SEC LOSS CORRELATION

I. INTRODUCTION

1. <u>Question</u>: ***In 2024, what statistical categories were the most indicative of a loss in the SEC***
2. <u>Independent Variables</u>: **30 statistical categories listed in section III**
3. <u>Dependent Variable</u>: ***Losses***
4. <u>Methodology</u>
   - Exclusions
     - UGA & Texas were excluded because data included the SEC Championship Game
     - I didn't normalize stats based on opponent strength.
   - Assumptions:
     - 8 Games were enough to see a trend, given 30 statistical categories.
     - Removing UGA and Texas would not skew because between those two teams, there was only 1 reg-season loss.
   - Limitations:
     - Non-SEC applicability may be limited
     - Effects on analysis if UGA and UT were added is unknown

II. DATA OVERVIEW

1. <u>Source</u>
   - CFB stats (https://cfbstats.com/2024/conference/911/index.html)
2. <u>Organization</u>
   - Losses were in own column, to allow iteration for correlation in Python
   - Most categories are averages to help normalize, see snippet below

Team	Losses	Loser Points Per Game	Loser Total Yards Per Play	Loser Total Yards Per Game
Mississippi State	8	21.1	5.31	367
Vanderbilt	5	21.9	4.96	290
Arkansas	5	23.8	5.94	399.8
Kentucky	7	14.1	4.42	281.3
Oklahoma	6	16.5	4.2	278.6
Missouri	3	23.5	4.96	347.5
Mississippi	3	28.8	6.37	446.6
Texas A&M	3	29.4	5.76	385.5
Auburn	6	19.1	6.1	403.6
Florida	4	27.3	5.9	360.1
Alabama	3	28.4	6.16	395.6
LSU	3	26.5	5.6	407.4
South Carolina	3	29.1	5.78	379.4
Tennessee	2	25	5.45	397.1

III. DATA MANIPULATION/STATISTICAL PROCESS

1. Data Manipulation
   - Removed columns that stored strings, and stored integer-only data in a separate data frame.
   - Each of the individual stat columns were correlated to the loss column (Opponent QB Rating correlated with Losses, Loser Chunk Plays correlated with Losses, etc.)
2. Statistical Process
   - I utilized a hypothesis test to determine if there was statistical significance
   - Null Hypothesis: No linear relationship between the respective statistical category and a loss
   - Alternative Hypothesis: There is a linear relationship between the respective statistical categories and losses
   - I converted the correlations to t statistics, with a sample size of 14 (Number of teams examined), and 12 degrees of freedom.
   - Using the two-tailed t-distribution, I computed the P-Values shown below
   - If the P-value was less than 0.05, I rejected the null. Notice the asterisk next to the "Opponent Rushin Attempts Per Game" stat. This was because even though it has significance, teams typically rush more when they are winning to drain the clock – therefore I concluded that this stat does not help address potential causation of a loss.

	Variable	Correlation to Losses	P Value
0	Allowed Yards Per Pass Attempt	0.811676	0.000424
1	Loser Points Per Game	-0.795048	0.000677
2	Opponent QB Rating	0.776091	0.001101
3	Loser Average Turnover Margin	-0.741290	0.002413
4	Pass Completion Percentage Allowed	0.728756	0.003111
5	Points Per Game Allowed	0.728547	0.003124
6	*Opponent Rushing Attempts Per Game	0.597696	0.023985
7	Loser Pass Rating	-0.547603	0.042662
8	Loser Passing Completion Percentage	-0.525142	0.053819
9	Yards Per Game Allowed	0.519092	0.057154
10	Loser Total Yards Per Game	-0.515237	0.059356
11	Loser Rush Yards Per Game	-0.511478	0.061560
12	Yards Per Play Allowed	0.496393	0.071000
13	Opponent Rushing Yards Per Game	0.493376	0.073005
14	Loser Pass Yards Per Attempt	-0.477670	0.084093
15	Sacks Allowed Per Game	0.474415	0.086531
16	Opponent Pass Attempt Per Game	-0.470882	0.089233
17	Opponents Sacks Per Game	-0.465748	0.093263
18	Loser Total Yards Per Play	-0.452280	0.104434
19	Loser Average Rush Per Attempt	-0.435958	0.119165
20	Loser Time of Possession Per Game	-0.396040	0.160975
21	Chunk Plays Allowed	0.386650	0.172046
22	Loser Chunk Plays	-0.380457	0.179612
23	Loser Red Zone Score Percentage	-0.328313	0.251783
24	Opponent Pass Yards Per Game	0.318199	0.267550
25	Loser Pass Yards Per Game	-0.317881	0.268055
26	Loser Rush Attempts Per Game	-0.308958	0.282458
27	Opponent Average Per Carry	0.278452	0.335053
28	Oppenent Red Zone Conversion Percentage	0.063397	0.829525
29	Loser Pass Attempts Per Game	-0.036578	0.901201

IV. DATA VISUALIZATION

1. <u>Thought Process</u>
   - I wanted an easy way to see strength and contrast all in one graph
   - The first chart is meant to highlight the contrast of strength and direction
   - The second chart is meant to highlight the statistical categories that are significant (P-value of 0.05 or less)
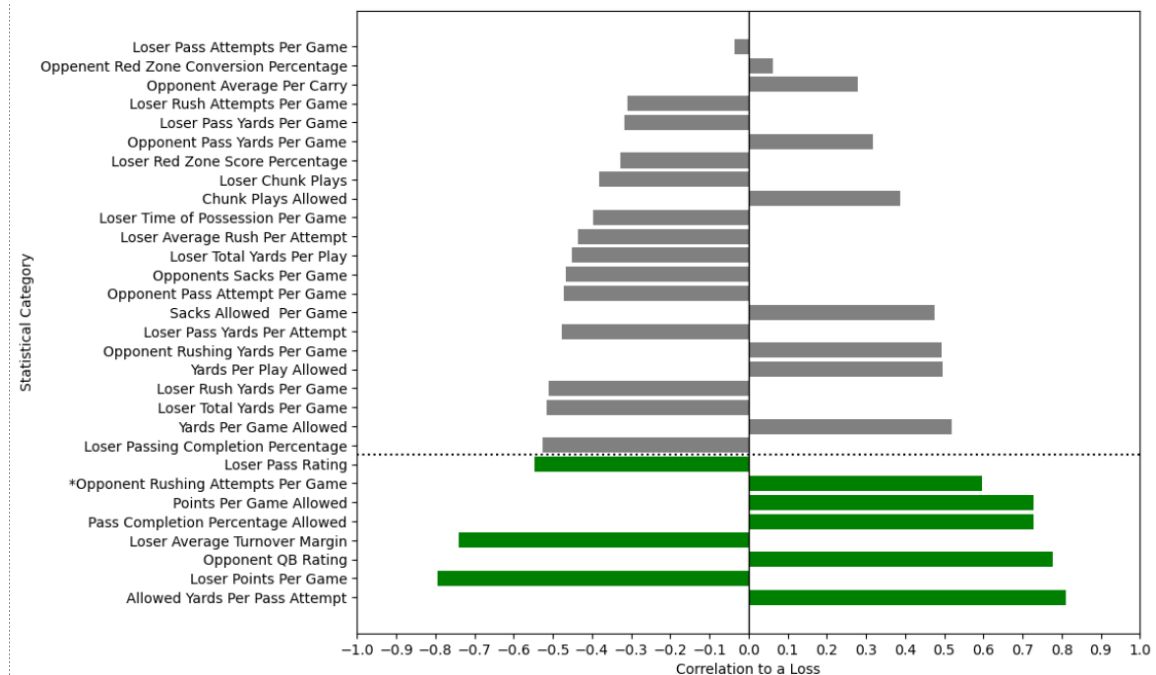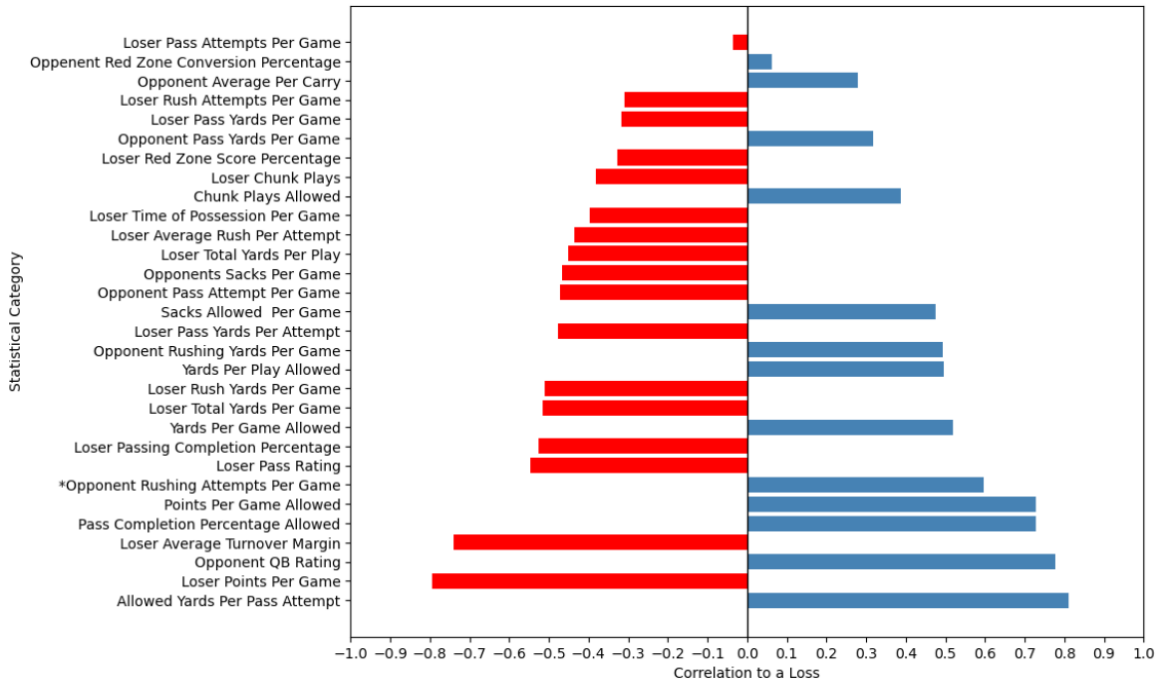
V. CONCLUSION

1. <u>Findings</u>
   - 8 of 30 statistical categories had a P-value of 0.05 or less
   - Again, I deemed the stat 'Opponent Rushing Attempts Per Game' irrelevant.
   - These 7 categories pass the commonsense test as well, consider 'Loser Pass Rating'. If your team has a poor pass rating, you didn't pass the ball effectively. In the SEC, one-dimensional 'run only' teams typically do not find sustained success because defenses can predict your plays and are athletic enough to sell out on defending the run – a lower passer rating appropriately negatively correlates to a loss.

2. <u>Most Significant Statistical Categories</u>
   - Original Question: In 2024, what statistical categories were the most indicative of a loss in the SEC

   - Top three significant statistical categories, in descending order
     - 'Allowed Yards Per Pass Attempt' – Passing is arguably the most effective way to get chunk plays (20+ yard gains on offense) since you can "go around" a defense and manipulate open space/favorable matchups. Whereas running the ball you typically must 'go through' a defense. So, if your team allows a lot of yards per pass attempt, the opponent has more opportunities to score, because they can effectively move into scoring position, by essentially going around your defense/exploiting defensive holes, with ta play that gains more yardage than a run typically does. This correlation indicates that the more passing yardage you allow on average, the greater the opportunities you give your opponent to score

     - 'Loser Points Per Game' – The main objective in football is to score more points than your opponent. This stat captures how many points your team scores every week, on average. This statistical category can be indicative of a deeper issue, offensive inefficiency. If your offense struggles to make trips to the redzone/endzone, you generally score less points. You also put pressure on your defense by forcing them to keep the game within reach by limiting opponent scores, while your defense grows in fatigue from staying on the field longer.

     - 'Opponent QB Rating' – This correlation makes sense again for reasons listed in the 'Allowed Yards Per Pass Attempt' explanation. If passing is arguably the most effective way to gain yardage into scoring position, having a QB who passes the ball very efficiently bolsters the significance of passing statistical categories. The higher your opponent's QB rating, the more effective your opponent is at passing the ball. The more effective your opponent is at passing the ball, the more yardage they gain without going through your defense, while simultaneously moving into scoring position.