Sports Data Analysis

Week 5

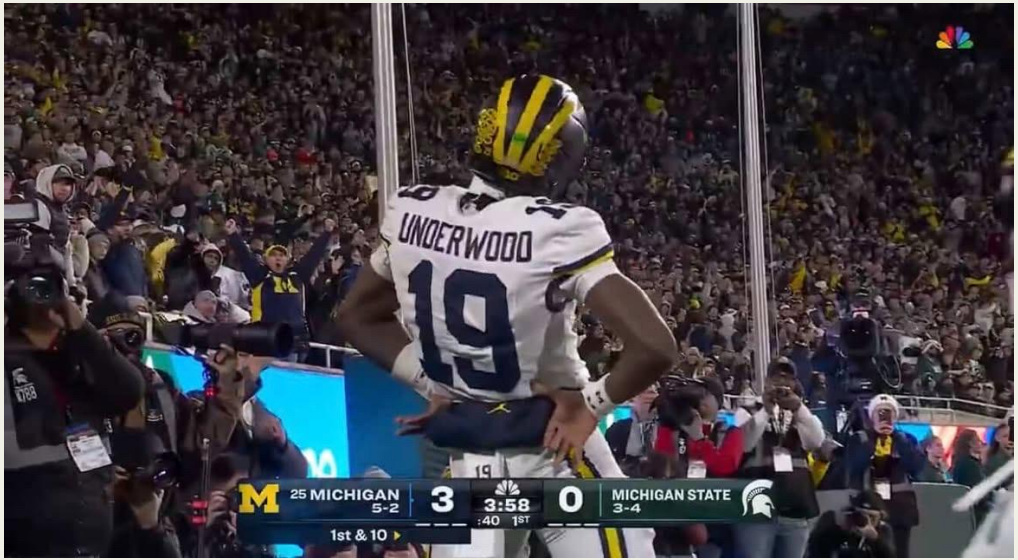


TABLE OF CONTENTS

Icebreaker

Best Hot Take

Aggregation

Often you'll want to compare data between **groups**

- Group up your data based on a **category** (that you can make if needed)
- Calculate mean/median/standard deviation/other stats related to these groups for **comparison**
- Can also group by **multiple** categories for extended comparisons

Correlations

Correlations can be extremely useful for model creation

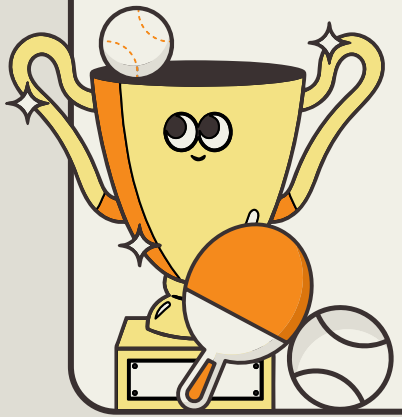
- Ranges from [-1, 1], the closer the value is to +/- 1 the **stronger** the correlation
- Variables that correlate more strongly with the response variable you chose will make a **better model** for prediction
- Visualize correlations with **heatmaps** to easily tell what to look at

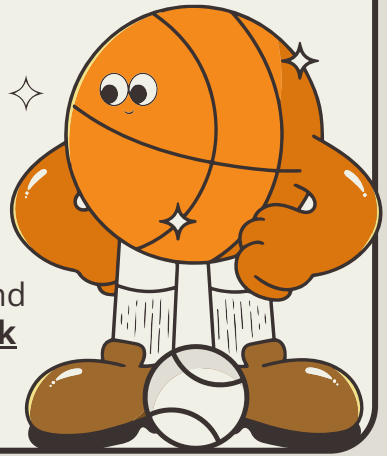
Example Code for Topics Discussed + Additional/Advanced Visualization Examples

https://colab.research.google.com/drive/13AbE7y3EH4oI3bzYNf0DKTzZcS0yo60D?usp=sharing

https://github.com/MichiganDataScienceTeam/F25-SportsDataAnalysis/blob/main/Data/Lions%202024/lions_2024_season.csv

Questions?





Goals for Today

- Use **grouping and aggregation** to compare groups and generate insights
- Make a **couple more plots** (try some fancy ones)
- Calculate **correlations** and visualize them
- **Pick** a response variable and what variables you'll use to predict it (even if you can't really make a model, we'll practice it anyway)