MAT 259: Project 3

MLB Pitchers: Winning Percentage Across Other Metrics

•••

Brianna Griffin

Concept Overview

I would like to utilize the data set on MLB starting pitchers to analyze relationships between several different variables. Specifically, I am interested in the interaction between variables and ‘winning percentage’. I have recently completed a Machine Learning Final Project in which I built 5 different regression models that predicted the ‘winning percentage’ as an outcome variable. I found that 2 of the most prominent variables in predicting ‘winning percentage’ were ‘best’ and ‘team_win_loss_percentage’. I am curious to see if the same results will be displayed in my 3D visualization. Going into the project, I also wanted to use curves in my visualization, with cubes connecting their endpoints. This is how I began thinking of my concept.

Data

mlb_data

age	tm	ip	nd	wchp	ltuf	wtm	ltm	tm_w_l_	wlst	lsv	cg	sho	qs	qs_p	gm_sc	best	wrst	bqr	s_dr	l_dr	rs_gs	ip_gs	pit_gs	x80	x80_9	x100	x120	max	year	cy_sta	w_pct	arm
25-29	ARI	213.1	14	2	2	19	13	0.594	7	4	2	0	20	63	50.1	78	12	22	0	13	5.1	6.5	93	5	18	7	2	130	2000	no	0.333333	left
19-24	PIT	144	10	1	3	9	17	0.346	3	5	1	0	10	38	44.2	78	20	19	1	17	4.7	5.4	87	6	14	3	3	128	2000	no	0.185185	left
19-24	STL	175	12	2	1	18	12	0.6	3	5	0	0	16	53	55.3	75	15	8	1	16	5.3	5.8	101	1	13	15	1	120	2000	no	0.354838	left
19-24	MON	95	1	2	1	8	9	0.471	0	0	0	0	7	41	50.5	81	23	19	0	9	4.6	5.6	88	5	8	4	0	112	2000	no	0.411764	right
25-29	CHC	32.2	1	0	0	1	3	0.25	1	0	0	0	0	0	32.8	45	23	12	1	2	3.2	3.8	73	2	2	0	0	92	2000	no	0	right
30-34	COL	101.1	5	1	2	9	10	0.474	1	3	0	0	7	37	42.9	64	18	9	1	9	5.6	5.3	94	4	6	9	0	119	2000	no	0.263157	right
19-24	PIT	71.2	6	0	0	5	7	0.417	1	3	0	0	3	25	41.3	73	2	7	2	6	5.7	4.8	81	5	6	1	0	102	2000	no	0.1	right
30-34	TOT	199.1	6	3	1	15	16	0.484	1	2	3	1	14	45	46.5	84	12	16	0	14	4.6	6.4	95	5	15	9	2	140	2000	no	0.387096	right
30-34	PHI	101.1	5	0	0	7	9	0.438	0	2	1	0	7	44	42.8	67	12	7	0	9	4.5	6.3	100	2	5	8	1	124	2000	no	0.25	right
30-34	ATL	98	1	3	1	8	7	0.533	1	0	2	1	7	47	50.4	84	21	9	0	5	4.8	6.5	91	3	10	1	1	140	2000	no	0.533333	right
30-34	COL	196.1	11	5	2	15	17	0.469	5	3	3	0	14	44	48.4	85	14	23	3	9	6	6.1	100	5	7	12	8	131	2000	no	0.375	right
19-24	CIN	140.1	11	2	2	11	15	0.423	2	9	1	0	11	42	48.5	70	26	14	0	14	5.5	5.4	86	9	10	7	0	116	2000	no	0.269230	right
30-34	COL	1.1	1	0	0	0	1	0	0	1	0	0	0	0	16	16	16	2	0	1	10	1.3	45	1	0	0	0	45	2000	no	0	left
30-34	STL	166	7	1	2	14	13	0.519	5	1	1	0	18	67	49.7	66	4	9	1	12	5.9	6	102	3	6	17	1	123	2000	no	0.4	right
25-29	PIT	217.2	10	0	3	13	19	0.406	4	6	2	1	19	59	53.9	87	20	14	0	16	4.4	6.8	108	2	3	22	5	128	2000	no	0.3125	right
25-29	MIL	115	7	0	0	8	12	0.4	2	3	0	0	11	55	48.1	77	5	15	0	10	4.5	5.8	102	1	8	6	5	127	2000	no	0.3	right
30-34	SDP	15.2	0	0	0	0	3	0	1	0	0	0	0	0	38.7	57	25	7	0	2	2	4.4	82	1	2	0	0	89	2000	no	0	right
30-34	COL	177	6	2	2	14	12	0.538	2	4	2	1	11	42	45.2	79	10	24	1	9	6.8	6.1	98	8	2	13	3	139	2000	no	0.294117	left
30-34	PHI	44	5	0	0	2	6	0.25	0	2	1	1	4	50	50.3	78	17	0	0	4	3.6	5.5	92	2	0	5	1	127	2000	no	0.125	right
30-34	PHI	93.1	2	0	0	0	5	0	1	0	0	0	2	40	50.2	73	26	11	0	4	2.6	6.2	101	0	1	4	0	108	2000	no	0	right
35-39	LAD	230	14	0	3	21	12	0.636	4	6	5	1	26	79	62.5	91	13	4	1	7	4.7	7	105	2	5	21	5	124	2000	no	0.393939	right
35-39	ATL	134.1	9	4	0	13	9	0.591	3	5	0	0	9	41	46	71	15	17	1	15	5.7	5.4	92	3	12	7	0	115	2000	no	0.290322	right
19-24	FLA	82.2	3	0	1	4	9	0.308	3	0	0	0	7	54	48.4	71	14	9	0	10	4.1	6.4	104	0	4	9	0	114	2000	no	0.230769	right
25-29	PHI	83	5	0	2	5	10	0.333	3	3	0	0	8	53	43.3	70	18	8	0	9	3.8	5.4	91	3	6	5	1	125	2000	no	0.117647	right
19-24	PHI	94.1	8	3	2	7	8	0.467	3	4	0	0	6	40	55.7	83	33	7	0	7	4.5	6.3	97	3	3	8	1	126	2000	no	0.2	left
25-29	SDP	205	4	5	5	17	17	0.5	1	1	0	0	16	47	47.2	81	10	30	0	10	4.6	6	104	1	13	16	4	135	2000	no	0.382352	right
19-24	PHI	27	3	2	0	3	2	0.6	1	2	0	0	1	20	41.6	51	37	0	0	3	8	5.4	97	0	3	1	1	121	2000	no	0.4	right
25-29	PIT	95	3	1	1	8	9	0.471	2	0	0	0	6	35	45.7	78	12	15	0	9	5	5.5	89	4	7	5	1	123	2000	no	0.333333	right
30-34	FLA	125	8	1	1	6	15	0.286	3	3	0	0	9	43	45.3	73	10	23	0	13	3.6	5.8	88	7	7	7	0	106	2000	no	0.136363	right
25-29	SDP	38.1	2	0	0	1	2	0.333	1	1	0	0	1	33	35	53	26	11	0	2	6.4	4.4	93	1	0	2	0	101	2000	no	0	right
19-24	MIL	162.1	4	0	3	15	8	0.652	1	2	1	1	18	78	58.1	83	24	10	0	11	3.9	7.1	102	0	9	14	0	117	2000	no	0.521739	right
25-29	TOT	167	6	1	8	7	21	0.25	3	1	0	0	14	50	41.9	67	8	22	1	13	4.2	5.7	93	5	14	9	0	115	2000	no	0.125	left
25-29	ARI	96	5	1	2	4	12	0.25	2	1	0	0	5	31	38.1	67	11	22	1	7	4.5	5.5	89	3	9	4	0	114	2000	no	0.1	left
25-29	PHI	71	1	0	6	3	9	0.25	1	0	0	0	9	75	47	60	8	0	0	6	3.7	5.9	98	2	5	5	0	115	2000	no	0.166666	left
25-29	MIL	73.2	1	0	0	1	1	0.5	0	0	0	0	0	0	35.5	62	9	17	0	1	3.5	4.3	75	1	1	0	0	87	2000	no	0	left
25-29	SFG	17.1	0	1	0	1	0	1	0	0	0	0	0	0	42	42	42	0	0	0	10.1	5	87	0	1	0	0	87	2000	no	0.111111	right

- 3,762 rows
- 33 columns
- 2000-2022

Link to Data Source

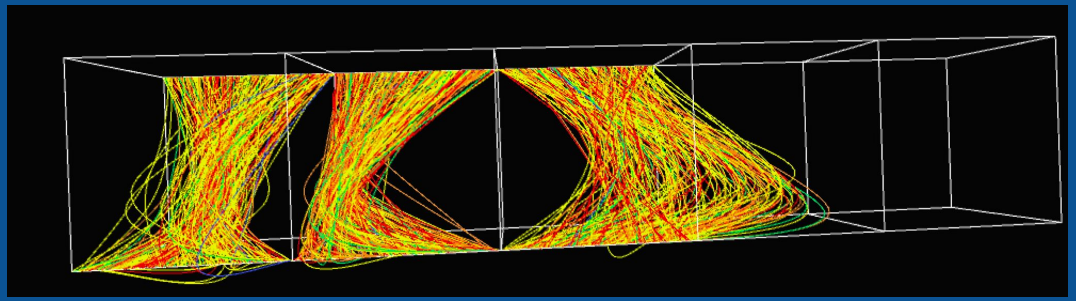
Code Book

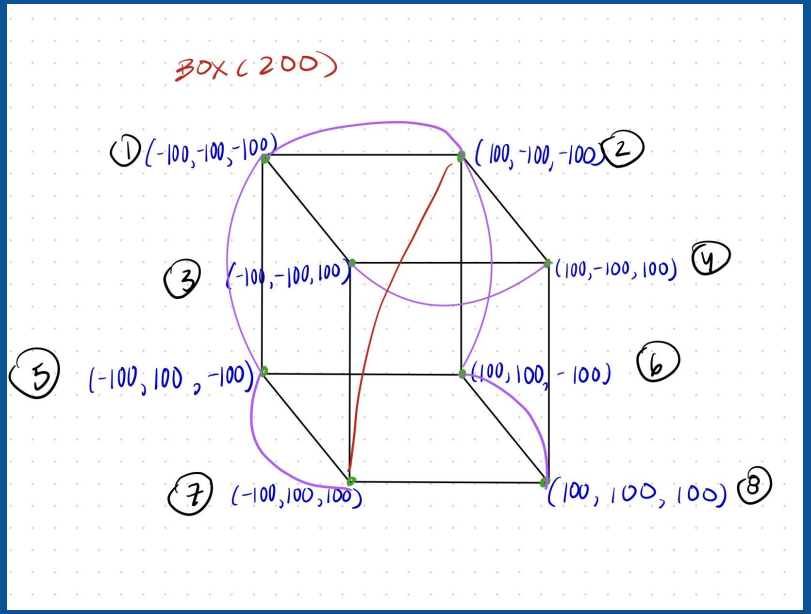
Since baseball is not a topic that everyone may be familiar with, I have created a codebook explaining the variables that are included in my 3D visualization. Hopefully this provides clarity and helps clear up any confusion for those interacting with my project.

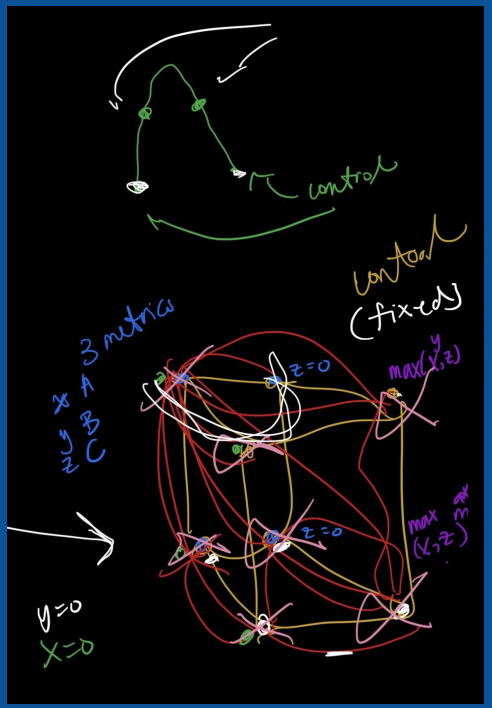
Link to Code Book PDF

Sketches

Here are some of the initial sketches that I had for Project # 3.







Final Idea

- 5 cubes with different x & y axes per cube
- Each cube measures the same z axis: winning percentage (key variable)
- Create a curve for each observation that goes through each cube
  - Begins at leftmost cube
  - Ends at rightmost cube
  - The curve is influenced by the differing x, y metrics of each cube
- Curves are colored by the age of the pitcher
  - I grouped pitcher’s age into 6 categories
  - Each category has a range of 5 years

Functions Added to 3D Visualization

- 'DIRECT' - plots points on the correspond x/y/z axes showing their respective direct relationships
    a. This will remove the curves from the graph
- 'LABELS' - removes the labels on the axes
- 'CUBES' - removes the cubes
- 'AGE_19_24' - removes points with 19≤age≤24 years old
- 'AGE_25_29' - removes points with 25≤age≤29 years old
- 'AGE_30_34' - removes points with 30≤age≤34 years old
- 'AGE_35_39' - removes points with 35≤age≤39 years old
- 'AGE_40_44' - removes points with 40≤age≤44 years old
- 'AGE_45_49' - removes points with 45≤age≤49 years old
- 'ROTATE_CAM' - rotates the camera view along the x-axis
- 'INITIAL' - positions the camera at the initial point of view

Final 3D Visualization