For my project, my original focus is to investigate the relationship between various football statistics and the outcomes of Kansas City Chiefs' games, with a specific focus on the seasons since 2022. After doing a linear regression I saw all the variables were not significant. Because of this, I needed to expand the data to have more points for a better fit and better results. I changed from just the 2022 chiefs season to data for the 2018 season through the 2022 season. I chose this time frame to capture the team's performance under consistent quarterback leadership, which is a critical factor in game outcomes. To answer my research question, I used linear regression, logistic regression, and k-fold cross-validation.

Linear Regression: I used linear regression to explore the relationship between different game statistics (like first downs, passing yards, rushing yards, turnovers) and the scoring margin (the difference between the Chiefs' score and their opponents'). This technique is good for quantifying the strength and direction of relationships between variables. In my analysis, we found that none of the predictors were statistically significant at the 0.05 level, indicating a need to refine our model. This is why I expanded our data.

```
Call:
lm(formula = ScoringMargin ~ OffPassY + OffRushY + OffTO + DefPassY +
    DefRushY + DefTO, data = chiefs_data)

Residuals:
    Min      1Q  Median      3Q     Max
-10.153  -3.574  -1.610   4.928  11.983

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.05956   15.28551  -0.200    0.844
OffPassY     0.05620    0.03327   1.689    0.115
OffRushY     0.06369    0.03960   1.608    0.132
OffTO       -3.23310    2.34664  -1.378    0.192
DefPassY    -0.02584    0.02382  -1.085    0.298
DefRushY    -0.06260    0.05894  -1.062    0.308
DefTO        2.31432    2.00833   1.152    0.270

Residual standard error: 7.387 on 13 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.3967,    Adjusted R-squared:  0.1183
F-statistic: 1.425 on 6 and 13 DF,  p-value: 0.2776
```

Logistic Regression: This technique is used to predict the binary outcome of a game (win or loss). Logistic regression is appropriate for our analysis as it deals with binary dependent variables and can provide insights into which factors are significant predictors of winning or losing. I'm still working on this part of my analysis to be usable.

```
Call:
glm(formula = WinLoss ~ OffPassY + OffRushY + OffTO + DefPassY +
    DefRushY + DefTO, family = "binomial", data = chiefs_data)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.140e+02  3.390e+05       0        1
OffPassY     6.775e-01  2.097e+03       0        1
OffRushY     7.594e-01  1.638e+03       0        1
OffTO       -3.950e+01  8.885e+04       0        1
DefPassY    -2.616e-01  8.071e+02       0        1
DefRushY    -3.047e-01  1.610e+03       0        1
DefTO        2.271e+01  1.001e+05       0        1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1.6908e+01  on 19  degrees of freedom
Residual deviance: 6.8218e-10  on 13  degrees of freedom
  (1 observation deleted due to missingness)
AIC: 14

Number of Fisher Scoring iterations: 25
```

K-Fold Cross-Validation: To assess the reliability and performance of my models and to lessen the risk of overfitting, we are using k-fold cross-validation. This method ensures that my model's performance is robust across different subsets of the data. I'm still working on this part of my analysis to be usable.

```
Generalized Linear Model

20 samples
 6 predictor

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 16, 16, 16, 16, 16
Resampling results:

  RMSE      Rsquared   MAE
  0.443567  0.4902677  0.2657195

>
```