Beyond the Net: Analyzing Critical Performance Drivers for Orlando City Men's Soccer

Abigail Wright, Georgina Wu, Oma Persaud, and Shivani Ananth

1. Abstract
2. Introduction and Motivation
3. Data Description
4. Exploratory Data Analysis
5. Model Diagnostics
6. Model Selection
7. Model Reliability
8. Results, Summary, and Interpretations
9. Conclusion and Limitations
10. Appendix

1. Abstract

2. Introduction

2.1 History of Soccer

The complex game coined as soccer in the United States has been traced back more than 2000 years ago in ancient day Rome, China, and Greece (https://www.bundesliga.com/en/faq/all-you-need-to-know-about-soccer/the-history-of-soccer-10560). Though the game has developed with the growth in popularity key concepts still remain today. Across the globe soccer has taken many forms and names throughout the history of this sport. England shows roots of football dating by to the 19th century where it was later spread to Europe and so on (https://www.bundesliga.com/en/faq/all-you-need-to-know-about-soccer/the-history-of-soccer-10560). According to the USA Soccer Forward Foundation, in 1862 Gerritt Miller Smith organized the first soccer club in America, The Oneidas of Boston. This study will dive into the key factors influencing the outcome; win, loss, or draw for the 2024 Regular Season of the Orlando City Men's Soccer Team. The direction of this research strives to provide a clear basis of understanding concerning the game of soccer as well as the influences game statistics have on the outcome.

2.2 Study Motivation

As of August 2024, 37% of adults 18 years and older considered themselves soccer fans in the US according for Forbes (https://www.forbes.com/sites/vitascarosella/2024/08/06/the-arms-race-for-the-hearts-and-minds-of-american-soccer-fans/). Since soccer is a complex sport with various levels and factors influencing the outcome of the game. Creating this foundation of understanding can be beneficial to fans, team members, coaching staff, as well as many others. The research team utilized this motivation to investigate the significance of over twenty variables and determine each predictor's significance to the game outcome.

3. Data Description

3.1 Introduction to Dataset

This dataset was carefully compiled from various online statistical logs by our team. Some of these sources included the Orlando City Men's Soccer Team website, ESPN, among many others. The purpose of this research is to provide a better understanding of the factors that significantly impact the outcome of a game for the 2024 Regular Season of the Orlando City Men's Soccer Team. These matches began in February of 2024 and concluded in October of that same year. US soccer matches consist of 11 players from each team at the beginning of the game, with 5 allotted player substitutions within the 90-minute period. This data encompasses multiple players for the 34 recorded matches. With this data in mind the research team determined to include player evidence or absence in each match pertaining to the top three scoring members and top three assisting members of the Orlando City Team. Though this dataset does not include the entire statistical analysis for the Orlando City roaster, including these six significant members could shed light on this complex topic.

3.2 Predictors and Dependent Variables

The predictors in this research project are: Home/ Away, Season, Possession, Shots, Shots on Goal, Blocked Shots, Total Passes, Passing Accuracy, Corners, Total Crosses, Offsides, Aerial Duals Wons, Goalkeeper Saves, Clearances, Fouls, Yellow Cards, Red Cards, OP Team, Season, Facundo Torres, Duncan McGuire, Ramiro Enrique, Martín Ojeda, Nicolás Lodeiro and Iván Angulo. The dependent variable this study is striving to investigate is Final, pertaining to the final score of the game and if a winner resulted from the game. The definition and explanation of these variables will be thoroughly outlined within the Data Dictionary section of the Appendix.

4. Exploratory Data Analysis

4.1 Variables

For the purposes of Exploratory Data Analysis and Model Diagnostics some variables were generalized to minimize the number of factors/ levels.

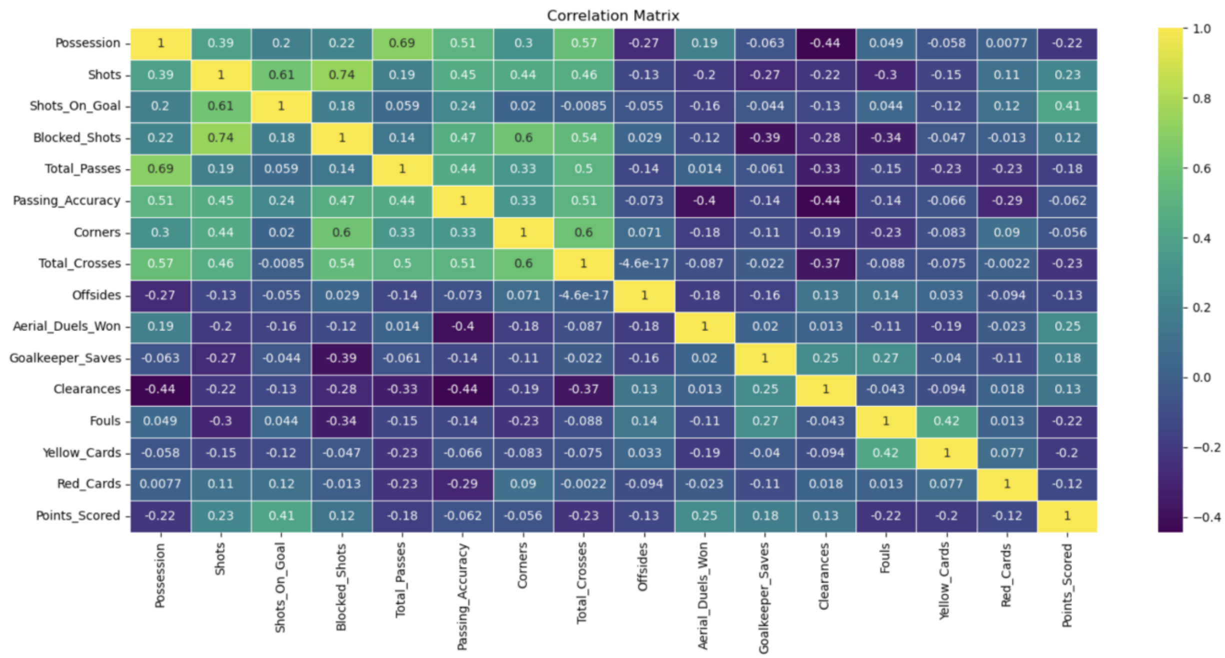
Opposing Teams
Teams
US Team
Non-US Team

Game Day
Seasons
Winter
Spring

	Summer
	Fall

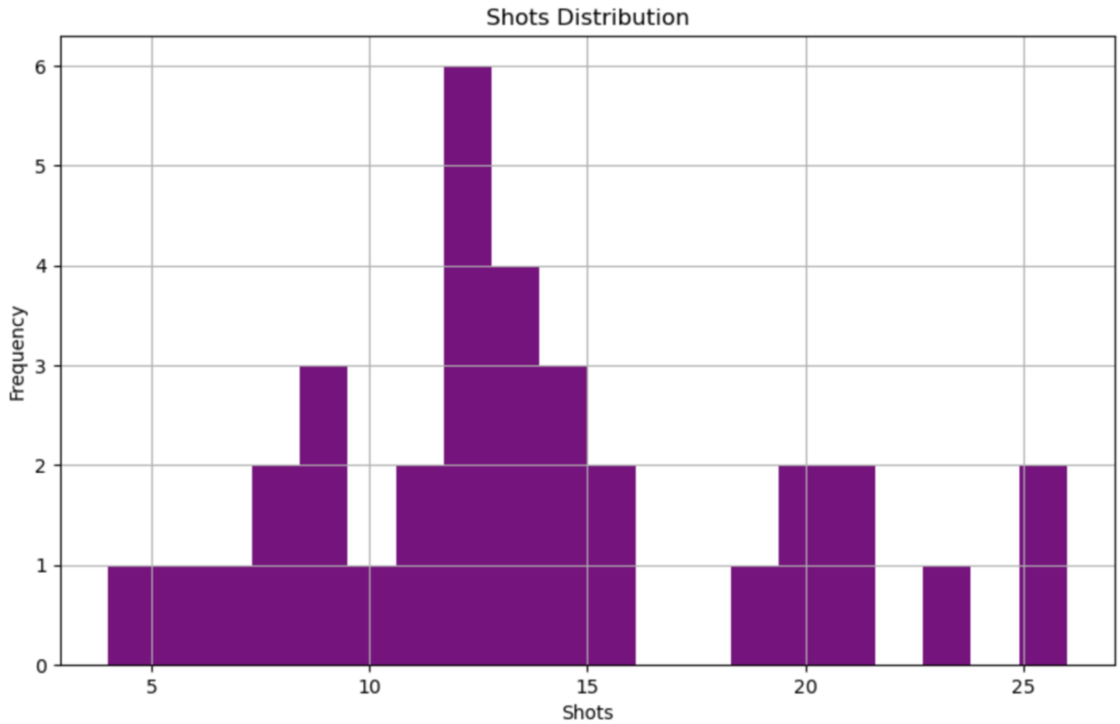
Multiple other variables were in the original dataset

4.2 Plots and Importance



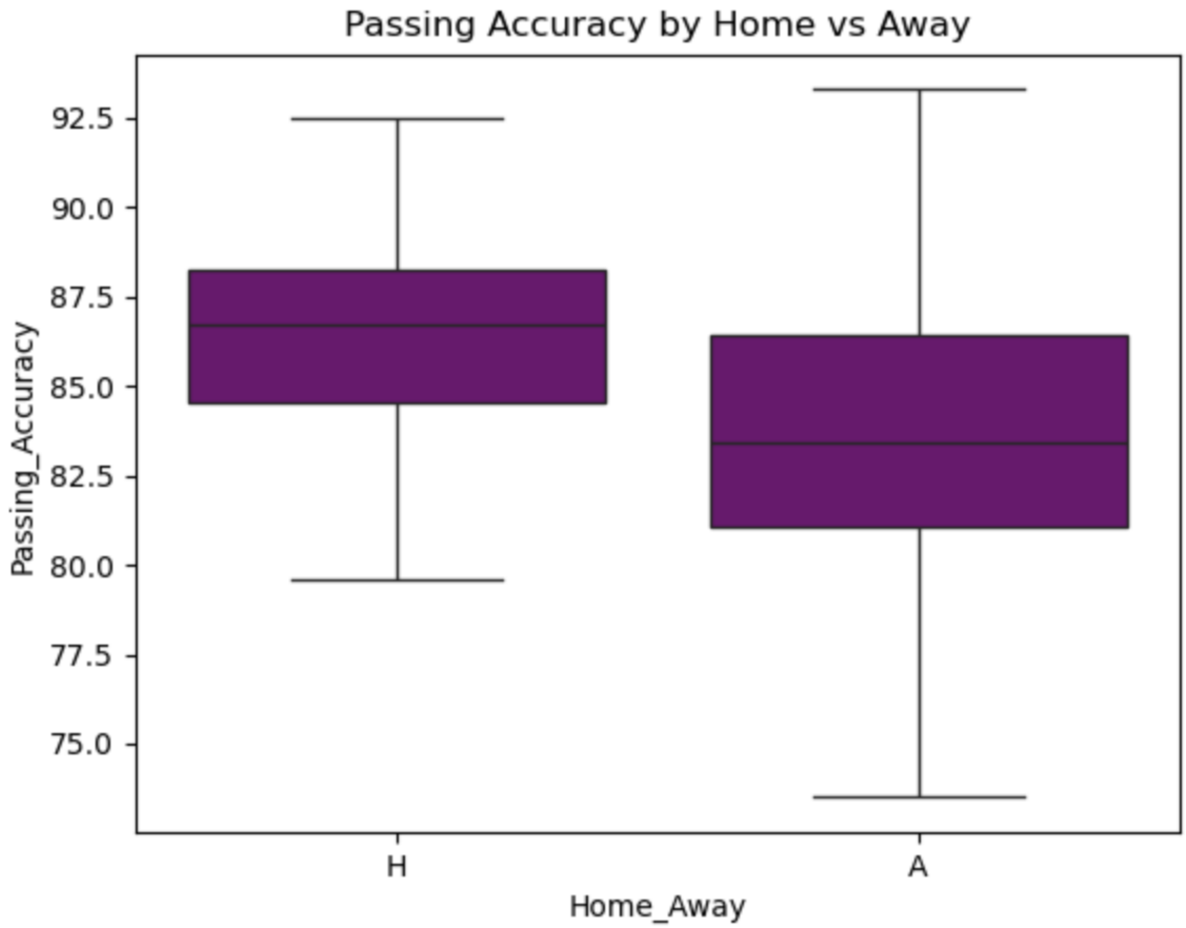
Correlation Matrix of Numeric Variables:

Correlation matrices help to understand the relationship present in the dataset and allows to find potential issues such as multicollinearity and redundancy between variables. Our highest correlation is 74% between the number of blocked shots and the total amount of shots. Since the correlation between the numeric variables does not exceed 80%, there is not likely to be multicollinearity issues or redundancy where two or more variables may be able to be combined.



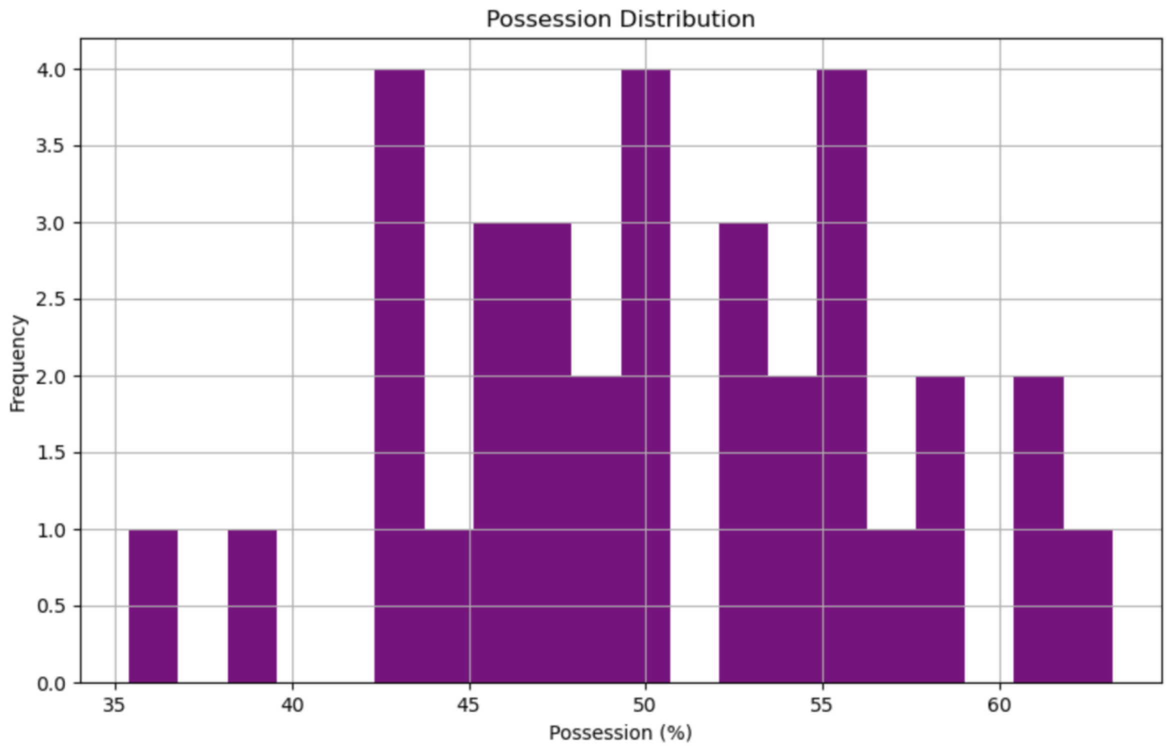
Shots Distribution Bar Chart:

The bar chart captures the frequency of how many shots are taken in each game. So, for each game recorded in the dataset, we see how many shots are more regularly taken. In the case of Orlando City Soccer, the most frequent number of shots that they take is between 10 to 15 shots per game. This could be a large deciding factor in the model and predicting if they can win a game because if they only are able to take about 13 shots consistently, then the team would need to make those shots count by making it into the goal.



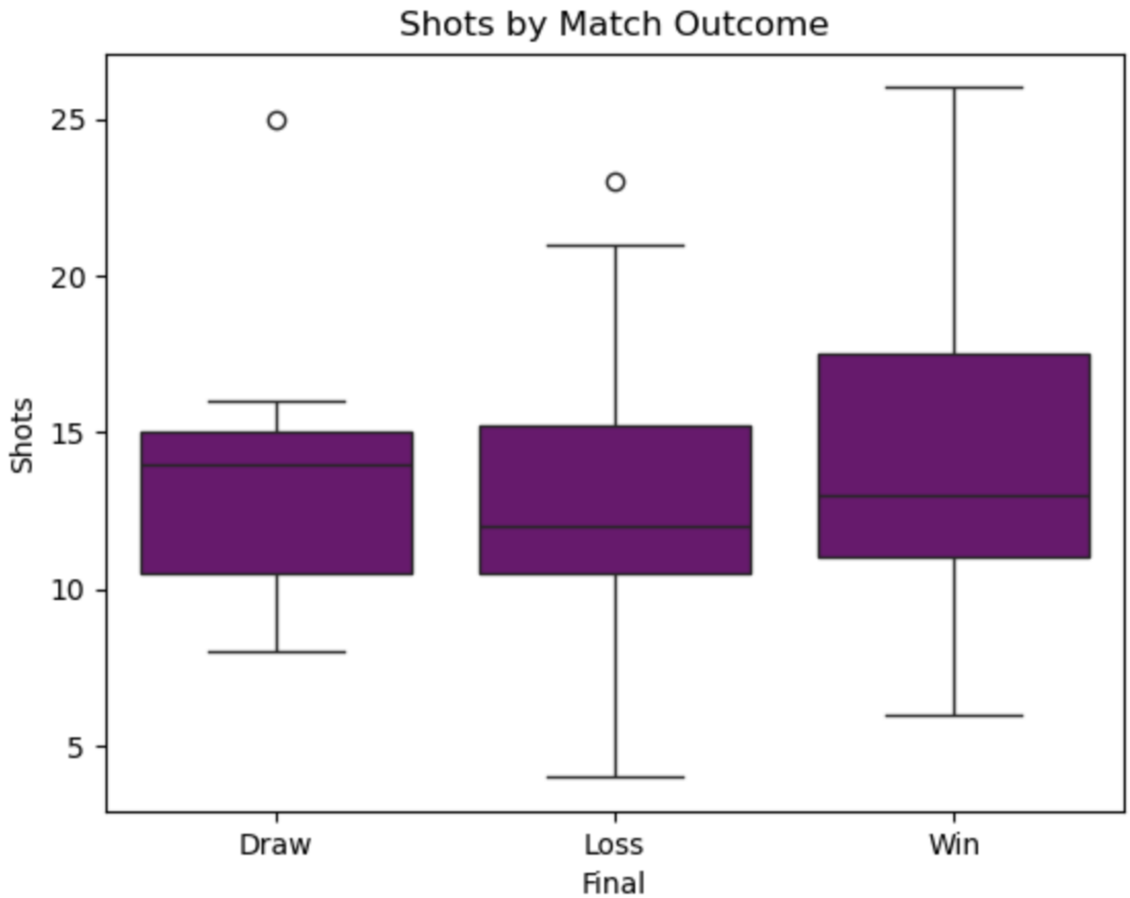
Passing Accuracy Based on If It's an Away or Home Game Box Plot:

The passing accuracy is high for both away and home games, however there is a significant difference from when the team is playing in their home field versus away. In the away games, there is a consistent passing accuracy of about 80% while home games have a much higher regular rate of about 88%. Depending on how many passes the team is making in one game, it could be the difference between a win and a loss.



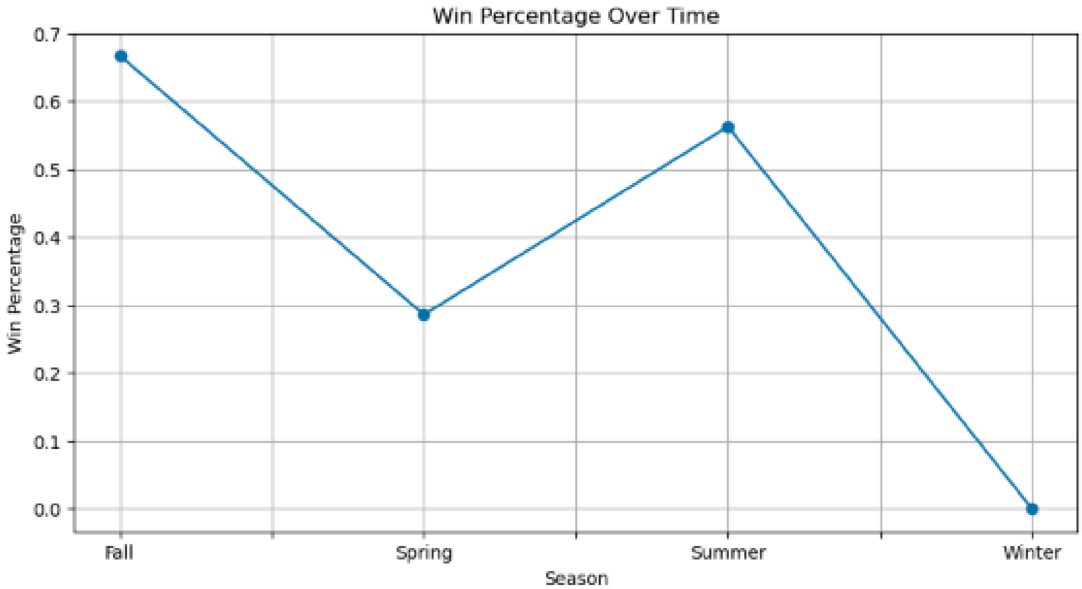
Possession Distribution Bar Chart:

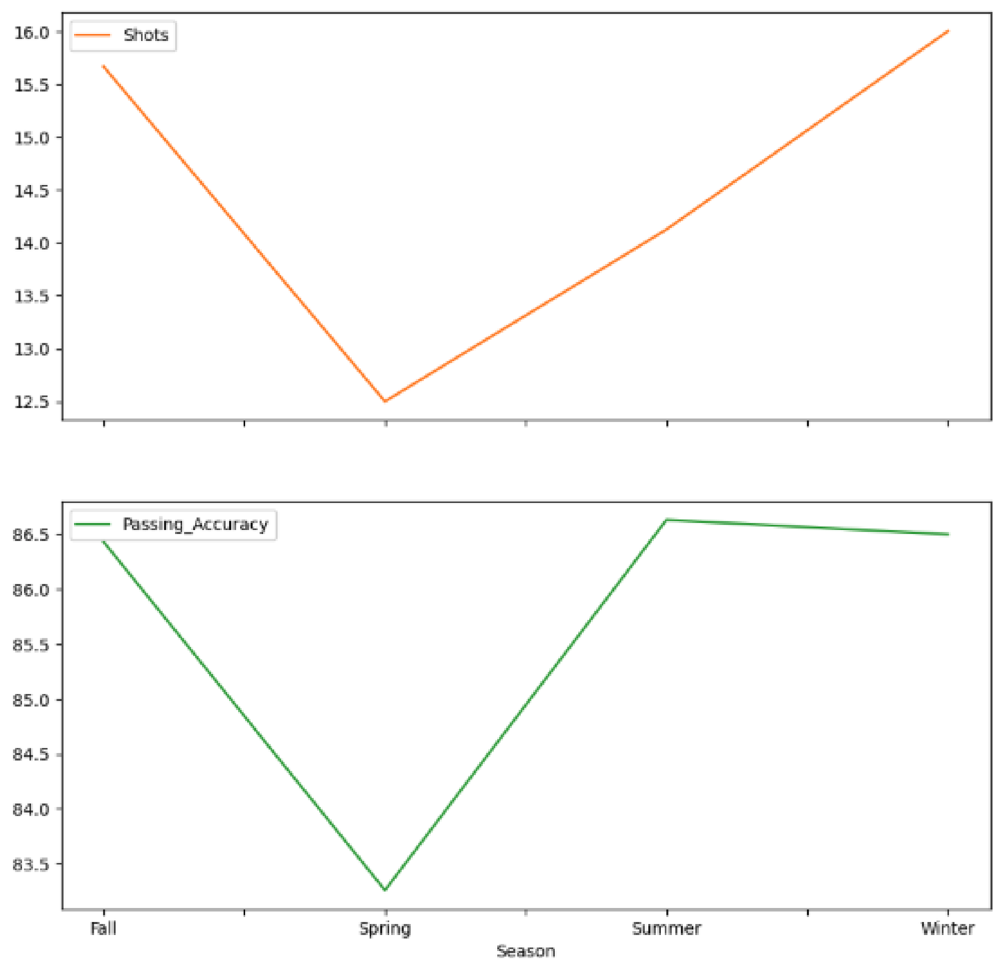
The amount of time that the team has the ball in their playing time is another important factor that should be considered since if a team does not have much time attempting to score, the less likely they'd be to get enough points to win. The spread of possession percentage does not have one consistent range as they have been as low as only having the ball for 35% to as high as over 60% of the game. This could be due to who the opposing team they are playing is or which of the key players they have on the field at the time and will be worth looking further into to see if this really is a significant attribute to winning a game.



Shots by Match Outcome Box Plot:

While we know from the Shots Distribution Bar Chart that the typical number of shots for a game is about 10 to 15 shots, this looks at whether the number of shots really does matter to winning a game. In the case of a game being won, there is a larger spread of the number of shots taken, however the most common number of shots is still about 13. The case of the game being lost typically seems to have a lower number of shots more consistently, but in the event of a draw, the number of shots on average is higher than in both cases of if it was a loss or win.


Win Percentage Over Time
Win Percentage
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Fall
Spring
Summer
Winter
Season


Shots
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
Passing_Accuracy
83.5
84.0
84.5
85.0
85.5
86.0
86.5
Fall
Spring
Summer
Winter
Season

```python
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report

features = data[['Possession', 'Shots_On_Goal', 'Passing_Accuracy']]
target = data['Win']  # Binary outcome: 1 for win, 0 for loss/draw
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.2, random_state=42)

model = LogisticRegression()
model.fit(X_train, y_train)

# Predicting and evaluating the model
predictions = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, predictions))
print(classification_report(y_test, predictions))
```

```
Accuracy: 0.5714285714285714
              precision    recall  f1-score   support

       False       0.67      0.50      0.57         4
        True       0.50      0.67      0.57         3

    accuracy                           0.57         7
   macro avg       0.58      0.58      0.57         7
weighted avg       0.60      0.57      0.57         7
```

The results shown in the image indicate that a logistic regression model was used to predict whether a soccer team would win based on three game statistics: possession, shots on goal, and passing accuracy. The model's accuracy, or the proportion of correct predictions it made, is about 57.1%, which means it correctly predicted the outcome (win or not win) for approximately 57 out of every 100 games. When breaking down the predictions by actual outcomes, the model was correct about 67% of the time when predicting a team would not win, and about 58% of the time when predicting a win. The F1-score, which balances precision and recall, is around 0.57 for both outcomes, suggesting the model is moderately effective but could benefit from improvements to increase its predictive accuracy.

4.3 Methodology

5. Model Diagnostics

5.1 Utilized Variables

5.2 Outlier Detection and Removal

```
Outliers for Possession are located at indices: Index([], dtype='int64')
Outliers for Shots are located at indices: Index([17, 21], dtype='int64')
Outliers for Shots_On_Goal are located at indices: Index([], dtype='int64')
Outliers for Blocked_Shots are located at indices: Index([21], dtype='int64')
Outliers for Total_Passes are located at indices: Index([8], dtype='int64')
Outliers for Passing_Accuracy are located at indices: Index([], dtype='int64')
Outliers for Corners are located at indices: Index([], dtype='int64')
Outliers for Total_Crosses are located at indices: Index([16, 17], dtype='int64')
Outliers for Offsides are located at indices: Index([0, 13, 29], dtype='int64')
Outliers for Aerial_Duels_Won are located at indices: Index([], dtype='int64')
Outliers for Goalkeeper_Saves are located at indices: Index([18], dtype='int64')
Outliers for Clearances are located at indices: Index([], dtype='int64')
Outliers for Fouls are located at indices: Index([], dtype='int64')
Outliers for Yellow_Cards are located at indices: Index([], dtype='int64')
Outliers for Red_Cards are located at indices: Index([2, 9], dtype='int64')
Outliers for Points Scored are located at indices: Index([], dtype='int64')
Detected outliers in the dataset:
   Final Home_Away  Possession  Shots  Shots_On_Goal  Blocked_Shots  \
0   Draw         H        57.7     16              4              5
2   Loss         H        55.0     23              8              4
8   Draw         A        46.7      9              4              2
9   Loss         H        46.2      9              4              3
13  Loss         H        42.6     12              5              4
16  Loss         H        55.7     19              1              8
17  Draw         A        60.5     25              9              7
18   Win         H        42.4      7              5              0
21   Win         H        55.6     26             10             10
29  Loss         A        35.4     12              4              4

    Total_Passes  Passing_Accuracy  Corners  Total_Crosses  ...  \
0            579              86.5        9             25  ...
2            451              79.6        8              9  ...
8            746              80.9        8              6  ...
9            362              80.7        5             13  ...
13           435              88.5        9             12  ...
16           627              92.5        9             31  ...
17           654              93.3       11             28  ...
18           382              83.8        1              1  ...
21           561              91.3       12             15  ...
29           382              83.5        3              5  ...

    Facundo_Torres  Duncan_McGuire  Ramiro_Enrique  Martín_Ojeda  \
0                Y               Y               Y             Y
2                N               Y               Y             Y
8                Y               Y               N             Y
9                Y               Y               N             Y
13               Y               N               Y             Y
16               Y               Y               N             Y
17               Y               Y               N             Y
18               Y               Y               N             Y
21               Y               Y               Y             Y
29               Y               Y               Y             Y

    \tNicolás_Lodeiro  Iván_Angulo     OP_Team   Game_Day Points Scored    Win
0                   Y            Y    Montreal 2024-02-24             0  False
2                   Y            Y   Minnesota 2024-03-09             2  False
8                   Y            Y    Montreal 2024-04-30             2  False
9                   Y            Y  Cincinnati 2024-05-04             0  False
13                  Y            Y    Columbus 2024-05-25             0  False
16                  Y            Y        LAFC 2024-06-15             1  False
17                  Y            Y   Charlotte 2024-06-19             2  False
18                  Y            Y     Chicago 2024-06-22             4   True
21                  Y            Y        D.C. 2024-07-06             5   True
29                  Y            Y    Columbus 2024-09-21             3  False
```

Specific Outliers Detected:

- **Shots**: Games at indices 17 and 21 had unusually high numbers of shots.
- **Blocked Shots**: Game at index 21 had an unusually high number of blocked shots.
- **Total Passes**: Game at index 8 had an unusually high number of total passes.
- **Total Crosses**: Games at indices 16 and 17 had unusually high numbers of total crosses.
- **Offsides**: Games at indices 0, 13, and 29 had unusually high numbers of offsides.

- **Goalkeeper Saves**: Game at index 18 had an unusually high number of goalkeeper saves.
- **Red Cards**: Games at indices 2 and 9 had red cards, which are less common events.

Outliers in our dataset likely represent real and meaningful events during Orlando City's matches, such as an unusually high number of shots, crosses, or a red card. These moments can significantly impact the outcome of a game and reflect strategic or situational changes rather than data errors. Since we're working with a small number of matches, removing these points could hide important insights. Instead of deleting them, it's more useful to keep them in the data and consider their influence when analyzing patterns or building models.

5.3 Transformations and Interactions

6. Model Selection

6.1 Full Model

To establish a baseline for evaluating the impact of feature selection, we constructed models using four different classification algorithms and one clustering algorithm: Logistic Regression, Naive Bayes, Support Vector Machines (SVM), Random Forest Model, and Hierarchical Clustering, respectively. Each of these models were trained using the complete set of predictor variables, to predict whether the Orlando City Men's Soccer Team would 'Win', 'Lose', or 'Draw'.

These are the libraries we imported to create and evaluate the models.

```python
import pandas as pd
from sklearn.model_selection import train_test_split, KFold, cross_val_score
from sklearn.preprocessing import LabelEncoder, StandardScaler, OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.feature_selection import RFE
from sklearn.metrics import accuracy_score, classification_report
import statsmodels.api as sm  # For p-values in Logistic Regression
from sklearn.svm import SVC
from sklearn import metrics
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import GridSearchCV
```

6.1.1 Logistic Regression

A Logistic Regression model was implemented using the 'liblinear' solver. We can see from the results that the accuracy on the full test set was 0.545, or 54.5%. This is a slight deterioration in accuracy compared to the earlier Logistic Regression model fitted in Section 4. The model predicted winning with 100% precision, but losing with only 43% precision, and a draw with 0% precision. The recall scores are almost opposite for win and lose, but the same for draw. Ultimately, the low accuracy indicates the same conclusion we came to in Section 4, which is that the model could benefit from improvements to increase its predictive accuracy.

```python
# --- Logistic Regression Model with All Predictors ---

# Create preprocessor pipeline
preprocessor_pipeline = Pipeline(steps=[('preprocessor', preprocessor)])

X_train_processed = preprocessor_pipeline.fit_transform(X_train)
X_test_processed = preprocessor_pipeline.transform(X_test)

# Evauate the model
print("\n--- Logistic Regression Model with All Predictors ---")
logistic_model_all = LogisticRegression(solver='liblinear', random_state=42,
                                        multi_class='auto')
logistic_model_all.fit(X_train_processed, y_train)
y_pred_all = logistic_model_all.predict(X_test_processed)
accuracy_all = accuracy_score(y_test, y_pred_all)
print(f"Accuracy on the test set: {accuracy_all:.3f}")
print(classification_report(y_test, y_pred_all))
```

```
--- Logistic Regression Model with All Predictors ---
Accuracy on the test set: 0.545
              precision    recall  f1-score   support

        Draw       0.00      0.00      0.00         3
        Loss       0.43      1.00      0.60         3
         Win       1.00      0.60      0.75         5

    accuracy                           0.55        11
   macro avg       0.48      0.53      0.45        11
weighted avg       0.57      0.55      0.50        11
```

6.1.2 Naive Bayes Model

A Gaussian Naive Bayes model was implemented using the 'GaussianNB' function from 'sklearn'. We can see from the results that this model had an accuracy of 0.545, the same as the above Logistic Regression Model. The precision and recall resulted slightly differently, with draw no longer being 0, but everything else has a slightly lower precision and recall. The f1-score for Win is, however, higher than that of the logistic regression model. If we are focusing on win only, the Naive Bayes Model performs slightly better than the Logistic regression model even though they have the same accuracy. The Naive Bayes Model could still benefit from improvements to increase its accuracy.

```python
# Create the Naive Bayes model pipeline
naive_bayes_pipeline = Pipeline(steps=[('preprocessor', preprocessor),
                                       ('classifier', GaussianNB())])

# Train the Naive Bayes model
naive_bayes_pipeline.fit(X_train, y_train)

# Make predictions on the test set
y_pred = naive_bayes_pipeline.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print("\n--- Naive Bayes Model with All Predictors ---")
print(f"Accuracy on the test set: {accuracy:.3f}")
print("\nClassification Report:")
print(classification_report(y_test, y_pred))
```

```
--- Naive Bayes Model with All Predictors ---
Accuracy on the test set: 0.545

Classification Report:
              precision    recall  f1-score   support

        Draw       0.25      0.33      0.29         3
        Loss       0.33      0.33      0.33         3
         Win       1.00      0.80      0.89         5

    accuracy                           0.55        11
   macro avg       0.53      0.49      0.50        11
weighted avg       0.61      0.55      0.57        11
```

6.1.3 Support Vector Machine (SVM) Model

The SVM model was implemented using the ‘SVC’ function from ‘sklearn’ with ‘random_state’ set to 42 for consistency. The SVM model also had the same accuracy as the above two models, performing in-between that of Logistic Regression and Naive Bayes. The SVM model did not predict Draw with any accuracy, but Loss with some precision (0.40), and Win with more precision (0.67). The f1-score for Win is the lowest of the three models so far. Again, we conclude that the SVM Model could benefit from improvements to increase its accuracy.

```python
# Create the SVM model pipeline
svm_pipeline = Pipeline(steps=[('preprocessor', preprocessor),
                               ('classifier', SVC(random_state=42))])

# Train the SVM model
svm_pipeline.fit(X_train, y_train)

# Make predictions on the test set
y_pred = svm_pipeline.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print("--- Support Vector Machine (SVM) Model with All Predictors ---")
print(f"Accuracy on the test set: {accuracy:.3f}")
print("\nClassification Report:")
print(classification_report(y_test, y_pred))
```

```
--- Support Vector Machine (SVM) Model with All Predictors ---
Accuracy on the test set: 0.545

Classification Report:
              precision    recall  f1-score   support

        Draw       0.00      0.00      0.00         3
        Loss       0.40      0.67      0.50         3
         Win       0.67      0.80      0.73         5

    accuracy                           0.55        11
   macro avg       0.36      0.49      0.41        11
weighted avg       0.41      0.55      0.47        11
```

6.1.4 Random Forest Model

The Random Forest Model was implemented using the ‘RandomForestClassifier’ function from ‘sklearn’. The Random Forest Model resulted in the lowest accuracy, 0.455, and also performed the poorest in every other evaluation measure. As such, although it could be improved, it is unlikely that we will move forward with this model method.

```python
# Create the Random Forest model pipeline
random_forest_pipeline = Pipeline(steps=[('preprocessor', preprocessor),
                                  ('classifier', RandomForestClassifier(random_state=42))])

# Train the Random Forest model
random_forest_pipeline.fit(X_train, y_train)

# Make predictions on the test set
y_pred = random_forest_pipeline.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print("--- Random Forest Model with All Predictors ---")
print(f"Accuracy on the test set: {accuracy:.3f}")
print("\nClassification Report:")
print(classification_report(y_test, y_pred))
```

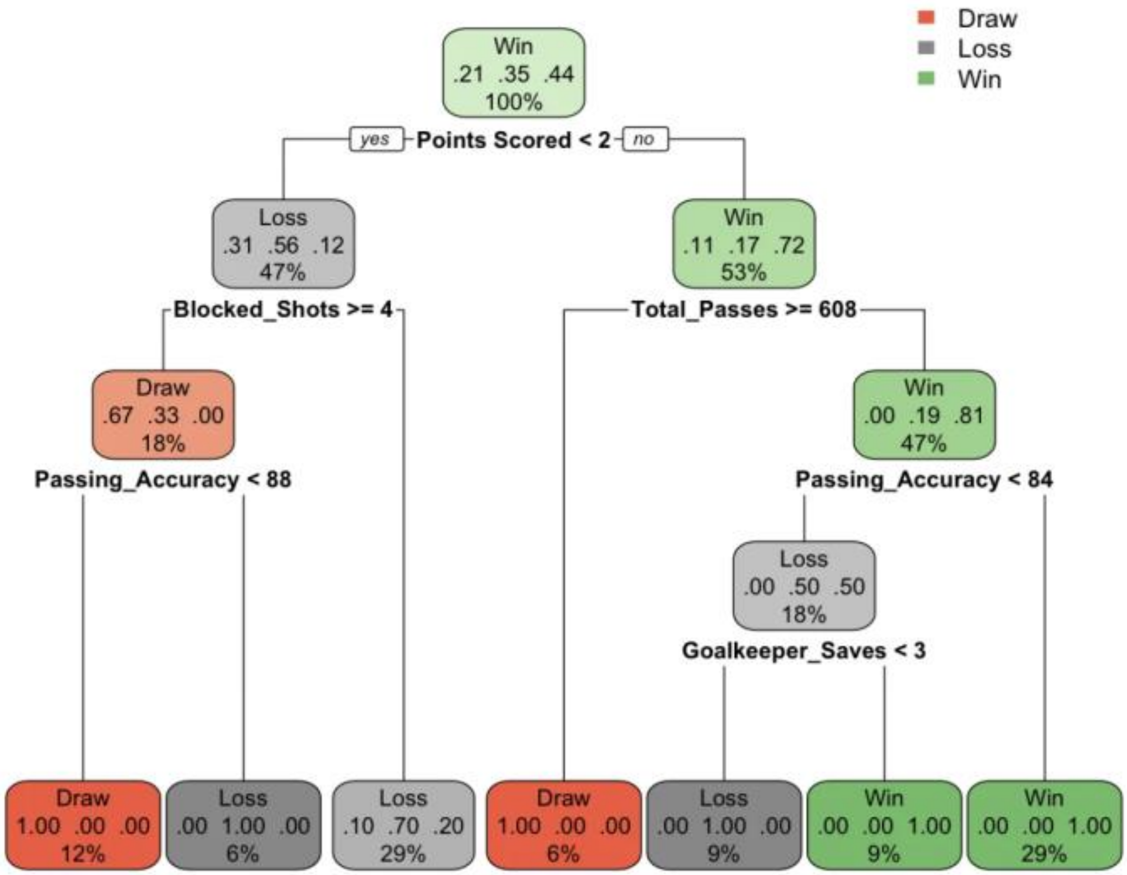
```
--- Random Forest Model with All Predictors ---
Accuracy on the test set: 0.455

Classification Report:
              precision    recall  f1-score   support

        Draw       0.00      0.00      0.00         3
        Loss       0.29      0.67      0.40         3
         Win       0.75      0.60      0.67         5

    accuracy                           0.45        11
   macro avg       0.35      0.42      0.36        11
weighted avg       0.42      0.45      0.41        11
```

6.1.5 Hierarchical Clustering



6.2 Reduced Model

After the EDA, we are finalizing the features with feature selection

7. Model Reliability

7.1 Probability Value

7.2 Model Efficacy

7.3 Pertinent Values & Variables

8. Results & Summary

8.1 Interpretations

9. Conclusion

9.1 Final Thoughts & Motivations

9.2 Study Limitations

Though this study highlighted many significant predictors for determining match outcome, many limitations affect this conclusion. Beginning the study containing over 20 predictive variables expanded the bounds of the predictive measures however countless other factors could be at play. In terms of conducting a further extensive study additional statistical data recommended include:

1. Team roaster- Orlando City and Opposing Teams:
    a. The duration of team member's time on current team.
    b. The career length pertaining to US Soccer/ soccer in general.
    c. Individual career statistics for each member during the time on the Orlando City team/ OP Team.
    d. Detailed statistics for the entirety of a player's soccer career.
2. Injury:

   a. Data pertaining to Orlando City team members and opposing team members that were injured prior to match, during the match, or injury occurred as a result of the match.
3. Coaching staff:
   a. The length of career concerning coaching for soccer.
   b. Match results during the coaching career.
   c. Player retention and turnover rate under individual's coaching.
4. Weather Conditions:
   a. As Florida is known for the heat and high humidity levels, this game day data could be beneficial. This is in tandem with conditions for Away games located in Northern regions where rain, snow, and sleet are common occurrences during the early and later months of the year.

10. Appendix

10.1 Work Cited

"A New Metric for Evaluating 1v1 Ability." *Stats Perform*, www.statsperform.com/resource/a-new-metric-for-evaluating-1v1-ability/.

"Glossary of Soccer Terms." *Marcus Jewish Community Center of Atlanta*, www.atlantajcc.org/clientuploads/PDF/SF/YLC/SOCCER/Fall%20Soccer/GlossarySoccerTerms.pdf.

"Opta Event Definitions." *Stats Perform*, www.statsperform.com/opta-event-definitions/.

"Opta Football Stats Definitions." *The Analyst*, www.theanalyst.com/2024/07/opta-football-stats-definitions/.

"Orlando City SC Player: Duncan McGuire." *Orlando City Soccer Club*, www.orlandocitysc.com/players/duncan-mcguire/.

"Orlando City SC Player: Facundo Torres." *Orlando City Soccer Club*, www.orlandocitysc.com/players/facundo-torres/.

"Orlando City SC Player: Iván Angulo." *Orlando City Soccer Club*, www.orlandocitysc.com/players/ivan-angulo/.

"Orlando City SC Player: Martín Ojeda." *Orlando City Soccer Club*, www.orlandocitysc.com/players/martin-ojeda/.

"Orlando City SC Player: Nicolás Lodeiro." *Orlando City Soccer Club*, www.orlandocitysc.com/players/nicolas-lodeiro/.

"Orlando City SC Player: Ramiro Enrique." *Orlando City Soccer Club*, www.orlandocitysc.com/players/ramiro-enrique/.

"Orlando City SC Schedule." *Orlando City Soccer Club*, www.orlandocitysc.com/schedule/.

"Orlando City SC Stats." *Orlando City Soccer Club*, www.orlandocitysc.com/stats/.

"Orlando City SC Team Statistics." *ESPN*, www.espn.com/soccer/team/stats/_/id/12011/league/USA.1/season/2024/view/scoring.

"Seasons in Florida." *Florida Smart*, www.floridasmart.com/seasons-in-florida.

"Soccer Statistics Manual." *National Collegiate Athletic Association*, fs.ncaa.org/Docs/stats/Stats_Manuals/Soccer/2009ez.pdf.

"Soccer Terms." *US Soccer Players*, www.ussoccerplayers.com/soccer-terms.

"The History of Soccer." *Bundesliga*, www.bundesliga.com/en/faq/all-you-need-to-know-about-soccer/the-history-of-soccer-10560.

"U.S. Soccer History Timeline." *U.S. Soccer*, www.ussoccer.com/history/timeline.

"Weather in Orlando." *TripSavvy*, www.tripsavvy.com/weather-in-orlando-2546241.

McKinley, Eliot. "Reinventing the Passing Wheel: What Determines a Good Passer?" *American Soccer Analysis*, www.americansocceranalysis.com/home/2018/10/11/whatdeterminesagoodpasser.

Scarosella, Vita. "The Arms Race for the Hearts and Minds of American Soccer Fans." *Forbes*, 6 Aug. 2024, https://www.forbes.com/sites/vitascarosella/2024/08/06/the-arms-race-for-the-hearts-and-minds-of-american-soccer-fans/.

New sources:

-

11.2 Data Dictionary

Predictors:

1. **Home / Away:**
    a. Home meaning the game is held at the Orlando City Inter&Co Stadium.
    b. Away meaning the soccer game is held at the opposing team's stadium.
2. **Possession:** "The percentage of one or more sequences in a row belonging to the same team. A possession is ended by the opposition gaining control of the ball. (https://www.statsperform.com/opta-event-definitions/#:~:text=A%20completed%20pass%20is%20a,Dribbles/Take%2Dons)"
3. **Shots:** "A shot is any attempt to score by a team. The shot can result in a wide or high ball (over the goal), a keeper save, a deflection by another player or the post, or a goal. A shot is not necessarily a shot on goal. (https://ussoccerplayers.com/soccer-terms)"
4. **Shots On Goal:** "A shot on goal is a shot that is on net. The results of a shot on goal must be either a save by the goalkeeper or defending team or a goal by the attacking

team. A shot that hits the post or crossbar without being deflected by a goalkeeper or defender and does not cross the goal line is not a shot on goal. (WEBSITE)"

5. **Blocked Shots:** "Is defined as an attempt to score including:
   - An attempt on target that is blocked by an outfield player, where other defenders or a goalkeeper are behind the 'blocker'.
   - A shot blocked unintentionally by the shooter's own teammate. (https://theanalyst.com/2024/07/opta-football-stats-definitions#:~:text=A%20blocked%20shot%20is%20defined,by%20the%20shooter's%20own%20teammate.)"
6. **Total Passes:** "Is known as the attempted delivery of the ball from one player to another player on the same team. A player can use any part of their body (permitted in the laws of the game), to execute a pass. Event categorization includes open play passes, Goal kicks, corners and free kicks played as a pass. Crosses, keeper throws and throw ins do not count as a pass. Opta adds a whole range of qualifiers to each pass event, so that various things can be measured.
   - Chipped Pass – A lofted ball with an intended recipient. Must be over shoulder height and using the passes height to avoid opposition players.
   - Headed Pass – A header when there is an intended recipient.
   - Launch – A long high ball into space or an area for players to chase or challenge the ball.
   - Flick-On – A glancing pass with head or foot onto a teammate when the ball is helped on in the same general direction.
   - Pull back – A pass inside the penalty area which is pulled back.

- Lay-Off – A first time pass away from Goal when there is pressure on the passer (typically played by a forward) with one touch when they have their back to Goal from the Goal-line to the center of the penalty area.
- Through Ball – A pass splitting the defense for a team mate to run on to.
- Tap Pass – A short pass after a dead ball situation which cannot have a lost outcome. (https://www.statsperform.com/opta-event-definitions/#:~:text=A%20completed%20pass%20is%20a,Dribbles/Take%2Dons)"

7. Passing Accuracy: "Is determined by calculating the percentage of completed passes out of the total attempted passes, which is a simple formula of successful passes divided by total attempts. (https://www.americansocceranalysis.com/home/2018/10/11/whatdeterminesagoodpasser#:~:text=DPOE%20contextualizes%20Per100%20by%20indicating,or%20bad%20at%20completing%20passes.&text=Let's%20look%20at%20some%20examples,rates%20in%20for%20forward%20passes.)"
8. **Corners**: "Awarded to the attacking team if the defense knocks the ball out of bounds over their own end line. The kick is taken from the corner arc nearest where the ball went out of bounds. Opponents must be at least ten yards away from the ball when the kick is taken (modified for small-sided games). A goal can be scored directly from a corner kick.(WEBSITE)"
9. **Total Crosses:** "A pass in which the ball is kicked from one side of the field to the other side. (WEBSITE)"
10. **Offsides**: "Occurs when a player positions himself nearer to the opponent's goal line than both the ball and the second-to-last opponent except when the ball is in play from a goal kick, a corner kick, or a throw, or if the player is in his/her defending half of the field. No

fewer than two defenders (usually the goalkeeper and one other defender) must be nearer to the goal line than the attacker. The person advancing with the ball must be the first to cross the line of defense. A player in an offside position is only penalized if, at the moment the ball is played by a teammate, he is, in the opinion of the referee, involved in active play, interfering with play or any opponent, or gaining an advantage by being in that position. When a player who is in an offside position receives the ball from a teammate or is involved directly in the play, an offside is called and an indirect free kick is awarded to the defense. (WEBSITE)"

11. **Aerial Duels Won:** "Aerial duels occur when two players contest a ball in the air; this is a symmetrical event because neither player starts with possession. (WEBSITE)"
12. **Goalkeeper Saves**: "A save is awarded to a goalkeeper only if a shot otherwise would have gone into the goal. A goalkeeper can be credited with a save without catching the ball. If the goalkeeper blocks the ball or punches it wide or over the goal, that goalkeeper can be credited with a save, provided the ball would have otherwise gone into the goal. To receive a save, the play must be a shot on goal. A goalkeeper cannot receive credit for a save on a cross. (WEBSITE)"
13. **Clearances**: "Happens when a team kicks the ball out of its defensive zone, ending an offensive threat by the opposing team. (WEBSITE)"
14. **Fouls:** "A stop in play when the referee judges a violation against an opposing player. The team that suffers the foul is awarded a free kick unless the foul is committed by a defensive player inside his own penalty area, in which case the foul results in a penalty kick. (WEBSITE)"

15. **Yellow Cards:** “Also called a caution or booking. Shown to a player by the referee for dangerous or unsportsmanlike behavior. If a player is shown two yellow cards in one game, it results in a red card and that player is ejected from the game. (WEBSITE)”
16. **Red Cards:** “When a player receives this, he is immediately ejected from the game. The team may not replace this player and will play down a man for the remainder of the game. Results from serious misconduct, violent play, offensive language, or intentionally denying a goal. (WEBSITE)”
17. **OP Team:** This refers to the opposing team competing against the Orlando City team. For interpretation purposes this column was generalized to either US team or non-US team.

    US Team:

    a. Atlanta
    b. Austin
    c. Charlotte
    d. Chicago
    e. Cincinnati
    f. Columbus
    g. D.C
    h. Dallas
    i. Kanas City
    j. LAFCA
    k. Miami
    l. Minnesota
    m. Nashville

n. New England

o. New York

p. New York City

q. Philidelphia

r. San Jose

Non-US Team

a. Montreal

b. Toronto

18. **Season**:

Pertaining to the Orlando Florida weather and respective season.

a. Winter: December to February

b. Spring: March to May

c. Summer: June to August

d. Fall: September to November

19. Facundo Torres

- Y: Played in game
- N: Did not play in game

20. Duncan McGuire

- Y: Played in game
- N: Did not play in game

21. Ramiro Enrique

- Y: Played in game
- N: Did not play in game

22. Martín Ojeda
    - Y: Played in game
    - N: Did not play in game
23. Nicolás Lodeiro
    - Y: Played in game
    - N: Did not play in game
24. Iván Angulo
    - Y: Played in game
    - N: Did not play in game

Dependent Variables:

1. **Final Score:**
    a. Draw
    b. Loss
    c. Win