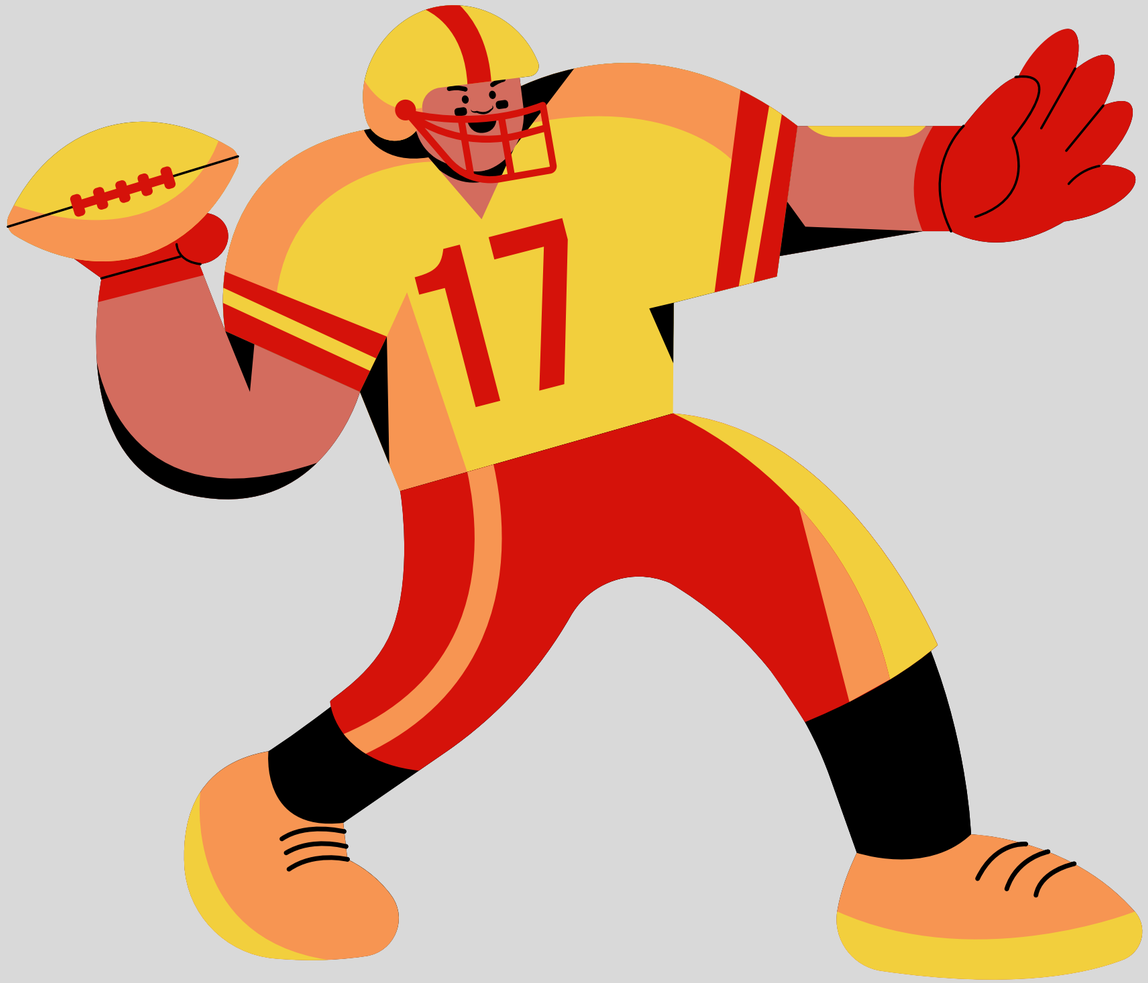


SUPERBOWL PREDICTIONS

Team: Ankit Jain, Marianna Carini, Karl Hickel, Allen Lee, Mira Daya

Model Summary (Ankit)

Methodologies:

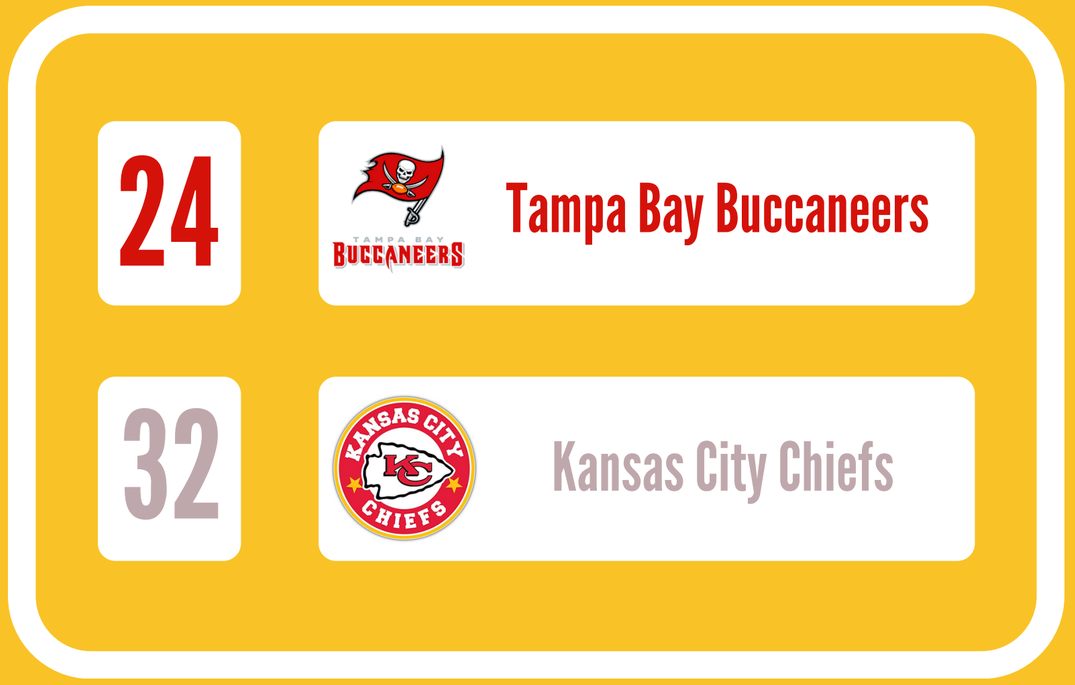
- Polynomial Regression, with different degrees of some variables based on linear patterns, noticed.
- Variables based on statistics of team and opponent based on matches previously played in the season.

Model Metrics:

- Score_Chiefs: R2=0.999, MAPE = 2.23%
- Score_Bucs: R2=0.7507, MAPE = 10.80%

Next Steps:

1. Consider data about players and previous seasons
2. Fine-tune model to predict to score of new teams with zero previous statistics



Model Summary (Mira)

Methodologies:

1. Feature selection using forward stepwise linear regression to determine important variables
2. ETS, simple exponential smoothing, to predict each variable inputs for the final regression model
3. Linear regression models for each team using selected features and forecasted variable inputs

Model Metrics:

- Score_Chiefs: R2=0.918 MSE=14.370 MAPE = 6.067%
- Score_Bucs R2=0.9241 MSE=10.747 MAPE = 23.49%

Next Steps:

1. Gather more data, ex: consider home/away
2. Fine-tune model to include more significant variables
3. Watch SB55!

26	Tampa Bay Buccaneers
32	Kansas City Chiefs

Model Summary (Allen)

Methodologies:

1. Used Boruta algorithm to identify important predictors
2. Played around with different predictors to build linear models
3. Utilized ARIMA to predict for selected predictors then fed into linear model for score prediction

Model Metrics:

- LM_Chiefs: R2=0.89 MAE=1.78 MAPE = 6.7
- LM_Bucs: R2=0.82 MAE=3.97 MAPE = 20

Next Steps:

1. Consider using more data; for TB model only 2020 data was used since it depended heavily on QBR.
2. Try to capture more aspects of the game; right now models are mainly focused on offense.

28	Tampa Bay Buccaneers
30	Kansas City Chiefs

Model Summary (Karl)

Methodologies:

1. Combined data from 2018-2020 seasons for both Tampa and Chiefs.
2. Used association by correlation to determine statistically significant variables and used them in a regression.
3. Used Arima on variables that were applied in linear regression to get predicted inputs for the game on Sunday and plugged them in.

Model Metrics:

- Score_Chiefs: R2=0.38 MAPE = 15.3%
- Score_Bucs R2=0.76 MAPE = 25.2%

Next Steps:

1. Gather more data from previous seasons.
2. Try a different form of regression.
3. Consider different variables like weather.

29	Tampa Bay Buccaneers
31	Kansas City Chiefs

Model Summary: Success Rate + In-game stats (Marianna)

Methodologies:

1. Manually selected variables using p-values and stepwise linear regression
2. Built model for each team using success rate metrics along with in-game stats for 2020 season
3. Prediction based on season averages for in-game statistics (running, passing, turnovers and current success rates

Model Metrics:

- Score_Chiefs: R2=0.40 MAPE = 0.135
- Score_Bucs R2=0.82 MAPE = 0.238

Next Steps:

1. Consider adding more variables in model that help with KC model R2
2. Impute in-game yards using ARIMA and/or ETS

30	Tampa Bay Buccaneers
22	Kansas City Chiefs

Model Summary: Success Rate Only (Marianna)

Methodologies:

1. Manually selected variables using p-values and stepwise linear regression
2. Built model for each team using success rate metrics only for 2020 season
3. Prediction based on current success rates

Model Metrics:

- Score_Chiefs: R2=0.47 MAPE = 0.184
- Score_Bucs R2=0.68 MAPE = 0.285

Next Steps:

1. Consider adding more variables in model that help with KC model R2
2. Gather historic success rate data
3. Make model for all teams

34 Tampa Bay Buccaneers

29 Kansas City Chiefs

Model Summary: "Lazy Man's" Ensemble (Marianna)

Methodologies:

1. Collected expert score predictions for Super Bowls back to 2015
2. Reviewed MAPE by source and year. Determined that the average of the sources is the most reliable
3. Prediction based on weighted average depending on source's historic MAPE

Model Metrics:

- MAPE = 0.171

Next Steps:

1. Add more historic data on score predictions
2. Try linear regression with source as variable

26	Tampa Bay Buccaneers
31	Kansas City Chiefs

Ensemble Modelling

- Weighted Average based on MAPE

Model	Source	KC	Tampa	MAPE	Weight KC	Weight TB
Lazy Man's (Avg.)	Marianna	32	26	0.17	0.12	0.13
Lazy Man's (MAPE Avg.)	Marianna	31	26	0.17	0.12	0.13
Success Rate Only	Marianna	29		0.18	0.12	0.00
Success Rate Only	Marianna		34	0.29		0.11
Success Rate + In-game Stats (Avg.)	Marianna	22		0.14	0.12	0.00
Success Rate + In-game Stats (Avg.)	Marianna		30	0.24		0.12
Poly. Reg	Ankit	32		0.02	0.14	
Poly. Reg	Ankit		24	0.11		0.14
Lin. Reg	Mira	32		0.06	0.13	
Lin. Reg	Mira		26	0.23		0.12
Lin. Reg	Karl	31		0.15	0.12	
Lin. Reg	Karl		29	0.25		0.12
Lin. Reg	Allen	30		0.07	0.13	
Lin. Reg	Allen		28	0.20		0.13

Score	30	28

Ensemble Model Prediction

28	Tampa Bay Buccaneers
30	Kansas City Chiefs

Kansas City Chiefs will win Superbowl 55 against Tampa Bay Buccaneers with a score of 30 to 28.