

In order to analyze the Rockies performance at home, we first need to establish a baseline to compare them against. If we average all the MLB, we can establish metrics to compare against by park. This will allow us to create a “profile” for each park by measuring the percent difference in BA, OBP, SLG, and ISO ($ISO = SLG - BA$, *Isolated Power*).

To start, we must compute the MLB averages with the following query:

```
#Compute averages for BA, OBP, SLG, and ISO for the MLB
drop table if exists mlb_stats;
create table mlb_stats as
  select
    round((sum(case when EVENT_CD > 19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0
end)),3) as 'BA',
    round(((sum(case when EVENT_CD > 19 then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0
end)+sum(case EVENT_CD when 15 then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end))/(sum(case
AB_FL when 'T' then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0 end)+sum(case EVENT_CD when 15
then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end)+sum(case SF_FL when 'T' then 1 else 0
end))),3) as 'OBP',
    round(((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end))),3) as 'SLG',
    round(((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end))),3)-round((sum(case when
EVENT_CD > 19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0 end)),3) as 'ISO'
  from events;
```

This gives us the following MLB averages:

BA: .255
OBP: .320
SLG: .401
ISO: .146

Then we can compute the averages for each ballpark in the MLB, which is very similar to the original query, but organized by home team:

```
#Compute averages for BA, OBP, SLG, and ISO for away teams all stadiums
drop table if exists stadium_stats;
create table stadium_stats as
  select
    HOME_TEAM_ID as 'Team',
    round((sum(case when EVENT_CD > 19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0
end)),3) as 'BA',
    round(((sum(case when EVENT_CD > 19 then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0
end)+sum(case EVENT_CD when 15 then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end))/(sum(case
AB_FL when 'T' then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0 end)+sum(case EVENT_CD when 15
then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end)+sum(case SF_FL when 'T' then 1 else 0
end))),3) as 'OBP',
    round(((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end))),3) as 'SLG',
```

```

round(((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end))),3)-round((sum(case when
EVENT_CD > 19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0 end)),3) as 'ISO'
from events
where BAT_HOME_ID = 0 and HOME_TEAM_ID != 'HOM'
group by HOME_TEAM_ID;

```

This gives us a list of the BA, OBP, SLG, and ISO for the away team at every stadium. We eliminate the home team statistics to eliminate for good/bad team bias, as over the length of these statistics, some teams have been much better than others. By only measuring the away team, we can eliminate this bias.

Then we can compute the percentage difference for each statistic, as well as an average score, for each ballpark centered around 100. For example, a core of 105 means that at that ballpark, that specific statistic is 5% higher than the MLB average:

```

#Compute a score (as percent difference, centered at 100) for each stadium
drop table if exists stadium_factors;
create table stadium_factors as
select
    Team as 'Team',
    round((((BA/0.255)+(OBP/0.320)+(SLG/0.401)+(ISO/0.146))/4)*100,0) as 'Offensive Park Factor',
    round((BA/0.255)*100,0) as 'BA Factor',
    round((OBP/0.320)*100,0) as 'OBP Factor',
    round((SLG/0.401)*100,0) as 'SLG Factor',
    round((ISO/0.146)*100,0) as 'ISO Factor'
from stadium_stats
order by (((BA/0.255)+(OBP/0.320)+(SLG/0.401)+(ISO/0.146))/4)*100 DESC;

```

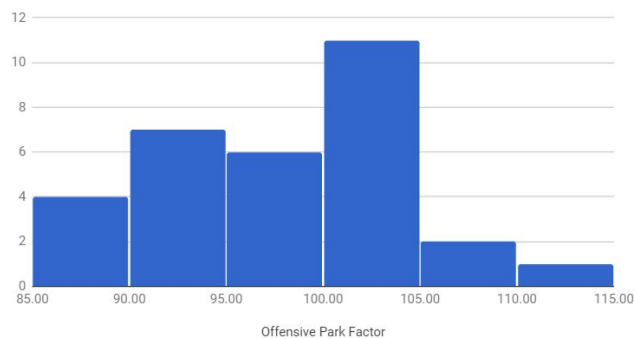
Which gives us the following results:

Team	Offensive Park Factor	BA Factor	OBP Factor	SLG Factor	ISO Factor
COL	111	111	107	112	114
BAL	108	104	103	108	116
TOR	108	102	102	108	118
HOU	104	102	103	104	108
ARI	104	103	99	105	109
MIN	103	106	101	104	101
MIL	103	102	100	104	108
KCA	102	104	103	102	100
CHN	102	100	103	101	103
NYA	102	99	98	102	108
BOS	101	101	102	101	101
CHA	101	100	100	101	103
TEX	101	99	98	101	104
CIN	101	97	97	101	108
CLE	99	100	100	99	97
FLO	97	98	100	96	92
PHI	96	96	94	97	99
DET	96	99	98	96	90
WAS	95	99	97	96	90
NYN	95	96	98	94	90
MIA	94	100	99	94	84
ANA	94	96	96	94	90
PIT	92	98	98	92	82
SEA	92	95	93	93	89
SDN	91	91	94	90	87
TBA	90	89	91	90	92
LAN	90	93	95	89	83
SLN	89	94	94	89	79
ATL	88	93	93	88	80
SFN	88	91	94	87	81
OAK	88	91	92	87	81

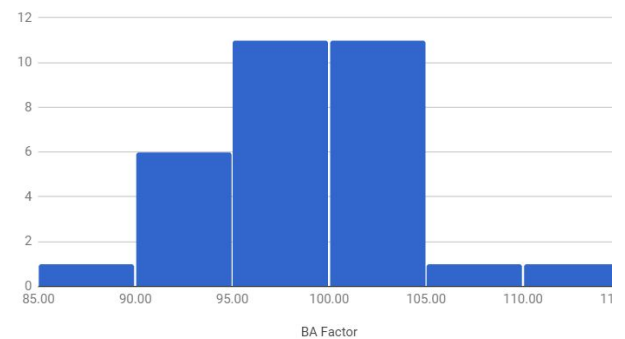
As you can see from the table, the Rockies have the ballpark with the most elevated stats, on average, 11% higher than the MLB average. The A's have the ballpark with the most deflated stats, on average, 12% lower than the MLB average. By looking at the 4 individual factors (BA, OBP, SLG, ISO), we can create a Park Profile. For example, The Reds (CIN) have an unusual set. The BA and OBP are both 3% below average, but the ISO is 8% above average. This indicates that getting hits is harder, but homeruns are easier (which tells us that there is probably a small outfield and softer turf/ground).

Below we can see the histograms for the 5 factors:

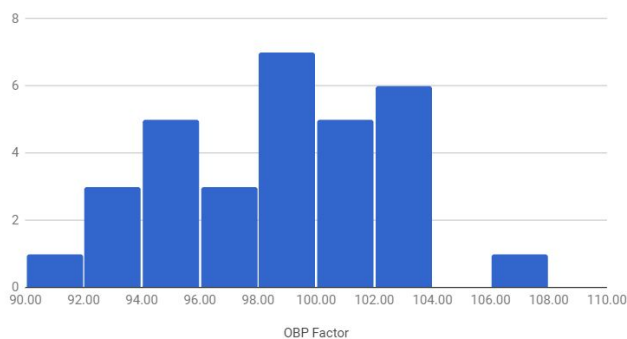
Offensive Park Factor



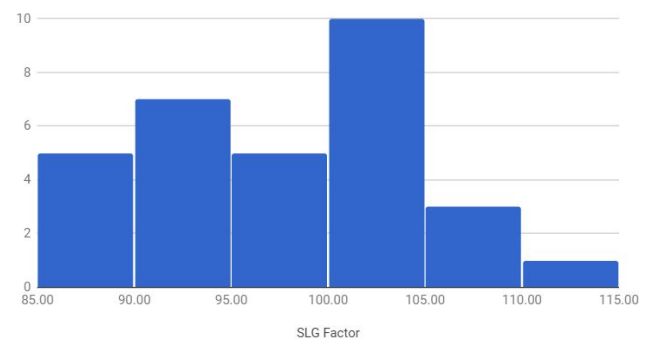
BA Factor



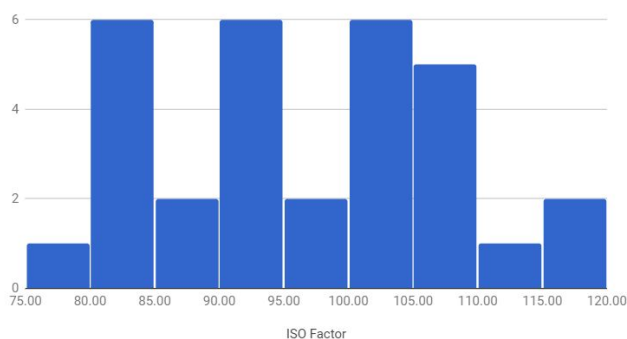
OBP Factor



SLG Factor



ISO Factor



Now we need to look at the Rockies performance. In order to do this, we can adjust their home stats for their park factor (all home BA will be divided by 1.11 as defined by the table above for COL, and so on) and compare their home stats to their away stats. This should allow us to identify players who, adjusted, player better/worse at home compared to on the road. We don't really need to adjust for away games because we can assume that any park factor bias evens out after enough games are played at enough different parks.

Here are the queries that compile the away and home stats, and then compare them:

```
#Congregate all of the home stats for Rockies players adjusting for Park Factor above
drop table if exists COL_home_stats;
create table COL_home_stats as
select
    BAT_ID as 'BatterID',
    count(*) as 'Home_AB',
    round(((sum(case when EVENT_CD > 19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0
end))/1.11),3) as 'Home_Adj_BA',
    round((((sum(case when EVENT_CD > 19 then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0
end)+sum(case EVENT_CD when 15 then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end))/(sum(case
AB_FL when 'T' then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0 end)+sum(case EVENT_CD when 15
then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end)+sum(case SF_FL when 'T' then 1 else 0
end)))/1.07),3) as 'Home_Adj_OBP',
    round((((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end)))/1.12),3) as 'Home_Adj_SLG',
    round((((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end)))-sum(case when EVENT_CD >
19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0 end)/1.14),3) as 'Home_Adj_ISO'
from events
where BAT_TEAM_ID = 'COL' and HOME_TEAM_ID = 'COL'
group by BAT_ID;

#Congregate all of the away stats for Rockies players
drop table if exists COL_away_stats;
create table COL_away_stats as
select
    BAT_ID as 'BatterID',
    count(*) as 'Away_AB',
    round((sum(case when EVENT_CD > 19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0
end)),3) as 'Away_BA',
    round(((sum(case when EVENT_CD > 19 then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0
end)+sum(case EVENT_CD when 15 then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end))/(sum(case
AB_FL when 'T' then 1 else 0 end)+sum(case EVENT_CD when 14 then 1 else 0 end)+sum(case EVENT_CD when 15
then 1 else 0 end)+sum(case EVENT_CD when 16 then 1 else 0 end)+sum(case SF_FL when 'T' then 1 else 0
end))),3) as 'Away_OBP',
    round(((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end))),3) as 'Away_SLG',
    round(((sum(H_CD)/sum(case AB_FL when 'T' then 1 else 0 end))),3)-round((sum(case when
EVENT_CD > 19 then 1 else 0 end)/sum(case AB_FL when 'T' then 1 else 0 end)),3) as 'Away_ISO'
from events
where BAT_TEAM_ID = 'COL' and HOME_TEAM_ID != 'COL'
group by BAT_ID;

#Compare the home and away stats
select distinct n.FIRST_NAME_TX as 'First', n.LAST_NAME_TX as 'Last', chs.BatterID,
    (chs.Home_Adj_BA-cas.Away_BA) as 'BA Diff',
    (chs.Home_Adj_OBP-cas.Away_OBP) as 'OBP Diff',
    (chs.Home_Adj_SLG-cas.Away_SLG) as 'SLG Diff',
```

```

      (chs.Home_Adj_ISO-cas.Away_ISO) as 'ISO Diff'
from COL_home_stats as chs
join COL_away_stats cas on chs.BatterID = cas.BatterID
join rosters n on n.PLAYER_ID = chs.BatterID and n.PLAYER_ID = cas.BatterID
where chs.Home_AB > 50 and cas.Away_AB > 50
order by
(chs.Home_Adj_BA-cas.Away_BA)+(chs.Home_Adj_OBP-cas.Away_OBP)+(chs.Home_Adj_SLG-cas.Away_SLG)+(chs.Home_Adj_ISO-cas.Away_ISO);

```

If we look at the final resulting table, we have that:

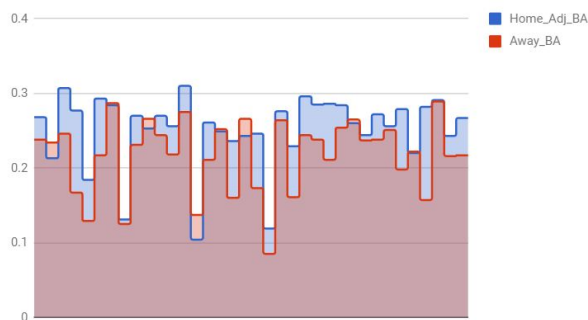
First	Last	BatterID	BA Diff	OBP Diff	SLG Diff	ISO Diff
Jason	Hammel	hammi002	-0.033	0.001	-0.096	-0.046
Ian	Stewart	stewi001	-0.002	0.004	-0.082	-0.034
Clint	Barnes	barmc001	-0.021	0.020	-0.043	0.021
Jonathan	Herrera	herri002	-0.023	-0.009	-0.012	0.055
Jorae	de la Rosa	delai001	0.006	0.012	0.005	0.020
Tv	Widinton	wiaot001	0.027	0.052	-0.010	0.016
Trov	Tulowitzki	tulot001	0.002	0.012	0.002	0.069
Wilin	Rosario	rosaw001	-0.005	0.017	0.015	0.084
Todd	Helton	heltt001	-0.003	0.001	0.030	0.088
Michael	Cuddver	cuddm001	-0.003	0.020	0.018	0.088

These are the player that perform worse at the Rockies home stadium when adjusting for park factor in Colorado. The difference is *Home Stat - Away Stat*, so a low score indicates that they underperform at home. If we look at the best players at home we get:

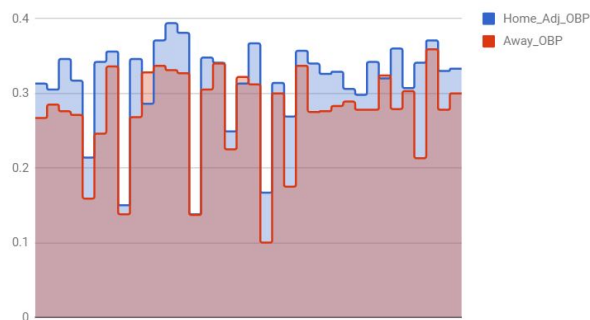
First	Last	BatterID	BA Diff	OBP Diff	SLG Diff	ISO Diff
Corev	Dickerson	dickc002	0.039	0.078	0.257	0.293
Tvler	Colvin	colvt001	0.076	0.096	0.198	0.197
Brad	Hawpe	hawpb001	0.050	0.043	0.227	0.244
Andrew	Brown	browa003	0.110	0.046	0.215	0.174
Chris	Iannetta	iannc001	0.073	0.055	0.207	0.198
Ramon	Hernandez	herrn002	0.076	0.024	0.206	0.186
Yorvit	Torrealba	torrv001	0.125	0.128	0.154	0.077
Miguel	Olivo	olivm001	0.075	0.050	0.174	0.166
Rvan	Soilborhds	soilr001	0.081	0.081	0.126	0.104
Carlos	Gonzalez	gonzc001	0.035	0.054	0.101	0.142

Where these are the players that perform better at Coors Field than at other stadiums when adjusting for park factor. Below, we can see the histograms for all Rockies players who have at least 50 home AB and 50 away AB. Any column in which red is higher than blue indicates that they perform better on the road than at home. From these, we can see that most players perform better at Coors, even when accounting for the park factor.

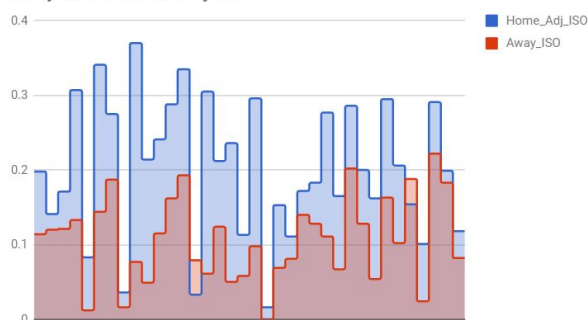
Away BA v. Home Adj. BA



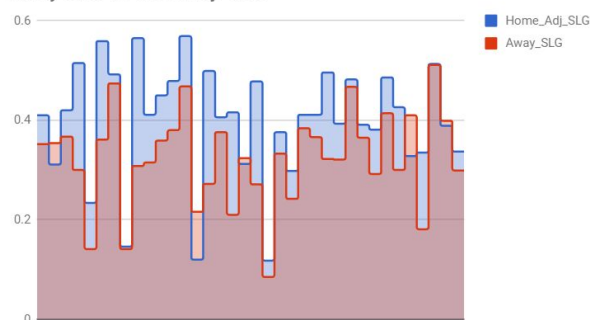
Away OBP v. Home Adj. OBP



Away ISO v. Home Adj. ISO



Away SLG v. Home Adj. SLG



When looking at player to resign/trade away, we have to consider the fact that half of the games are played at home and half are played on the road. So there are player who play better at home/away when accounting for park factor, but the contribution evens out over the season.

I think the more important analysis to take from this is looking at platooning certain players (generally used in terms of batting a right versus a lefty pitcher and vice versa), but instead platooning based on home/away. Corey Dickerson, for example, should always be starting at home, but can probably be benched during away games (he's on the Pirates now). Jason Hammel, on the other hand, should probably be benched at home, but could start on the road (he's actually a pitcher, but hit enough to count here, and he's on the Royals).

The same analysis could be done in reverse, by looking at pitching. A pitcher's ERA, HR/9, K/9, and BB/9 could be adjusted for their home park to see who you should look to start on the road and who to start at home (or who to bring in for relief duty if a player has major home/away splits).