Altitude Effects on Baseball Statistics and Park Factor

Sam Busser Kevin Nguyen Armen Arsenian Blaise Page



Questions Sought To Answer

- The main question we are seeking to answer is whether or not altitude really has a significant effect on the game of baseball.
- We answered these questions by looking at offensive statistics such as batting average, runs per 9 innings, the ratio of home runs to total runs, and several more offensive stats.

Data Preparation

<u>Data Cleaning:</u> In order to prepare the data, we had to combine the statistics for Florida and Miami since they are the same team but changed cities.

<u>Data Reduction:</u> We had to go through our data and remove columns that we did not need, such as who was on base during the hit, or what the pitch count was. We had to group by stadiums for some statistics, so we had to combine all the results of statistics that had to do with each stadium. This made the data much easier to work with

<u>Data Integration:</u> We had to merge two tables, one with all the baseball data and one with all the altitude data, so that we had a table with both. This allowed us to test our correlations.

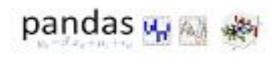
List of Tools

iPython Notebook to process our data Pandas to manipulate and clean our data set Numpy to test the data for correlation MatPlotLib to visually represent our findings.







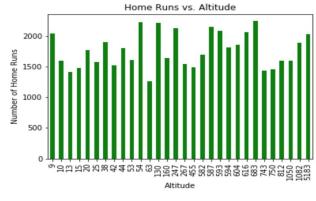


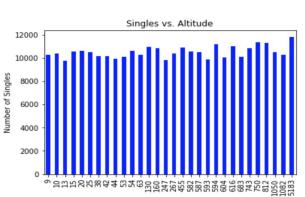
Procedure

- Preprocess the data
- Run statistical tests on the dataset
- Determine key event correlations
- Evaluate results

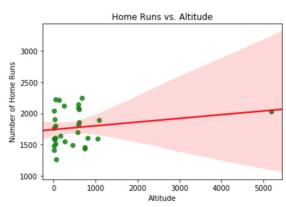
Results: Highest Correlation of Events

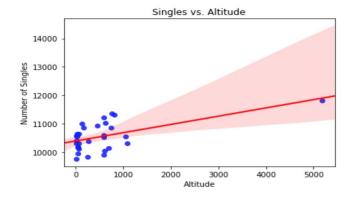
event correlation 1 strikeouts -0.235383 4 caught stealing -0.0922677 7 passed ball -0.0839017 10 foul error -0.0586253 13 hit by pitch -0.0477827 2 stolen bases -0.0329117 15 interference -0.0126699 11 walk 0.0613119 14 generic outs 0.100284 0 generic outs 0.100284 16 error 0.117782 3 defensive indifference 0.117391 4 wild pitch 0.191274 21 home runs 0.20928 8 balk 0.2893827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.555642			
4 caught stealing -0.0922677 7 passed ball -0.0839017 10 foul error -0.0586253 13 hit by pitch -0.0477827 2 stolen bases -0.0329117 15 interference -0.0126699 11 walk 0.0613119 14 generic outs 0.100284 0 generic outs 0.100284 16 error 0.1177182 3 defensive indifference 0.117391 6 wild pitch 0.191274 21 home runs 0.19932 12 intentional walks 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608		event	correlation
7 passed ball -0.0839017 10 foul error -0.0586253 13 hit by pitch -0.0477827 2 stolen bases -0.0329117 15 interference -0.0126699 11 walk 0.0613119 14 generic outs 0.100284 0 generic outs 0.100284 16 error 0.117182 3 defensive indifference 0.117391 6 wild pitch 0.191274 21 home runs 0.19932 12 intentional walks 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	1	strikeouts	-0.235383
10 foul error -0.0586253 13 hit by pitch -0.0477827 2 stolen bases -0.0329117 15 interference -0.0126699 11 walk 0.0613119 14 generic outs 0.100284 0 generic outs 0.100284 16 error 0.117182 13 defensive indifference 0.117391 16 wild pitch 0.191274 11 home runs 0.19932 12 intentional walks 0.20928 18 balk 0.283827 19 other advance 0.289123 17 fielder's choice 0.289535 18 doubles 0.400911 20 triples 0.517608	4	caught stealing	-0.0922677
13 hit by pitch -0.0477827 2 stolen bases -0.0329117 15 interference -0.0126699 11 walk 0.0613119 14 generic outs 0.100284 0 generic outs 0.100284 16 error 0.117782 3 defensive indifference 0.117391 6 wild pitch 0.191274 21 home runs 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	7	passed ball	-0.0839017
2 stolen bases -0.0329117 15 interference -0.0126699 11 walk 0.0613119 14 generic outs 0.100284 0 generic outs 0.100284 16 error 0.117182 3 defensive indifference 0.117391 6 wild pitch 0.191274 21 home runs 0.19932 12 intentional walks 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	10	foul error	-0.0586253
15 interference -0.0126699 11 walk 0.0613119 14 generic outs 0.100284 0 generic outs 0.100284 16 error 0.117391 6 wild pitch 0.191274 21 home runs 0.19932 12 intentional walks 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	13	hit by pitch	-0.0477827
11 walk 0.0613119 14 generic outs 0.100284 0 generic outs 0.100284 16 error 0.117182 3 defensive indifference 0.117391 6 wild pitch 0.191274 21 home runs 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	2	stolen bases	-0.0329117
14 generic outs 0.100284 0 generic outs 0.100284 16 error 0.117182 3 defensive indifference 0.117391 6 wild pitch 0.191274 21 home runs 0.29932 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	15	interference	-0.0126699
0 generic outs 0.100284 16 error 0.117182 3 defensive indifference 0.117391 6 wild pitch 0.191274 21 home runs 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	11	walk	0.0613119
16 error 0.117182 3 defensive indifference 0.117391 6 wild pitch 0.191274 21 home runs 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	14	generic outs	0.100284
3 defensive indifference 0.117391 6 wild pitch 0.191274 21 home runs 0.19932 12 intentional walks 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	0	generic outs	0.100284
6 wild pitch 0.191274 21 home runs 0.19932 12 intentional walks 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	16	error	0.117182
21 home runs 0.19932 12 intentional walks 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	3	defensive indifference	0.117391
12 intentional walks 0.20928 8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	6	wild pitch	0.191274
8 balk 0.283827 9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	21	home runs	0.19932
9 other advance 0.289123 17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	12	intentional walks	0.20928
17 fielder's choice 0.289535 5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	8	balk	0.283827
5 pickoff 0.365965 19 doubles 0.400911 20 triples 0.517608	9	other advance	0.289123
19 doubles 0.400911 20 triples 0.517608	17	fielder's choice	0.289535
20 triples 0.517608	5	pickoff	0.365965
	19	doubles	0.400911
18 singles 0.555642	20	triples	0.517608
	18	singles	0.555642



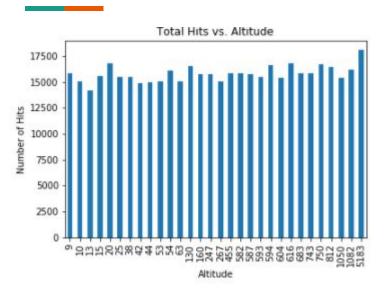


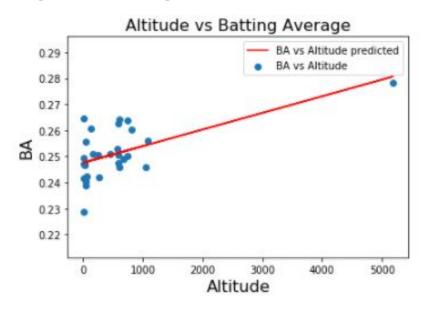
Altitude





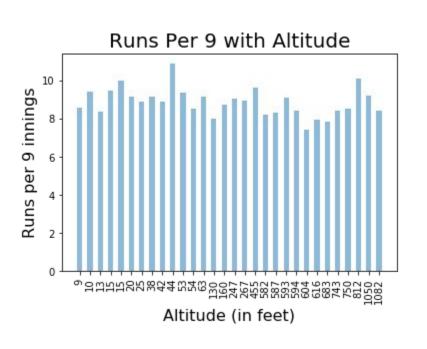
Results: Total Hits & Batting Average vs. Altitude

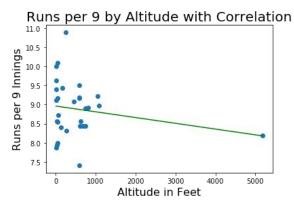


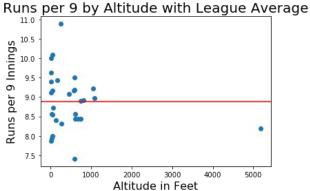


The correlation between BA and Altitude is 0.613893390432

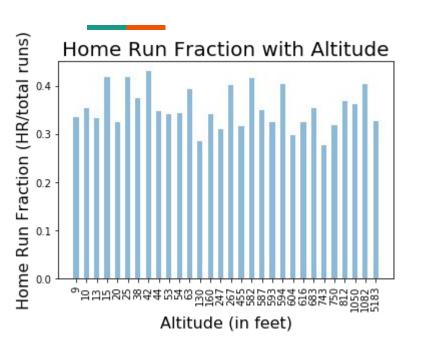
Results: Runs Per 9

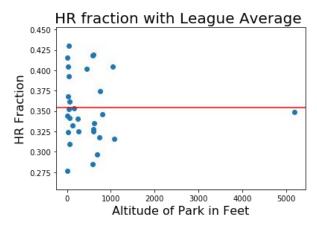


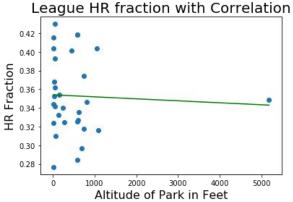




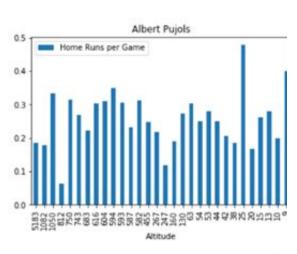
Results: HR Fraction

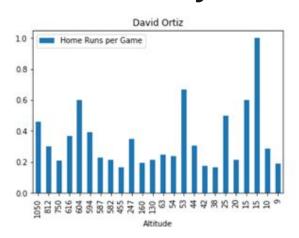


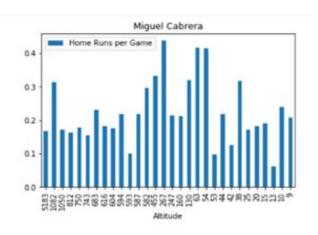


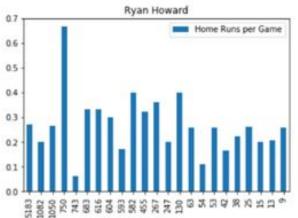


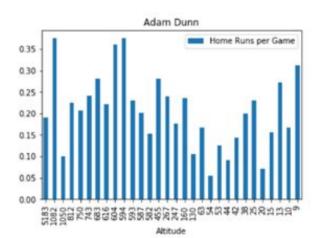
Results: Player Stats



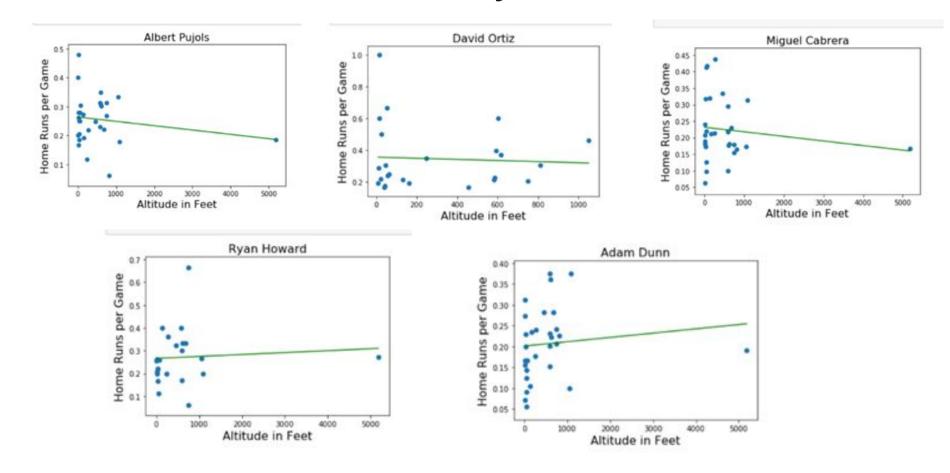








Results: Player Stats



How the Results Can Be Applied

- Humidor is working based on our research.
- Scientifically the offensive stats should be higher, but our results say that altitude doesn't play much of a factor
- Look into other factors:
 - Park dimensions, wind speed, etc.