**Column Names of “retro\_final\_PA\_1990-2020b.csv”**

**• each row in the dataset is an “event”**

**• the data is arranged in order by year, team, game, plate appearance**

“row\_idx” === row number of dataset, to preserve order when rearranging

“GAME\_ID" == string identifier of this game

"PARK" == string identifier of the park

"DATE" == date of the game

"YEAR" == year of the game

"HOME\_TEAM\_ID" == string identifier of the home team

"AWAY\_TEAM\_ID" == string identifier of the away team

"INNING" == inning of the plate appearance

"BAT\_HOME\_IND" == 1 if this batter is on the home team else 0

"BAT\_ID" == string identifier of the batter

"BAT\_NAME" == string name of the batter

"BAT\_HAND" == string handedness of the batter {L,R,B}

"FIELD\_POS" == integer code for the primary fielding position of this batter

"PIT\_ID" == string identifier of the pitcher

"PIT\_NAME"== string name of the pitcher

"PIT\_HAND" == string handedness of the pitcher {L,R}

"COUNT" == string count at the occurrence of each event (e.g., “11” or “32”)

"PITCH\_SEQ\_TX" == string indicating the pitches thrown during this plate appearance

"EVENT\_TX" == string indicating the event of this plate appearance (e.g., K or WP)

"HIT\_VAL" == 1 if single, 2 if double, 3 if triple, 4 if HR, else 0

"HIT\_BINARY" == 1 if this plate appearance is a hit else 0

"consecutive.bat.row" == 1 if this row in the dataset has the same batter as the previous row, else 0

"BATTER\_SEQ\_NUM" == given a game and a home/away team, the integer batter number of the current batter {1,2,3,…}

"ORDER\_CT" == the integer number of the batter in the order {1,2,3,…,9}

"HAND\_MATCH" == 1 if BAT\_HAND and PIT\_HAND match, else 0

"SP\_IND" == 1 if the current pitcher is the starting pitcher

"HOME\_DIV" == string indicating the division of the home team (e.g., “AL\_E” or “NL\_C”)

"AWAY\_DIV"" == string indicating the division of the away team

"HOME\_LEAGUE" == string indicating the league of the home team {“AL”, “NL”}

"AWAY\_LEAGUE"== string indicating the league of the away team

"IN\_DIV"== 1 if the HOME\_DIV matches AWAY\_DIV, else 0

"IN\_LEAGUE" == 1 if the HOME\_LEAGUE matches AWAY\_LEAGUE, else 0

"EVENT\_OUTS\_CT" == integer number of events during this plate appearance {0,1,2,3}

"OUTS\_CT" == integer outs count prior to this plate appearance within a given half-inning {0,1,2}

"EVENT\_CODE" == string {“W”, “HP”, “IW”, “other”} helper column for WOBA

"PA\_IND" == 1 if this row is a plate appearance, else 0

"AB\_IND" == 1 if this row is an at-bat, else 0

"WOBA\_APP" == 1 if this row appears in wOBA, else 0

"EVENT\_WOBA" == the wOBA of this event

"EVENT\_RUNS" == integer number of runs generated by this plate appearance {0,1,2,3,4}

"EVENT\_PITCH\_COUNT" == integer number of pitches during this plate appearance

"PITCH\_COUNT\_CUMU" == within a game, the cumulative number of pitches by the current pitcher prior to this plate appearance

"PITCH\_COUNT\_FINAL" == within a game, the final pitch count of the current pitcher

"DAYS\_SINCE\_SZN\_START" == integer number of days since the season has started

"PIT\_REST" == integer number of days of rest a pitcher has prior to this game, with a maximum of 6 {1,2,3,4,5,6}

"EVENT\_WOBA\_19" === EVENT\_WOBA using 2019 wOBA weights for each season

“BASE\_STATE” === {000,100,010,001,110,101,011,111} indicates the men on base, as a string, at the start of this plate appearance

*Discontinued: need to be fixed and added back*

"EVENT\_ER\_CT" == integer number of runs earned during this plate appearance [slight errors have been noticed in this column, i.e. values of -1 due to error in earned runs recording, to be fixed soon. For now please use EVENT\_RUNS which is ok.]

"EVENT\_RBI\_CT" == integer number of RBIs during this plate appearance [slight errors have been noticed in this column, i.e. values of -1 due to error in rbi recording, to be fixed soon. For now please use EVENT\_RUNS which is ok.]

“MAN\_ON\_1ST” === 1 if there is a man on first base at the start of this plate appearance, else 0

“MAN\_ON\_2ND” === 1 if there is a man on second base at the start of this plate appearance, else 0

“MAN\_ON\_3RD” === 1 if there is a man on third base at the start of this plate appearance, else 0

*Discontinued: do it yourself, and on a smaller subset of the data.*

"NUM\_WOBA\_APP\_BAT" === number of wOBA appearances during the season for a given batter up to this point in the season

“NUM\_WOBA\_APP\_FINAL\_BAT” === final number of wOBA appearances for this batter this season

"NUM\_WOBA\_APP\_PIT" === number of wOBA appearances during the season for a given pitcher up to this point in the season

“NUM\_WOBA\_APP\_FINAL\_PIT” === final number of wOBA appearances for this pitcher this season

"WOBA\_AVG\_BAT" == within a season, the avg. wOBA of this batter prior to this plate appearance

"WOBA\_AVG\_BAT\_19" === WOBA\_AVG\_BAT using 2019 wOBA weights for each season

“WOBA\_FINAL\_BAT” === final avg. wOBA of this batter during this season

“WOBA\_FINAL\_BAT\_19” === final avg. wOBA of this batter during this season, using 2019 woba weights

"WOBA\_AVG\_PIT" == within a season, the avg. wOBA of this pitcher prior to this plate appearance

"WOBA\_AVG\_PIT\_19" === WOBA\_AVG\_PIT using 2019 wOBA weights for each season

“WOBA\_FINAL\_PIT” === final avg. wOBA of this pitcher during this season

“WOBA\_FINAL\_PIT\_19” === final avg. wOBA of this pitcher during this season, using 2019 woba weights