Mess Hall Crowd Analysis & Prediction

Prepared by: B Sanjai
Date: July 14, 2024

The Problem: Unpredictable Crowds

- **The Challenge:** Mess hall attendance is highly irregular. Some weeks are significantly overcrowded, leading to long waits and resource strain, while other weeks are underused, resulting in wasted food and staff time.
- **The Unknown:** The key factors driving these fluctuations were not well understood, making proactive planning difficult.
- **Our Goal:** To use historical data to uncover the hidden patterns behind crowd changes.

Our Objectives

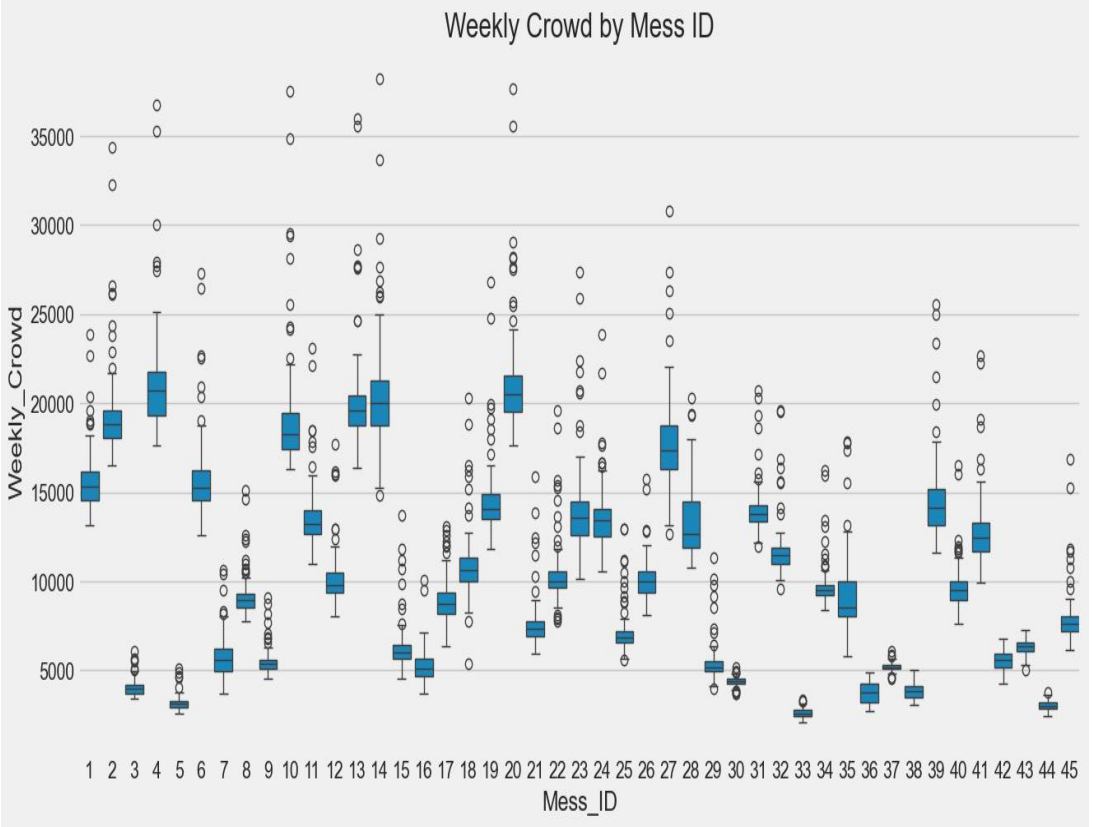
- Analyze historical data to understand weekly crowd variations.
- Identify the most important factors influencing these changes.
- Visualize the findings through clear and insightful graphs.
- Build a highly accurate predictive model to forecast future crowd levels.
- Recommend actionable, data-driven strategies for the administration.

Step 1 - Data Cleaning & Preparation

- Standardized Dates: Corrected inconsistent date formats (e.g., DD/MM/YYYY vs. DD-MM-YYYY) into a uniform, machine-readable format.
- Feature Engineering: Created new time-based features from the date to help the model understand seasonality:
  - Year
  - Month
  - Week_of_Year
- This foundational step ensured the data was clean, consistent, and ready for accurate analysis and modeling.

Step 2 - Key Finding 1: Not All Messes Are Equal

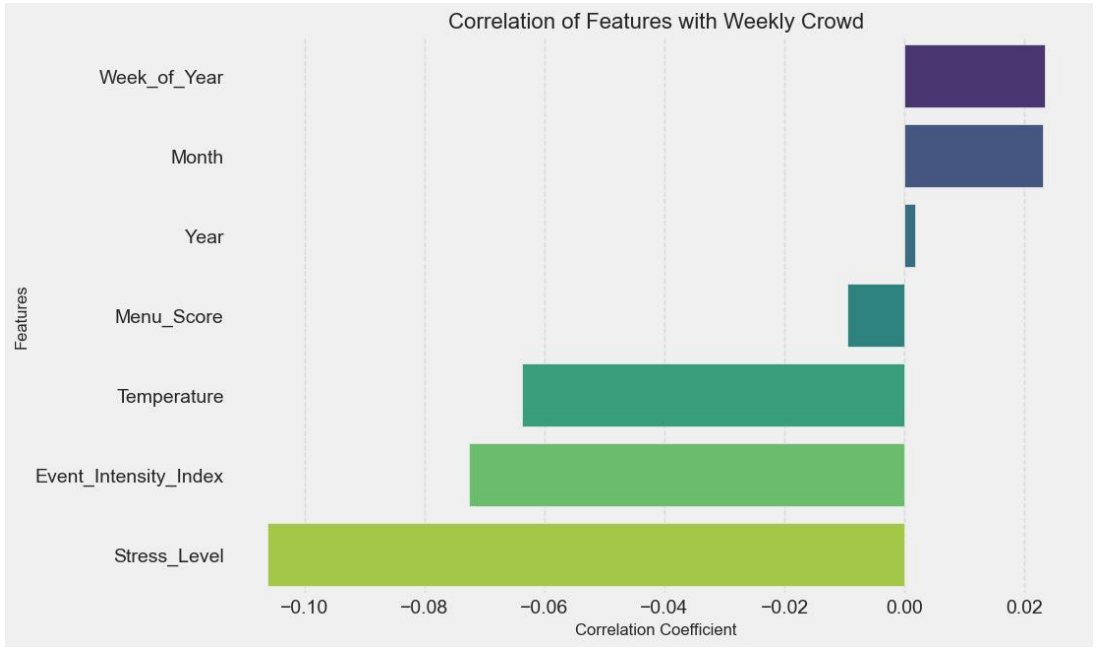
- Exploratory Data Analysis (EDA) immediately revealed that Mess_ID is the single most significant factor.
- Different messes have vastly different average crowd levels and variability.
- Insight: A "one-size-fits-all" approach to staffing and inventory is inefficient. Planning must be tailored to each specific mess hall.

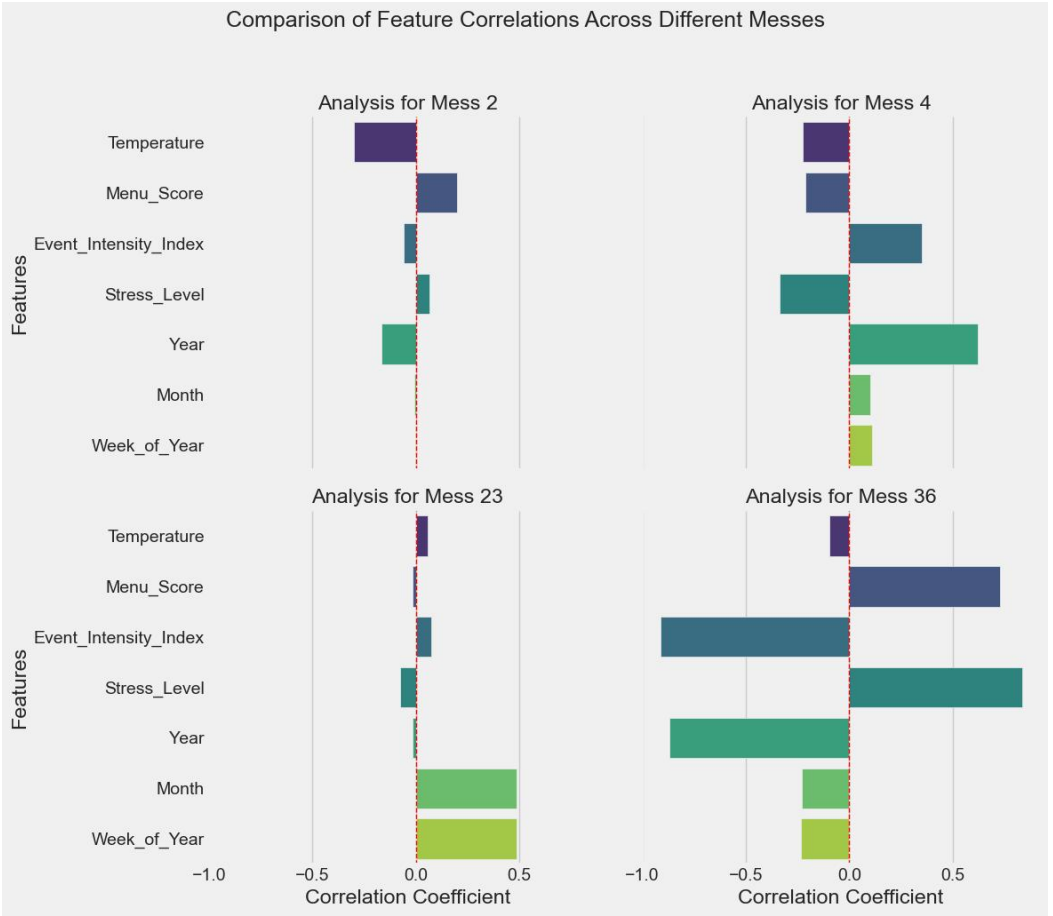


Step 2 - Key Finding 2: Campus Life Drives Demand

- Correlation analysis showed a strong relationship between campus activities and mess attendance.
- The most influential factors after Mess_ID are:
- Stress Level: Higher academic stress (e.g., exams) leads to significantly larger crowds.
- Menu Score: A more popular menu directly translates to more diners.
- Event Intensity: Weeks with more campus events also see a rise in mess attendance.

CORRELATION FOR WHOLE MESS VS ONE MESS



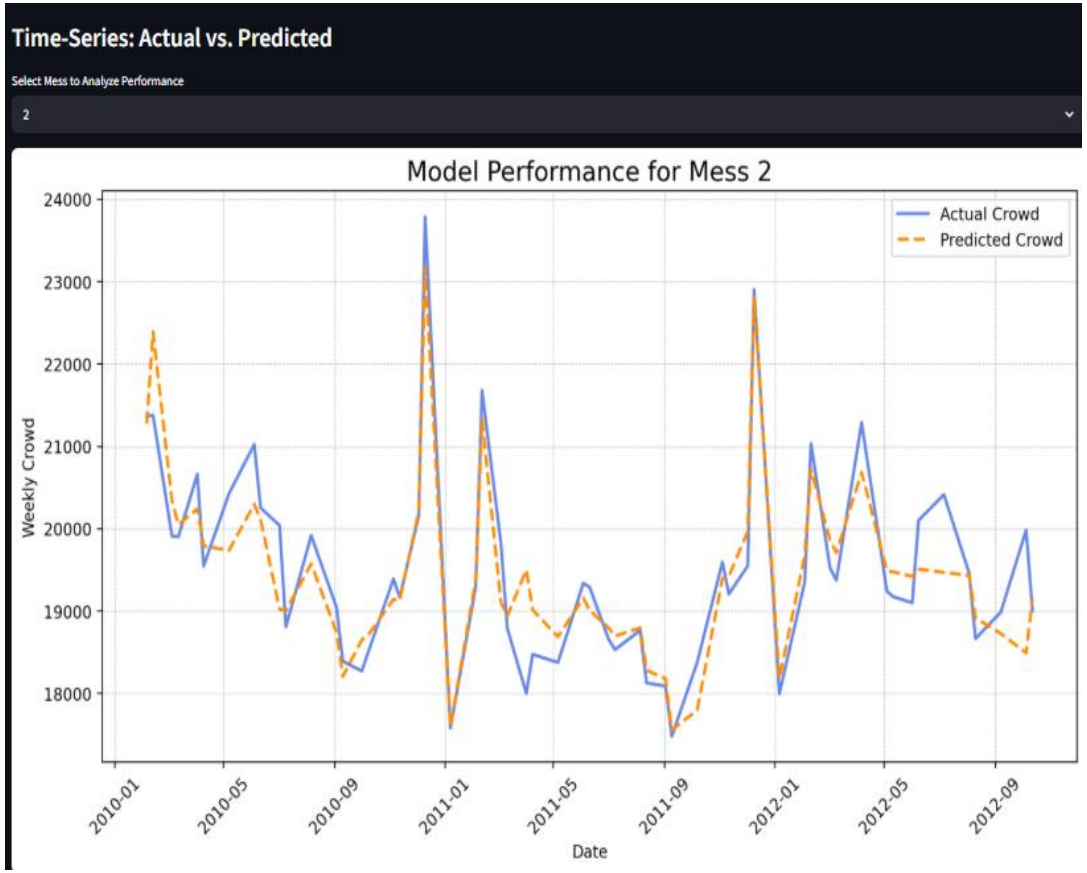


Step 3 - Building the Predictive Model

- Algorithm Chosen: We implemented a state-of-the-art XGBoost (Extreme Gradient Boosting) model. This algorithm is renowned for its high accuracy on tabular data.
- Features Used: The model was trained on all relevant factors to learn the complex relationships between them:
- Mess_ID, Date features (Year, Month, Week), Temperature, Menu_Score, Stress_Level, Event_Intensity_Index, and Is_Holiday.

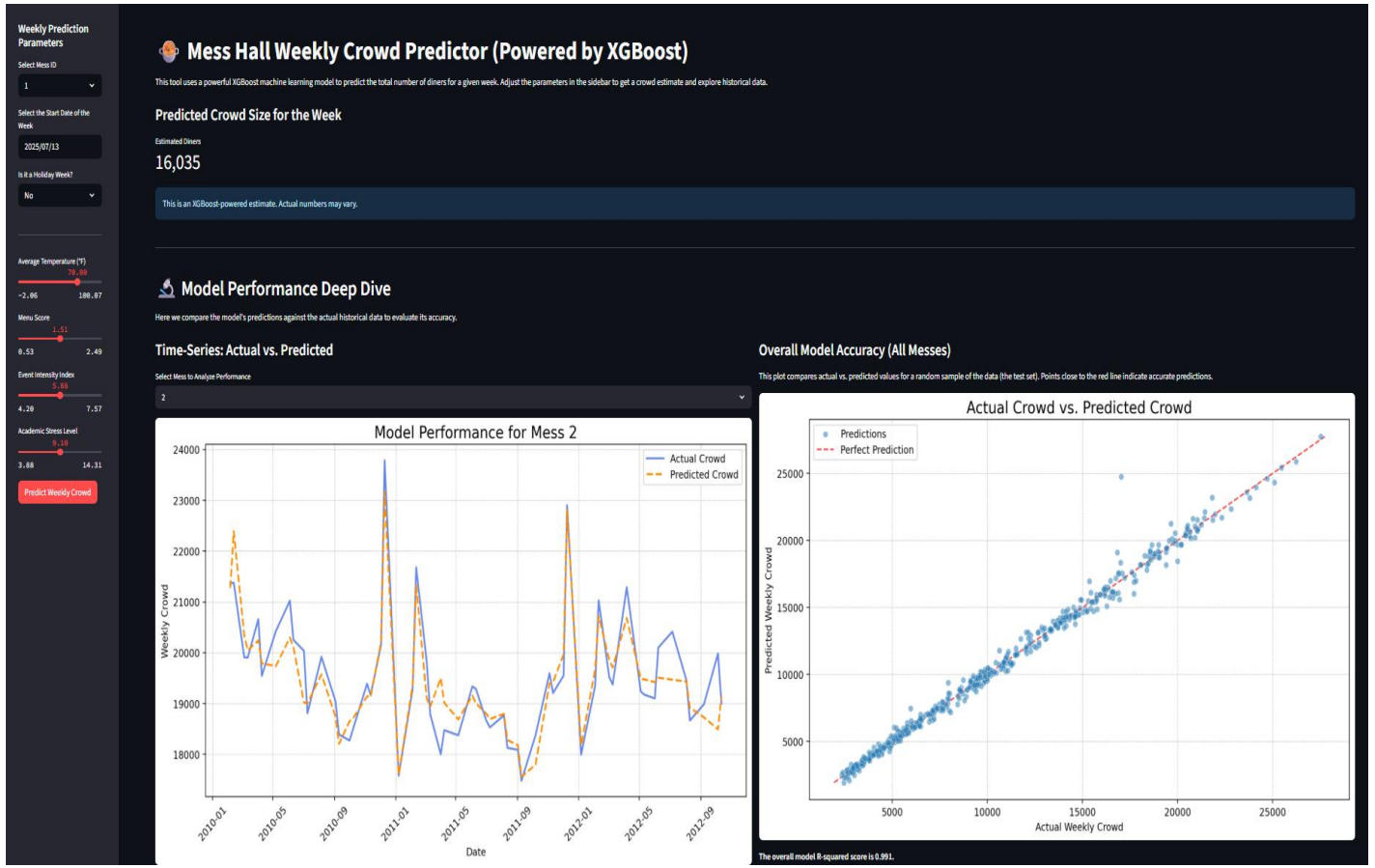
Step 4 - Model Performance: Highly Accurate

- The model's performance is excellent, achieving an R-squared (R^2) score of 0.947.
- In simple terms: This means our model can explain 94.7% of the variability in weekly crowd numbers, making it extremely reliable for forecasting.
- The charts show the model's predictions (orange) closely tracking the actual historical data (blue).



Step 5 - The Solution: A Predictive Dashboard

- To make the model usable for the mess administration, we built an interactive dashboard.
- Functionality:
- Managers can select a mess and input the expected conditions for an upcoming week (e.g., menu score, stress level).
- The dashboard provides an instant, data-driven prediction of the expected crowd size.
- This tool transforms our analysis into a practical decision-making asset.



Final Insights & Recommendations

- Insight 1: The "Average" Mess Doesn't Exist—Context is Everything.The single most powerful predictor of a mess hall's crowd is its own identity (Mess_ID)
- Insight 2: The Menu is Not Just Food; It's a Crowd Management Tool.Because Menu_Score has a direct and measurable impact on crowd size, it can be used strategically to manage demand rather than just satisfy hunger.Even Academic stress Menu Quality and Cmpus event affects Some mess.
- Insight 3: A Highly Accurate Predictive Model is Now Possible.The relationships between these factors are complex, but they are not random. Our final XGBoost model achieved an R-squared score of 0.947, meaning it can successfully explain 94.7% of the weekly variance in crowd size.