



# **Moneyball 2.0**

## **Predicting Pitcher Salaries**

TEAMSIX Final Project  
Segment 2

# Topic

This project will create a supervised learning model to predict MLB pitcher future salaries using MLB pitcher statistics

# Reasoning

Reasons for topic selection:

- The abundance of baseball statistics currently available online
- Ideal opportunity for machine learning
- WE LOVE BASEBALL!

# Data Sources

The data sources being used:

- Lahman's Baseball Database
  - MLB pitching statistics
  - Salary data
- Years 2016-2019
  - Did not use 2020-2021 due to the pandemic

# Why?

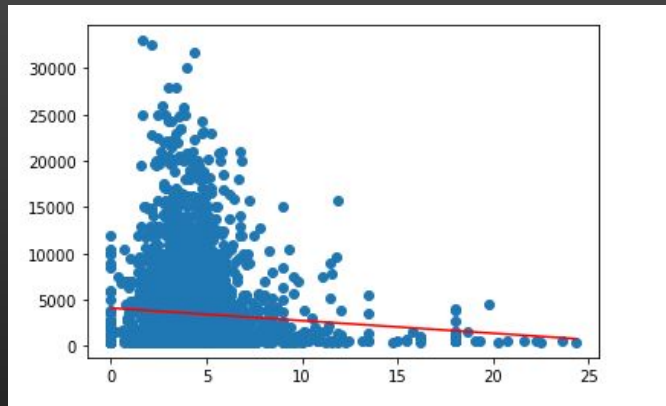
We are planning to use machine learning to answer the following questions:

- Is there a correlation between various pitching statistics and salaries?
  - Which statistics?
  - Is a single statistic a good indicator or multiple in combination?
- Which machine learning model creates the best prediction?
- If this works, which MLB team will hire us?

# Our first model...

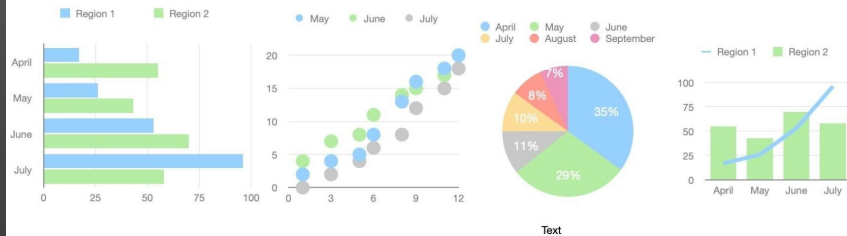
Our first model used linear regression and a very simple subset of our data:

- Salary and ERA only
  - 2010-2016
  - Linear regression
  - Removed outliers



**Analysis,  
analysis,  
analysis**

# The Dashboard



Text

