

## **Sabermetrics 101; An Introduction to Baseball Analytics**

### **What Is This Workshop About?**

Sabermetrics 101: Introduction to Baseball Analytics is an introduction to sabermetrics, baseball analytics, data science, the SAS Language, and SQL. It covers the theory and the fundamentals of the emerging science of Sabermetrics. We will discuss the game of baseball, not through consensus or a fan's conventional wisdom, but by searching for objective knowledge in baseball performance.

Other areas of sabermetrics will be analyzed and better understood with historical baseball data. The Workshop also serves as applied introduction to the basics of data science, that requires skills in computation (accessing data), statistics, and communicating results of analyses.

Using baseball data, the basics of descriptive statistics, the SAS Language, and the SAS Data Step (and SQL) will be covered.

### **Workshop Reminders**

Additional information regarding optional resources will be provided at the start of the Workshop.

### **Workshop Schedule**

#### **Introduction: Getting Started**

- Overview of Viya for Learners (VFL)
- Workshop Orientation
- Introduction to datasets and platforms

---

#### **Module 1: Workshop Introduction – Defining the Game**

- Key Concepts: Sabermetrics, Statistics, Data Science, Relational Databases
- Data Management Basics: Data Step and PROC SQL (SELECT Statements)
- Baseball History: Henry Chadwick

---

#### **Module 2: Measuring Hitting – Batting Analytics**

- Evaluating Batting Performance: Batting Average, OBP, SLG
  - Descriptive Statistics: Measures of Center and Spread, Distribution Shapes
  - Data Management Practice: Basic Queries
  - Baseball History: Hugh Fullerton
- 

### **Module 3: Advanced Queries and Run Estimation**

- Run Scoring: Relationship Between Runs, Runs Allowed, and Wins
  - Statistical Concepts: Correlation and Dependence
  - Merging Tables
  - Coding: Run Estimators
  - Baseball History: F.C. Lane
- 

### **Module 4: Pitching, Fielding, and SAS Language Foundations**

- Pitching and Fielding Metrics
  - Introduction to the SAS Language: Variables, Calculations, Graphing Basics
  - Baseball History: Alan Roth and Branch Rickey
- 

### **Module 5: Advanced Coding for Sabermetrics**

- Programming: Loops, Conditionals, Functions
  - Statistical Concepts: Regression to the Mean, Sample Size, Statistical Reliability
  - Advanced Hitting Analytics: Run Estimators, Pythagorean Regression
  - Baseball History: Earnshaw Cook
- 

### **Module 6: Putting It All Together – WAR and Contextual Adjustments**

- Adjusting for Park Factors and League Run Environments
- Defining Replacement Level and Understanding WAR

- Advanced Metrics: WAR for Offense and Defense
  - Baseball History: George Lindsay
- 

### **Final Project**

- Cumulative evaluation covering sabermetrics, SAS code, and statistical concepts.

### **Notes:**

This course can be developed with Python and/or R. It would be great to have an SQL “sandbox” from which to query the data.