Executive Summary

The machine learning / predictive analytics portion of the project involved two parts: 1. We used machine learning models & a number of formulas to predict probabilities of a given team winning against another team. 2. Using these probabilities we built out a simulation of the entire World Cup, and ran the simulation 100,000 times. The Netherlands is the team most likely to win the World Cup according to our simulations. Morocco, Mexico, Senegal, and Belgium are some of the teams that our simulation had winning the World Cup more often than what Odds Makers expect. England, France, Brazil, and Argentina are less likely to win the World Cup according to our simulations than when compared to Vegas Odds. The simulation only gives the United States a 40% chance of getting out of the Group Stages - we hope we're wrong here!

The Tableau section consisted of two separate sections; the findings of the Machine Learning and the Historical Tournaments data. The findings of the Machine Learning were showcased against the Vegas odds. Our odds favored less South America teams and had the Netherlands with the highest percentage to win the upcoming World Cup whereas the Vegas odds had Brazil with the best odds. As for our Historical data, we created a story for each of the AFC Asian Cup, AFCON, Copa America, and the Euros. For each story, we explored the attendance, the top scorers per competition, player and team statistics, and top winners by competition/tournament. The Historical data was separate from the Machine Learning and was sourced from Kaggle. The Historical section was more to have viewers' explore the world of Futbol/Soccer and how big the World Cup is in comparison to the individual tournaments hosted.

Our Database branch was based off of our Machine Learning results. First, the data was transformed via Python/Pandas. Then we were able to use our knowledge of SQLite and create a queryable Database that allows the user to select whichever team they would like and what round you want to see(in example, selecting the Qatar in the Round of 16). The Database table will be updated and show all simulated matches(from Machine Learning results). Below is an example image of how that would look.

Group	Stage	Squad 1	S1%	Seed	Squad 2	S2%	Seed	S1_Prob	S1_wins	Simulation	Winner
R16	Knockout	Qatar	0.6010889696	1A	Iran	0.6970299892	2B	0.3957536492	1	2	Qatar

(limited to 100 results). The table will include all other simulated Qatar vs *insert team here* and their respective statistics.

The Webpage segment of our project was mainly piecing together the Tableau, Database, and Machine Learning files into their respective HTML and app.py. To do so, we created an HTML page for each section. We also made sure that the app routes were properly in place so that our webpage would properly load. The biggest issue we faced was adding the Database and Machine Learning to our website. As for our color scheme, we wanted to use the upcoming World Cup colors and keep our webpage clean so it wouldn't take away from our findings.