INTRODUCTION

Club Membership Trend using Time Series Analysis and Regression Model

Objective & Motivation

- Explore student participation in clubs and events over time.
- Use time series analysis and regression techniques for trend analysis.
- Predict future membership trends.
- Provide insights for better resource allocation, event planning, and club management.
- Understand how student engagement in extracurricular activities impacts their overall experience.
- Optimize club activities to improve student involvement.

Attributes of Datasets & Its types

The dataset provided includes information about club membership trends over multiple semesters. Key attributes are:

- **Semester** – Represents the time period (Semester 1, 2, 3, and 4).
- **Club Name/ID** – Identifies the specific club (e.g., Club 1, Club 2, etc.).
- **Past Members** – Number of continuing members from the previous semester.
- **New Members** – Number of new participants who joined during that semester.
- **Total Members** – Sum of past and new members in the current semester.
- **Event Attendance** – Number of participants attending events hosted by the club.

Handing Categorical data

Why Handle It?

Machine learning models need **numerical inputs**, so **categorical data** (e.g., clubs, semesters) must be encoded.

Methods:

- **One-Hot Encoding**: Converts categories to binary columns.

```
import pandas as pd
df = pd.DataFrame({'Club': ['Club1', 'Club2', 'Club3'], 'Semester': [1, 2, 3]})
df_encoded = pd.get_dummies(df, columns=['Club'])
print(df_encoded)
```

Ordinal Encoding: Maps ordered categories to integers.

```
from sklearn.preprocessing import OrdinalEncoder

encoder = OrdinalEncoder()

df['Semester'] = encoder.fit_transform(df[['Semester']])

print(df)
```

Relationship Between Attributes

1. Past Members and New Members

- High past membership often correlates with increased new memberships due to positive club reputation and recurring participation.

2. Total Members and Event Attendance

- Clubs with larger memberships often see increased participation, suggesting a stronger sense of community and interest in events.

3. New Members and Event Attendance

- A steady increase in new members can predict higher event engagement over subsequent semesters.

4. Semester Progression and Membership Growth

- Membership growth trends over semesters can be analyzed using time series models to identify seasonality and patterns.

Dataset

Semester	Club	Past Members	New Members	Total Member	Event Attendance
Semester 1	Club 1	0	37	37	17
Semester 1	Club 2	0	38	38	18
Semester 1	Club 3	0	30	30	10
Semester 1	Club 4	0	49	49	19
Semester 1	Club 5	0	42	42	32
Semester 1	Club 6	0	31	31	21
Semester 1	Club 7	0	30	30	13
Semester 1	Club 8	0	40	40	32
Semester 1	Club 9	0	39	39	34
Semester 1	Club 10	0	47	47	35

Semester	Club	Past Members	New Members	Total Members	Event Attendance
Semester 2	Club 1	37	32	69	39
Semester 2	Club 2	38	37	75	38
Semester 2	Club 3	30	49	79	12
Semester 2	Club 4	49	49	98	97
Semester 2	Club 5	42	38	80	26
Semester 2	Club 6	31	33	64	51
Semester 2	Club 7	30	42	72	59
Semester 2	Club 8	40	49	89	77
Semester 2	Club 9	39	38	77	30
Semester 2	Club 10	47	33	80	17

Semester	Club	Past Members	New Members	Total Member	Event Attendance
Semester 3	Club 1	69	41	110	101
Semester 3	Club 2	75	38	113	85
Semester 3	Club 3	79	33	112	19
Semester 3	Club 4	98	37	135	74
Semester 3	Club 5	80	38	118	65
Semester 3	Club 6	64	44	108	67
Semester 3	Club 7	72	48	120	81
Semester 3	Club 8	89	43	132	27
Semester 3	Club 9	77	40	117	99
Semester 3	Club 10	80	48	128	89

Semester	Club	Past Members	New Members	Total Members	Event Attendance
Semester 4	Club 1	110	32	142	58
Semester 4	Club 2	113	45	158	122
Semester 4	Club 3	112	37	149	55
Semester 4	Club 4	135	42	177	49
Semester 4	Club 5	118	48	166	134
Semester 4	Club 6	108	48	156	155
Semester 4	Club 7	120	39	159	98
Semester 4	Club 8	132	44	176	123
Semester 4	Club 9	117	31	148	123
Semester 4	Club 10	128	47	175	127

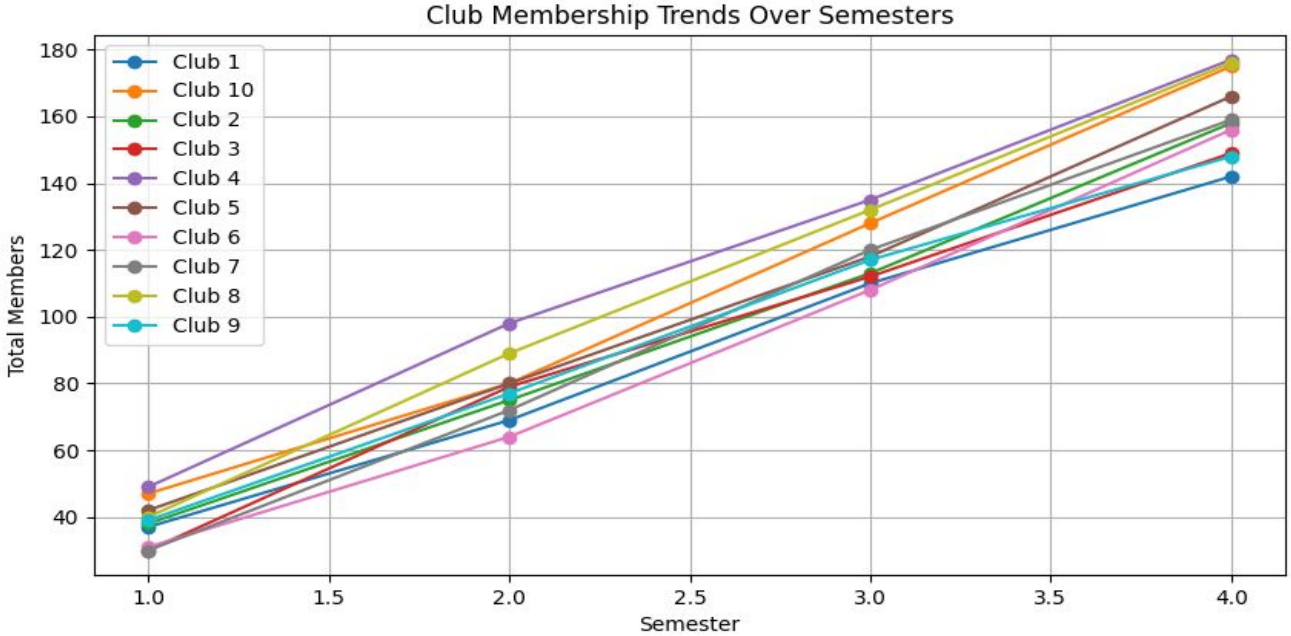
Model Required

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.linear_model import LinearRegression

# Load dataset
data_path = "club_membership.csv"
df = pd.read_csv(data_path)

#The dataset has columns: 'Semester', 'Club', 'Total Members', 'Event Attendance'
# Pivot data for visualization
membership_trends = df.pivot(index='Semester', columns='Club', values='Total Members')
event_trends = df.pivot(index='Semester', columns='Club', values='Event Attendance')

#Club Membership Trends Over Semesters
plt.figure(figsize=(10, 5))
for club in membership_trends.columns:
    plt.plot(membership_trends.index, membership_trends[club], marker='o', label=club)
plt.title("Club Membership Trends Over Semesters")
plt.xlabel("Semester")
plt.ylabel("Total Members")
plt.legend()
plt.grid()
plt.show()
```

```
#Event Attendance Trends Over Semesters
plt.figure(figsize=(10, 5))
for club in event_trends.columns:
    plt.plot(event_trends.index, event_trends[club], linestyle='dashed', marker='s', label=club)
plt.title("Event Attendance Trends Over Semesters")
plt.xlabel("Semester")
plt.ylabel("Event Attendance")
plt.legend()
plt.grid()
plt.show()

#Membership Trends with Future Predictions
future_semesters = np.array([5, 6]).reshape(-1, 1)
predictions = {}
plt.figure(figsize=(12, 6))
for club in membership_trends.columns:
    X = np.array(membership_trends.index).reshape(-1, 1)
    y = membership_trends[club].values
    model = LinearRegression()
    model.fit(X, y)
```

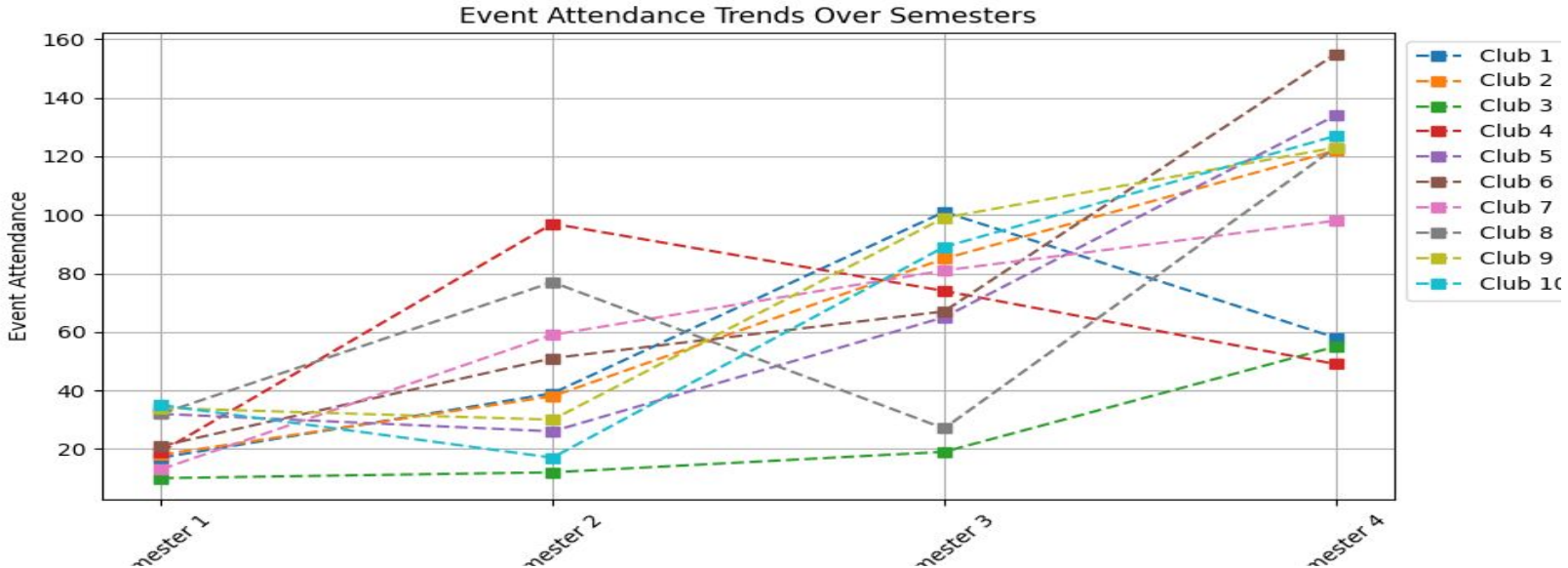
```
pred = model.predict(future_semesters)
    predictions[club] = pred
    plt.plot(X, y, marker='o', label=club)
    plt.plot(future_semesters, pred, linestyle='dashed')
plt.title("Club Membership Trends and Predictions")
plt.xlabel("Semester")
plt.ylabel("Membership Count")
plt.legend()
plt.grid()
plt.show()
#Actual vs Predicted Attendance Scatter Plot
X_attendance = event_trends.index.values.reshape(-1, 1)
y_attendance = event_trends.mean(axis=1).values  # Averaging attendance across clubs
model_attendance = LinearRegression()
model_attendance.fit(X_attendance, y_attendance)
predicted_attendance = model_attendance.predict(X_attendance)
plt.figure(figsize=(8, 6))
plt.scatter(y_attendance, predicted_attendance, alpha=0.7)
plt.xlabel("Actual Attendance")
plt.ylabel("Predicted Attendance")
plt.title("Regression Model: Actual vs Predicted Attendance")
plt.grid()
plt.show()
```

Interpretation of Model



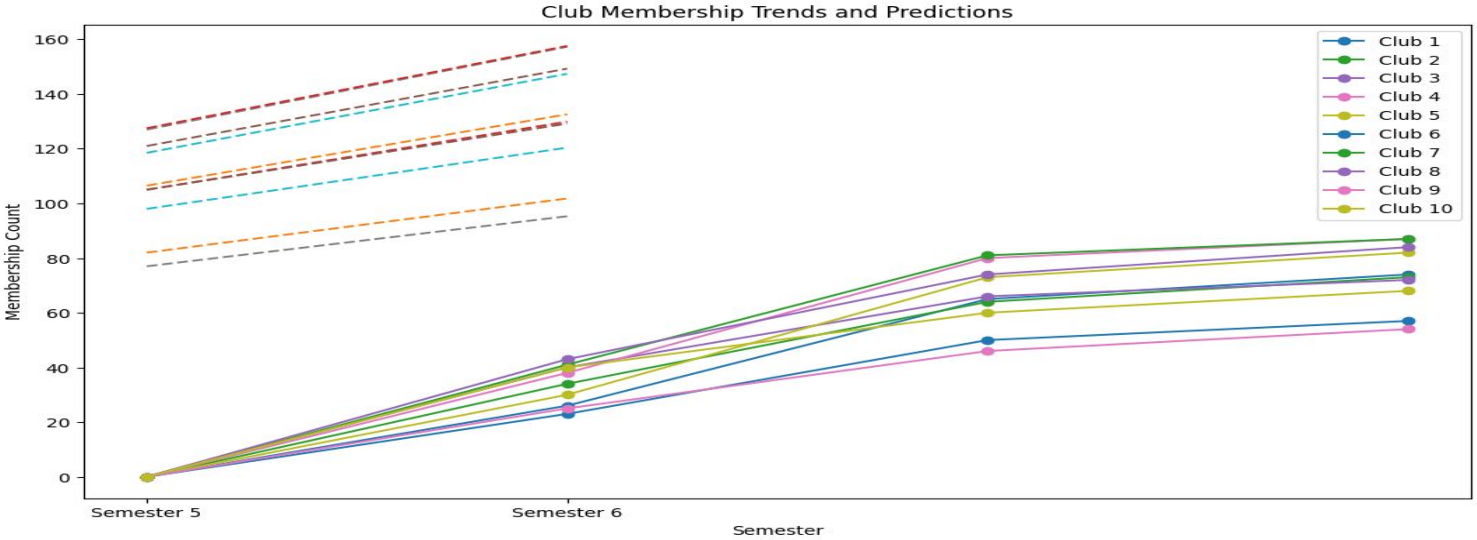
The line graph represents the **club membership trends over different semesters**. Each line corresponds to a different club, showing the **total number of members** across four semesters. The general upward trend in all clubs indicates **a consistent increase in membership over time**, suggesting growing student participation. Some clubs, such as **Club 1 and Club 6**, exhibit a **steeper growth rate**, implying higher recruitment or retention rates. The trend also suggests that no clubs experienced a decline in membership, which could indicate **effective engagement strategies** or a growing interest in club activities among students.

Interpretation of Model



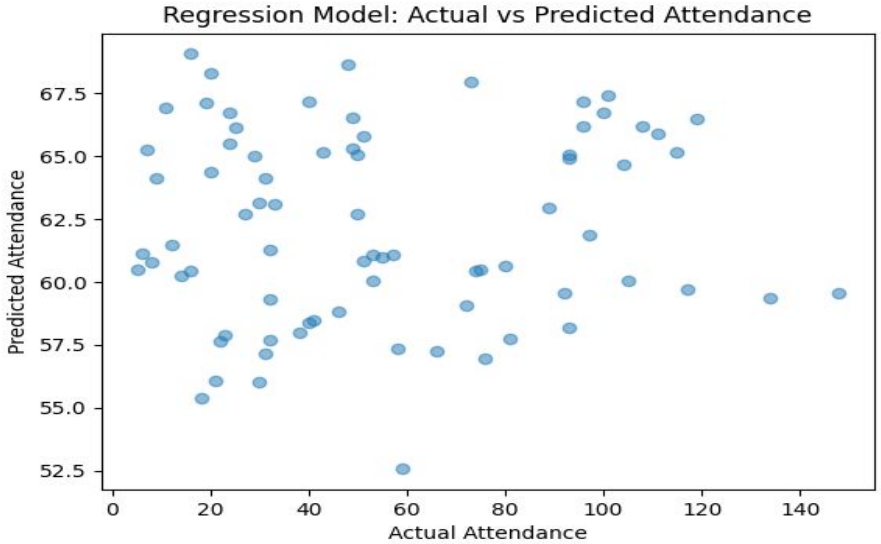
The line graph illustrates **event attendance trends over different semesters** for various clubs. Unlike the previous membership graph, this one exhibits more **fluctuations** in attendance patterns. While most clubs show a general increase, some clubs, such as **Club 4 and Club 5**, demonstrate significant **spikes and drops** in attendance across semesters, indicating varying levels of engagement. **Club 6 and Club 9** show a **sharp rise** in attendance by Semester 4, suggesting increased popularity or successful event planning. However, certain clubs, like **Club 3**, maintain relatively lower attendance, which may indicate **less effective outreach** or fewer events being conducted.

Interpretation of Model



The graph illustrates **club membership trends along with predictions** for future semesters. The **solid lines** represent actual data, while the **dashed lines** indicate projected membership counts. The trend shows a **steady increase** in club memberships over time, suggesting growing student engagement. The predicted values for upcoming semesters indicate that this trend is expected to continue, with some clubs experiencing **faster growth** than others. The **upper set of dashed lines** suggests that clubs with historically higher memberships are likely to maintain their dominance, while clubs with lower initial memberships may experience **gradual but consistent growth**.

Interpretation of Model



The scatter plot compares **actual event attendance** with the **predicted attendance** based on a regression model. Ideally, if the model had perfect accuracy, the points would align along a **45-degree diagonal line** (y = x). However, the spread of points indicates some **deviation** between actual and predicted values. While some predictions are relatively close, others show noticeable discrepancies, suggesting that the model might require **further tuning** to improve accuracy. Possible refinements could include incorporating additional variables or adjusting the regression parameters to minimize prediction errors.

Applicability

➳ This model can predict future membership trends which will help deciding club budget and expenditure.

➳ If certain events show consistently lower attendance than predicted, clubs can adjust marketing strategies or event formats to increase engagement.

➳ University administration and club leaders can use insights from the model to allocate funding, promote high-impact events, and improve overall student engagement.

➳ The model can help forecast club event attendance for future semesters, allowing organizers to plan resources, venues, and logistics more effectively.

Thank You

- Abhinav Anand (BTECH/10164/23)
- Aman Agrawal (BTECH/10166/23)
- Randhir Raj (BTECH/10216/23)
- Shivam Sharma (BTECH/10230/23)
- Mayank Raj (BTECH/10640/23)