**Phase 4: SQL + data visualizations**

In this phase, you will write SQL queries for the business questions that you created in Phase 3. Your SQL queries should be of varying complexity. You will be graded based on the complexity of queries and the insights you obtain from the queries. Please note that not all queries need to have visualizations.

1. **Which player hit the most HR for each team?**

TeamName	Name	HR
Atlanta Braves	Matt Olson	54
Philadelphia Phillies	Kyle Schwarber	47
New York Mets	Pete Alonso	46
Los Angeles Angels	Shohei Ohtani	44
Los Angeles Dodgers	Mookie Betts	39
Texas Rangers	Adolis García	39
Chicago White Sox	Luis Robert Jr.	38
New York Yankees	Aaron Judge	37
Miami Marlins	Jorge Soler	36
San Diego Padres	Juan Soto	35
Arizona Diamondbacks	Christian Walker	33
Boston Red Sox	Rafael Devers	33
Seattle Mariners	Julio Rodríguez	32
Detroit Tigers	Spencer Torkel...	31
Houston Astros	Yordan Alvarez	31
Tampa Bay Rays	Isaac Paredes	31
Kansas City Royals	Bobby Witt	30
Oakland Athletics	Brent Rooker	30
Baltimore Orioles	Gunnar Hende...	28
Baltimore Orioles	Anthony Sant...	28
Washington Nationals	Lane Thomas	28
St. Louis Cardinals	Nolan Gorman	27
Chicago Cubs	Christopher M...	26
Chicago Cubs	Cody Bellinger	26
Pittsburgh Pirates	Jack Suwinski	26
Toronto Blue Jays	Vladimir Guerr...	26
Cleveland Guardians	José Ramírez	24
Milwaukee Brewers	Willy Adames	24
Minnesota Twins	Max Kepler	24
Cincinnati Reds	Spencer Steer	23
Colorado Rockies	Ryan McMahon	23
San Francisco Giants	Wilmer Flores	23

Homerun leader from each team
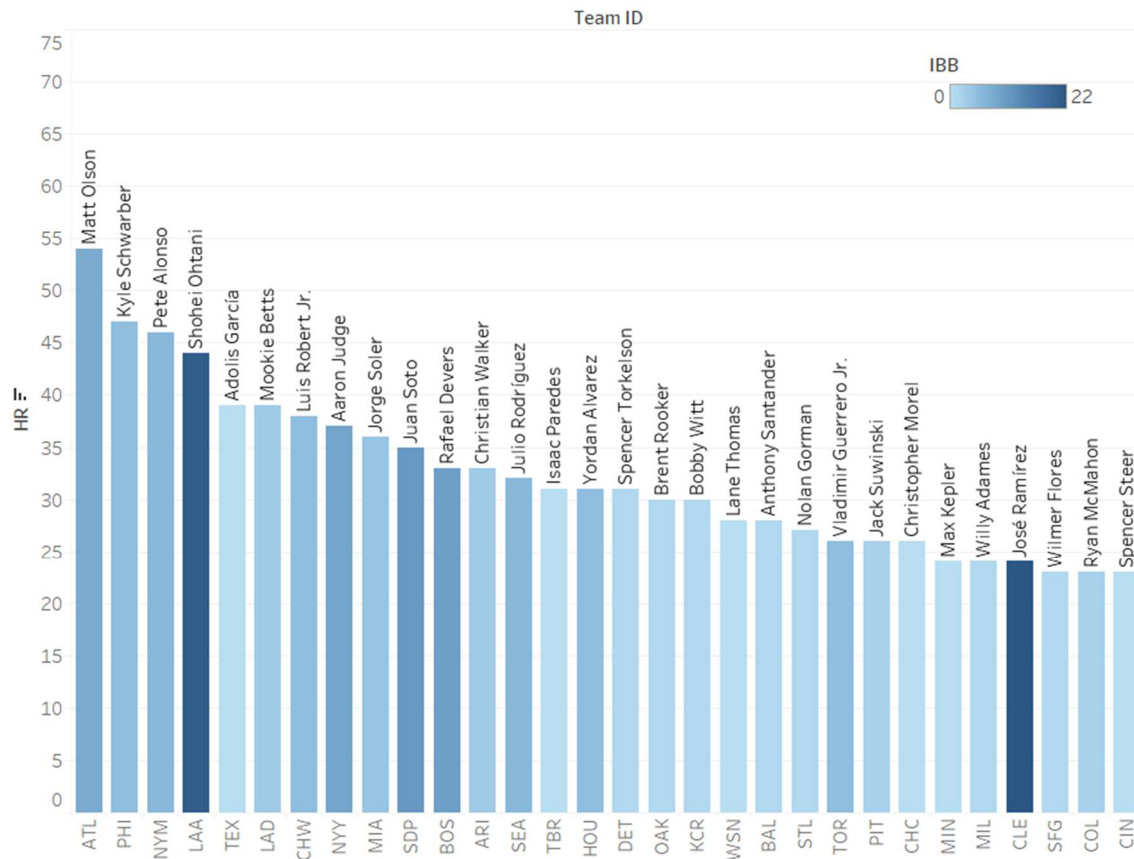
After knowing who the HR leader for each team is, I wanted to dive deeper and see how often these leaders are given free passes on base. In the graph, darker the blue, the more IBB the hitters received. We can see that HR leaders do not necessarily get more intentional base on balls (IBB), as there are many other factors need to be considered when giving out free passes.

## 2. Who is the most frequently traded player in the 2023 season?

playerID	Name	TeamCount	TradeCount
dejonpa01	Paul DeJong	3	2
marisja01	Jake Marisnick	3	2
tapiara01	Raimel Tapia	3	2
wynnsau01	Austin Wynns	3	2
giolilu01	Lucas Giolito	3	2
hearnta01	Taylor Hearn	3	2
leonedo01	Dominic Leone	3	2
lopezjo02	Jorge López	3	2
lopezre01	Reynaldo López	3	2
millety01	Tyson Miller	3	2
misiean01	Anthony Misiewicz	3	2
moorema02	Matt Moore	3	2
scottta02	Tayler Scott	3	2
weavelu01	Luke Weaver	3	2

I retrieved how many rows each player has in both batting and pitching tables. This number indicates how many teams each player has played for during the 2023 season. Subtract 1 from that number would be how many times this player has been traded during the season (of course this way would disregard the situation where a player did not play a single game before getting traded).

3. **Who has the highest hitting percentage in each age level? List out the leaders and the average of each age level.**

age	name	highestBA
19	Junior Caminero	0.235
20	Evan Carter	0.306
21	Noelvi Marte	0.316
22	Johan Rojas	0.302
23	Iván Herrera	0.297
24	Jeter Downs	0.4
25	Ronald Acuna Jr.	0.337
26	Cole Tucker	0.5
27	Alejo Lopez	0.5
28	Zack Collins	0.5
29	Seby Zavala	0.357
30	Michael Perez	0.5
31	Yandy Díaz	0.33
32	Jake Marisnick	0.4
33	Freddie Freeman	0.331
34	Charlie Culberson	1
35	Donovan Solano	0.282
36	Charlie Blackmon	0.279
37	Carlos Santana	0.249
38	Justin Turner	0.276
39	Yuli Gurriel	0.245
40	Miguel Cabrera	0.257
41	Adam Wainwright	0
42	Nelson Cruz	0.245

Batting Average Leader at each age
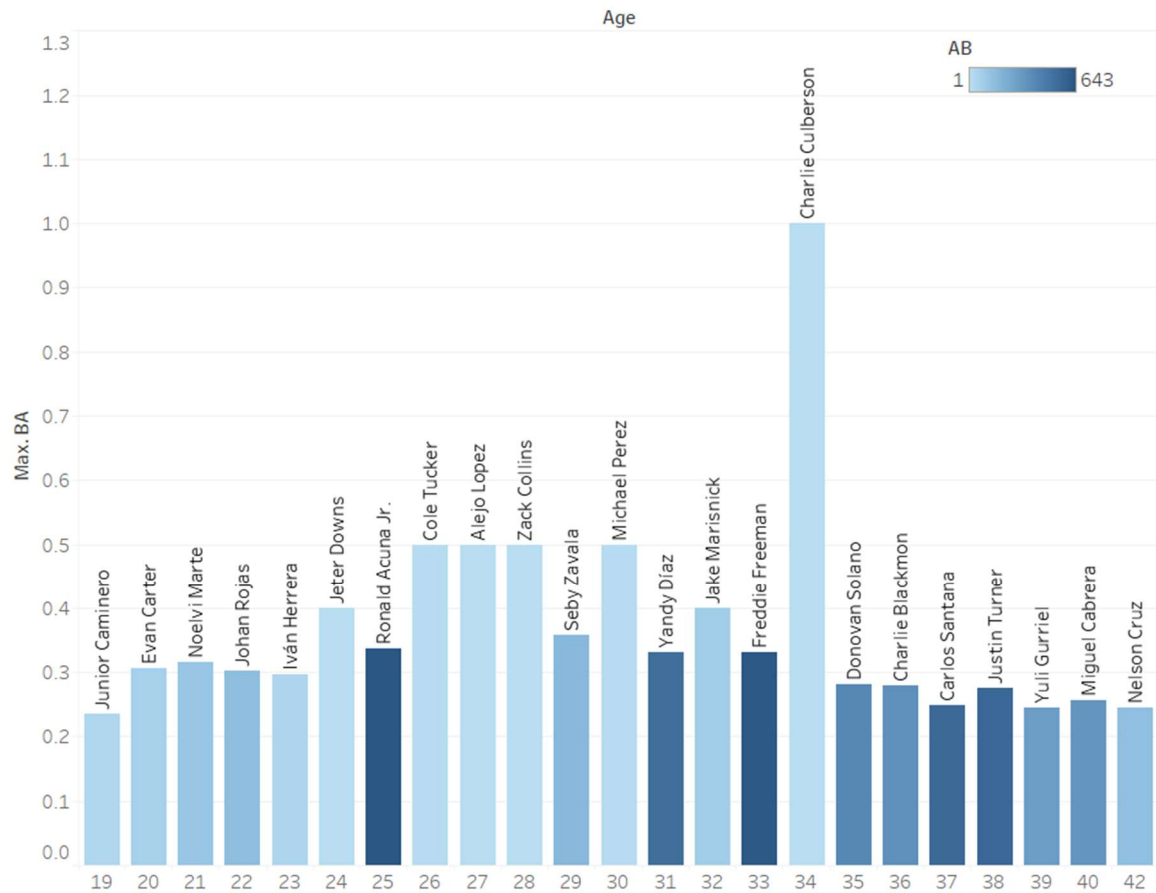
After retrieving the result, I want to see how many at-bats these players have. The darker the blue, the more at-bats the player has. We can see without filtering number of at-bats, batting average leaders can be someone who has only 1 AB, which is an extremely small sample size.

4. Does successful rate of challenges by the managers translate to wins? List out number of wins and calculate the rate.

Name	TeamID	WinRate	SuccessRate	SuccessRateRanking	WinRateRanking
Dave Martinez	WSN	0.4383	0.7586	1	25
Skip Schumaker	MIA	0.5185	0.7241	2	12
Torey Lovullo	ARI	0.5185	0.6667	3	12
Dusty Baker	HOU	0.5556	0.6429	4	6
Matt Quatraro	KCR	0.3457	0.6364	5	29
Aaron Boone	NYY	0.5062	0.6000	6	15
Gabe Kapler	SFG	0.4906	0.5652	7	18
Terry Francona	CLE	0.4691	0.5610	8	21
Brian Snitker	ATL	0.6420	0.5600	9	1
Alex Cora	BOS	0.4815	0.5263	10	19
Oliver Marmol	STL	0.4383	0.5217	11	25
Rob Thomson	PHI	0.5556	0.5000	12	6
Bud Black	COL	0.3642	0.5000	12	28
Dave Roberts	LAD	0.6173	0.4872	14	3
Mark Kotsay	OAK	0.3086	0.4872	14	30
Craig Counsell	MIL	0.5679	0.4828	16	5
Kevin Cash	TBR	0.6111	0.4727	17	4
Rocco Baldelli	MIN	0.5370	0.4706	18	11
Scott Servais	SEA	0.5432	0.4571	19	10
Bob Melvin	SDP	0.5062	0.4359	20	15
David Ross	CHC	0.5123	0.4348	21	14
Brandon Hyde	BAL	0.6235	0.4146	22	2
Derek Shelton	PIT	0.4691	0.4146	22	21
Pedro Grifol	CHW	0.3765	0.4000	24	27
David Bell	CIN	0.5062	0.3962	25	15
Bruce Bochy	TEX	0.5556	0.3793	26	6
Buck Showalter	NYM	0.4630	0.3704	27	23
John Schneider	TOR	0.5494	0.3462	28	9
A.J. Hinch	DET	0.4815	0.3243	29	19
Phil Nevin	LAA	0.4506	0.2143	30	24

From the data, we can see that success rate in challenges does not directly affect the probability of winning a game.

5. Does lower Batting Park Factor really tie with lower batting averages?

teamID	BattingParkFactor	BPF_low_Ranking	BA	BA_ranking	HR	HR_ranking
OAK	92	1	0.223	30	171	20
SDP	94	2	0.244	19	205	13
SEA	95	3	0.242	22	210	11
BAL	96	4	0.255	10	183	17
SFG	96	4	0.235	28	174	19
WSN	96	4	0.254	11	151	29
NYM	97	7	0.238	25	215	10
CLE	97	7	0.25	13	124	30
TBR	98	9	0.26	3	230	6
PHI	98	9	0.256	8	220	8
MIL	98	9	0.24	23	165	24
STL	99	12	0.25	13	209	12
TOR	99	12	0.256	8	188	16
ARI	99	12	0.25	13	166	22
PIT	99	12	0.239	24	159	28
HOU	100	16	0.259	4	222	7
MIN	101	17	0.243	21	233	3
MIA	101	17	0.259	4	166	22
KCR	101	17	0.244	19	163	26
LAD	102	20	0.257	7	249	2
NYY	102	20	0.227	29	219	9
DET	102	20	0.236	27	165	24
ATL	103	23	0.276	1	307	1
CIN	103	23	0.249	16	198	14
CHC	103	23	0.254	11	196	15
CHW	103	23	0.238	25	171	20
TEX	104	27	0.263	2	233	3
LAA	104	27	0.245	18	231	5
BOS	106	29	0.258	6	182	18
COL	113	30	0.249	16	163	26

This database is not enough to precisely answer this question as I am unable to distinguish home and away game statistics in this database. Therefore, from the data, there are no significant correlations between the Batting Park Factor and batting averages.

6. Do more fielding errors lead to more losses?

teamID	errors	error_rankings	WinRate	WinRate_rankings
ARI	56	1	0.5185	12
TEX	57	2	0.5556	6
MIN	66	3	0.5370	11
STL	67	4	0.4383	25
BAL	71	5	0.6235	2
TOR	71	5	0.5494	9
SEA	73	7	0.5432	10
SDP	73	7	0.5062	15
TBR	75	9	0.6111	4
LAD	76	10	0.6173	3
MIL	77	11	0.5679	5
KCR	80	12	0.3457	29
HOU	81	13	0.5556	6
ATL	82	14	0.6420	1
CLE	82	14	0.4691	21
COL	82	14	0.3642	28
NYM	88	17	0.4630	23
WSN	90	18	0.4383	25
CIN	91	19	0.5062	15
PIT	91	19	0.4691	21
PHI	92	21	0.5556	6
CHC	92	21	0.5123	14
LAA	95	23	0.4506	24
CHW	95	23	0.3765	27
NYY	96	25	0.5062	15
MIA	97	26	0.5185	12
DET	100	27	0.4815	19
BOS	102	28	0.4815	19
OAK	102	28	0.3086	30
SFG	117	30	0.4877	18

As we can see from the data, error rate does not have significant correlation with winning games. However, we do see most winningest teams have limited errors.

7. List out each team's Runs per game and Runs allowed per game. How does this margin impact the result of games?

teamID	RunScoresPerGame	RunsAgainstPerGame	PointDiff	Wins	Losses	WinRate	WinRate_rankings
ATL	5.85	4.42	1.43	104	58	0.6420	1
LAD	5.59	4.31	1.28	100	62	0.6173	3
TBR	5.31	4.1	1.21	99	63	0.6111	4
TEX	5.44	4.42	1.02	90	72	0.5556	6
BAL	4.98	4.19	0.79	101	61	0.6235	2
HOU	5.1	4.31	0.79	90	72	0.5556	6
MIN	4.8	4.07	0.73	87	75	0.5370	11
SDP	4.64	4	0.64	82	80	0.5062	15
SEA	4.68	4.07	0.61	88	74	0.5432	10
CHC	5.06	4.46	0.6	83	79	0.5123	14
MIL	4.49	3.99	0.5	92	70	0.5679	5
PHI	4.91	4.41	0.5	90	72	0.5556	6
TOR	4.6	4.14	0.46	89	73	0.5494	9
BOS	4.77	4.79	-0.02	78	84	0.4815	19
NYM	4.43	4.51	-0.08	75	87	0.4630	23
ARI	4.6	4.7	-0.1	84	78	0.5185	12
NYY	4.15	4.31	-0.16	82	80	0.5062	15
CLE	4.09	4.3	-0.21	76	86	0.4691	21
CIN	4.83	5.07	-0.24	82	80	0.5062	15
SFG	4.16	4.44	-0.28	79	83	0.4877	18
MIA	4.11	4.46	-0.35	84	78	0.5185	12
DET	4.08	4.57	-0.49	78	84	0.4815	19
LAA	4.56	5.12	-0.56	73	89	0.4506	24
PIT	4.27	4.88	-0.61	76	86	0.4691	21
STL	4.44	5.12	-0.68	71	91	0.4383	25
WSN	4.32	5.22	-0.9	71	91	0.4383	25
KCR	4.17	5.3	-1.13	56	106	0.3457	29
CHW	3.96	5.19	-1.23	61	101	0.3765	27
COL	4.45	5.91	-1.46	59	103	0.3642	28
OAK	3.61	5.7	-2.09	50	112	0.3086	30

In general, most teams do follow the pattern where the larger the point difference you have, the more games you win. However, quite interestingly, there are several teams in the middle who scored less than they gave up, but still winning more than losing.

8. **Does Runners Left on Base impact Runs scored per game? List these two columns out for each team.**

teamID	RunScoresPerGame	LOB	score_rankings
STL	4.44	1183	19
TOR	4.6	1170	14
TEX	5.44	1164	3
LAD	5.59	1163	2
SEA	4.68	1144	12
SDP	4.64	1144	13
MIN	4.8	1138	10
HOU	5.1	1134	5
SFG	4.16	1130	24
CHC	5.06	1127	6
PHI	4.91	1123	8
CLE	4.09	1121	27
BOS	4.77	1117	11
PIT	4.27	1111	22
CIN	4.83	1107	9
DET	4.08	1103	28
LAA	4.56	1102	16
MIA	4.11	1097	26
ARI	4.6	1091	14
NYM	4.43	1082	20
WSN	4.32	1082	21
MIL	4.49	1075	17
ATL	5.85	1062	1
TBR	5.31	1047	4
OAK	3.61	1043	30
BAL	4.98	1041	7
COL	4.45	1038	18
KCR	4.17	1009	23
CHW	3.96	1009	29
NYY	4.15	1007	25

From the data returned, we can see that number of runners left on base does not significantly affect the runs scored per game.

9. **Which player(non-pitcher) played the most game not starting?**

Name	TeamID	gamesNotStarting	Pos_Summary
Ryan McKenna	BAL	55	OF-P

10. List out any player who has appeared as both batter and pitcher and their stats.

Name	TeamID	AB	BA	IP	ERA	Pos_Summary
Lane Thomas	WSN	628	0.268	1	27	*98/HD1
Shohei Ohtani	LAA	497	0.304	132	3.14	*D1
Zach McKinstry	DET	464	0.231	1	18	549H67/8D1
Miguel Rojas	LAD	385	0.236	3	6	*6H/1D45
Martín Maldonado	HOU	362	0.191	1	36	*2/1
Willi Castro	MIN	358	0.257	2.1	11.57	785H4/69D1
Luke Raley	TBR	357	0.249	2.2	30.38	93D7H8/1
Isiah Kiner-Falefa	NYY	326	0.242	4	2.25	875H/9146
Alec Burleson	STL	315	0.244	1.2	21.6	7DH39/1
Christian Bethancourt	TBR	315	0.225	0.1	81	2/HD1
Rowdy Tellez	MIL	311	0.215	1	0	3DH/1
Brandon Crawford	SFG	283	0.194	1	0	6/H1

Here, I set a filter of minimum 100 at-bats or 50 innings pitched. 42 players satisfied one of the filters, but only Shohei Ohtani satisfied both filters simultaneously.

11. Ranking pitchers who created the most double plays

(found out it is impossible to get this insight using the current data).