

THE MLB PAYROLL PROBLEM

JOSH WOLF

CMSE 201, SECTION 008



INTRODUCTION

- FOR THIS PROJECT, I USED DATA FROM THE LAHMAN BASEBALL DATABASE
- I EXTRACTED THE “TEAMS”, “SALARIES”, AND “POSITIONS” TABLES
- AFTER FILTERING THE DATA, I ANALYZED YEARS FROM 1985-2016

1 teams																				
	yearID	IgID	teamID	franchID	divID	Rank	G	Ghome	W	L	... DP	FP	name	park	attendance	BPF	PPF	teamIDBR		
0	1871	NaN	BS1	BNA	NaN	3	31	NaN	20	10	... 24	0.834	Boston Red Stockings	South End Grounds I	NaN	103	98		BOS	
1	1871	NaN	CH1	CNA	NaN	2	28	NaN	19	9	... 16	0.829	Chicago White Stockings	Union Base-Ball Grounds	NaN	104	102		CHI	
2	1871	NaN	CL1	CFC	NaN	8	29	NaN	10	19	... 15	0.818	Cleveland Forest Citys	National Association Grounds	NaN	96	100		CLE	
3	1871	NaN	FW1	KEK	NaN	7	19	NaN	7	12	... 8	0.803	Fort Wayne Kekiongas	Hamilton Field	NaN	101	107		KEK	
4	1871	NaN	NY2	NNA	NaN	5	33	NaN	16	17	... 14	0.840	New York Mutuals	Union Grounds (Brooklyn)	NaN	90	88		NYU	
...		
2950	2020	NL	SLN	STL	C	3	58	27.0	30	28	... 46	0.983	St. Louis Cardinals	Busch Stadium III	0.0	97	96		STL	
2951	2020	AL	TBA	TBD	E	1	60	29.0	40	20	... 52	0.985	Tampa Bay Rays	Tropicana Field	0.0	96	95		TBR	
2952	2020	AL	TEX	TEX	W	5	60	30.0	22	38	... 40	0.981	Texas Rangers	Globe Life Field	0.0	102	102		TEX	
2953	2020	AL	TOR	TOR	E	3	60	26.0	32	28	... 47	0.982	Toronto Blue Jays	Sahlen Field	0.0	100	99		TOR	
2954	2020	NL	WAS	WSN	E	4	60	33.0	26	34	... 48	0.981	Washington Nationals	Nationals Park	0.0	103	102		WSN	

2955 rows × 48 columns

1 salaries					
	yearID	teamID	IgID	playerID	salary
0	1985	ATL	NL	barkele01	870000
1	1985	ATL	NL	bedrost01	550000
2	1985	ATL	NL	benedbr01	545000
3	1985	ATL	NL	campri01	633333
4	1985	ATL	NL	ceronri01	625000
...
26423	2016	WAS	NL	strasst01	10400000
26424	2016	WAS	NL	taylomi02	524000
26425	2016	WAS	NL	treinbl01	524900
26426	2016	WAS	NL	werthja01	21733615
26427	2016	WAS	NL	zimmery01	14000000

26428 rows × 5 columns

1 fielding																		
	playerID	yearID	stint	teamID	IgID	POS	G	GS	InnOuts	PO	A	E	DP	PB	WP	SB	CS	ZR
0	abercda01	1871	1	TRO	NaN	SS	1	1.0	24.0	1	3	2.0	0	NaN	NaN	NaN	NaN	
1	addybo01	1871	1	RC1	NaN	2B	22	22.0	606.0	67	72	42.0	5	NaN	NaN	NaN	NaN	
2	addybo01	1871	1	RC1	NaN	SS	3	3.0	96.0	8	14	7.0	0	NaN	NaN	NaN	NaN	
3	allisar01	1871	1	CL1	NaN	2B	2	0.0	18.0	1	4	0.0	0	NaN	NaN	NaN	NaN	
4	allisar01	1871	1	CL1	NaN	OF	29	29.0	729.0	51	3	7.0	1	NaN	NaN	NaN	NaN	
...	
144763	zimmebr02	2020	1	BAL	AL	P	2	1.0	21.0	0	0	0.0	0	NaN	NaN	NaN	NaN	
144764	zimmejo02	2020	1	DET	AL	P	3	2.0	17.0	0	1	0.0	0	NaN	NaN	NaN	NaN	
144765	zimmeky01	2020	1	KCA	AL	P	16	1.0	69.0	2	4	0.0	1	NaN	NaN	NaN	NaN	
144766	zuberty01	2020	1	KCA	AL	P	23	0.0	66.0	0	2	0.0	0	NaN	NaN	NaN	NaN	
144767	zuninmi01	2020	1	TBA	AL	C	28	25.0	647.0	244	15	1.0	2	5.0	NaN	11.0	4.0	

144768 rows × 18 columns

RESEARCH AIMS

- WHAT IS THE DISADVANTAGE SMALL MARKET TEAMS FACE
- IF THERE IS A CLEAR HANDICAP, WHERE SHOULD THESE TEAMS ALLOCATE THEIR SMALLER PAYROLLS?

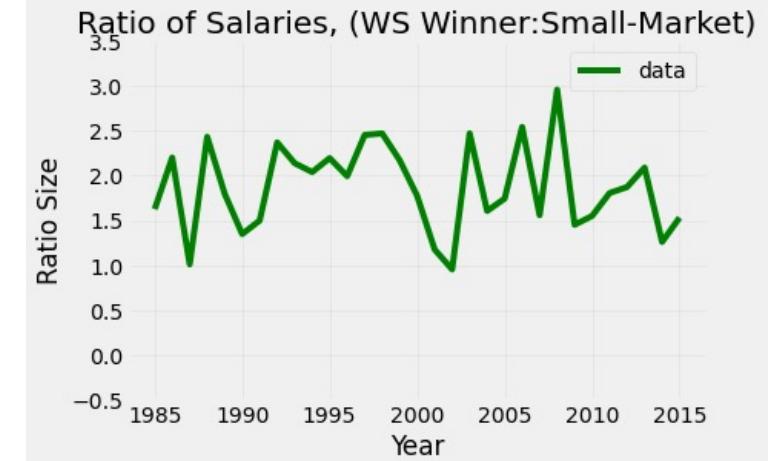
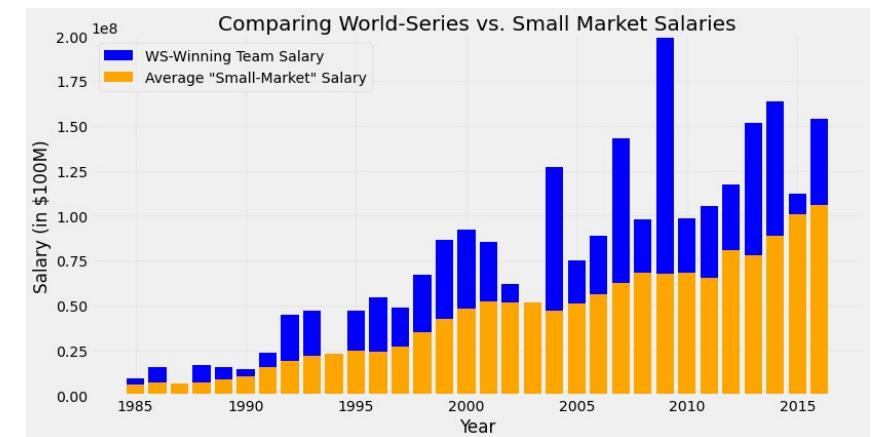


BACKGROUND INFORMATION

- SMALL MARKET TEAMS
 - T.V. MARKET SIZE, REGIONAL BROADCASTING RIGHTS, METROPOLITAN SIZE, OWNERSHIP WEALTH, STADIUM/FACILITIES
- DEFINING WHICH TEAMS ARE SMALL MARKET
 - KANSAS CITY ROYALS, OAKLAND A'S, TAMPA BAY RAYS, CINCINNATI REDS, MIAMI MARLINS, CLEVELAND INDIANS, MILWAUKEE BREWERS, DETROIT TIGERS, PITTSBURGH PIRATES, COLORADO ROCKIES, ARIZONA DIAMONDBACKS, MINNESOTA TWINS, BALTIMORE ORIOLES, SAN DIEGO PADRES, SEATTLE MARINERS
- RANDOMLY SELECTING FOUR CHAMPIONSHIP TEAMS
 - THIS ALLOWED ME TO ANALYZE HOW WORLD SERIES WINNERS SPENT THEIR PAYROLL VIA POSITION

RESULTS WITH VISUALIZATIONS

- OVER 31 YEARS, THE MEAN SMALL MARKET SALARY WAS GREATER THAN THE CHAMPIONSHIP PAYROLL ONLY 9.6% OF THE TIME
- ON AVERAGE, CHAMPIONSHIP TEAMS SPENT 1.86X MORE ON PAYROLL THAN SMALL MARKET TEAMS



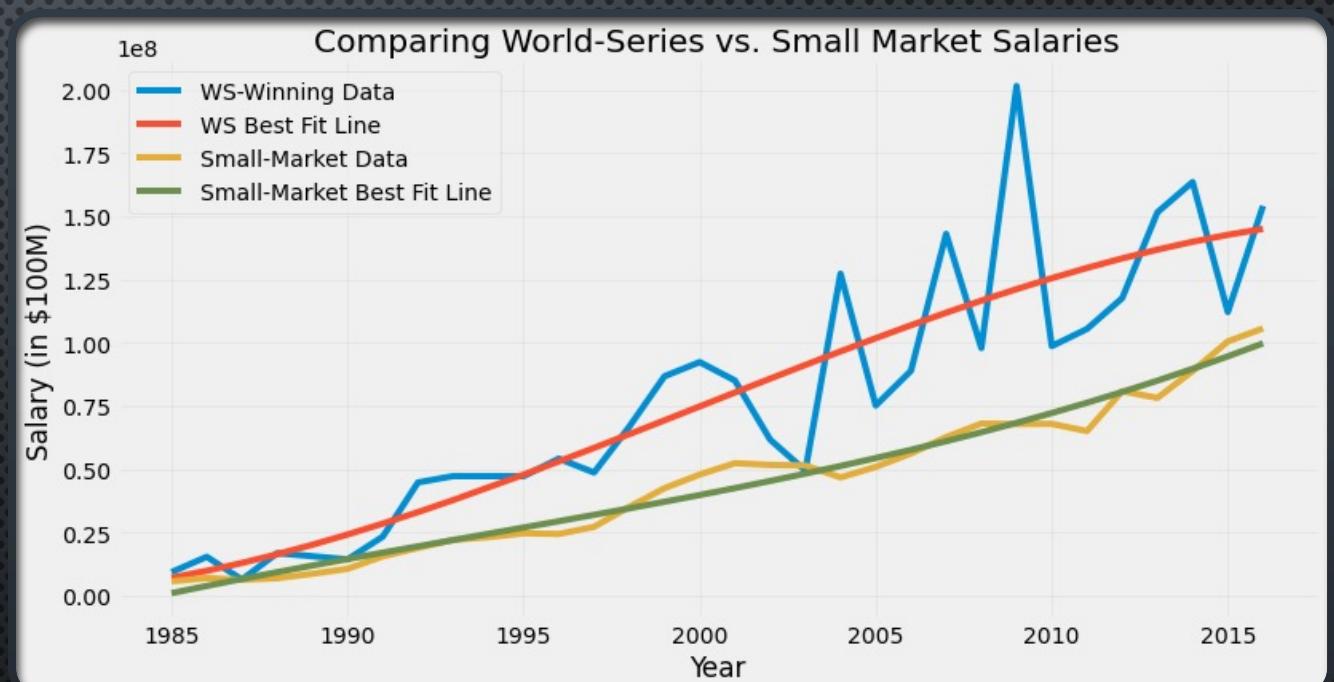
POLYFIT MODEL

- BEST FIT LINE FOR THE MEAN SALARY OF A SMALL-MARKET TEAM IS:

$$1876 x^3 - 1.122 \times 10^7 x^2 + 2.237 \times 10^{10} x - 1.487 \times 10^{13}$$

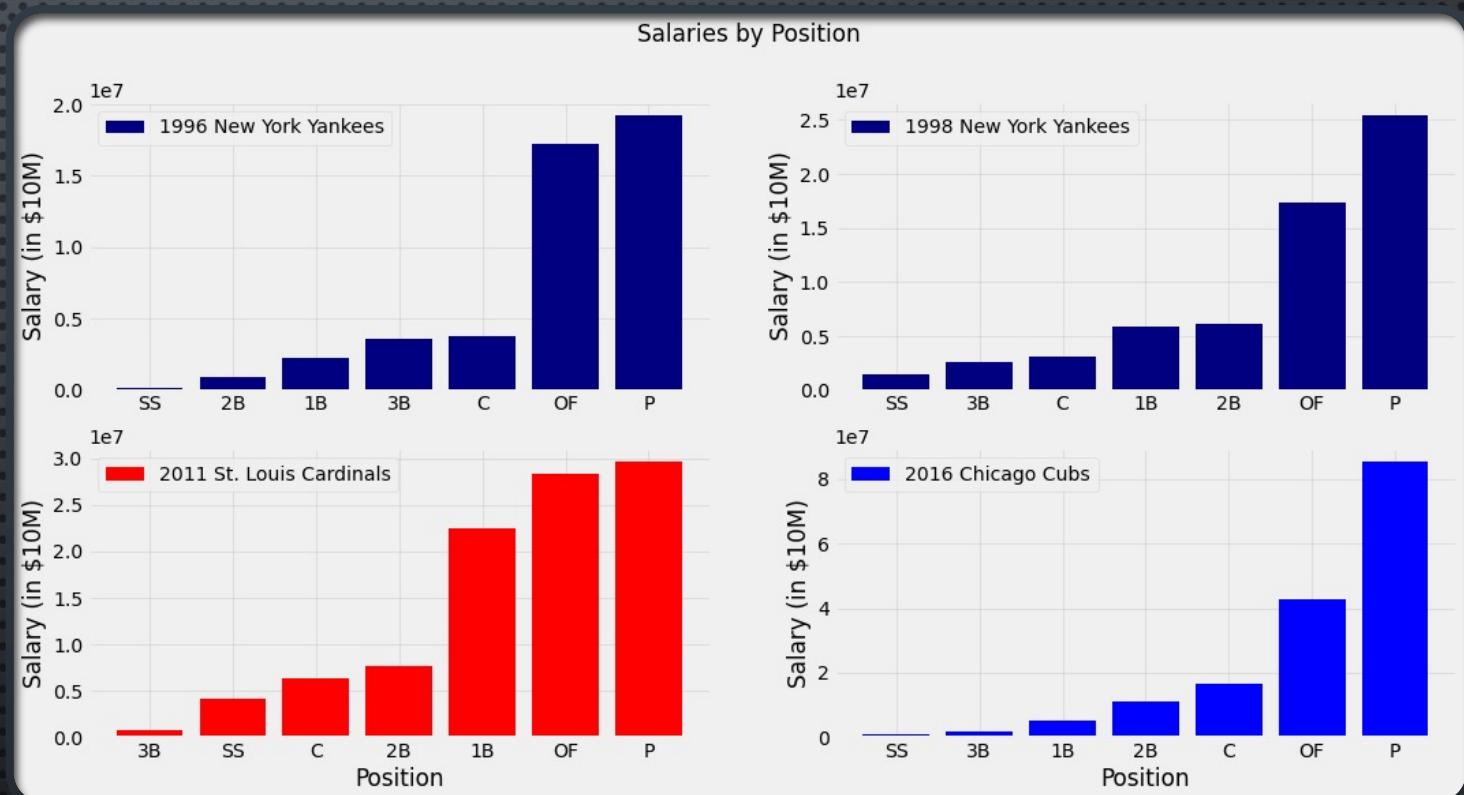
- BEST FIT LINE FOR THE SALARY OF WORLD SERIES WINNING TEAMS IS:

$$-4385 x^3 + 2.631 \times 10^7 x^2 - 5.261 \times 10^{10} x + 3.507 \times 10^{13}$$



SALARIES BY POSITION

- 4 RANDOMLY SELECTED CHAMPIONS
 - 1996, 1998, 2011, 2016
- MOST IMPORTANT POSITION(S)
 - PITCHERS, OUTFIELD, CATCHER
- LEAST IMPORTANT POSITION(S)
 - SHORTSTOP



CONCLUSION AND REFERENCES



Small market teams are at a major disadvantage

Must spend ~1.86x more than they currently do to compete



Focus on allocating salary by position

Primarily spend on pitchers, outfield, and catcher
Use remaining payroll on other positions

Works Cited

"Code Faster with Line-of-Code Completions, Cloudless Processing." Kite, www.kite.com/python/answers/how-to-group-a-pandas-dataframe-by-column-names-and-sum-over-each-group-in-python.

Lahman, Sean. "Download Lahman's Baseball Database." SeanLahman.com, 16 Feb. 2021, www.seanlahman.com/baseball-archive/statistics/

"Pandas.DataFrame.merge¶." Pandas.DataFrame.merge - Pandas 1.2.4 Documentation, pandas.pydata.org/docs/reference/api/pandas.DataFrame.merge.html.

Trueblood, Matt. "Power Ranking All 30 MLB Teams by Market Size." Bleacher Report, Bleacher Report, 3 Oct. 2017, bleacherreport.com/articles/961412-mlb-power-rankings-all-30-mlb-teams-by-market-size.