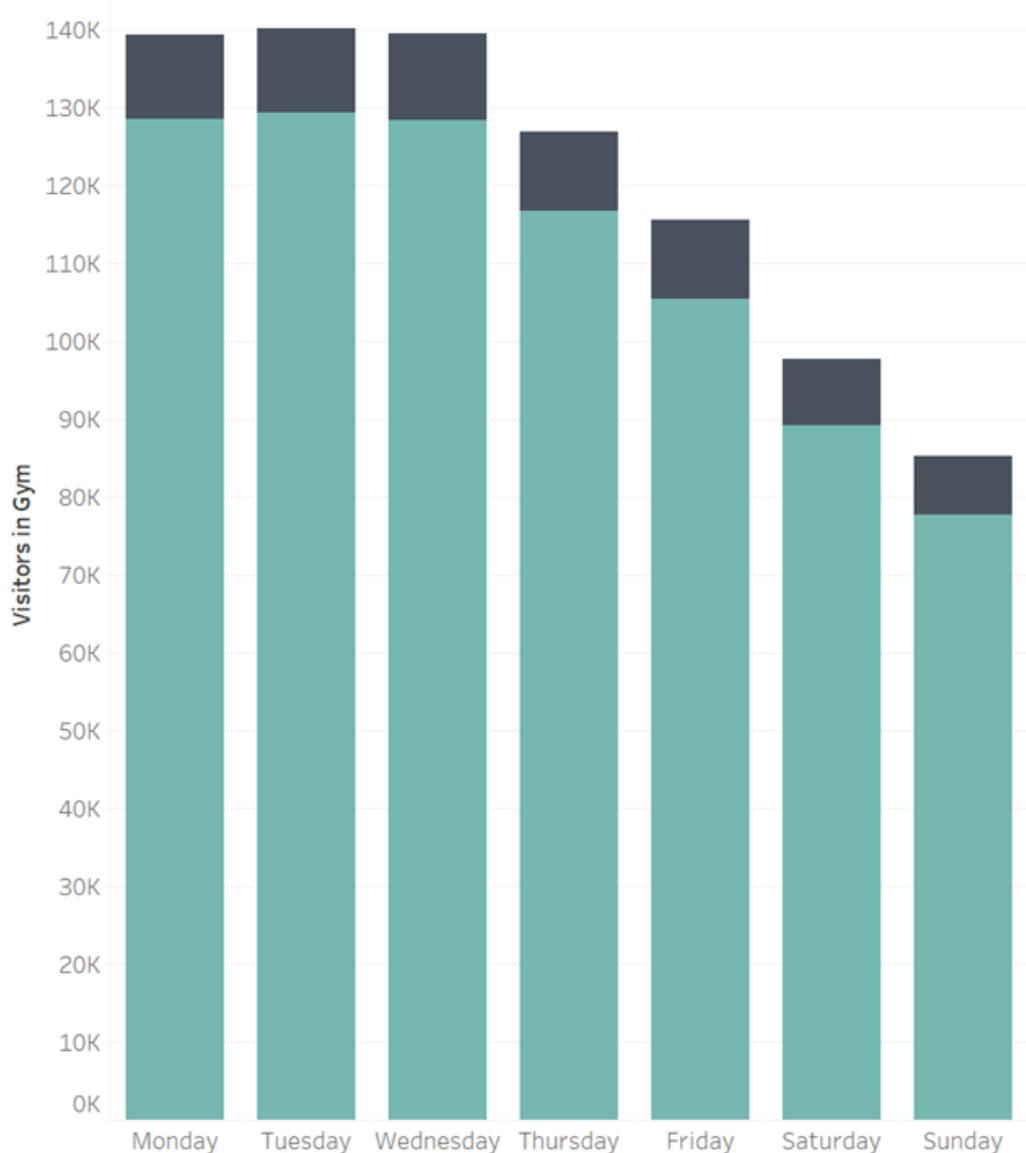
Our analytical focus is to contrast Off-Peak members' usage with that of all other membership types.



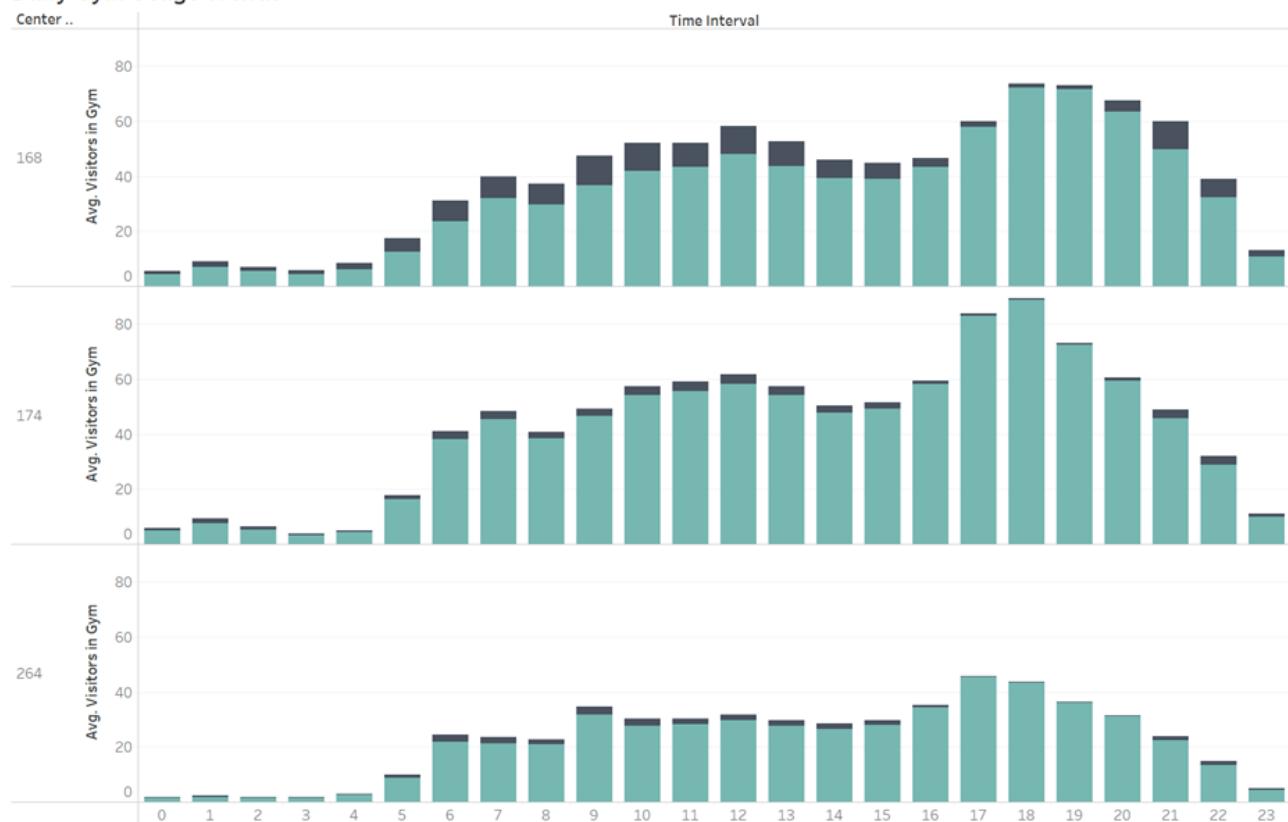
The analysis revealed a peak in gym attendance in the first quarter of the year, with a decline towards December. Weekdays showed higher attendance, particularly early in the week, with a significant drop during weekends.

## Daily Gym Usage Trends (All Centers)

Day Of Week



## Daily Gym Usage Trends



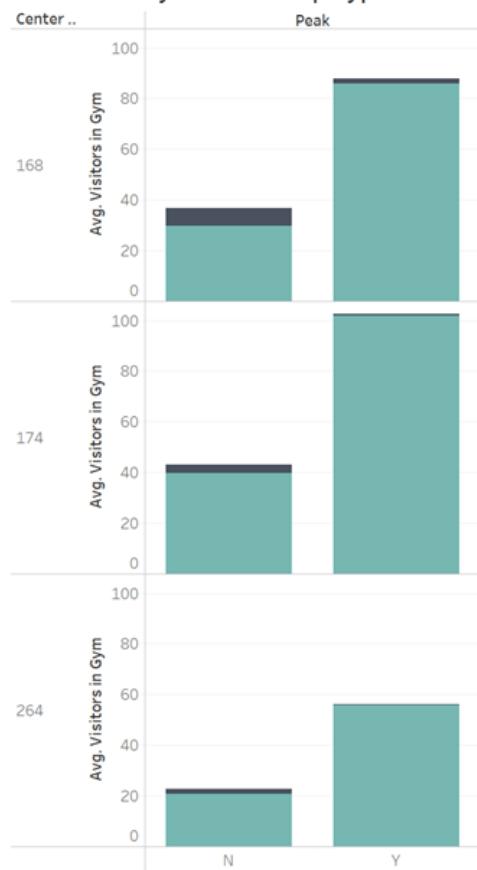
**Daily and Hourly Trends:** Peak gym usage was consistently noted from 5 to 9 PM across all examined gyms, aligning with typical workday schedules. There was a notable low in attendance from 11 PM to 5 AM.

## Daily Gym Usage Trends by Day of Week



- When analysing gym usage patterns by day of the week, notable trends emerge.
  - Noticeable increase in attendance on Saturday mornings, particularly in gyms 168 and 174, indicating a surge in member activity.
- Conversely, Friday evenings experience significantly lower usage from 5-9 PM compared to other weekdays.
  - Suggesting a shift in member preferences towards the end of the workweek.
- All gyms consistently show minimal activity between 11 PM and 5 AM, reinforcing the pattern of low overnight usage.

### % of Visits by Membership Type



- In Centre 168, Off-Peak membership accounted for a more significant portion of visits during designated Off-Peak hours, indicating a possible price sensitivity among members at this location.
- However, in 174 and 264, the Off-Peak membership showed minimal impact on overall gym business, suggesting that members prioritised flexibility over cost savings.

### 3. Class Popularity Analysis

**What class offering would deliver the best member experience and value?**

PureGym aims to refine its group exercise class offerings to enhance member experience and drive value. Understanding which classes are most popular among different member segments and at which locations is crucial for optimising class schedules and ensuring the best possible member experience.

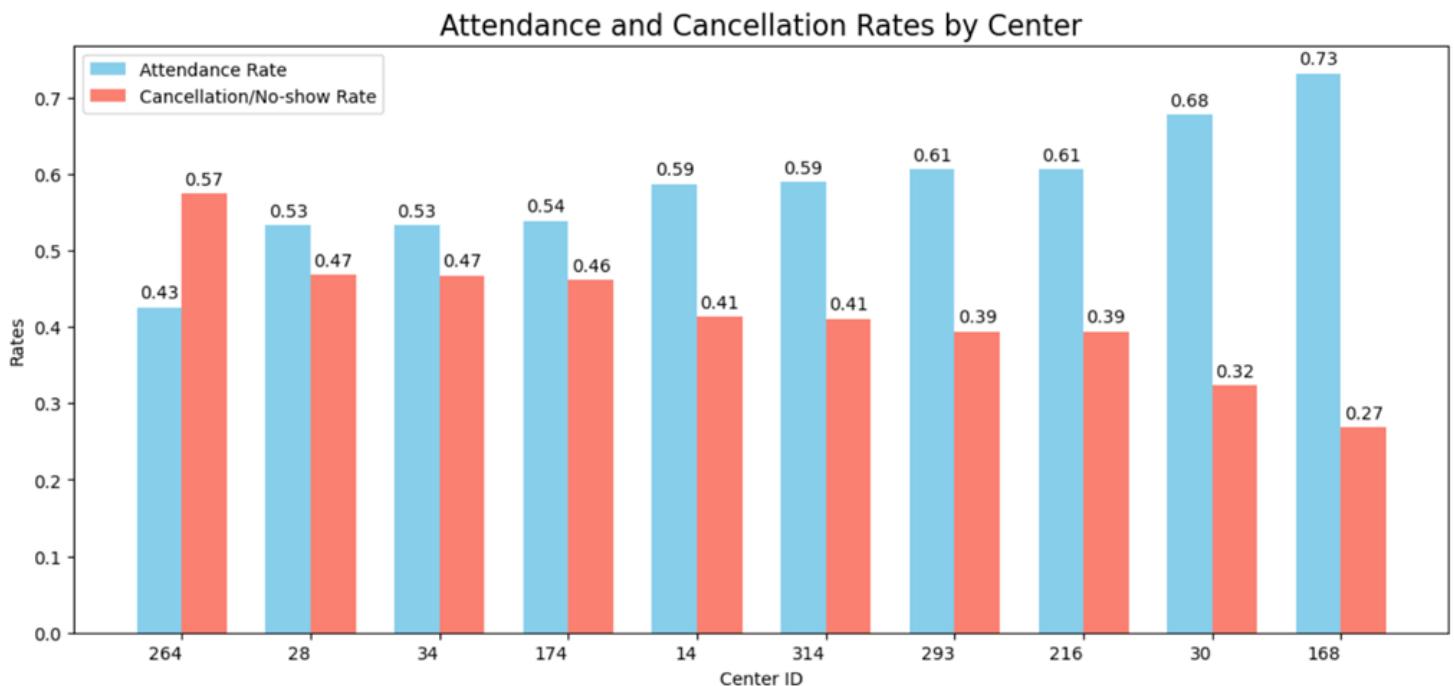
#### 3.1 Timetable Classes

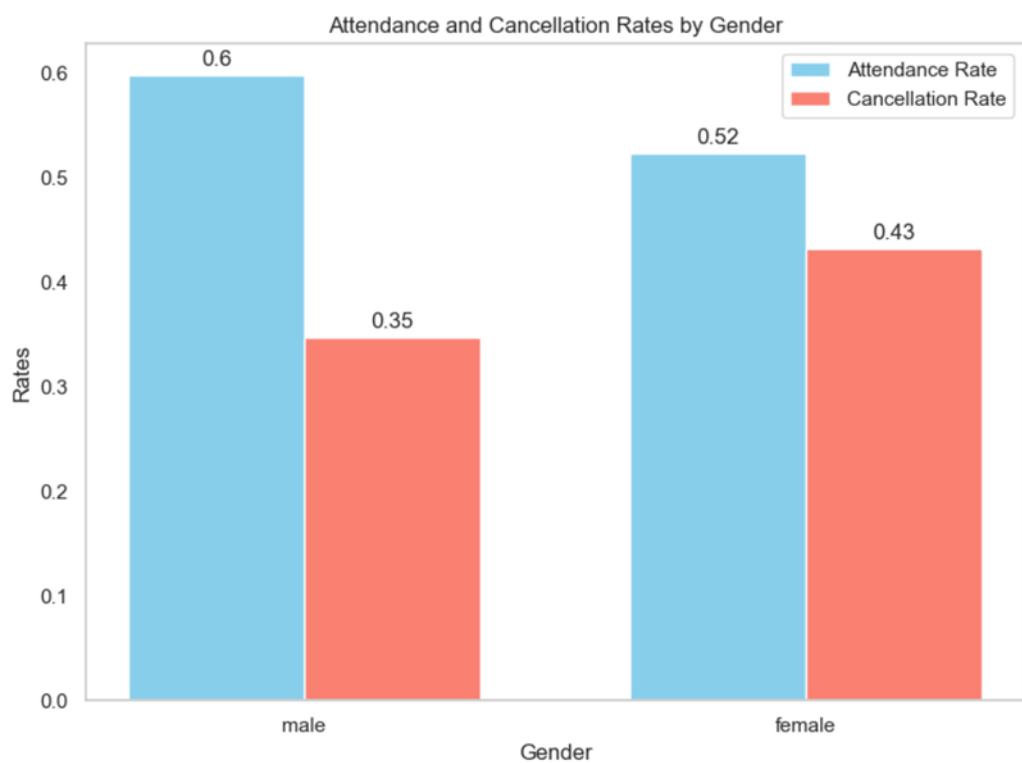
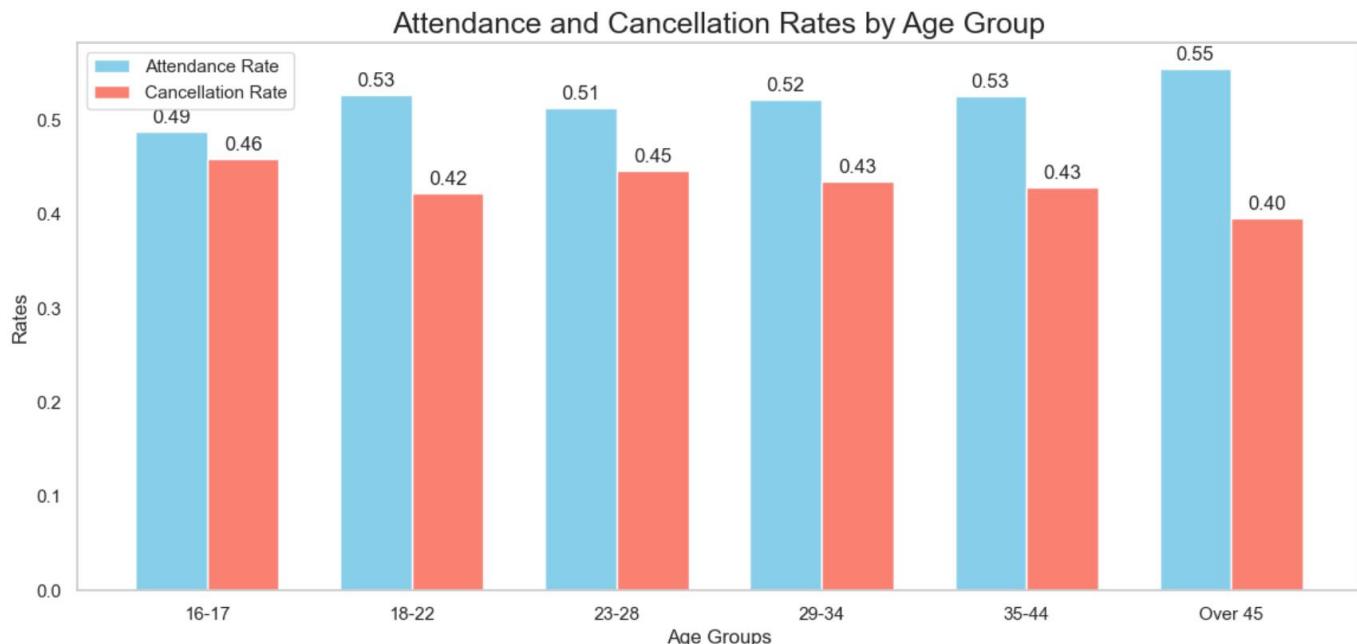
In which gyms and with which types of members are these classes most popular?

- Attendance rates by gym and demographics.
- Correlation of attendance with the time of day or week.
- Patterns that suggest optimal scheduling for specific demographics.
- Which classes are trending in terms of attendance over time?

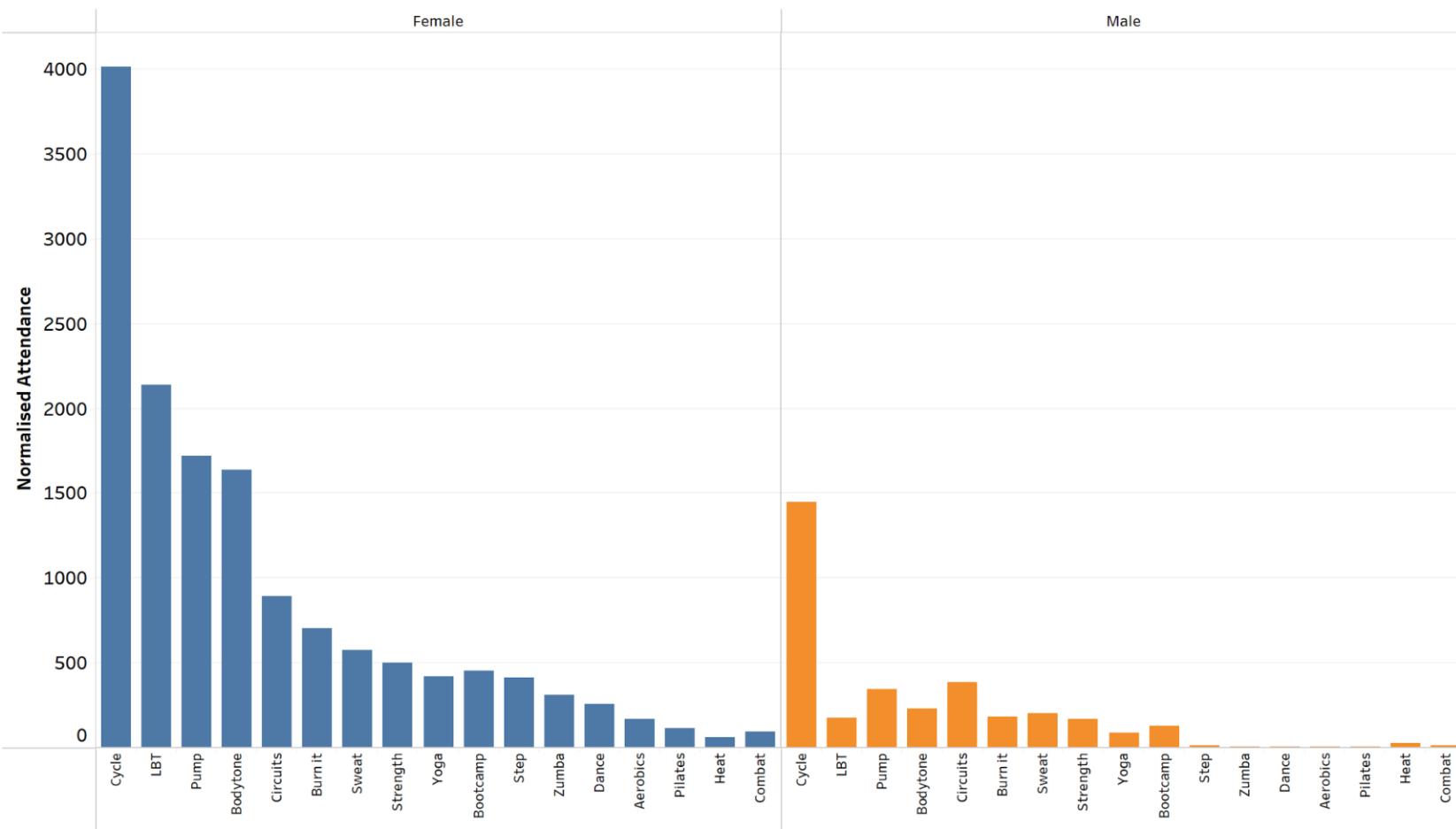
Analytical Approach – see Appendix

##### 3.1.1 Attendance rates by Center and demographics.





## Timetable Class Normalised Attendance by Gender

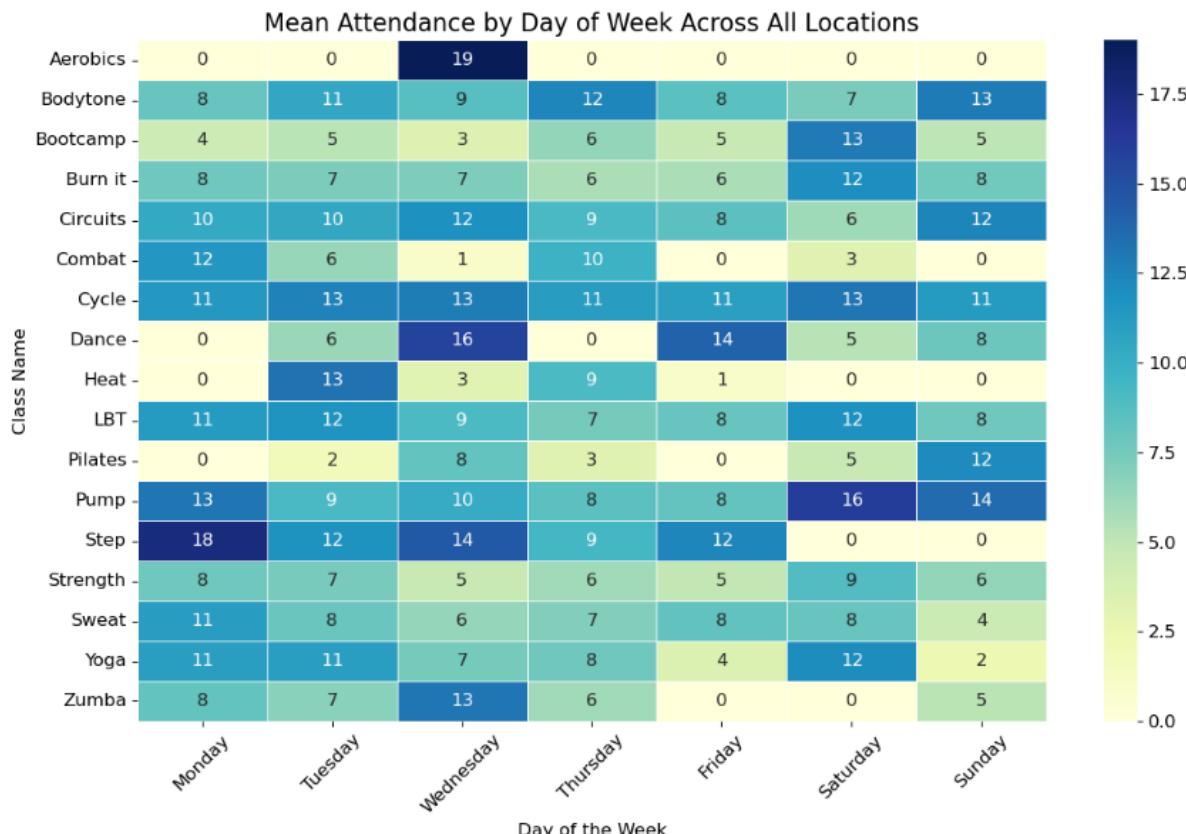
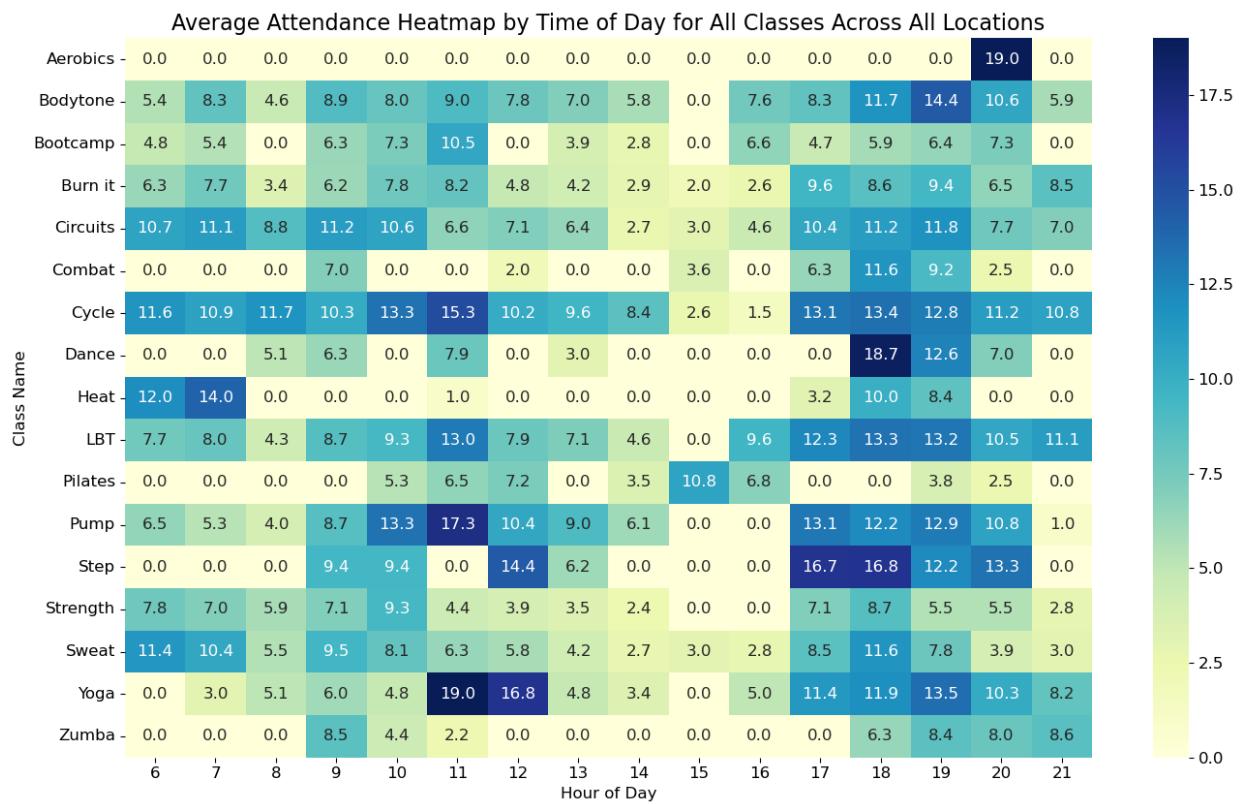




## Insights

- Centers 168 and 30 excel in attendance.
- Members over 45 consistently engaged; Younger demographics cancel often.
- Cycle class is most popular, with female participation dominating most classes, especially LBT, Bodytone and Pump

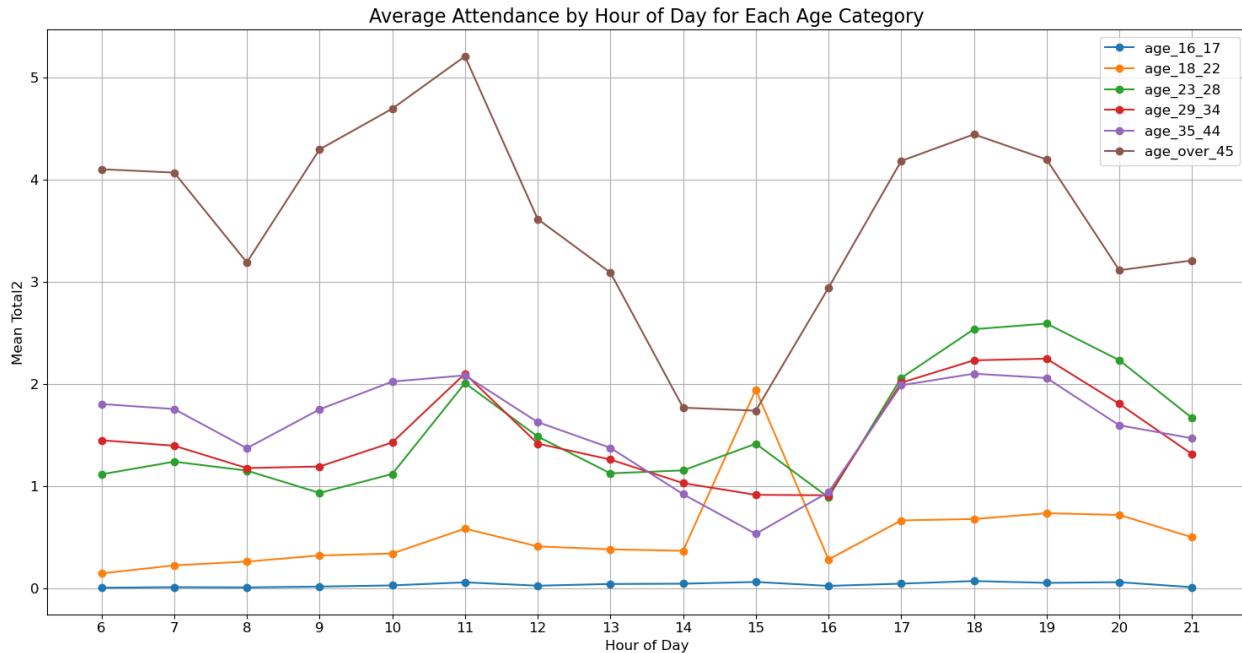
### 3.1.2 Correlation of attendance with time of day or week



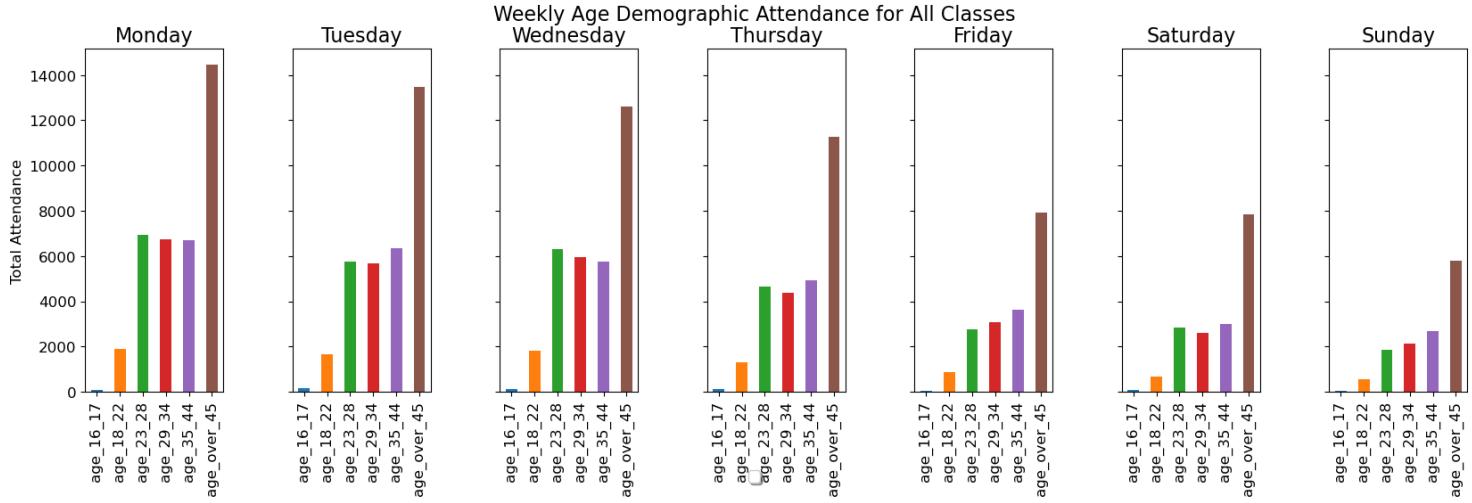
## Insights

- High mid-morning / late afternoon Cycle and Pump attendance
- Low Combat AM turnout; reschedule advised.
- Weekday Circuit and Yoga peaks; Pump weekend surge noted.

### 3.1.3 – Patterns that suggest optimal scheduling for certain demographics

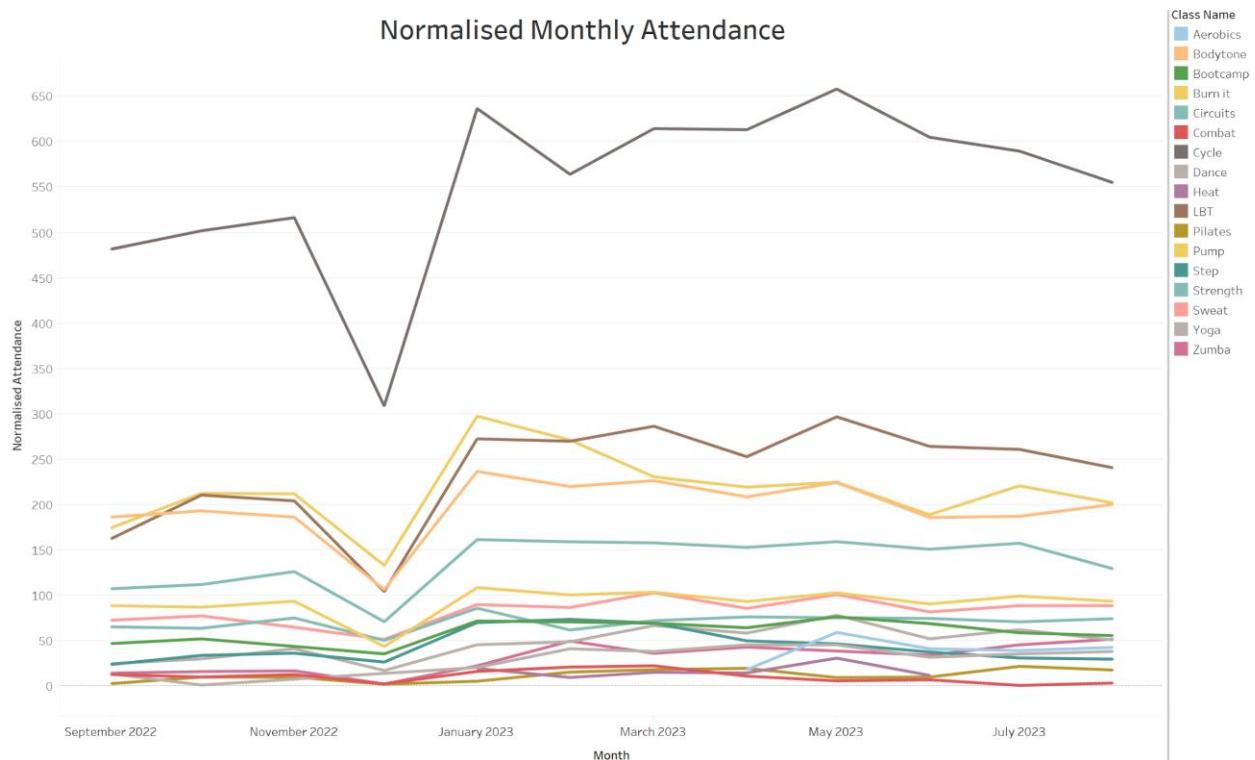


- Cycle, Pump, and Bodytone show high engagement across all ages, while LBT appeals to younger clients. Scheduling these classes later in the day could enhance accessibility.
- Circuits and Burn It evening class offerings can be seen to attract younger attendees, informing scheduling time.



- Highest attendance across all ages on Mondays , tapering off toward the weekend
- Over 45s dominate the classes, with consistent midweek strength in ages 23+
- Lower weekend engagement, opportunity for targeted age-group classes.

### 3.1.4 – Which classes are trending in terms of attendance over time?



- Cycle consistently leads in attendance, all classes dip in December, with a marked recovery for Cycle and Pump by mid-year.
- Combat shows a notable decline over time.

### 3.2 Class Space Analysis

**What are the differences in popularity between classes in different group exercise spaces?**

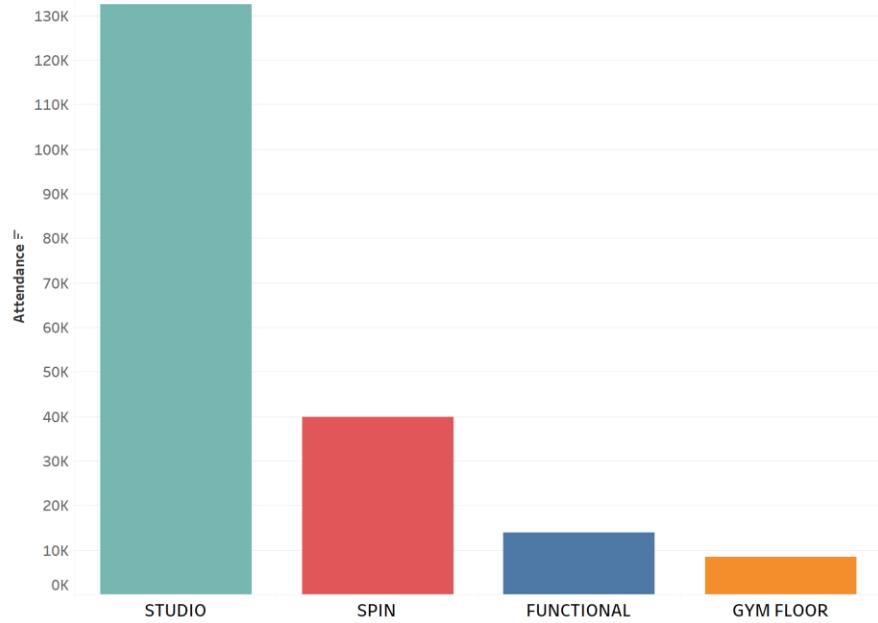
#### Initial Wrangling

Process	Action
Import Libraries	Import necessary libraries e.g. pandas, numpy, matplotlib, seaborn, etc.
Ignore Warnings	Optionally, ignore any warnings that might occur during execution.
Import Data	Import two CSV files, 'class_data.csv' and 'timetable_data.csv', into pandas DataFrame objects.
<b>Clean and Wrangle the Data Set</b>	
Convert Date and Time Columns	Convert 'start_date_time' and 'stop_date_time' columns in 'class_data' DataFrame to datetime.
Determine Time Span	Calculate the minimum and maximum dates in 'class_data' DataFrame to determine time span.
Normalize Data to Case Sensitivity	Convert class names to lowercase for normalization.
Replace Specific Class Names	Replace 'Legs, Bums & Tums' class name with 'LBT'.
Create Mapping Dictionary for Partial Matches	Create a mapping dictionary to map class names to space types.
Apply Mapping Function	Apply the mapping function to assign space types based on class names.
Check Unique Class Names	Check unique class names in timetable data and class data.
Replace and Drop Columns	Replace 'class_name' column with its normalized version and drop it from both DataFrames.
Group Data and Check Spaces Attributed	Group data by class name and space type to check the spaces attributed.
Calculate Participation Counts	Calculate participation counts based on gender.
Export Cleaned Data	Export cleaned data to CSV files ('class_data_clean.csv' and 'timetable_data_clean.csv').
In Excel add Total and Total2	Total - sum of male, female and unknown. Total2 sum of all age groups.

## Analytical Approach

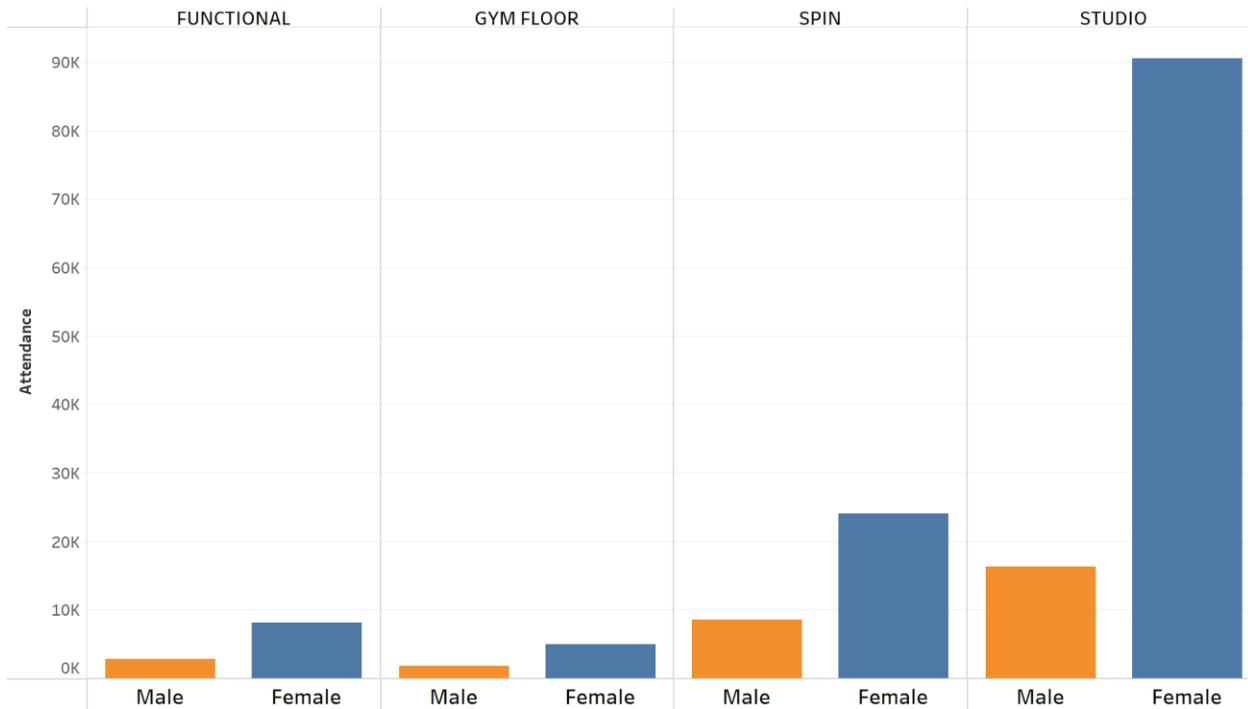
Process	Action	Visualization Tools Used
<b>3.2 Class Space Analysis</b>		
<b>3.2.1 Attendance by Space Type</b>	Filtered for attended sessions only. Grouped and summed attendance by space type. Sorted results by attendance.	Bar plot displaying attendance by space type.
<b>3.2.2 Space Attendance by Gender</b>	Calculated total and percentage attendance by gender within each space type. Selected relevant columns for presentation.	Bar plots showing gender attendance percentages.
<b>3.2.3 Space Attendance by Age Group</b>	Analyzed and calculated attendance by age groups within each space type. Computed percentage attendance for age groups.	Bar charts of percentage attendance by age group.
<b>3.2.4 Space Attendance by Gym</b>	Analyzed space usage by gym center. Added and sorted by total attendance per gym. Prepared data for heatmap visualization.	Stacked Bar Chart
<b>3.2.5 Attendance by Time of Day</b>	Defined and applied time of day categories. Grouped and analyzed space usage by time segments.	Heatmaps and line charts for time-based space usage.
<b>3.2.6 Seasonal Analysis of Space Usage</b>	Defined seasons and grouped data by season. Ordered and prepared data for seasonal analysis.	Heatmaps and line charts showing seasonal space usage.
<b>3.2.7 Weekly Attendance by Space</b>	Extracted and mapped day names from date. Grouped by day and space type, summed attendance. Ordered days for visual representation.	Heatmaps and line charts displaying weekly space usage.
<b>3.2.8 Space Utilization</b>	Compared actual attendance to class capacity. Calculated utilization percentage for each space type.	Utilization data presented in tabular form.
<b>3.2.9 Space Utilization with All Bookings</b>	Calculated total bookings without the 'ATTENDED' filter. Merged bookings with capacity data for utilization analysis.	Utilization percentages based on bookings displayed in tabular form.

### 3.2.1 Attended by Space Type



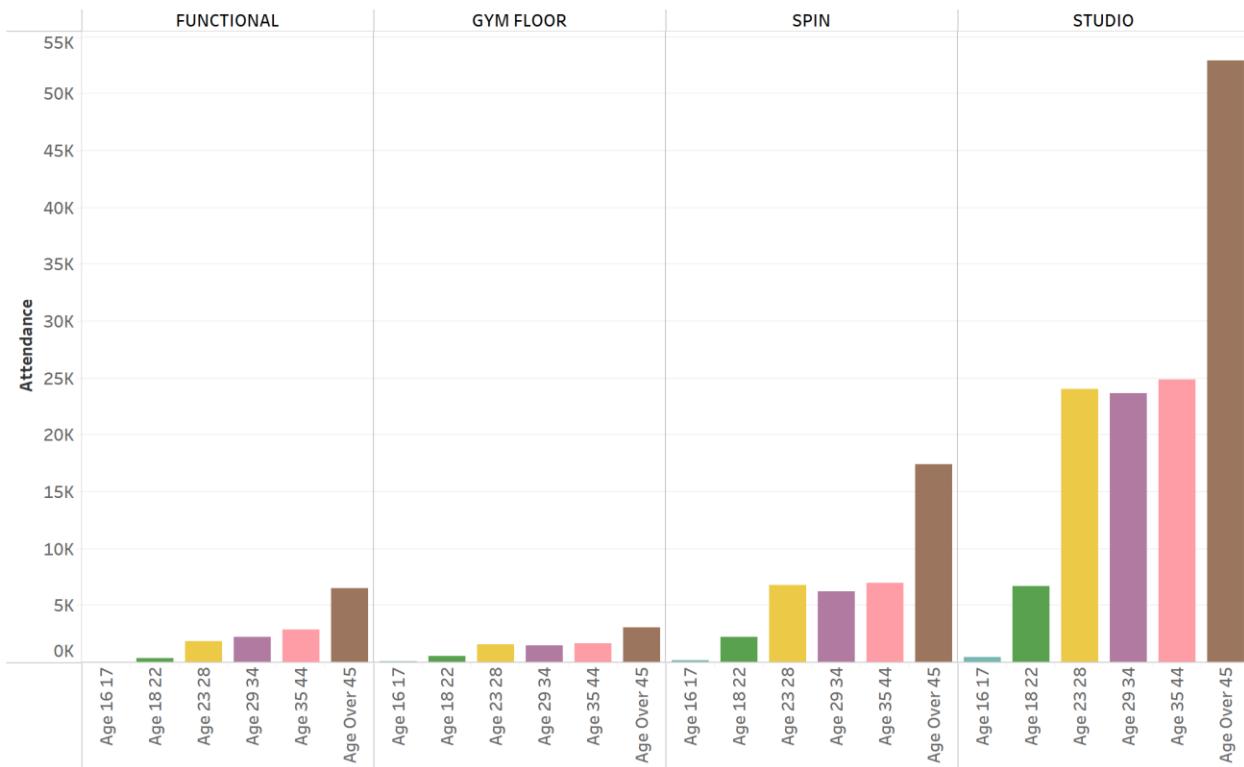
Studio space is by far the most attended.

### 3.2.2 Space Attendance by Gender



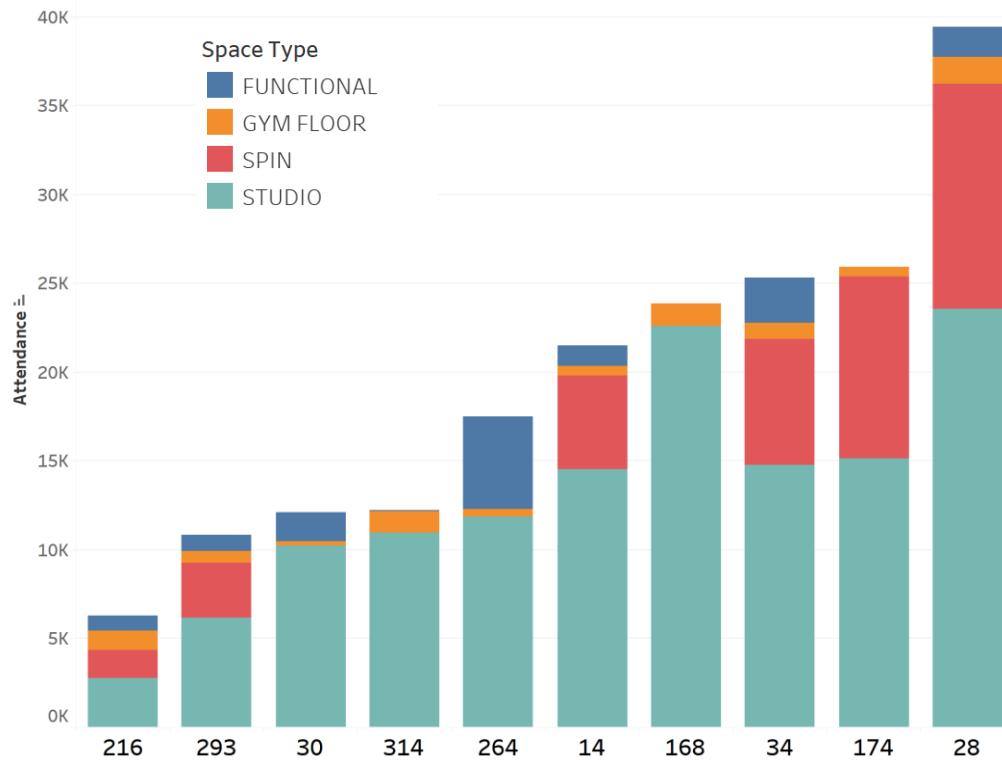
Females dominate attendance in all spaces.

### 3.2.3 Space Attendance by Age Group



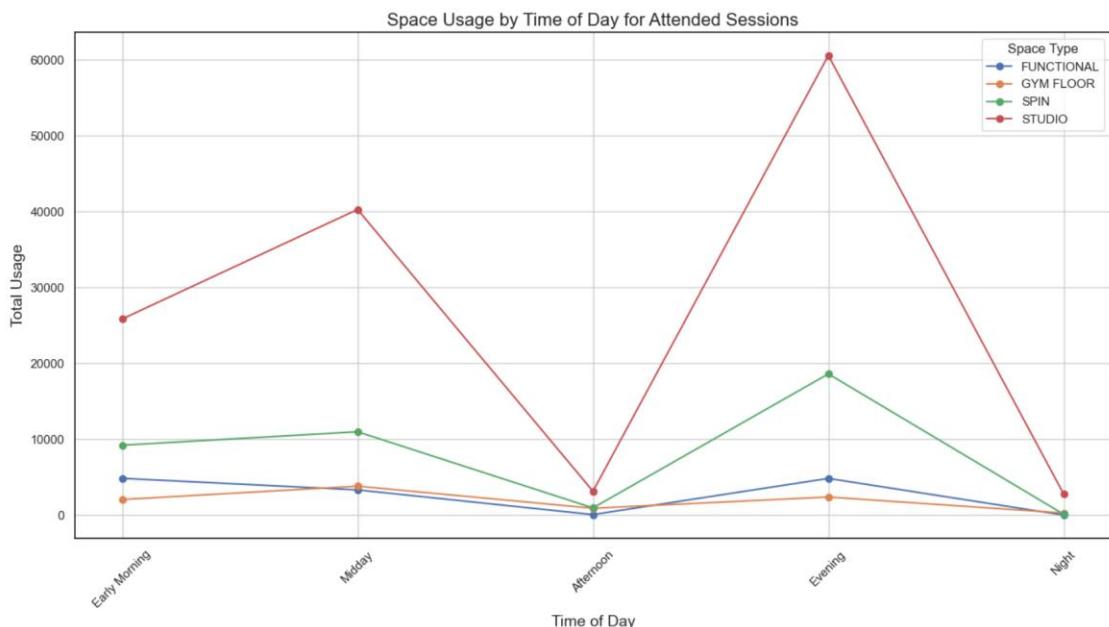
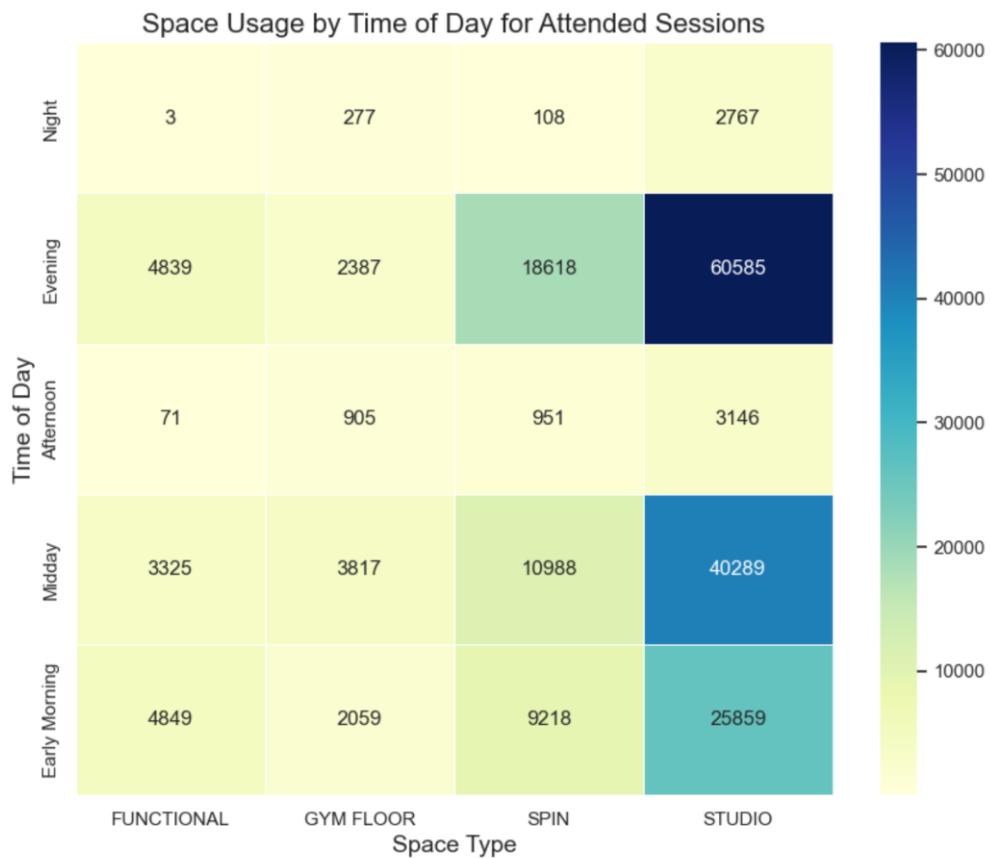
Over 45s constitute the largest attending age group in all spaces. Higher attendance is broadly in line with increasing age.

### 3.2.4 Space Attendance By Gym



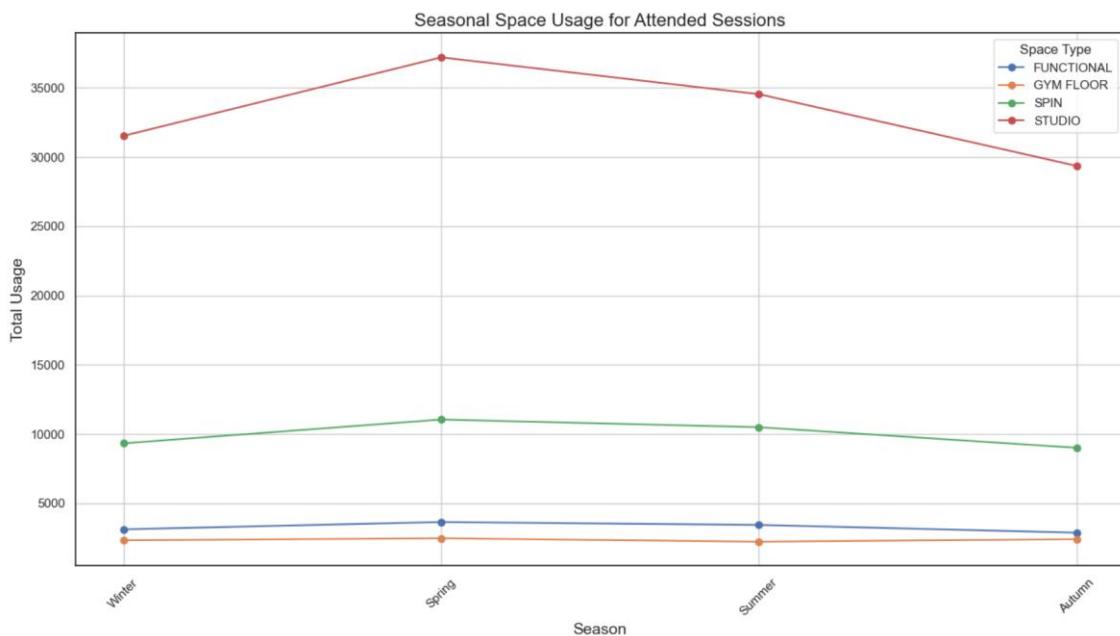
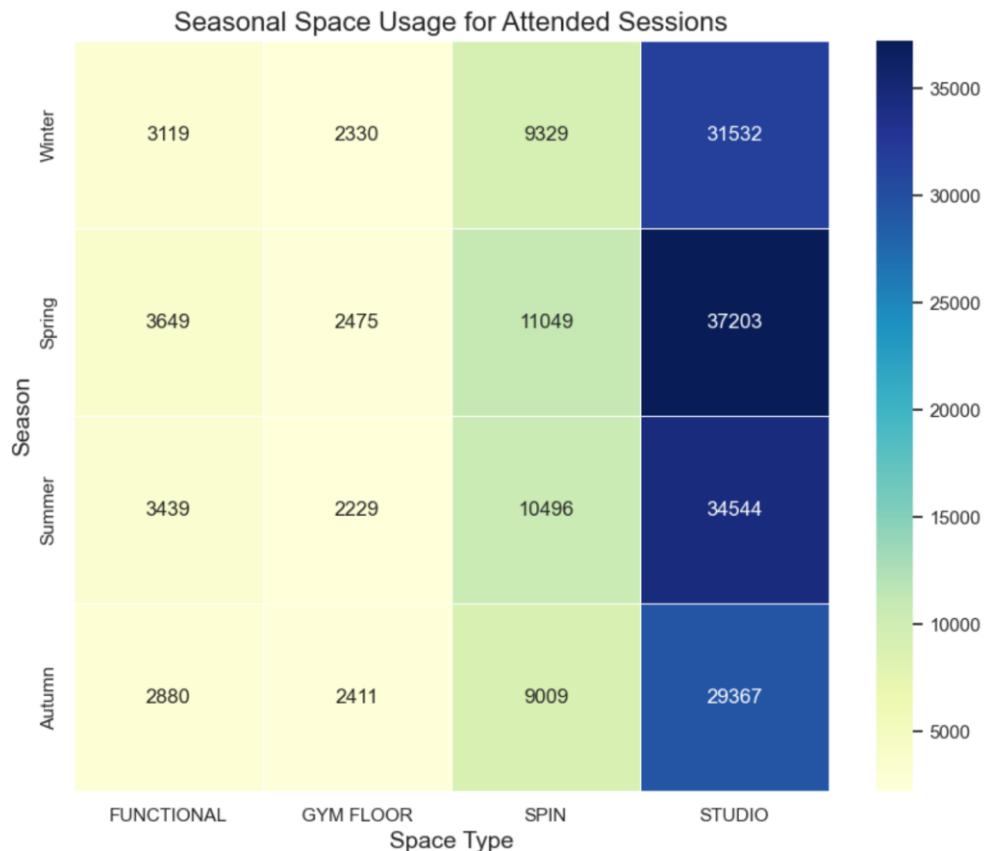
Center 28 has the highest attendance.

### 3.2.5 Attendance by time of day



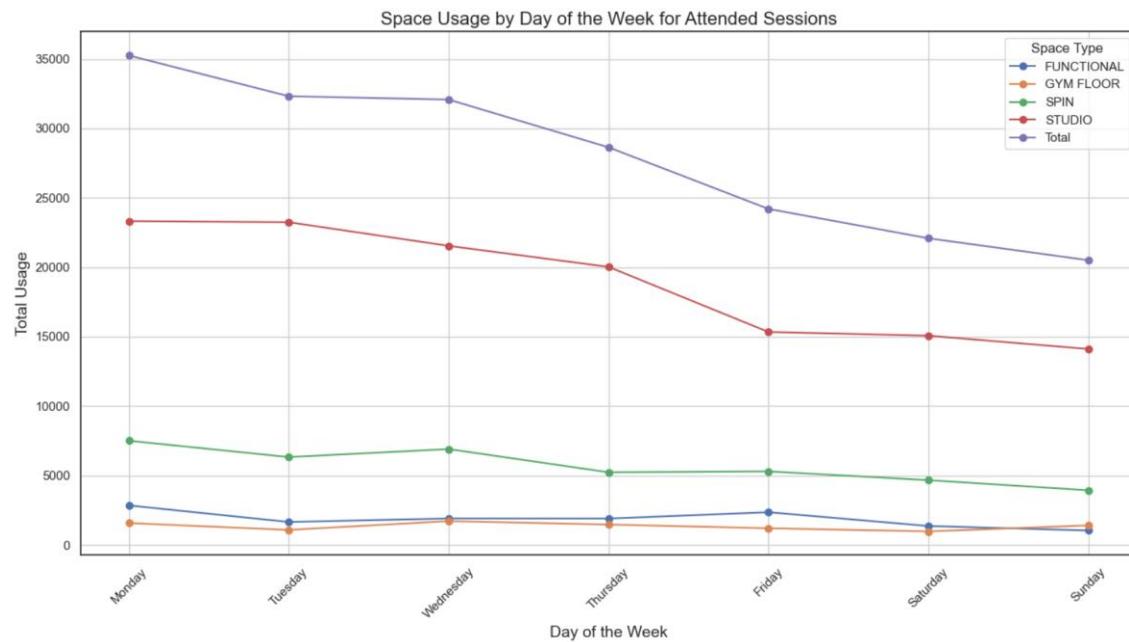
Evenings are the busiest times of day for Studio, Spin and functional. Gym floor attendance is higher during the early morning, midday and evening.

### 3.2.6 Seasonal Analysis of Space Usage



Notable upticks in attendance in all spaces from Winter to Spring, followed by a gradual decline.

### 3.2.7 Weekly attendance by space



Generally, attendance is highest early in the week.

### 3.2.8 Space Utilisation by Attendance

space_type	total_attendance	total_capacity	utilization_percentage
FUNCTIONAL	13087	28794	45.45
GYM FLOOR	9445	19373	48.75
SPIN	39883	66795	59.71
STUDIO	132646	348644	38.05

#### Spin Space Utilization:

- Spin has the highest utilization percentage at approximately 60%.
- This indicates high popularity and strong demand for spin classes.
- Considerations might include increasing capacity or adding more classes to meet demand.

#### Functional and Gym Floor Utilization:

- Functional space and Gym Floor have utilization percentages of 45% and 49%, respectively.
- Indicates a healthy interest but suggests potential for increased participation.
- Opportunities may exist to optimize scheduling and capacity to enhance utilization.

#### Studio Space Utilization:

- Studio space has the lowest utilization at 38%.
- Despite its versatility for hosting various classes, utilization is relatively low.
- This low figure could suggest an oversupply of studio space, a need for more diverse class offerings.
- Potential mismatches in class scheduling with member preferences or availability.

### 3.2.9 Space Utilisation by Total Bookings

space_type	total_capacity	total_bookings	utilization_percentage_bookings
FUNCTIONAL	28794	26009	90.33
GYM FLOOR	19373	16837	86.91
SPIN	66795	86481	129.47
STUDIO	348644	234068	67.14

#### Spin Utilization Based on Bookings:

- Spin class bookings show a utilization percentage of 129%, significantly exceeding the total capacity.
- This overbooking suggests a demand that surpasses available supply, possibly due to double bookings, waitlists, or non-attending members.
- This highlights the need to increase spin class capacity or improve booking and attendance management.

#### Gym Floor Utilization:

- Gym Floor shows a high utilization percentage of 86% based on bookings, nearing total capacity.
- This indicates a strong and consistent demand for weight training spaces, with little room for additional bookings without risking overcapacity.

#### Functional Space Utilization:

- Functional space booking utilization is 90%, closely approaching total capacity.
- This high demand shows that the space is being used effectively, with limited margin for growth without expanding capacity or optimizing the schedule.

#### Studio Utilization Based on Bookings:

- Studio space has a booking utilization of 67%, which is an improvement over its attendance-based utilization but still below full capacity.
- This suggests potential excess capacity in studio spaces or opportunities to increase bookings through enhanced marketing, diversifying class offerings.

### 3.3 Class Inclusion Score (CIS)

What classes are gyms providing that aren't currently included in the suggested timetable but are successful?

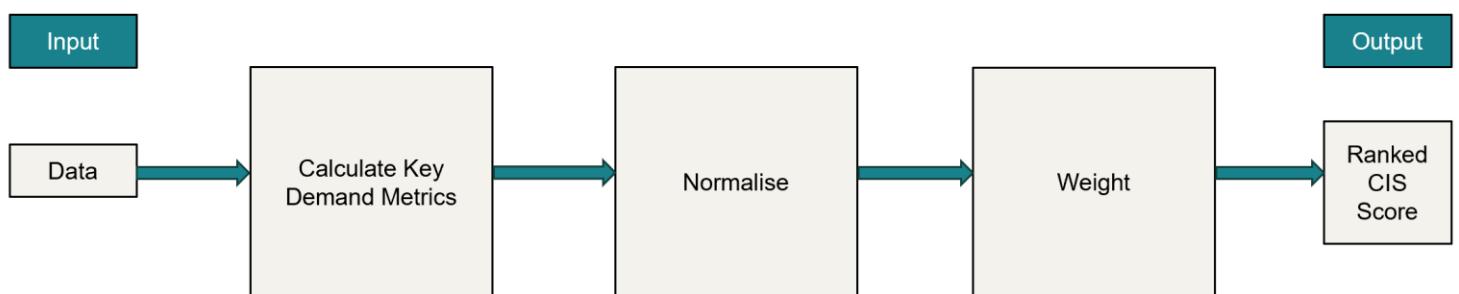
#### Aim

To provide an objective, quantifiable method of assessing class popularity through a data driven analysis of PureGym's class offerings.

#### Objectives

- Identify "Extra" classes that merit inclusion in the Timetable.
- Pinpoint underperforming Timetable classes for re-evaluation.

An algorithm that attempts to assess class popularity based on key demand metrics



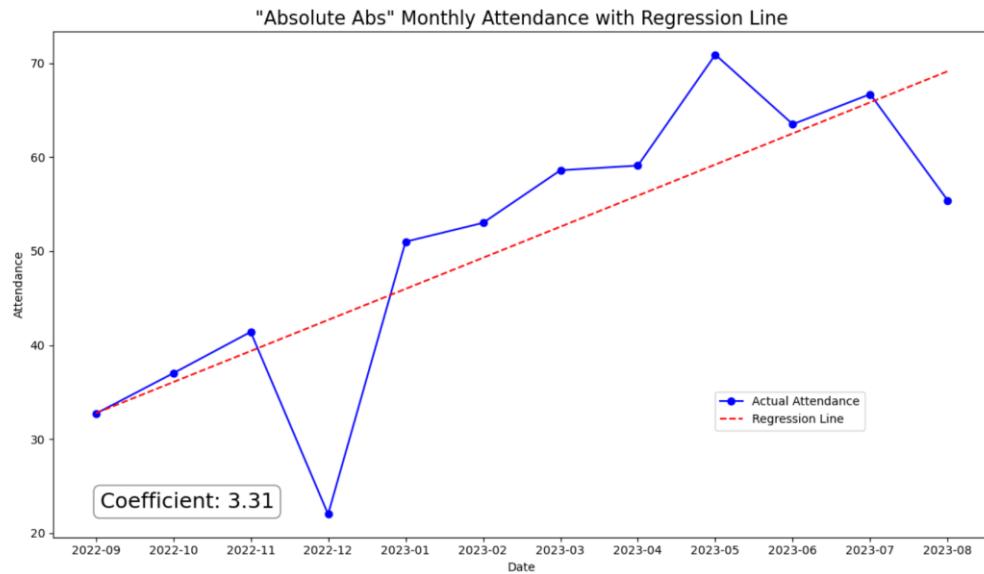
## Analytical Approach

Process	Action
<b>3.3 Class Inclusion Score</b>	
<b>3.3.1 Load and Explore Data</b>	
Import Libraries	Import necessary libraries e.g. pandas, numpy, matplotlib, seaborn, etc.
Ignore Warnings	Optionally, ignore any warnings that might occur during execution.
Import and View Dataset	Import the dataset from 'class_data_final.csv' and view its properties.
<b>3.3.2 Data Cleaning and Preparation</b>	
Capitalize Extra Class Names	Capitalize the first words of extra classes in the 'class_name' column.
Normalize Attendance Figures	Calculate normalized attendance figures based on the number of centers offering each class.
Remove Sporadic Classes	Remove classes specified in the classes_to_exclude list from the dataset.
<b>3.3.3 Calculate Metrics</b>	
Average Monthly Attendance to Capacity	Calculate average monthly attendance, average monthly capacity, and attendance to capacity ratio.
Growth Score - Monthly Attendance	Calculate coefficients and R <sup>2</sup> values for all classes using linear regression.
Example Monthly Attendance Regression	Perform linear regression and plot monthly attendance data for a specific class.
Monthly Average Attendance to Capacity Ratio	Calculate coefficients for the attendance to capacity ratio.
Coefficients	
Example Month Attendance Capacity Ratio	Perform linear regression and plot attendance to capacity ratio for a specific class.
Regression	
Waiting Lists	
Extract and Analyze Waiting Lists	Filter out entries with 'Total2' greater than 0 for 'WAITING LIST' participation status and analyze waiting lists.
Normalised Attendance	
Calculate Total Attendance	Calculate total attendance for each class.
<b>3.3.4 Merge Metrics</b>	
Merge Metrics into a Single DataFrame	Merge filtered metrics into a single DataFrame for analysis.
<b>3.3.5 Standardize Metrics</b>	
CIS derived from Min-Max Scaling and Weighted Sum Model	Use Min-Max Scaling to standardize the metrics and calculate the Class Inclusion Score (CIS).
<b>3.3.6 Weight Metrics</b>	
Use Weighted Sum Model to Weight Metrics	Assign weights to each metric and calculate the weighted sum to determine CIS.
<b>3.3.7 Rank Metrics</b>	
Rank Classes based on Each Factor	Rank classes based on CIS, attendance, growth score, capacity utilization, and waitlist ratio.
Summarize Rankings	Summarize rankings and identify the best and worst-performing classes.
<b>3.3.8 Assess Optimal Day and Time for inclusion of Absolute Abs and Metcon</b>	
Analyze Attendance by Day of Week	Analyze attendance for Absolute Abs and Metcon classes by day of the week.
Analyze Attendance by Time of Day	Analyze attendance for Absolute Abs and Metcon classes by time of day.

### Key Demand Metrics (3.3.3)

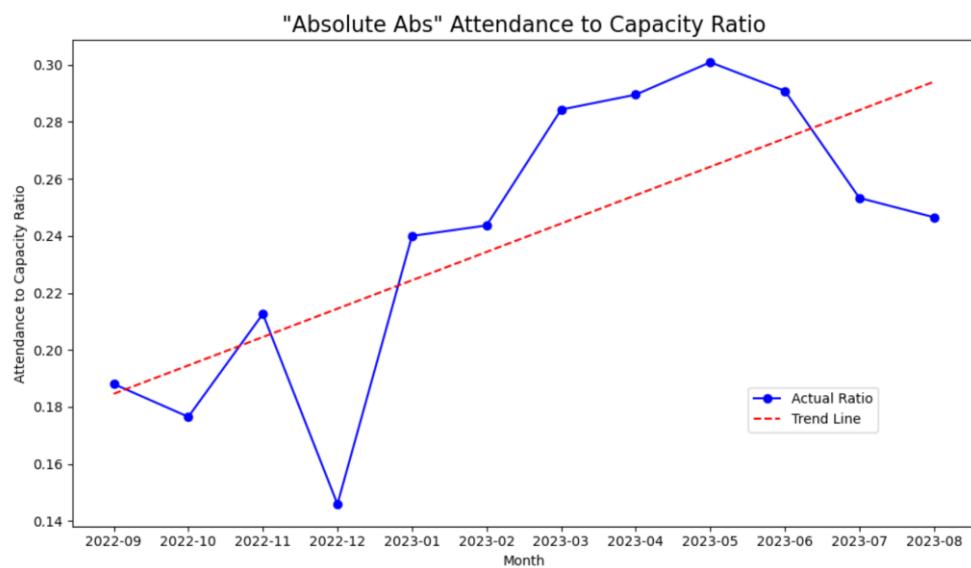
#### 1. Growth Score

- Coefficient of linear regression of the normalised monthly attendance in each class.
- A measure of the average number of members onboarded to a class per month in each centre.
- E.g. On average Absolute Abs onboards 3.31 members per month per centre.



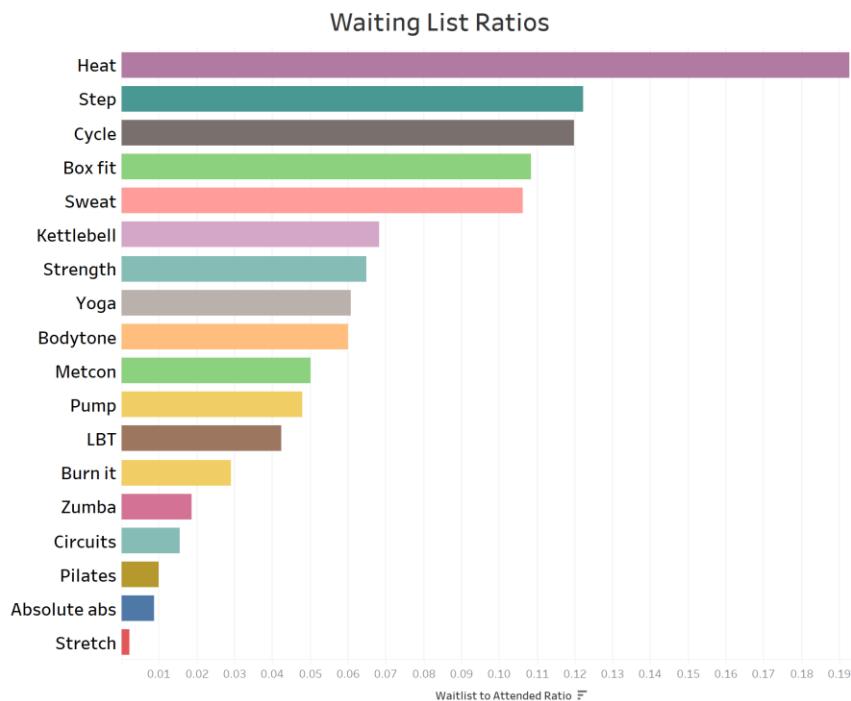
## 2. Capacity Utilisation score

- The coefficient of a linear regression analysis on the monthly trend in Attendance to Capacity Ratio.
- Indicates how efficiently space and resources are being used by examining attendance relative to class capacity.
- Quantifies the rate at which a class is filling its available capacity.



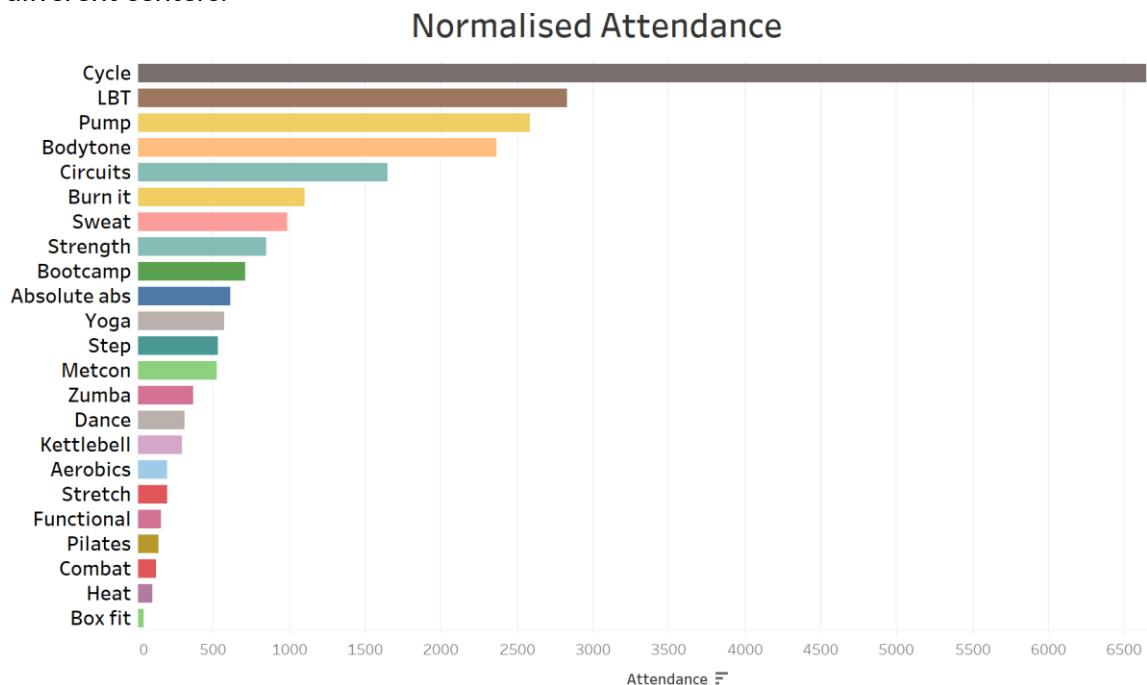
### 3. Waiting List Score

The proportion of members on a class waiting list to attendance.



### 4. Normalised Attendance Score

The overall attendance for each class, adjusted to consider variations in class availability across different centers.

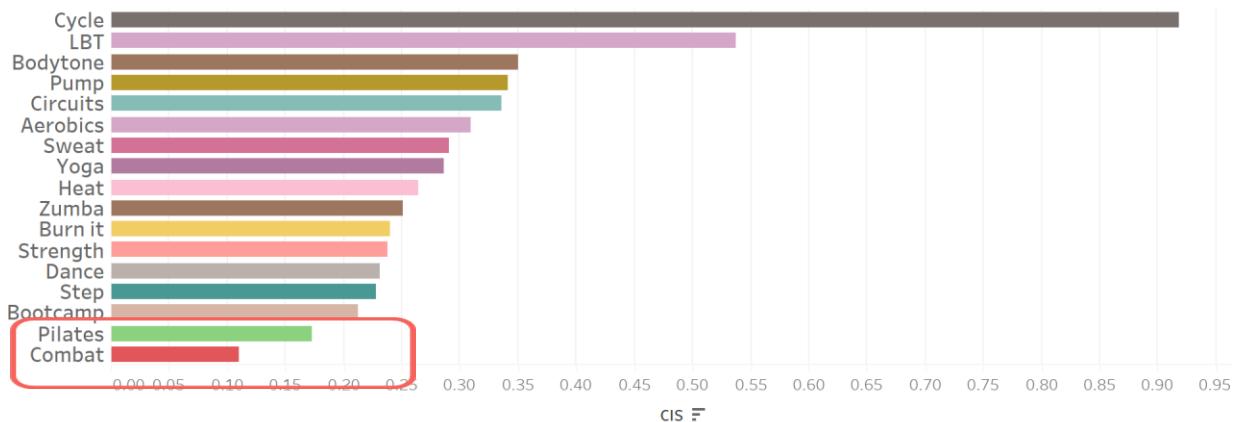


## High Ranking Extra Classes

Class Name	Normalised Attendance Rank (of 31)	Growth Score Rank (of 31)	Capacity Utilisation Rank (of 31)	Waitlist Ratio Rank (of 31)	CIS Rank (of 31)	CIS
Absolute abs	10	8	11	17	11	0.238
Metcon	13	10	7	10	12	0.237
Box fit	27	19	2	4	15	0.222
Kettlebell	16	17	10	6	19	0.197
Stretch	18	9	8	18	20	0.192

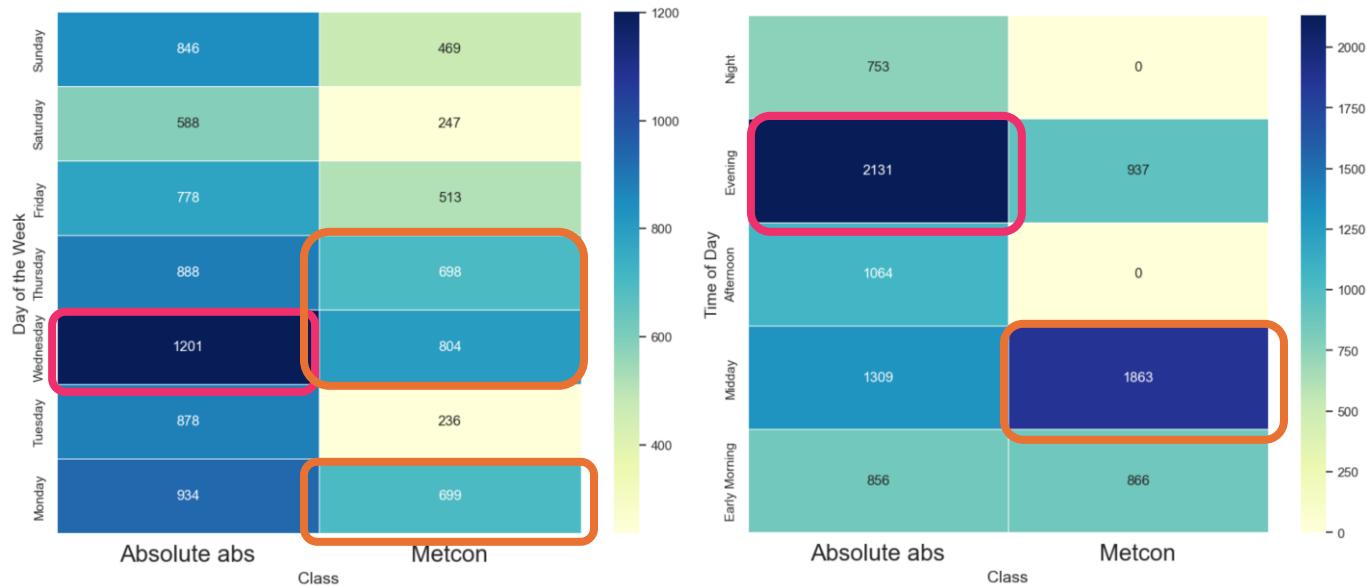
## Low Ranking Timetable Classes

Timetable Classes



Class Name	Normalised Attendance Rank (of 31)	Growth Score Rank (of 31)	Capacity Utilisation Rank (of 31)	Waitlist Ratio Rank (of 31)	CIS Rank (of 31)	CIS
Combat	22	30	27	25	29	0.098
Pilates	20	16	12	16	21	0.162

### Incorporating Absolute Abs and Metcon into the Timetable (3.3.8).



### CIS Model Limitations

Model Limitation	Explanation
Historic Data	The model relies heavily on historical data for predictions and assessments. While past trends can indicate future performance, they may not always capture sudden shifts in preferences or external factors (e.g., new fitness trends, seasonal variations, or unforeseen events like health crises).
Qualitative Data	Focusing on quantitative data like attendance, capacity, and waitlist numbers might overlook qualitative aspects that significantly affect class popularity and member satisfaction, such as class quality, instructor performance, or overall member experience.
Linear Regression	Assuming trends will continue linearly might not always hold true, especially in dynamic environments like fitness where new trends and preferences emerge regularly. The model might not quickly adapt to these nonlinear shifts. Similarly, the r-squared numbers vary in how well the model fits to the data.
Interdependencies	Classes might have interdependencies or complementary effects not accounted for in the model. For example, the popularity of one class could boost attendance in another (e.g., a strength class followed by a stretch class), which the model's current metrics might not capture.
Weighting Subjectivity	There is some subjectivity in the weighting with the WSM model. I have favored Growth Score and absolute attendance score over the other metrics. (other uses?)

## Conclusions

Conclusion	Explanation
<b>Class Scheduling and Planning</b>	The Metrics I've outlined guide the optimization of class schedules to align with member demand, ensuring popular classes are more accessible while reconsidering or reshaping those with declining interest.
<b>Resource Optimisation</b>	Insights from capacity utilization and attendance trends inform decisions on instructor allocation, space usage, and class size adjustments, aiming for optimal resource use that aligns with member preferences.
<b>Member Engagement and retention</b>	By continuously adapting to trends and member demands, the gym can enhance member satisfaction and engagement, potentially improving retention rates. Offering the right mix of classes encourages ongoing participation and attracts new members.
<b>Marketing and Promotion</b>	Understanding which classes have high demand but are at full capacity can guide targeted marketing efforts to attract members to similar classes with availability, maximizing attendance and utilization.
<b>Future Investments</b>	Long-term data on class performance and member preferences can inform decisions about investments in new facilities, equipment, or instructor training, ensuring they are aligned with what members want from their gym experience.
<b>Potential Improvements to the model.</b>	We believe the CIS methodology could be refined to incorporate analysis of retention rates, class duration, Instructor performance and cost per member. We would also consider the inclusion of qualitative metrics such as sentiment scores derived from analysis of customer and instructor reviews.

### 3.4 Extra Classes

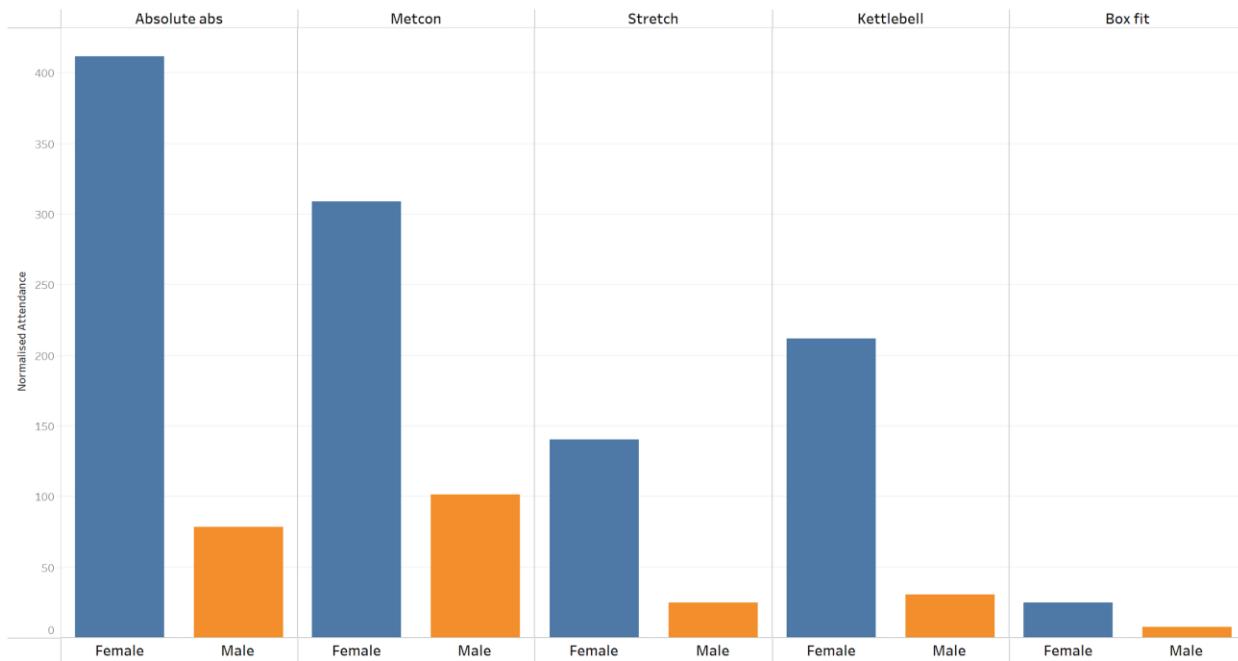
#### Analytical Approach

Process	Action
3.4 Extra Classes	Filter the class_data_cis to include only class names that exist in merged_data_scaled_extra. It then displays the filtered DataFrame.
3.4.1 Top Extra Classes by Gender	Aggregate total attendees by class name for females and males separately, then sort the results to see the most popular classes for each gender.
3.4.2 Top Extra Classes by Age	Aggregate participation by class and age group, then identify the most popular class by age. It excludes the total count of attendees from age groups and prepares the data for plotting.
3.4.3 Top extra classes - Attendance by Month	Convert the 'start_date_time' column to datetime format, extract months. Group by class name and month, sum the normalized_Total2 for monthly attendance. Plot the monthly attendance trends for the top extra classes.
3.4.4 Top extra classes - Attendance by day of week	Convert the 'start_date_time' column to datetime format, extract the day of the week, and map the day names. Group by class name and day name, sum the Total2 for daily attendance. Plot the daily attendance trends for the top extra classes with days of the week in the correct order.
Optimal Scheduling	Extracted the day of the week from the start_date_time and mapped it to day names (Monday to Sunday). Grouped the data by class_name along with day_name, summed the normalized_Total2 values for each group. Identified the category with the maximum summed normalized_Total2 for each class, indicating the 3 most effective days to schedule each class.
3.4.5 Top Extra Classes by time of day	Convert the 'start_date_time' column to datetime format, extract the time of day using function, and map the time of day. Group by class name and time of day, sum the Total2 for daily attendance. Plot the time of day attendance trends for the top extra classes in a heatmap.
Optimal Scheduling	Extracted the hour from start_date_time and categorized it into segments (Early Morning, Midday, Afternoon, Evening, Night). Grouped the data by class_name along with detailed_time_of_day, summed the normalized_Total2 values for each group. Identified the category with the maximum summed normalized_Total2 for each class, indicating the 2 most effective times to schedule each class.
3.4.6 Optimal Class Durations	Calculate Class Durations by finding the difference between start and end times and converting it to minutes. List All Class Durations - determined all unique durations for each class type to see the range of class lengths available. Determine Most Popular Class Lengths: Group data by class type and duration, and summing the Total2 values for these groups to find most popular class length for each class type based on participant numbers.

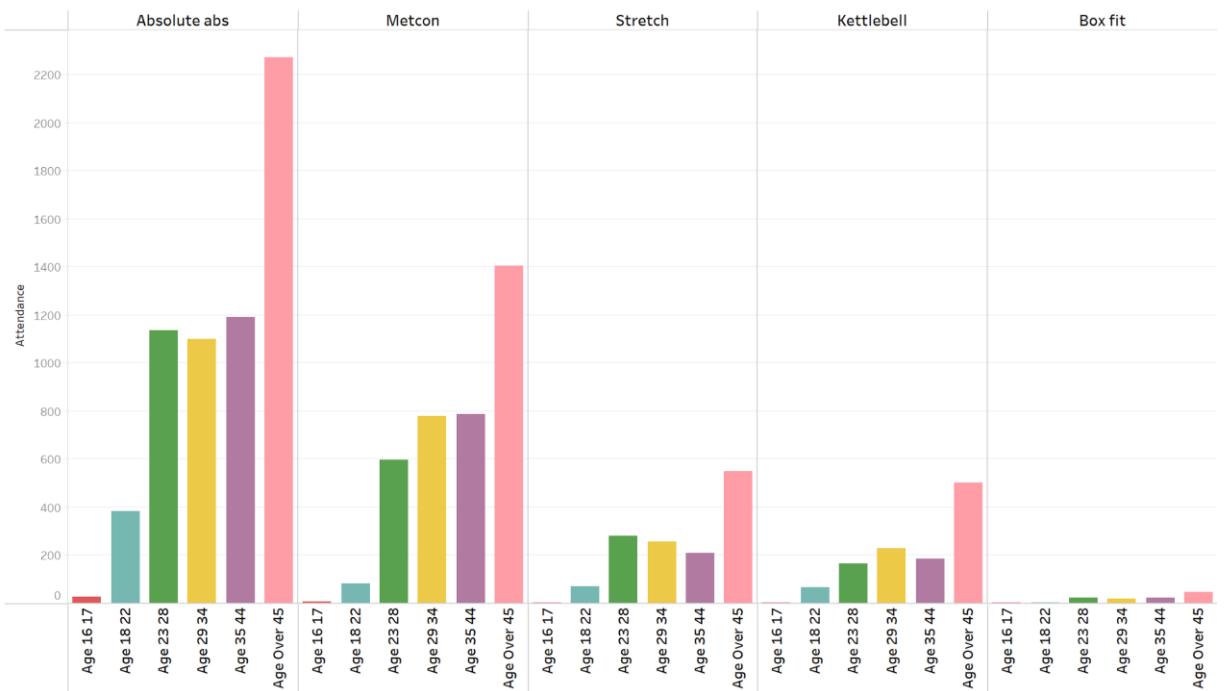
## Demographics

The top performing extra classes are broadly in line with Timetable classes in terms of demographics.

### 3.4.1 Top Extra Classes by Gender (Normalised)

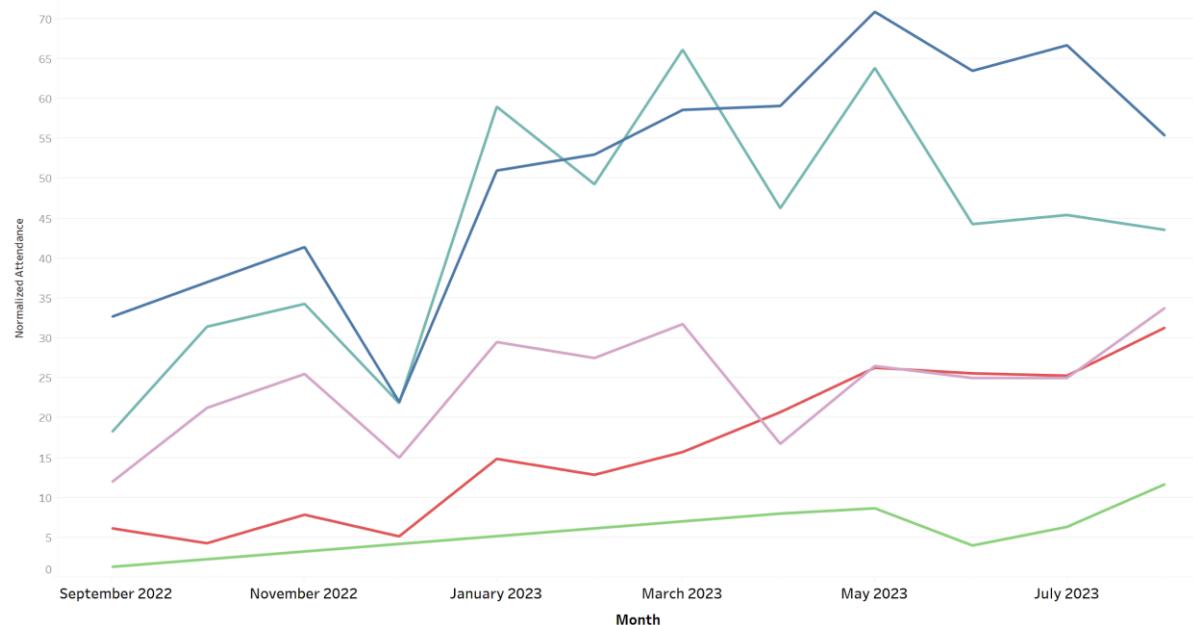


### 3.4.2 Top Extra Classes by Age

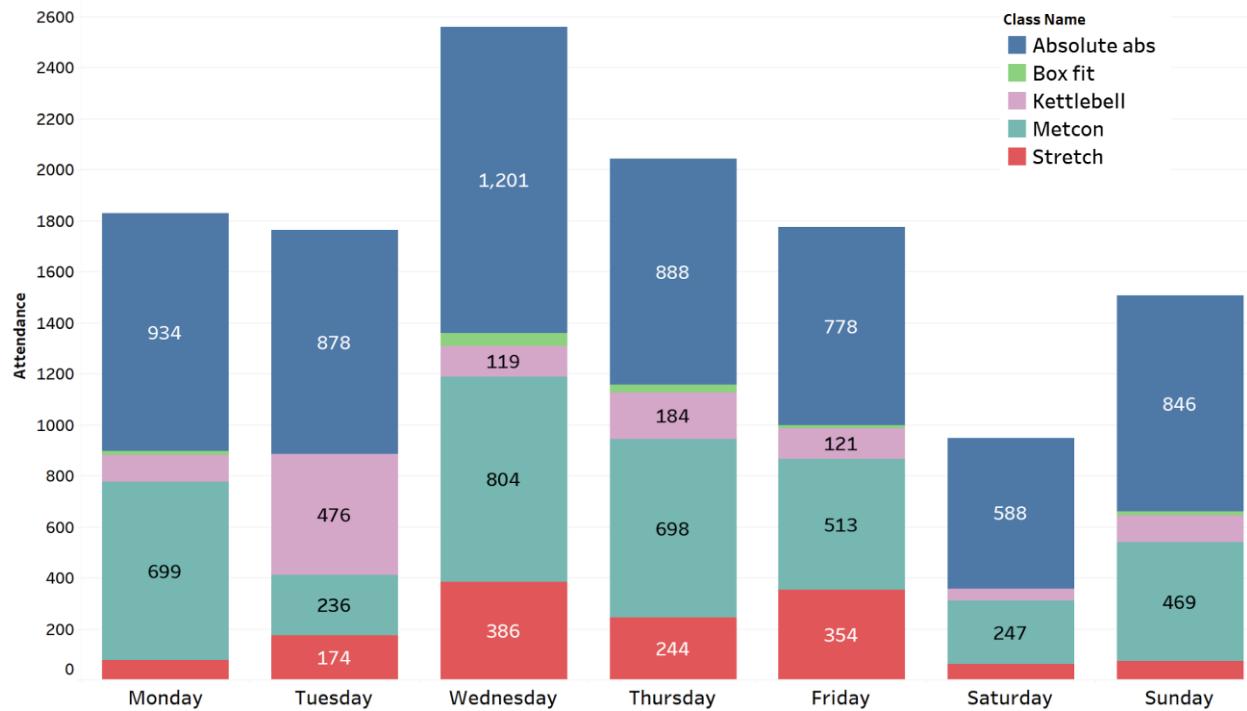


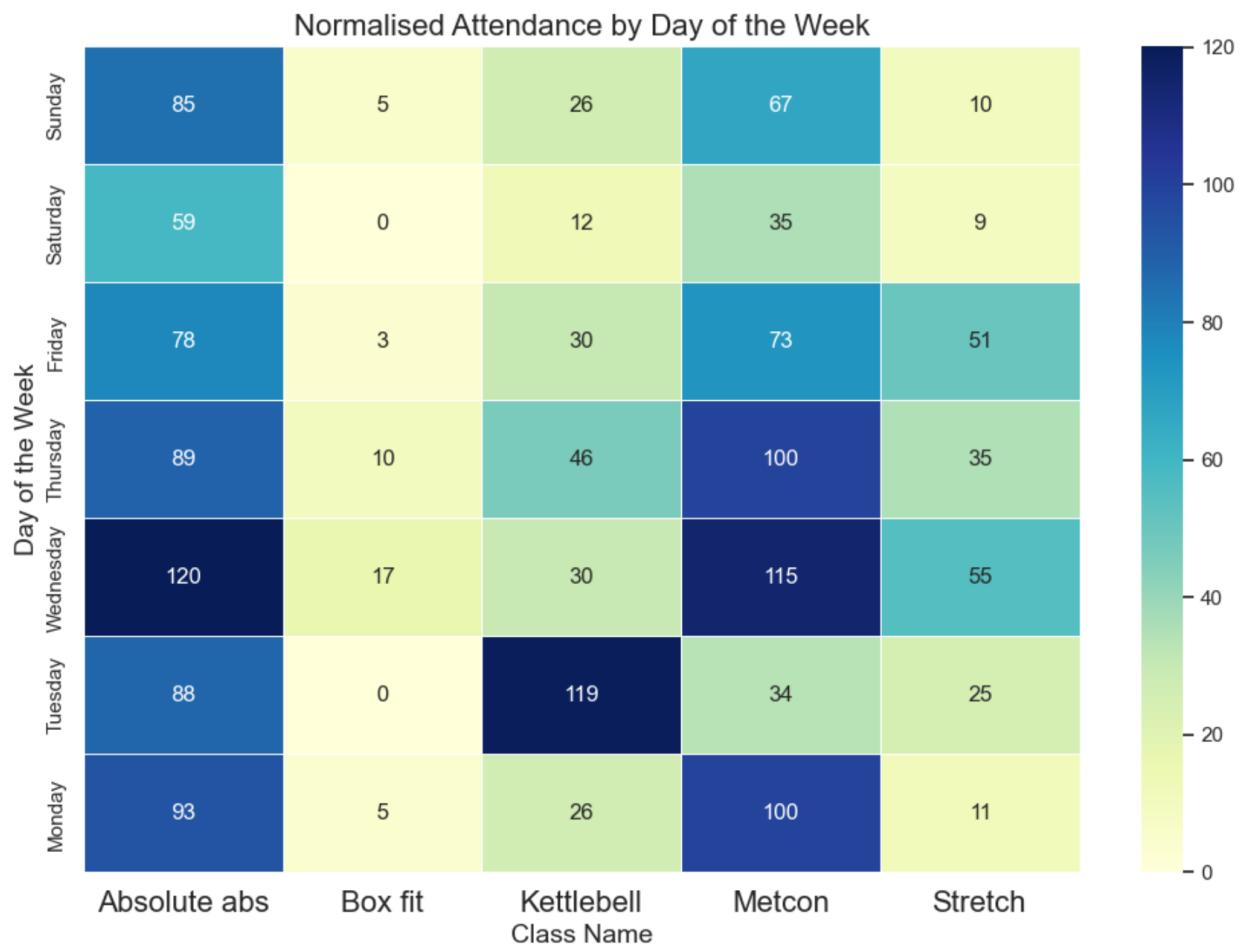
The top extra classes see a marked increase in January (note attendance figures normalised for comparison).

### 3.4.3 Top Extra Classes Attendance by Month



### 3.4.4 Top Extra Classes Attendance by Day of Week

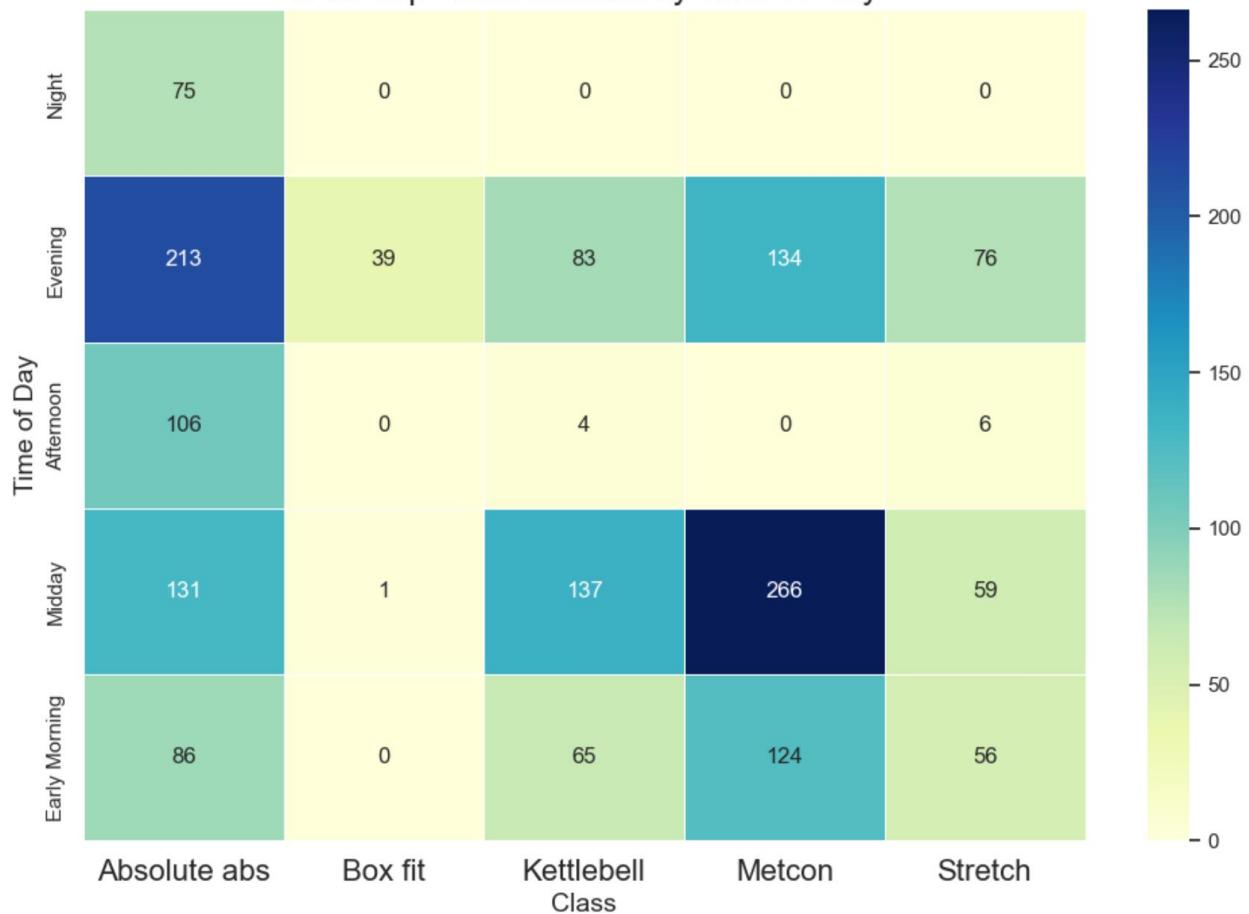




### Optimal Scheduling – Day of Week

class_name	Best Day	Second Best Day	Third Best Day
Absolute abs	Wednesday	Monday	Thursday
Box fit	Wednesday	Thursday	Sunday
Kettlebell	Tuesday	Thursday	Friday
Metcon	Wednesday	Monday	Thursday
Stretch	Wednesday	Friday	Thursday

### 3.4.5 Top Extra Classes by Time of Day



### Optimal Scheduling – Time of Day

class_name	Best Slot Time	Second Best Slot Time
Absolute abs	Evening	Midday
Box fit	Evening	Midday
Kettlebell	Midday	Evening
Metcon	Midday	Evening
Stretch	Evening	Midday

### 3.4.6 Optimal Class Duration

Most Classes have variable class lengths:

Class Name	Class Durations
Absolute abs	[15.0, 20.0, 30.0]
Box fit	[30.0, 45.0]
Kettlebell	[30.0, 45.0]
Metcon	[30.0, 45.0, 60.0]
Stretch	[15.0, 30.0]

The optimal class durations are:

Class Name	Optimal Class Duration
Absolute abs	30.0
Box fit	30.0
Kettlebell	30.0
Metcon	45.0
Stretch	30.0

## 4. Recommendations

### 4.1 Off-Peak Product

1. Centre 168: Implement targeted promotions to highlight the value of Off-Peak hours. Experiment with 'Off-Peak Fridays' to potentially extend Off-Peak benefits and assess member response.
2. Centres 174 and 264: Consider reassessing or discontinuing the Off-Peak membership based on its low usage and the negligible price difference to the Core product.
3. Extend the review to include all PureGym locations to better understand local competition and price sensitivity, ensuring that the strategy is tailored effectively.

The approach of trialling initiatives will ensure that we gain feedback and minimise risk to the key Core product members.

### 4.2 Operational Adjustments

1. 24-Hour Access: Conduct a cost-benefit analysis of maintaining 24-hour operations, especially during low-usage late-night hours.
2. Saturday Morning Peaks: Align staffing and resources to accommodate the surge in Saturday morning usage. Conduct further analysis to understand if it is driven by class availability and consider future inclusion in peak time schedule.

### 4.3 Class Adjustments

#### Primary

1. Incorporate Absolute Abs and Metcon into the Timetable in accordance with the optimal scheduling in 3.4.4, 3.4.5 and 3.4.6.
2. Remove Combat from the Timetable.
3. Consider removing Pilates from the Timetable / Monitoring for deteriorating metrics.

#### Secondary

1. Monitor Stretch for Potential Inclusion.
2. Extend Test Bed for Kettlebell and Box Fit.

#### 4.4 Conclusions

The analysis underscores the importance of aligning gym offerings with member preferences and local market dynamics. PureGym should consider both operational efficiencies and member satisfaction when adjusting membership types, facility access hours and class availability.

#### 4.5 Future Considerations

Continued analysis is recommended to refine the understanding of member preferences and to ensure that PureGym remains competitive and responsive to the needs of its members. The Core product stands out as the member preference in delivering PureGym's proposition, so it is important that this remains a primary focus in any analysis and change.

## 5. Appendix

### 5.1 Data Cleaning Checklist

Stage
<b>Data Accuracy</b>
Check for spelling errors. Values out of range Incorrect or invalid data types Blank cells or spaces Incorrect use of nulls Incorrect calculations Mistypes and other format errors.
<b>Data Completeness</b>
Blanks in a required field Partial or incomplete data Incorrect or invalid calculations Missing values.
<b>Data Consistency</b>
Precision Structure of data Case sensitivity Data type General Consistency
<b>Data Uniqueness</b>
Entries with the same spelling but in a different case. Entries with different spelling. <b>Different words but with the same meaning.</b> Words with alternate representations.
<b>Duplicate Values</b>
<b>Data timeliness</b>
Correct date format and type.

## 5.2 Analytical Approach to Timetable Classes

Process	Action
<b>Initial Imports</b>	<p>Import Libraries</p> <pre>import pandas as pd, import numpy as np, import matplotlib.pyplot as plt, import seaborn as sns.</pre> <p>Import Data</p> <pre>class_data_split = pd.read_excel('class_data_split.xlsx')</pre>
<b>3.1.0 – Preparing Class Data for Timetable Insights</b>	<p>Define Mapping</p> <p>Standardize Class Names</p> <p>Check Mapping of Timetable Classes</p> <p>Convert to Datetime</p> <p>Count Monthly Attendance</p> <p>'centre_id' to 'location' – Using the '.map()' function to map the centre ID to a new location column for ease of reference.</p> <p>Implementing a function to normalize and map timetable class names to create 'standardized_class_name' column.</p> <p>for cls in unique_classes: print(cls) to check mapping of standardized classes.</p> <p>Using 'pd.to_datetime()' to convert 'start_date_time' to datetime format and extract year-month for grouping</p> <p>Filtering for the 'ATTENDED' 'participation_status' and sum the 'Total2' column to get attendance counts by class, location, and month.</p> <p>Assigning pathway to path = to save then save the 'timetable_class_data_clean' DataFrame for future use</p>
<b>3.1.1 – Attendance Rates</b>	<p>Calculate rates by location.</p> <p>Analyse Rates by Age Group</p> <p>Sum 'Total2' for 'ATTENDED' and 'CANCELLED' statuses by location, calculate rates by class, and sort/print by location.</p> <p>Calculate and display attendance and cancellation rates by age group using 'Total2'.</p>

<p><b>3.1.2 – Correlation of Attendance with Time of Day / Day of Week</b></p> <p>Prepare Data for Heatmap</p> <p>Generate Heatmaps (All Locations and Specific)</p> <p>Generate Combined Heat Map</p> <p>Visualize Location-Specific Patterns</p>	<p><u>Time of Day</u></p> <p>Convert 'start_date_time' to datetime, extract hour, and create heatmap data table of median 'Total2' for 'ATTENDED' status across all locations</p> <p>Plot heatmaps for median attendance by time, adjusting figure size by class count, using 'plot_heatmap(heatmap_data, location)' for each location (1-10).</p> <p>Utilise 'prepare_data_for_location(location)' to filter and pivot data for mean attendance by time for classes at each specific location</p> <p><u>Day of Week</u></p> <p>Convert 'start_date_time' to extract 'hour_of_day' and 'day_of_week' and group data by class, time, and day.</p> <p>Map 'day_of_week' to a numerical using 'pd.categorical()' for visualization.</p> <p>Pivot data to visualize mean attendance by class and day with 'sns.heatmap()'. Loop through locations, pivot data, , and plot heatmaps showing mean attendance by day for each class.</p>
--	---

<p><b>3.1.3 – Patterns that Suggest Optimal Scheduling for Certain Demographics</b></p> <p>Time Series Analysis</p> <p>Demographic Attendance Distribution</p> <p>Weekly Attendance Patterns</p>	<p>Extract 'hour_of_day' from 'start_date_time', calculate mean attendance per age group, and plot with 'plt.plot()'.</p> <p>Aggregate attendance by class and hour, and visualize as stacked bar charts by age bands using 'groupby', 'sum', and 'ax.bar'</p> <p>Filter data by location and day, and visualize total attendance by age demographic as stacked bars for each day, looping over 'locations' and 'day_order'</p>
<p><b>3.1.4 – Which Classes are Trending in Terms of Attendance Over Time?</b></p> <p>Daily Average Attendance</p> <p>Seasonal Attendance Trends</p> <p>Waiting List Sizes</p> <p>Waiting List Trends Over Time</p> <p>Average Attendance per Session</p>	<p>Convert 'start_date_time' to datetime, sum 'Total2' by date, and plot daily totals and a 7-day rolling average using 'rolling().mean()' and 'plot(kind='line')</p> <p>Convert 'year_month' to datetime, extract and order months for a specific timeframe, and create line plots for each locations on a subplot grid using 'sns.lineplot()'</p> <p>Group by 'location' and 'standardized_class_name' to sum 'Total2' for waiting lists and visualize using 'sns.barplot()'</p> <p>Filter for 'WAITING LIST', plot waiting list sizes over time for each class using 'sns.lineplot()'</p> <p>Aggregate attendance and session count by 'standardized_class_name', calculate average per session, and identify the bottom five classes using 'sort_values()' and 'head()'</p> <p>Group by 'location' and 'standardized_class_name', calculate and sort the average attendance per session, and display the worst-performing classes for each location</p>

## 5.3 Workbook submissions

### 5.3.1 Products

- Jupyter Notebook: PureAnalytics\_LSE\_EP\_Assignment3\_Product

Section	Subsection	Description
1. Import the Visits Data		Perform basic exploration before data cleaning
	1.1 Import the Data	Basic data importation steps and initial checks
	1.2 Check for missing values and duplicates	Procedures to identify and handle missing values and duplicates
	1.3 Remove missing values and duplicates	Steps taken to clean the data from missing values and duplicates
	1.4 Convert to Datetime	Convert data fields to appropriate DateTime formats
	1.5 Add duration calculation	Calculate duration based on datetime fields
	1.6 Check data integrity	Ensure data consistency and accuracy
	1.7 Check distribution of duration data	Analyze the spread and characteristics of duration data
2. Import the Gyms Data		Perform basic exploration before data cleaning
	2.1 Import the Data	Import gyms data and initial handling
	2.2 Clean the data	Data cleaning specific to gyms data
	2.3 Remove and/or edit columns	Modify or remove unnecessary columns
3. Merge the Visit and Gym Data		Into single dataframe
	3.1 Merge using a Left Join	Method of merging and rationale
	3.2 Edit Data Types in the merged dataframe	Adjust data types post-merge to ensure consistency
	3.3 Add a new columns for Day of Week and Peak/Off-Peak	Enhance data with additional informative columns
	3.4 Final clean of columns and names	Final adjustments to column names and data cleanup
	3.5 Extract to CSV	Export the cleaned and merged data to CSV format
4.0 Data Exploration		Exploratory data analysis of the merged dataset
	4.1 Basic Exploration	Initial exploratory analysis
	4.2 Peak vs Off-Peak sage	Analysis of gym usage during different times
	4.3 Gym Visits based on Facilities	Influence of gym facilities on visitor counts
	4.4 Visitor Age	Analysis based on the age distribution of visitors
	4.5 Visitors by Gender	Gender distribution among gym visitors
	4.6 Gender and Functional Preferences	Preferences for gym activities based on gender
	4.7 Peak visits by Off-Peak members at 168	Specific analysis of off-peak membership usage during peak times
	4.8 Gym Busyness	Assessment of gym busyness at various times
5.0 Extraction of data to Tableau		Data preparation for further analysis in Tableau

- Tableau Workbooks:
  - Off\_Peak\_Analysis\_vF
  - Visit Analysis

Workbook	Sheet	Description
Off_Peak_Analysis_vF		Left Join of the Python Extract “final_average_visitor_counts” and “visit_clean”) to visualise Off-Peak trends
	Peak Months	Visualise visit trends by month
	Peak Days	Visualise visit trends by day of week
	Day of Week	Visualise visit trends by center
	Day of Week_2	Visualise visit trends by center by day of week
	Facilities	Visualise visits based on facilities available
Visit Analysis		Uses the “visit_clean” extract to visualise additional insights on Gym demographics
	Gender	Visualise visits by gender
	Duration	Visualise visit duration by demographics

- Excel Workbooks: PureAnalytics\_LSE\_EP\_Assignment3\_PricingResearch.xlsx
  - Product research across 130 gyms

### 5.3.2 Class Analysis

- Jupyter Notebook: PureAnalytics\_LSE\_EP\_Assignment3\_Class Analysis.ipynb
- Tableau Workbooks: PureAnalytics\_LSE\_EP\_Assignment3\_Class\_Analysis.twbx  
PureAnalytics\_LSE\_EP\_Assignment3\_Class\_Space\_Analysis.twbx