# Data Science 2: Statistics for Data Science

## Report on the Analysis & Predictive Modelling of Strikeouts vs Home Runs in Major League Baseball Players 1871 - 2000

### Submitted by Group 8

Han Chen

Andrew Lagree

Rebecca Manning

Charlie Wang

Nathan Ye

Johanna Yu

### Supervised By:

Dr. Sergiy Nokhrin

Dr. Amanda Easson

Dr. Prakriti Sodhi

December 2021

# Objectives

What are you setting out to prove or predict? What is your rationale for there being a correlation in the data that you’re looking to confirm and/or exploit?

# Data Preparation

What was your data source (e.g., web scraping, corporate data, a standard machine learning dataset, open data, etc.)? How good was the data quality? What did you need to do to procure it? What tools or code did you need to use to prepare it for analysis? What challenges did you face?

# Analysis or model

If you are conducting an inference test, explain the analysis you performed clearly and include well-labelled diagrams to make your points. If you chose to do a predictive model, explain the model, how you trained and tested it, and how well it works. How did you confirm that the data met the requirements for the test or modelling technique to be valid?

# Conclusions

Did you prove/disprove your hypothesis or create a useful model? What did you learn about your data set?