

# **ANALYSIS OF LOYOLA RAMBLERS GAMES (2016 - PRESENT)**

**CHARLES HWANG (LUC '20)**

# BACKGROUND

- LOYOLA RAMBLERS' "HEAD-TO-HEAD" PROGRAMS INCLUDE WOMEN'S SOFTBALL AND BOTH WOMEN'S AND MEN'S SOCCER, VOLLEYBALL, AND BASKETBALL
  - CROSS COUNTRY, TRACK & FIELD, GOLF, SPIRIT TEAMS EXCLUDED
- I WANTED TO RUN VARIOUS ANALYSES ON GAME DATA BECAUSE IT SEEMED THAT NO PREVIOUS STATISTICAL ANALYSES HAD BEEN PERFORMED

# VARIABLES INCLUDED

- TEAM
- SEASON
- TYPE (REGULAR SEASON, CONFERENCE TOURNAMENT, POSTSEASON, ETC.)
- LOCATION (HOME, AWAY, OR NEUTRAL)
- OPPONENT
- OPPONENT CONFERENCE
- OUTCOME (WIN, DRAW, OR LOSS)
- DAY OF WEEK
- MONTH
- SCORE FOR
- SCORE AGAINST
- LENGTH (REGULATION VS. OVERTIME, ETC.)
- OUTCOME OF PREVIOUS GAME
  - STREAK PRIOR TO GAME
- WHETHER OR NOT I ATTENDED

# EDA

- ALL DATA WERE COMPILED MANUALLY BY HAND IN EXCEL THEN READ INTO R AS A .CSV FILE
- THERE WERE 1,171 GAMES BETWEEN FALL 2016 AND APRIL 24, 2022
- VARIABLES SELECTED FROM THOSE EASILY AVAILABLE ON LOYOLARAMBLERS.COM
- CONFERENCE AT TIME OF GAME RECORDED
- DIFFERENT CONFERENCES PER SPORT
- ACCURACY CHECKS
  - MISSING DATA (NONE)
  - IMPOSSIBLE VALUES
  - SIMILAR VALUES (NORTHERN IOWA VS. NORTHERN ILLINOIS, MVC VS. MAC, ETC.)

# HYPOTHESES/QUESTIONS

1. WHICH VARIABLE(S) ARE THE MOST SIGNIFICANT FOR EACH TEAM?
2. DID EACH TEAM PERFORM DIFFERENTLY DURING THE SEASONS BEFORE, AFTER, AND THE SUCCESSIVE SEASON AFTER THE ONSET OF THE COVID-19 PANDEMIC?
3. DID TEAMS PERFORM DIFFERENTLY BEFORE AND AFTER A HEAD COACHING CHANGE?
  - WVB (AFTER 2017), SB (AFTER 2019), MBB (AFTER 2020-21)
4. DO TEAMS PERFORM DIFFERENTLY AT HOME THAN AWAY?
5. DO TEAMS PERFORM DIFFERENTLY WHEN I ATTEND?

# PREDICTIONS

1. LOCATION IS SIGNIFICANT
  - SEASON AND WIN/LOSS STREAK MAY BE SIGNIFICANT
2. TEAMS GENERALLY PERFORMED DIFFERENTLY IN EACH OF THE THREE SEASONS
3. WVB DID BETTER, SB AND MBB HAD NO STATISTICALLY SIGNIFICANT DIFFERENCE
4. TEAMS PERFORM BETTER AT HOME THAN AWAY
5. TEAMS PERFORM BETTER WHEN I ATTEND

# ANALYSES PERFORMED

1. GLM (STEPWISE SELECTION WITH AIC, BOTH DIRECTIONS)

2-3. (A LOT OF) TWO-PROPORTION Z-TESTS ( $\alpha = 0.05$ )

4-5.  $\chi^2$ -TESTS OF INDEPENDENCE ( $\alpha = 0.01$ )

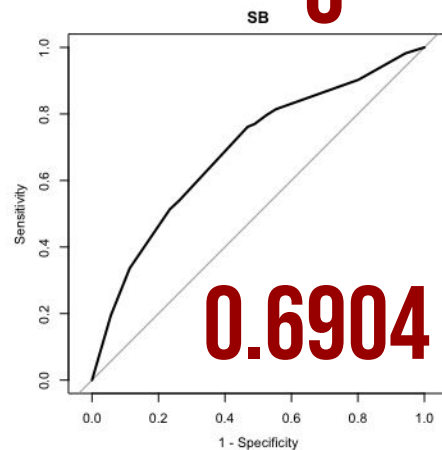
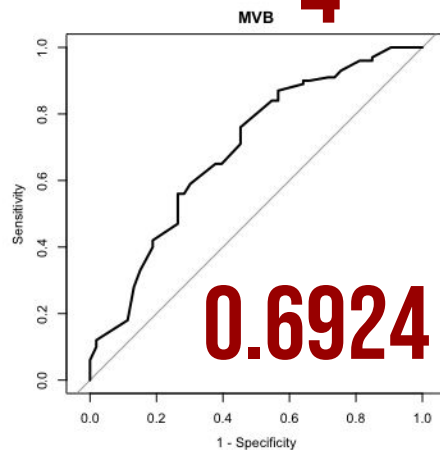
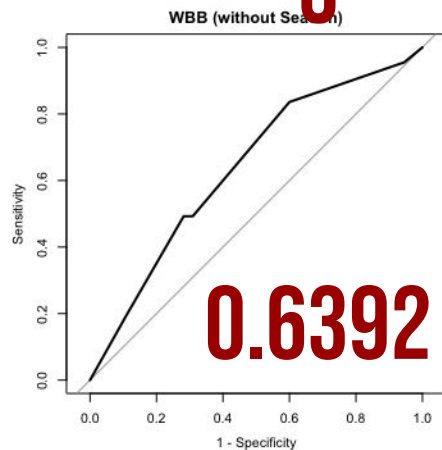
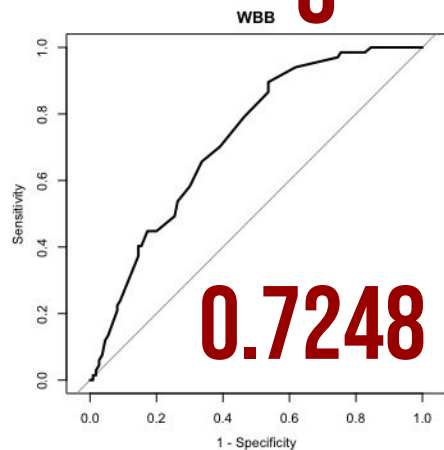
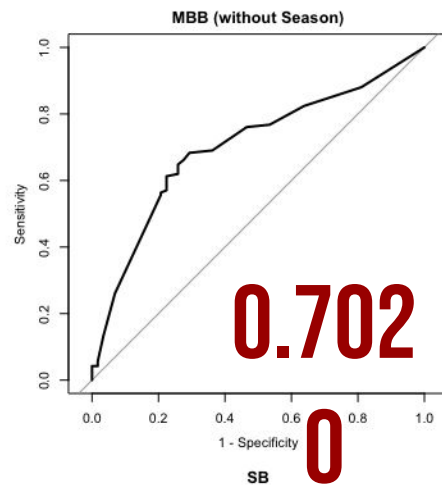
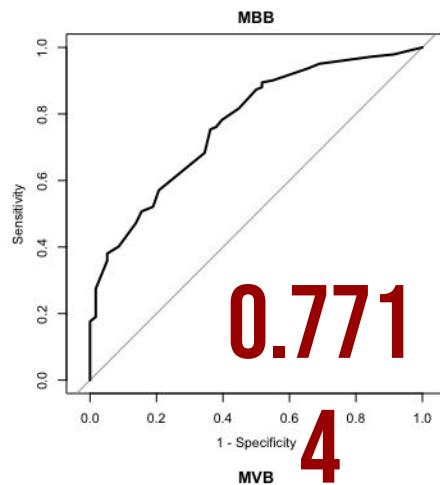
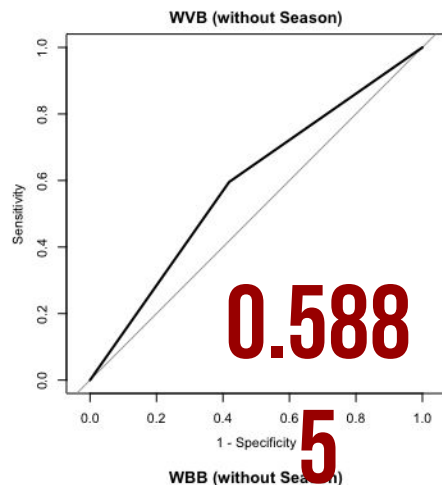
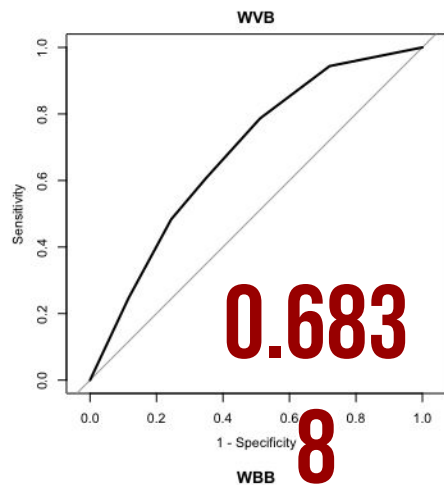
5. LIKELIHOOD RATIO  $\chi^2$  G-TEST, FISHER'S EXACT TEST

(NO MSOC OR WSOC)	SEASON	TYPE	LOCATION	MONTH	MATCH LENGTH	PREVIOUS OUTCOME	WIN/LOSS STREAK #	DID I ATTEND?	STEPS
WVB	'17, '21								9
WVB*						<u>0.019926</u>			8
MBB	(3)		<u>HOME</u> , N					0.084409	8
MBB*			<u>HOME</u> , N				0.121238		8
WBB	(5)						0.080987		7
WBB*			<u>HOME</u> , N			<u>0.008321</u>			6
MVB	(0)	REGULAR			<u>4, 5</u> SETS			0.055525	5
SOFTBALL			<u>HOME</u> , N	<u>0.016517</u>					9

\*WITHOUT SEASON VARIABLE

SCORE WAS EXCLUDED FROM ALL, OPPONENT AND CONFERENCE WERE EXCLUDED FROM SOME





## 2. TWO-PROPORTION Z-TESTS ON WIN % (YEARS 0, 1, AND 2 OF PANDEMIC)

	MSOC	WSOC	WVB	MBB	WBB*	MVB*	SB*
0 VS. 1	0.72546	0.77139	0.44357	<u>0.09625</u>	0.58592	<u>0.04106</u>	<u>0.04518</u>
0 VS. 2	0.61404	0.70755	0.79007	0.36922	0.52211	<u>0.06620</u>	<u>0.0879</u> †
*1+ WBB, 9+ MVB, 31+ SB GAMES CANCELLED IN 2020 † 8+ 2022 SB GAMES NOT INCLUDED (SEASON STILL ONGOING)	0.90572	0.97476	0.31647	0.42010	0.24023	0.72402	0.7561 †

### 3. TWO-PROPORTION Z-TESTS ON WIN % (BEFORE & AFTER HC CHANGE) <sup>†</sup>

	X-STATISTIC*	P-VALUE	WIN % BEFORE	WIN % AFTER	95% C.I.
WVB (2017)	14.5548	<u>0.0001361411</u>	0.3114754	0.6140351	(-0.449,-0.156 )
SB (2019)	3.199205	<u>0.07367408</u>	0.4901961	0.3762376	(-0.009, 0.237)
MBB (2021)	* $\chi^2(1) = Z^2$ <sup>†</sup> NO GAMES FOR MEN'S SOCCER OR WOMEN'S BASKETBALL	0.5006121	0.7005988	0.7575758	(-0.219, 0.105)

4-5. TWO-WAY FREQUENCY TABLES

	HOME	NEUTRAL <sup>1</sup>	AWAY		YES	NO	TOTAL
WIN	331	99	204	WIN	94	540	634
DRAW	10	<u>0</u>	9	DRAW <sup>2</sup>	<u>0</u>	19	19
LOSS	176	77	265	LOSS	43	475	518
TOTAL	517	176	478	TOTAL	137	1,034	1,171
WIN %	0.64990	0.56250	0.43619	WIN %	0.6861	0.5314	0.5495

	Outcome		
	W	D	L
Location			
H			
N			
A			

	Outcome		
	W	D	L
Attend			
Y			
N			

## 4-5. TWO-WAY FREQUENCY TABLES (MODIFIED)

	HOME	N+AWAY <sup>1</sup>
WIN	331	303
DRAW	10	9
LOSS	176	342
TOTAL	517	654
WIN %	0.64990	0.47018

	YES	NO	TOTAL
WIN+D <sup>2</sup>	94	559	653
LOSS	43	475	518
TOTAL	137	1,034	1,171
WIN %	0.6861	0.5406	0.5576

ODDS RATIO:  $(94 \times 475) / (559 \times 43) \approx$   
1.85755294

## 4-5. $\chi^2$ -TESTS OF INDEPENDENCE, LR $\chi^2$ G-TEST, FISHER'S EXACT TEST

	OUTCOME VS. LOCATION	OUTCOME VS. LOCATION <sup>1</sup>	LOCATION (LR $\chi^2$ G-TEST <sup>1</sup> )	OUTCOME VS. ATTENDANCE	OUTCOME VS. ATTENDANCE <sup>2</sup>	ATTENDANCE <sup>2</sup> (FISHER'S)
STATISTIC	50.15222	38.14837	39.37174	14.30914	10.38377	
P-VALUE	< 0.0000001	0.00000000 6	0.00000000 3	0.00078128 8	0.00127127 5	0.00131014 4
SIGNIF. CODE	***	***	***	***	**	**

<sup>1</sup>NEUTRAL AND AWAY GAMES GROUPED TOGETHER

<sup>2</sup>WINS AND DRAWS GROUPED TOGETHER

# CONCLUSIONS

1. LOCATION IS THE MOST SIGNIFICANT VARIABLE OVERALL (SEE #4) ✓
2. MOST TEAMS DID NOT PERFORM DIFFERENTLY BEFORE AND AFTER THE PANDEMIC  
✗
  - EXCEPTIONS: MVB ( $P = 0.04106$ ) AND SB ( $P = 0.04518$ ), 2020 VS. 2021
3. WVB DID BETTER, MBB HAD NO DIFFERENCE, SB DID WORSE (NOT SIGNIFICANT) ✓
4. TEAMS DID BETTER AT HOME ( $P < 0.00000001$ ) ✓
5. TEAMS DID BETTER WHEN I ATTENDED ( $P = 0.001310144$ ) ✓



# FUTURE IMPROVEMENTS

- MANUALLY CREATING THE DATASET WAS TEDIOUS BUT ALSO PRONE TO HUMAN ERROR
- ADDING OTHER VARIABLES AND INTERACTION TERMS
  - PRIVATE VS. PUBLIC, DAYS SINCE PREVIOUS GAME, DATE, TIME, DURATION, OFFICIAL ATTENDANCE, NATIONAL POLL RANKINGS, DISTANCE FROM LOYOLA, TICKET PRICE, ETC.
- ADDING OTHER TEAMS (FINAL RANKING/POSITION AS OUTCOME)
- FIGURING OUT A WAY TO CODE DRAWS IN GLM FOR SOCCER (1/2?)
- REVIEWING TEST ASSUMPTIONS (RANDOMNESS, INDEPENDENCE, ETC.)
- LOOKING INTO THE UNDERLYING REASONS FOR RESULTS/FURTHER ANALYSIS
- PREDICTIONS/MACHINE LEARNING

QUESTIONS?

