MLB Data (Tables in Excel)

Executive Summary: [¼ page description of the project highlights. What are the key results?]

We will focus on importing data and do some first exploratory analysis.

Table of Contents

1. Frame the problem ........................................................................................................................................2
2. Collect the needed data ....................................................................................................................................2
   Data Sources .................................................................................................................................................... 2
   Import Data .................................................................................................................................................... 2
3. Prepare and explore the data ..........................................................................................................................2
   Format as Table ............................................................................................................................................. 3
   Split data in Columns .................................................................................................................................. 2
   Data Validation ............................................................................................................................................. 3
   Empty cells .................................................................................................................................................... 3
   Functions and Calculations ............................................................................................................................ 3
   Pivot Table .................................................................................................................................................... 3
4. Models and Algorithms ..................................................................................................................................4
5. Communicate the results and/or implement a data-driven product ..............................................................4
6. Evaluate the value of the project .....................................................................................................................4
1. **Frame the problem**

[What it the question you can answer? Why are they important?]

Here are some questions:

- Do some teams have more younger players?
- Does the weight of a player have an impact on their performance?

Found patterns can be used to inform selecting new players for a team. (Note: I have no idea about baseball!)

2. **Collect the needed data**

[Data source, data quality and reliability.]

**Data Sources**

I found these potential data sources using a quick web search:

- [http://mlb.mlb.com/stats/sortable.jsp#elem=%5BObject%5D&tab_level=child&click_text=Sortable+Player+hitting&game_type='R'&season=2018&season_type=ANY&league_code='MLB'&sectionType=sp&statType=hitting&page=1&ts=1547740208757](http://mlb.mlb.com/stats/sortable.jsp#elem=%5BObject%5D&tab_level=child&click_text=Sortable+Player+hitting&game_type='R'&season=2018&season_type=ANY&league_code='MLB'&sectionType=sp&statType=hitting&page=1&ts=1547740208757)

**Import Data**

Data comes in different formats

- Text: CSV, fixed-width text, XML
- Application specific: Excel, SAS, SPSS, Stata, etc.
- Web: HTML, XHTML
- Relational database

You can try:

- Copy & paste into Excel (use CRTL and SHIFT for marking the table)
- Excel provides: **Data > From Web**
- Use other tools. For example, Firefox: Install Add-ons > Extensions > Table to Excel (use red icon next to URL in Browser)

3. **Prepare and explore the data**

[Clean and connect the data.]

**Split data in Columns**

Add empty columns to the right.

**Data > Text to Columns**
Clean the player name. Careful, there are some players with middle names or "Jr." Use sorting to find them.

**Data Validation**

Use sorting or Data > Data Validation

**Empty cells**

Find cells with Home > Find & Select > Go to Special and select Blanks.

Selected cells can be replaced with entering a value and CTRL + Enter

**Format as Table**

Select data for table with CTRL + Shift + Arrow keys and use Home > Format as Table

Use Table Tools > Design (after clicking inside the table) to set the name. There is also the Formulas > Name Manager.

Filter by team or position.

Insert > Slicer for more convenient filtering.

Tables can be selected using CTRL + a and provide sorting and filtering and a context menu (when selected).

Select columns in a table with CTRL + a and then select in the columns with the mouse.

Use the context menu to add averages, conditional formatting, etc.

Home > Conditional Formatting for more choices.

**Hint:** Always look at the menus that appear when you select a table or chart.

**Functions and Calculations**

Rows/Columns, references, absolute references ($). For example, =A1 or =$A$1.

Structured references in tables often look like =Tablename[@[Columnname]]

or =Tablename[Columnname]

Formulas > Text (use substitute to clean the position name)

Function result depend on the original data. Convert function results into fixed values: Copy + Home > Paste > Value (Note: you may have to convert the table back to ranges and make a new table)

Use Formulas > More > Statistical > CountIF to calculate the proportion of players younger than 25 years.

Are some players under/overweight? Calculate the Body Mass Index.

**Pivot Table**

Reorganize and summarize selected columns to create a report. Example (from Wikipedia):
Compare the average height, weight and age by position and/or team.

Highlight the table and use Insert > Pivot Table

Define columns, rows and the aggregation function.

Use conditional formatting (check conditional formatting rules), sorting, slicer (team, age, etc.), add a chart.

4. Models and Algorithms
[What type of problem do we have? Forecast values, make yes/no decisions, compare data?]

We will learn about models and algorithms later in this course.

5. Communicate the results and/or implement a data-driven product
[Prepare report, visualizations, real-time dash boards. Implement decision support tools or incorporate algorithms in apps and web sites.]

6. Evaluate the value of the project
[Does the project answer the initial questions?]

References