Grace Coccagna  
Professor Eren Bilen  
Data 400  
1/30/2025

The Economics of Weather: Analyzing MLB Attendance Trends

Through my advancement in the Data Analytics program at Dickinson College, I had the opportunity to complete a concentration in Economics. I chose this topic due to my passion for understanding the reasoning behind how businesses and corporations operate. One strong area of interest I have developed through my experiences is a deep passion for baseball operations and the economics behind the sport. For my project, my goal is to analyze how weather conditions impact attendance and ticket sales for Major League Baseball (MLB) games. I will be utilizing data from the years 2015 to 2018 for all 30 MLB teams. This study will identify key trends and develop predictive models to forecast attendance and revenue based on weather forecasts. While gathering data for this project, I encountered challenges due to the lack of publicly available game-by-game ticket price data. As a solution, I manually created and merged a dataset containing the average ticket prices for each team across the years. By integrating weather, attendance, and ticket sales data, this project aims to provide insights that can inform decision-making for teams, ticket vendors, and stadium managers.

The datasets I will be using for this project were sourced from Kaggle, with additional manual creation and merging to incorporate ticket price information from Statista. The weather data includes variables such as temperature, rainfall, and wind conditions, while the games dataset captures attendance numbers and average ticket prices for each game. An Exploratory Data Analysis (EDA) will be conducted to identify trends in weather conditions and their relationship with attendance. To achieve the project objectives, I plan to use machine learning algorithms such as linear regression to identify linear relationships between weather variables and attendance. Decision trees and random forests may also be considered to capture more complex, non-linear relationships and perform feature importance analysis. Additionally, I will create visualizations to effectively communicate my findings, either utilizing Python or a Power BI dashboard. This project will explore the development of a predictive model that could estimate game attendance based on weather forecasts, providing valuable insights into potential revenue changes.

This project is important to various stakeholders, such as baseball teams, ticket vendors, and stadium operations managers. It will carefully consider ethical and societal implications, particularly regarding data privacy and ensuring that predictive models are transparent and explainable. Additionally, while weather is a natural variable, this model could play a significant role in operational planning by identifying its effects on revenue changes. I aim to deliver actionable insights and a predictive model that demonstrates the relationship between weather conditions and baseball game attendance, ultimately enabling better decision-making for key stakeholders.