

GW Data Science Capstone Presentation

# NHL Data API

What does it take to be a  
championship team?

---

# Introduction

---

- ❖ Personal introduction
- ❖ The NHL publishes all of its robust data through an API on their website, the data includes play-by-play data dating back to 1917.
  - ❖ I intended to collect data back to the start of the Ovechkin era(2005-2006 season) in Washington, D.C.
  - ❖ The Washington Capitals have dominated the regular season for the majority of this time, yet fell short in the playoffs. The 2016 team, on paper, was the most talented team by far, yet they still failed to get past the second round of the playoffs. Then the 2017-2018 team filled with many new and young faces was written off, as a rebuilding year for the team, but they won it all. My goal is to see if the data can tell us what made them different.

---

# Challenges

---

- ❖ I had some initial challenges harvesting the data, the payload was heavily nested JSON.
  - ❖ I chose a dataset that while heavily numeric needed a lot of transformation and cleaning so that I could run models on it.
- ❖ I decided to do this project in Python, but I had to teach myself Python first....18 data camp courses later.
- ❖ My time was challenged this semester due to family emergencies .

---

# Problem Definition

---

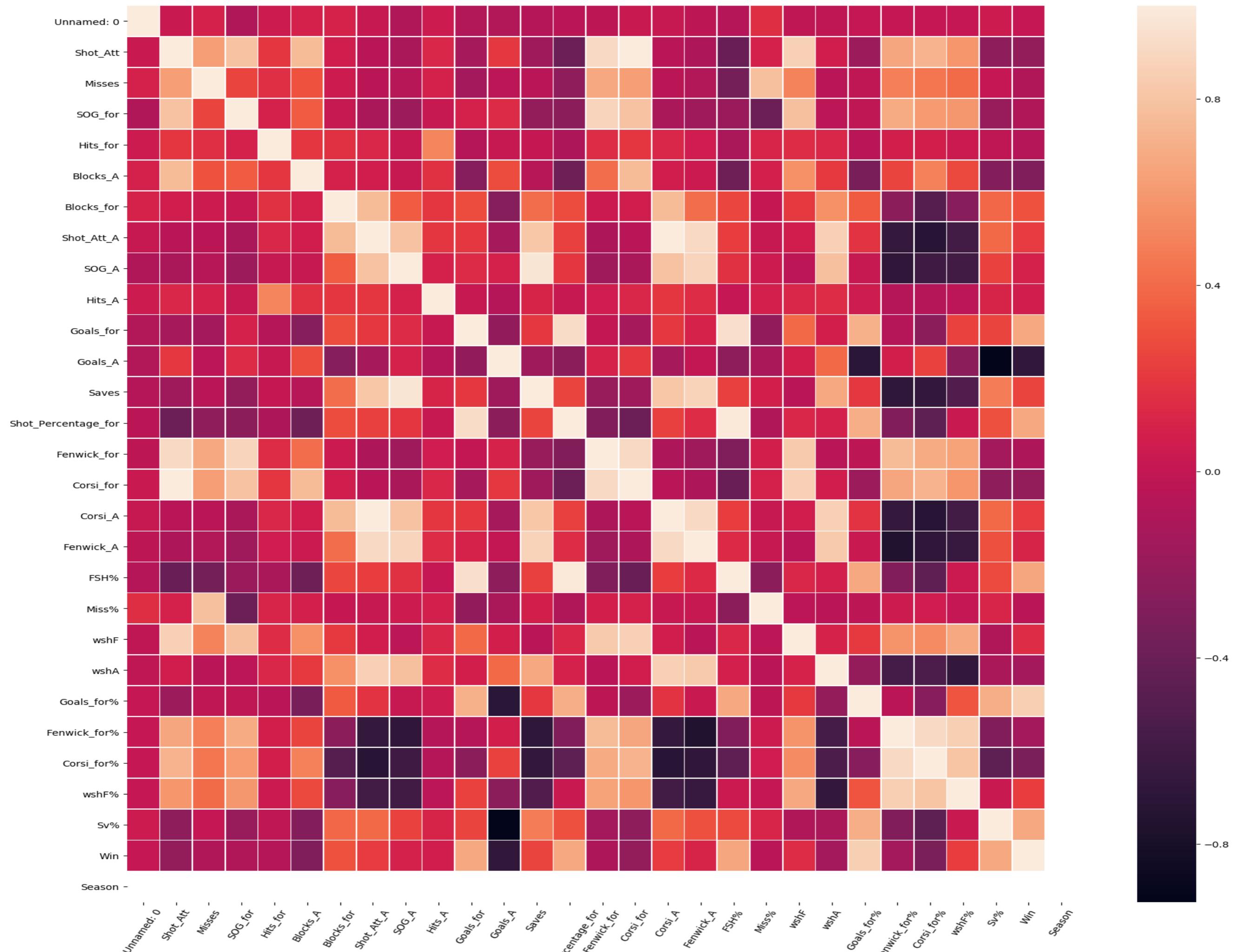
- ❖ What variables are most important for a team to win the Stanley Cup?
  - ❖ Common data collected that could contribute:
    - ❖ Offensive - Goals, points(I'll define this but it is different than goals), shooting percentage.
    - ❖ Defensive - Hits, Blocked Shots, Goalie percentage
    - ❖ Win percentage(both home/away)

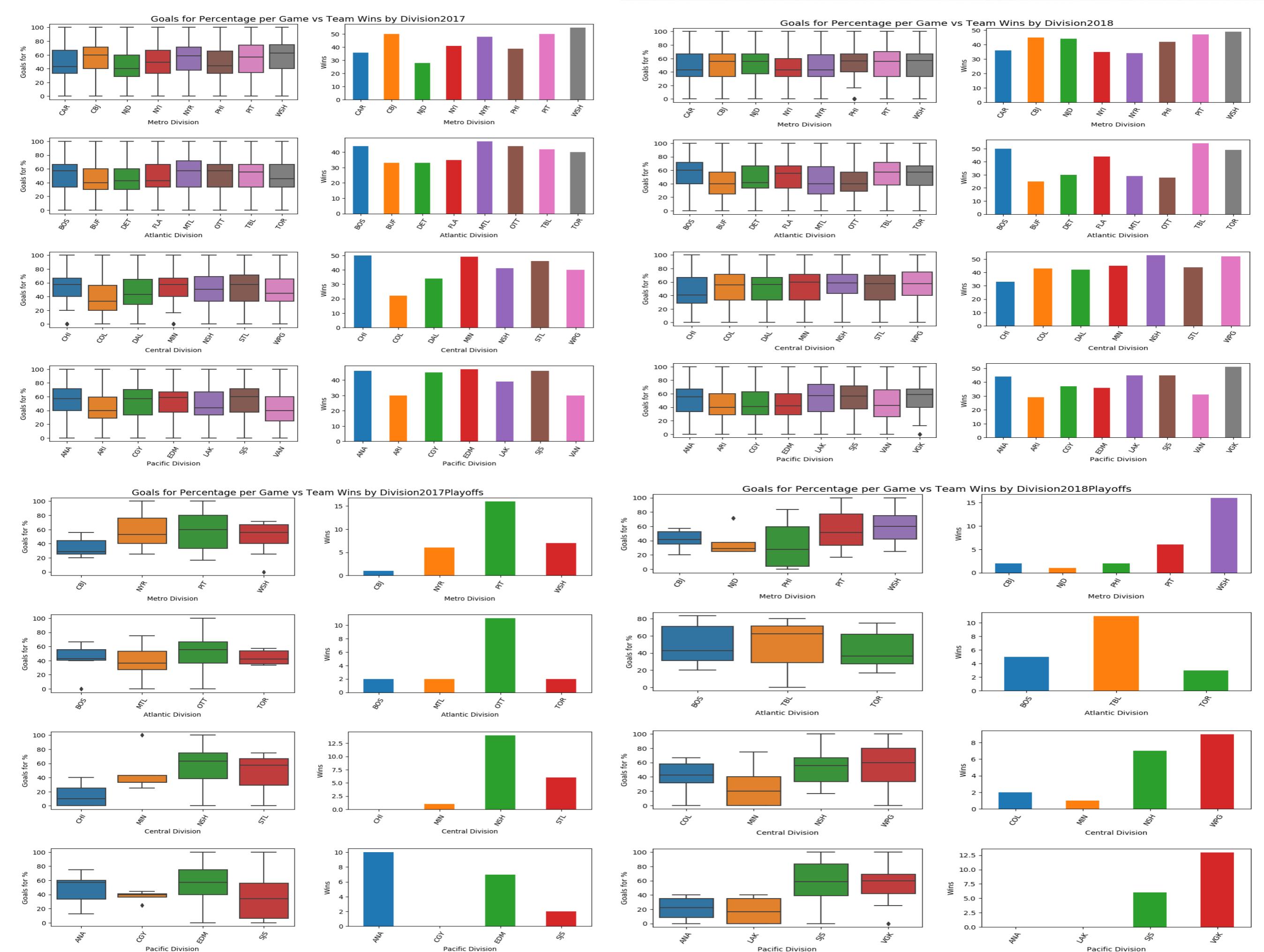
---

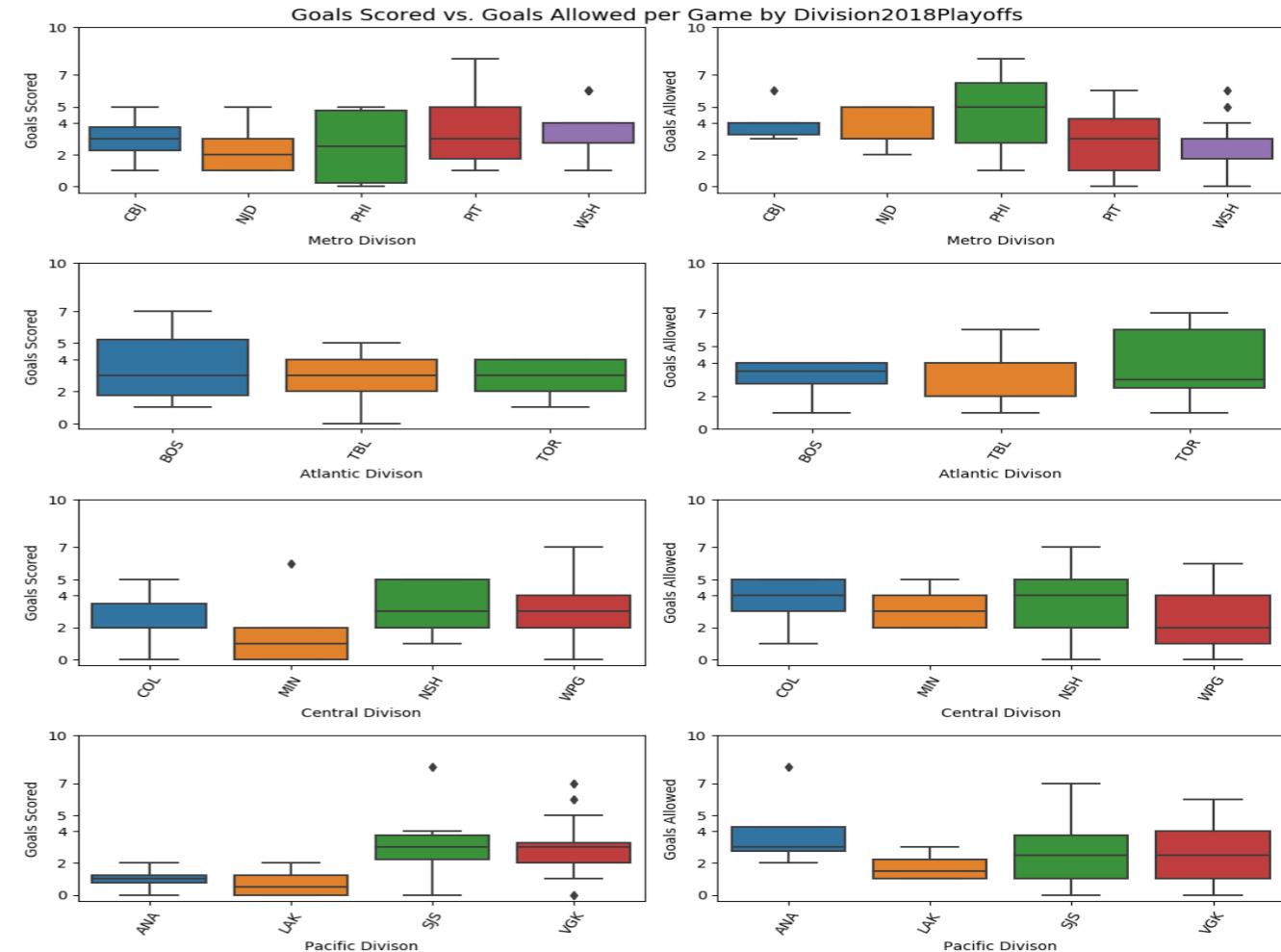
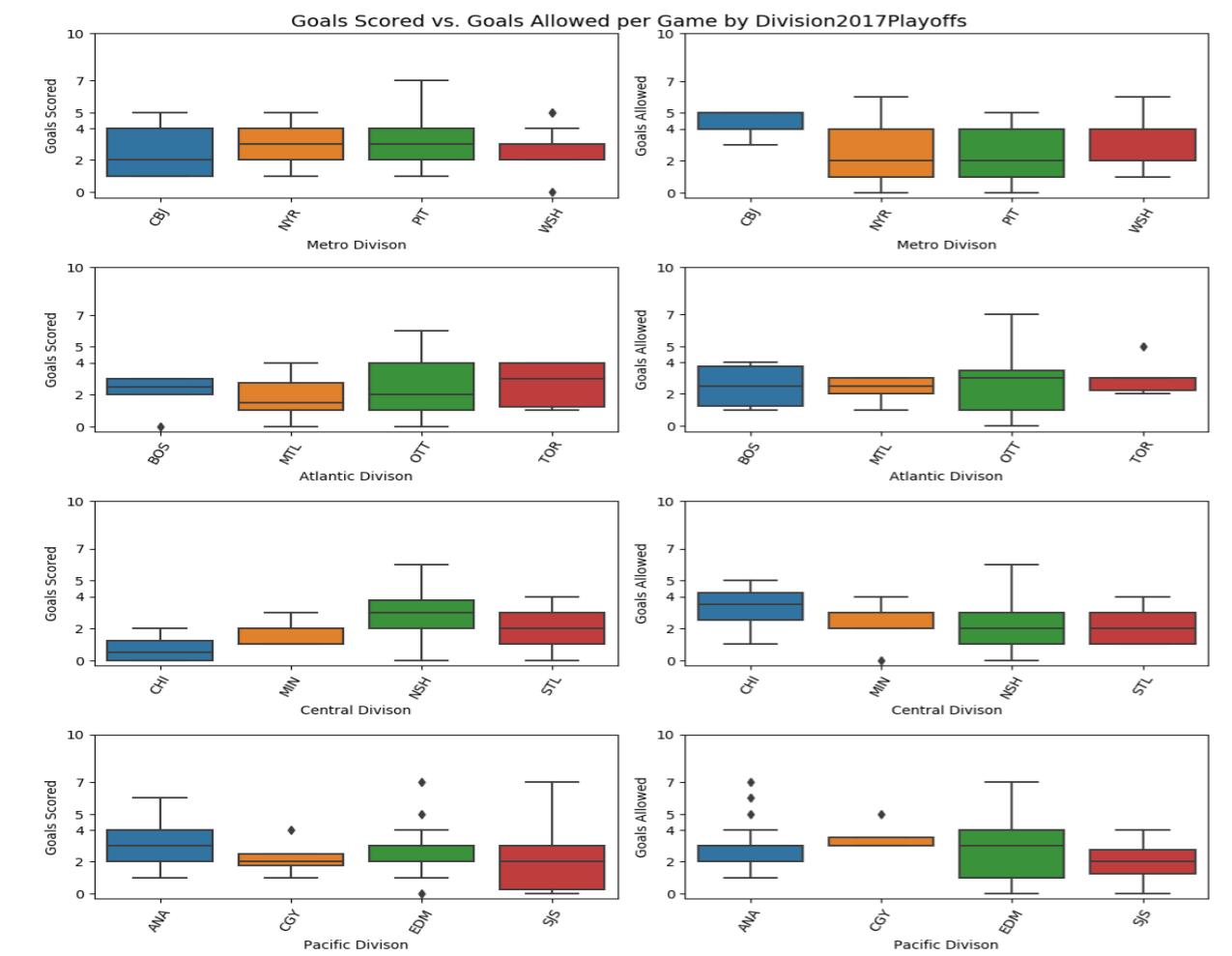
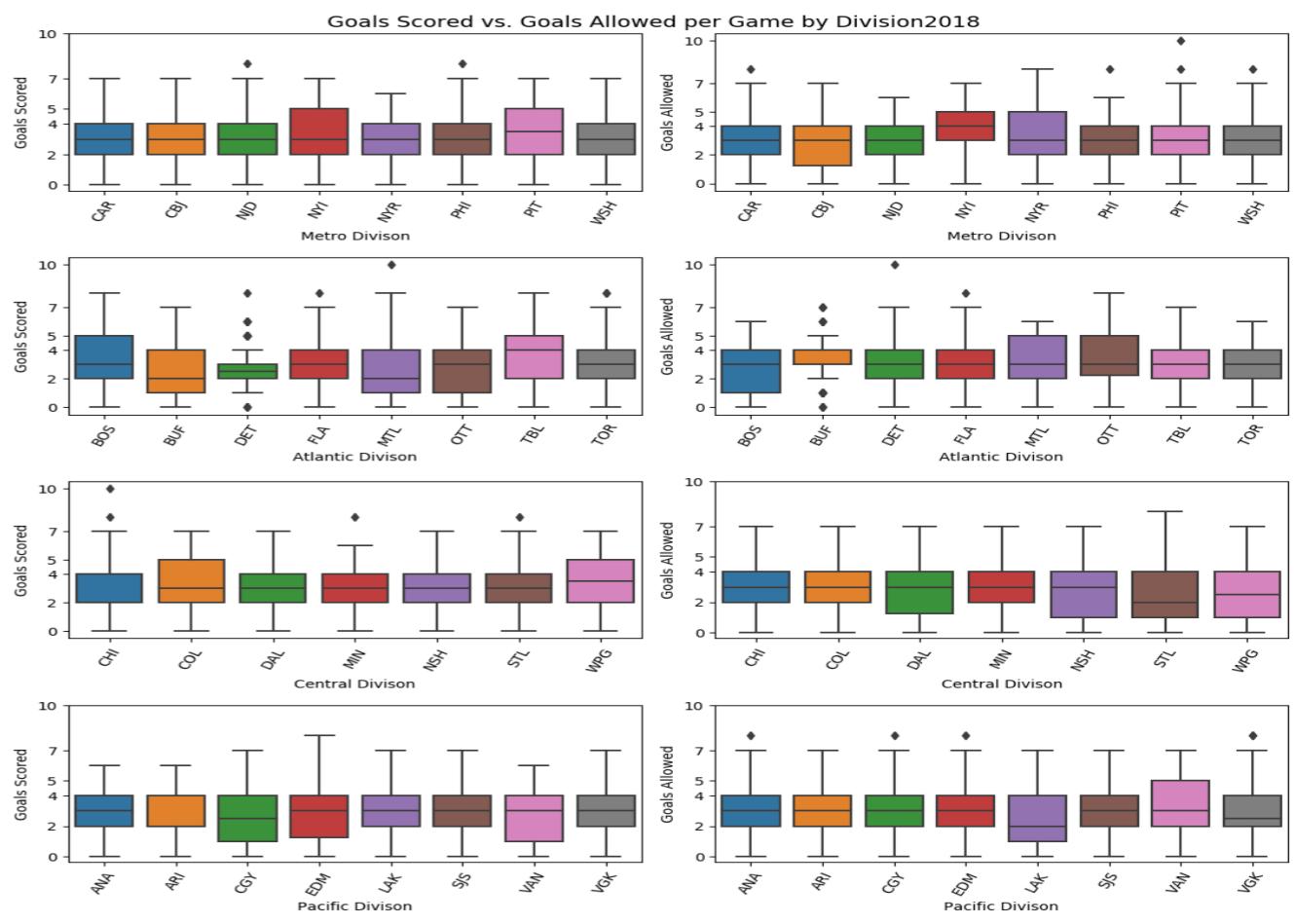
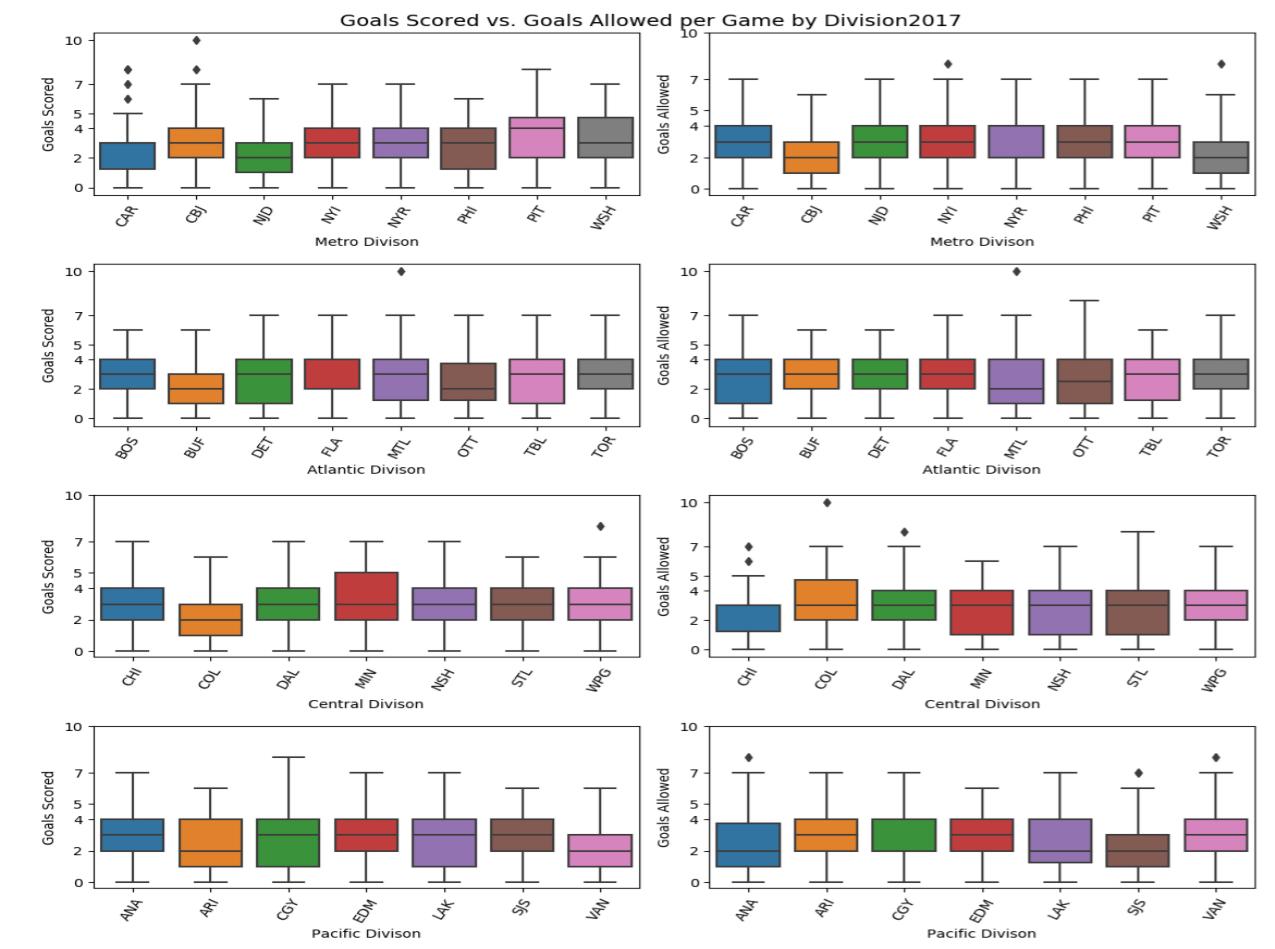
# Methodology

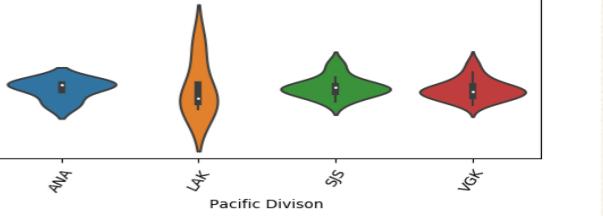
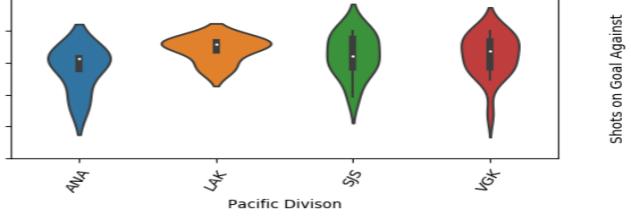
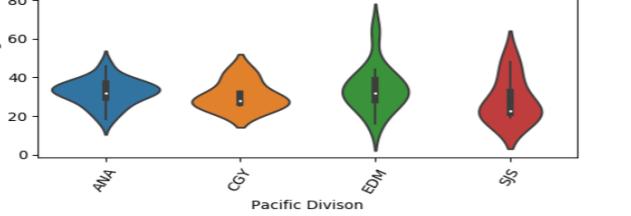
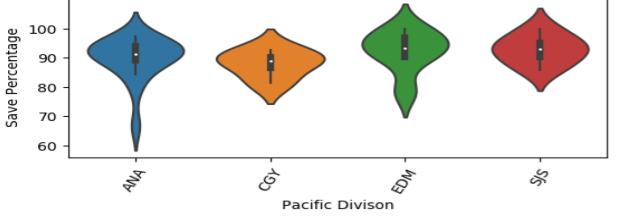
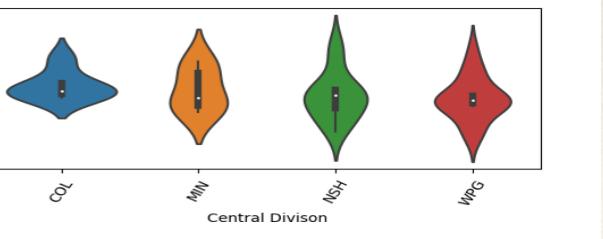
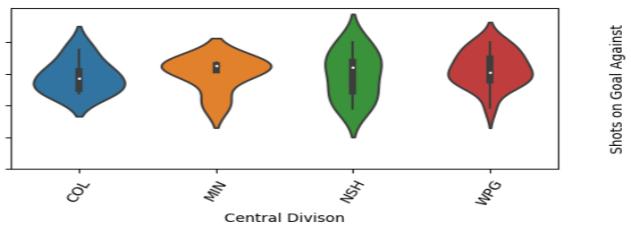
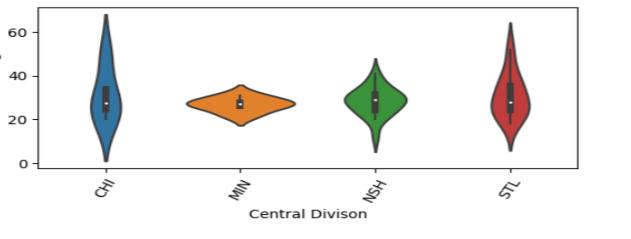
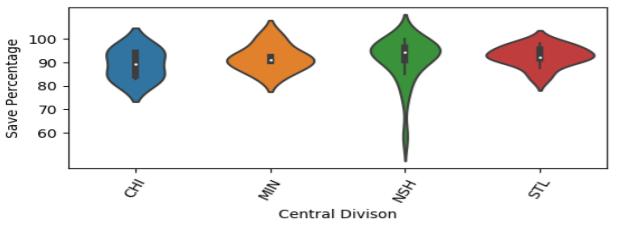
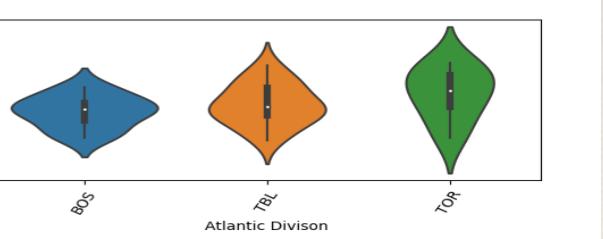
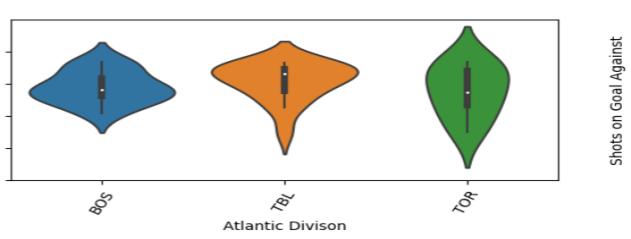
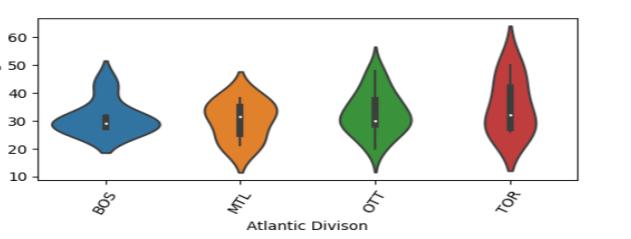
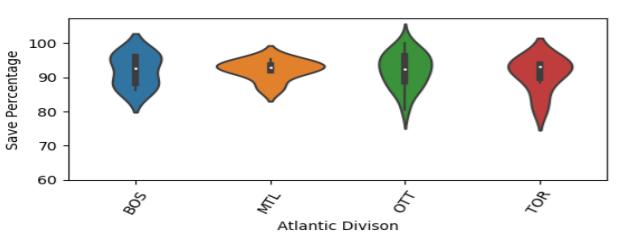
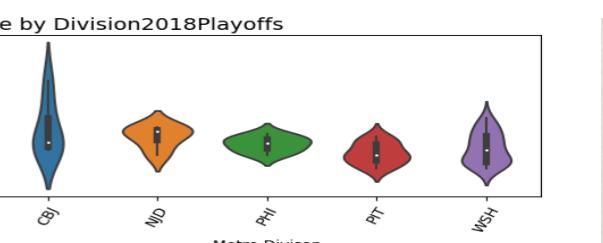
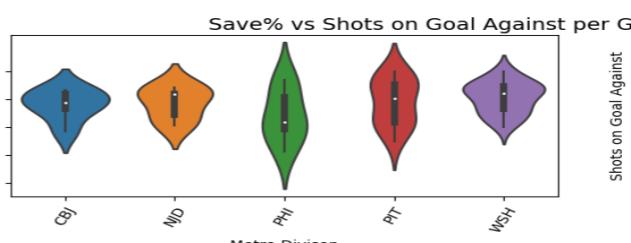
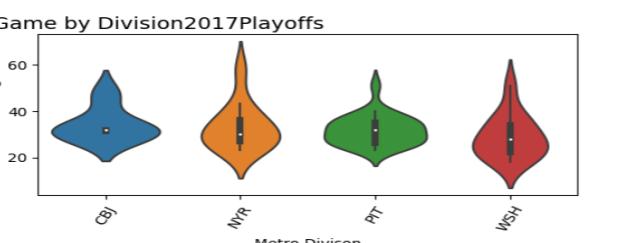
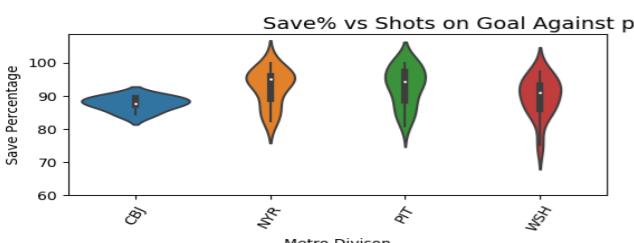
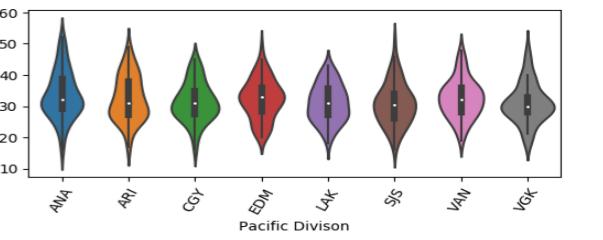
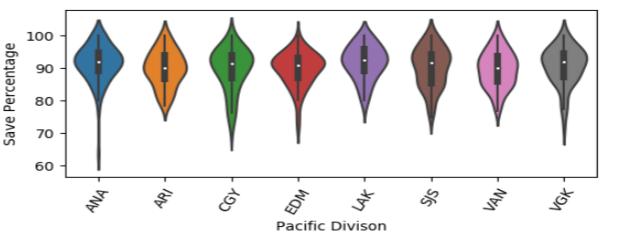
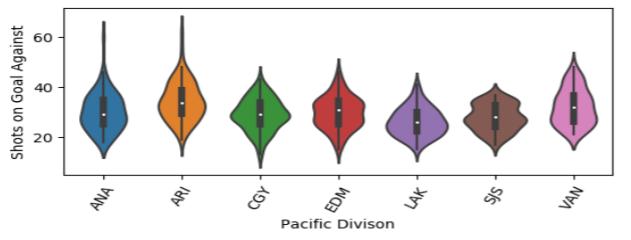
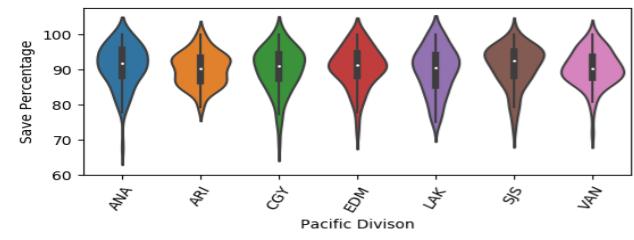
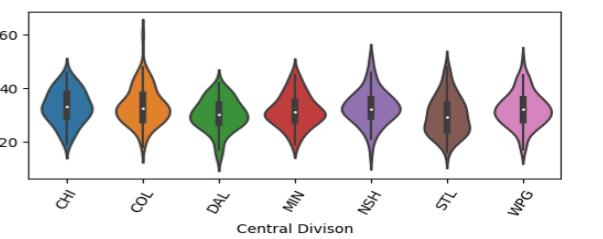
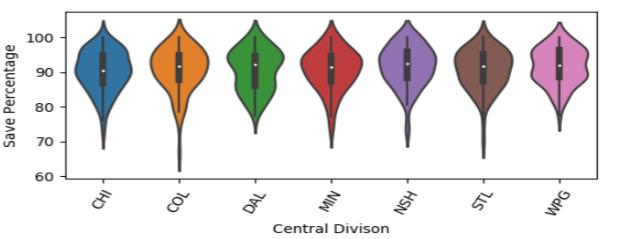
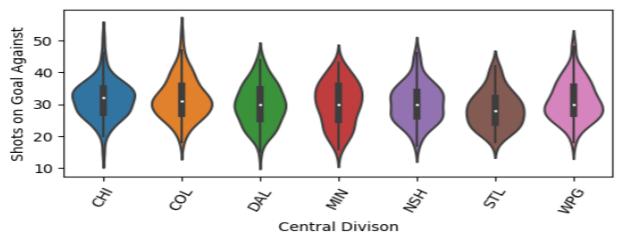
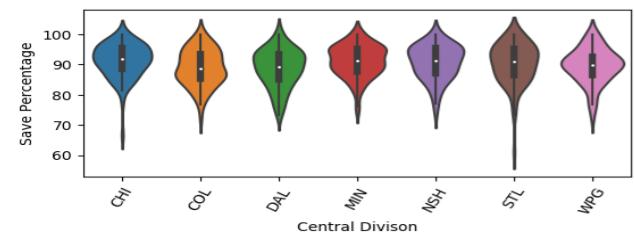
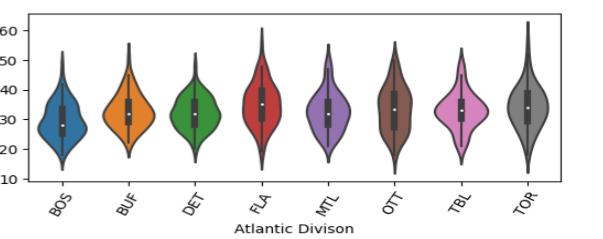
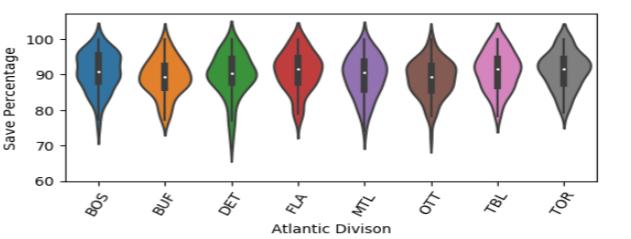
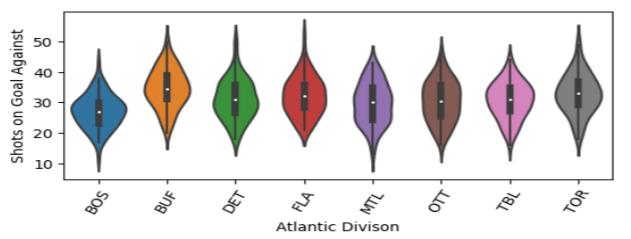
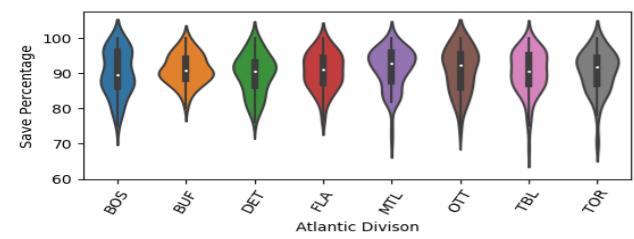
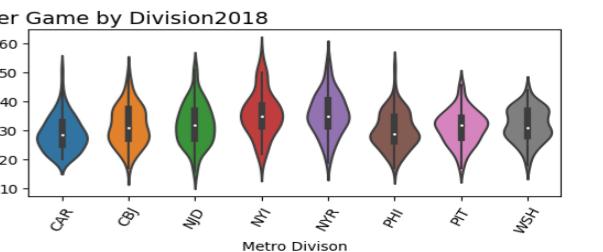
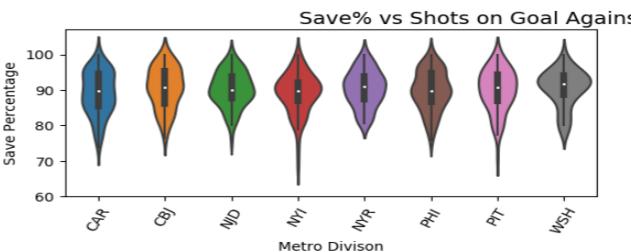
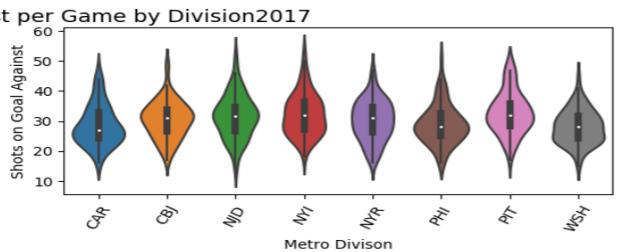
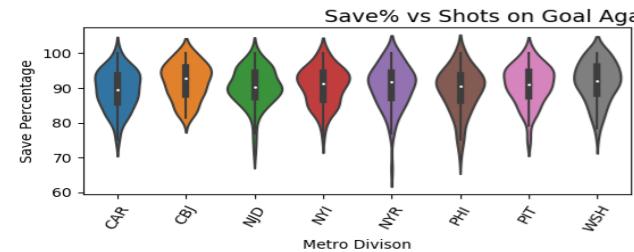
---

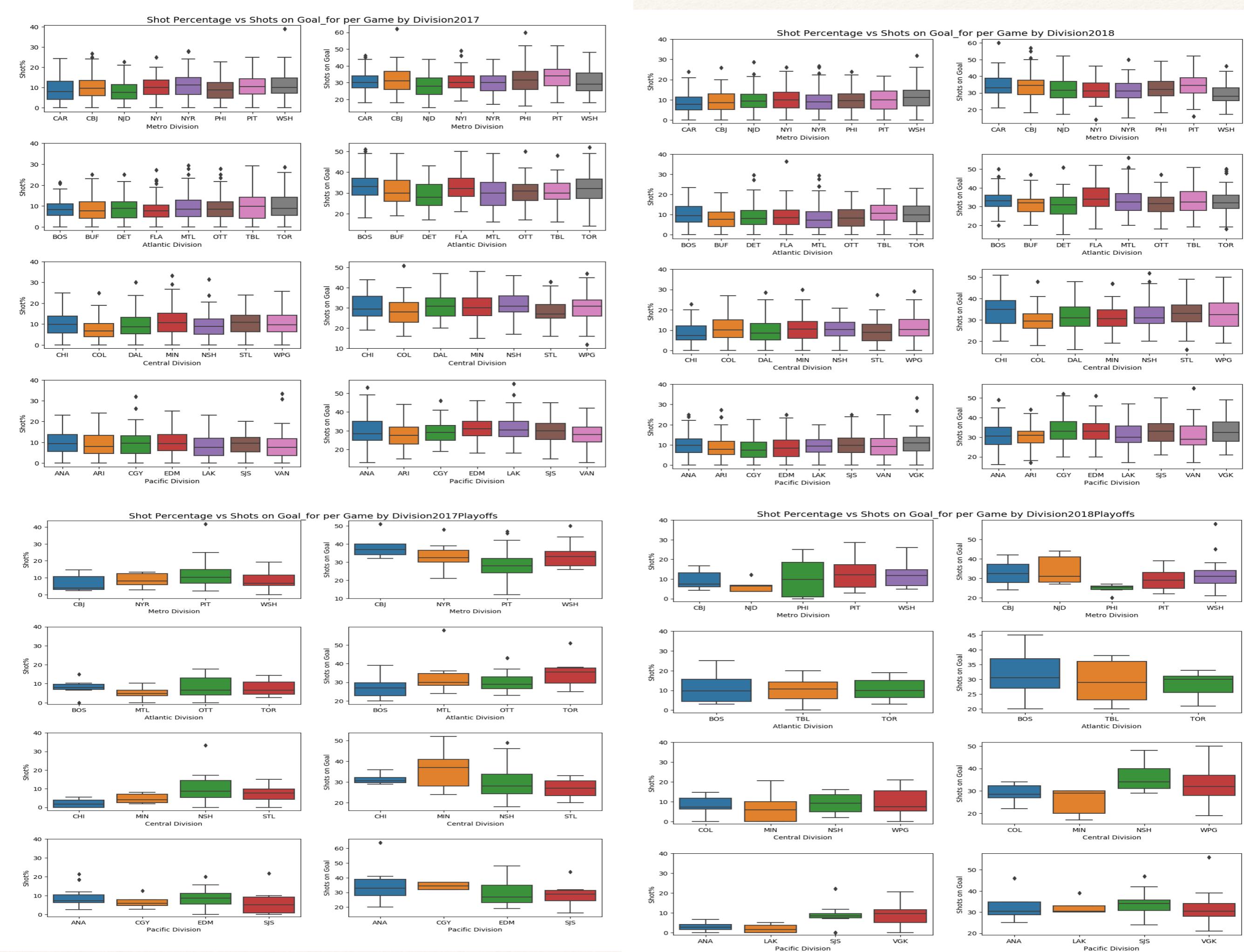
- ❖ Using the NHL API I will harvest play by play data, and perform exploratory data analysis to identify predictor variables within the data.
- ❖ I will use a Generalized Linear Model/Logistic Regression to see if I can build a predictor model based on the aggregated statistical values taken from the play by play data.
- ❖ Additionally I used a Naive Bayes with Gaussian distribution to compare models.
- ❖ Finally as close to the final presentation as possible I will harvest and clean the most up to date information for this season to run through both prediction models; this only worked for data that had happened though, I could not figure out how to forecast games that did not occur yet.

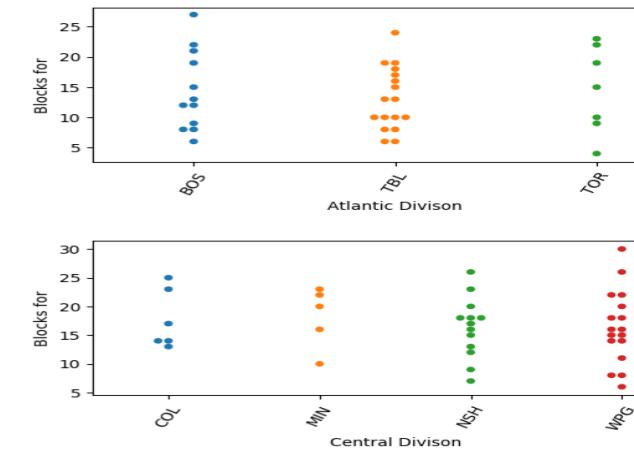
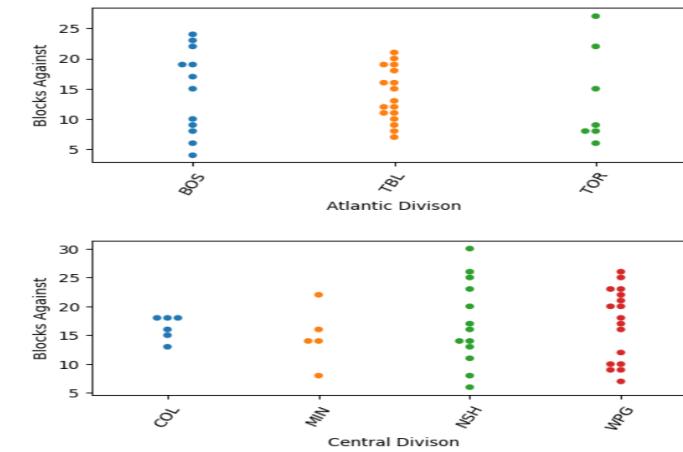
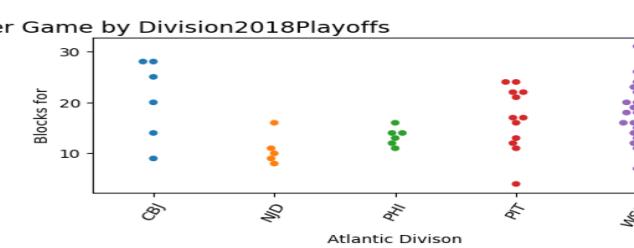
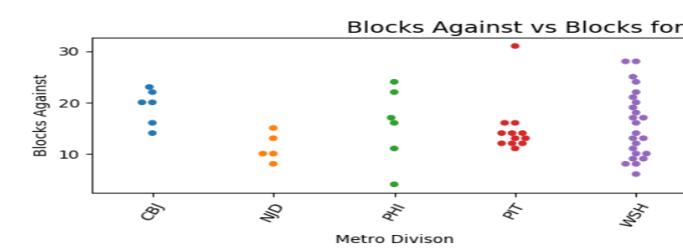
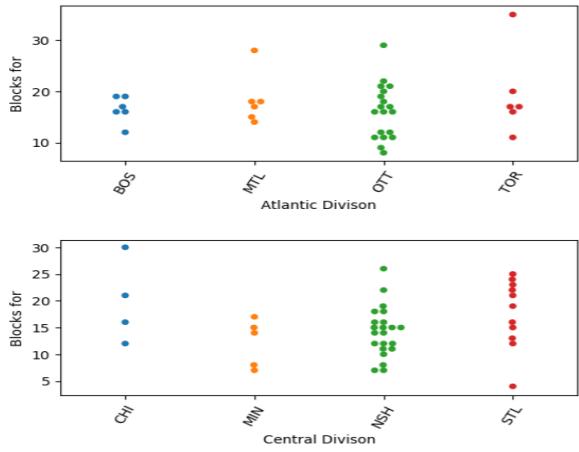
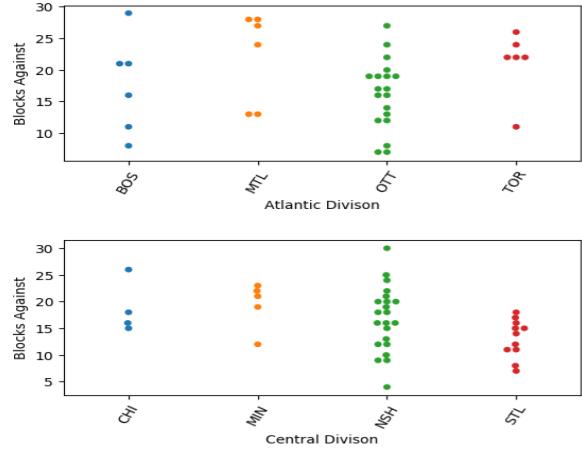
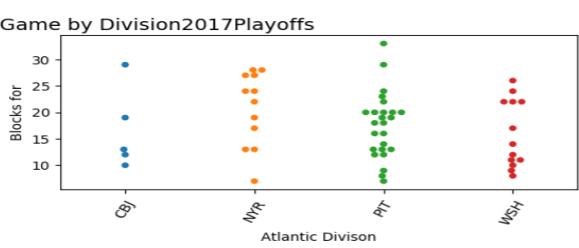
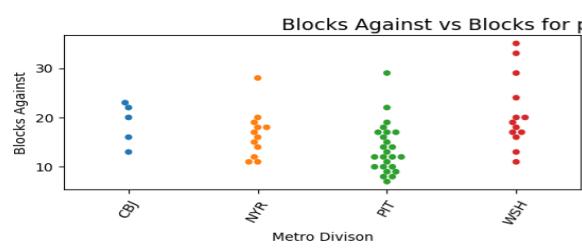
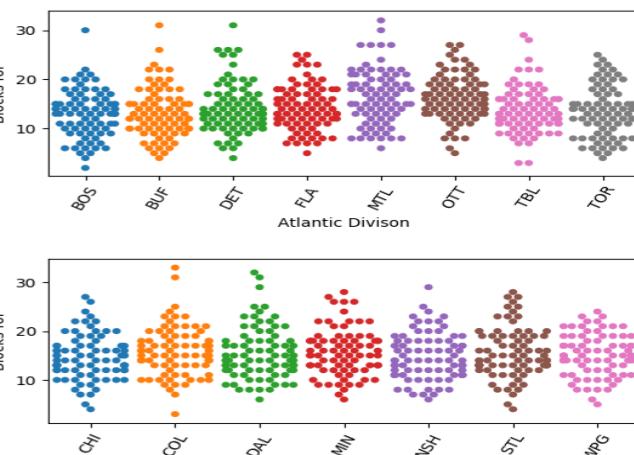
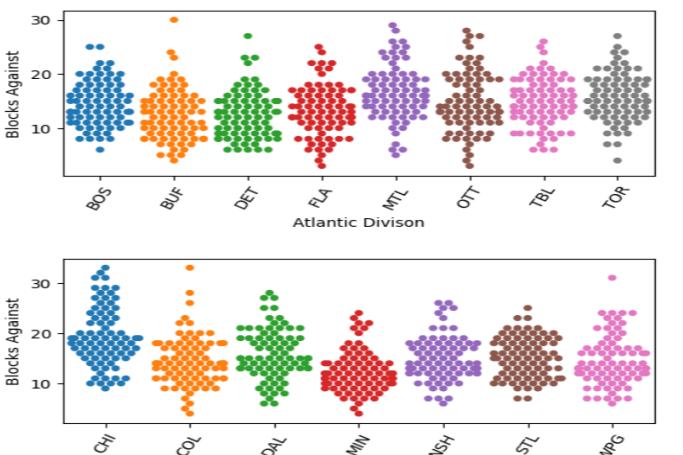
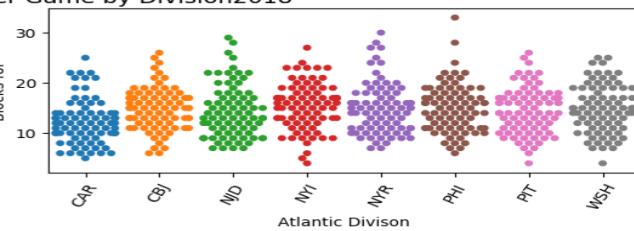
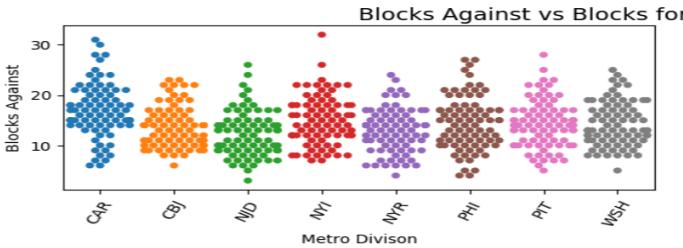
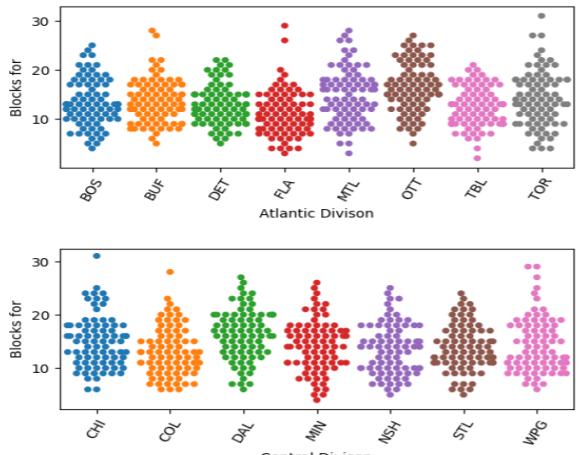
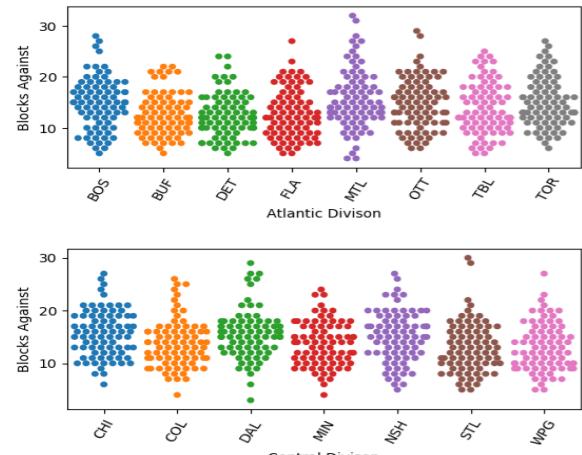
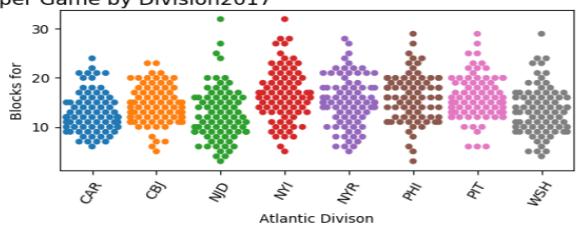
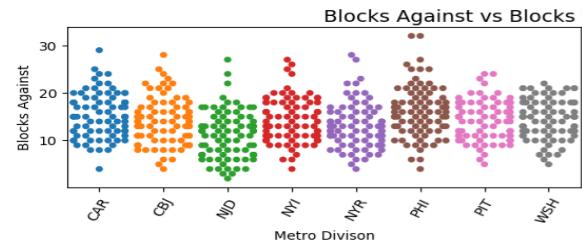




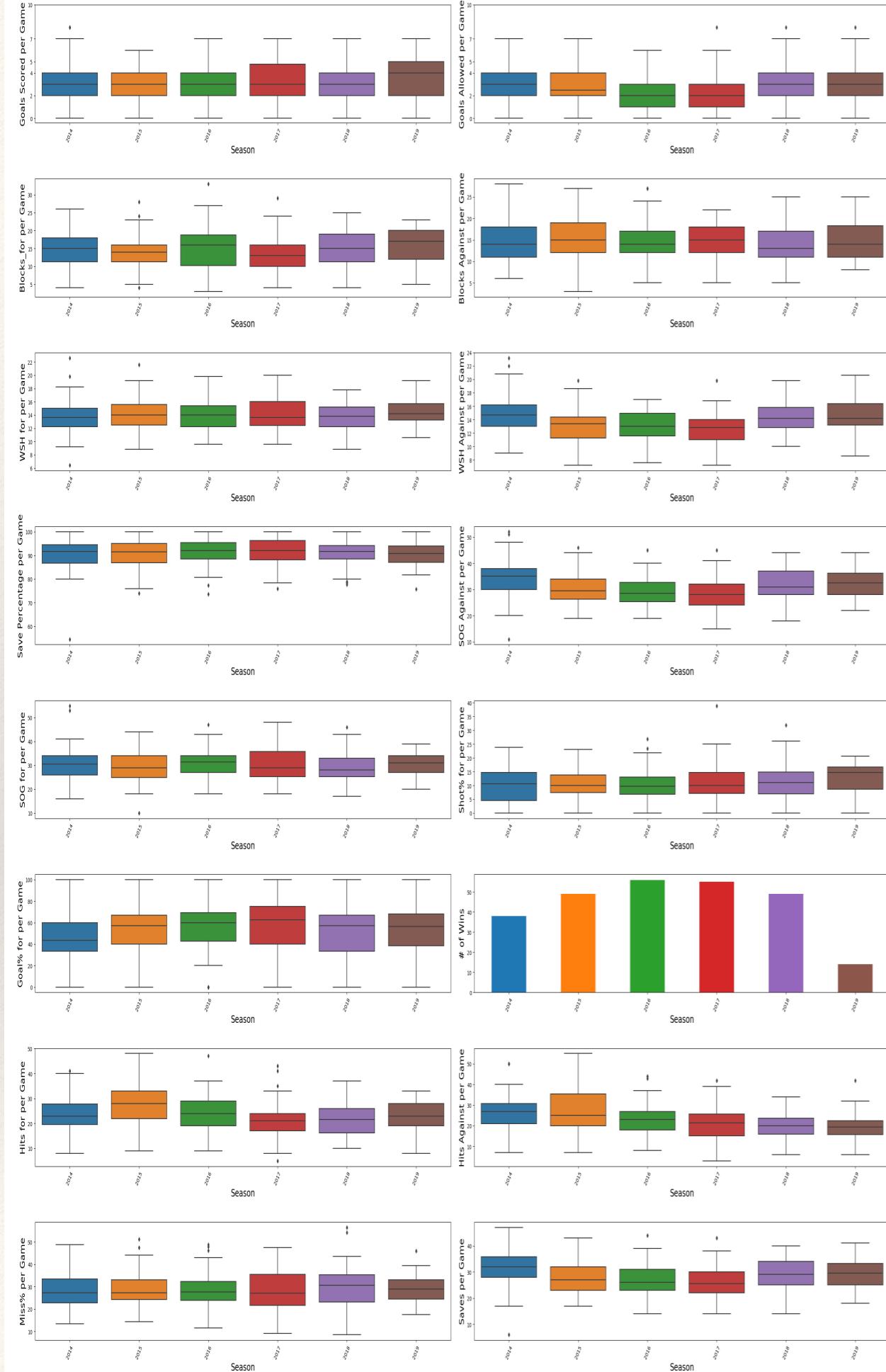




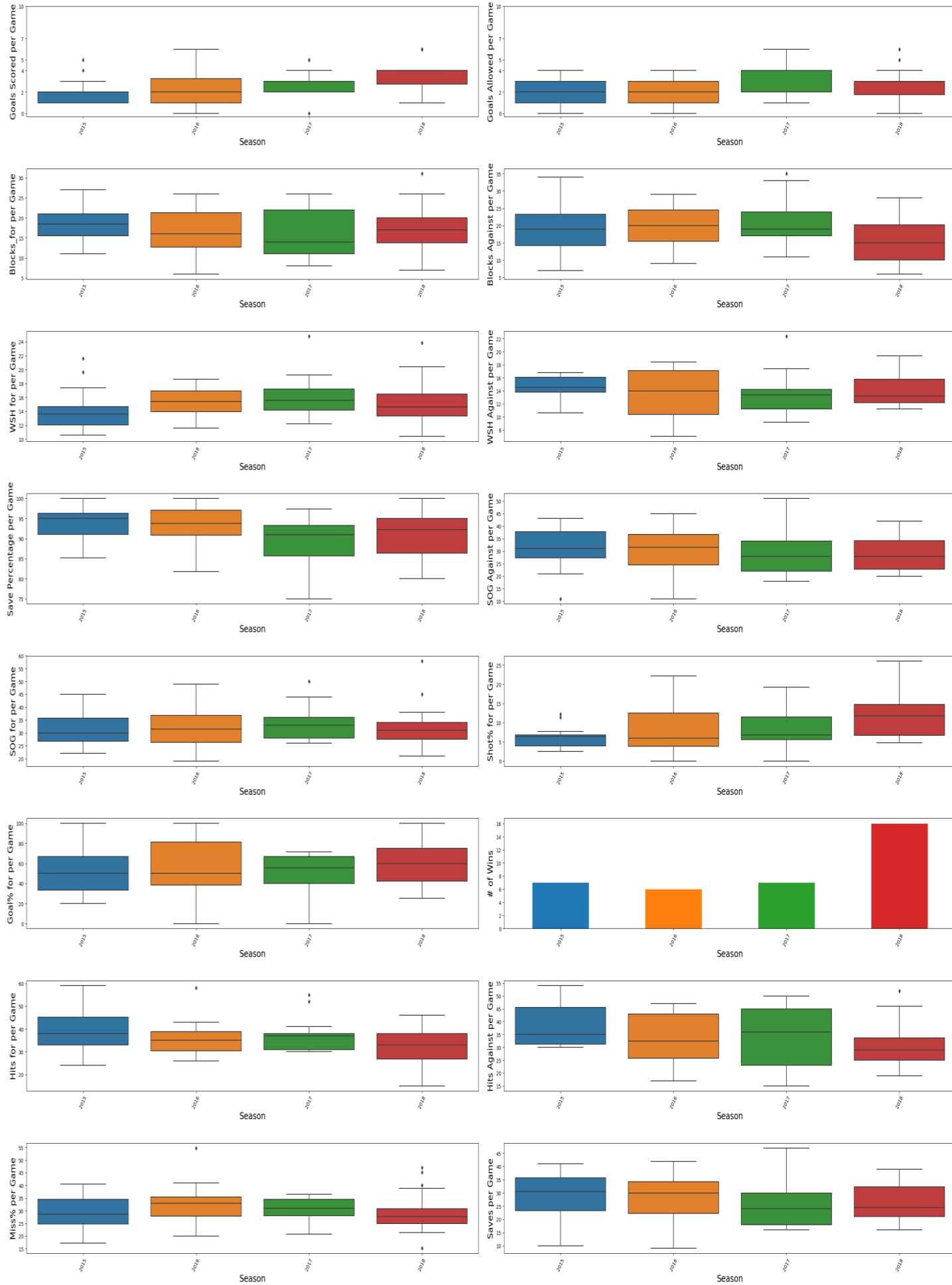




Stats by Season by Game for Washington

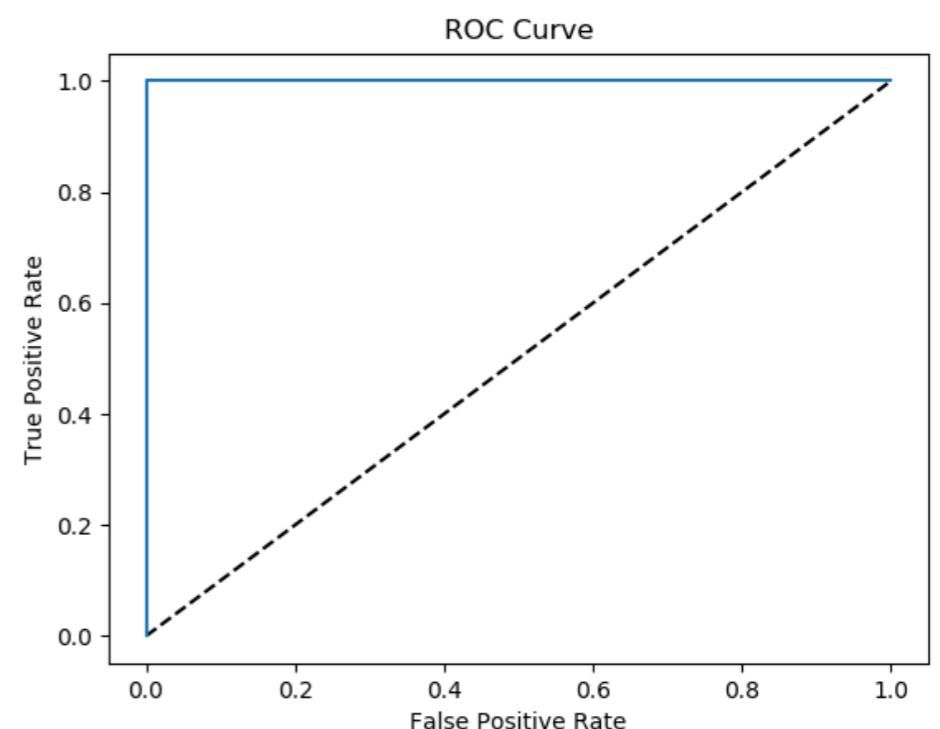
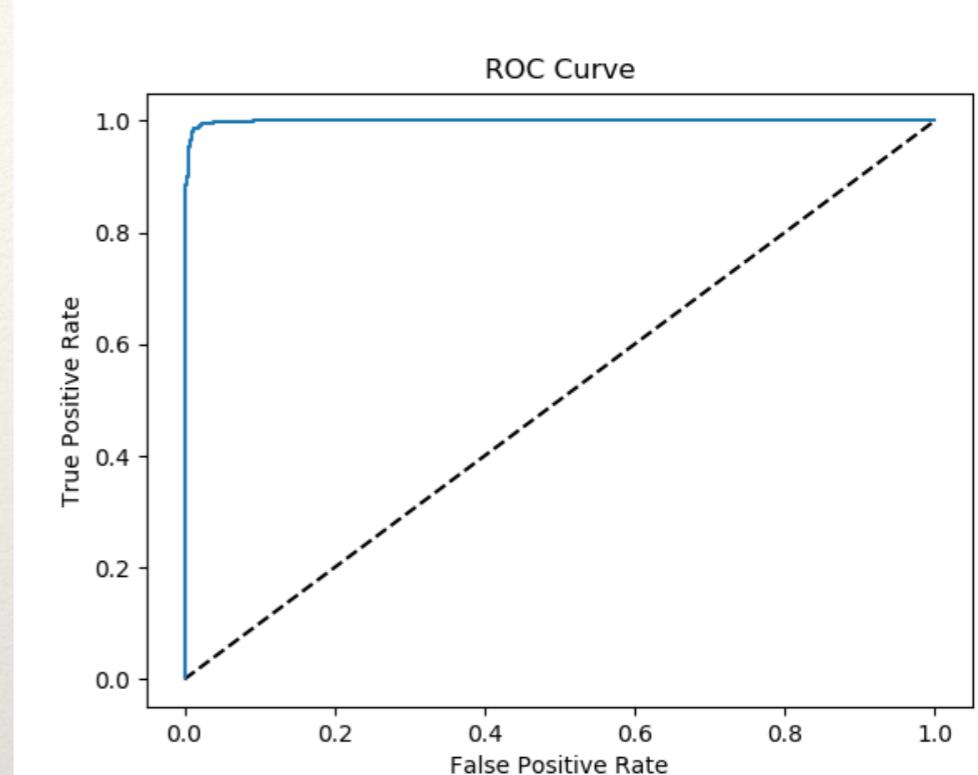


Playoff Stats by Season for Washington



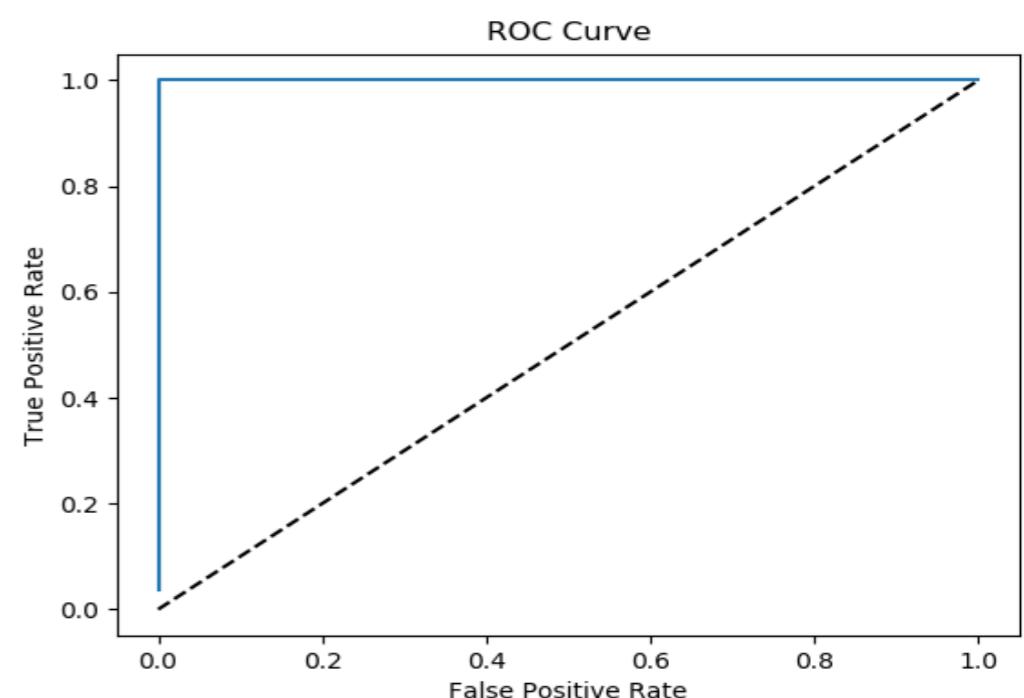
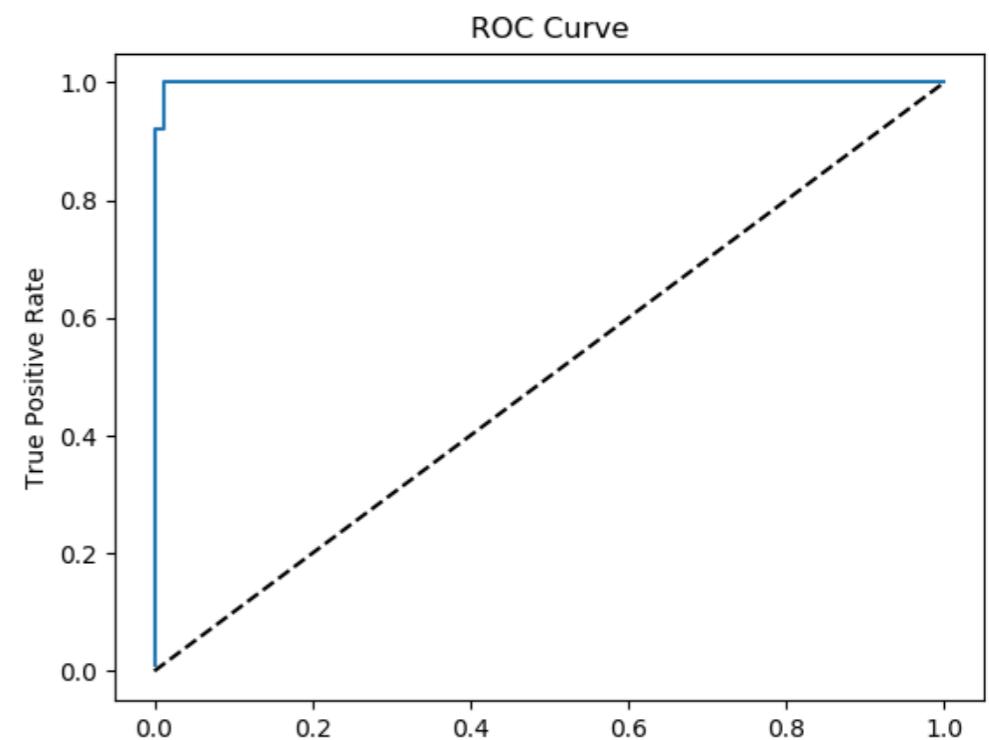
# The Models

- ❖ The Naive Bayes with a Gaussian distribution performed between 67 and 69% accuracy.
- ❖ The Logistic Regression performed between a 96 and 100% accuracy
- ❖ I think there were still too many features derived from Goals\_for. The models cannot forecast, I need more time to understand how to build that type of model.



# The Models II

- ❖ The playoff models were similar to the regular season
  - ❖ Bayes models performed between 57% and 61% on the final run.
  - ❖ LogReg still over performing between 98 and 100% on the 2018 season.





Home Game 1 of the Stanley Cup Finals

“We're not gonna be F\*&%ing suck this year!  
We're Stanley Cup Champions!”

*–Alexander Ovechkin at the Capitals Championship Parade*