Class Engagement and Gym Visitor Analysis of GoodLife Fitness

Presented By

Shireesha Thyaranahalli Narayana
Shih-Chieh Ku

Introduction

- To solve the problem related to the absence of gym class, and analyze t visitors to the gym based on the weather, time and date. GoodLife is fitness club chain in Canada. They want to optimize their Gym utilizati and hence maximize client satisfact and profits.

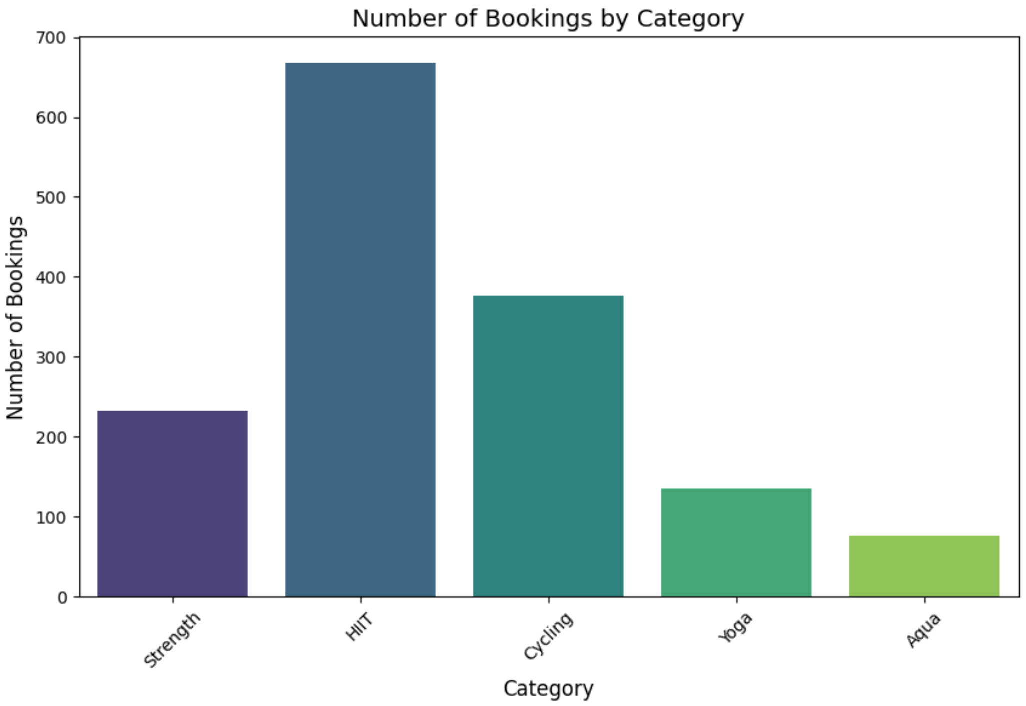
Dataset #1 – Class Attendance Record

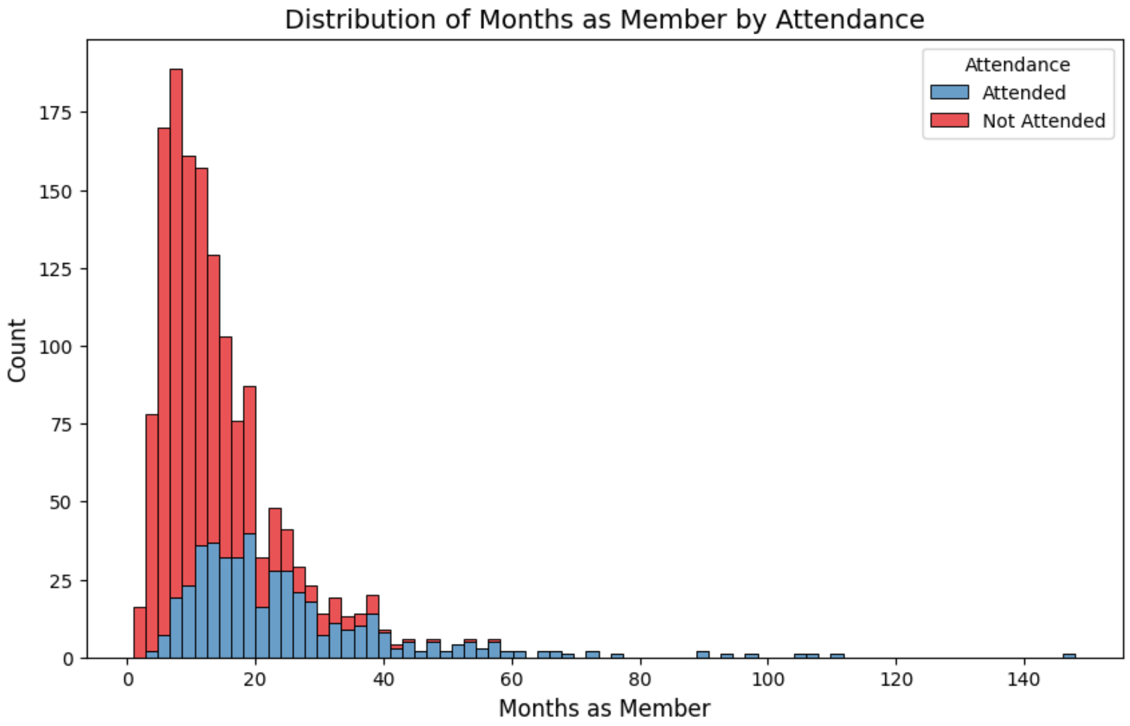
Attribute	Description
booking_id	Nominal. The unique identifier of the booking.
months_as_member	Discrete. The number of months as this fitness club member, minimum 1 month.
weight	Continuous. The member's weight in kg, rounded to 2 decimal places.
days_before	Discrete. The number of days before the class the member registered,
day_of_week	Nominal. The day of the week of the class.
time	Ordinal. The time of day of the class. Either AM or PM
category	Nominal. The category of the fitness class.
attended	Nominal. Whether the member attended the class (1) or not (0)

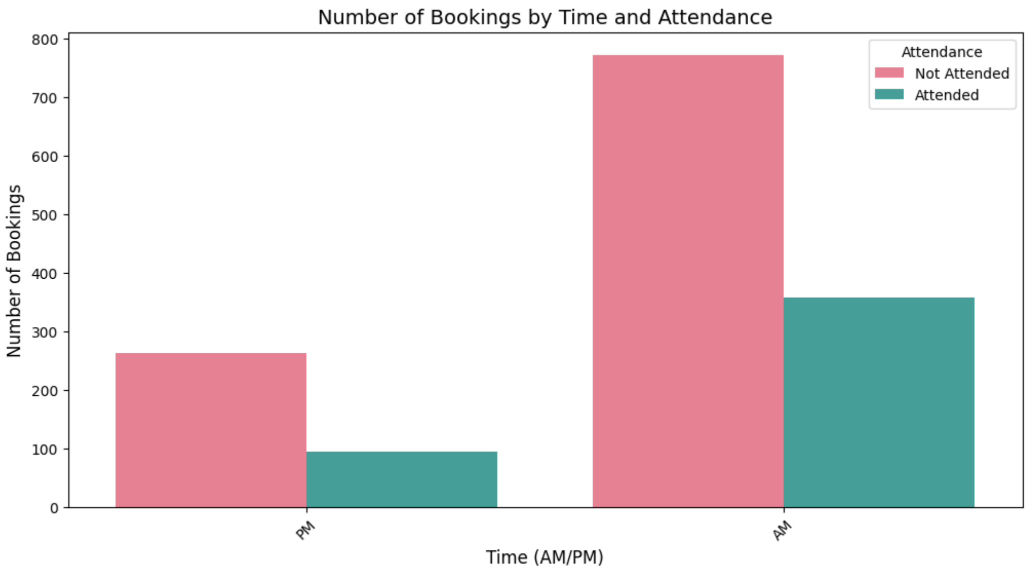
Dataset #2 – Visitors to gym by each 10 minutes

Attribute	Description
date	Datetime of data
timestamp	Number of seconds since beginning of day
day_of_week	0 [monday] - 6 [sunday]
is_weekend	If 1, it's either saturday or sunday, otherwise 0
is_holiday	If 1 it's a federal holiday, 0 otherwise
temperature	degrees fahrenheit
is_start_of_semester	If 1 it's the beginning of a school semester, 0 otherwise
month	1 [jan] - 12 [dec]
hour	0 - 23

Exploring the data – Class Attendance Record



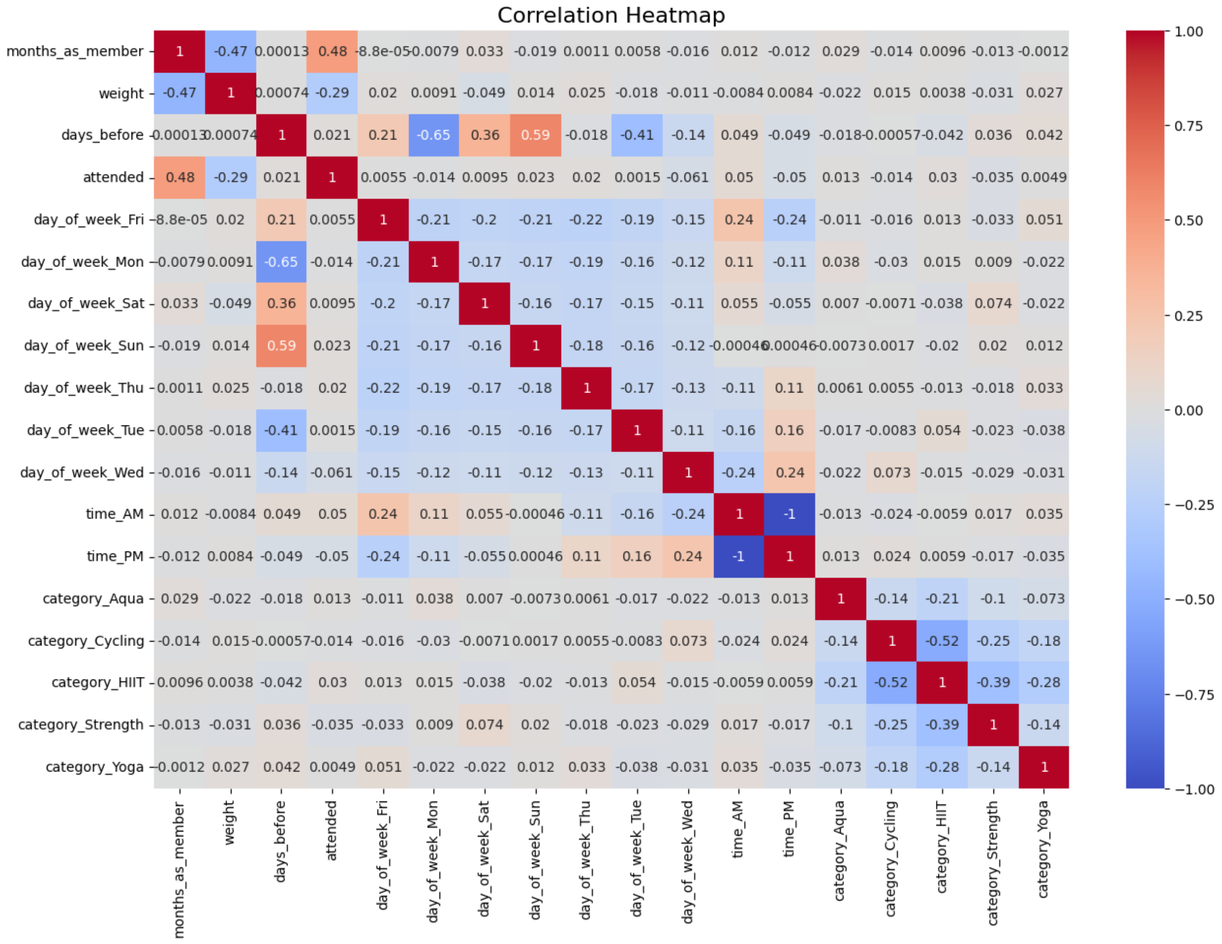




According to the graphic shown, the class is in the morning, it has more absences, it is nearly 2.5 times than the attendance.

Also, the majority of those who did not attend the class usually became members not longer than 20 months.

The HIIT class has the highest number of bookings, it is nearly 6 times number of Aqua class.



Correlation Heatmap

	months_as_member	weight	days_before	attended	day_of_week_Fri	day_of_week_Mon	day_of_week_Sat	day_of_week_Sun	day_of_week_Thu	day_of_week_Tue	day_of_week_Wed	time_AM	time_PM	category_Aqua	category_Cycling	category_HIIT	category_Strength	category_Yoga
months_as_member	1	-0.47	0.00013	0.48	-8.8e-05	-0.0079	0.033	-0.019	0.0011	0.0058	-0.016	0.012	-0.012	0.029	-0.014	0.0096	-0.013	-0.0012
weight	-0.47	1	0.00074	-0.29	0.02	0.0091	-0.049	0.014	0.025	-0.018	-0.011	-0.0084	0.0084	-0.022	0.015	0.0038	-0.031	0.027
days_before	0.00013	0.00074	1	0.021	0.21	-0.65	0.36	0.59	-0.018	-0.41	-0.14	0.049	-0.049	-0.018	-0.00057	-0.042	0.036	0.042
attended	0.48	-0.29	0.021	1	0.0055	-0.014	0.0095	0.023	0.02	0.0015	-0.061	0.05	-0.05	0.013	-0.014	0.03	-0.035	0.0049
day_of_week_Fri	-8.8e-05	0.02	0.21	0.0055	1	-0.21	-0.2	-0.21	-0.22	-0.19	-0.15	0.24	-0.24	-0.011	-0.016	0.013	-0.033	0.051
day_of_week_Mon	-0.0079	0.0091	-0.65	-0.014	-0.21	1	-0.17	-0.17	-0.19	-0.16	-0.12	0.11	-0.11	0.038	-0.03	0.015	0.009	-0.022
day_of_week_Sat	0.033	-0.049	0.36	0.0095	-0.2	-0.17	1	-0.16	-0.17	-0.15	-0.11	0.055	-0.055	0.007	-0.0071	-0.038	0.074	-0.022
day_of_week_Sun	-0.019	0.014	0.59	0.023	-0.21	-0.17	-0.16	1	-0.18	-0.16	-0.12	-0.00046	0.00046	-0.0073	0.0017	-0.02	0.02	0.012
day_of_week_Thu	0.0011	0.025	-0.018	0.02	-0.22	-0.19	-0.17	-0.18	1	-0.17	-0.13	-0.11	0.11	0.0061	0.0055	-0.013	-0.018	0.033
day_of_week_Tue	0.0058	-0.018	-0.41	0.0015	-0.19	-0.16	-0.15	-0.16	-0.17	1	-0.11	-0.16	0.16	-0.017	-0.0083	0.054	-0.023	-0.038
day_of_week_Wed	-0.016	-0.011	-0.14	-0.061	-0.15	-0.12	-0.11	-0.12	-0.13	-0.11	1	-0.24	0.24	-0.022	0.073	-0.015	-0.029	-0.031
time_AM	0.012	-0.0084	0.049	0.05	0.24	0.11	0.055	-0.00046	-0.11	-0.16	-0.24	1	-1	-0.013	-0.024	-0.0059	0.017	0.035
time_PM	-0.012	0.0084	-0.049	-0.05	-0.24	-0.11	-0.055	0.00046	0.11	0.16	0.24	-1	1	0.013	0.024	0.0059	-0.017	-0.035
category_Aqua	0.029	-0.022	-0.018	0.013	-0.011	0.038	0.007	-0.0073	0.0061	-0.017	-0.022	-0.013	0.013	1	-0.14	-0.21	-0.1	-0.073
category_Cycling	-0.014	0.015	-0.00057	-0.014	-0.016	-0.03	-0.0071	0.0017	0.0055	-0.0083	0.073	-0.024	0.024	-0.14	1	-0.52	-0.25	-0.18
category_HIIT	0.0096	0.0038	-0.042	0.03	0.013	0.015	-0.038	-0.02	-0.013	0.054	-0.015	-0.0059	0.0059	-0.21	-0.52	1	-0.39	-0.28
category_Strength	-0.013	-0.031	0.036	-0.035	-0.033	0.009	0.074	0.02	-0.018	-0.023	-0.029	0.017	-0.017	-0.1	-0.25	-0.39	1	-0.14
category_Yoga	-0.0012	0.027	0.042	0.0049	0.051	-0.022	-0.022	0.012	0.033	-0.038	-0.031	0.035	-0.035	-0.073	-0.18	-0.28	-0.14	1

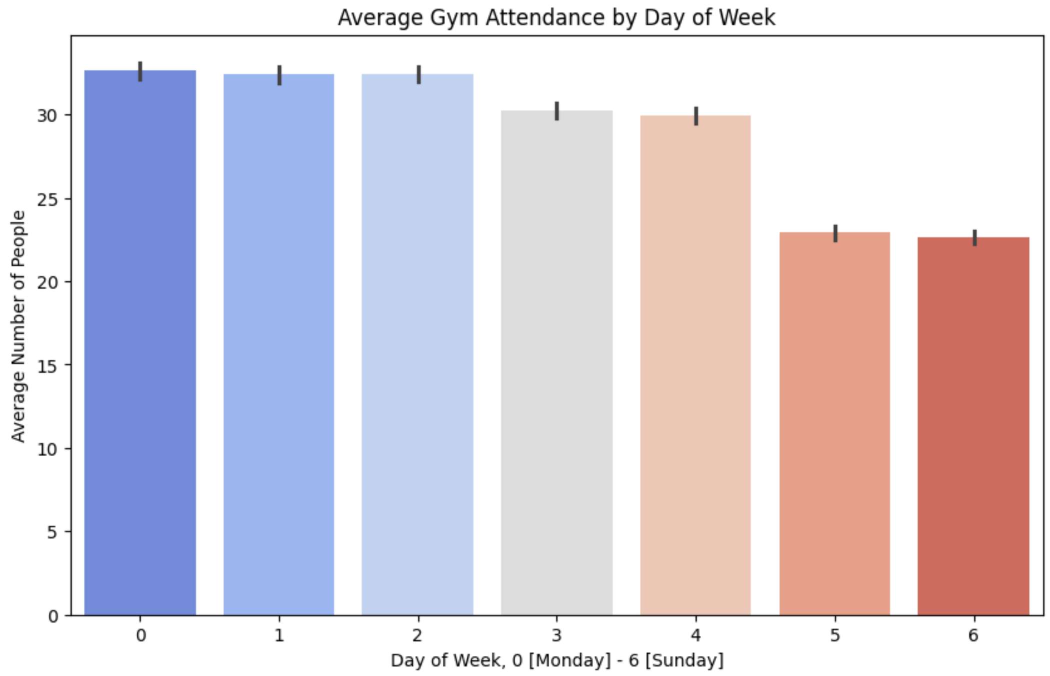
Months as Member:

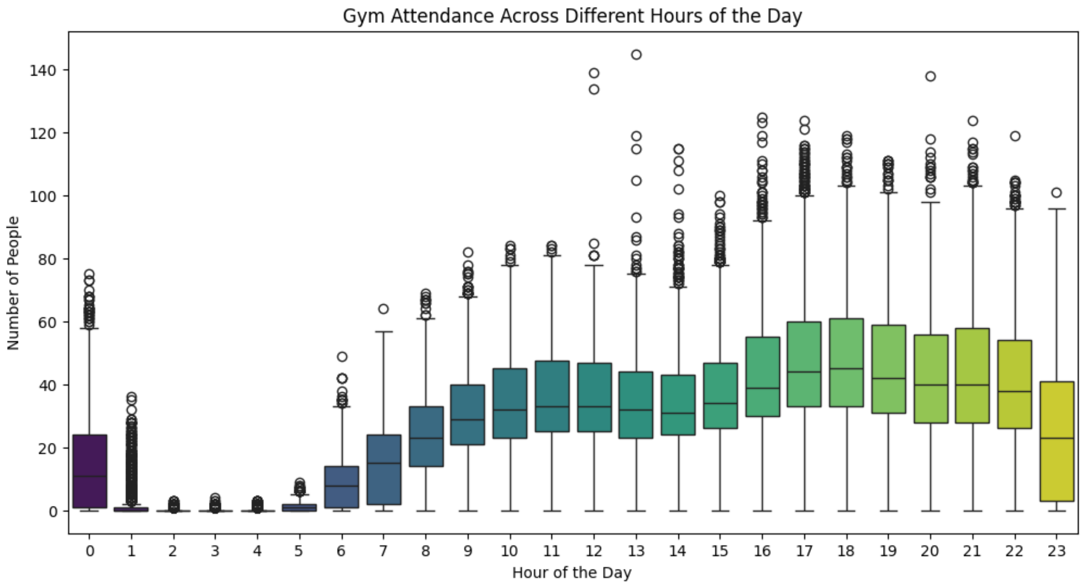
- **Correlation (0.48)**: This indicates that individuals who have been members for a longer period are more likely to attend. This could be due to established habits or a higher commitment level.

Weight:

- Correlation (-0.29): There is a moderate negative correlation between weight and attendance. This could indicate that lighter individuals attend more frequently or that heavier individuals face more barriers to regular attendance.

Exploring the data – Visitors to gym / 10 mins



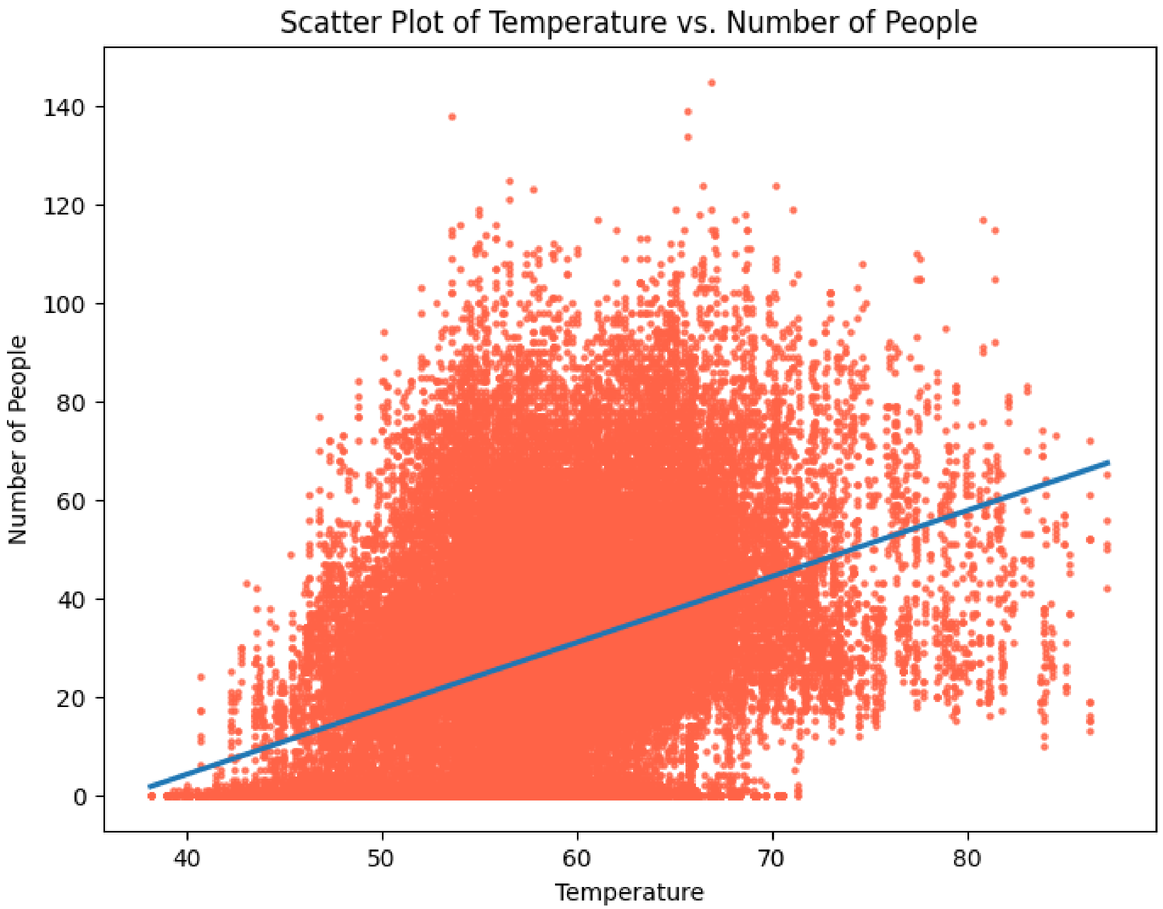


Weekday and weekend:

- Usually, fewer people go to the gym during the weekend. Instead, from Monday to Wednesday, there is exactly the same number (over 30) of people going to the gym, then when the day approaches the weekend, the number of visitors decreases.

Daytime and night:

- For the visitors of distribution during the day, the visitors of daytime are significantly higher than at night. The two turning points are 11 p.m. and 6 a.m., so the number of people going to the gym would be increased at 6 in the morning, increasing until 11the midnight.



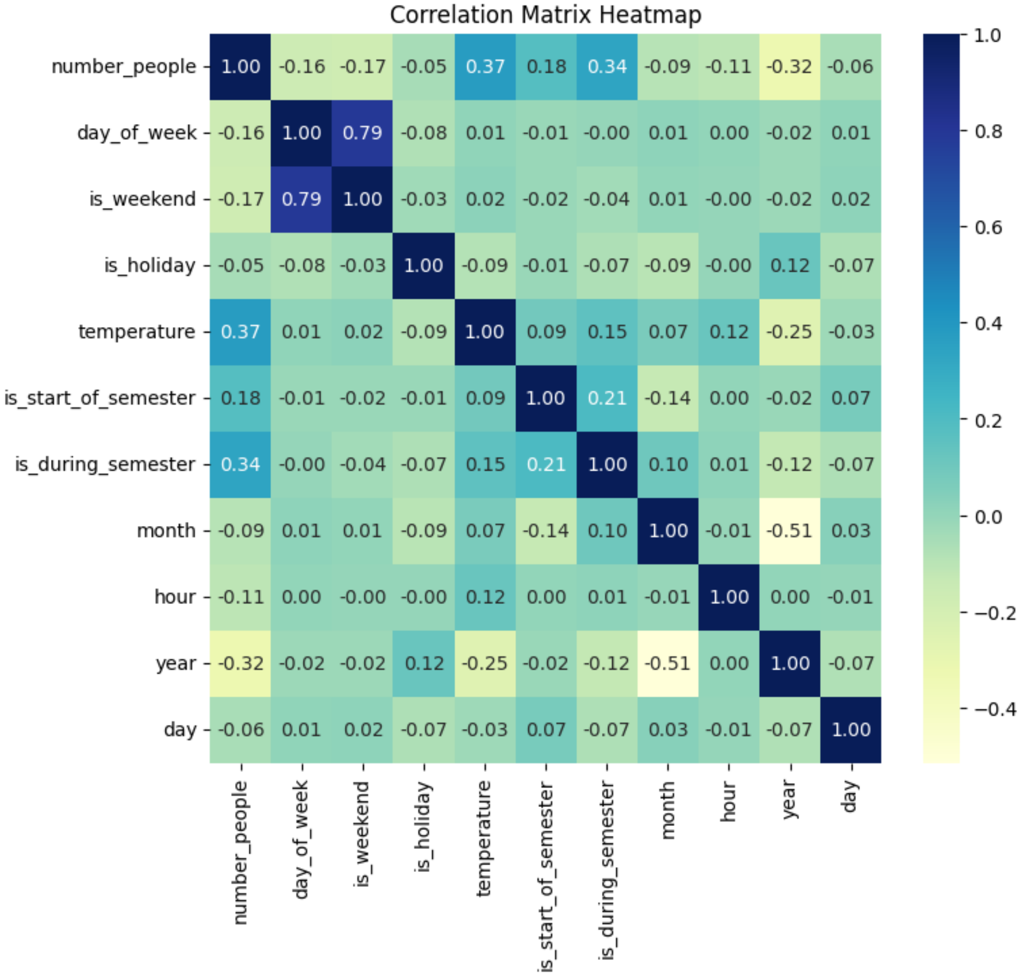
Correlation between Temperature and Number of People:

- When the temperature stays relatively comfortable and a nice degree, and there are more visitors at the gym. This suggests that as the temperature increases, the number of people also tends to increase.

- There is a higher concentration of data points between temperatures of approximately 50˚F to 70˚F and the number of people ranging from 20 to 80.

Correlation:

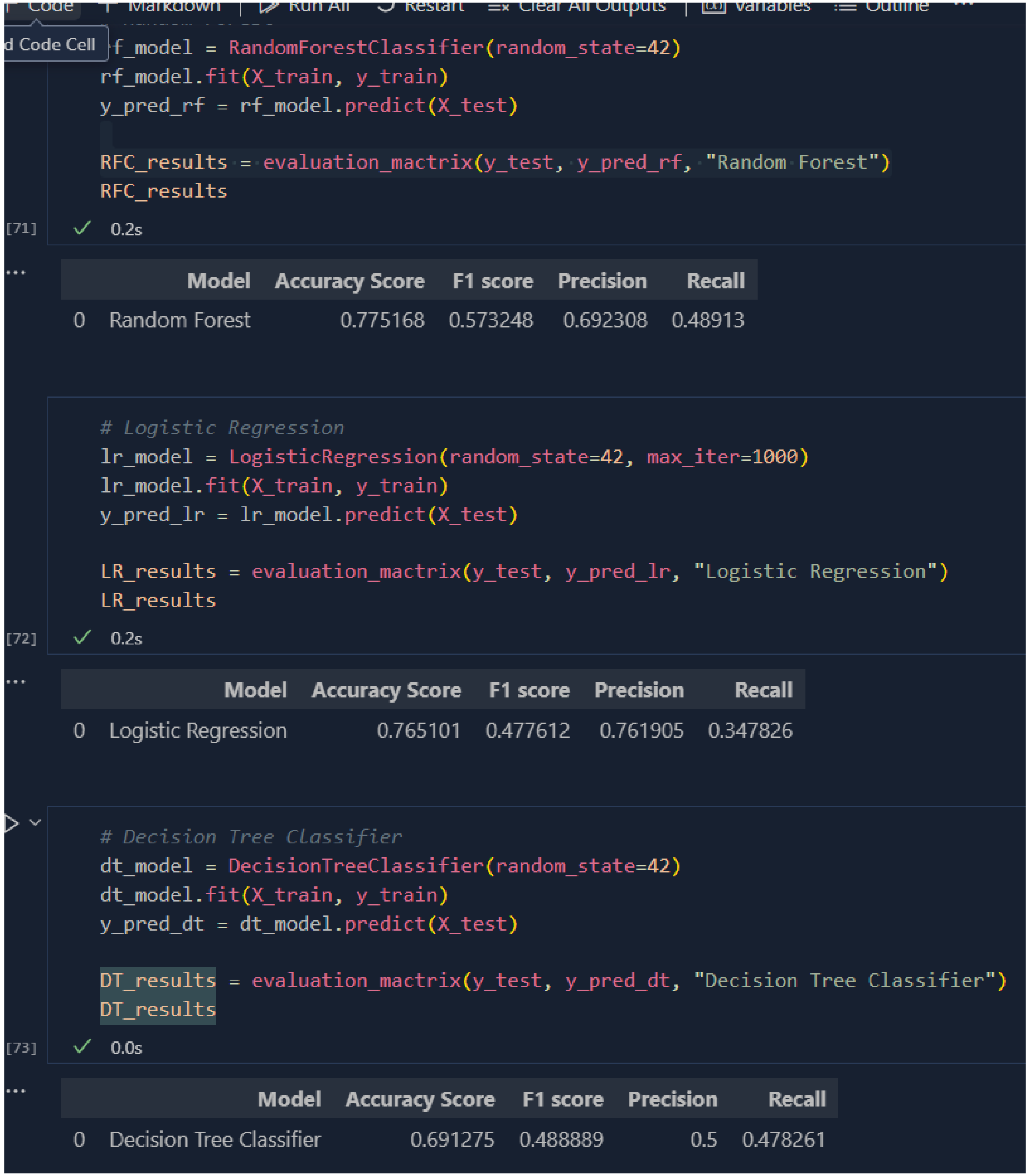
- There are three attributes which have the positive correlation with `number_people`, they are `temperature`, `is_start_of_semester`, and `is_during_semester`. Based on the results, we can assume: this gym might be located near the college or university or that the majority of their customers are students.



	number_people	day_of_week	is_weekend	is_holiday	temperature	is_start_of_semester	is_during_semester	month	hour	year	day
number_people	1.00	-0.16	-0.17	-0.05	0.37	0.18	0.34	-0.09	-0.11	-0.32	-0.06
day_of_week	-0.16	1.00	0.79	-0.08	0.01	-0.01	-0.00	0.01	0.00	-0.02	0.01
is_weekend	-0.17	0.79	1.00	-0.03	0.02	-0.02	-0.04	0.01	-0.00	-0.02	0.02
is_holiday	-0.05	-0.08	-0.03	1.00	-0.09	-0.01	-0.07	-0.09	-0.00	0.12	-0.07
temperature	0.37	0.01	0.02	-0.09	1.00	0.09	0.15	0.07	0.12	-0.25	-0.03
is_start_of_semester	0.18	-0.01	-0.02	-0.01	0.09	1.00	0.21	-0.14	0.00	-0.02	0.07
is_during_semester	0.34	-0.00	-0.04	-0.07	0.15	0.21	1.00	0.10	0.01	-0.12	-0.07
month	-0.09	0.01	0.01	-0.09	0.07	-0.14	0.10	1.00	-0.01	-0.51	0.03
hour	-0.11	0.00	-0.00	-0.00	0.12	0.00	0.01	-0.01	1.00	0.00	-0.01
year	-0.32	-0.02	-0.02	0.12	-0.25	-0.02	-0.12	-0.51	0.00	1.00	-0.07
day	-0.06	0.01	0.02	-0.07	-0.03	0.07	-0.07	0.03	-0.01	-0.07	1.00

Model training –

Class Attendance Record



```
f_model = RandomForestClassifier(random_state=42)
rf_model.fit(X_train, y_train)
y_pred_rf = rf_model.predict(X_test)

RFC_results = evaluation_mactrix(y_test, y_pred_rf, "Random Forest")
RFC_results
```

```
[71]  ✓ 0.2s

              Model  Accuracy Score  F1 score  Precision   Recall
0     Random Forest        0.775168  0.573248   0.692308  0.48913
```

```
# Logistic Regression
lr_model = LogisticRegression(random_state=42, max_iter=1000)
lr_model.fit(X_train, y_train)
y_pred_lr = lr_model.predict(X_test)

LR_results = evaluation_mactrix(y_test, y_pred_lr, "Logistic Regression")
LR_results
```

```
[72]  ✓ 0.2s

                    Model  Accuracy Score  F1 score  Precision    Recall
0     Logistic Regression        0.765101  0.477612   0.761905  0.347826
```

```
# Decision Tree Classifier
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)
y_pred_dt = dt_model.predict(X_test)

DT_results = evaluation_mactrix(y_test, y_pred_dt, "Decision Tree Classifier")
DT_results
```

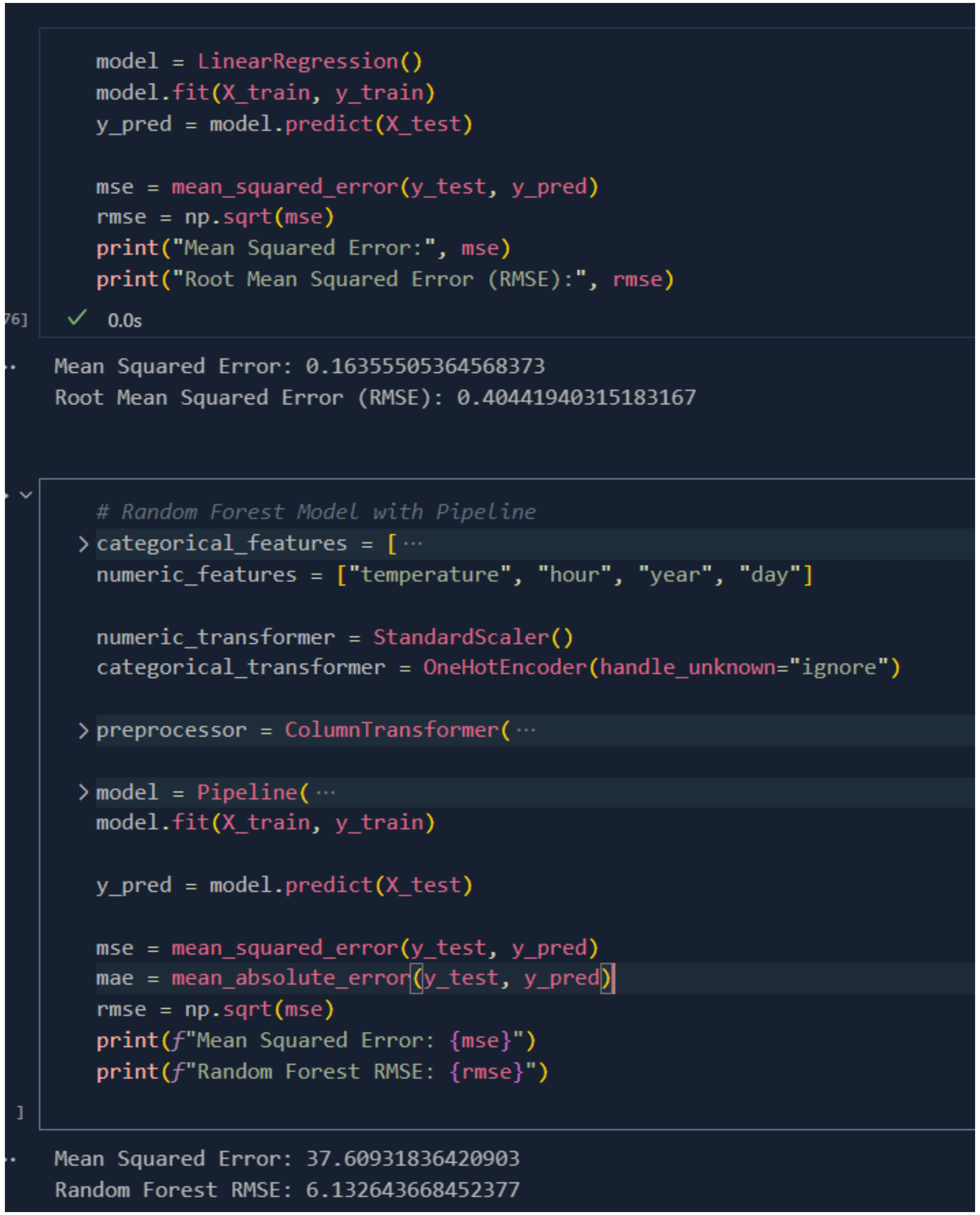
```
[73]  ✓ 0.0s

                         Model  Accuracy Score  F1 score  Precision    Recall
0     Decision Tree Classifier        0.691275  0.488889        0.5  0.478261
```

Model Selection:

- **Random Forest** appears to be the best overall performer with the highest accuracy and a good balance between precision and recall. It is recommended for scenarios where a higher overall accuracy is preferred.
- **Logistic Regression** might be suitable in cases where minimizing false positives is critical due to its high precision, despite its lower recall.
- **Decision Tree** can be considered for simpler models or when interpretability is crucial, but it may not be the best choice based on performance metrics.

Model training –

Visitors to gym / 10 mins



```
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
print("Mean Squared Error:", mse)
print("Root Mean Squared Error (RMSE):", rmse)
```

76] ✓ 0.0s

```
Mean Squared Error: 0.16355505364568373
Root Mean Squared Error (RMSE): 0.40441940315183167
```

```
# Random Forest Model with Pipeline
categorical_features = [ ···
numeric_features = ["temperature", "hour", "year", "day"]

numeric_transformer = StandardScaler()
categorical_transformer = OneHotEncoder(handle_unknown="ignore")

preprocessor = ColumnTransformer( ···

model = Pipeline( ···
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mse)
print(f"Mean Squared Error: {mse}")
print(f"Random Forest RMSE: {rmse}")
```

```
Mean Squared Error: 37.60931836420903
Random Forest RMSE: 6.132643668452377
```

Overall Conclusion

1. **Performance**: Linear Regression significantly outperforms the Random Forest Regressor in terms of both MSE and RMSE. The low error values for Linear Regression indicate it is more suited for this specific task of this dataset.

2. **Model Suitability**:
   1. **Linear Regression**: This model is appropriate for this dataset, as it provides accurate predictions with minimal error. It suggests that the relationship between the features and the target variable is linear, which Linear Regression captures effectively.
   2. **Random Forest Regressor**: This model is not suitable for this dataset, as it fails to provide accurate predictions. The high error metrics suggest that the model might be overfitting or not capturing the underlying patterns in the data correctly.

Business Decision and Marketing Strategies

- Class and equipment arrangement: Adjust class schedules based on predicted attendance to maximize space utilization. Simultaneously, gym equipment should be allocated dynamically based on real-time crowd predictions to ensure availability.
- Cost management: Schedule staff shifts based on predicted peak times to maintain service quality. Also, adjust the opening hours based on the real-time prediction of coming people. Based on the previous modification, re-arrange the number of staff or coaches, hire more part-time employees, but maintain highly qualified service and control the operational cost.
- Promotional Offers: Design and implement targeted promotions to encourage gym visits and class attendance, especially for off-peak hours.
- Member Communication: Develop communication strategies to inform members about available slots or remind them of the class approaching, and implement the punishment for absence.

expenditure

Improvements after Implementing Business & Marketing Strategies

- By adjusting the class schedule and allocation of equipment, we can reduce waste. Instead, it produces more efficiency and loyalty.
- By re-arranging staff allocation and opening hours, we can control and reduce the cost, avoiding the waste of human resources. Additionally, closing the gym at off-peak hours would reduce the hydro bill, operational expenditure, and salary of employees.
- By offering tailor promotions, it can balance the number of gym visits either on weekdays or weekends, including every period of a day to create more revenue.
- By developing communication channels with customers, the class vacancy would be filled, and reaching a better attendance rate.

Thank you