

BAIS:3200

Gym Membership Project Report

Ella Zaputil, Baylie Schnieder, Cassady Jackson, and Lauren Brodsky

Introduction: Many people around the world spend hours in the gym every week, what are they doing while they're there? In this project, we are diving into the commonalities among the individuals who are going to the gym. What day of the week is the most popular? What is the average age of those who favor each class? This database will help gym owners or exercise enthusiasts understand the trends happening within gyms. It will be useful in understanding what will draw members in, and what should be prioritized. This database contains data that has been cleaned for ease of analysis.

Data: Our team is utilizing the 2024 Gym Membership Dataset from Kaggle¹ for this project. This dataset contains synthetic data that simulates the average membership data of a gym. The original data contained 17 columns, including those detailed in *Table 1* below, and those we found not prevalent in our study including the use of the drink bar, favorite drink flavor, use of the sauna and check-in and check-out times. We also added a surrogate key, TrainerID, to uniquely identify each personal trainer employed at the gym. We cleaned the data and are utilizing Microsoft Excel to store the cleaned Gym Membership data². The data contains 1,000 unique instances, each instance identified by the Member ID number. Additional fields, data types, and a description of the data are detailed in *Table 1*.

Table 1 - Data Dictionary

<u>Field</u>	<u>Data Type</u>	<u>Description</u>
MemberID	Numeric	MemberID Number that uniquely identifies each gym member
First_Name	Text	Gym member's first name
Gender	Text	Gym member's gender
Birthday	Date	Gym member's date of birth
Abonoment_Type	Text	Specifies type of gym membership the member has
Visit_per_Week	Numeric	Number of days per week the

¹ [Gym Membership Dataset \(kaggle.com\)](https://www.kaggle.com/datasets/andrewyeh/2024-gym-membership-dataset)

² [Clean Gym Membership Data Spreadsheet](#)

		member visits the gym
Days_per_Week	Text	Specifies the days during the week the member attended the gym
Attend_Group_Lesson	Binary	1 indicates the member attended a group lesson, 0 indicates they did not
Fav_Group_Lesson	Text	Gym member's favorite group class
Avg_Time_In_Gym	Numeric	Returns the average time the member spends at the gym, in minutes
Personal_Training	Binary	1 indicates the member has a personal trainer, 0 indicates they do not
Trainer_ID	Alphanumeric	ID that uniquely identifies personal trainers
Name_Personal_Trainer	Text	First name of the personal trainer

This data model contains two entities: MEMBER and TRAINER. The MEMBER entity is identified by a strong primary key, MemberID. The TRAINER entity is also identified by a strong primary key, TrainerID, which was created as a surrogate key to uniquely identify trainers.

Most attributes in both entities are mandatory and simple. However, there are two multivalued attributes in MEMBER, DaysPerWeek and FavGroupLesson. FavGroupLesson is also an optional attribute.

The relationship between the two entities is a mandatory many to optional one relationship. In other words, each member is not required to have a trainer, but a trainer must have at least one member, and can train multiple members.

Entity-Relationship Diagram

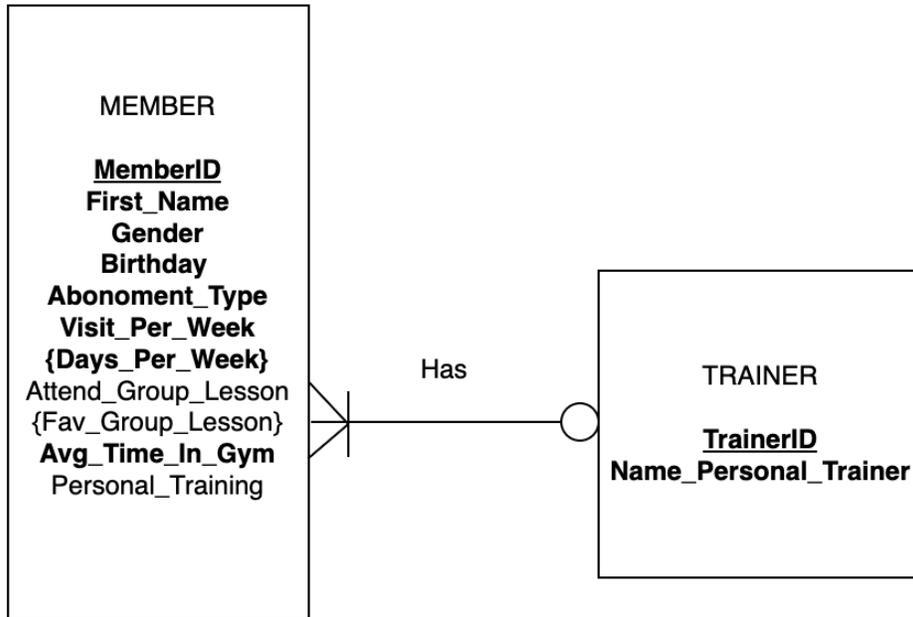


Figure 1: Entity-Relationship Diagram for Gym Membership dataset

From the ERD, the data was normalized to create a relational schema with four tables. The two multivalued attributes, Days_Per_Week and Fav_Group_Lesson, were resolved by creating two additional tables: DAYS and LESSONS. The TRAINER table serves as the parent table to MEMBER, while MEMBER acts as the parent table to the child tables DAYS and LESSONS.

Graphical Relational Schema

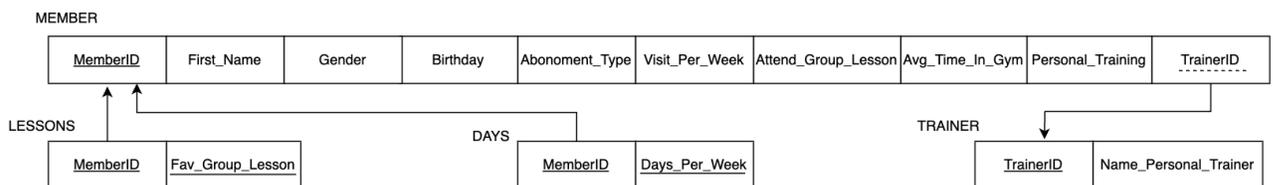


Figure 2: Graphical Relational Schema For Gym Membership dataset

Data Implementation:

To implement the database in APEX, we wrote CREATE TABLE commands for each table in the relational schema. Since TRAINER is the parent table to MEMBER, it was created first. Next, the MEMBER table was created as it serves as the parent table for DAYS and LESSONS. Then, DAYS and LESSONS were created in no particular order.

TRAINER

```
CREATE TABLE TRAINER (  
    TrainerID VARCHAR2(5)  
    Name_Personal_Trainer VARCHAR2(10) NOT NULL  
    CONSTRAINT TRAINER_PK PRIMARY KEY (TrainerID));
```

MEMBER

```
CREATE TABLE MEMBER (  
    MemberID NUMBER(3, 0),  
    First_Name VARCHAR2(50) NOT NULL,  
    gender VARCHAR2(10) NOT NULL,  
    Birthday DATE,  
    Abonement_Type VARCHAR2(20) NOT NULL,  
    Visit_Per_Week NUMBER NOT NULL,  
    Attend_Group_Lesson NUMBER(1),  
    avg_time_in_gym NUMBER,  
    personal_training NUMBER(1),  
    TrainerID VARCHAR2(5),  
    CONSTRAINT MEMBER_PK PRIMARY KEY (MemberID)  
    CONSTRAINT FK_Trainer FOREIGN KEY (TrainerID) REFERENCES  
    TRAINER(TrainerID) );
```

DAYS

```
CREATE TABLE DAYS (  
    MemberID NUMBER,  
    days_per_week VARCHAR2(10) NOT NULL,  
    CONSTRAINT DAYS_PK PRIMARY KEY (MemberID, days_per_week),  
    CONSTRAINT DAYS_FK FOREIGN KEY (MemberID) REFERENCES  
    MEMBER(MemberID)  
);
```

LESSONS

```
CREATE TABLE LESSONS (  
    MemberID NUMBER,  
    fav_group_lesson VARCHAR2(20) NOT NULL,  
    CONSTRAINT PK_Lessons PRIMARY KEY (MemberID, fav_group_lesson),  
    CONSTRAINT FK_Lessons FOREIGN KEY (MemberID) REFERENCES  
    MEMBER(MemberID)  
);
```

Each table uses appropriate data types and field sizes to ensure data is accurate and logical. For example, Visit_Per_Week is a number that must fall within realistic limits, and binary fields like personal_training are restricted to 0 or 1. These constraints help maintain consistent and valid data.

After all the tables were created, the normalized data was loaded into APEX. The TRAINER table contains 4 rows of data. Next, MEMBER was loaded into APEX and returned 1,000 rows of data. DAYS was created and returned 2,680 rows of data. Finally, LESSONS was added and returned 1,040 rows of data. Below is an example of using the INSERT command for each of the four tables.

TRAINER

```
INSERT INTO TRAINER (TrainerID, Name_Personal_Trainer)
VALUES ('T1', 'Chantal');
```

MEMBER

```
INSERT INTO MEMBER (Member_ID, First_Name, gender, Birthday, Abonement_Type,
Visit_Per_Weew, Attend_Group_Lesson, avg_time_in_gym, personal_training, Trainer_ID)
VALUES (1, 'Fey', 'Female', TO_DATE('1997-04-18', 'YYYY-MM-DD'), 'Premium', 4, 1, 131,
NULL, NULL);
```

DAYS

```
INSERT INTO DAYS (MemberID, days_per_week)
VALUES (71, 'Fri');
```

LESSONS

```
INSERT INTO LESSONS (MemberID, fav_group_lesson)
VALUES (911, 'XCore');
```

Analysis

This analysis helps gym owners understand trends to improve services. Each question further explains those who are attending the gym, and what activities they favor.

Question 1 Demographics:

What percentage of Premium members are male and female? We wrote a subquery that calculated the percentage of Premium members for each gender and displayed the results as a percent rounded to two decimal places.

```

SELECT gender, ROUND ((COUNT(*) * 100.0) / (SELECT COUNT(*) FROM MEMBER
WHERE Abonement_Type = 'Premium'), 2) AS percent_premium
FROM MEMBER
WHERE Abonement_Type = 'Premium'
GROUP BY gender;

```

GENDER	PERCENT_PREMIUM
Male	50.91
Female	49.09

Figure 3: Percentage of Premium Members by Gender

As reflected in Figure 3, the percentage of premium subscribers by gender reveals a fairly even split between male and female subscribers. This reflects that 50.91% of the premium members are males, while 49.09% were females. Therefore, we can conclude that both genders value the additional benefits of a premium membership at similar rates.

Question 2 Trainer Usage:

What percentage of Premium and Standard members use personal trainers? To answer this question, we grouped members by subscription type and calculated the percentage of members in each subscription type who use a personal trainer. This was done by dividing the sum of the personal_training column by the total count of members in each subscription type. The results were then multiplied by 100 to show as percentages and rounded to two decimal places for clarity.

```

SELECT Abonement_Type AS Subscription_Type, ROUND((SUM(personal_training) * 100.0)/
COUNT(*), 2) AS percent_uses_Personal_trainer
FROM MEMBER
GROUP BY Abonement_Type;

```

SUBSCRIPTION_TYPE	PERCENT_USES_PERSONAL_TRAINER
Premium	54.16
Standard	49.51

Figure 4: Personal Training Usage by Subscription Type

Based on the results from Figure 4, a higher percentage of Premium members use a personal trainer (54.16%) compared to Standard members (49.51%). This indicates that Premium members are more likely to take advantage of additional services like personal training.

Additionally, Gym management could consider creating programs to boost personal training usage, as only about half of each subscription type currently utilizes this service.

Question 3 Lesson Preference:

What percentage of premium members prefer the top 5 classes? This query answers what the top 5 classes are for premium members at the gym. This required a join query, between the Lessons table and the Member table, joined on MemberIDs. The query then filtered the data based on Premium membership status, and grouped the members based on their favorite group lesson. It also includes a subquery, used to calculate the percentage of premium members that were represented in each favorite lesson category. It then fetched only the first 5 rows from the data to show what the top 5 classes are for premium members.

```
SELECT Fav_Group_Lesson, ROUND((COUNT(*)*100.0)/(SELECT COUNT(*) FROM
MEMBER WHERE Abonement_Type = 'Premium'),2) AS PercentPremium
FROM LESSONS JOIN MEMBER ON LESSONS.MemberID = MEMBER.MemberID
WHERE Abonement_Type = 'Premium'
GROUP BY Fav_Group_Lesson
ORDER BY PercentPremium DESC
FETCH FIRST 5 ROWS ONLY;
```

FAV_GROUP_LESSON	PERCENTPREMIUM
BodyPump	13.39
Yoga	11.56
Kickboxen	11.56
Spinning	9.94
XCore	9.94

Figure 5: Lesson Preference by Percentage of Premium Members

The query results show that there is an even percentage of premium members who prefer Yoga and Kickboxing, as well as an even percentage of premium members who prefer Spinning and XCore. The gym owners could use this information to provide more of premium member’s favorite classes, Body Pump to draw more members in.

Question 4 Gym Usage by Day:

This query answers which day of the week is the busiest at the gym. This query involved using multiple case statements within the select clause to check for each day of the week, assign them as 0 or 1, and then summing each column to output the total number of gym users per weekday.

```
SELECT
SUM(CASE WHEN Days_Per_Week = 'Sun' THEN 1 ELSE 0 END) AS Sunday,
```

```

SUM(CASE WHEN Days_Per_Week = 'Mon' THEN 1 ELSE 0 END) AS Monday,
SUM(CASE WHEN Days_Per_Week = 'Tue' THEN 1 ELSE 0 END) AS Tuesday,
SUM(CASE WHEN Days_Per_Week = 'Wed' THEN 1 ELSE 0 END) AS Wednesday,
SUM(CASE WHEN Days_Per_Week = 'Thu' THEN 1 ELSE 0 END) AS Thursday,
SUM(CASE WHEN Days_Per_Week = 'Fri' THEN 1 ELSE 0 END) AS Friday,
SUM(CASE WHEN Days_Per_Week = 'Sat' THEN 1 ELSE 0 END) AS Saturday
FROM DAYS;

```

SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
379	402	394	381	359	379	386

Figure 6: Gym Usage by Day of the Week

The query results show that Monday has the most gym usage with 402 attendees and Thursday has the least with 359 uses. Gym use seems to spike on Monday, then decrease throughout the week, then increase again during the weekend. The gym owners could take this information and create special workout classes or deals to draw people in on Wednesdays and Thursdays when the gym is less busy.

Question 5 Lesson by Age:

What is the average age for each lesson? This query answers the question of the average age of the people who take each lesson that is offered by the gym. This query completes that task by taking the age of each individual from their Birthday information and then taking the floor then the average of each birthday and dividing it by 365 days of the year. It is grouped by Favorite group lesson to see what the average age of each birthday is.

```

SELECT Fav_Group_Lesson, FLOOR(AVG(((SYSDATE - Birthday)/365))) as AvgAge
FROM MEMBER JOIN LESSONS ON MEMBER.MemberID = LESSONS.MemberID
WHERE Fav_Group_Lesson IS NOT NULL
GROUP BY Fav_Group_Lesson;

```

FAV_GROUP_LESSON	AVGAGE
Spinning	30
Zumba	32
Running	30
HIT	32
BodyBalance	30
XCore	30
Yoga	30
BodyPump	30
Pilates	32
LesMiles	30
Kickboxen	31

Figure 7: Average Age of Lesson Groups

The query results show the highest average age of any of the members is 32 with the lowest at 30. Individuals in the 32 range who attend the gym appear to enjoy Zumba, HIT and Pilates for their favorite lessons. Individuals 30 appears to enjoy Spinning, Running, BodyBalance, XCore, Yoga, BodyPump and LesMiles. The owners can take from these results that the average attendee's of their classes is in their 30's and promote towards that age demographic.

Question 6 Membership Type Distribution:

How many members are standard and premium? This query answers the question of what the membership distribution looks like for the gym. It does this by counting the number of members who have premium or standard memberships, then summarizing the count of each type.

```
SELECT Abonement_Type as Subscription_Type, COUNT(Abonement_Type) as  
Count_Members  
FROM MEMBER  
GROUP BY Abonement_Type  
ORDER BY Count_Members DESC;
```

SUBSCRIPTION_TYPE	COUNT_MEMBERS
Standard	507
Premium	493

Figure 8: Count of Members by Subscription Type

The query results show that there is a pretty even distribution between premium and standard memberships at this gym. The gym owners could use this information to decide which membership preferences they would like to focus on. They could try to draw in more premium members by highlighting their preferences, or could focus on retention of standard members.

Web Design

Home Page:

The home page of this web application for the Gym Membership database includes a brief project description with a hyperlink to the original Kaggle dataset. Navigation menus were created to organize and group tables together. On the homepage, there is an illustrated image of a gym. This image is from FREEP!K. In creating the web design, we ensured consistency in colors, theme, and format. *Figure 9* shows a screenshot of the homepage.

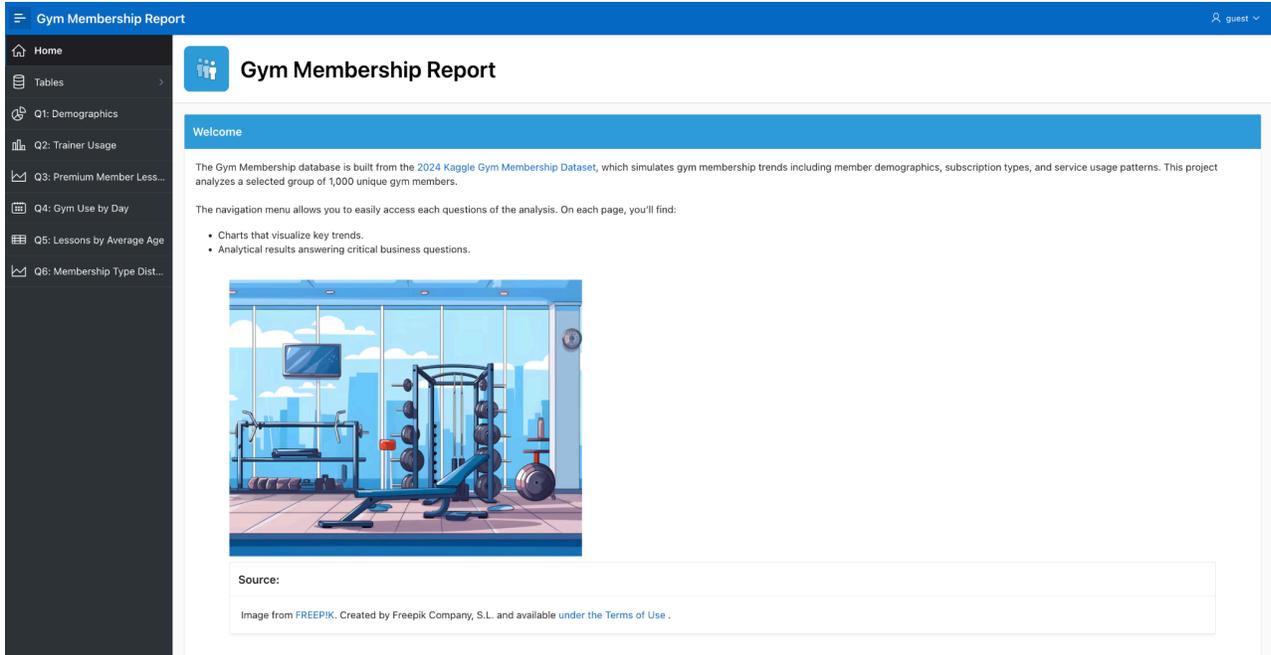


Figure 9: Home Page

Tables:

Tables were created in the web application (Figure 10-13). Users are able to filter, search, and group the data. The column headers were also updated for readability. Every page also has a brief description of the data and its importance.

TRAINER Table

The TRAINER table contains information about personal trainers employed at the gym. Each record includes a unique Trainer ID and the trainer's name. This table is essential for identifying trainers assigned to gym members.

Trainer ID	Personal Trainer Name
T1	Chantal
T2	Mike
T3	Jeffrey
T4	Hanna

Figure 10: TRAINER

MEMBER Table

The MEMBER table contains detailed information about gym members. Each record includes a unique Member ID, personal details (e.g., name, gender, and birthdate), and gym activity information such as subscription type, weekly visits, participation in group lessons, average time spent in the gym, and personal training status. Additionally, the table links members to their assigned trainer through the Trainer ID.

Member ID	First Name	Gender	Birthdate	Subscription Type	Visits Per Week	Attend Group Lesson	Average Time In Gym	Personal Training	Trainer ID
148	Hazlett	Male	3/22/1988	Premium	2	1	174	1	T2
149	Andrew	Male	6/29/1998	Standard	2		112		
150	Cirilo	Male	6/27/2005	Premium	1		82	1	T2
151	Alberto	Male	2/2/1995	Premium	2	1	56		
152	Marion	Male	9/25/1997	Premium	2		49	1	T4
153	Gail	Male	7/22/1977	Premium	2	1	35	1	T1
154	Burke	Male	3/16/1985	Standard	2	1	104	1	T2
155	Moe	Male	8/27/1977	Standard	1		33	1	T4
156	Jocelyne	Female	2/1/2004	Premium	2		70	1	T1
157	Jillie	Female	8/30/2004	Standard	4		82	1	T3
158	Torrance	Male	2/2/2002	Standard	2		85	1	T2
159	Ruthann	Female	6/3/2002	Standard	3	1	142	1	T4
160	Millie	Female	3/19/2000	Premium	3	1	91	1	T4
161	Adriane	Female	12/22/2006	Standard	3	1	154	1	T1
162	Peder	Male	1/27/1995	Standard	1		166		
163	Cesaro	Male	11/24/1987	Premium	3		154	1	T3
164	Quincy	Male	1/29/2004	Standard	5		147		
165	Christian	Male	5/22/1995	Standard	2	1	162	1	T1
166	Bordie	Male	9/19/1992	Premium	3		49	1	T1
167	Nessie	Female	10/15/1981	Premium	4	1	169		
168	Desi	Male	10/5/2010	Standard	3	1	69		

Figure 11: MEMBER

LESSONS Table

The LESSONS table captures information about the favorite group lessons attended by gym members. Each entry is linked to a unique Member ID and specifies the preferred group lessons, such as yoga, spinning, or kickboxing. This data helps identify trends in class preferences among members.

Member ID	Favorite Group Lesson
1	Kickboxen
3	LesMiles
3	Spinning
3	XCore
5	BodyBalance
5	Running
7	LesMiles
7	XCore
8	BodyPump
8	Running
8	Yoga
9	BodyPump
12	Kickboxen
12	LesMiles
12	Pilates
13	LesMiles
13	Spinning
13	XCore
16	LesMiles
16	Spinning
19	Pilates
20	BodyBalance

Figure 12 LESSONS

DAYS Table

The DAYS table records the specific days of the week that each gym member visits the gym. Each entry is linked to a unique Member ID and lists one or more days per member, providing insights into gym attendance patterns and scheduling preferences.

Member ID	Days Per Week
536	Fri
536	Mon
536	Thu
537	Fri
537	Sat
537	Thu
538	Fri
538	Thu
539	Fri
540	Mon
541	Mon
541	Tue
542	Tue
543	Sat
543	Sun
544	Fri
544	Mon
545	Tue
546	Wed
547	Fri
547	Sat
548	Mon

Figure 13: DAYS

Queries:

The results of the first question were presented as a pie chart so that viewers can easily see the breakdown of the percentage of men and women who have a Premium membership. Question 1 was placed on its own page (Figure 14). Additionally, above the pie chart is a text box explaining the business question, and below it, the answer is stated. This way, viewers can quickly understand and compare the gender distribution among Premium members.

Question 1 Demographics:

Business Question: What percentage of Premium members are male and female?

Answer: As reflected in the pie chart below, the percentage of Premium subscribers by gender reveals a fairly even split between male and female subscribers. This reflects that 51% of the Premium members are males, while 49% are females. Therefore, we can conclude that both genders value the additional benefits of a Premium membership at similar rates.

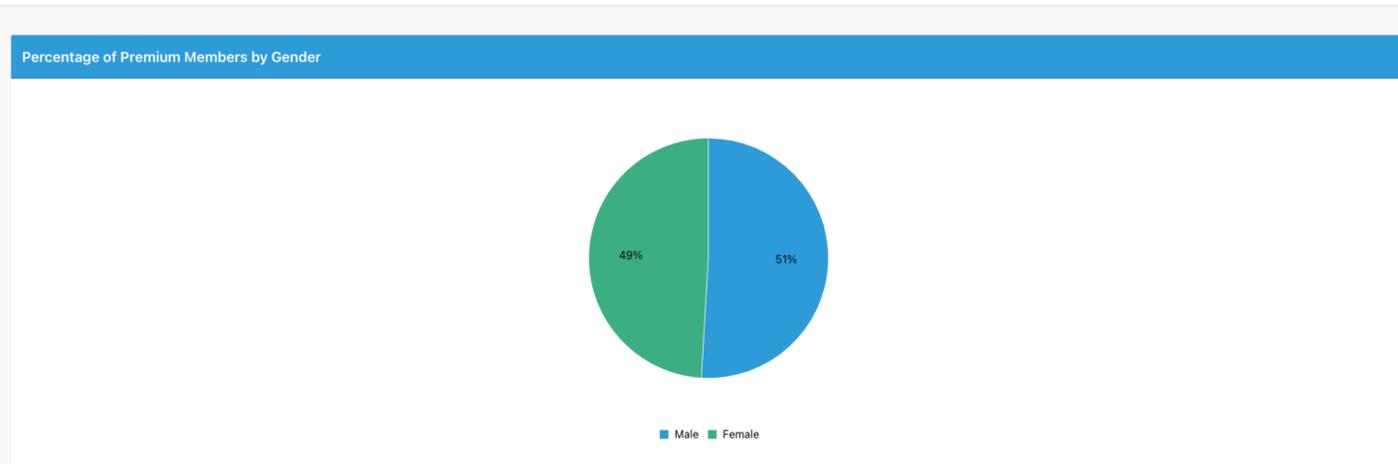


Figure 14: Question 1 Demographics

The second question was also presented as a bar chart to easily compare the percentages of Premium and Standard members using personal trainers. This question was placed on its own page titled "*Personal Trainer Usage by Subscription Type (%)*" (Figure 15). The values displayed on the bar chart represent percentages, making it clear for viewers to interpret the data. Additionally, a text box above the bar chart contains the business question and the answer so that viewers can easily understand the findings.

Question 2 Trainer Usage:

Business Question: What percentage of Premium and Standard members use personal trainers?

Answer: Based on the results of the bar chart, a higher percentage of Premium members use a personal trainer (54%) compared to Standard members (50%). This indicates that Premium members are more likely to take advantage of additional services like personal training.

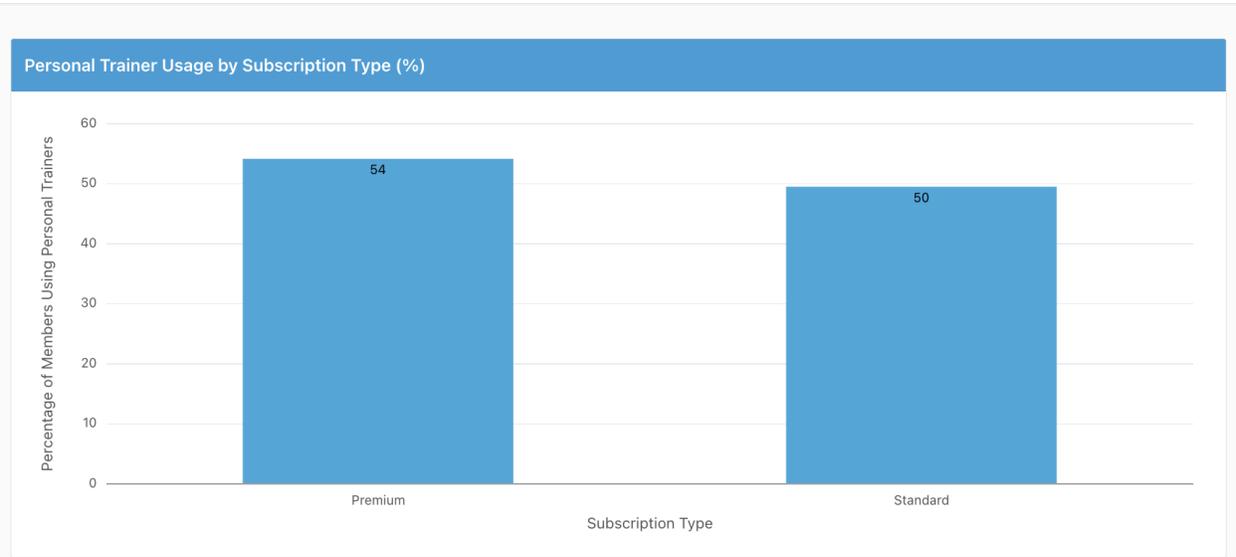


Figure 15: Question 2 Trainer Usage

The results of the third question were also represented as a bar chart to easily compare the top classes for premium members to each other. This question was placed on its own page titled *Q3: Premium Member Lesson Preference (Figure 16)*. The values displayed on the bar chart represent percentages, making it easy to read the data. There is also a text box to the left of the chart that shares the business question and interpretation based on this chart.

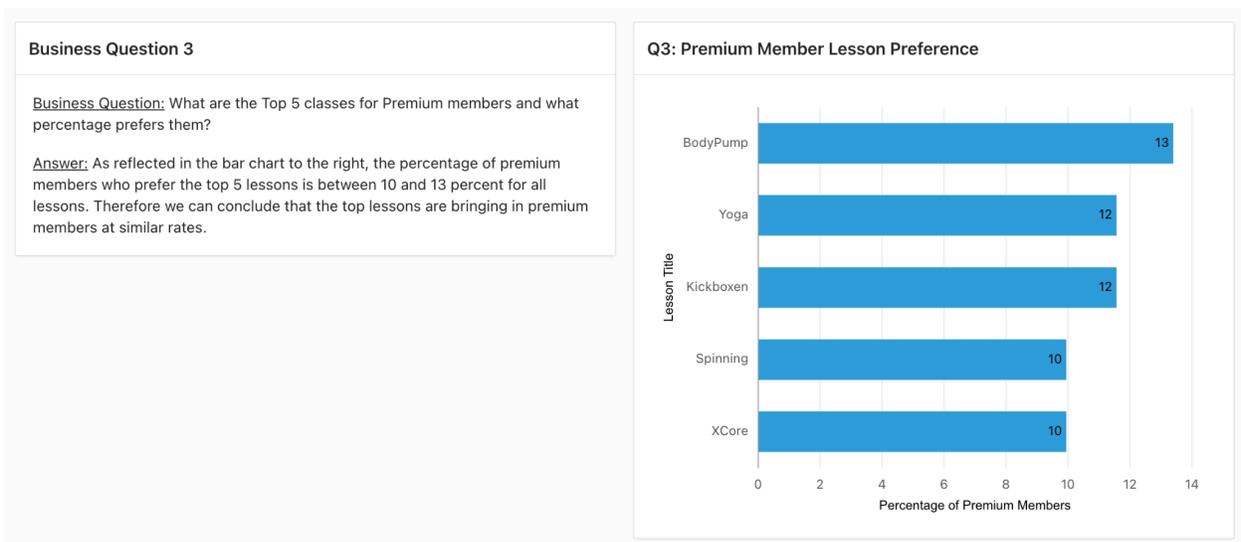


Figure 16: Question 3 Lesson Preference

The fourth question was displayed with a classic report, showing each day of the week and the number of gym attendees on that day (*Figure 17*). The report has its own page, titled Gym Use by Day, and displays the business question, answer, and query results. We found that a table was the easiest to understand the data and see how the gym uses changes through the week.

Gym Use by Day

Business Question: What days of the week are the busiest at the gym?

Answer: The query results show that Monday has the most gym usage with 402 attendees and Thursday has the least with 359 uses. Gym use seems to spike on Monday, then decrease throughout the week, then increase again during the weekend. The gym owners could take this information and create special workout classes or deals to draw people in on Wednesdays and Thursdays when the gym is less busy.

Gym Use by Day

Sunday ↑	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
379	402	394	381	359	379	386

1 - 1

Figure 17: Gym Use by Day

The fifth question was presented as a classic report of data showing the Average Age of each favorite lesson group. This question is placed on its own page titled *Q5: Lessons by Average Age* (*Figure 18*). The classic report displays all the options for Favorite Group Lessons and each average age of them, the average overall for lesson goers are people in their 30's since it only came back as 11 rows of data, making it one of the easiest queries to represent visually as a classic report. Classic Reports makes it easier to visually look at the data for the average age of each lesson. In the text above the report, it includes the business question and answering that question in an analysis.

Business Question: What is the average age for each lesson?

Answer: As you can see in the report the average age range for all the lessons at the gym was between 30 - 32. We can conclude the average age of individuals who attend the gym that take lessons are in their 30's.

Favorite Group Lesson ↑	Average Age
BodyBalance	30
BodyPump	30
HIT	32
Kickboxen	31
LesMiles	30
Pilates	32
Running	30
Spinning	30
XCore	30
Yoga	30
Zumba	32

Figure 18: Question 5 Lessons by Average Age

The sixth question was represented as a donut chart for ease of comparison between membership types. This question was placed on its own page titled *Q6: Membership Type Distribution* (Figure 19). The donut chart displays the values, and shows that there is a near even distribution of standard and premium members. There is also a text box included above the data to show the business question it is answering, as well as the analysis.

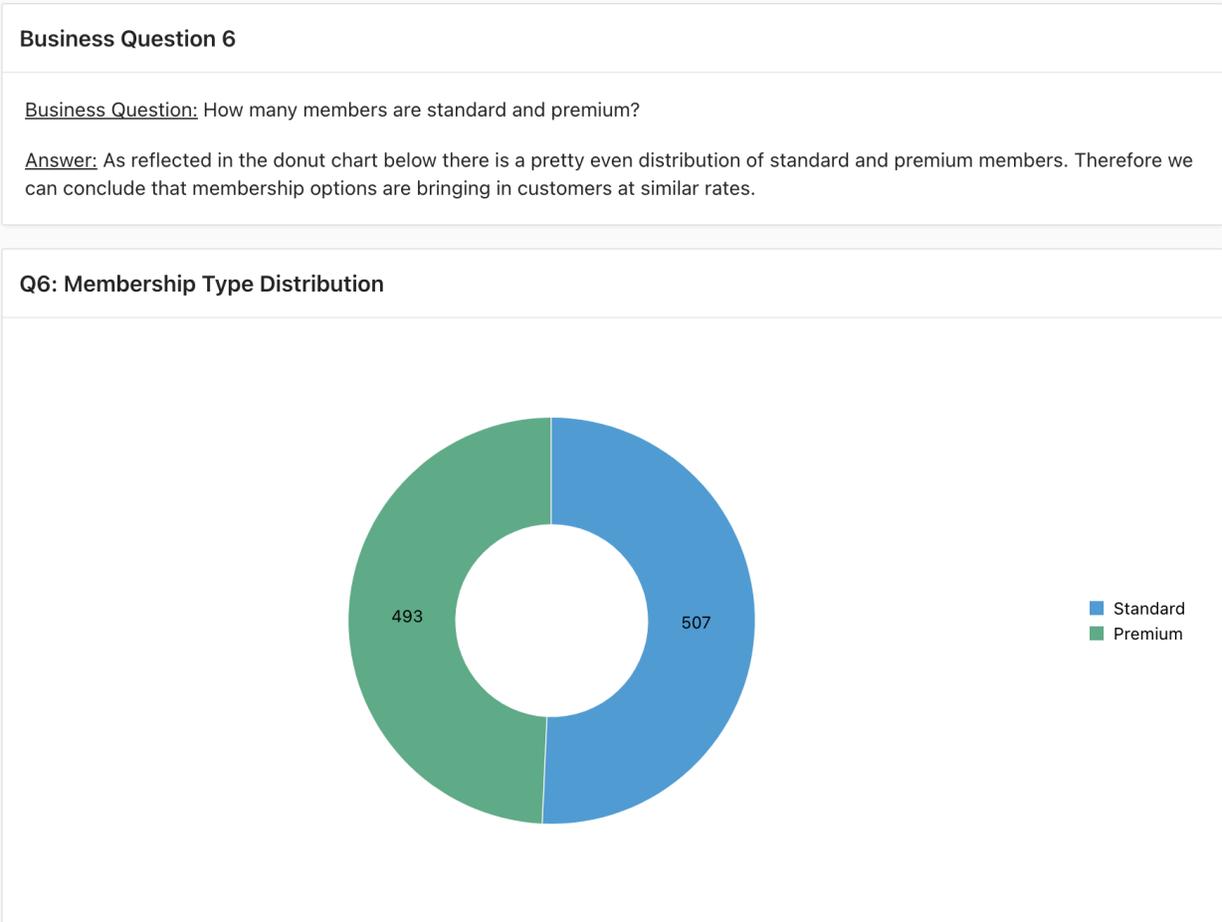


Figure 19: Membership Type Distribution