

# Data-analytics Hackathon

**Name:** Anumula Madhava  
**College:** IIITDM Kancheepuram  
**Roll.No:** CS23B1008  
**Year:** 2<sup>nd</sup> year  
**Branch:** Computer Science and endgineering  
**Mail:** cs23b1008@iiitdm.ac.in

# Data-analytics Hackathon

## Challenge Brief:

The mess hall crowd is highly irregular — some weeks it's overcrowded, others it's underused. This happens despite fixed meal timings and a standard menu cycle. The reasons behind these fluctuations aren't well understood. You are provided with a data set that simulates real-world mess usage over time. It includes weekly crowd counts and possible influencing factors like holidays, temperature, menu scores, event intensity, and academic stress. Your goal is to find the patterns behind this irregular crowding, identify what factors impact it, and design a data-driven approach to predict or manage crowd levels better.

## Data Understanding and EDA(Exploratory Data Analysis):

- Data Understanding:

The data set contains weekly mess hall crowd counts across various messes, along with influencing factors. Each row represents one week in one mess hall.

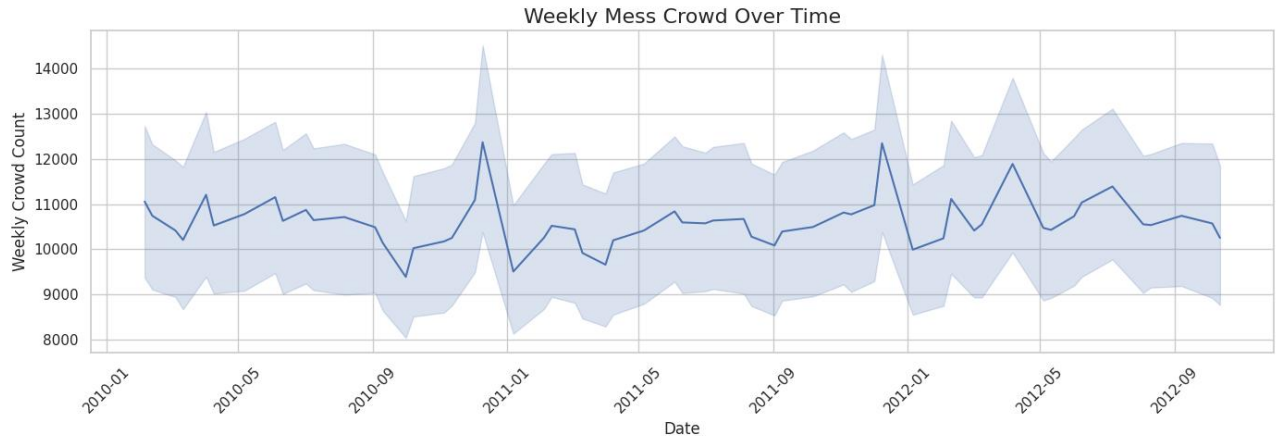
**Columns: (7 features + 1 target(weekly crowd))**

Column	Description
Date	Week start date
Weekly_Crowd	Number of people who ate in the mess that week
Mess_ID	Unique ID of the mess hall
Is_Holiday	1 if it was a holiday week, 0 otherwise
Menu_Score	Rating (0–10) for menu quality that week
Temperature	Average temperature during the week
Event_Intensity_Index	Scale of events held on campus that week
Stress_Level	Academic stress level (exam periods etc.)

# Data-analytics Hackathon

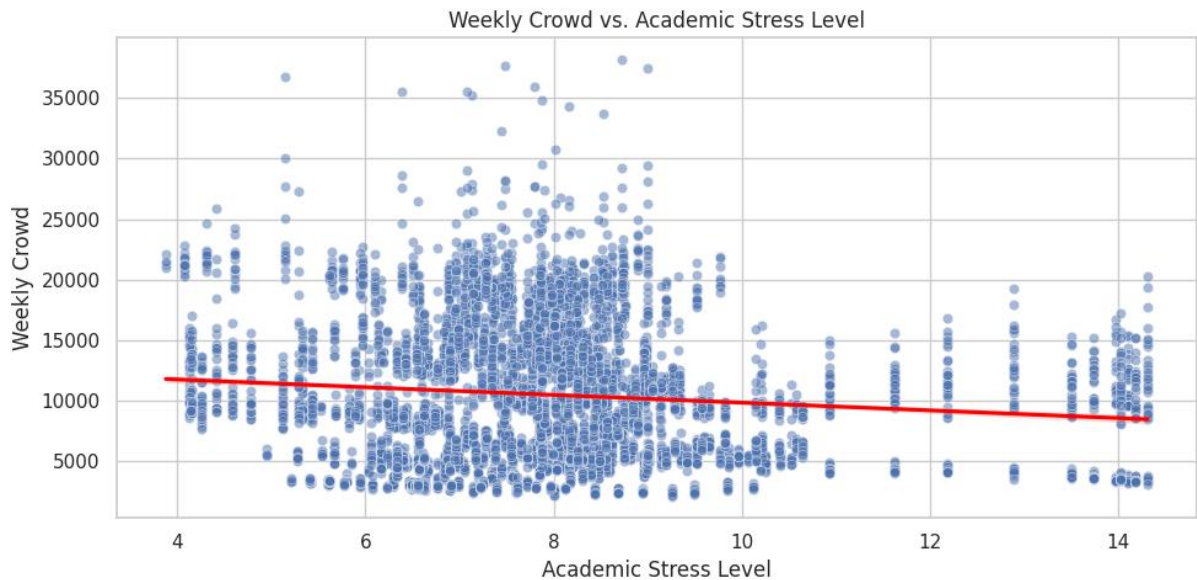
- Exploratory Data Analysis:

## 1. Crowd Over Time:



**Insight:** Crowd fluctuates significantly — some weeks show high peaks, others low dips.

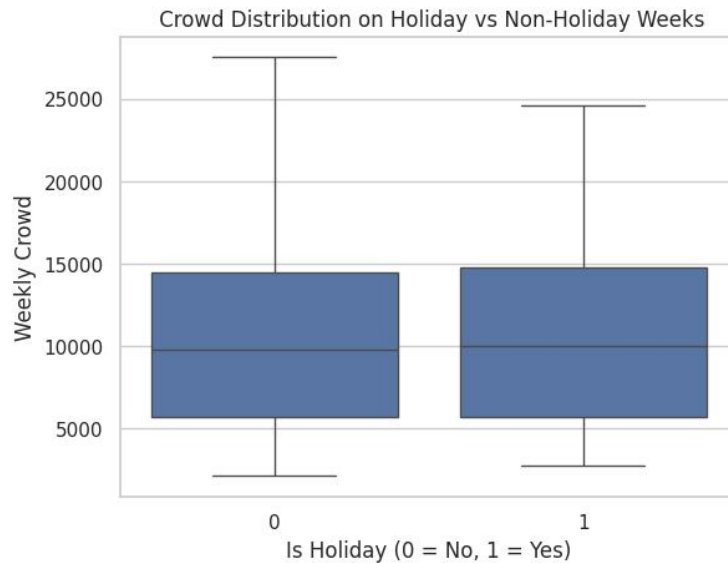
## 2. Distribution of Weekly Crowd:



**Insight:** That red line (regression line) shows a downward trend, supporting our correlation findings.(point number 5)

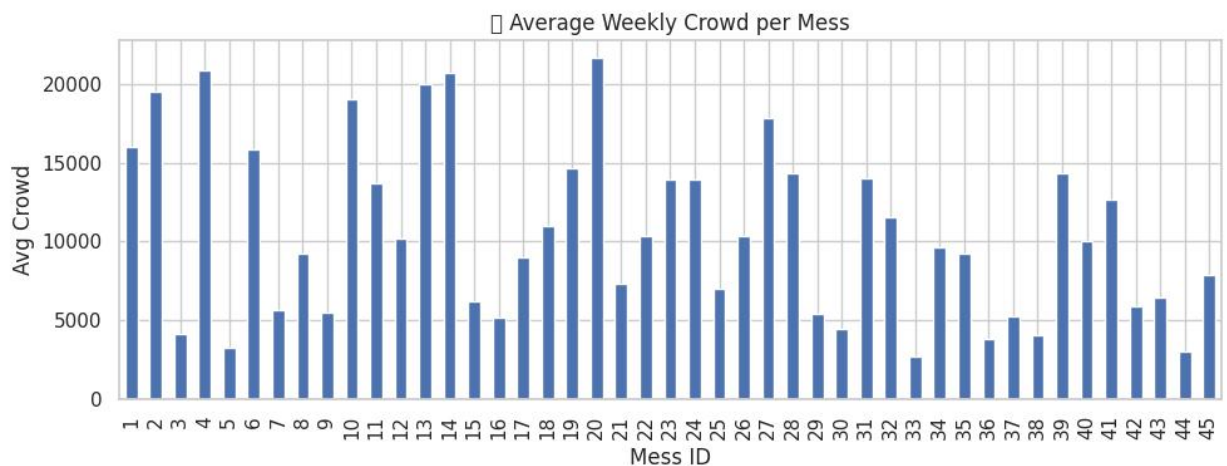
# Data-analytics Hackathon

## 3. Boxplot: Crowd vs. Holiday:



**Insight:** Holidays do **not** significantly increase or decrease crowd.

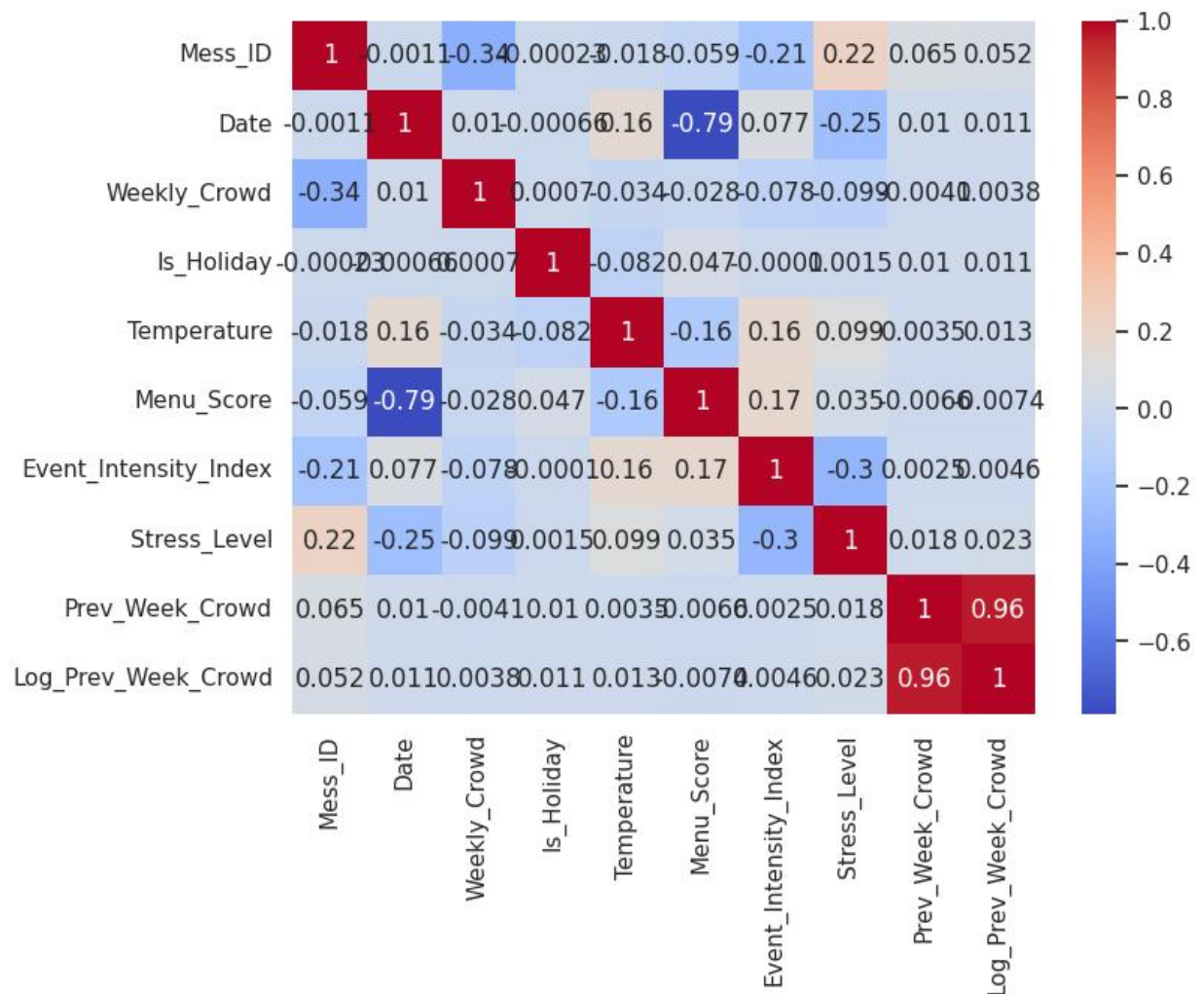
## 4. Average Crowd per Mess



**Insight:** Some messes consistently have higher attendance. This explains why Mess\_ID is our top feature.

# Data-analytics Hackathon

## 5. Correlation Heatmap:



### Correlation with Weekly Crowd:

Weekly_Crowd	1.000000
Is_Holiday	0.036891
Menu_Score	-0.009464
Temperature	-0.063811
Event_Intensity_Index	-0.072654
Stress_Level	-0.106176
Mess_ID	-0.335331

### Insight::

Mess\_ID and Stress\_Level are the **most correlated** with Weekly\_Crowd

Holiday, Temperature, and Menu\_Score have **low direct correlation**.

# Data-analytics Hackathon

## Summary of EDA Insights

- Weekly mess crowd fluctuates heavily, not explained by just holidays or menu.
- Academic stress and event intensity play a moderate role.
- Some mess halls are inherently more popular (importance of Mess\_ID).
- The **target variable (Weekly\_Crowd) is skewed** — justifying the use of **log transformation** before modeling.

# Data-analytics Hackathon

## Summary of Approach and Findings

### Data Understanding & Cleaning:

- Analyzed data set with weekly records across 45 mess halls.
- Each record included: Mess\_ID, Weekly\_Crowd, Is\_Holiday, Menu\_Score, Temperature, Event\_Intensity\_Index, and Stress\_Level.
- Identified that Weekly\_Crowd is highly skewed → applied log1p transformation to stabilize variance.

### Exploratory Data Analysis (EDA)

- Found that Mess\_ID, Stress\_Level, and Event\_Intensity\_Index had the most influence on mess crowd.
- Weak correlations from Menu\_Score, Holiday, and Temperature.
- Visualized trends over time and distribution across mess halls.

### Modeling

- Trained a RandomForestRegressor using features: ['Is\_Holiday', 'Menu\_Score', 'Event\_Intensity\_Index', 'Stress\_Level', 'Mess\_ID', 'Temperature']
- Used log-transformed target variable (Weekly\_Crowd) to improve learning.
- Achieved high accuracy with:
- MAE: 601.78 people
- $R^2$ : 0.9714 (97.14% variance explained)

### Feature Importance

- Mess\_ID was the most important factor (mess popularity matters)
- Followed by Event\_Intensity\_Index and Stress\_Level
- Menu Score and Holiday had negligible effect

# Data-analytics Hackathon

## Findings

- Mess popularity (Mess\_ID) drives attendance more than menu, temperature, or holidays.
- Event-heavy weeks slightly reduce crowd, likely due to off-campus distractions.
- Stressful academic periods reduce mess usage — students may skip meals.
- Menu Score and Holiday flag had minimal effect on actual mess attendance.

## Suggested Actions

Insight	Recommendation
Stressful weeks reduce turnout	Offer packed meals or flexible hours during exams
Mess popularity varies	Balance load by improving underutilized messes
Weak influence of holidays/menu	Focus on external scheduling/events more than tweaking the menu
Accurate prediction model now available	Use dashboard to forecast weekly needs (staffing, food stock)