I. INTRODUCTION

a. Question: ***In the College Football Playoff (CFP) era, can a collection of ESPN TOP 150 recruits be used to predict on-field success?***

- In College Football (CFB), the larger brand teams (Alabama, Ohio State, etc.) typically have a large stockpile of talent on their roster, but not every star-studded team wins its league. Consider Texas A&M 2022 recruiting class, the highest in that cycle – over the next 3 years they had an overall record of 20-18 and fired their head coach. Only one team can win a conference per year, only a handful of teams make the playoffs, and only one can be a national champion. So, with that in mind does having a large stockpile of talent correlate to a successful season?

b. Variables: (Independent) Weighted Talent Score, (Dependent) Success Categories

- Weighted Talent Score (WTS) is how I quantified the quality of a team given its ESPN Top 150 Recruit (ET150) collection and maturation, at a single point in time.
- The Success Categories I correlated the WTS to were metrics that I assumed would define "good" to "elite" success in any level of college football. From "9+ Win Season Count" to "CFP Appearances".

II. METHEDOLOGY

a. Data Overview / Manipulation

- Sources. The main source used was ESPN's Top 150 Recruit database, which can be accessed here. This database is a widely accepted collection of the best 150 football recruits in the nation. I used every single ET150 recruit from 2011 to 2024, to exclude recruits who did not play on the FBS level (Only FBS teams can participate in the CFP) or never enrolled. Reason being is that if I wanted to see a team's talent level in 2014 (the first year of the CFP era) I had to use the recruits from the 2011-2014 recruiting cycle because they all could have impacted the results of the 2014 season for a team. I also used the CFP website appearance table and the College Football Data API.
- Construction Process. Since I used every recruit from the database over the course of fourteen years, and each year houses 150 recruits – there were manyof data points. There were a total of 2100 recruits and over 70 teams examined in this study, which brought a challenge to data wrangling, cleaning, and storage. I started with combining every single recruit into one large database, which I referred to as "recruit_db". Then I would have to create a "success" database so I could compare the talent, and eventually correlate WTS to the success metrics. I referred to this database as "success_db". The last database worth mentioning that I had to create was a master database that housed important recruit and success information, and that was used for the correlation was called "correlation_df". You can reference snippets of each database below.

	Rank	Player	Class	Pos	Hometown	Home_State	Ht	Wt	\
0	3	Cameron Robinson	2014	OT	West Monroe	LA	6'5''	330	
1	6	Da'Shawn Hand	2014	DE	Woodbridge	VA	6'4''	262	
2	3	Eyabi Anoma	2018	DE	Baltimore	MD	6'6''	223	
3	5	Patrick Surtain Jr.	2018	CB	Plantation	FL	6'2''	192	
4	8	Tony Brown	2014	CB	Beaumont	TX	6'0''	196	

	Stars	Grade	School
0	5	95	Alabama
1	5	94	Alabama
2	5	94	Alabama
3	5	93	Alabama
4	5	92	Alabama

“recruit_db”

	School	ET150_Count	5_Star_Count	4_Star_Count	\
0	Michigan	62	3	59	
1	LSU	106	10	96	
2	Clemson	101	13	88	
3	Ohio State	127	20	107	
4	Georgia	141	25	116	

	80_Percent_ Win_Seasons_2014-2024	Conference_Chip_App_Since_CFP_Era	\
0	7	2	
1	7	2	
2	11	8	
3	10	5	
4	9	7	

	Conference_Chip_Wins_CFP_Era	CFP_Berths	CFP_Chip_Wins
0	2	3	1
1	1	1	1
2	8	7	2
3	5	6	2
4	3	4	2

“success_db”

	School	Weighted Talent Score	ET150_Count	5_Star_Count	4_Star_Count	\
0	Alabama	145.119318	194.0	33.0	161.0	
1	Arizona	1.404545	4.0	0.0	4.0	
2	Arizona St	1.159091	11.0	0.0	11.0	
3	Arkansas	3.343548	20.0	0.0	20.0	
4	Auburn	23.602381	70.0	7.0	63.0	

	80_Percent_ Win_Seasons_2014-2024	Conference_Chip_App_Since_CFP_Era	\
0	11.0	7.0	
1	2.0	1.0	
2	2.0	1.0	
3	1.0	0.0	
4	2.0	1.0	

	Conference_Chip_Wins_CFP_Era	CFP_Berths	CFP_Chip_Wins
0	7.0	8.0	3.0
1	0.0	0.0	0.0
2	1.0	1.0	0.0
3	0.0	0.0	0.0
4	0.0	0.0	0.0

“correlation_df”

b. Quantification

- To note, the "Weighted Talent Score" column in the data frame directly above, is an average WTS for a *season*, over the CFP era. It is used to determine the WTS for a team in a respective *season,* not for a recruiting class – the team's WTS for a season is simply a summation. To create the score, I used the following formula:

$$\textit{Weighted Talent Score}_{\textit{Team, Season}} = \sum \textit{Weighted Talent Score}_{\textit{Previous 4 Recruiting Cycles}}$$

$$\text{(i.e. } WTS_{2020\ \text{Season}} = WTS_{2017\ \text{Class}} + WTS_{2018\ \text{Class}} + WTS_{2019\ \text{Class}} + WTS_{2020\ \text{Class}} \text{)}$$

- The reason for a distinctive Season WTS and a Recruiting Class WTS, is so I could observe how talented a team is projected to be at any point in the CFP era, and to build a CFP era average per team. Having a separate WTS just for a team's recruiting class allows segmented observations, and the average I was looking for. The talent weights are 1.0 for 5-star recruits, and 0.75 for 5-star recruits. This is because 5-star recruits are expected to be transformative talents, while 4-star recruits are just on the edge of the elite. Below is how the WTS for a is calculated for a *recruiting class*:

$$\textit{Weighted Talent Score}_{\textit{Team, Recruiting Class}} = \textit{(Total ET150 Recruits) (Talent Factor) (Maturation Weight)}$$

Total ET150 Recruits = Total count of ET150 recruits for a team in a recruiting class
Talent Factor = Total of 5- and 4-star recruits multiplied by talent weight (5 stars x 1.0, 4 stars x 0.75)
Maturation Weight = Total 5- and 4-star recruits multiplied by maturation weight (Freshman x 0.7, Sophomore x 0.9, Junior x 1.0, Senior x 1.0)

- The maturation weight added extra steps to my analysis. Since recruiting classes for a team are observed as an object via the WTS, I needed a way to apply the necessary maturation weights. To do this, I multiplied every calculated **recruiting class WTS** by the years that were relevant for this study. For example, for Auburn's 2020 recruiting class WTS, to get the effect of a recruit's maturation – I had to multiply Auburn's talent factor of the 2020 class by 0.7, 0.9, 1.0, and 1.0. To do that I subtracted the difference between the recruiting class year and the actual season. If that difference was between 0 and 3, that means the resulting WTS includes the full maturation of a recruiting class:

	Class	School	4	5	Talent Factor	Season	Year Diff	\
0	2011	Alabama	10.0	1.0	8.50	2014	3	
1	2011	Arkansas	3.0	0.0	2.25	2014	3	
2	2011	Auburn	7.0	1.0	6.25	2014	3	
3	2011	California	5.0	0.0	3.75	2014	3	
4	2011	Clemson	5.0	2.0	5.75	2014	3	

	Maturation Weight	Weighted Talent Score
0	1.0	93.50
1	1.0	6.75
2	1.0	50.00
3	1.0	18.75
4	1.0	40.25

c. Scope / Limitations / Assumptions

- Scope. This study uses WTS season averages to find correlation with totaled frequencies. Since that's the case, the data and analysis can only be used to address an overall correlation (2014-2024) and not a year over year analysis. This study is not meant to predict yearly correlations, rather just an overall correlation over the CFP era. It is a predictive study.
- Limitations. I did not have the capability to gather transfer portal information for 2000+ recruits, therefore WTS for teams do not consider the loss of a ET150 recruits (this will be addressed in findings). Additionally, only one iteration of the 12-team playoff model has been conducted, so findings may be less representative as more iterations of the new playoff model occur.
- Assumptions. The maturation weight is used to account for a recruit's development overtime. To note, the maturation weights of juniors and seniors are both 1.0. This is because I assume that as an ET150 recruit matures into a junior, they are as likely to have an impact on their team as a senior. This is because when ET150 recruits are brought into a program, their talent allows them to see the field sooner, rather than having to wait 4 years (consider Ryan Williams, Jeremiah Smith, Travis Hunter, etc.). Additionally, since my data does not track the progression of players (whether they transferred, declared for draft, redshirted), I had to group a class all as one. So, in my data, the entire 2020 class for Auburn all theoretically graduated after the 2024 season and did not transfer. I assumed this would not skew my data because of the large quantity of players, teams, and seasons. Also, my scope is prediction, trying to assume if a team's projected talent would correlate to future winning. I also assumed that quality players, at large, do not need to transfer, and if they declared for the draft early, I weighed them equally to seniors so there is not a large discrepancy.

d. Statistical Process

- Hypothesis Test. I used a two-tailed hypothesis test to determine if there was enough evidence to support an association between ET150 recruit acquisition and on-field success. See below for null and alternative hypotheses:

H_0 = There is no association between a team's WTS and on-field success metrics

H_A = There is association between a team's WTS and on-field success metrics

(Success metrics include seasons with 80 percent win percentage or better, conference championship appearances, conference championship wins, CFP berths, and CFP championship wins)

- Coding Steps and Results. I used a model from a previous study to determine significance. With my correlation data frame, as shown in step b as 'correlation_df," I was able to correlate the WTS with the success metrics using the ".corrwith" function, and the results where stored into another data frame called "correlation_results_df". Using a previously used model of mine, I ran the correlations through a for loop to get the t statistic. The CDF function was then used to calculate my p values, which I stored in one list. See below for correlation and p value computation:

V) CORRELATE AVG SEASON WTS TO SUCCESS METRICS

[10]:
```
#1) Drop all columns except WTS and success metrics
correlation_df2 = correlation_df.copy()
correlations = correlation_df2.drop(columns = ['Weighted Talent Score','School','ET150_Count','5_Star_Count','4_Star_Count'])\
.corrwith(correlation_df2['Weighted Talent Score'])
# print(correlations)

#2) Tidy
correlations_results = correlations.copy()
correlations_results_df = correlations_results.reset_index()
correlations_results_df.columns = ['Stat','Correlation']
print(correlations_results_df)
```

```
                              Stat  Correlation
0  80_Percent_ Win_Seasons_2014-2024     0.596077
1  Conference_Chip_App_Since_CFP_Era     0.557501
2       Conference_Chip_Wins_CFP_Era     0.631791
3                         CFP_Berths     0.803782
4                      CFP_Chip_Wins     0.862081
```

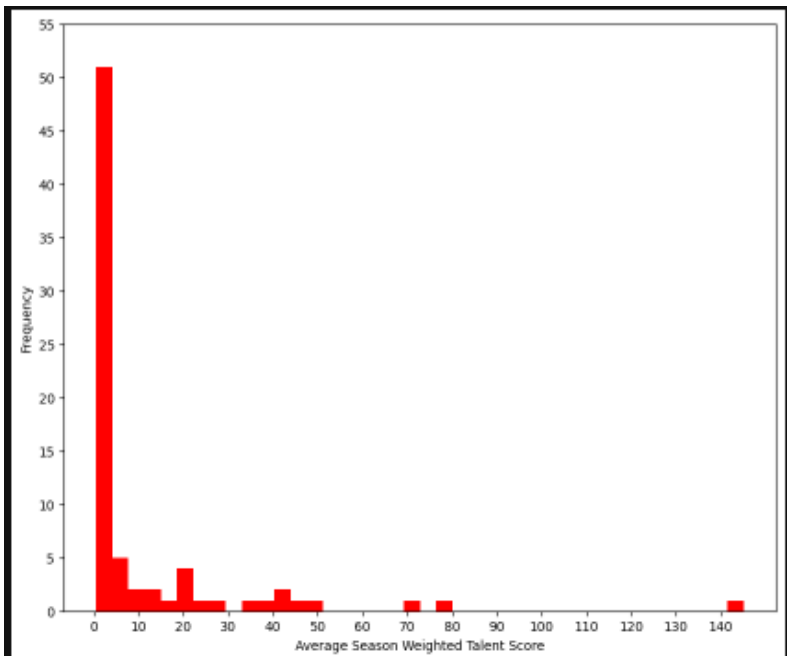
VI) STASTICAL SIGNIGFICANCE

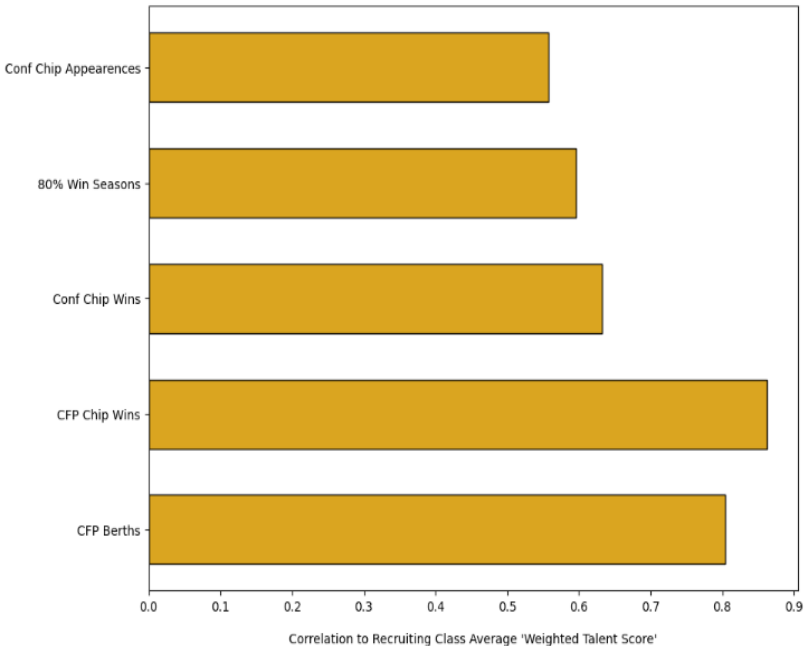
[11]:
```
#1) t test to see significance
pvalues = []

for corr in correlations_results_df['Correlation']:
    r = corr                                          # correlation coefficients
    n = 73                                            # teams examined
    tstat = r * math.sqrt((n - 2) / (1 - r ** 2))     # T-stat
    dfree = n - 2                                     # degree of freedom
    p = 2 * (1 - t.cdf(abs(tstat), dfree))            # P value
    pvalues.append(p)

correlations_results_df['P Value'] = pvalues
correlations_results_df = correlations_results_df.copy()
sorted_cm_df = correlations_results_df.reset_index(drop = True)
sorted_cm_df = correlations_results_df.sort_values(by = 'P Value', ascending = True)
print(sorted_cm_df) #all stats below shown to be significant at 5% level
```

	Stat	Correlation	P Value
3	CFP_Berths	0.803782	0.000000e+00
4	CFP_Chip_Wins	0.862081	0.000000e+00
2	Conference_Chip_Wins_CFP_Era	0.631791	2.049229e-09
0	80_Percent_ Win_Seasons_2014-2024	0.596077	2.637946e-08
1	Conference_Chip_App_Since_CFP_Era	0.557501	3.001166e-07

III. DATA VISUALIZATION/FINDINGS

a. Visual Analysis. As shown in the graphic above, all statistical categories are significant. The average WTS per team is extremely left skewed, while all statistical categories had a moderate or greater correlation. The interesting thing about this observation is that it shows that most teams don't recruit ET150 recruits at a large, while there is an exclusive elite category of teams that have high WTS. The teams that do have elite level WTS are also the teams that typically win the win championships (UGA, Alabama, Ohio State, etc). Here's what the data looks like:





b. Analysis. Here is what can be inferred from the data

- Amateurism. College football is not a professional sport, not just legally but also based on play. NFL players are more technically sound football players because they've played high level football (College and Pros) longer, have better training resources, and have more time to dedicate solely to the game. Compare this to a college football player who at the minimum is a couple months removed from high school and still must balance education with their training. Given that college football players are younger in their careers and have time constraints, the more talented players a team collects, the more they can widen the talent gap between themselves and their opponents. Also consider, recruiting is not leveled like in the NFL where teams have a salary cap and a draft system. In college, the teams with the most money rake in the best recruits because players are attracted to winning programs and the benefits that come with being in those programs – like winning national championships. This helps explain why CFP berths and CFP Championship wins are so strongly correlated with WTS. The higher the count of ET150 recruits, the higher the WTS – and the higher the WTS, the higher the ability to predict if a team will be successful. Coach Nick Saban, the greatest CFB coach once said, *"I figured out a long time ago that I couldn't coach bad players to play good, so the key to the drill was to go get some good player."* Even in the transfer portal era where recruits can quickly leave a team if they don't immediately play, this saying of coach Saban's stands true, specifically teams that are elite winners. This is because those teams can build a stockpile of talented players who are prepared to play. Consider 4 teams: Ohio State, Georgia, Auburn, Penn State. Ohio State and Georgia have an approximate average WTS between 71 and 77 – both teams have won multiple CFP national championships. Now compared to Auburn and Penn State, who both have an approximate average WTS between 13 and 23. Neither Auburn nor Penn State have won a CFP championship. Additionally, in the CFP era, Georgia and Ohio State have won their conferences 3 and 5 times respectively, while Auburn and Penn State have won their conferences 0 and 1 time respectively – Georgia and Auburn are in the same conference, and Ohio State and Penn State are in the same conference. In college football, during the CFP era, having the most talented players, especially at a higher proportion, can accurately predict on field success because good players rise to the moments of winning, especially against other players they are more talented than.

IV. REFERENCES

a. https://www.espn.com/college-sports/football/recruiting/rankings/scnext300boys/_/class/2015/view/espnu150/order/true
b. https://collegefootballplayoff.com/sports/2021/9/7/team-appearances
c. https://api.collegefootballdata.com/