## APPENDIX A – LAHMAN DATASET

Table Name/ Description					
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?	
playerID	Player ID code	Unique Values: 1867 Type: Object	None	Yes/No	
yearID	Year	Min: 1933 Max: 2019 Unique Values: 87 Type: float64	Missing Values:		
gameNum	Game number (for years in which more than one game was played)	Min: 0 Max: 2 Unique Values: 3 Type: float64	Missing Values:		
gameID	Game ID code	Unique Values: 90 Type: Object	Missing Values: 50		
teamID	Team; a factor	Unique Values: 44 Type: Object	None		
lgID	League; a factor with levels AL, NL	Unique Values: 2 Type: Object	None		
GP	Game played (zero if player did not appear in game)	Min: 0 Max: 1 Unique Values: 2 Type: int64	None		
startingPos	If the player started, what position he played	Min: 1 Max: 10 Unique Values: 10 Type: float64	Missing Values: 3684		
Table Name/ Description	Appearances Description: Data on player ap A data frame with 107357 obs		ables		
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?	
yearID	Year	Min: 1871 Max: 2019 Unique Values: 149 Type: int64	None		
teamID	Team; a factor	Unique Values: 149 Type: Object	None		
lgID	League; a factor with levels AA, AL, FL, NL, PL, UA	Unique Values: 6 Type: Object	Missing Values: 737		
playerID	Player ID code	Unique Values: 19690 Type: Object	None		

		Min: 1936		
		Type: Object		
awardID	Name of award won	Unique Values: 2	None	
playerID	Manager (player) ID code	Unique Values: 83 Type: Object	None	Yes/No
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
Table Name/ Description	AwardsManagers Description: Award information A data frame with 179 observat	ions on the following 6 variables		
Table Name/	AwardsManagars	Unique Values: 50	8962	
G_pr	Games as pinch runner	Min-Max: 0 – 92	Missing Values:	
C	Company	Unique Values: 94	8962	
G_ph	Games as pinch hitter	Min-Max: 0 – 109	Missing Values:	
		Unique Values: 157	1264	
G_dh	Games as designated hitter	Min-Max: 0 – 162	Missing Values:	
G_of	Games as outfielder	Min-Max: 0 – 164		
G_rf	Games as right fielder	Min-Max: 0 – 162		
G_cf	Games as centerfielder	Min-Max: 0 – 162		
G_lf	Games as leftfielder	Min-Max: 0 – 163	None	
G_ss	Games as shortstop	Min-Max: 0 – 165	None	
G_3b	Games as third baseman	Min-Max: 0 – 164		
G_2b	Games as second baseman	Min-Max: 0 – 163		
G_1b	Games as first baseman	Min-Max: 0 – 162	None	
		Unique Values: 157		
G_c	Games as catcher	Min-Max: 0 – 160	None	
<del></del>	·	Unique Values: 95		
G_p	Games as pitcher	Min-Max: 0 – 106	None	
		Type: int64		
		Unique Values: 166		
	appeared on defense	Max: 165	7698	
G_defense	Games in which player	Min: 0	Missing Values:	
		Type: int64		
		Unique Values: 166		
~~	Sames in which player batted	Max: 165	110.10	
G_batting	Games in which player batted	Min: 0	None	
		Type: int64		
		Unique Values: 165	0302	
G3	Games started	Max: 164	Missing Values: 8962	
GS	Games started	Type: int64 Min: 0	Missing Values	
		Unique Values: 165		
		Max: 165		
G_all	Total games played	Min: 1	None	
- "	<del> </del>	Type: Object	50	
gameID	Game ID code	Unique Values: 90	Missing Values:	
		Type: float64		
	was played)	Unique Values: 3		
	which more than one game	Max: 2	1	
gameNum	Game number (for years in	Min: 0	Missing Values:	

Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
Table Name/ Description		ations on the following 7 variables		1
notes	Notes about the award	Unique Values: 29 Type: Object	Missing Values: 1468	
		туре. Објест	Null values should be converted to N or zeros	
tie	AL, NL, ML, NL Award was a tie (Y or N)	Type: Object Unique Values: 1 Type: Object	Missing Values: 6187	
lgID	League; a factor with levels	Type: int64 Unique Values: 4	None	
yearID	Year	Min: 1936 Max: 2016 Unique Values: 81	None	
awardID	Name of award won	Unique Values: 29 Type: Object	None	
playerID	Player ID code	Unique Values: 1360 Type: Object	None	Yes/No
Variable Name	•	vations on the following 6 variable  Statistical Summary/Type	Errors	Used in final dataset?
Table Name/ Description	AwardsPlayers Description: Award information			
notes	Notes about the award	Unique Values: 1 Type: Object	Missing Values:	
		Type: Object	177 Null values should be converted to N or zeros	
tie teamID	Team; a factor  Award was a tie (Y or N)	Unique Values: 44 Type: Object Unique Values: 1	None  Missing Values:	
gameID	Game ID code	Unique Values: 90 Type: Object	Missing Values: 50	
gameNum	Game number (for years in which more than one game was played)	Min: 0 Max: 2 Unique Values: 3 Type: float64	Missing Values: 1	
lgID	League; a factor with levels AL, NL	Unique Values: 2 Type: Object	None	
		Max: 2016 Unique Values: 81 Type: int64		

awardID	name of award votes were received for	Unique Values: 1 Type: Object	None	Yes/No
voorID		Min: 1983	None	
yearID	Year		None	
		Max: 2016		
		Unique Values: 34		
		Type: float64		
lgID	League; a factor with levels	Unique Values: 2	None	
	AL, NL	Type: Object		
playerID	Manager (player) ID code	Unique Values: 112	None	
		Type: Object		
pointsWon	Number of points received	Min: 1	None	
		Max: 154		
		Unique Values: 118		
		Type: int64		
pointsMax	Maximum number of points	Min: 24	Missing Values:	
	possible	Max: 160	11	
		Unique Values: 6		
		Type: float64		
votesFirst	Number of first place votes	Min: 0	None	
	·	Max: 30		
		Unique Values: 30		
		Type: int64		
	AwardsSharePlayers Description: Award voting for m			
Table Name/ Description	Description: Award voting for m A data frame with 6879 observe	ations on the following 7 variable		Head in
Description	Description: Award voting for n		es Errors	Used in final dataset?
Description  Variable Name	Description: Award voting for m A data frame with 6879 observe	Statistical Summary/Type		final
Description  Variable Name	Description: Award voting for m A data frame with 6879 observa	Statistical Summary/Type  Unique Values: 3	Errors	final dataset?
Variable Name awardID	Description: Award voting for m A data frame with 6879 observe  Description  name of award votes were received for	Statistical Summary/Type  Unique Values: 3 Type: Object	Errors None	final dataset?
Variable Name awardID	Description: Award voting for m A data frame with 6879 observa  Description  name of award votes were	Statistical Summary/Type  Unique Values: 3 Type: Object Min: 1911	Errors	final dataset?
Variable Name awardID	Description: Award voting for m A data frame with 6879 observe  Description  name of award votes were received for	Statistical Summary/Type  Unique Values: 3 Type: Object Min: 1911 Max: 2016	Errors None	final dataset?
Variable Name awardID	Description: Award voting for m A data frame with 6879 observe  Description  name of award votes were received for	Statistical Summary/Type  Unique Values: 3 Type: Object Min: 1911 Max: 2016 Unique Values: 98	Errors None	final dataset?
Variable Name awardID yearID	Description: Award voting for m A data frame with 6879 observa  Description  name of award votes were received for  Year	Statistical Summary/Type  Unique Values: 3 Type: Object Min: 1911 Max: 2016 Unique Values: 98 Type: int64	None None	final dataset?
Variable Name awardID yearID	Description: Award voting for many Adata frame with 6879 observed  Description  name of award votes were received for Year  League; a factor with levels AL	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911 Max: 2016 Unique Values: 98 Type: int64 Unique Values: 3	Errors None	final dataset?
Variable Name  awardID  yearID	Description: Award voting for made and the second s	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911 Max: 2016 Unique Values: 98 Type: int64 Unique Values: 3 Type: Object	None None None	final dataset?
Variable Name  awardID  yearID	Description: Award voting for many Adata frame with 6879 observed  Description  name of award votes were received for Year  League; a factor with levels AL	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911  Max: 2016 Unique Values: 98 Type: int64 Unique Values: 3 Type: Object Unique Values: 3 Type: Object Unique Values: 2441	None None	final dataset?
Variable Name  awardID  yearID  IgID  playerID	Description: Award voting for m A data frame with 6879 observa  Description  name of award votes were received for  Year  League; a factor with levels AL ML NL Player ID code	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911  Max: 2016 Unique Values: 98 Type: int64  Unique Values: 3 Type: Object  Unique Values: 3 Type: Object  Unique Values: 2441 Type: Object	None None None None	final dataset?
Variable Name  awardID  yearID  IgID  playerID	Description: Award voting for made and the second s	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911 Max: 2016 Unique Values: 98 Type: int64 Unique Values: 3 Type: Object Unique Values: 3 Type: Object Unique Values: 2441 Type: Object Min: 1	None None None	final dataset?
Variable Name  awardID  yearID  IgID  playerID	Description: Award voting for m A data frame with 6879 observa  Description  name of award votes were received for  Year  League; a factor with levels AL ML NL Player ID code	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911 Max: 2016 Unique Values: 98 Type: int64 Unique Values: 3 Type: Object Unique Values: 2441 Type: Object Min: 1 Max: 154	None None None None	final dataset?
Variable Name  awardID  yearID  IgID  playerID	Description: Award voting for m A data frame with 6879 observa  Description  name of award votes were received for  Year  League; a factor with levels AL ML NL Player ID code	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911 Max: 2016 Unique Values: 98 Type: int64 Unique Values: 3 Type: Object  Unique Values: 3 Type: Object  Unique Values: 2441 Type: Object  Min: 1 Max: 154 Unique Values: 349	None None None None	final dataset?
Variable Name  awardID  yearID  lgID  playerID  pointsWon	Description: Award voting for m A data frame with 6879 observa  Description  name of award votes were received for Year  League; a factor with levels AL ML NL Player ID code  Number of points received	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911 Max: 2016 Unique Values: 98 Type: int64 Unique Values: 3 Type: Object  Unique Values: 3 Type: Object  Unique Values: 2441 Type: Object  Min: 1 Max: 154 Unique Values: 349 Type: float64	None None None None None None	final dataset?
Variable Name  awardID  yearID  lgID  playerID  pointsWon	Description: Award voting for m A data frame with 6879 observe  Description  name of award votes were received for  Year  League; a factor with levels AL ML NL Player ID code  Number of points received  Maximum number of points	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911  Max: 2016 Unique Values: 98 Type: int64  Unique Values: 3 Type: Object  Unique Values: 3 Type: Object  Unique Values: 2441 Type: Object  Min: 1  Max: 154 Unique Values: 349 Type: float64  Min: 24	None None None None	final dataset?
	Description: Award voting for m A data frame with 6879 observa  Description  name of award votes were received for Year  League; a factor with levels AL ML NL Player ID code  Number of points received	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911  Max: 2016  Unique Values: 98 Type: int64  Unique Values: 3 Type: Object  Unique Values: 2441 Type: Object  Min: 1  Max: 154  Unique Values: 349 Type: float64  Min: 24  Max: 160	None None None None None None	final dataset?
Variable Name  awardID  yearID  lgID  playerID  pointsWon	Description: Award voting for m A data frame with 6879 observe  Description  name of award votes were received for  Year  League; a factor with levels AL ML NL Player ID code  Number of points received  Maximum number of points	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911 Max: 2016 Unique Values: 98 Type: int64 Unique Values: 3 Type: Object Unique Values: 2441 Type: Object Min: 1 Max: 154 Unique Values: 349 Type: float64  Min: 24 Max: 160 Unique Values: 21	None None None None None None	final dataset?
Variable Name  awardID  yearID  lgID  playerID  pointsWon  pointsMax	Description: Award voting for m A data frame with 6879 observation  Description  name of award votes were received for Year  League; a factor with levels AL ML NL Player ID code  Number of points received  Maximum number of points possible	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911 Max: 2016 Unique Values: 98 Type: int64 Unique Values: 3 Type: Object  Unique Values: 3 Type: Object  Unique Values: 2441 Type: Object  Min: 1 Max: 154 Unique Values: 349 Type: float64  Min: 24 Max: 160 Unique Values: 21 Type: int64	None None None None None None None	final dataset?
Variable Name  awardID  yearID  lgID  playerID  pointsWon  pointsMax	Description: Award voting for m A data frame with 6879 observe  Description  name of award votes were received for  Year  League; a factor with levels AL ML NL Player ID code  Number of points received  Maximum number of points	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911 Max: 2016 Unique Values: 98 Type: int64 Unique Values: 3 Type: Object  Unique Values: 3 Type: Object  Unique Values: 2441 Type: Object  Min: 1 Max: 154 Unique Values: 349 Type: float64  Min: 24 Max: 160 Unique Values: 21 Type: int64  Min: 0	None None None None None None None None	final dataset?
Variable Name  awardID  yearID  lgID  playerID  pointsWon	Description: Award voting for m A data frame with 6879 observation  Description  name of award votes were received for Year  League; a factor with levels AL ML NL Player ID code  Number of points received  Maximum number of points possible	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911  Max: 2016  Unique Values: 98 Type: int64  Unique Values: 3 Type: Object  Unique Values: 2441 Type: Object  Min: 1  Max: 154  Unique Values: 349 Type: float64  Min: 24  Max: 160  Unique Values: 21 Type: int64  Min: 0  Max: 30	None None None None None None None	final dataset?
Variable Name  awardID  yearID  lgID  playerID  pointsWon  pointsMax	Description: Award voting for m A data frame with 6879 observation  Description  name of award votes were received for Year  League; a factor with levels AL ML NL Player ID code  Number of points received  Maximum number of points possible	Statistical Summary/Type  Unique Values: 3 Type: Object  Min: 1911 Max: 2016 Unique Values: 98 Type: int64 Unique Values: 3 Type: Object  Unique Values: 3 Type: Object  Unique Values: 2441 Type: Object  Min: 1 Max: 154 Unique Values: 349 Type: float64  Min: 24 Max: 160 Unique Values: 21 Type: int64  Min: 0	None None None None None None None None	final dataset?

Table Name/ Description	Batting Description: Batting table - batt A data frame with 107429 obse	ting statistics ervations on the following 22 vari	ables	
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
playerID	Player ID code	Unique Values: 19689 Type: Object	None	Yes/No
yearID	Year	Min: 1933 Max: 2019 Unique Values: 87 Type: float64	None	
stint	player's stint (order of appearances within a season)	Min: 1 Max: 5 Unique Values: 5 Type: int64	None	
teamID	Team; a factor	Unique Values: 44 Type: Object	None	
lgID	League; a factor with levels AA AL FL NL PL UA	Unique Values: 2 Type: Object	Missing Values: 738	
G	Games: number of games in which a player played	Type: int64	None	
AB	At Bats	Type: int64	None	
R	Runs	Type: int64	None	
Н	Hits: times reached base because of a batted, fair ball without error by the defense	Type: int64	None	
2B	Doubles: hits on which the batter reached second base safely	Type: int64	None	
3B	Triples: hits on which the batter reached third base safely	Type: int64	None	
HR	Homeruns	Type: int64	None	
RBI	Runs Batted In	Type: float64	Missing Values: 756	
SB	Stolen Bases	Type: float64	Missing Values: 2368	
CS	Caught Stealing	Type: float64	Missing Values: 23541	
ВВ	Base on Balls	Type: int64	None	
SO	Strikeouts	Type: float64	Missing Values: 2100	
IBB	Intentional walks	Type: float64	Missing Values: 36651	
НВР	Hit by pitch	Type: float64	Missing Values: 2817	
SH	Sacrifice hits	Type: float64	Missing Values: 6069	
SF	Sacrifice flies	Type: float64	Missing Values: 36104	

GIDP	Grounded into double plays	Type: float64	Missing Values: 25441	
Table Name/ Description	BattingPost Description: Post season battin A data frame with 14750 obser	g statistics vations on the following 22 varia	bles	
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
playerID	Player ID code	Unique Values: 19689 Type: Object	None	Yes/No
yearID	Year	Min: 1933 Max: 2019 Unique Values: 87 Type: float64	None	
stint	player's stint (order of appearances within a season)			
teamID	Team; a factor	Unique Values: 44 Type: Object	None	
lgID	League; a factor with levels  AA AL NL	Unique Values: 2 Type: Object	None	
G	Games: number of games in which a player played	Type: int64	None	
AB	At Bats	Type: int64	None	
R	Runs	Type: int64	None	
Н	Hits: times reached base because of a batted, fair ball without error by the defense	Type: int64	None	
2B	Doubles: hits on which the batter reached second base safely	Type: int64	None	
3B	Triples: hits on which the batter reached third base safely	Type: int64	None	
HR	Homeruns	Type: int64	None	
RBI	Runs Batted In	Type: float64	None	
SB	Stolen Bases	Type: float64	None	
CS	Caught Stealing	Type: float64	Missing Values: 201	
ВВ	Base on Balls	Type: int64	None	
SO	Strikeouts	Type: float64	None	
IBB	Intentional walks	Type: float64	None	
НВР	Hit by pitch	Type: float64	Missing Values: 201	
SH	Sacrifice hits	Type: float64	Missing Values: 201	
SF	Sacrifice flies	Type: float64	Missing Values: 201	
GIDP	Grounded into double plays	Type: float64	Missing Values: 201	

Table Name/ Description		nools' players attended, by playe vations on the following 3 variab		
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
playerID	Player ID code	Unique Values: 6575 Type: Object	None	Yes/No
schoolID	school ID code	Unique Values: 1038 Type: Object	None	
yearID	Year player attended school	Min: 1864 Max: 2014 Unique Values: 151 Type: int64	None	
Table Name/ Description	Fielding Description: Fielding table A data frame with 143046 obse	rvations on the following 18 vari	ables	
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
playerID	Player ID code	Unique Values: 19491 Type: Object	None	Yes/No
yearID	Year	Min: 1871 Max: 2019 Unique Values: 149 Type: int64	None	
stint	player's stint (order of appearances within a season)	Min: 1 Max: 5 Unique Values: 5 Type: int64	None	
teamID	Team; a factor	Unique Values: 149 Type: Object	None	
lgID	League; a factor with levels AA AL FL NL PL UA	Unique Values: 6 Type: Object	Missing Values: 1513	
G	Games	Type: int64	None	
GS	Games Started	Type: float64	Missing Values: 46157	
InnOuts	Time played in the field expressed as outs	Type: float64	Missing Values: 29929	
PO	Putouts	Type: int64	None	
A	Assists	Type: int64	None	
E	Errors	Type: float64	Missing Values:	
DP	Double Plays	Type: int64	None	
РВ	Passed Balls (by catchers)	Type: float64	Missing Values: 131443	
WP	Wild Pitches (by catchers)	Type: float64	Missing Values: 141877	
SB	Opponent Stolen Bases (by catchers)	Type: float64	Missing Values: 134230	

CS	Opponents Caught Stealing (by catchers)	Type: float64	Missing Values: 134230	
ZR	Zone Rating	Type: float64	Missing Values: 141877	
Table Name/ Description		ta: information about positions prations on the following 18 varia	played in the outfield	
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
playerID	Player ID code	Unique Values: 3513 Type: Object	None	Yes/No
yearID	Year	Min: 1933 Max: 2019 Unique Values: 85 Type: int64	None	
stint	player's stint (order of appearances within a season)	Min: 1 Max: 5 Unique Values: 5 Type: int64	None	
Glf	Games played in left field	Min: 0 Max: 156 Unique Values: 157 Type: float64	Missing Values: 37	
Gcf	Games played in center field	Min: 0 Max: 162 Unique Values: 159 Type: float64	Missing Values: 37	
Grf	Games played in right field	Min: 0 Max: 160 Unique Values: 159 Type: float64	Missing Values: 43	
Table Name/ Description		lata: information about positions ations on the following 8 variable	•	
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
playerID	Player ID code	Unique Values: 3529 Type: Object	None	Yes/No
yearID	Year	Min: 1954 Max: 2019 Unique Values: 66 Type: int64	None	
stint	player's stint (order of appearances within a season)	Min: 1 Max: 4 Unique Values: 4 Type: int64	None	
teamID	Team; a factor	Unique Values: 149 Type: Object	None	

lgID	League; a factor with levels	Unique Values: 6	Missing Values:	
o .	AA AL FL NL PL UA	Type: Object	1513	
POS	Position	Unique Values: 3		
	1	Type: Object		
G	Games	Type: int64	None	
GS	Games Started	Type: int64	None	
InnOuts	Time played in the field	Type: int64	None	
	expressed as outs	7,5		
РО	Putouts	Type: int64	None	
A	Assists	Type: int64	None	
E	Errors	Type: int64	None	
DP	Double Plays	Type: int64	None	
РВ	Passed Balls (by catchers)	Type: float64	Missing Values: 33279	
WP	Wild Pitches (by catchers)	Type: float64	Missing Values: 33279	
SB	Opponent Stolen Bases (by catchers)	Type: float64	Missing Values: 33279	
CS	Opponents Caught Stealing (by catchers)	Type: float64	Missing Values: 33279	
ZR	Zone Rating	Type: float64	Missing Values: 33279	
Description	nominated for the Baseball Hal	. This is composed of the voting r I of Fame.		.5
Variable Name	nominated for the Baseball Hal	<del>_</del>		Used in
	nominated for the Baseball Hal A data frame with 4191 observ	l of Fame. ations on the following 9 variable	25	Used in final
Variable Name	nominated for the Baseball Hal A data frame with 4191 observ <b>Description</b>	l of Fame. ations on the following 9 variable Statistical Summary/Type	Errors	Used in final dataset?
	nominated for the Baseball Hal A data frame with 4191 observ	I of Fame. ations on the following 9 variable Statistical Summary/Type Unique Values: 1279	25	Used in final
<b>Variable Name</b> playerID	nominated for the Baseball Hal A data frame with 4191 observ Description  Player ID code	l of Fame. ations on the following 9 variable Statistical Summary/Type	Errors	Used in final dataset?
<b>Variable Name</b> playerID	nominated for the Baseball Hal A data frame with 4191 observ <b>Description</b>	I of Fame. ations on the following 9 variable Statistical Summary/Type  Unique Values: 1279 Type: Object Min: 1933	Errors  None	Used in final dataset?
<b>Variable Name</b> playerID	nominated for the Baseball Hal A data frame with 4191 observ Description  Player ID code	I of Fame. ations on the following 9 variable Statistical Summary/Type  Unique Values: 1279 Type: Object Min: 1933 Max: 2019	Errors  None	Used in final dataset?
<b>Variable Name</b> playerID	nominated for the Baseball Hal A data frame with 4191 observ Description  Player ID code	I of Fame. ations on the following 9 variable Statistical Summary/Type  Unique Values: 1279 Type: Object Min: 1933 Max: 2019 Unique Values: 80	Errors  None	Used in final dataset?
Variable Name playerID yearID	nominated for the Baseball Hall A data frame with 4191 observed Description  Player ID code  Year of ballot	I of Fame. ations on the following 9 variable Statistical Summary/Type  Unique Values: 1279 Type: Object Min: 1933 Max: 2019 Unique Values: 80 Type: float64	Errors  None	Used in final dataset?
Variable Name	nominated for the Baseball Hall A data frame with 4191 observed Description  Player ID code  Year of ballot  Method by which player was	I of Fame. ations on the following 9 variable Statistical Summary/Type  Unique Values: 1279 Type: Object Min: 1933 Max: 2019 Unique Values: 80 Type: float64 Unique Values: 9	Errors  None  None	Used in final dataset?
playerID yearID votedBy	nominated for the Baseball Hall A data frame with 4191 observed Description  Player ID code  Year of ballot	I of Fame. ations on the following 9 variable Statistical Summary/Type  Unique Values: 1279 Type: Object Min: 1933 Max: 2019 Unique Values: 80 Type: float64	Errors  None  None	Used in final dataset?
Variable Name playerID yearID	nominated for the Baseball Hall A data frame with 4191 observed Description  Player ID code  Year of ballot  Method by which player was voted upon	I of Fame. ations on the following 9 variable Statistical Summary/Type  Unique Values: 1279 Type: Object Min: 1933 Max: 2019 Unique Values: 80 Type: float64 Unique Values: 9 Type: Object	Errors  None  None  None  Missing Values:	Used in final dataset?
Variable Name  playerID  yearID  votedBy  ballots	nominated for the Baseball Hall A data frame with 4191 observed Description  Player ID code  Year of ballot  Method by which player was voted upon Total ballots cast in that year  Number of votes needed for	I of Fame. ations on the following 9 variable Statistical Summary/Type  Unique Values: 1279 Type: Object Min: 1933 Max: 2019 Unique Values: 80 Type: float64 Unique Values: 9 Type: Object Unique Values: 74	None  None  None  None  Missing Values: 197  Missing Values:	Used in final dataset?
Variable Name  playerID  yearID  votedBy  ballots  needed	nominated for the Baseball Hall A data frame with 4191 observed Description  Player ID code  Year of ballot  Method by which player was voted upon  Total ballots cast in that year  Number of votes needed for selection in that year	I of Fame. ations on the following 9 variable Statistical Summary/Type  Unique Values: 1279 Type: Object Min: 1933 Max: 2019 Unique Values: 80 Type: float64 Unique Values: 9 Type: Object Unique Values: 74  Unique Values: 74	None  None  None  Missing Values: 197  Missing Values: 354  Missing Values:	Used in final dataset?
votedBy ballots needed votes inducted	nominated for the Baseball Hall A data frame with 4191 observed Description  Player ID code  Year of ballot  Method by which player was voted upon  Total ballots cast in that year  Number of votes needed for selection in that year  Total votes received  Whether player was inducted by that vote or not (Y or N)  Category of candidate; a factor with levels Manager Pioneer/Executive Player	I of Fame. ations on the following 9 variable Statistical Summary/Type  Unique Values: 1279 Type: Object Min: 1933 Max: 2019 Unique Values: 80 Type: float64 Unique Values: 9 Type: Object Unique Values: 74  Unique Values: 65  Unique Values: 367  Unique Values: 2	None  None  None  Missing Values: 197  Missing Values: 354  Missing Values: 197	Used in final dataset?
Variable Name  playerID  yearID  votedBy  ballots  needed  votes	nominated for the Baseball Hall A data frame with 4191 observed Description  Player ID code  Year of ballot  Method by which player was voted upon  Total ballots cast in that year  Number of votes needed for selection in that year  Total votes received  Whether player was inducted by that vote or not (Y or N)  Category of candidate; a factor with levels Manager	I of Fame. ations on the following 9 variable Statistical Summary/Type  Unique Values: 1279 Type: Object Min: 1933 Max: 2019 Unique Values: 80 Type: float64 Unique Values: 9 Type: Object Unique Values: 74  Unique Values: 65  Unique Values: 367  Unique Values: 2 Type: Object Unique Values: 4	None  None  None  Missing Values: 197  Missing Values: 354  Missing Values: 197  None	Used in final dataset?

Table Name/	HomeGames					
Description	Description: Data mapping teams to the stadiums they played regular season games in as the					
Description	home team;					
	A data frame with 3108 observations on the following 9 variables					
Variable Name	Description Description	Statistical Summary/Type	Errors	Used in final dataset?		
year .key	Year	Min-Max: 1871 - 2019 Unique Values: 149 Type: int64	None	Yes/No		
league .key	League; a factor with levels AA AL FL NL PL UA	Unique Values: 6 Type: object	Missing Values: 77			
team .key	Team; a factor	Unique Values: 148 Type: object	None			
park .key	Unique identifier for each ballpark	Unique Values: 254 Type: object	None			
span .first	First date the park began acting as home field for the team	Unique Values: 1099 Type: object	Inconsistencies in time/date formats			
span .last	Last date the park began acting as home field for the team	Unique Values: 1137 Type: object	Inconsistencies in time/date formats			
games	Total games in this time span	Min-Max: 1 - 89 Unique Values: 87 Type: int64	None			
openings	openings Total opening in this time span	Min-Max: 0 - 83 Unique Values: 84 Type: int64	None			
attendance	Total attendance in this time span	Unique Values: 2665 Type: int64	None			
Table Name/ Description	and some basic statistics for the	formation about individual team nose teams in each year ations on the following 10 variable		managed		
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?		
playerID	Manager ID code	Unique Values: 714 Type: Object	None	Yes/No		
yearID	Year	Min: 1871 Max: 2019 Unique Values: 149 Type: float64	Missing Values: 1			
teamID	Team; a factor	Unique Values: 44 Type: Object	None			
lgID	League; a factor with levels AA AL FL NL PL UA	Unique Values: 6 Type: Object	Missing Values: 67			
inseason	Managerial order. Zero if the individual managed the team the entire year. Otherwise denotes where the manager appeared in the managerial	Unique Values: 9 Type: int64	None			

- 165 es: 165  - 116	
es: 165  -116     None es: 114  -120     None es: 119  None es: 12     None es: 2  None	
es: 165  -116     None es: 114  -120     None es: 119  None es: 12     None es: 2  None	
-116 None -120 None -120 None -1219 None -120 None -120 None -120 None -120 None	
es: 114  - 120 es: 119  Pes: 12 None es: 2 None  Dowing 10 variables	
es: 114  - 120 es: 119  Pes: 12 None es: 2 None  Dowing 10 variables	
None es: 119 None es: 2 None owing 10 variables	
es: 119  es: 12  None  es: 2  None  owing 10 variables	
es: 119  es: 12  None  es: 2  None  owing 10 variables	
es: 12 None es: 2 None owing 10 variables	
es: 2 None  owing 10 variables	
es: 2 None  owing 10 variables	
owing 10 variables	
owing 10 variables	
mmary/Type Frrors	
	Used in
" "	final
	dataset?
es: 54 None	Yes/No
	,
ues : 1892 - 1981 None	
es: 2	
es: 33 None	
es: 2 None	
es: 2 None	
- 165 None	
es: 40	
- 116 None	
- 120 None	
es: 12 None	1
: <u>e</u> : e	es: 2

Table Name/ Description	Parks Description: Name and location A data frame with 255 observat	data for baseball stadiums ions on the following 6 variables		
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
park .key	unique identifier for each ballpark	Unique Values: 255 Type: Object	None	Yes/No
park .name	the name of the ballpark	Unique Values: 243 Type: Object	None	
park .alias	a semicolon delimited list of other names for the ballpark if they exist	Unique Values: 57 Type: Object	Missing Values: 196	
city	city where the ballpark is located	Unique Values: 87 Type: Object	None	
state	state where the ballpark is located	Unique Values: 37 Type: Object	Missing Values: 1	
country	country where the ballpark is located	Unique Values: 7 Type: Object	None	
Table Name/ Description	details about players listed in th by playerID	er names, DOB, and biographical in Batting, Pitching, and other file wations on the following 26 variab	s where players are ide	_
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
playerID	A unique code assigned to each player. The playerID links the data in this file with records on players in the other files	Unique Values: 20087 Type: Object	None	Yes/No
birthYear birthMonth birthDay	Year, Month, and Day player was born	Type: float64 Unique Values: 169 Unique Values: 12 Unique Values: 31	Missing Values: 114 252 424	
birthCountry birthState birthCity	Country, State, and City where player was born	Type: Object Unique Values: 57 Unique Values: 296 Unique Values: 4883	Missing Values: 61 555 172	
deathYear deathMonth deathDay	Year, Month, and Day player died	Type: float64 Unique Values: 149 Unique Values: 12 Unique Values: 31	Missing Values: 10253 10254 10255	
deathCountry deathState deathCity	Country, State, and City where player died	Type: Object Unique Values: 25 Unique Values: 107 Unique Values: 2669	Missing Values: 10255 10305 10260	
nameFirst	Player's first name	Type: Object Unique Values: 2529	Missing Values: 37	

nameLast	Player's last name	Unique Values: 10237	None	
nameGiven	Player's given name (typically	Unique Values: 13337	37	
	first and middle)			
weight	Player's weight in pounds	Unique Values: 153	Missing Values: 817	
height	Player's height in inches	Unique Values: 23	Missing Values: 737	
bats	a factor: Player's batting hand	Unique Values: 3	Missing Values:	
	(left (L), right (R), or both (B))	Type: Object	1180	
throws	a factor: Player's throwing hand (left(L) or right(R))	Unique Values: 3 Type: Object	Missing Values: 977	
debut	Date that player made first	Unique Values: 10572	Missing Values:	
uebut	major league appearance	Type: Object	195	
finalGame	Date that player made first	Unique Values: 9479	Missing Values:	
illialGaille	major league appearance (blank if still active)	Type: Object	195	
retroID	ID used by retrosheet	Unique Values: 20030	Missing Values:	
	12 4304 27 10110311001	Type: Object	57	
bbrefID	ID used by Baseball Reference	Unique Values: 20084	Missing Values: 3	
	website	Type: Object		
birthDate	Player's birthdate, in as Date	Type: date/time	Inconsistencies	
	-		in date formats	
	format			
deathDate		Type: date/time		
deathDate  Table Name/	Player's death date, in as Date format  Pitching	Type: date/time	Inconsistencies in date formats	
	Player's death date, in as Date format  Pitching Description: Pitching table	Type: date/time  vations on the following 30 varial	Inconsistencies in date formats	
Table Name/	Player's death date, in as Date format  Pitching Description: Pitching table		Inconsistencies in date formats	Used in final dataset?
Table Name/ Description  Variable Name	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description	vations on the following 30 varial  Statistical Summary/Type	Inconsistencies in date formats  bles  Errors	final dataset?
Table Name/ Description  Variable Name	Player's death date, in as Date format  Pitching  Description: Pitching table  A data frame with 47628 obser	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845	Inconsistencies in date formats	final
Table Name/ Description  Variable Name  playerID	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object	lnconsistencies in date formats  bles  Errors  None	final dataset?
Table Name/ Description	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object Unique Values: 149	Inconsistencies in date formats  bles  Errors	final dataset?
Table Name/ Description  Variable Name  playerID  yearID	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code  Year	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object  Unique Values: 149 Type: int64	Inconsistencies in date formats  bles  Errors  None  Missing Values: 1	final dataset?
Table Name/ Description  Variable Name  playerID	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code  Year  player's stint (order of	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object Unique Values: 149	lnconsistencies in date formats  bles  Errors  None	final dataset?
Table Name/ Description  Variable Name  playerID  yearID	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code  Year	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object  Unique Values: 149 Type: int64  Unique Values: 5	Inconsistencies in date formats  bles  Errors  None  Missing Values: 1	final dataset?
Table Name/ Description  Variable Name  playerID  yearID  stint	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code  Year  player's stint (order of appearances within a season)	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object  Unique Values: 149 Type: int64  Unique Values: 5 Type: int64	Inconsistencies in date formats  bles  Errors  None  Missing Values: 1  None	final dataset?
Table Name/ Description  Variable Name  playerID  yearID  stint  teamID	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code  Year  player's stint (order of appearances within a season)	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object  Unique Values: 149 Type: int64  Unique Values: 5 Type: int64  Unique Values: 149	Inconsistencies in date formats  bles  Errors  None  Missing Values: 1  None  None	final dataset?
Table Name/ Description  Variable Name  playerID  yearID  stint	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code  Year  player's stint (order of appearances within a season) Team; a factor	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object  Unique Values: 149 Type: int64  Unique Values: 5 Type: int64  Unique Values: 149 Type: Object	Inconsistencies in date formats  bles  Errors  None  Missing Values: 1  None	final dataset?
Table Name/ Description  Variable Name  playerID  yearID  stint  teamID	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code  Year  player's stint (order of appearances within a season) Team; a factor  League; a factor with levels	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object  Unique Values: 149 Type: int64  Unique Values: 5 Type: int64  Unique Values: 149 Type: Object  Unique Values: 16	Inconsistencies in date formats  bles  Errors  None  Missing Values: 1  None  None  Missing Values:	final dataset?
Table Name/ Description  Variable Name  playerID  yearID  stint  teamID  lgID	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code  Year  player's stint (order of appearances within a season) Team; a factor  League; a factor with levels AA AL FL NL PL UA	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object  Unique Values: 149 Type: int64  Unique Values: 5 Type: int64  Unique Values: 149 Type: Object  Unique Values: 6 Type: Object	Inconsistencies in date formats  bles  Errors  None  Missing Values: 1  None  None  Missing Values: 132	final dataset?
Table Name/ Description  Variable Name  playerID  yearID  stint  teamID  lgID	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code  Year  player's stint (order of appearances within a season) Team; a factor  League; a factor with levels AA AL FL NL PL UA	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object  Unique Values: 149 Type: int64  Unique Values: 5 Type: int64  Unique Values: 149 Type: Object  Unique Values: 6 Type: Object  Unique Values: 54 Type: int64  Unique Values: 54 Type: int64  Unique Values: 54 Type: int64	Inconsistencies in date formats  bles  Errors  None  Missing Values: 1  None  None  Missing Values: 132	final dataset?
Table Name/ Description  Variable Name  playerID  yearID  stint  teamID  IgID  W	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code  Year  player's stint (order of appearances within a season) Team; a factor  League; a factor with levels AA AL FL NL PL UA  Wins	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object  Unique Values: 149 Type: int64  Unique Values: 5 Type: object  Unique Values: 6 Type: Object  Unique Values: 54 Type: int64  Unique Values: 94	Inconsistencies in date formats  bles  Errors  None  Missing Values: 1  None  None  Missing Values: 132  None	final dataset?
Table Name/ Description  Variable Name  playerID  yearID  stint  teamID  IgID  W	Player's death date, in as Date format  Pitching Description: Pitching table A data frame with 47628 obser  Description  Player ID code  Year  player's stint (order of appearances within a season) Team; a factor  League; a factor with levels AA AL FL NL PL UA  Wins  Losses	vations on the following 30 varial  Statistical Summary/Type  Unique Values: 9845 Type: Object  Unique Values: 149 Type: int64  Unique Values: 5 Type: int64  Unique Values: 149 Type: Object  Unique Values: 6 Type: Object  Unique Values: 54 Type: int64  Unique Values: 54 Type: int64  Unique Values: 43 Type: int64	Inconsistencies in date formats  bles  Frrors  None  Missing Values: 1  None  None  Missing Values: 132  None  None  None	final dataset?

SHO	Shutouts	Unique Values: 15	None	
		Type: int64		
SV	Saves	Unique Values: 57	None	
		Type: int64		
<b>IPouts</b>	Outs Pitched (innings pitched	Unique Values: 1311	None	
	x 3)	Type: int64		
Н	Hits	Unique Values: 534	None	
		Type: int64		
ER	Earned Runs	Unique Values: 216	None	
		Type: int64		
HR	Homeruns	Unique Values: 48	None	
		Type: int64		
BB	Walks	Unique Values: 209	None	
		Type: int64		
SO	Strikeout	Unique Values: 338	None	
		Type: int64		
ВАОрр	Opponent's Batting Average	Unique Values: 448	Missing Values:	
		Type: float64	4441	
ERA	Earned Run Average	Unique Values: 1168	Missing Values:	
	_	Type: float64	94	
IBB	Intentional Walks	Unique Values: 22	Missing Values:	
		Type: float64	14578	
WP	Wild Pitches	Unique Values: 61	None	
		Type: int64		
НВР	Batters Hit By Pitch	Unique Values: 42	Missing Values:	
	,	Type: float64	734	
BK	Balks	Unique Values: 16	None	
		Type: int64		
BFP	Batters faced by Pitcher	Unique Values: 1736	None	
	,	Type: float64		
GF	Games Finished	Unique Values: 78	None	
		Type: int64		
R	Runs Allowed	Unique Values: 345	None	
		•		
	1	Type: int64		
SH	Sacrifices by opposing batters	Type: int64 Unique Values: 25	Missing Values:	
SH	Sacrifices by opposing batters	Unique Values: 25	Missing Values: 19187	
SH SF		Unique Values: 25 Type: float64	19187	
	Sacrifice flies by opposing	Unique Values: 25 Type: float64 Unique Values: 18	19187 Missing Values:	
SF	Sacrifice flies by opposing batters	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64	19187 Missing Values: 19187	
	Sacrifice flies by opposing batters Grounded into double plays	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64 Unique Values: 44	19187 Missing Values: 19187 Missing Values:	
SF GIDP	Sacrifice flies by opposing batters Grounded into double plays by opposing batter	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64	19187 Missing Values: 19187	
SF GIDP Table Name/	Sacrifice flies by opposing batters Grounded into double plays by opposing batter Pitching Post	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64 Unique Values: 44 Type: float64	19187 Missing Values: 19187 Missing Values:	
SF GIDP	Sacrifice flies by opposing batters Grounded into double plays by opposing batter Pitching Post Description: Post season pitching	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64 Unique Values: 44 Type: float64	19187 Missing Values: 19187 Missing Values: 23018	
SF GIDP Table Name/ Description	Sacrifice flies by opposing batters Grounded into double plays by opposing batter Pitching Post Description: Post season pitching A data frame with 5798 observ	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64 Unique Values: 44 Type: float64  Ing statistics ations on the following 30 variables	19187 Missing Values: 19187 Missing Values: 23018	
SF GIDP Table Name/	Sacrifice flies by opposing batters Grounded into double plays by opposing batter Pitching Post Description: Post season pitching	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64 Unique Values: 44 Type: float64	19187 Missing Values: 19187 Missing Values: 23018	Used in
SF GIDP Table Name/ Description	Sacrifice flies by opposing batters Grounded into double plays by opposing batter Pitching Post Description: Post season pitching A data frame with 5798 observ	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64 Unique Values: 44 Type: float64  Ing statistics ations on the following 30 variables	19187 Missing Values: 19187 Missing Values: 23018	final
SF GIDP Table Name/ Description Variable Name	Sacrifice flies by opposing batters Grounded into double plays by opposing batter Pitching Post Description: Post season pitching A data frame with 5798 observ  Description	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64 Unique Values: 44 Type: float64  Ing statistics ations on the following 30 variables  Statistical Summary/Type	19187 Missing Values: 19187 Missing Values: 23018  Errors	final dataset?
SF GIDP Table Name/ Description	Sacrifice flies by opposing batters Grounded into double plays by opposing batter Pitching Post Description: Post season pitching A data frame with 5798 observ	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64 Unique Values: 44 Type: float64  Ing statistics ations on the following 30 variables  Statistical Summary/Type  Unique Values: 1829	19187 Missing Values: 19187 Missing Values: 23018	final
SF GIDP Table Name/ Description Variable Name	Sacrifice flies by opposing batters Grounded into double plays by opposing batter Pitching Post Description: Post season pitching A data frame with 5798 observ  Description  Player ID code	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64 Unique Values: 44 Type: float64  Ing statistics ations on the following 30 variables  Statistical Summary/Type  Unique Values: 1829 Type: Object	19187 Missing Values: 19187 Missing Values: 23018  Errors  None	final dataset?
SF GIDP Table Name/ Description Variable Name	Sacrifice flies by opposing batters Grounded into double plays by opposing batter Pitching Post Description: Post season pitching A data frame with 5798 observ  Description	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64 Unique Values: 44 Type: float64  Ing statistics ations on the following 30 variables  Statistical Summary/Type  Unique Values: 1829 Type: Object Unique Values: 123	19187 Missing Values: 19187 Missing Values: 23018  Errors	final dataset?
SF GIDP Table Name/ Description Variable Name	Sacrifice flies by opposing batters Grounded into double plays by opposing batter Pitching Post Description: Post season pitching A data frame with 5798 observ  Description  Player ID code	Unique Values: 25 Type: float64 Unique Values: 18 Type: float64 Unique Values: 44 Type: float64  Ing statistics ations on the following 30 variables  Statistical Summary/Type  Unique Values: 1829 Type: Object	19187 Missing Values: 19187 Missing Values: 23018  Errors  None	final dataset?

		Type: int64	
teamID	Team; a factor	Unique Values: 48	None
		Type: Object	
lgID	League; a factor with levels	Unique Values: 3	None
	AA AL FL NL PL UA	Type: Object	
W	Wins	Unique Values: 5	None
		Type: int64	
L	Losses	Unique Values: 5	None
		Type: int64	
G	Games	Unique Values: 8	None
		Type: int64	
GS	Games Started	Unique Values: 8	None
		Type: int64	
CG	Complete Games	Unique Values: 8	None
	·	Type: int64	
SHO	Shutouts	Unique Values: 4	None
		Type: int64	
SV	Saves	Unique Values: 5	None
		Type: int64	
IPouts	Outs Pitched (innings pitched	Unique Values: 90	None
	x 3)	Type: int64	1.0.00
Н	Hits	Unique Values: 39	None
		Type: int64	1.10.10
ER	Earned Runs	Unique Values: 20	None
	Edified Rails	Type: int64	Tronc
HR	Homeruns	Unique Values: 6	None
	Tromerans	Type: int64	Tronc
BB	Walks	Unique Values: 19	None
55	Walks	Type: int64	None
SO	Strikeout	Unique Values: 31	None
30	Strikeout	Type: int64	None
ВАОрр	Opponent's Batting Average	Unique Values: 306	Missing Values:
влорр	opponent 3 batting Average	Type: float64	71
ERA	Earned Run Average	Unique Values: 334	Missing Values:
LIVA	Larried Kurr Average	Type: float64	33
IBB	Intentional Walks	Unique Values: 5	Missing Values:
IDD	intentional warks	Type: float64	50
WP	Wild Pitches	Unique Values: 6	Missing Values:
VVF	Wild Fitches	Type: int64	50
LIDD	Dattors Hit Dy Ditah	Unique Values: 5	
HBP	Batters Hit By Pitch	· · · · · ·	Missing Values:
DI	D-III-	Type: float64	50
ВК	Balks	Unique Values: 2	Missing Values:
DED	Datta and for a dilate Bit also as	Type: int64	50
BFP	Batters faced by Pitcher	Unique Values: 109	Missing Values:
C.F.	Compa Finish	Type: float64	50
GF	Games Finished	Unique Values: 7	None
_		Type: int64	<u> </u>
R	Runs Allowed	Unique Values: 28	None
		Type: int64	
SH	Sacrifices by opposing batters	Unique Values: 8	Missing Values:
		Type: float64	50

SF	Sacrifice flies by opposing	Unique Values: 4	Missing Values:	1
31	batters	Type: float64	50	
GIDP	Grounded into double plays	Unique Values: 7	Missing Values:	
	by opposing batter	Type: float64	50	
Table Name/ Description	Salaries Description: Player salary data	a ervations on the following 5 variables	les	
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
yearID	Year	Min-Max: 1985 – 2016 Unique Values: 32 Type: int64	None	Yes/No
teamID	Team; a factor	Unique Values: 35 Type: Object	None	
lgID	League; a factor	Unique Values: 2	None	
		Type: Object		
playerID	Player ID code	Unique Values: 5149 Type: Object	None	
salary	Salary	Min-Max: 0 - 33000000 Unique Values: 3393 Type: int64	A few values with 0's; considered as missing values	
Table Name/	Schools	1 - 1/2	1 8	
Description		chools players attended, by school		
		vations on the following 5 variable		
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
schoolID	school ID code	Unique Values: 1207 Type: Object	None	Yes/No
name_full	school name	Unique Values: 1199 Type: Object	None	
city	city where school is located	Unique Values: 856 Type: Object	None	
state	state where school's city is located	Unique Values: 49 Type: Object	None	
country	country where school is located	Unique Values: 1 Type: Object	None	
Table Name/ Description	SeriesPost Description: Post season serie A data frame with 343 observa	es information ations on the following 9 variables		
Variable Name	Description	Statistical Summary/Type	Errors	Used in final dataset?
			1	
yearID	Year	Unique Values: 123 Type: Object	None	Yes/No
yearID	Year  Level of playoffs		None	Yes/No

teamIDwinner	Team ID of the team that	Unique Values: 42	None	
teamb winter	won the series; a factor	Type: Object	Tronc	
lgIDwinner	League ID of the team that	Unique Values: 3	None	
.8.2	won the series; a factor with	Type: Object	1.0	
	levels AL NL	.,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		
teamIDloser	Team ID of the team that lost	Unique Values: 45	None	
	the series; a factor	Type: Object	1.00.00	
lgIDloser	League ID of the team that	Unique Values: 3	None	
.8	lost the series; a factor with	Type: Object	1.00.00	
	levels AL NL	7,1-1-3,1-1		
wins	Wins by team that won the	Unique Values: 6	None	
	series	Type: int64		
losses	Losses by team that won the	Unique Values: 6	None	
	series	Type: int64		
ties	Tie games	Unique Values: 2	None	
		Type: int64		
Table Name/	Teams	71		
Description	Description: Yearly statistics an	d standings for teams		
	•	ations on the following 48 variable	es	
Variable Name	Description	Statistical Summary/Type	Errors	Used in
				final
10		1074 0040		dataset?
yearID	Year	Min-Max: 1871 – 2019	None	
		Unique Values: 149		
1.15		Type: int64	N. II. 50	
lgID	League; a factor with levels AA	Unique Values: 6	Nulls: 50	
. 15	AL FL NL PL UA	Type: Object		
teamID	Team; a factor	Unique Values: 149	None	
f 115		Type: Object		
franchID	Franchise (links to	Unique Values: 120	None	
1: 15	TeamsFranchises table)	Type: Object	N. II. 4547	
divID	Team's division; a factor with	Unique Values: 3	Nulls: 1517	
	levels C E W	Type: Object		
Rank	Position in final standings	Unique Values: 13	None	
	+-	Type: int64		
G	Games	Unique Values: 13	None	
	<u> </u>	Type: int64		
Ghome	Games played at home	Unique Values: 38	Nulls: 399	
***	1,1,0	Type: float64		
W	Wins	Unique Values: 112	None	
	<u> </u>	Type: int64	1	
L	Losses	Unique Values: 114	None	
		Type: int64		
DivWin	Division Winner (Y or N)	Unique Values: 2	Nulls: 1545	
		Type: Object		
WCWin	Wild Card Winner (Y or N)	Unique Values: 2	Nulls: 2181	
		Type: Object		
LgWin	League Champion(Y or N)	Unique Values: 2	Nulls: 28	
		Type: Object		
WSWin	World Series Winner (Y or N)	Unique Values: 2	Nulls: 357	
		Type: Object		1

R	Runs scored	Unique Values: 618	None	
		Type: int64		
AB	At bats	Unique Values: 1106	None	
		Type: int64		
Н	Hits by batters	Unique Values: 732	None	
25		Type: int64	<del></del>	
2B	Doubles	Unique Values: 309	None	
		Type: int64		
3B	Triples	Unique Values: 126	None	
		Type: int64		
HR	Homeruns by batters	Unique Values: 259	None	
		Type: int64		
BB	Walks by batters	Unique Values: 572	Nulls: 1	
		Type: float64		
SO	Strikeouts by batters	Unique Values: 1104	Nulls: 16	
		Type: float64		
SB	Stolen bases	Unique Values: 322	Nulls: 126	
		Type: float64		
CS	Caught stealing	Unique Values: 137	Nulls: 832	
		Type: float64		
НВР	Batters hit by pitch	Unique Values: 97	Nulls: 1158	
	μ	Type: int64		
SF	Sacrifice flies	Unique Values: 56	Nulls: 1541	
<b>5</b> .	Sucrimee mes	Type: float64	1141131 13 11	
RA	Opponents runs scored	Unique Values: 604	None	
IVA	Opponents runs scoreu	Type: int64	None	
ER	Earned runs allowed	Unique Values: 640	None	
LIX	Lamed runs anowed	Type: int64	None	
ERA	Earned run average	Unique Values: 392	None	
LIVA	Lamed full average	Type: float64	None	
CG	Complete games	Unique Values: 147	None	
CG	Complete games	Type: int64	None	
SHO	Chutoute	Unique Values: 32	None	
SHO	Shutouts	•	None	
6) /	Carra	Type: int64	News	
SV	Saves	Unique Values: 66	None	
	0 . 5:: 1 . 1/:	Type: int64	<del></del>	
IPouts	Outs Pitched (innings pitched x	Unique Values: 830	None	
	3)	Type: int64		
НА	Hits allowed	Unique Values: 744	None	
		Type: int64		
HRA	Homeruns allowed	Unique Values: 244	None	
		Type: int64		
BBA	Walks allowed	Unique Values: 563	None	
		Type: int64		
SOA	Strikeouts by pitchers	Unique Values: 1130	None	
		Type: int64		
E	Errors	Unique Values: 455	None	
		Type: int64		
DP	Double Plays	Unique Values: 199	None	
		Type: int64		
FP	Fielding percentage	Unique Values: 166	None	
	<del>-</del>	•	<u> </u>	

		Type: float64		
name	Team's full name	Unique Values: 139	None	
		Type: object		
park	Name of team's home ballpark	Unique Values: 217	Nulls: 34	
	'	Type: object		
attendance	Home attendance total	Unique Values: 2638	Nulls: 279	
		Type: float64		
BPF	Three-year park factor for	Unique Values: 44	None	
	batters	Type: int64		
PPF	Three-year park factor for	Unique Values: 42	None	
	pitchers	Type: int64		
teamIDBR	Team ID used by Baseball	Unique Values: 101	None	
	Reference website	Type: object		
teamIDlahman45	Team ID used in Lahman	Unique Values: 148	None	
	database version 4.5	Type: object		
teamIDretro	Team ID used by Retrosheet	Unique Values: 151	None	
-	,	Type: object		
Table Name/	TeamFranchises			
Description	Description: Information about t	eam franchises		
	A data frame with 120 observation			
Variable Name		Statistical Summary/Type	Errors	Used in
variable ivame	Description	Statistical Summary/Type	Errors	final
				dataset?
franchID	Franchise ID; a factor	Unique Values: 120	None	Yes/No
ITATICITID	Franciise ib, a factor	Type: object	None	TES/NO
franchName	Franchise name	Unique Values: 99	None	
Iranchiname	Franchise name	Type: object	None	
active	Whether team is currently	Unique Values: 2	Nulls: 25	
active	active (Y or N)	Type: object	Nulls. 23	
NAassoc	ID of National Association	Unique Values: 12	Nulls: 108	
NAdSSUC	team franchise played as	Type: object	Nulls. 106	
Table Name/	TeamsHalf	Type. Object		
Description	Description: Split season data for	rtoams		
Description	A data frame with 52 observation			
	A data frame with 32 observatio			
Variable Name	Description	Statistical Summary/Type	Errors	Used in
				final
				dataset?
yearID	Year	Min-Max: 1981 - 1981	None	Yes/No
		Unique Values: 1		
		Type: int64		
lgID	League; a factor with levels AL	Unique Values: 2	None	
	NL	Type: object		
teamID	Team; a factor	Unique Values: 26	None	
		Type: object		
half	First or second half of season	Unique Values: 2	None	
		Type: int64		
divID	Division	Unique Values: 2	None	
		Type: object		
		1	1	
DivWin	Won Division (Y or N)	Unique Values: 1 Type: object	None	

Rank	Team's position in standings	Unique Values: 7	None
	for the half	Type: int64	
G	Games played	Unique Values: 13	None
		Type: int64	
W	Wins	Unique Values: 20	None
		Type: int64	
L	Losses	Unique Values: 18	None
		Type: int64	