BASKETBALL PLAYOFFS QUALIFICATION

AC 2023/2024

Bruna Marques - up202007191
Diogo Babo - up202004950
Inês Oliveira - up202103343

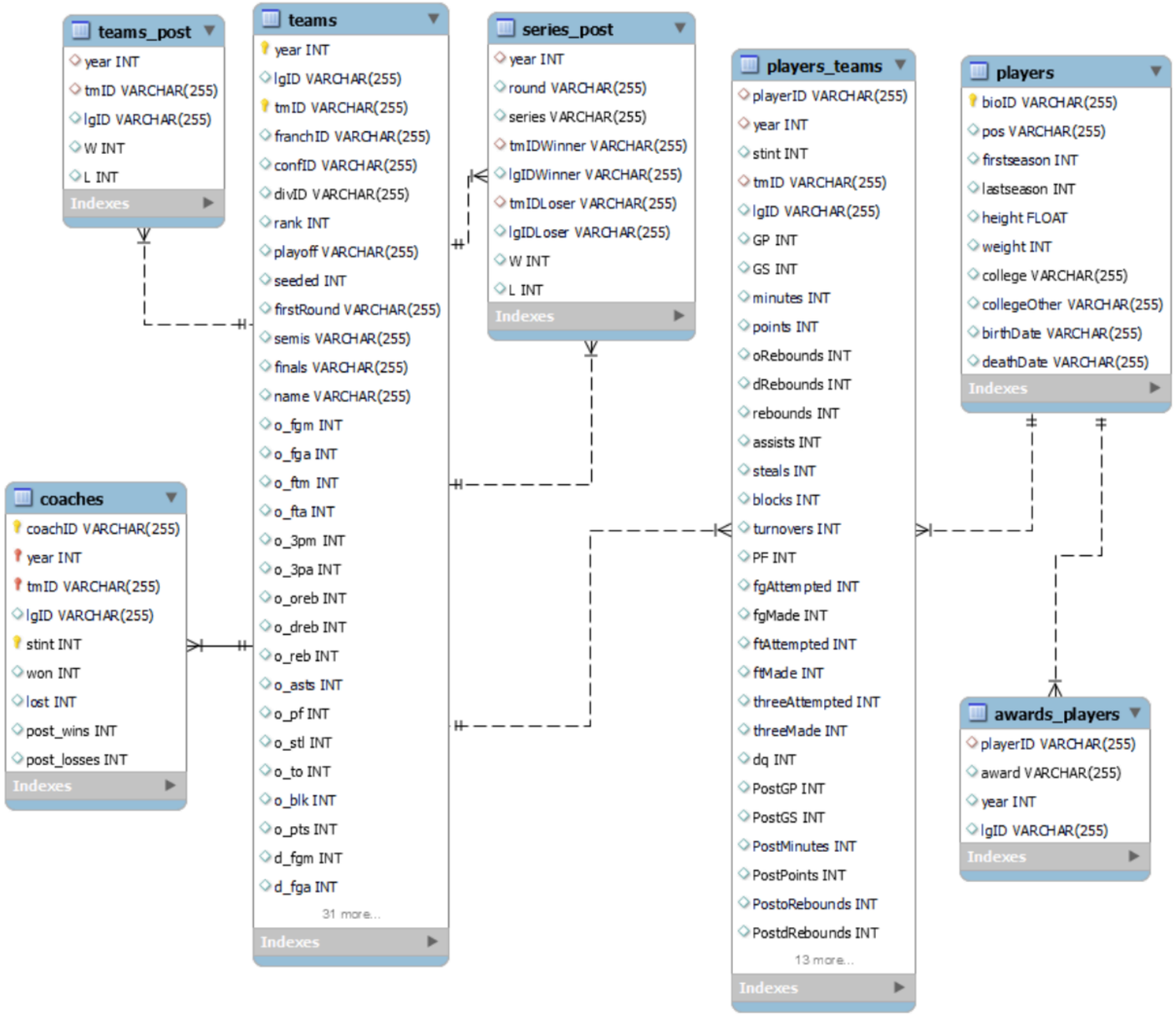
U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Domain Description

The data was provided in .csv files, containing information about teams, players, coaches, and their relationships.

There are:

- 893 players
- 20 teams
- 57 coaches
- 12 awards



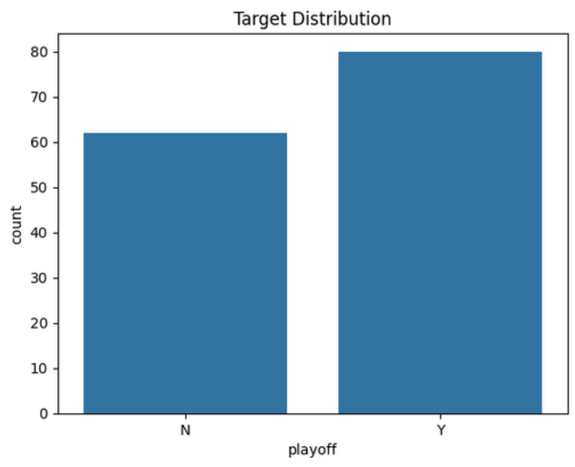
Data Quality

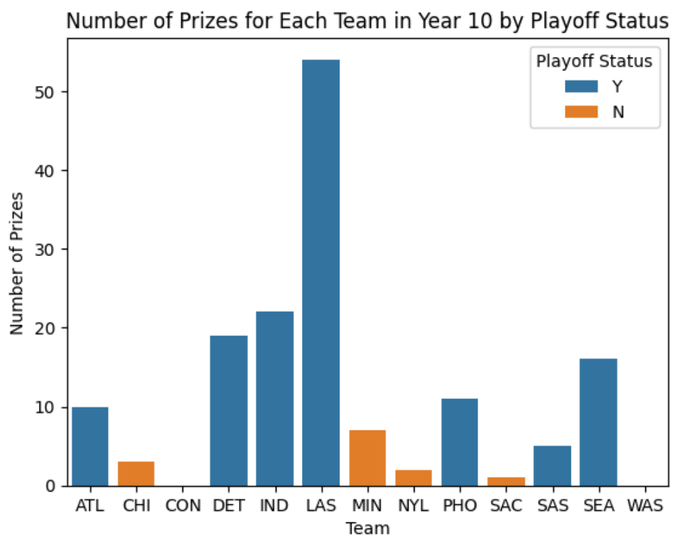
- **Accuracy**: There are data points with low accuracy, for example, there is a player with a height of 9, while the average is 65.5.
- **Integrity**: The dataset contains some attributes with a large number of missing values, such as “divID” and "collegeOther”.There are also attributes with a significant portion of data points being 0, such as "lastseason" and "firstseason".
- **Consistency**: The data must be consistent throughout the database. However, there are several cases of inconsistency. For example, in year 6 there are two teams with a "W" (winner) in the championship final.

Exploratory data analysis

Interesting observations:

- The data is practically balanced.
  - **43% NO** and **57% YES**.
- There are teams in which players and the coach change during the course of a year.
- There are missing values (table players).
- There are teams that do not participate every year.





Problem Definition

Objective: use the data provided (relating to players, teams, coaches, matches and the relationships between them) to predict which teams will qualify for the playoffs the following season **(binary classification problem)**.

Methodology: Train a series of machine learning models and compare their performances to identify the best outcome.

Evaluation Metric: The primary measure used was ROC AUC, with additional consideration given to accuracy and f-measures.

Data Preparation

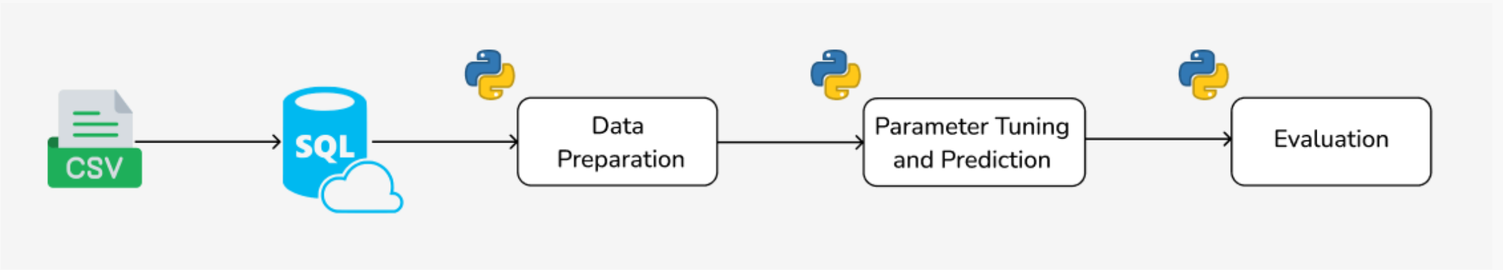
- **Drop** irrelevant **features** and with large number of null values.
- Removed redundant data: ldID column.
- Removed players with height == 0 because the rest of the columns were also 0 or null.
- **Replaced zeros** with the **average** in the weight column.
- **players_prizes** and **coaches_prizes**: created two columns with the number of awards that the players and coaches of each team have. Each award has a weight of importance.

Data Preparation

- **Players statistics:** gathered and summed the previous year's player statistics for each team.
  - If a player was on two teams in one year, the following year's statistics are the sum of these two appearances.
- **Age**: new column - calculated the age of each player according to the year of the season.
- **mean_age, mean_weight, mean_height:** new columns - calculated this averages about the players of each team.
- Z-score normalization.
- Encode categories.

Experimental setup

- Data preparation
- Parameter Tuning - GridSearch with Cross Validation
- Predictions on different models
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - Random Forest
  - K-Nearest Neighbors
  - Naive Bayes
- Evaluation



Results

Train set: years 1-9
Test set: year 10

Model	Accuracy	AUC
Logistic Regression	0.77	0.75
Decision Tree	0.77	0.74
K-Nearest Neighbors	0.77	0.88
Support Vector Machine	0.77	0.70

Models performance:

- Average Accuracy: 0.73
- Average ROC AUC: 0.78

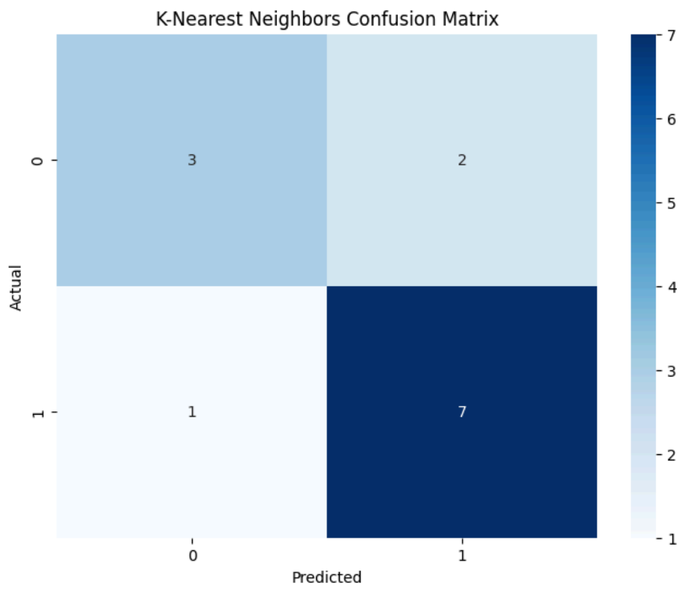
Models best Parameters:

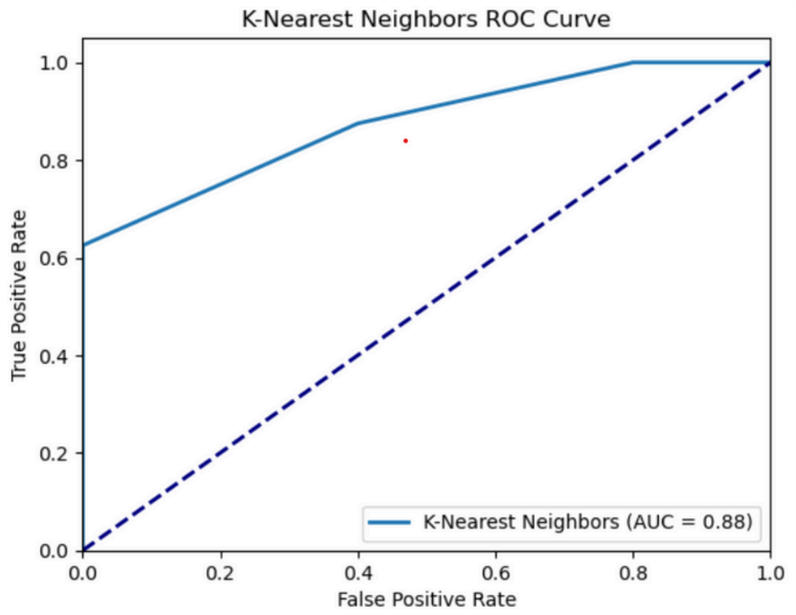
- Logistic Regression
  - C: 10
- Support Vector Machine
  - C: 10
- Decision Tree
  - max_depth: None
- K-Nearest Neighbors
  - n_neighbors: 7

Results

K-Nearest Neighbors

Confusion Matrix





Classification Report

	precision	recall	f1-score	support
0	0.75	0.60	0.67	5
1	0.78	0.88	0.82	8
accuracy			0.77	13
macro avg	0.76	0.74	0.75	13
weighted avg	0.77	0.77	0.76	13

Conclusions and Future Work

- We can fairly determine whether a team reaches playoffs within different models, ranging from 61% to 77% percent of accuracy and AUC(Area under the curve) 70% to 88%.
- There is a range of improvements in the statistics used.
- There is a range of improvements in the setup used where we could have had over/undersampling.

BASKETBALL PLAYOFFS QUALIFICATION

Report

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Business Understanding

Analysis of requirements with the end user

- Group players based on their performance metrics, playing style, and historical data.
- Develop a predictive system to understand new players and predict their impact on game outcomes.
- Provide a tool for coaches to evaluate the risk associated with different player combinations during a match.

Definition of business goals

The business goal is to design a predictive system that:

- Provide predictions for all teams.
- Analyze the results obtained.
- Enhance the accuracy of predictions to achieve at least a 70% success rate.

Translation of business goals into data mining goals

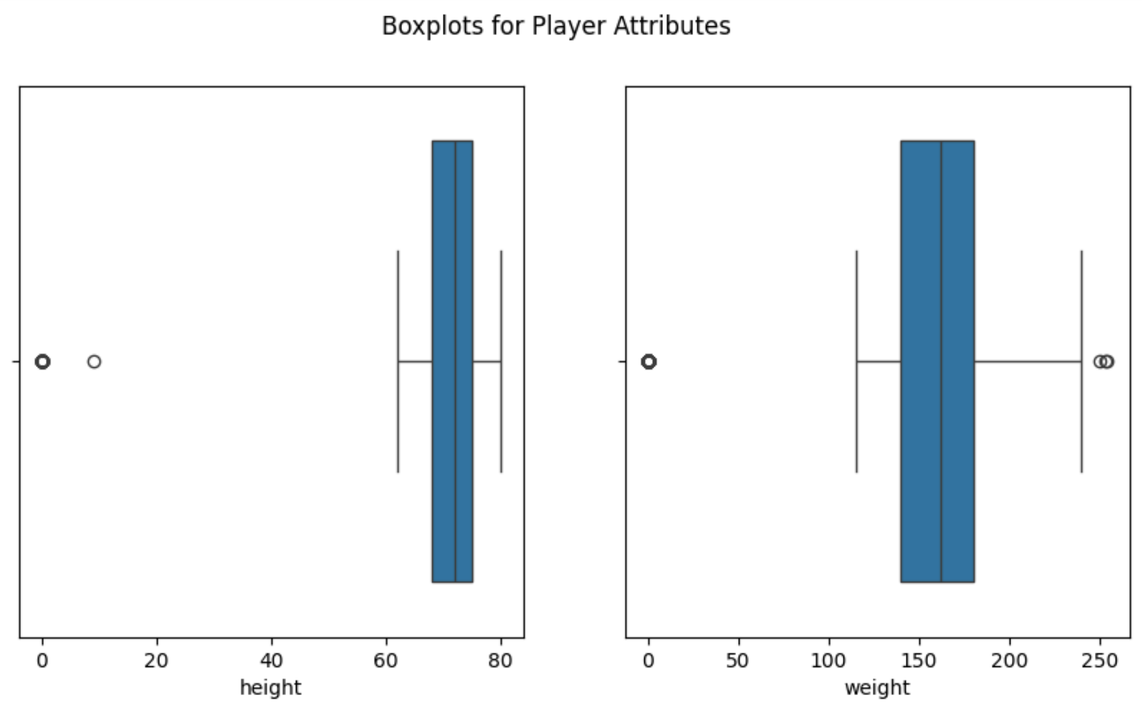
Develop a system that predicts which teams for a certain year will qualify for the playoffs. 'Y' indicates qualifying team and 'N' for those that won't make the playoffs.

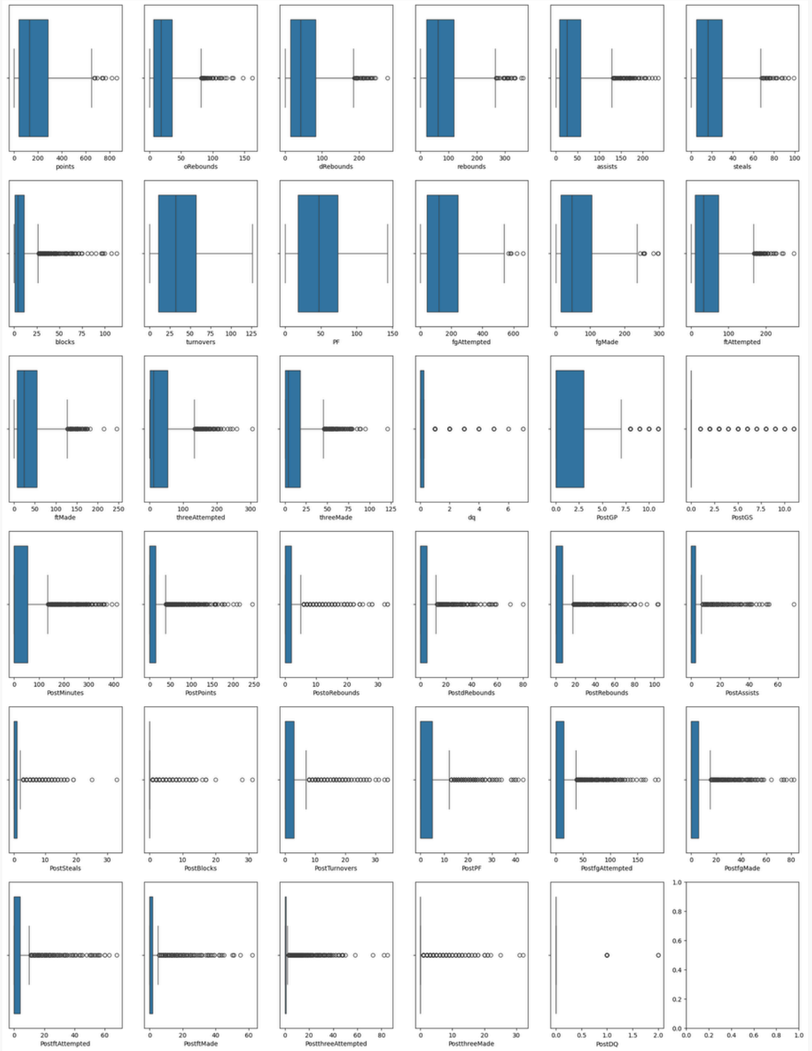
Data Quality

- **Accuracy**: There are data points with low accuracy, for example, there is a player with a height of 9, while the average is 65.5.
- **Integrity**: The dataset contains some attributes with a large number of missing values, such as “divID” and "collegeOther”.There are also attributes with a significant portion of data points being 0, such as "lastseason" and "firstseason".
- **Consistency**: The data must be consistent throughout the database. However, there are several cases of inconsistency. For example, in year 6 there are two teams with a "W" (winner) in the championship final.
- **Completeness:** There are missing values in several columns, for example, pos: 78 missing values, college: 167 missing values, collegeOther: 882 missing values. Thus, there are tables that are not very complete.
- **Uniqueness:** There are no duplicate records.
- **Timeliness:** The dataset shows real information from 10 different years of competitions, from 2000 to 2009.
- **Validity**: The data accurately reflects the real-world information (WNBA) that intendeds to represent.

Outliers

We didn´t find outliers. Some players had height ou weight as 0, and one case wich the height was 9, but we considered them as errors of the database (replaced by the mean).

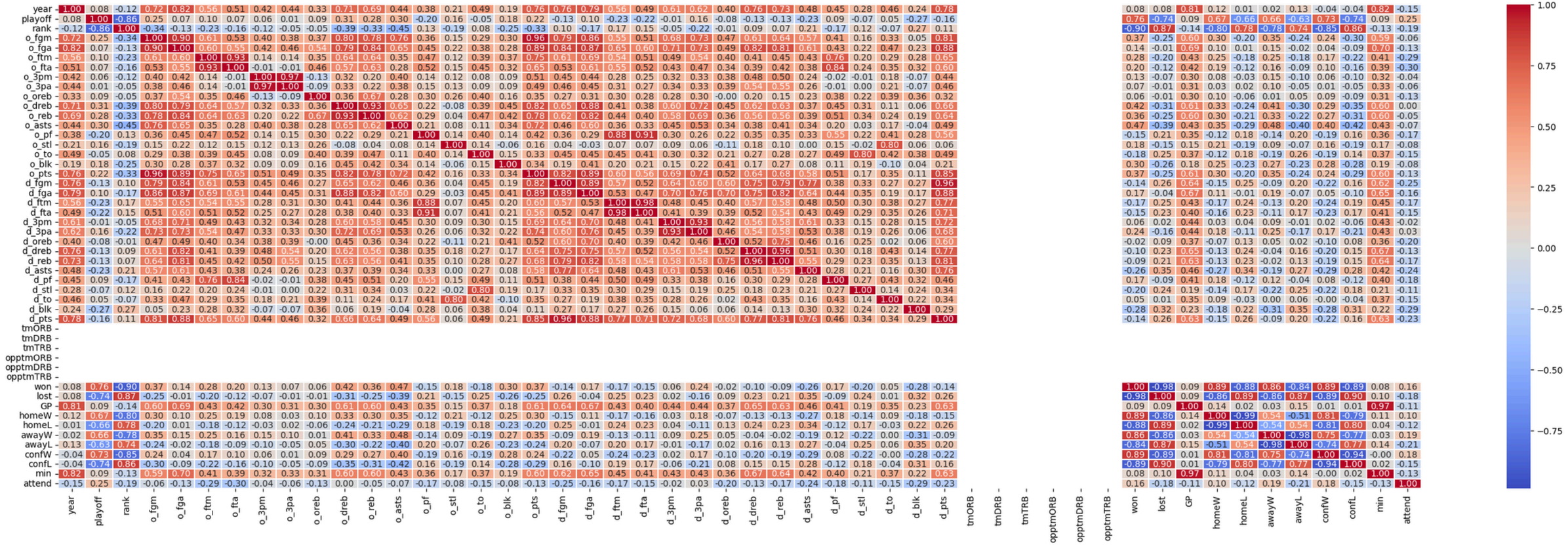




An attempt to find more outliers

Correlation Matrix

With the correlation matrix, we could observe the relationship between the attributes of the Teams table. Upon analyzing it, it appears that the game statistics of the teams have minimal influence on their playoff qualification.



Unused Data Preparation Strategies

Some transformations were tested but were not utilized in the final models:

- Number of wins and defeats, number of attendees and rank of the previous year of the teams.
- Teams statistics as the sum of their players' statistics calculated as the average of performances of the last 2/3/4/5 last years.
- Teams statistics as the average of their players' statistics calculated as the average of performances of the last 2/3/4/5 last years.
- Teams statistics as the teams statistics of the previous year.
- Number of wins and defeats of the season and post season of the coach.
- Without information about the age, height and weight of the players.
- Without normalization.

New Changes

After the presentation and critical evaluation of the results, it became clear that the predicted number of teams qualifying for the playoffs didn't align with the practical constraint of only eight teams making it (four from each conference).

To address this issue, the train/test data were split into the two conferences, and a threshold was assigned. Only the top four teams, meaning the four most likely to qualify, were effectively classified as playoff contenders, while the remaining teams were labeled as non-qualifiers.

Our Setup

We worked with GitHub, SQLite3 and several Python libraries such as Pandas, Scikit-learn (Sklearn), Matplotlib, datetime, Numpy and Seaborn. These tools helped us with collaboration, transfer data from the CSV files to a relational database, analyze data, apply machine learning methods, create clear visualizations and calculate elapsed time.

To utilize our data with machine learning models, we employed label encoding to make the data compatible. Per example, this involves converting categorical variables, such as strings, into a format suitable for analysis.

Six different models were applied and fine-tuned their parameters using GridSearch with Cross Validation for each one.

Logistic Regression

Logistic Regression is a binary classification algorithm well-suited for predicting the probability of an instance belonging to a particular class. Therefore, it fits into the goals of this project.

West

```
Feature importance for WE:
prev_finals: 0.0366237442433258
weight: 0.01151214326984592
playoff: 0.0073746898735137165
prev_firstRound: 0.004979981850290682
sum_steals: 0.004611984021391438
sum_rebounds: 0.003461389159196231
sum_threeMade: 0.003291531932861763
sum_dRebounds: 0.0032906079805125185
sum_oRebounds: 0.0032739268608393845
sum_points: 0.0032126972844962014
prev_semis: 0.0030120053109315673
sum_GP: 0.0028746306540689583
sum_fgAttempted: 0.002613119841309738
height: 0.002517287370330547
year: 0.0024793042348940867
sum_minutes: 0.002245093121336047
tmID: 0.0021876881097430047
sum_assists: 0.0019690060954229037
sum_PF: 0.0019032424322699543
players_prizes: 0.0018527873770786626
sum_fgMade: 0.0013773117989839788
sum_GS: 0.0012777999884262374
sum_ftAttempted: 0.001230367020809194
sum_threeAttempted: 0.0009971121325681577
...
```

Train set	Test set	Accuracy	AUC	Time
1-9	10	0.67	1	0.15
1-6 & 8-10	7	0.71	0.92	0.15
1-5 & 7-10	6	0.43	0.83	0.14
1-8	9-10	0.69	0.73	0.13

East

```
Feature importance for EA:
age: 0.30703986470647415
sum_dq: 0.25162107163319847
sum_ftMade: 0.18310103305676648
prev_firstRound: 0.17970710859248454
sum_rebounds: 0.14167283499392594
prev_finals: 0.12886452750110766
players_prizes: 0.11070171228966626
sum_turnovers: 0.10810811280354284
sum_threeMade: 0.10797216690679255
year: 0.10493382379348104
height: 0.10393374813340059
sum_threeAttempted: 0.09913524654020406
sum_GP: 0.09871714510191365
sum_oRebounds: 0.08044135561498073
sum_blocks: 0.07911021044600895
sum_assists: 0.05447768949800991
sum_points: 0.051938732216419016
sum_steals: 0.048423991198579215
sum_PF: 0.04615209592718269
prev_semis: 0.04493139552452867
sum_dRebounds: 0.03917615067043772
playoff: 0.03205156751293739
weight: 0.029204205379914375
sum_fgAttempted: 0.016593887205290838
...
```

Train set	Test set	Accuracy	AUC	Time
1-9	10	0.43	0.66	0.23
1-6 & 8-10	7	0.71	0.92	0.26
1-5 & 7-10	6	1.0	1.0	0.26
1-8	9-10	0.43	0.67	0.20

Support Vector Machine

SVMs are also good at solving binary classification problems, and has similar results to Logistic Regression.

West

Train set	Test set	Accuracy	AUC	Time
1-9	10	0.67	1.0	0.21
1-6 & 8-10	7	0.43	1.0	0.24
1-5 & 7-10	6	0.71	0.83	0.20
1-8	9-10	0.69	0.67	0.18

East

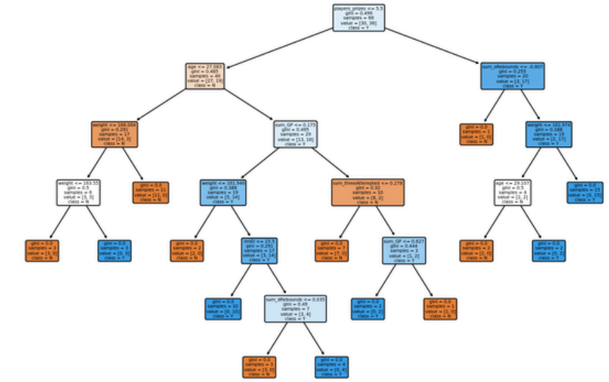
Train set	Test set	Accuracy	AUC	Time
1-9	10	0.43	0.67	0.32
1-6 & 8-10	7	1.0	1.0	0.30
1-5 & 7-10	6	1.0	1.0	0.62
1-8	9-10	0.43	0.67	0.18

Decision Tree

The decision tree model makes decisions by recursively splitting the dataset into subsets based on the values of features. Decision trees are often considered white box models because they offer transparency, interpretability, and a clear representation of decision-making processes.

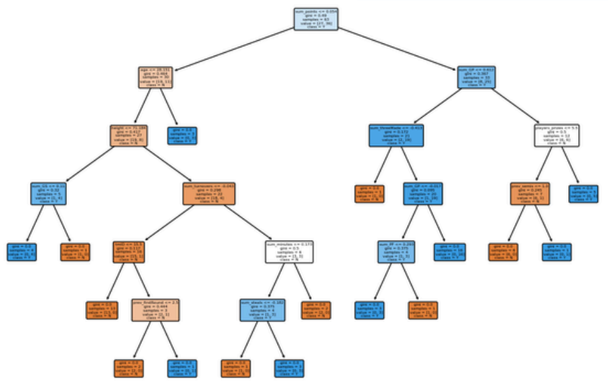
West

Train set	Test set	Accuracy	AUC	Time
1-9	10	0.67	0.75	0.13
1-6 & 8-10	7	0.71	0.88	0.11
1-5 & 7-10	6	0.71	0.71	0.13
1-8	9-10	0.54	0.51	0.11



East

Train set	Test set	Accuracy	AUC	Time
1-9	10	0.71	0.71	0.13
1-6 & 8-10	7	0.43	0.58	0.19
1-5 & 7-10	6	0.33	0.25	0.14
1-8	9-10	0.57	0.56	0.12



	precision	recall	f1-score
0	0.67	0.67	0.67
1	0.75	0.75	0.75

Random Forest

It builds multiple decision trees during training and combines their predictions to reach a single result. The West conference accuracy varies across folds, indicating that the model's performance is not consistent on different subsets of the data and is sensitive to the specific subsets of data used in each fold (hight variance: 23% to 71%).

West

Train set	Test set	Accuracy	AUC	Time
1-9	10	0.67	0.88	6.15
1-6 & 8-10	7	1.0	1.0	9.15
1-5 & 7-10	6	0.71	0.58	10.98
1-8	9-10	0.69	0.63	5.29

East

Train set	Test set	Accuracy	AUC	Time
1-9	10	0.71	0.5	6.73
1-6 & 8-10	7	1.0	1.0	6.0
1-5 & 7-10	6	0.67	0.75	9.00
1-8	9-10	0.43	0.33	5.38

cross-validation:

```
Fold 1: Accuracy = 0.7143
Fold 2: Accuracy = 0.6154
Fold 3: Accuracy = 0.4615
Fold 4: Accuracy = 0.2308
Fold 5: Accuracy = 0.6154
Mean Accuracy: 0.5275
```

```
Fold 1: Accuracy = 0.5385
Fold 2: Accuracy = 0.5385
Fold 3: Accuracy = 0.4615
Fold 4: Accuracy = 0.5000
Fold 5: Accuracy = 0.5000
Mean Accuracy: 0.5077
```

	precision	recall	f1-score
0	0.67	0.67	0.67
1	0.75	0.75	0.75

K-Nearest Neighbors

It is a non-parametric, instance-based learning algorithm, uses proximity to make predictions about the grouping of an individual data point.

West

Train set	Test set	Accuracy	AUC	Time
1-9	10	1.0	1.0	0.12
1-6 & 8-10	7	0.71	0.92	0.17
1-5 & 7-10	6	0.71	0.83	0.45
1-8	9-10	0.69	0.81	0.11

	precision	recall	f1-score
0	1.00	1.00	0.65
1	1.00	1.00	0.20

East

Train set	Test set	Accuracy	AUC	Time
1-9	10	0.43	0.63	0.75
1-6 & 8-10	7	0.71	0.54	0.13
1-5 & 7-10	6	0.67	0.88	0.17
1-8	9-10	0.43	0.80	0.12

	precision	recall	f1-score
0	0.33	0.33	0.63
1	0.50	0.50	0.20

Naive Bayes

Naive Bayes is based on Bayes' theorem (finds the probability of an event occurring given the probability of another event that has already occurred) with the "naive" assumption of independence between every pair of features and therefore takes less time.

West

Train set	Test set	Accuracy	AUC	Time
1-9	10	0.67	1.0	0.04
1-6 & 8-10	7	0.71	1.0	0.06
1-5 & 7-10	6	0.71	0.58	0.12
1-8	9-10	0.54	0.68	0.03

	precision	recall	f1-score
0	0.50	0.50	0.63
1	0.75	0.75	0.15

East

Train set	Test set	Accuracy	AUC	Time
1-9	10	0.71	0.67	0.06
1-6 & 8-10	7	1.0	1.0	0.13
1-5 & 7-10	6	0.67	0.63	0.04
1-8	9-10	0.43	0.82	0.02

	precision	recall	f1-score
0	0.67	0.67	0.67
1	0.75	0.75	0.75

Conclusions

As expected, in cases where the training set contains years that come after the test set year, predictions tend to be more accurate. This is because having knowledge of the future assists in estimating the past.

As expected, when the train set have less years, predictions tend to be less reliable since it has less information.

K-Nearest Neighbors turns out to be the best model for the **West conference**. Takes a short amount of time with great results.

In the **East conference**, three models have similar performance. However, they differ in the AUC, which is an important metric to binary classification problems once it summarizes the ability to discriminate between the positive and negative classes. Consequently, the Decision Tree stands as the best choice in this scenario.

Annex

Annex

PLAYERS:

	firstseason	lastseason	height	weight
count	893.000000	893.0	893.000000	893.000000
mean	0.001120	0.0	65.500560	145.415454
std	0.033464	0.0	20.940425	61.275703
min	0.000000	0.0	0.000000	0.000000
25%	0.000000	0.0	68.000000	140.000000
50%	0.000000	0.0	72.000000	162.000000
75%	0.000000	0.0	75.000000	180.000000
max	1.000000	0.0	80.000000	254.000000

COACHES:

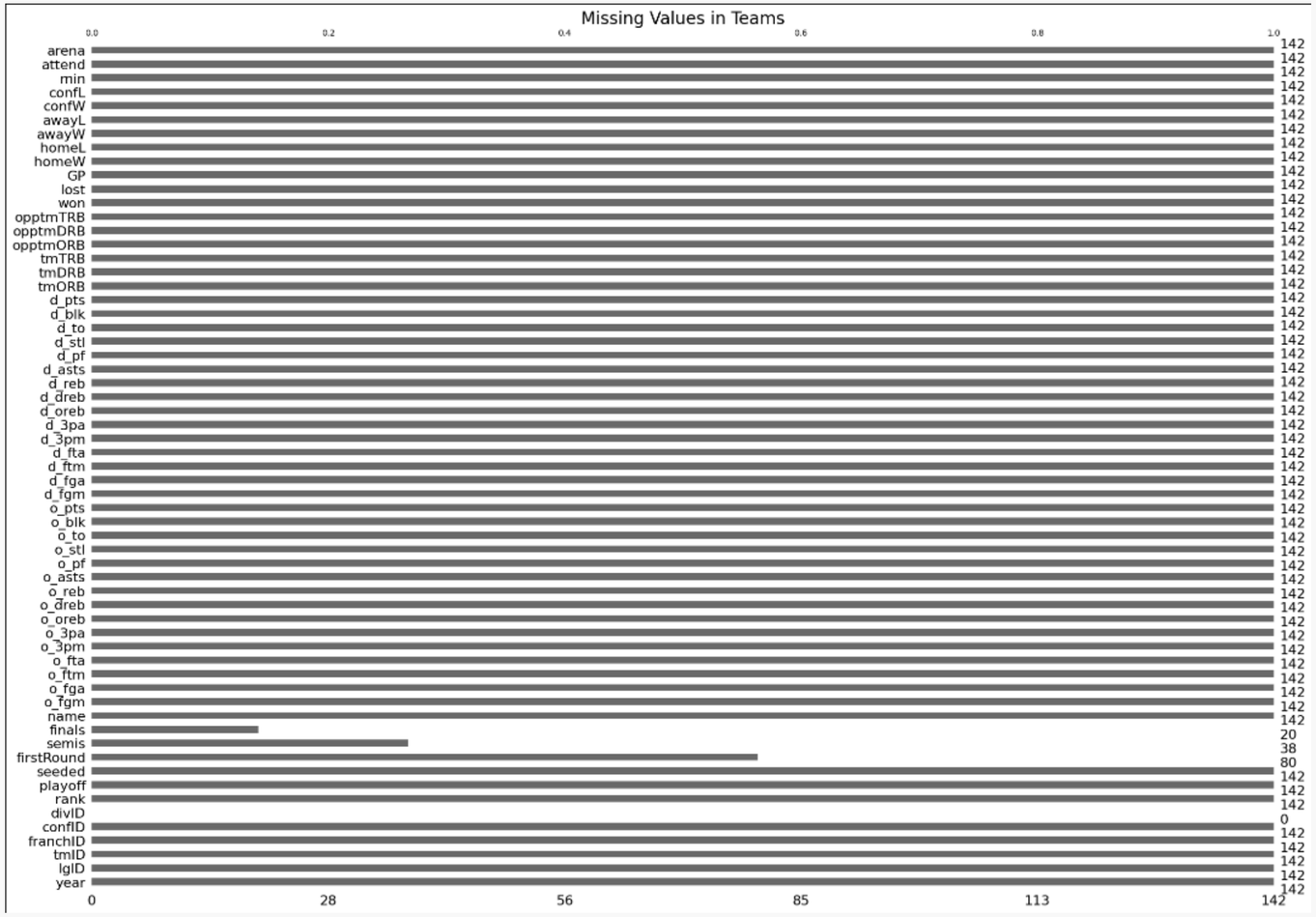
	year	stint	won	lost	post_wins	post_losses
count	162.000000	162.000000	162.000000	162.000000	162.000000	162.000000
mean	5.314815	0.364198	14.672840	14.623457	1.166667	1.172840
std	2.896715	0.693861	6.403445	5.678789	1.953656	1.316782
min	1.000000	0.000000	0.000000	2.000000	0.000000	0.000000
25%	3.000000	0.000000	10.000000	11.000000	0.000000	0.000000
50%	5.000000	0.000000	16.000000	15.000000	0.000000	0.000000
75%	8.000000	0.000000	18.750000	18.000000	1.000000	2.000000
max	10.000000	2.000000	28.000000	30.000000	7.000000	5.000000

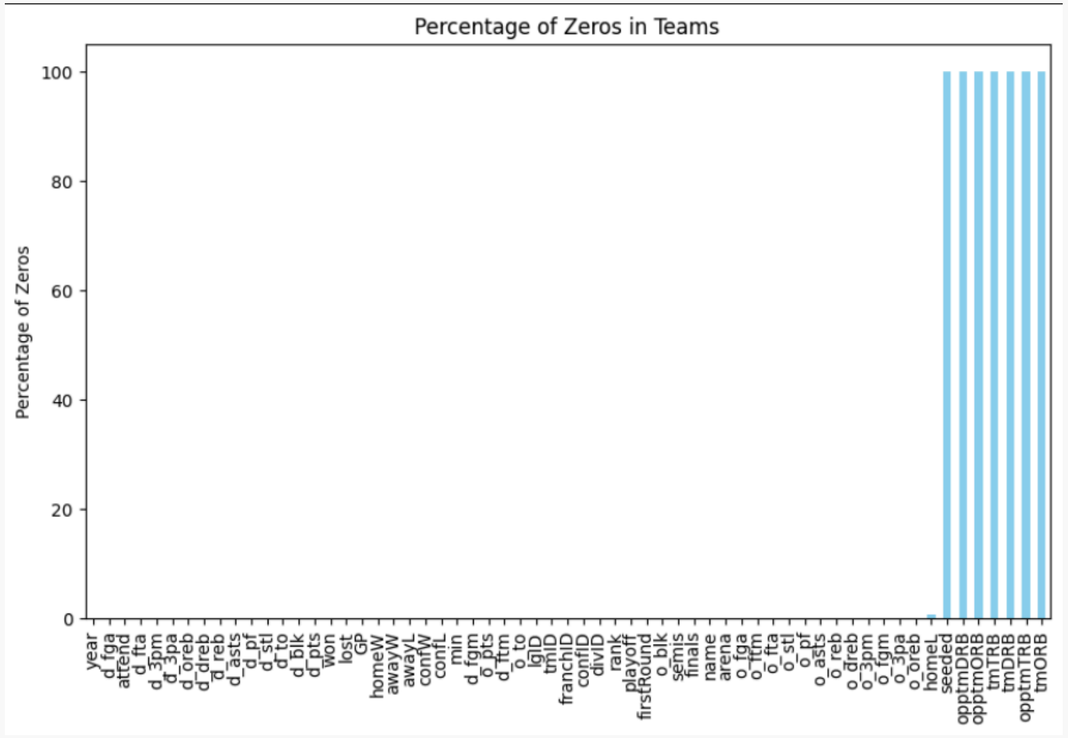
SERIES_POST:

	year	W	L
count	70.00000	70.000000	70.000000
mean	5.50000	2.071429	0.614286
std	2.89302	0.259399	0.572127
min	1.00000	2.000000	0.000000
25%	3.00000	2.000000	0.000000
50%	5.50000	2.000000	1.000000
75%	8.00000	2.000000	1.000000
max	10.00000	3.000000	2.000000

TEAMS_POST:

	year	W	L
count	80.000000	80.000000	80.000000
mean	5.500000	2.350000	2.350000
std	2.890403	2.228129	0.843441
min	1.000000	0.000000	0.000000
25%	3.000000	1.000000	2.000000
50%	5.500000	1.500000	2.000000
75%	8.000000	3.250000	3.000000
max	10.000000	7.000000	5.000000


Missing Values in Teams
0.0
0.2
0.4
0.6
0.8
1.0
arena 142
attend 142
min 142
confL 142
confW 142
awayL 142
awayW 142
homeL 142
homeW 142
GP 142
lost 142
won 142
opptmTRB 142
opptmDRB 142
opptmORB 142
tmTRB 142
tmDRB 142
tmORB 142
d_pts 142
d_blk 142
d_to 142
d_stl 142
d_pf 142
d_asts 142
d_reb 142
d_dreb 142
d_oreb 142
d_3pa 142
d_3pm 142
d_fta 142
d_ftm 142
d_fga 142
d_fgm 142
o_pts 142
o_blk 142
o_to 142
o_stl 142
o_pf 142
o_asts 142
o_reb 142
o_dreb 142
o_oreb 142
o_3pa 142
o_3pm 142
o_fta 142
o_ftm 142
o_fga 142
o_fgm 142
name 142
finals 20
semis 38
firstRound 80
seeded 142
playoff 142
rank 142
divID 0
confID 142
franchID 142
tmID 142
lgID 142
year 142
0
28
56
85
113
142


Percentage of Zeros in Teams
Percentage of Zeros
100
80
60
40
20
0
year
d_fga
attend
d_fta
d_3pm
d_3pa
d_oreb
d_dreb
d_reb
d_asts
d_pf
d_stl
d_to
d_blk
d_pts
won
lost
GP
homeW
awayW
awayL
confW
confL
min
d_fgm
o_pts
d_ftm
o_to
lgID
tmID
franchID
confID
divID
rank
playoff
firstRound
o_blk
semis
finals
name
arena
o_fga
o_ftm
o_fta
o_stl
o_pf
o_asts
o_reb
o_dreb
o_3pm
o_fgm
o_3pa
o_oreb
homeL
seeded
opptmDRB
opptmORB
tmTRB
tmDRB
opptmTRB
tmORB

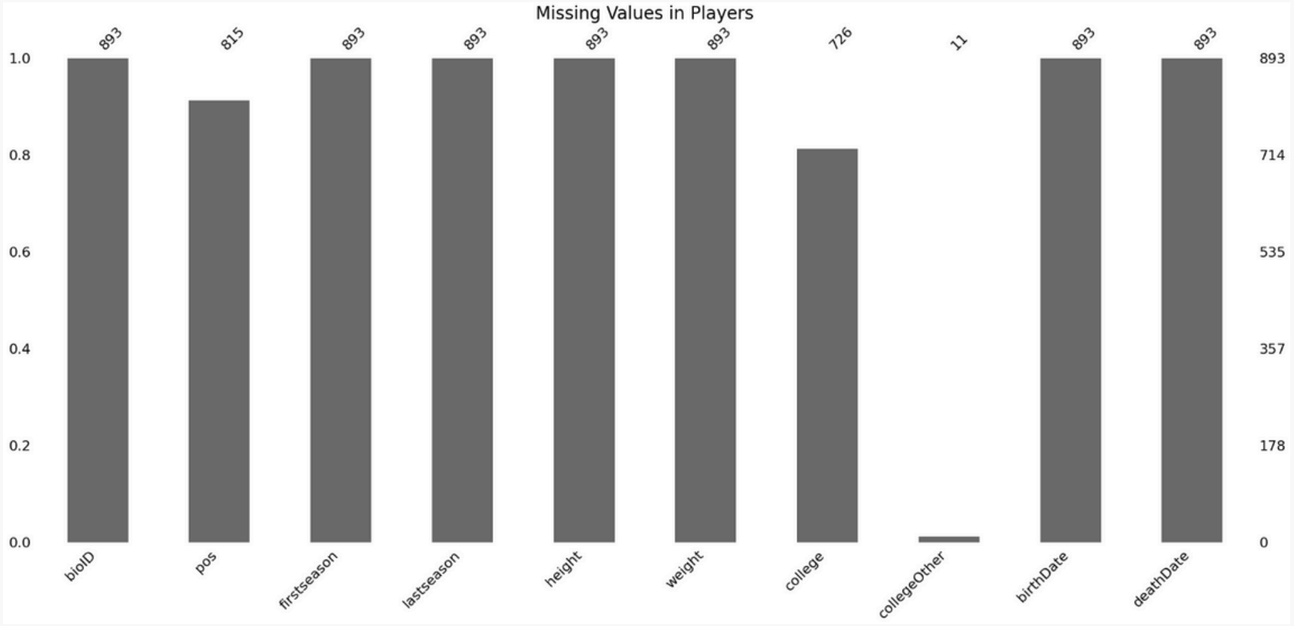
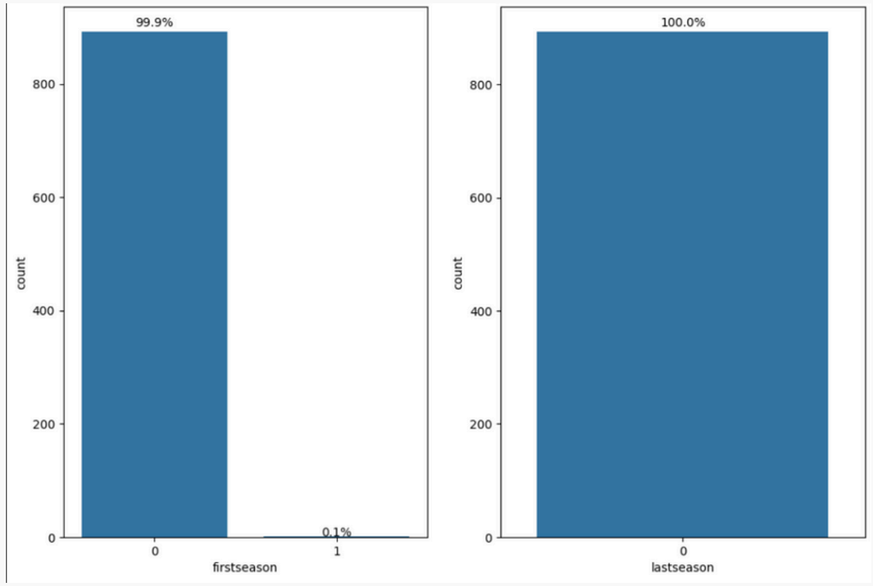
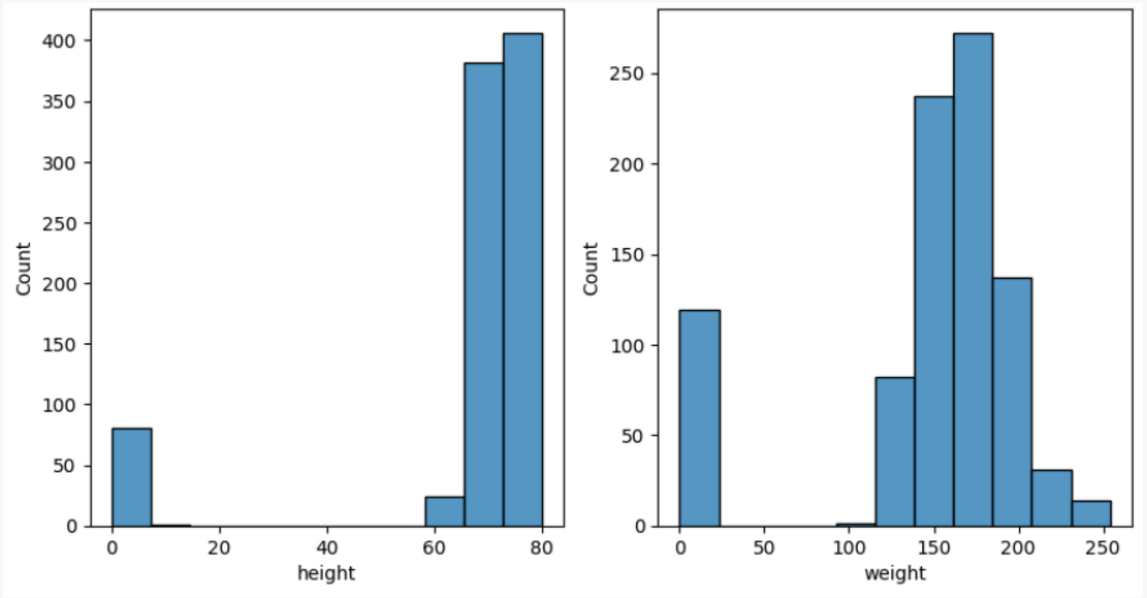
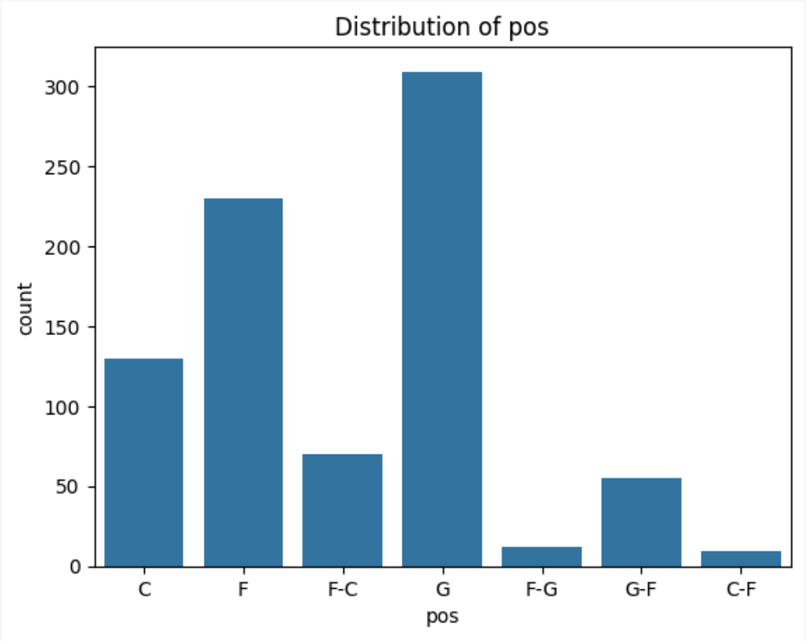

Missing Values in Players
893
815
893
893
893
893
726
11
893
893
1.0
0.8
0.6
0.4
0.2
0.0
893
714
535
357
178
0
bioID
pos
firstseason
lastseason
height
weight
college
collegeOther
birthDate
deathDate

table Players







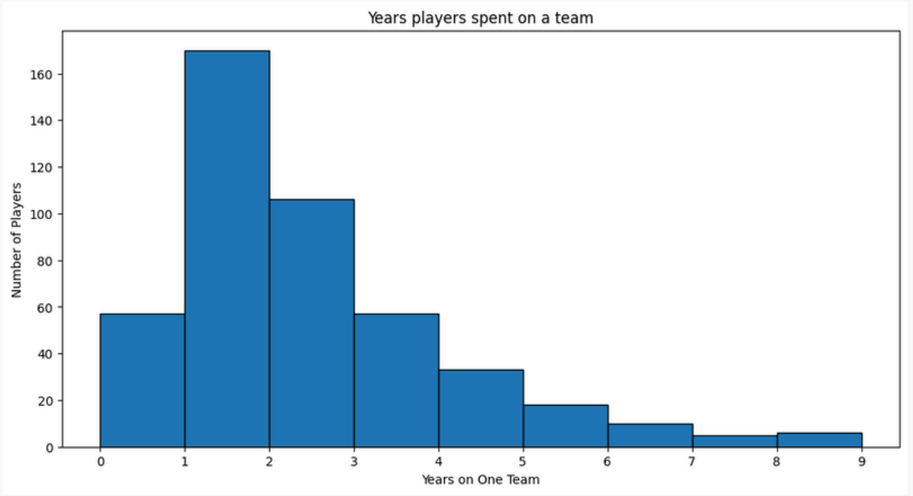
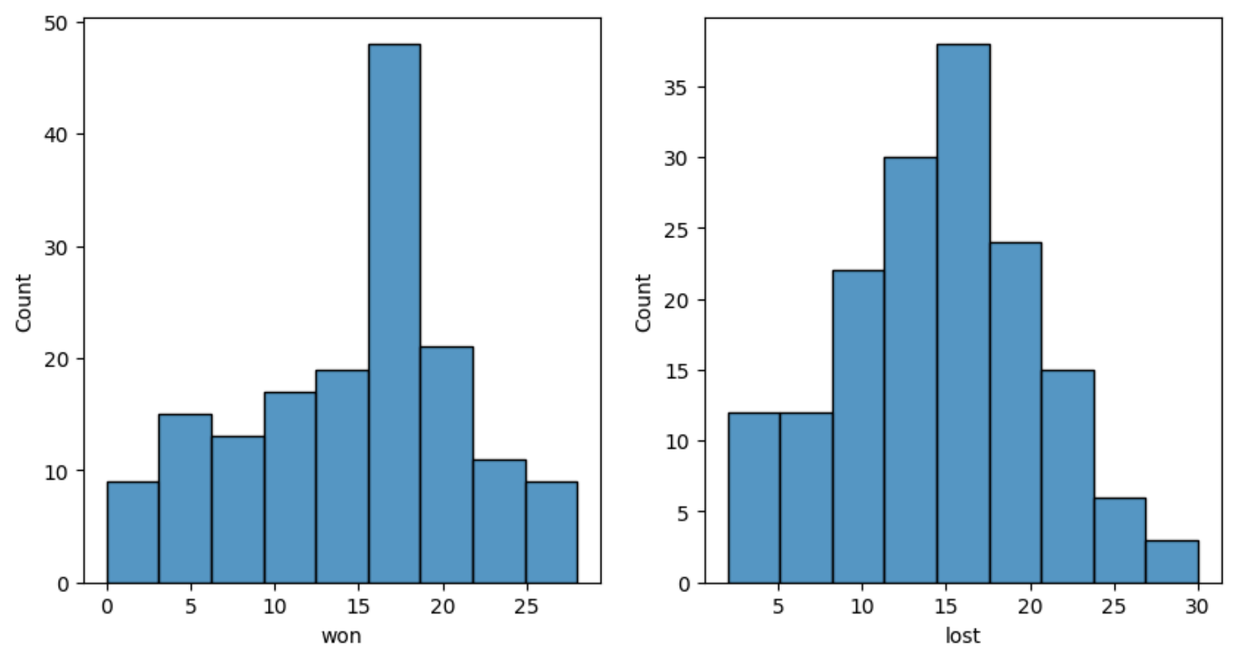
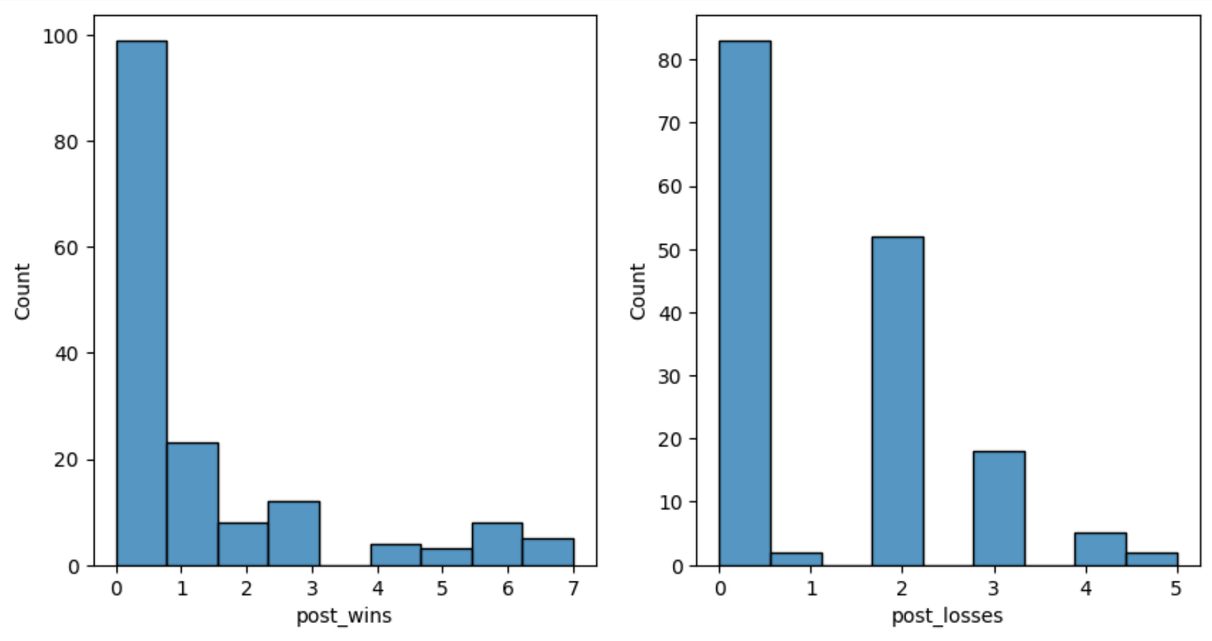
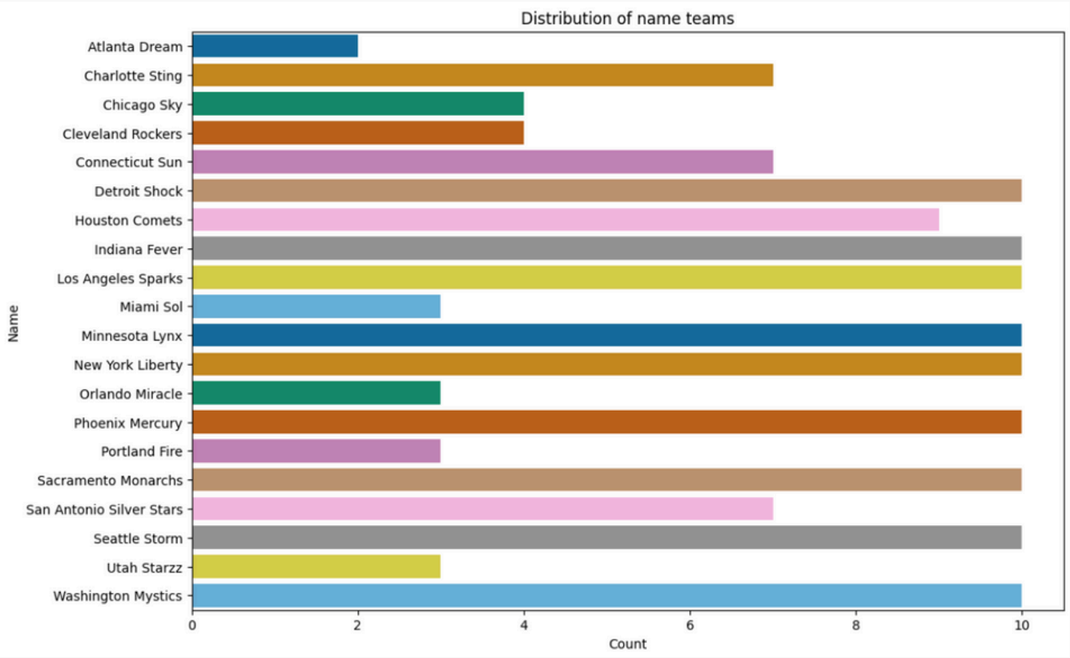
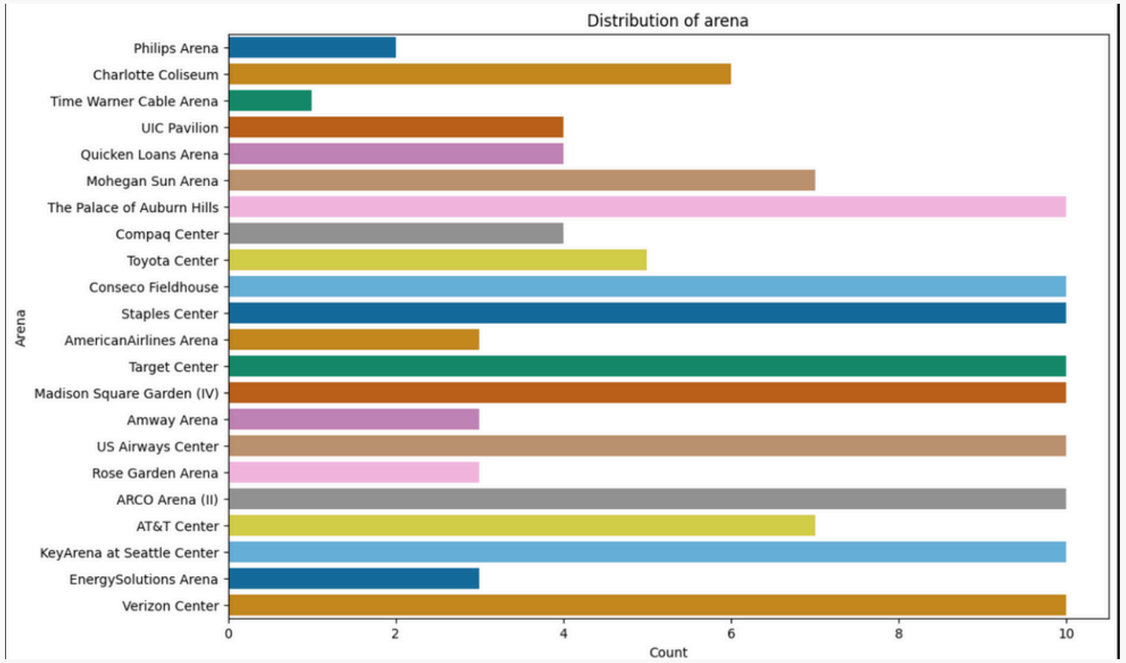


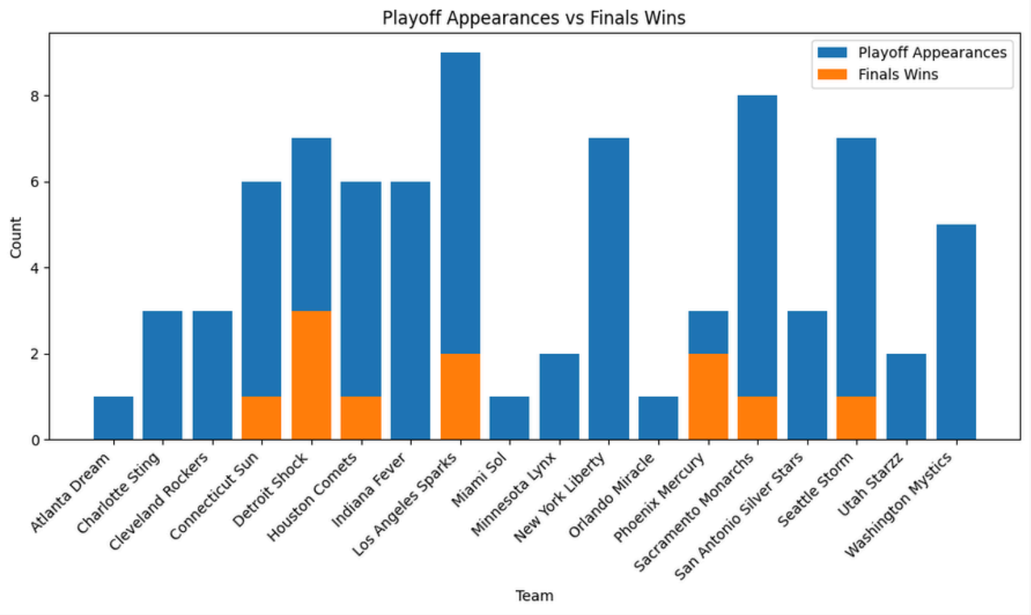
table Coaches

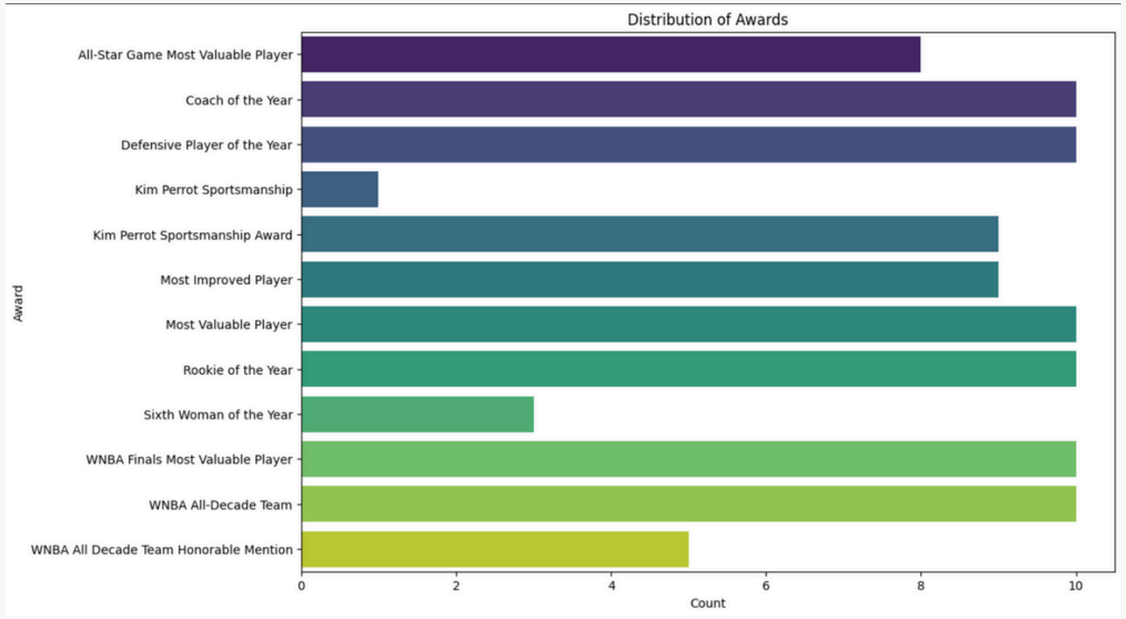


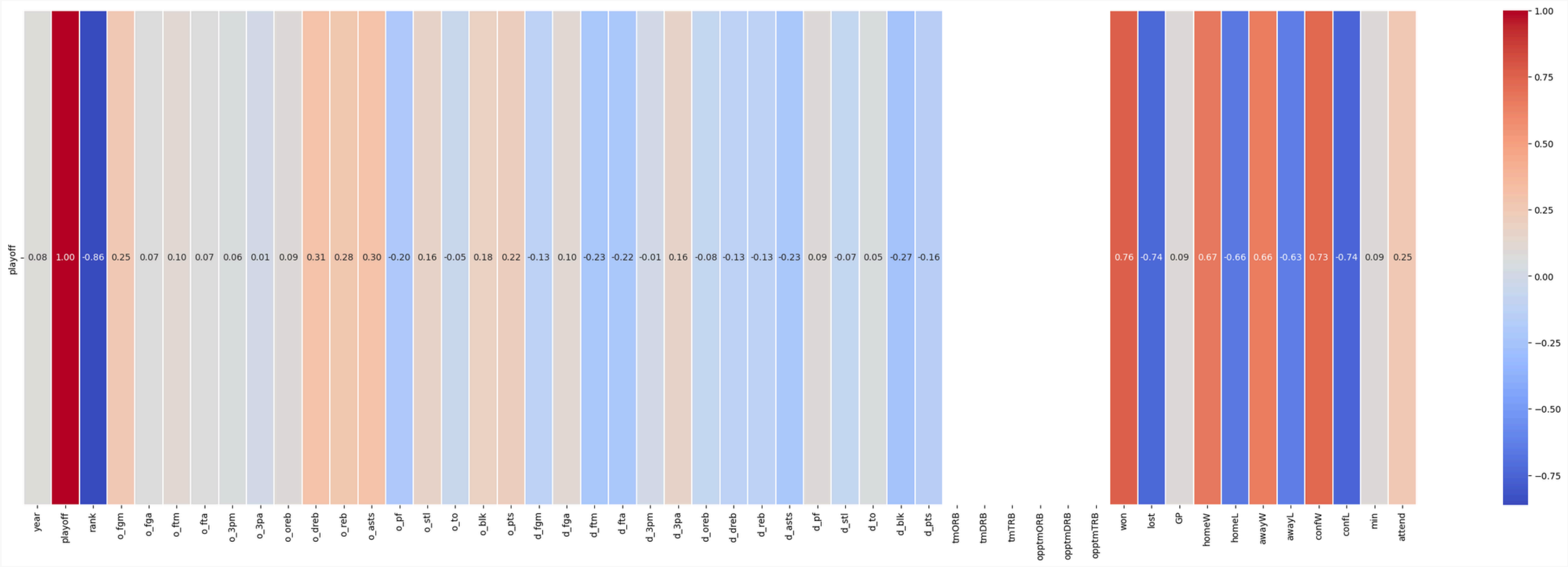



Distribution of name teams
Atlanta Dream
Charlotte Sting
Chicago Sky
Cleveland Rockers
Connecticut Sun
Detroit Shock
Houston Comets
Indiana Fever
Los Angeles Sparks
Miami Sol
Minnesota Lynx
New York Liberty
Orlando Miracle
Phoenix Mercury
Portland Fire
Sacramento Monarchs
San Antonio Silver Stars
Seattle Storm
Utah Starzz
Washington Mystics
Name
Count
0
2
4
6
8
10

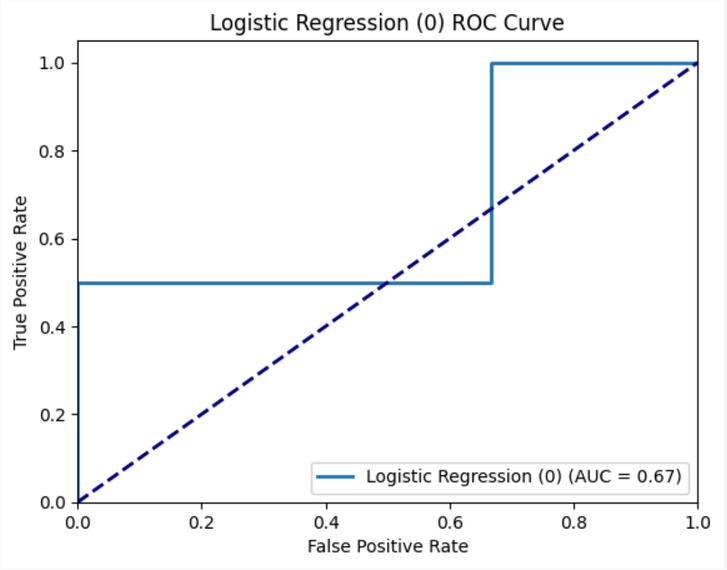

Distribution of arena
Philips Arena
Charlotte Coliseum
Time Warner Cable Arena
UIC Pavilion
Quicken Loans Arena
Mohegan Sun Arena
The Palace of Auburn Hills
Compaq Center
Toyota Center
Conseco Fieldhouse
Staples Center
AmericanAirlines Arena
Target Center
Madison Square Garden (IV)
Amway Arena
US Airways Center
Rose Garden Arena
ARCO Arena (II)
AT&T Center
KeyArena at Seattle Center
EnergySolutions Arena
Verizon Center
Arena
Count
0
2
4
6
8
10

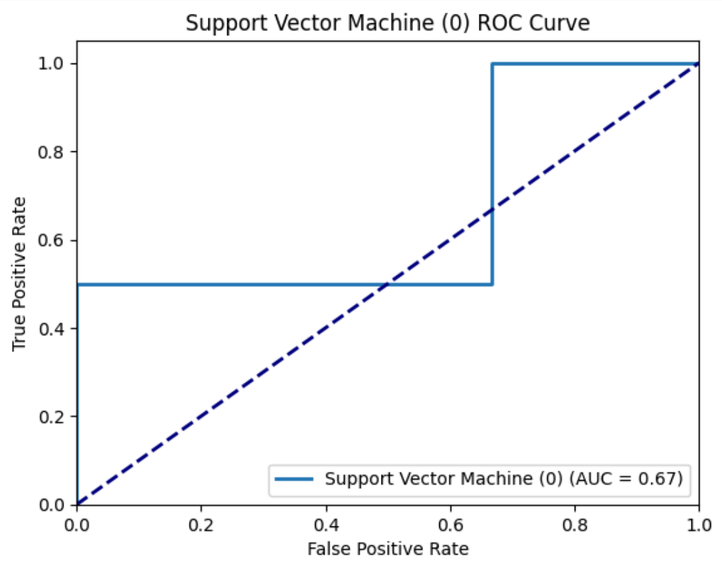

Playoff Appearances vs Finals Wins
Playoff Appearances
Finals Wins
Count
0
2
4
6
8
Atlanta Dream
Charlotte Sting
Cleveland Rockers
Connecticut Sun
Detroit Shock
Houston Comets
Indiana Fever
Los Angeles Sparks
Miami Sol
Minnesota Lynx
New York Liberty
Orlando Miracle
Phoenix Mercury
Sacramento Monarchs
San Antonio Silver Stars
Seattle Storm
Utah Starzz
Washington Mystics
Team

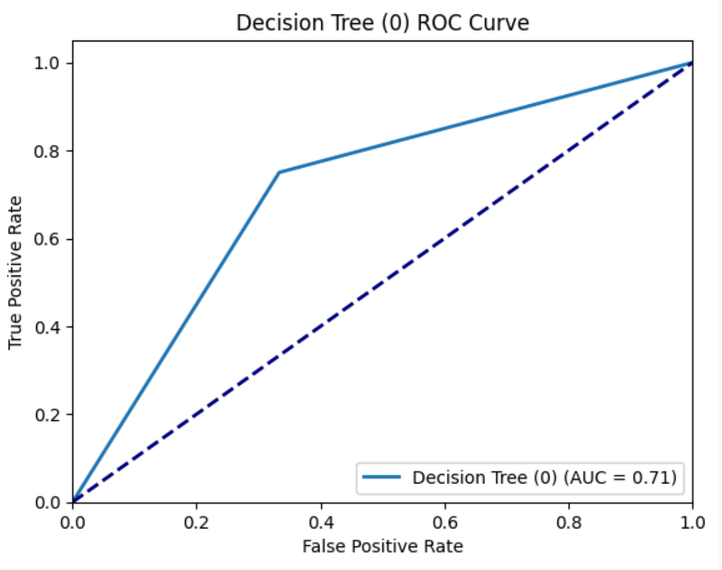

Distribution of Awards
All-Star Game Most Valuable Player
Coach of the Year
Defensive Player of the Year
Kim Perrot Sportsmanship
Kim Perrot Sportsmanship Award
Most Improved Player
Most Valuable Player
Rookie of the Year
Sixth Woman of the Year
WNBA Finals Most Valuable Player
WNBA All-Decade Team
WNBA All Decade Team Honorable Mention
Award
Count
0
2
4
6
8
10

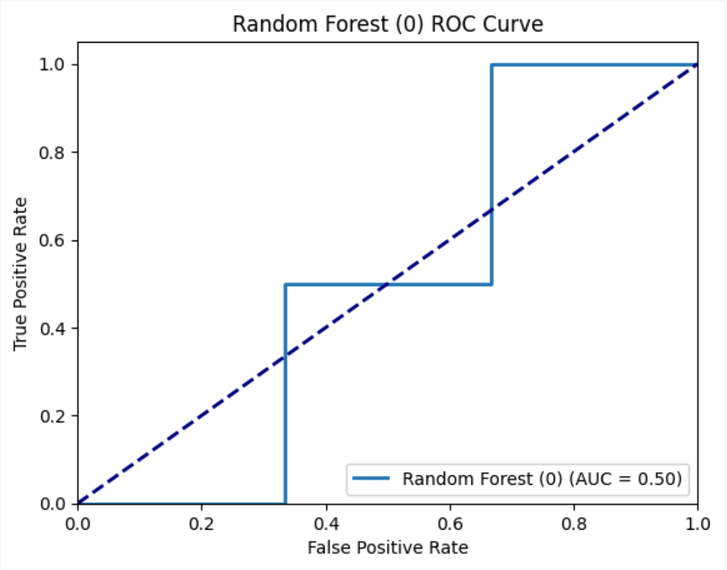

playoff
0.08 1.00 -0.86 0.25 0.07 0.10 0.07 0.06 0.01 0.09 0.31 0.28 0.30 -0.20 0.16 -0.05 0.18 0.22 -0.13 0.10 -0.23 -0.22 -0.01 0.16 -0.08 -0.13 -0.13 -0.23 0.09 -0.07 0.05 -0.27 -0.16
0.76 -0.74 0.09 0.67 -0.66 0.66 -0.63 0.73 -0.74 0.09 0.25
year playoff rank o_fgm o_fga o_ftm o_fta o_3pm o_3pa o_oreb o_dreb o_reb o_asts o_pf o_stl o_to o_blk o_pts d_fgm d_fga d_ftm d_fta d_3pm d_3pa d_oreb d_dreb d_reb d_asts d_pf d_stl d_to d_blk d_pts tmORB tmDRB tmTRB opptmORB opptmDRB opptmTRB won lost GP homeW homeL awayW awayL confW confL min attend
1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75

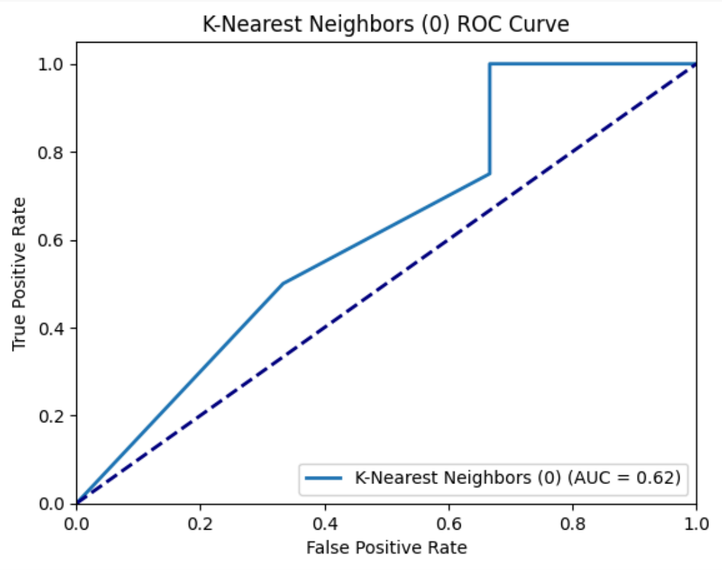
East

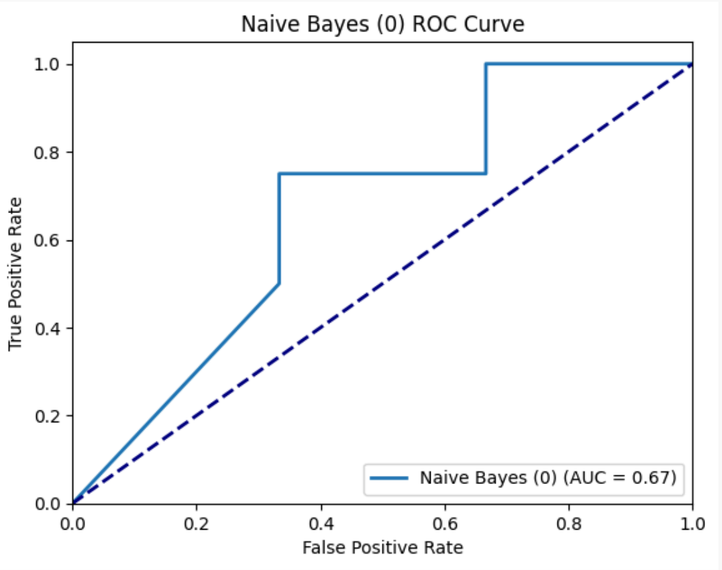

Logistic Regression (0) ROC Curve
True Positive Rate
False Positive Rate
Logistic Regression (0) (AUC = 0.67)


Support Vector Machine (0) ROC Curve
True Positive Rate
False Positive Rate
Support Vector Machine (0) (AUC = 0.67)

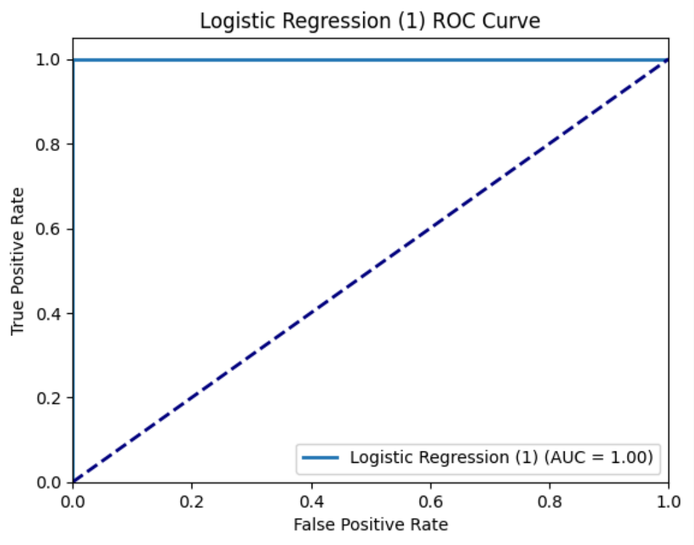

Decision Tree (0) ROC Curve
True Positive Rate
False Positive Rate
Decision Tree (0) (AUC = 0.71)

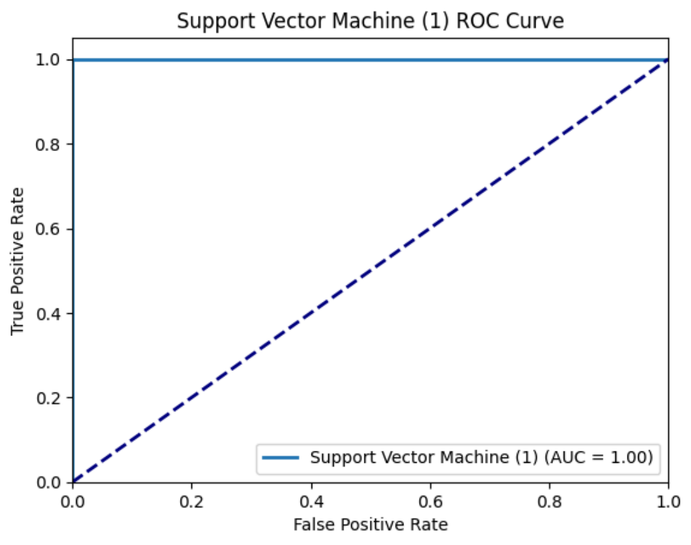

Random Forest (0) ROC Curve
True Positive Rate
False Positive Rate
Random Forest (0) (AUC = 0.50)

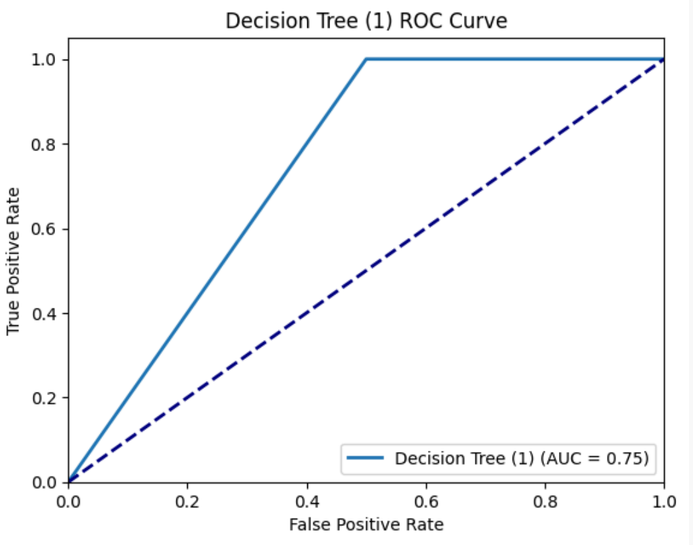

K-Nearest Neighbors (0) ROC Curve
True Positive Rate
False Positive Rate
K-Nearest Neighbors (0) (AUC = 0.62)

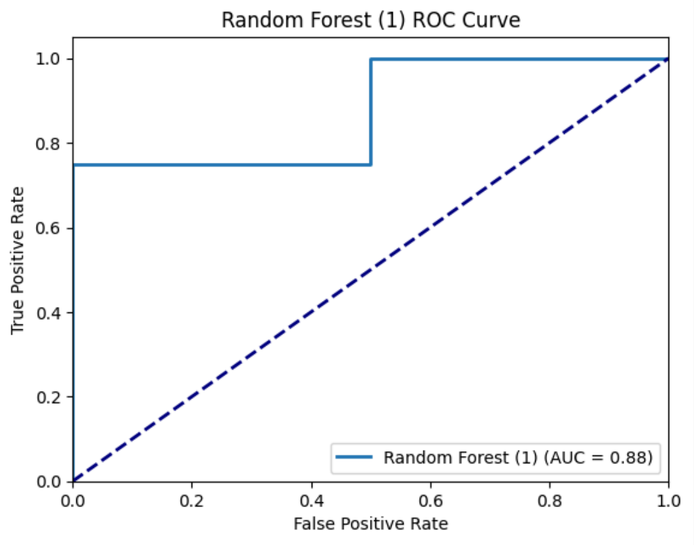

Naive Bayes (0) ROC Curve
True Positive Rate
False Positive Rate
Naive Bayes (0) (AUC = 0.67)

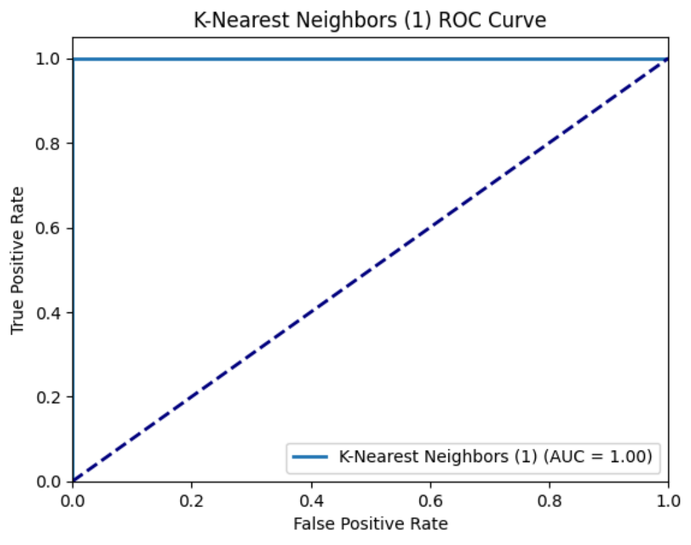
West

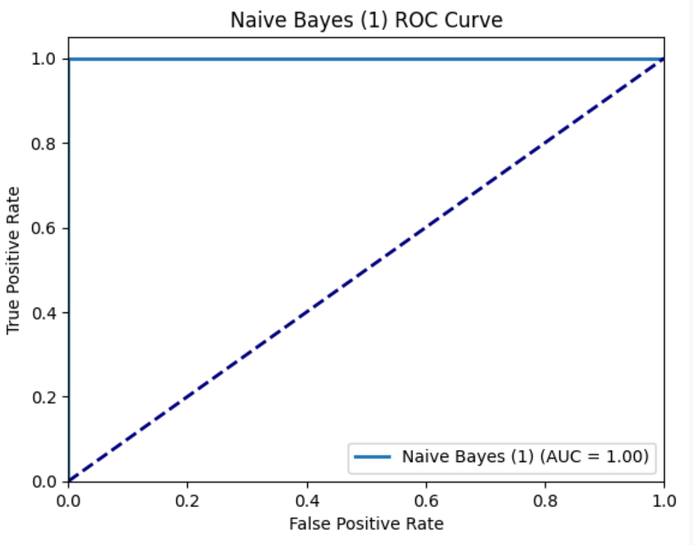

Logistic Regression (1) ROC Curve
True Positive Rate
False Positive Rate
Logistic Regression (1) (AUC = 1.00)

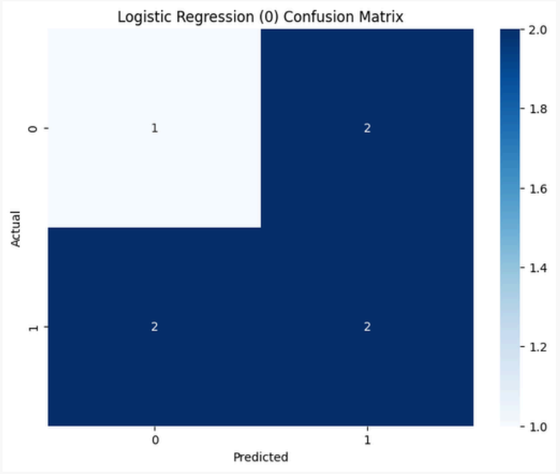

Support Vector Machine (1) ROC Curve
True Positive Rate
False Positive Rate
Support Vector Machine (1) (AUC = 1.00)

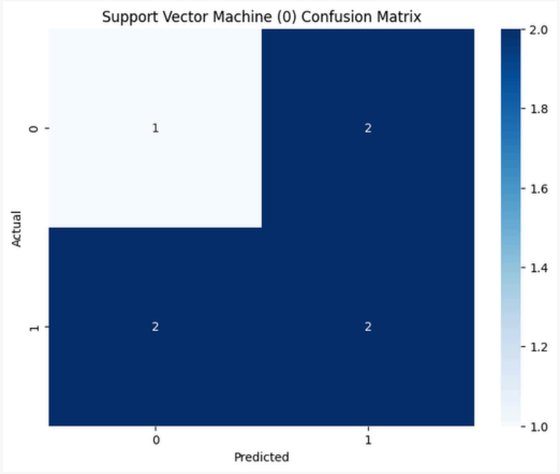

Decision Tree (1) ROC Curve
True Positive Rate
False Positive Rate
Decision Tree (1) (AUC = 0.75)

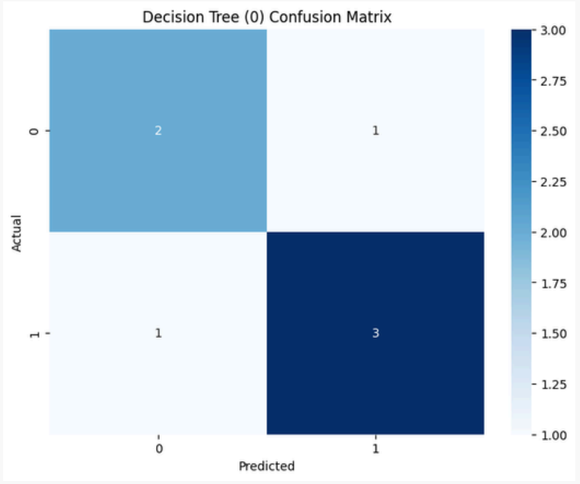

Random Forest (1) ROC Curve
True Positive Rate
False Positive Rate
Random Forest (1) (AUC = 0.88)

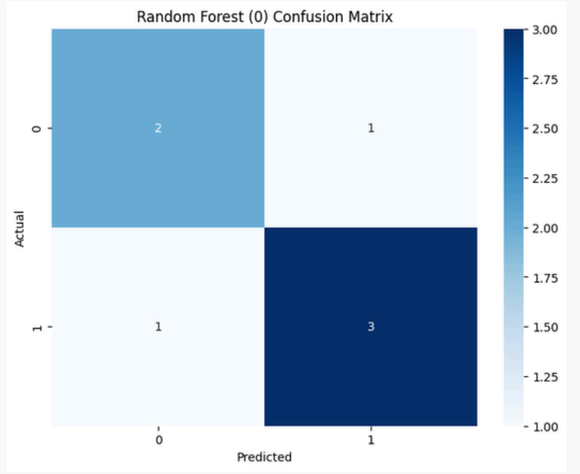

K-Nearest Neighbors (1) ROC Curve
True Positive Rate
False Positive Rate
K-Nearest Neighbors (1) (AUC = 1.00)

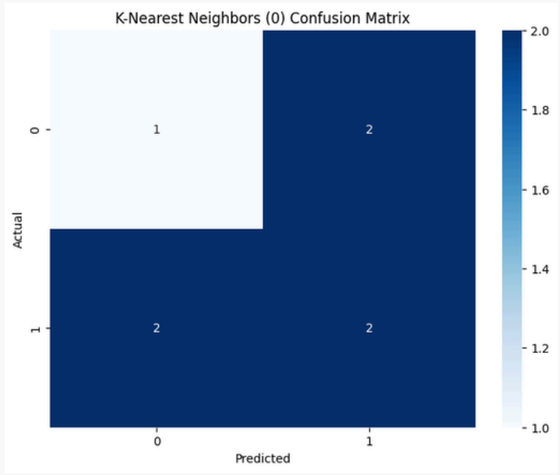

Naive Bayes (1) ROC Curve
True Positive Rate
False Positive Rate
Naive Bayes (1) (AUC = 1.00)

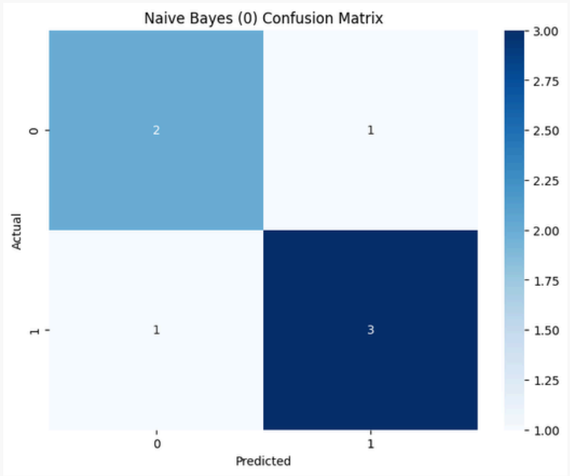

Logistic Regression (0) Confusion Matrix
Actual
Predicted
0
1
1
2
2
2

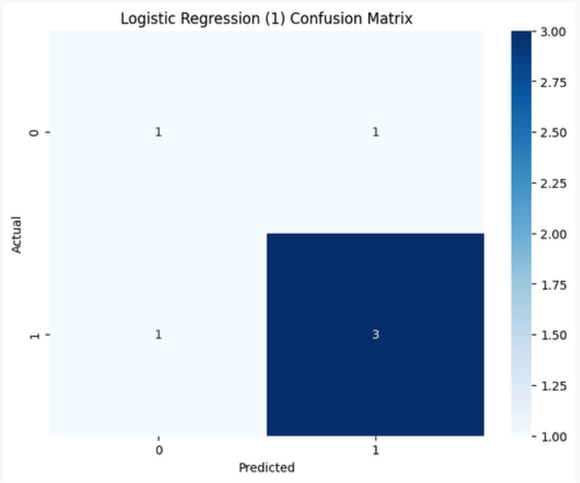

Support Vector Machine (0) Confusion Matrix
Actual
Predicted
0
1
1
2
2
2

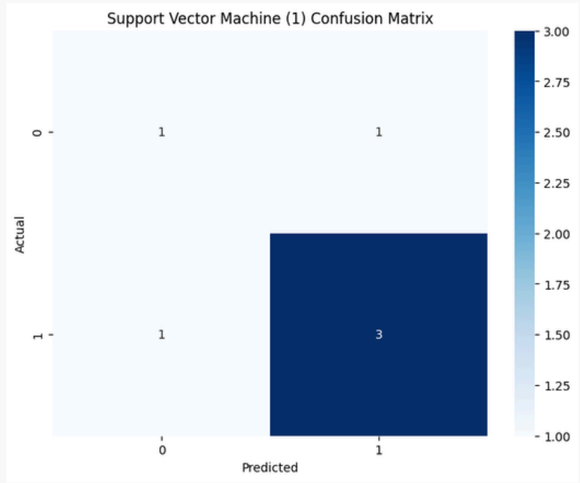

Decision Tree (0) Confusion Matrix
Actual
Predicted
0
1
2
1
1
3

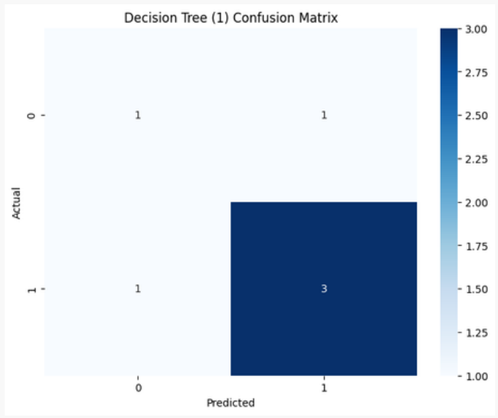

Random Forest (0) Confusion Matrix
Actual
Predicted
0
1
2
1
1
3

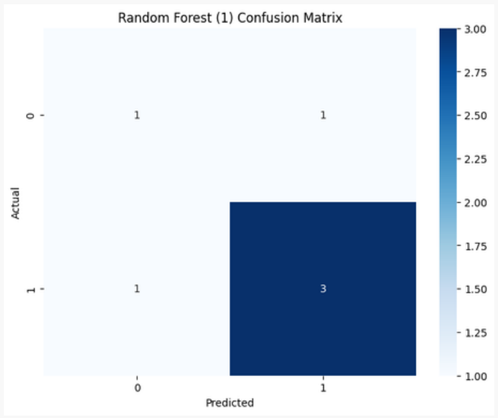

K-Nearest Neighbors (0) Confusion Matrix
Actual
Predicted
0
1
1
2
2
2

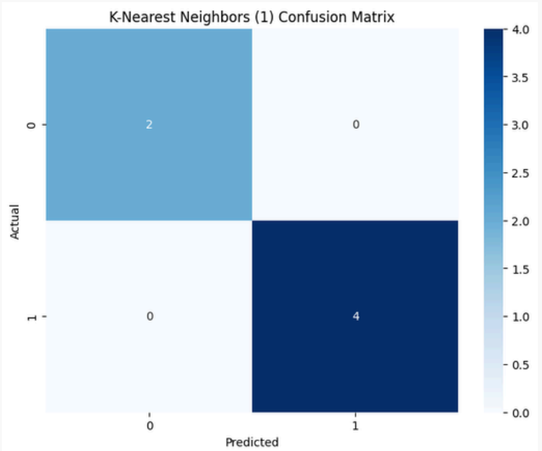

Naive Bayes (0) Confusion Matrix
Actual
Predicted
0
1
2
1
1
3

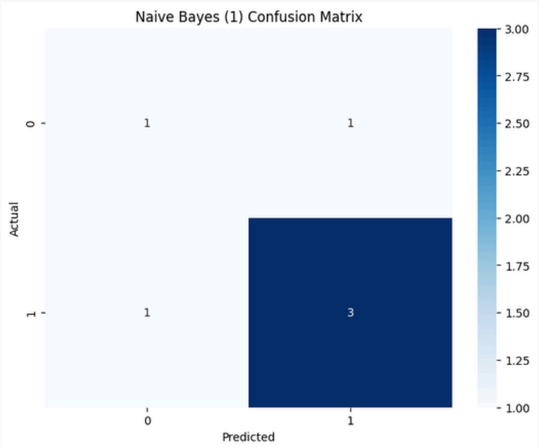

Logistic Regression (1) Confusion Matrix
Actual
Predicted
0
1
1
1
1
3


Support Vector Machine (1) Confusion Matrix
Actual
Predicted
0
1
1
1
1
3


Decision Tree (1) Confusion Matrix
Actual
Predicted
0
1
1
1
1
3


Random Forest (1) Confusion Matrix
Actual
Predicted
0
1
1
1
1
3


K-Nearest Neighbors (1) Confusion Matrix
Actual
Predicted
0
1
2
0
0
4


Naive Bayes (1) Confusion Matrix
Actual
Predicted
0
1
1
1
1
3