Predictability of Heart Rate

What can we look at for heart rate responses? A research study on the DII Men's Soccer Team at Seton Hill University.

By Jonah Zembower

Outline

- Initial Problem
- Data Preparation
- Methodology
- Results
- Conclusion
- Final Thoughts

Initial Research

Current research is limited on the effects of training stimulus and game stimulus on the heart rate of men's soccer players at colleges.

Very little research focuses on differences in training stimulus based on positions on the soccer team

Midfielders

Forwards

Center Backs

Outside Backs

Goalkeepers

Study Aims

- Assess positional differences
- Assess baseline measurement differences
- Look at predictive factors of heart rate
- Understand the variability of heart rate
- Find improvements and why they occur

Data Collection

Polar H10 Heart Rate Monitors:

- Assess two-second interval heart rates

Baseline Measurements:

- Bioelectrical Impedance Analysis (BIA)
- Cooper 1.5 Mile Run Test

Datasets

Heart Rate

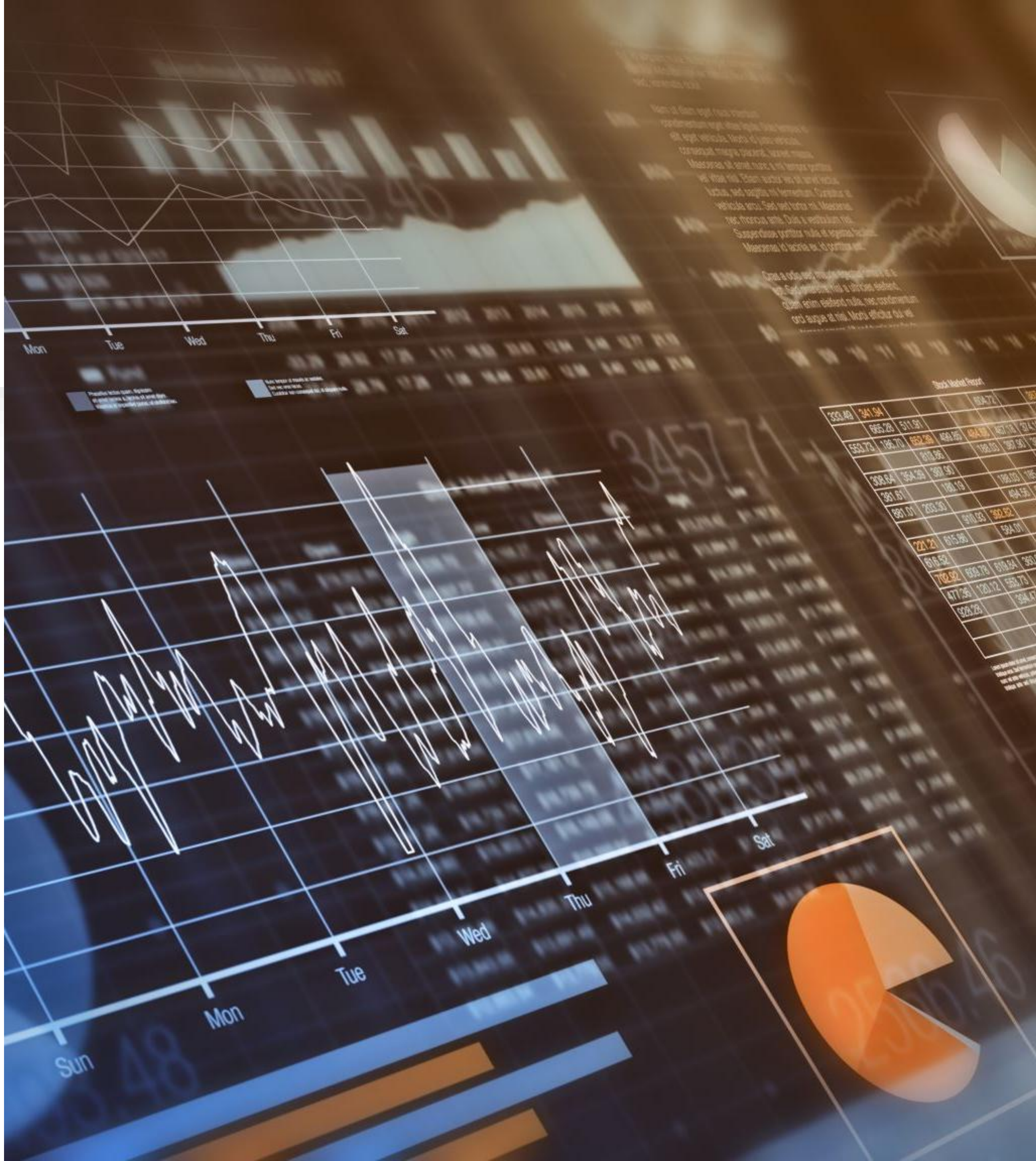
2024-01-24	2024-01-26	2024-01-27	2024-01-28	2024-02-02	2024-02-03	2024-02-04
146	143	136	119	131	116	108
146	143	136	119	131	120	105
146	142	138	119	131	120	104
145	135	140	119	131	120	105
137	127	145	119	131	120	105
132	124	146	119	130	120	105
130	125	146	121	130	117	105
128	127	146	121	127	117	102
127	127	144	124	128	117	102
124	128	145	125	135	118	100
119	134	144	125	135	118	103
123	134	144	123	135	118	103
126	135	147	123	135	117	103
128	127	147	123	135	118	103
123	127	146	122	135	119	103
121	128	146	126	135	120	118
113	133	145	127	135	122	118
117	134	145	127	135	122	119
122	129	145	127	135	122	119
125	127	145	127	135	123	119
125	128	149	127	143	123	109
130	130	152	130	143	123	112
117	129	155	133	141	122	114
112	130	156	136	139	121	115
112	131	159	136	137	119	115
115	129	159	136	133	119	120

Baseline Measurements

Height	Age	Weight (lbs)	Fat%	Fat Mass (lbs)	FFM (lbs)	Muscle Mass (lbs)	TBW (lbs)	TBW%
5'6	18	177.4	19.6	34.8	142.6	135.6	102.8	57.9
5'11	19	163.8	12	19.6	144.2	137	101	61.7
5'6	19	133	11.1	14.8	118.2	112.2	86.4	65
6'0	19	161.4	11.7	18.8	142.6	135.6	98.6	61.1
6'0	18	146.2	9.7	14.2	132	125.4	90.8	62.1
5'11	19	154.4	10.9	16.8	137.6	130.8	96.4	62.4
5'7	19	136.6	11.6	15.8	120.8	114.8	86.8	63.5
5'3	19	141.4	18.1	25.6	115.8	110	85.8	60.7
6'0	18	156.2	10.6	16.6	139.6	132.6	96.8	62
6'0	19	153.2	11.7	18	135.2	128.4	92.4	60.3
5'11	19	175	13.9	24.4	150.6	143.2	105.2	60.1
5'11	20	169.2	15.1	25.6	143.6	136.4	99.2	58.6
5'11	20	184.8	19.1	35.2	149.6	142.2	103.6	56.1

Data Challenges

- Inconsistent Data
- Missing data points/incorrect readings of the heart rate
- Keeping baseline measurement collection consistent
- Synthesizing data correctly for analysis purposes



Ethics of Data

IRB Approved Research Study

Informed Consent Given

Data Privacy

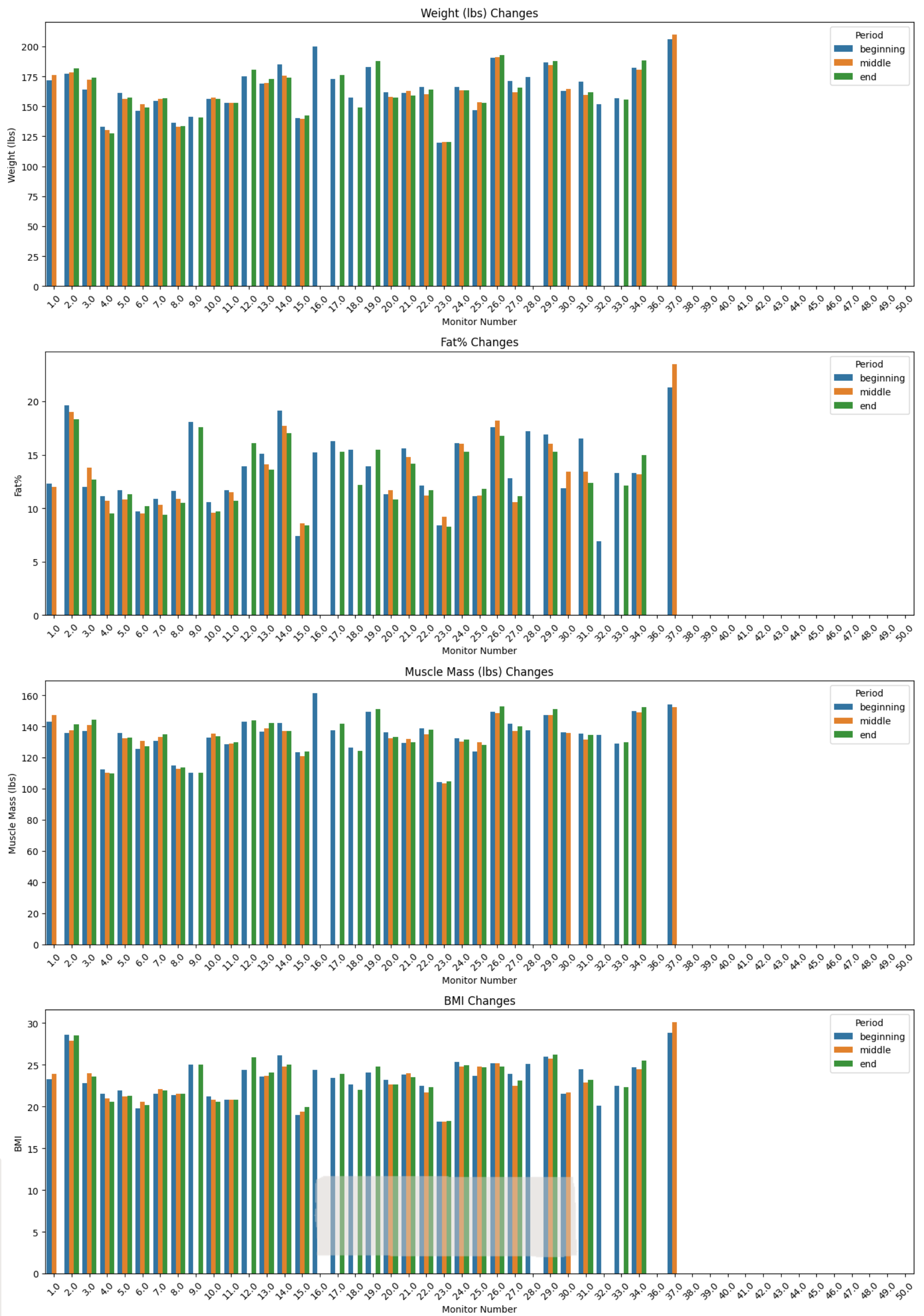
- Secured data
- Only authorized personnel from informed consent had access to data

So…Solving the Problem

- How do we understand that there appears to be positional differences?
- How can we find the baseline differences?
- What are the ways to understand the predictability of the heart rate?
- **Ho: There is no difference in heart rate between each position.**
- **Ha: There is a difference in heart rate between each position.**

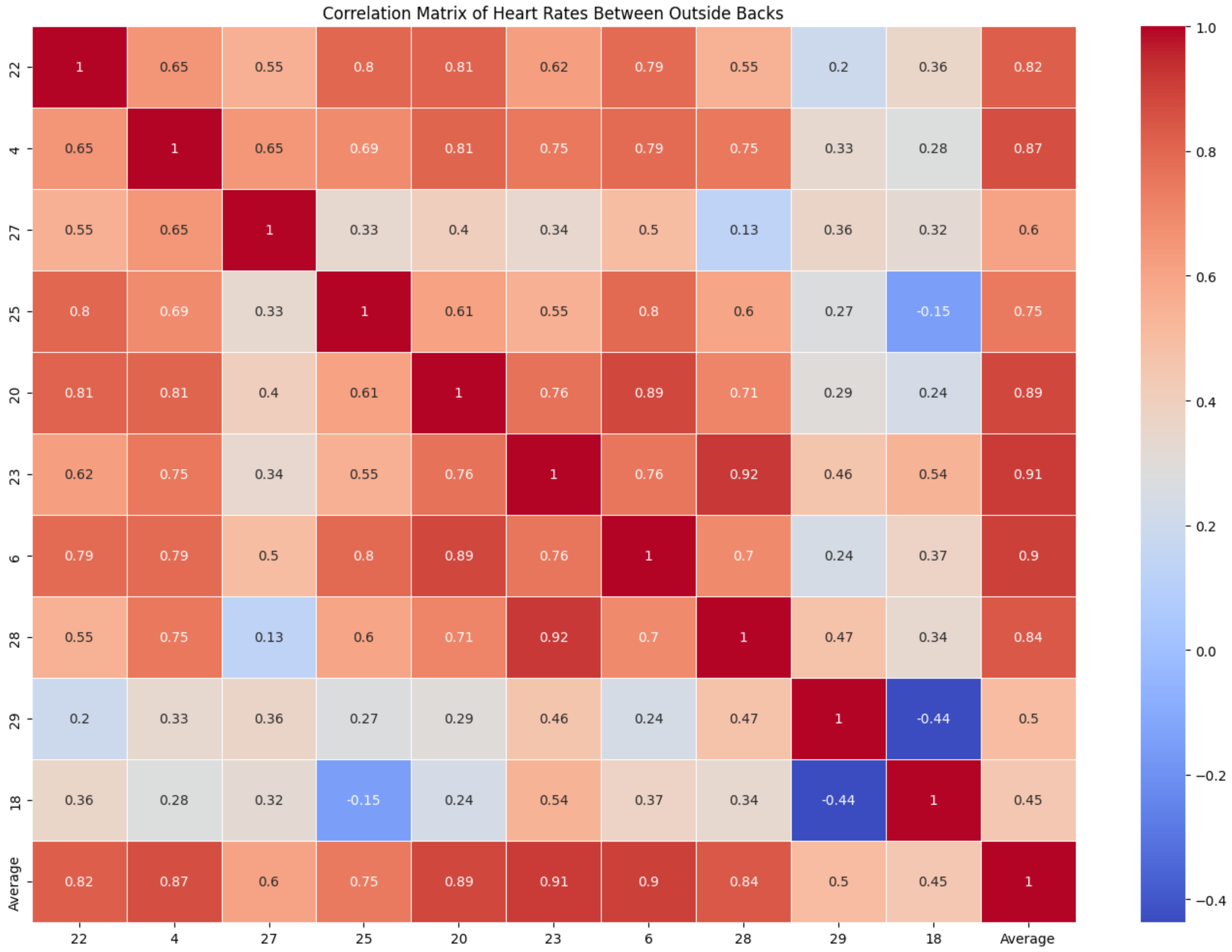
Baseline Measurements EDA

- Body composition metrics
    - Weight
    - Fat%
    - Muscle Mass (lbs)
    - BMI Changes

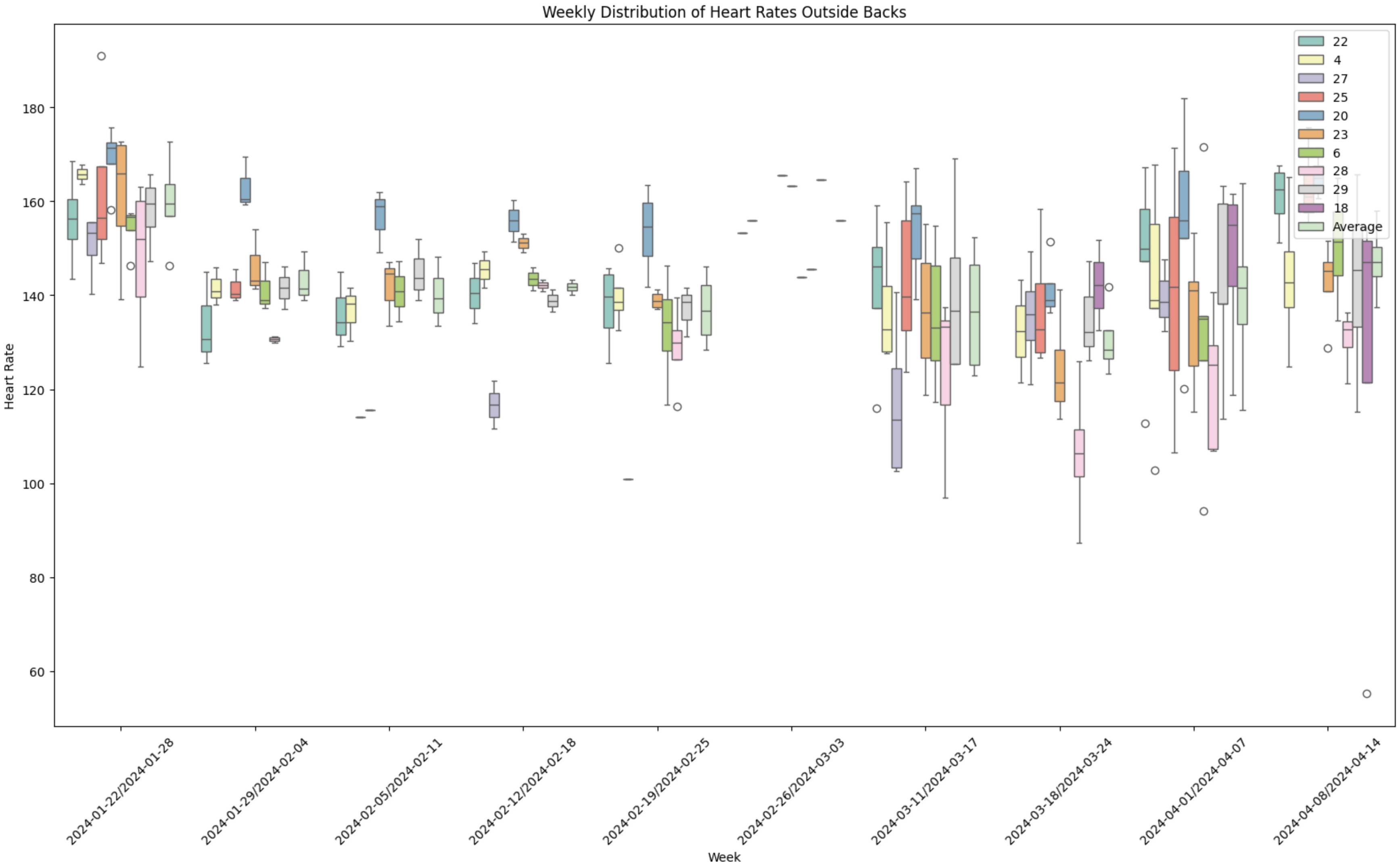


Heart Rate EDA

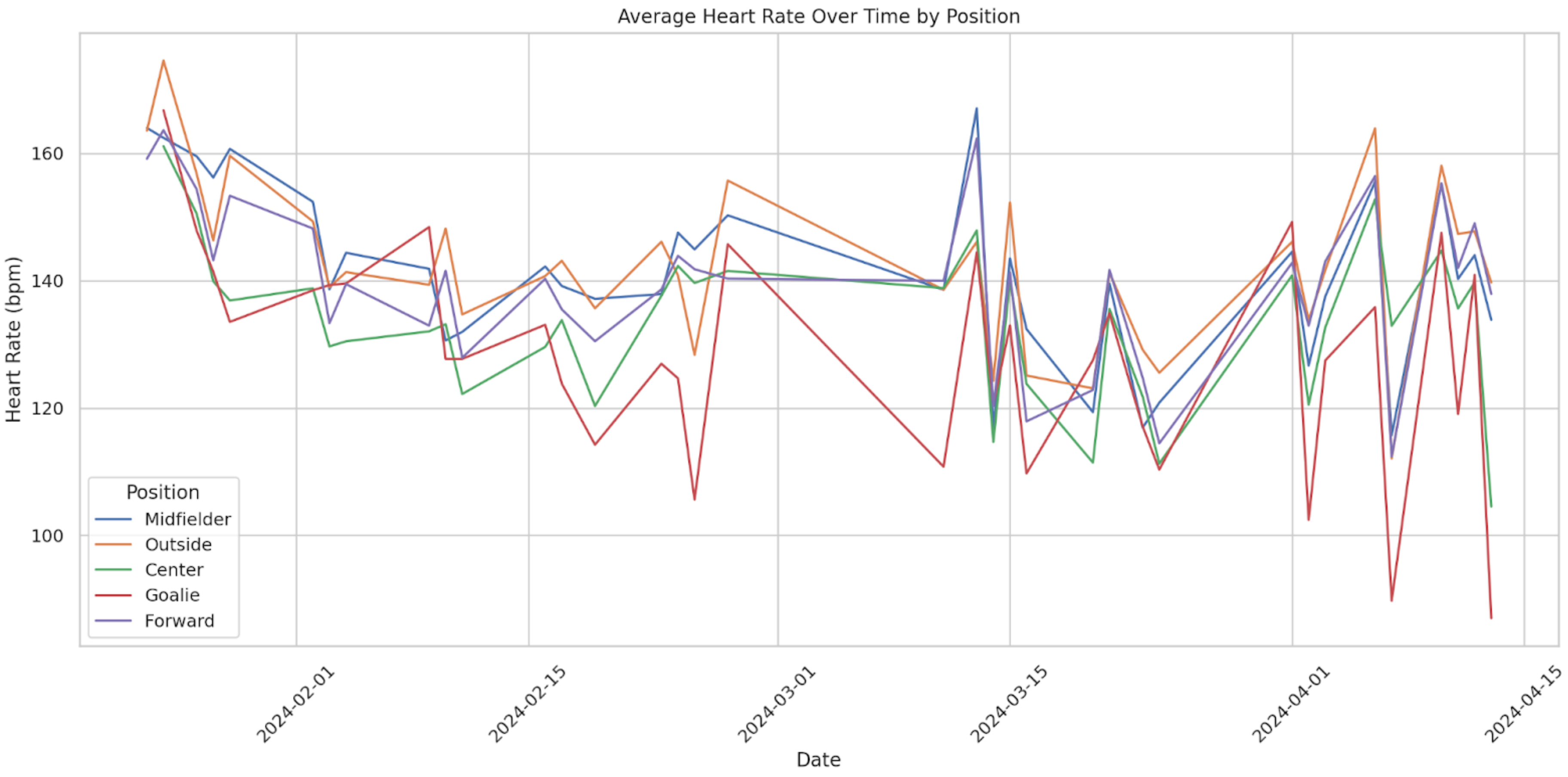
Correlations of Outside Backs



Weekly Distributions of Heart Rate



Average Heart Rates Over Time



Functions/Code

Use of .describe()

- This gives the mean, std, min, max, quartiles, etc. for each column.

.isnull().sum()

- This was to look at the different null values that I needed to account for.

I also used an extensive amount of these packages:

- Matplotlib
- Pandas
- Scikit learn
- Seaborn

Feature Engineering

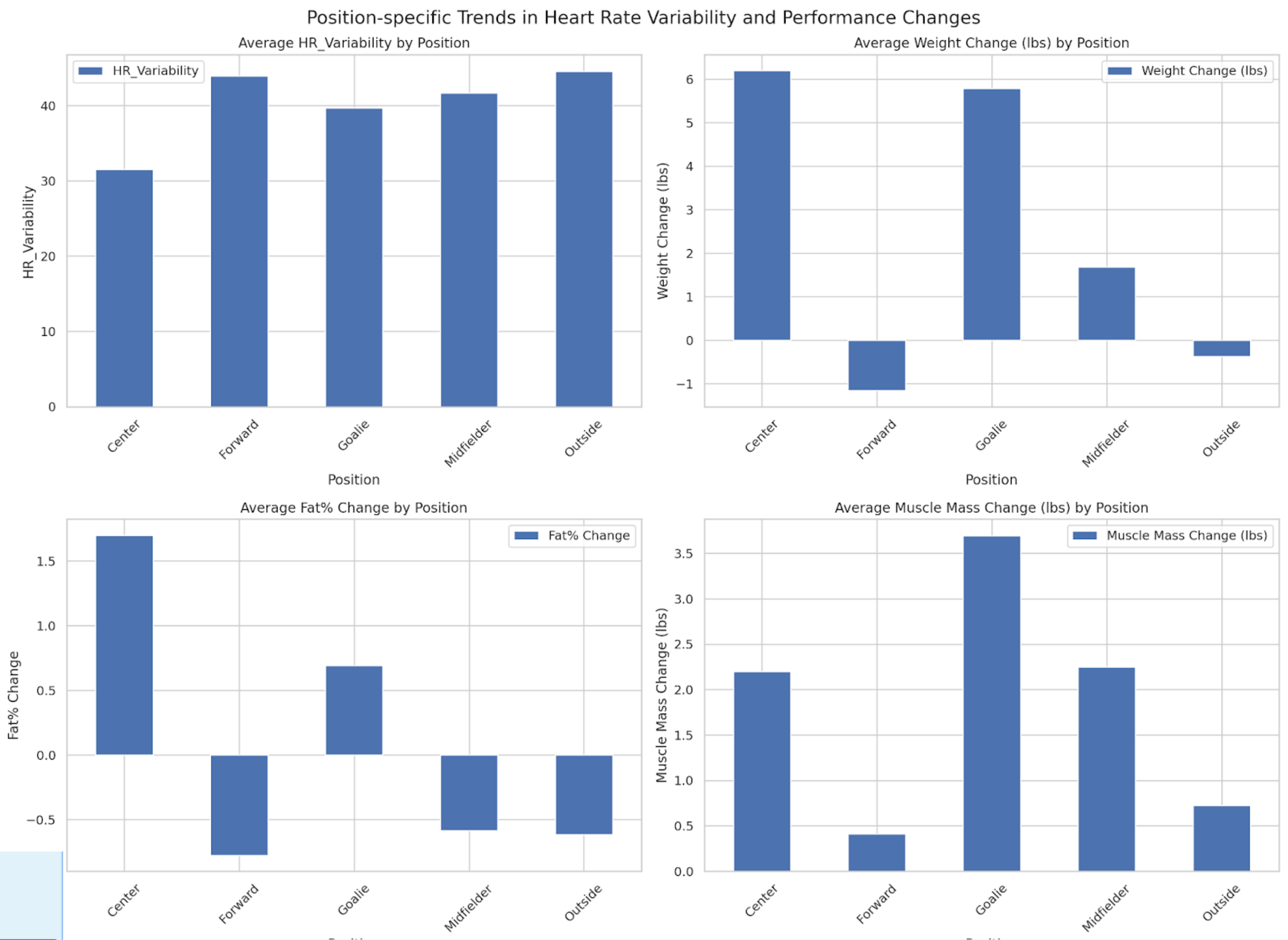
Focused on heart rate of each session using each value instance or average heart rate.

Focused on baseline measurement results

Looked at AR, MA, Variance, and Theoretical HRV

Trends in HRV and Baselines

- Average HRV around 40 bpm
- Weight seemed to increase or decrease depending on the position a little
- Fat% generally decreased
- Muscle Mass increased



P-values for Baseline Body Composition Measurements

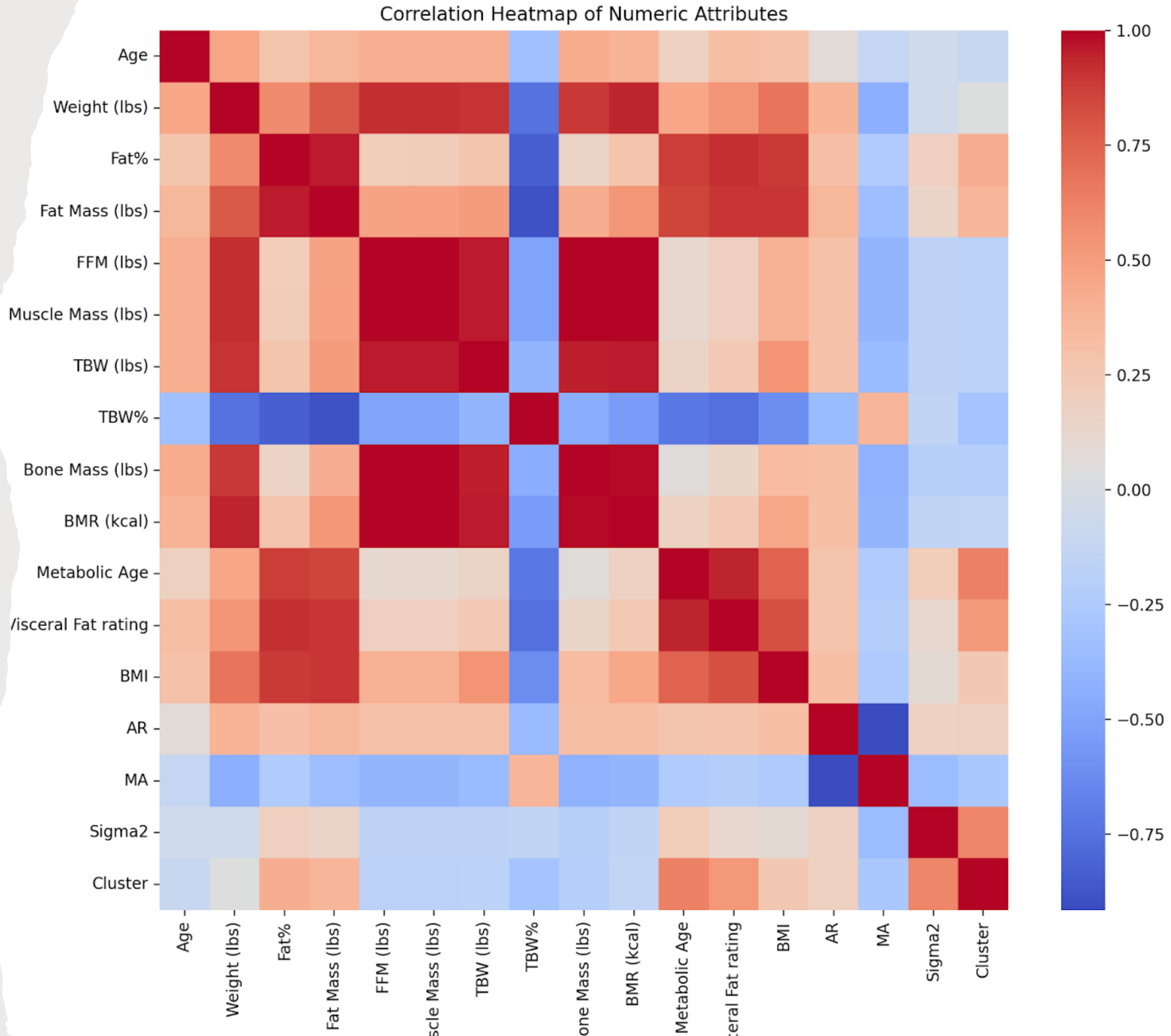
Metric	p-value
Fat%	0.0108906829
Fat Mass (lbs)	0.0411494906
FFM (lbs)	0.07772360932
Muscle Mass (lbs)	0.07858809668
TBW (lbs)	0.007675595323
TBW%	0.004836955053
Bone Mass (lbs)	0.09014848157
BMR (kcal)	0.1636817277
Metabolic Age	0.06449231728
Visceral Fat rating	0.05619167866
BMI	0.9569173959

Cohen's d Effect Sizes for Body Composition Changes

Metric	Cohen's d Effect Size
Weight Change	0.1191
Fat% Change	-0.1222
Muscle Mass Change	0.2487
BMR Change	0.2312

Correlation of Tests and HR

- Low correlation for AR, MA values
- High correlations between baseline measurements



Heart Rate Significant Differences

- Found statistically significant different heart rates from the beginning to end of the spring semester
- Found statistically significant differences across the heart rates of positions

ANOVA Results Across Positions

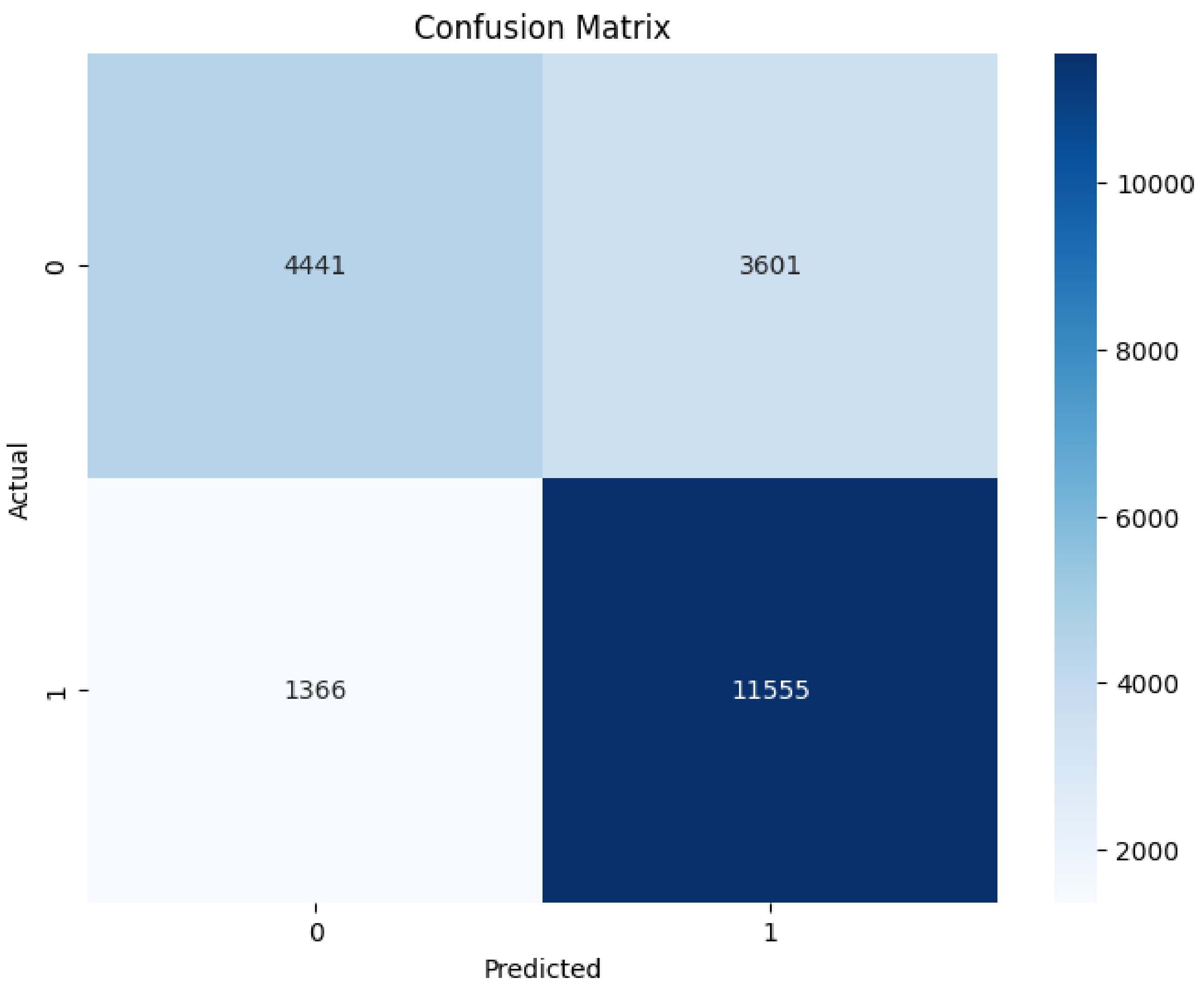
Statistic	Value
F-statistic	1144.96
P-value	0.0

T-Test Results by Position

Position	t-statistic	p-value
Midfielder	31.758	3.620086120454103e-189
Outside Back	41.809	0.0
Forward	43.987	0.0

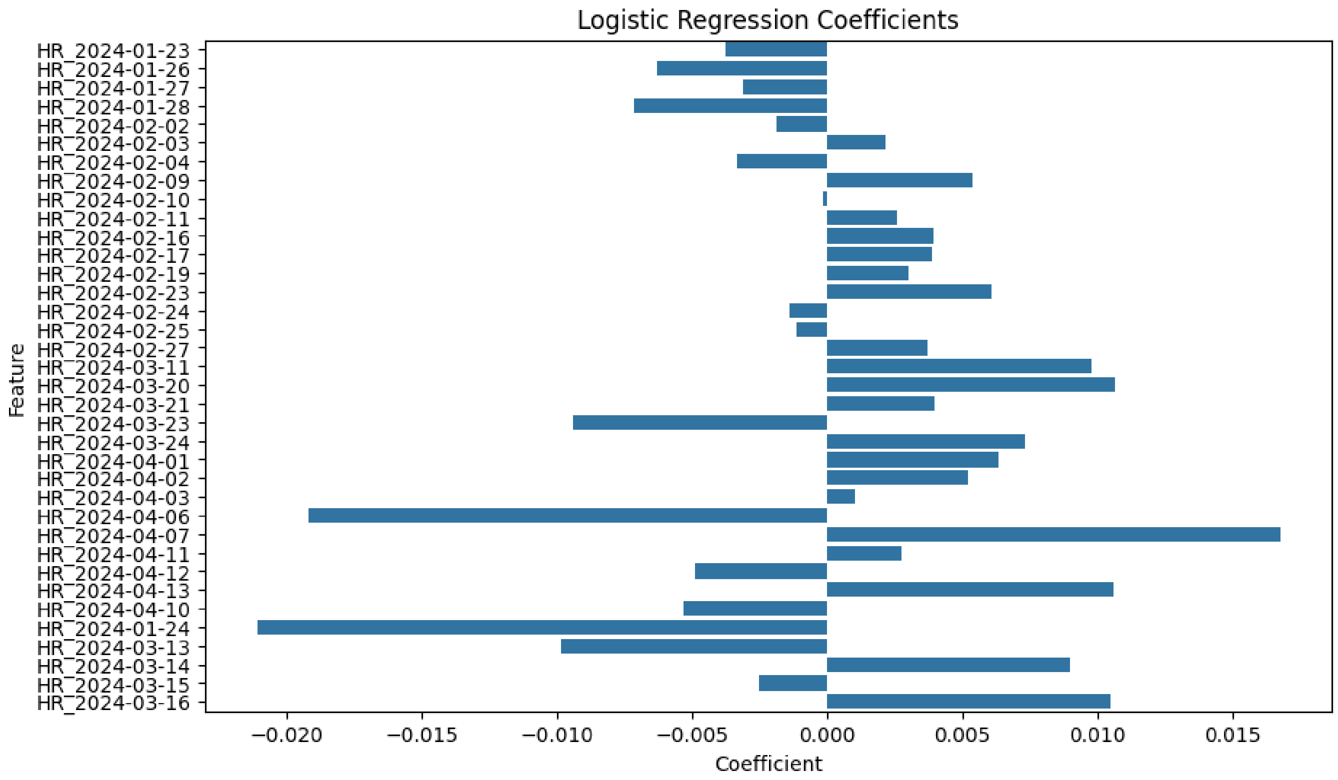
Logistic Regression Modeling

- True Positives: 11,555
- False Positives: 3,601
- True Negatives: 4,441
- False Negatives: 1,366



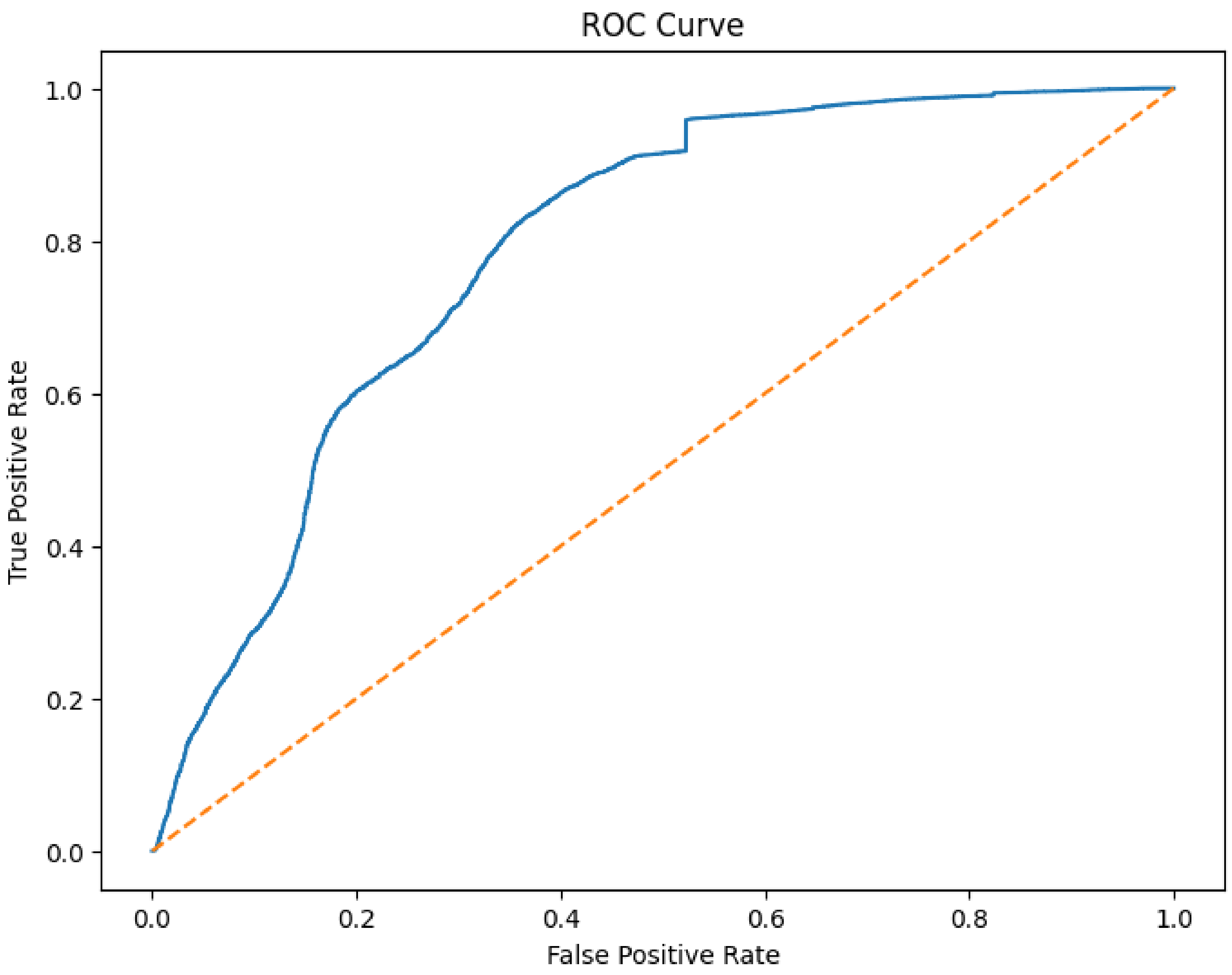
Logistic Regression Coefficients

- Positive increases likelihood of target class.
- Negative decreases the likelihood of target class.



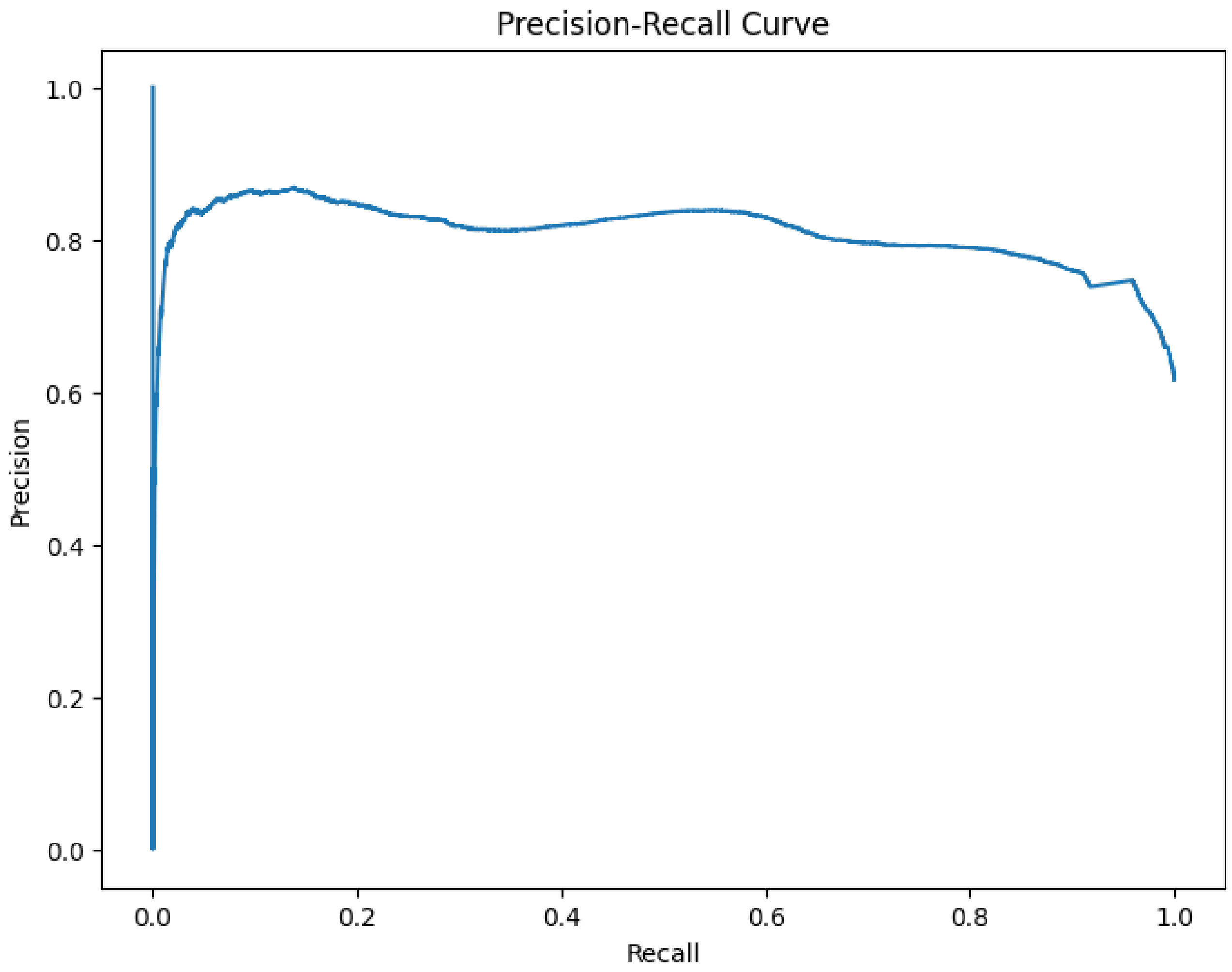
ROC Curve

- The TPR appears to be about 0.8 at a FPR of about 0.2.
- Showcases some potential predictability.



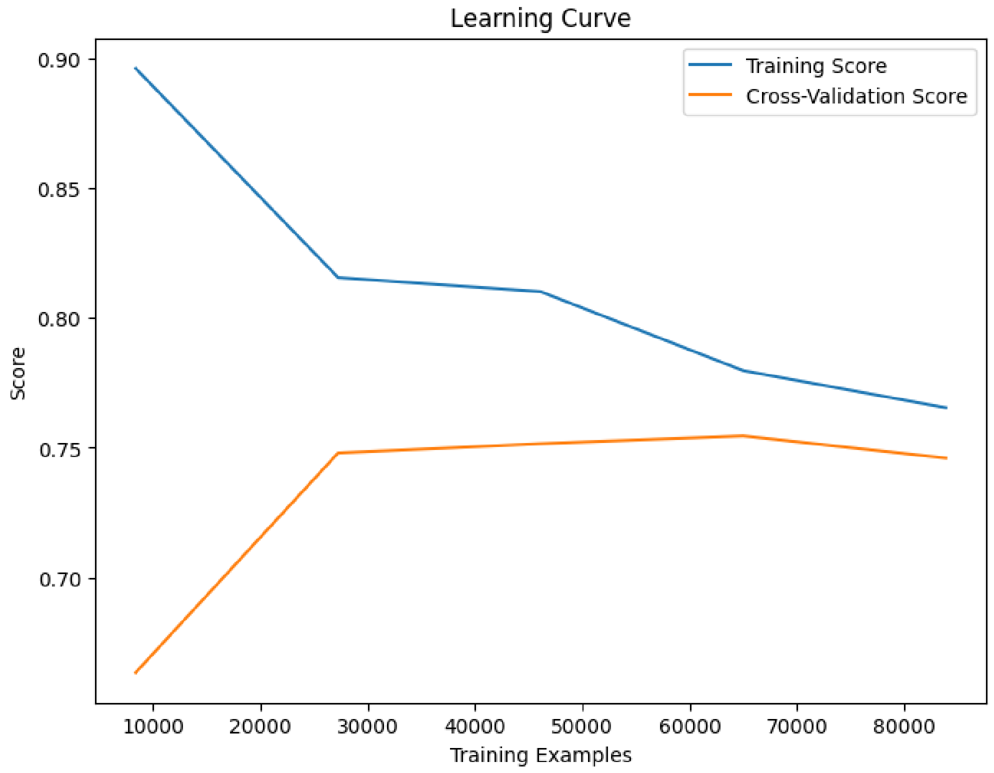
Precision-Recall Curve

- Stabilizes at about 0.8 precision throughout recall
- Significant drop in precision as it approaches recall values of 1.0.



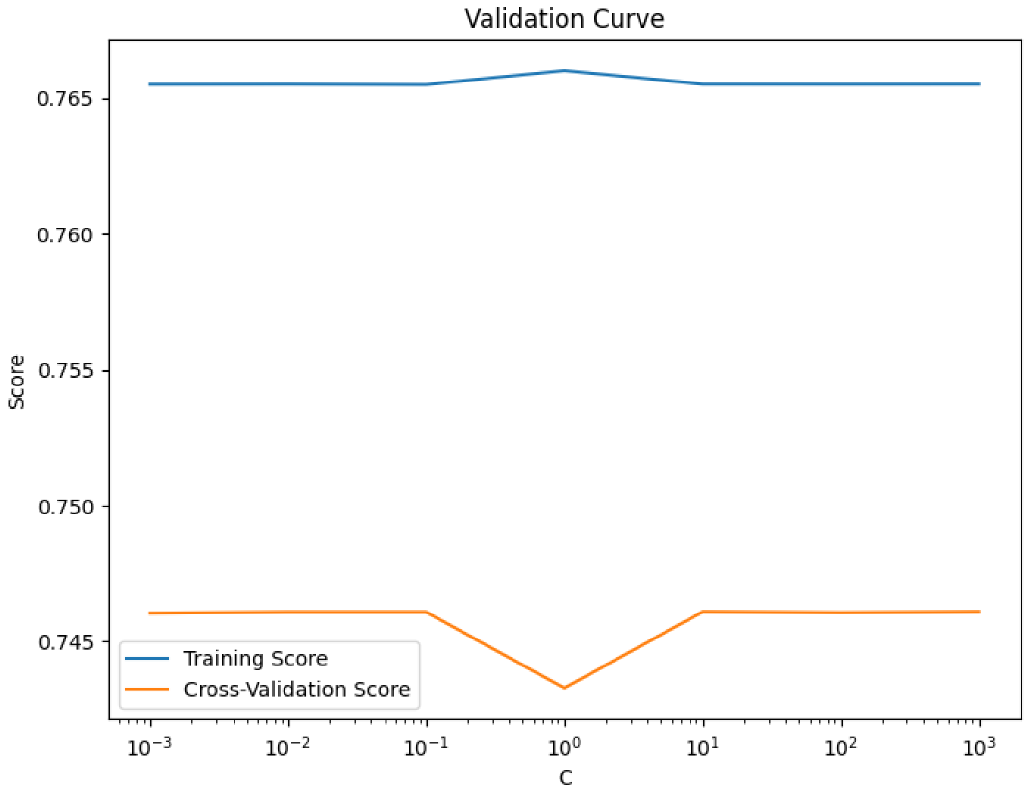
Learning Curve

- As training samples increase, the training score goes down and is bounded with the cross-validation score at about 0.75-0.8.
- Slight overfitting since the training score does better than the cross-validation score.



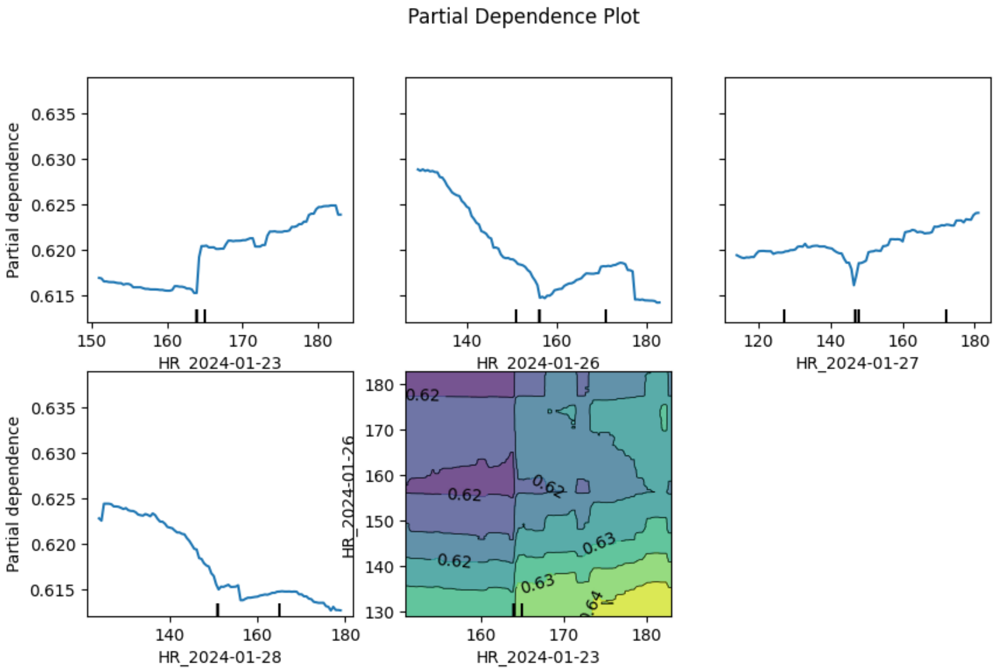
Validation Curve

- Hyperparameter of C
- Model consistently fits training data well
- Slightly drops during cross-validation
- There is some appearance of overfitting
- We want a parameter that is not in the middle since the curves indicate more overfitting at that point.



Partial Dependence Plot

- First four showcases the relationship between HRV Group and the values of individual features.
- The bottom right shows the interaction effect between two features on the predicted outcome.



Model Evaluation

R^2 = 0.763058722511091

RMSE = 0.4867661425047032

MAE = 0.23694127748890903

MSE = 0.23694127748890903

Conclusion

- Model was effective in predictions at an accuracy of 76.3%
- Model was much better at identifying individuals in the high variability group compared to low variability group.
- AUC Score of 0.787 showcases discriminative ability
- Increased potential for further optimization of the parameters.

Key Takeaways

- Heart rate appears to be different for players of different positions.
- Soccer players tend to reduce fat percentage throughout the semester at the college level.
- Recovery strategies should be different based on the position a player is in along with their heart rate variability group.
- Heart rate showed adapted response by the end of the semester for this training.

Things to Study in the Future

- Utilize the type of practice as a feature in the prediction of heart rate variability groups.
- Look at other baseline measurements in relation to the heart rate.
- Assess heart rate at different levels of soccer to see if there is similar predictive potential.

Lessons Learned

- Be confident.
- Try multiple techniques even if it feels like it won’t ever work.
- Sometimes there is no perfectly right answer.
- Be curious.