GAME CHANGER

Sports Wager Analytics

- *Capitalizing on online sports betting market inefficiencies*

Link to presentation recording

Prepared for
2023 Spring | MSDS 498 | Team 55

Prepared by
Timothy Steed
Timothy Chiu
Christopher Kradjian
Garrett Lynch

Meet The Team

Timothy Steed

Tim S. is a researcher and aspiring data scientist at the Financial Industry Regulatory Authority (FINRA), with experience in derivative modelling and trading roles at various banks and hedge funds in Chicago.

Timothy Chiu

Tim C. is in a hybridized design development and project planning/management role within the operations organization of health diagnostics division in Abbott Laboratories' US headquarter.

Christopher Kradjian

Chris received his first Masters degree in Information Management Systems from Harvard Extension School and leads the Revenue Management team at Universal Studios Hollywood.

Garrett Lynch

An aspiring data scientist, pivoting after he completed his MBA at UC Irvine. He currently works on satellite proposal efforts in the space sector.

TS	Primary Role: Data Researcher	Secondary Role: Data Modeler
TC	Primary Role: Data Modeler	Secondary Role: Data Analyst
CK	Primary Role: Solution Architect	Secondary Role: Storyteller
GL	Primary Role: Project Manager	Secondary Role: Technical Writer

Overview

Business Model Overview

The team seeks to achieve the following:

- Identify team and game characteristics (i.e. features) that rank the most impact to scoring metrics.
- Build a classification model which would analyze NBA games and classify specific contests as recommended bets based on user-defined confidence thresholds.
- Devise an optimal betting strategy for the most popular wager types in NBA games.

Problem Statement

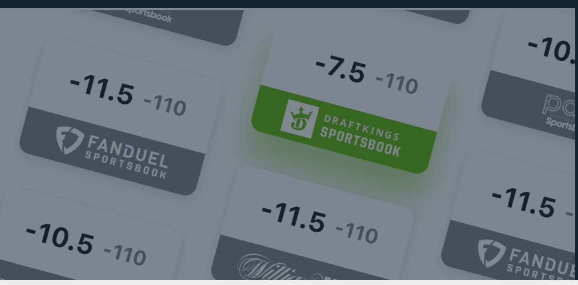
The NBA is a fast-paced game where game flow and line changes occur within a few minutes. For the most typical issued odds of -110, bettors will need to win 52.38% of times or more to break even or profit.

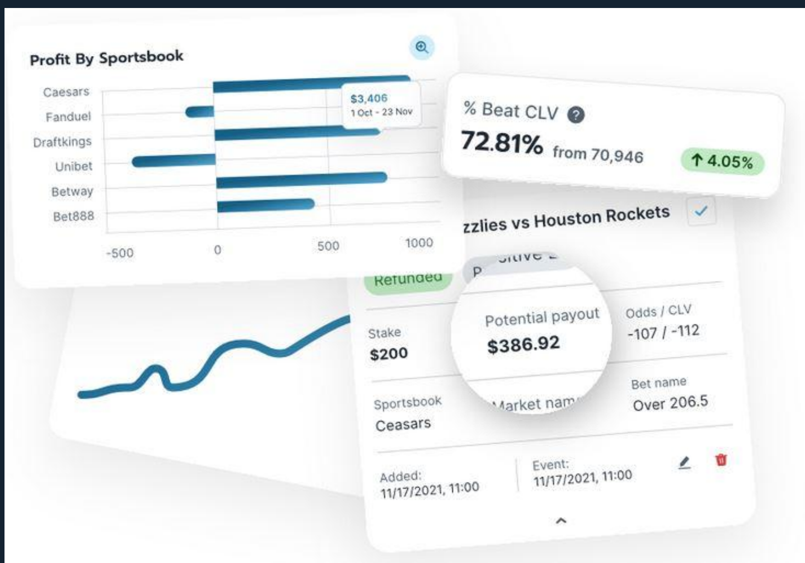
On one hand we plan to model these moving lines, to see if at any point before or during the game, they shift enough to allow for confidently predicting a better than ~52.4% success rate.

On the other hand we plan to build models which would take in upcoming NBA matches and classify specific contests as "profitable bets" based on user-set confidence thresholds.

Opportunity

We seek to design and develop a new method of improving the accuracy (and returns) from predicting the outcomes from sporting events. We will focus this initial phase around the (NBA), and focus specifically on the betting lines - which team will win, or betting point spread - the winning or losing team will cover the published margin of victory or defeat.

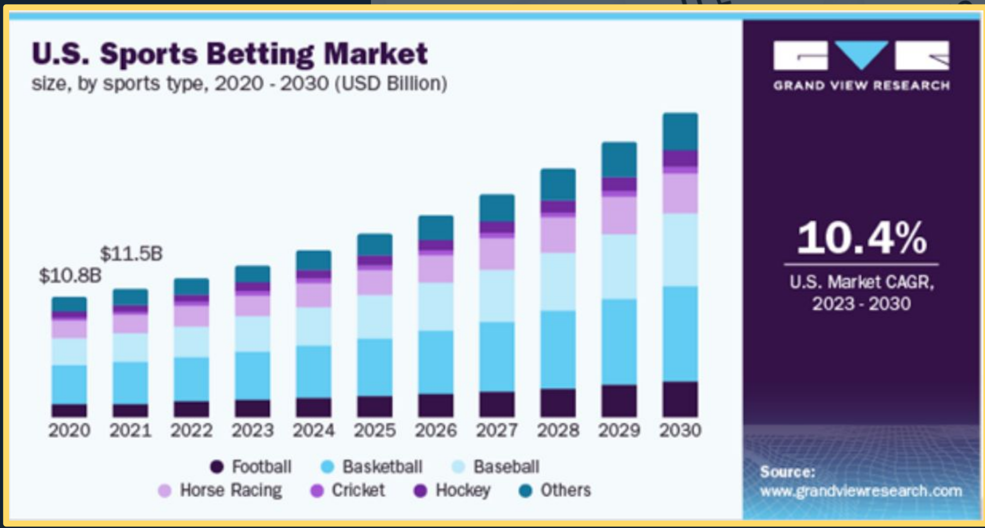




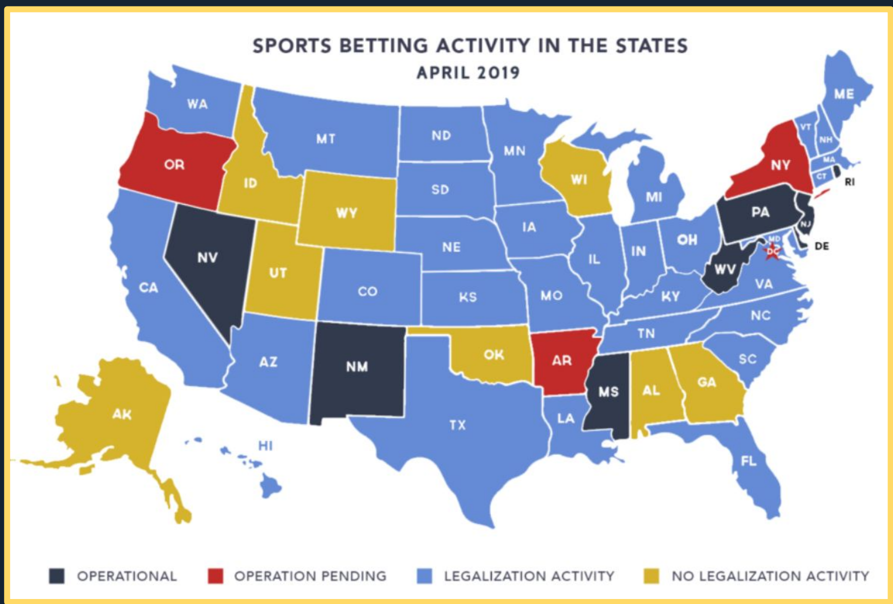
Market Analysis

U.S. sports betting market was already worth USD $10.8 billion in 2020. Basketball has been the dominant sport the American public wagers online.

The world of basketball sports betting offers a wide variety of wager types, among which the most popular are “Moneyline”, “Point Spread”, and “Point Total”.



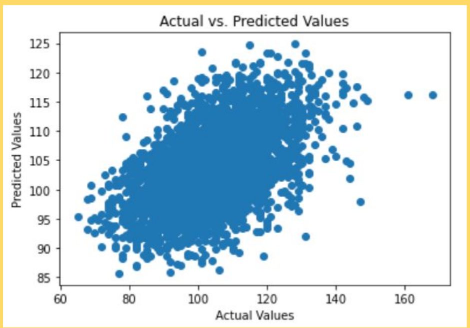
Taxes are based on federal government excise tax of 25% and Nevada state income tax of 6.75%. We decided upon Nevada because it had one of the lowest taxes for this industry.



Deliverables

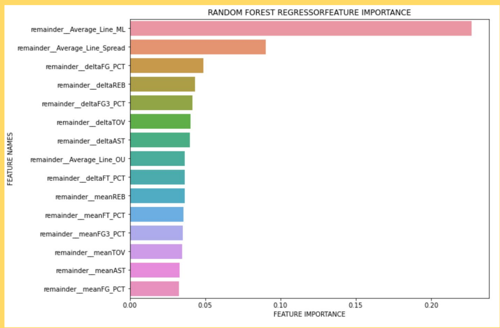
Approach 1

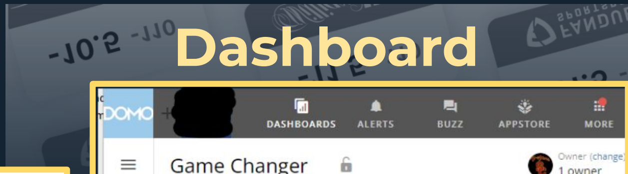
Regression Models



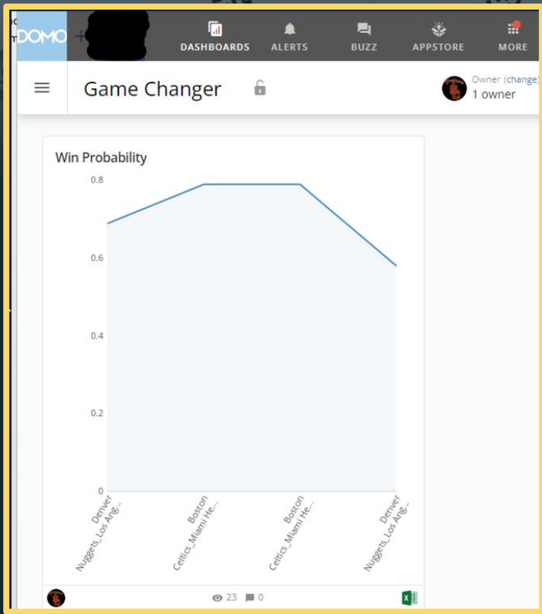
Approach 2

Classification and RF feature modeling

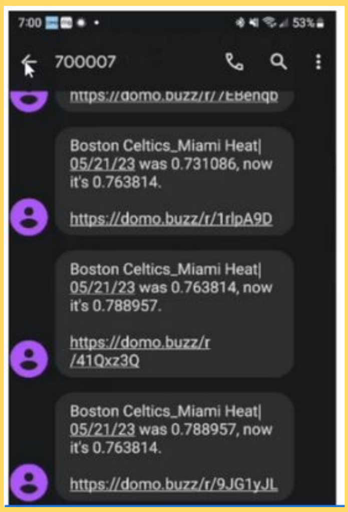




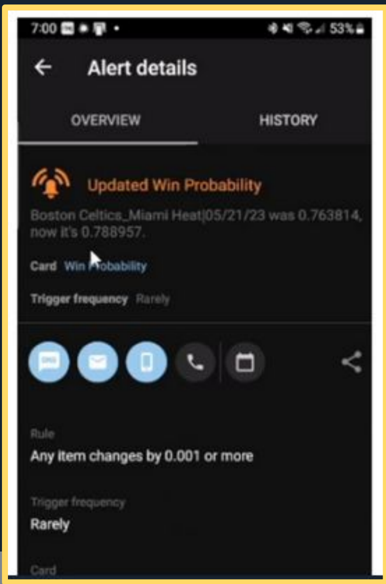
Dashboard



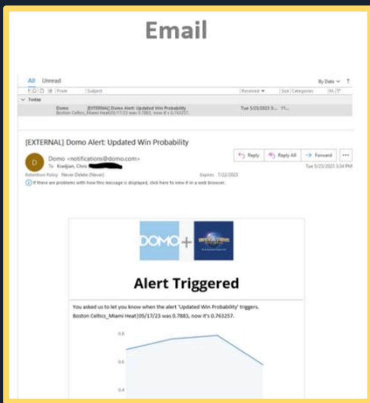
SMS Text



Push Notifications



Email



Data Overview

Source: Kaggle

NBA Odds and Scores

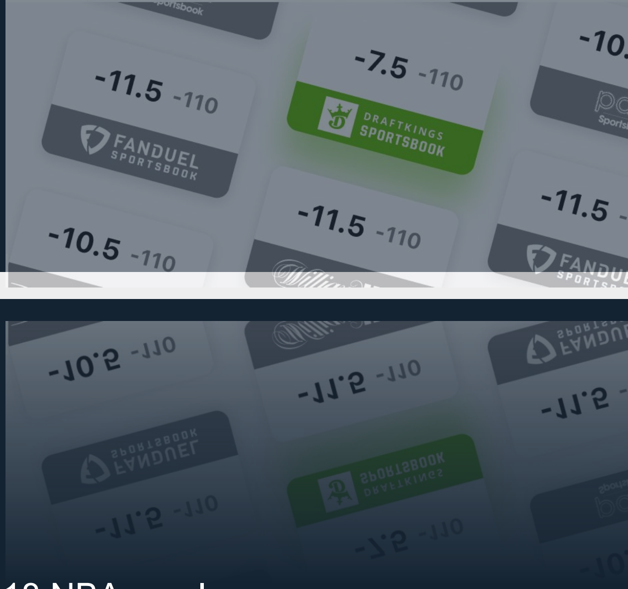
- 3 sets of text files (7 each) compiled from web scraping
  1. Box score data: 2012-2019 NBA regular seasons among 30 teams
  2. Lines and Odds data: Regular season major game bets (moneyline, spreads, and over/under) from 5 different sports books (Pinnacle, Bovada, Betonline, Heritage, and 5dimes)
  3. Lines and Odds data: Playoffs major game bets (moneyline, spreads, and over/under) from 5 different sports books (Pinnacle, Bovada, Betonline, Heritage, and 5dimes)

- Original box score dataset contains features in raw scores and box score statistics category. *Continuous numeric, date/time, and categorical variables are present..*

- Both Lines and Odds datasets contain features in lines and odds from individual sportsbooks and average values. *Continuous numeric, date/time, and categorical variables are present.*

	Dataset 1	**Dataset 2**	**Dataset 3**
Description	NBA regular season box score	NBA regular season lines and odds	NBA playoffs lines and odds
Type	Raw text	Raw text	Raw text
# of records	17226	17208	1182
# of fields	28	58	58

Data Overview



Data munging

- Discard Playoffs lines & odds data - focus on 2012-2018 NBA regular seasons
- Create following labels from existing labels to provide more clarity to box score data

Labels	Season	Team_Wins	Team_Losses	Location
Description	NBA season year	Number of wins as of record entry in the season	Number of losses as of record entry in the season	Home or visiting

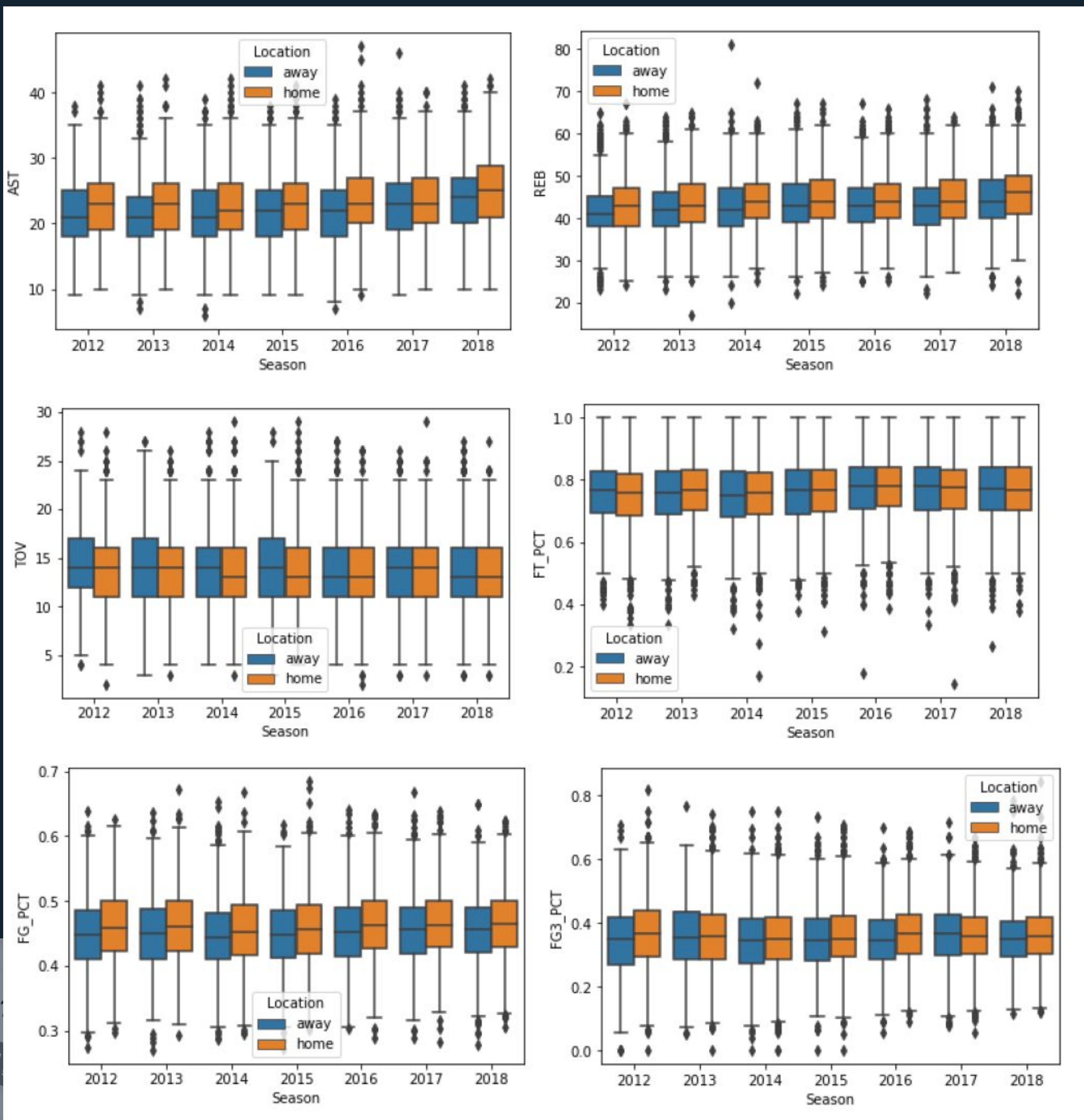
- Drop 2 null entries from box score data (Canceled game on April 16, 2013 between Boston Celtics and Indiana Pacers due to Boston Marathon bombings)
- Engineered per-game features (with a 1 game offset) for approach 2: cumulative averages of assists, rebounds, turnovers, field goal %, free throw %, 3 point %. Further engineered matchup deltas of per-game features after dropping first game of each team in every season (30 teams x 7 seasons = 210 records)
- Inner join regular season box score and line&odds data on Date, Game ID and Team ID, resulting in 16968 total records
- Drop 6 null entries from merged dataset, resulting in 16962 records before model training

- Identify target variables:
  - Score (approach 1)
  - Spread (approach 2)
  - Cover spread (approach 2)
  - Wins/Losses (approach 2)
  - Over/Under (approach 2)

Data Overview

EDA: Home advantage

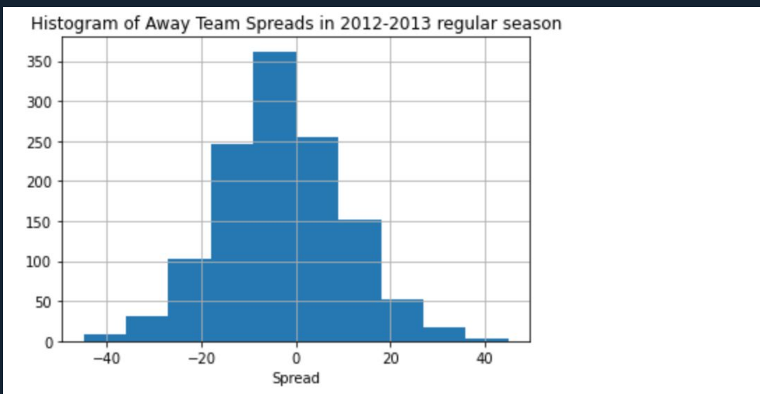
Box score statistics



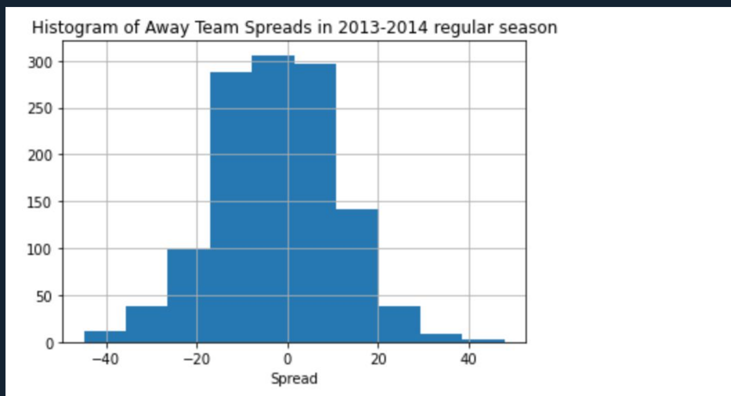
Data Overview

EDA: Home advantage

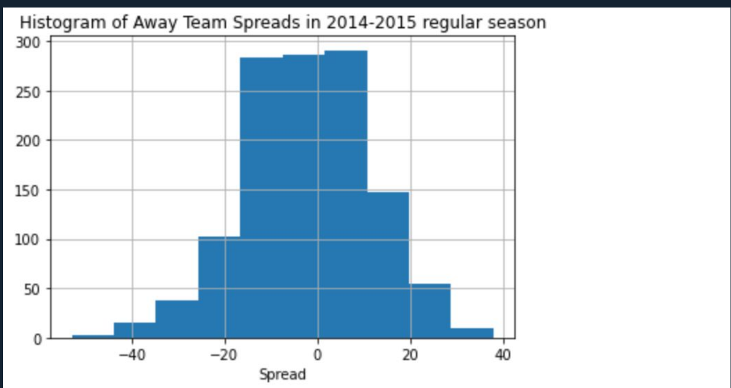
Home team point margin (spread)



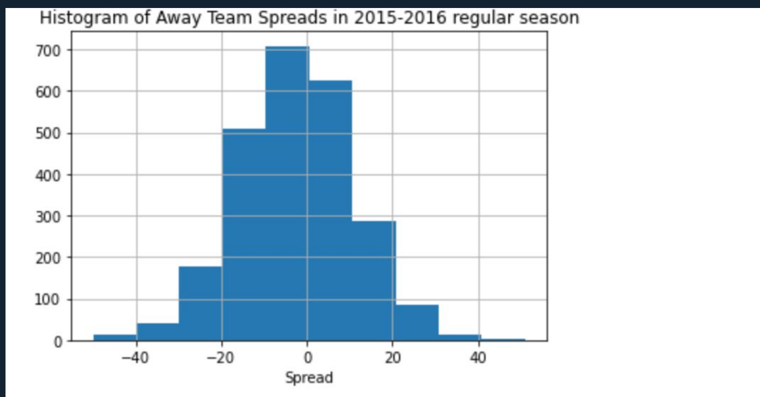
Average Spread is −3.23 and this is how much home court is worth.



Average Spread is −2.6 and this is how much home court is worth.



Average Spread is −2.41 and this is how much home court is worth.

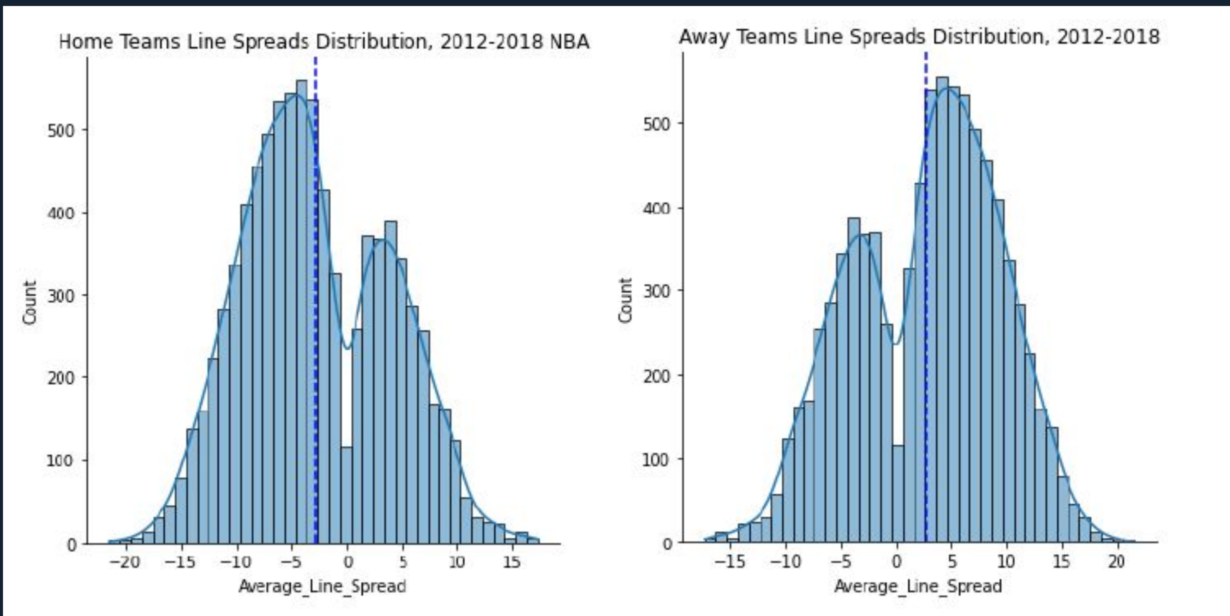


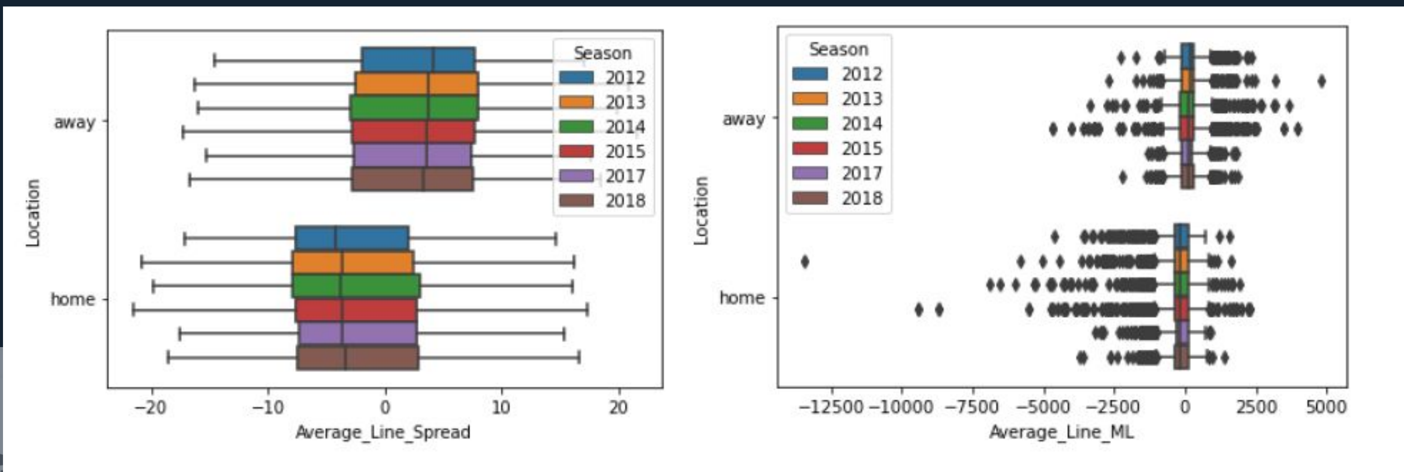
Average Spread is −2.67 and this is how much home court is worth.

Data Overview

EDA: Home advantage

Lines & Odds data (Sports books perception)

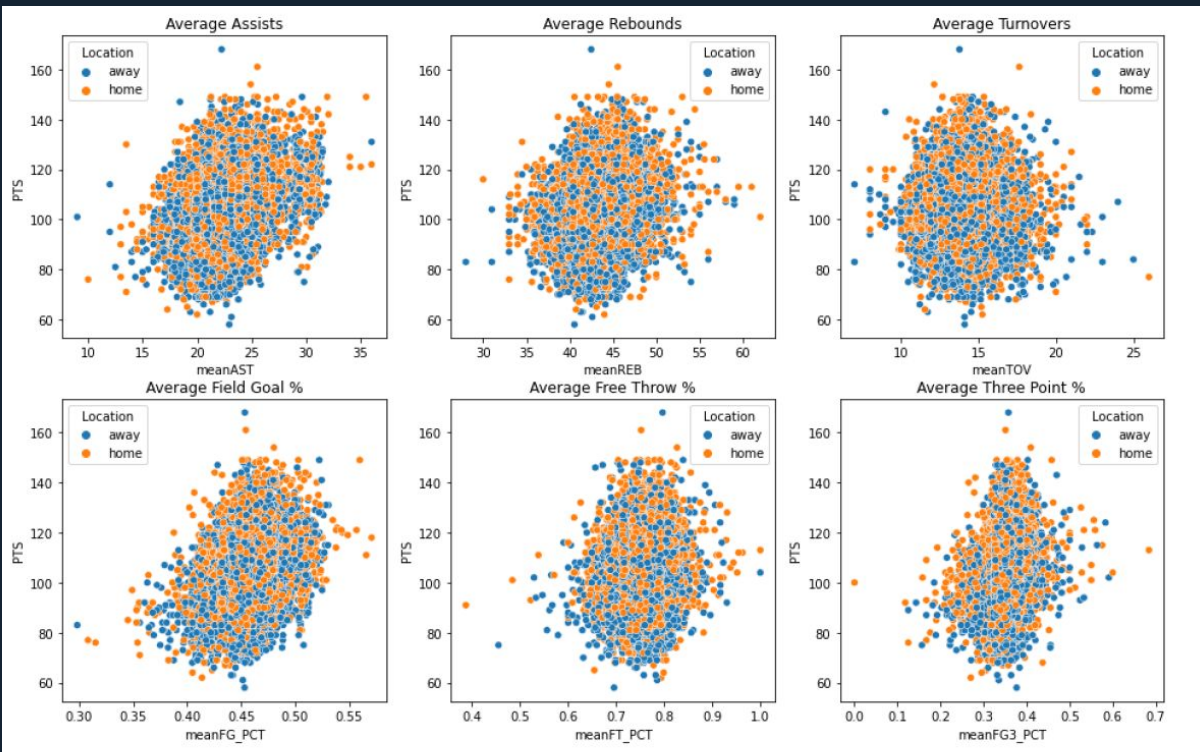


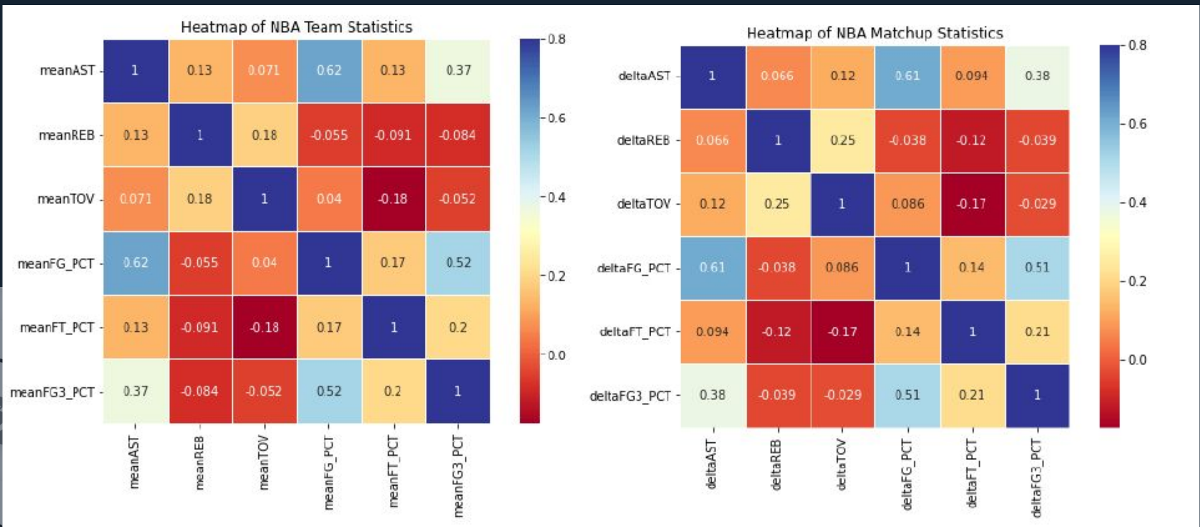


Data Overview

EDA: Feature Correlation

Box score statistics (incl. engineered features)





Data Overview

EDA: Team segmentation

1. **Floor sweepers**
   - GSW, SAS were top winners
2. **The let down**
   - PHX, PHI were top losers
   - MIL huge comeback in 5 years
3. **Scoring machines**
   - not necessarily teams with the most wins
4. **Odds beaters**
   - major occurrences in 2015
5. **The floppers**
   - major occurrences in 2015

1

		Team_Wins
		max
Season	TEAM_ABBREVIATION	
2015	GSW	73
2014	GSW	67
2015	SAS	67
2016	GSW	67
2012	MIA	66
2017	HOU	65
2013	SAS	62
2016	SAS	61
2018	MIL	60
2012	OKC	60
2014	ATL	60
2017	TOR	59
2013	OKC	59
2017	GSW	58
2012	SAS	58
2018	TOR	58
	GSW	57
2015	CLE	57
2012	DEN	57
2013	LAC	57
	IND	56

2

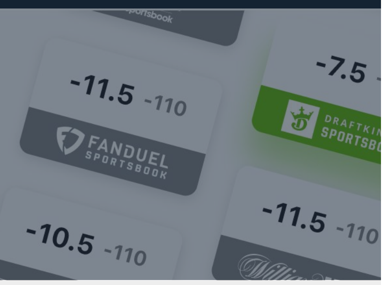
		Team_Losses
		max
Season	TEAM_ABBREVIATION	
2015	PHI	72
2013	MIL	67
2014	MIN	66
	NYK	65
2015	LAL	65
2018	NYK	65
2014	PHI	64
2018	PHX	63
2013	PHI	63
2018	CLE	63
2016	BKN	62
2012	ORL	62
	CHA	61
2015	BKN	61
2014	LAL	61
2017	PHX	61
2018	CHI	60
2017	MEM	60
2013	ORL	59
2015	PHX	59
2017	ATL	58

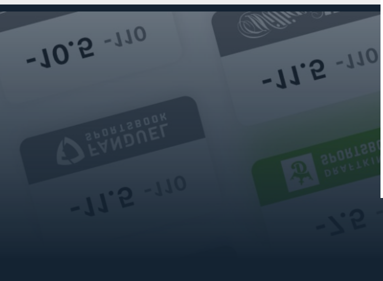
4

		upset
		sum
Season	TEAM_ABBREVIATION	
2015	DEN	44
	BKN	38
	NOP	34
	MIN	34
	MIL	34
	DET	34
	NYK	32
	WAS	30
	POR	30
	SAC	30
	DAL	30
	LAL	30
	CHA	30
	ORL	26
	CHI	26
	HOU	24
2013	CHA	24
2018	ATL	24
	ORL	23
2015	TOR	22
2013	LAL	21

5

		flop
		sum
Season	TEAM_ABBREVIATION	
2015	ATL	46
	HOU	44
	CLE	40
	BOS	38
	OKC	36
	DET	34
	CHI	32
	MIA	32
	IND	32
	UTA	30
2017	OKC	29
2015	PHX	28
	SAC	28
	TOR	28
	WAS	28
	LAC	26
	MEM	26
2018	OKC	25
2017	CLE	25
2015	CHA	24
	MIN	24





3

		meanPTS	
		max	min
Season	TEAM_ABBREVIATION		
2017	IND	140.000000	105.500000
2018	NOP	140.000000	115.240506
2017	BKN	131.000000	104.925926
2016	IND	130.000000	103.375000
	SAS	129.000000	101.181818
2018	POR	128.000000	110.777778
	LAL	126.750000	111.311111
	NYK	126.000000	104.573171
	SAC	125.666667	112.980000
2015	OKC	125.500000	107.269231
2013	GSW	125.000000	101.562500
2018	GSW	125.000000	108.000000
	MIN	125.000000	108.000000
2017	GSW	124.500000	113.463415
	POR	124.000000	101.611111
2018	UTA	123.000000	105.000000
2017	TOR	122.500000	107.777778
2015	NYK	122.000000	96.058824
2016	BOS	122.000000	103.571429
2017	HOU	122.000000	103.857143
2018	MIL	121.888889	113.000000

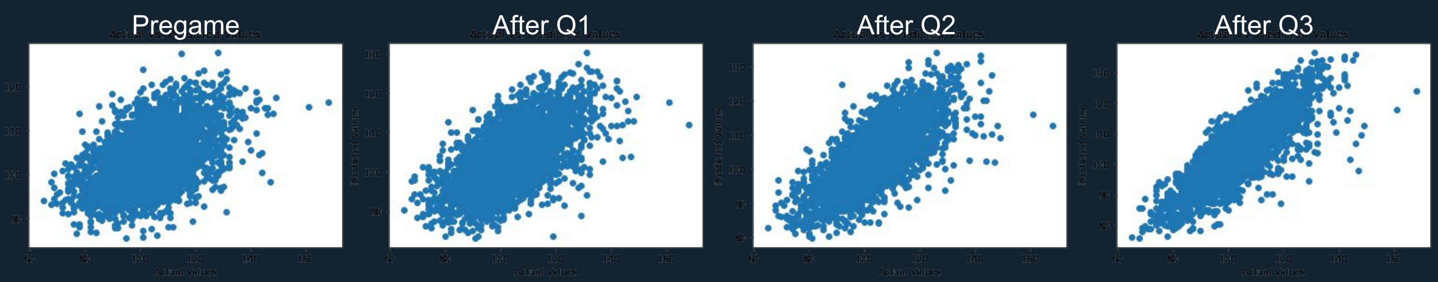
Predictive Models

Approach 1 - Regression Model Selection & Training

Regression Model

For Approach one we built four random forest regressors (one model fit with pregame data, one with data after Q1, one after Q2, & one after Q3 to predict the individual final scores of each team, and alert us via a Domo push alert the difference between our prediction and the available betting line exceeds a certain threshold.

The difference between the final outcome and our predicted outcome was more accurate than the gambling line, even prior to the game starting, and becomes increasingly more accurate as the game progresses. Prior to the games starting our model was able to reduce the average difference between predicted outcome vs actual outcome by 12.12% vs the Las Vegas line and 24.84% vs the Las Vegas over under. The results predictively improve after each quarter as the model is being fed more information about what is occuring in the game, peaking with an error reduction between our predicted outcome and the actual outcome of ~77% in the spread and ~93% in the point total (it is worth noting that the pregame spread was used as a feature in our model).



The 4 above charts represent the models prediction on the y axis vs the actual outcome on the x axis. The more information about game in progress the model has, the more noticeably linear the relationship between prediction and outcome becomes, which we can take advantage of via the live line. Predicted point totals were derived from summing the models predicted outcome for each individual team, and were more accurate than the individual predictions. After three quarters the mean difference between our models predicted point total and the actual total was .039 points with a standard deviation of ~12 points, vs the available pregame total's average difference of 1.15 points vs the actua outcome with a standard deviation of ~18 points.

Sports Lines Prediction

Approach 2

Regression models
Target: Spread

- Trained and tested 3 types of model: Random Forest, a Support Vector Machine (Epsilon-Support Vector Regression), and Gradient Boosting Tree (eXtreme Gradient Boosting)
- Mean absolute error (MAE) is selected as model error metric for its direct interpretability and robustness to outliers
- One-hot encoding of categorical variable
- Wrapped in SKlearn pipeline
- Performed 10-fold cross validation with resampling on SVM pipeline
- Further compare model performance with dataset partitioning into visiting team and home team
- Best performing model: **Random Forest on home teams data**

Sports Lines Prediction

Approach 2

Regression models
Target: Spread

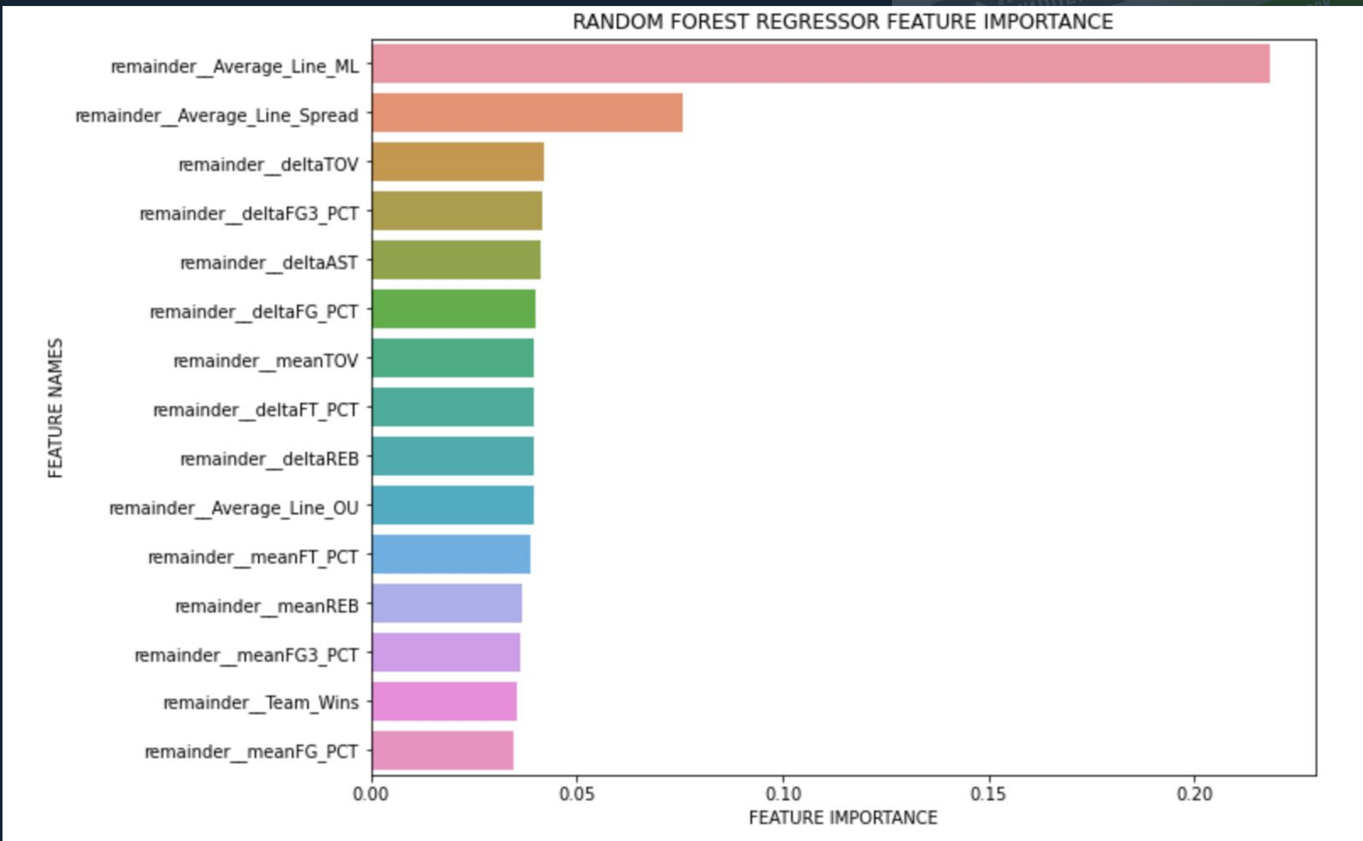
- Result summary

Target	Dataset	Model	Metrics	Score
Point Spread	All Teams	RF	MAE R^2	8.26 0.343
Point Spread	Home Teams	RF	MAE R^2	7.99 0.354
Point Spread	Away Teams	RF	MAE R^2	7.99 0.354
Point Spread	Home Teams	SVM	MAE R^2	9.33 (mean) 0.217 (mean)
Point Spread	Home Teams	XGB	MAE R^2	8.59 0.265

- Best MAE of 7.99 is beyond average sports books line spreads

Sports Lines Prediction

- Feature Importance (RF on home teams data)



Top 15 features in weight

The most important features are found to be the sportsbooks pre-game published moneyline and line spread. No surprise as excellent predictors given resources of sportsbooks to leverage supercomputers in determining the optimal line and odds to maintain balanced bets on both sides.

Most of the matchup deltas outrank the per-game statistics as more important model features. We think that the raw value of a statistic would matter less for prediction of one team's superiority if the other team also had a high value in that metric. Deltas paint a more accurate matchup.

Sports Lines Prediction

Approach 2

Classification models
Targets: Cover spread, Win/Loss (Moneyline), Over/Under

- Trained and tested 3 types of model: Gradient Boosting Tree (eXtreme Gradient Boosting), Random Forest, Random forest embedding paired with logistic regression

- Medium tree depth: 6, large number of trees: 500

- One-hot encoding of categorical variable

- Wrapped in SKlearn pipeline

- Further compare model performance with dataset partitioning into visiting team and home team

- Best performing model for each target:
    - **Cover Spread: XGB on away teams data**
    - **Win/Loss: XGB on away teams data**
    - **Over/Under: XGB on away teams data**

Sports Lines Prediction

Approach 2

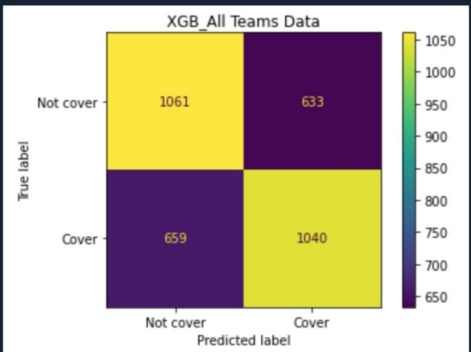
Classification models
Target: Cover spread

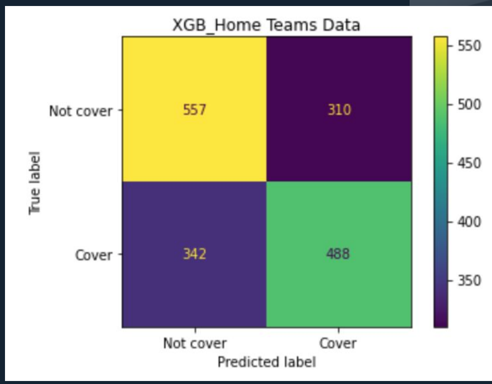
- Result summary

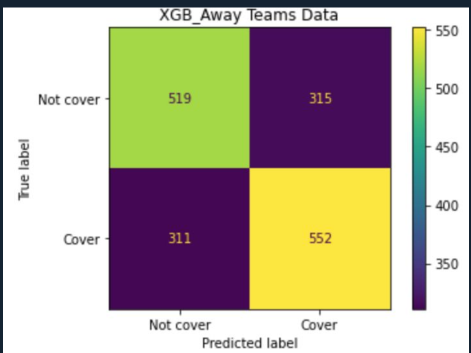
Target	Dataset	Model	Metrics	Score
Cover Spread	All Teams	XGB	Accuracy AUC	61.92%
Cover Spread	Home Teams	XGB	Accuracy	61.58%
Cover Spread	Away Teams	XGB	Accuracy AUC	63.11% 0.70
Cover Spread	Away Teams	RF	Accuracy AUC	56.04% 0.58
Cover Spread	Away Teams	RT + LR	Accuracy AUC	61.17% 0.63

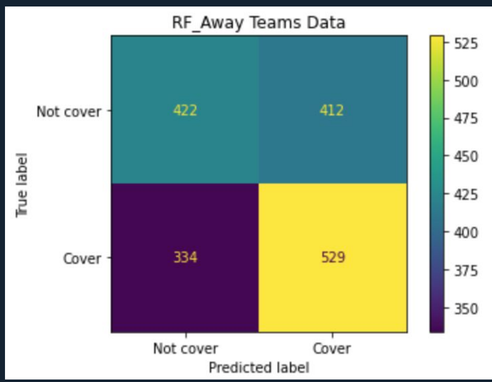
Sports Lines Prediction

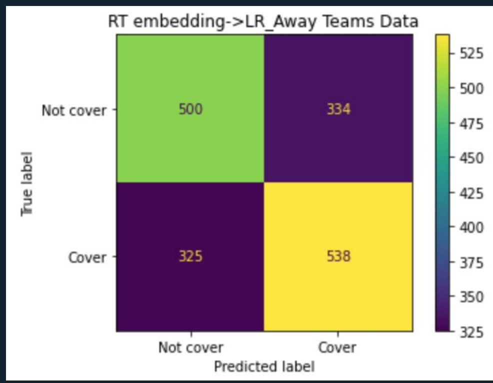
- Confusion matrices



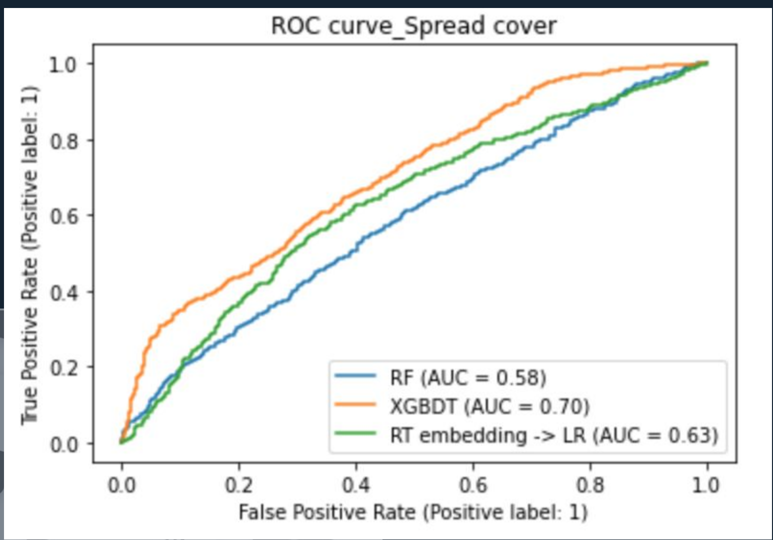






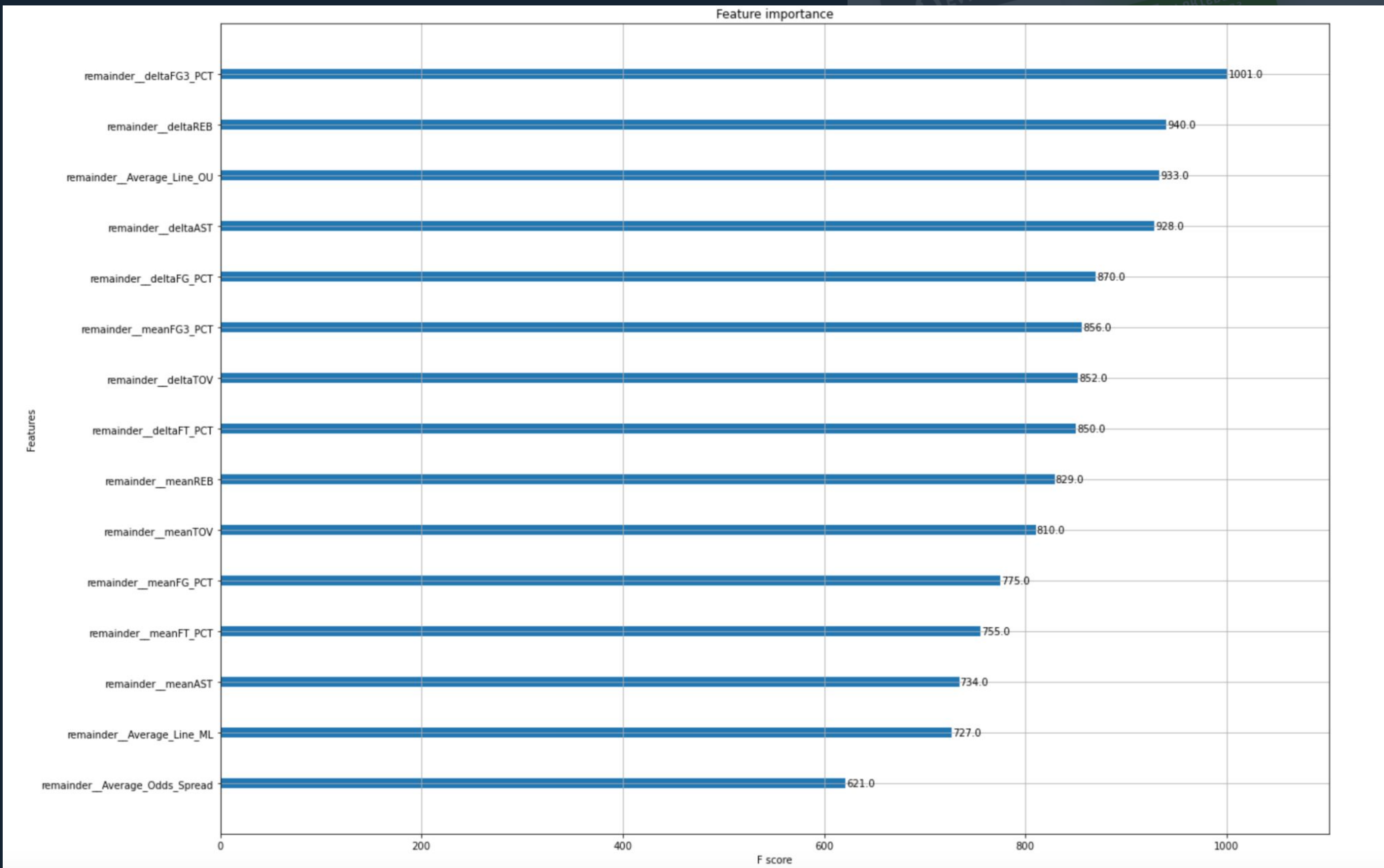


- ROC Curve (Away Teams Data)



Sports Lines Prediction

- Feature importance (XGB on away teams data)



Top 15 features in weight

Dominated by matchup deltas. Interesting observation: Sports books lines OU appears as 3rd most important feature.

Sports Lines Prediction

Approach 2

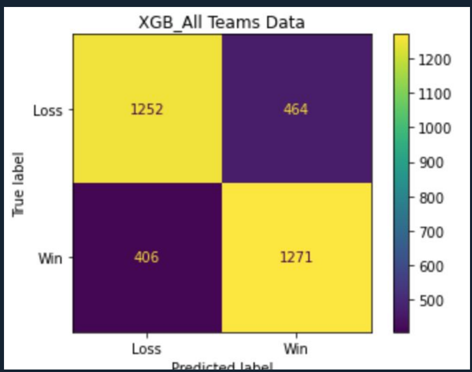
Classification models
Target: Win/Loss

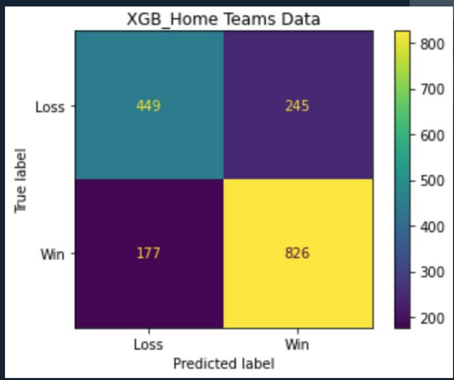
- Result summary

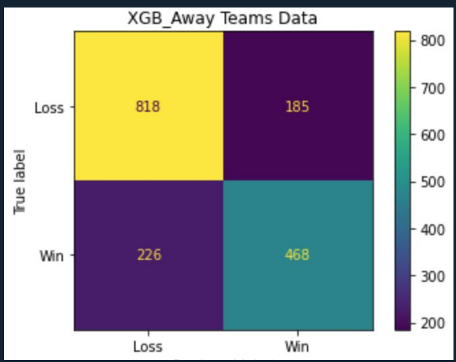
Target	Dataset	Model	Metrics	Score
Win/Loss	All Teams	XGB	Accuracy AUC	74.36%
Win/Loss	Home Teams	XGB	Accuracy	75.13%
Win/Loss	Away Teams	XGB	Accuracy AUC	75.78% 0.83
Win/Loss	Away Teams	RF	Accuracy AUC	70.54% 0.77
Win/Loss	Away Teams	RT + LR	Accuracy AUC	70.77% 0.75

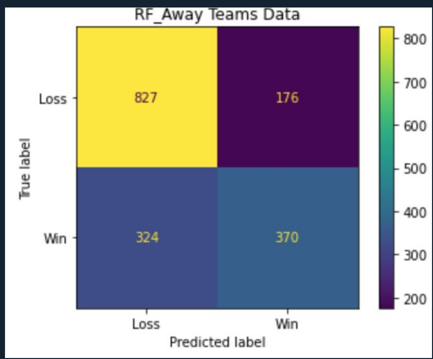
Sports Lines Prediction

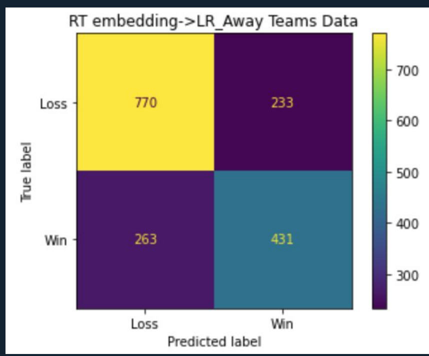
- Confusion matrices



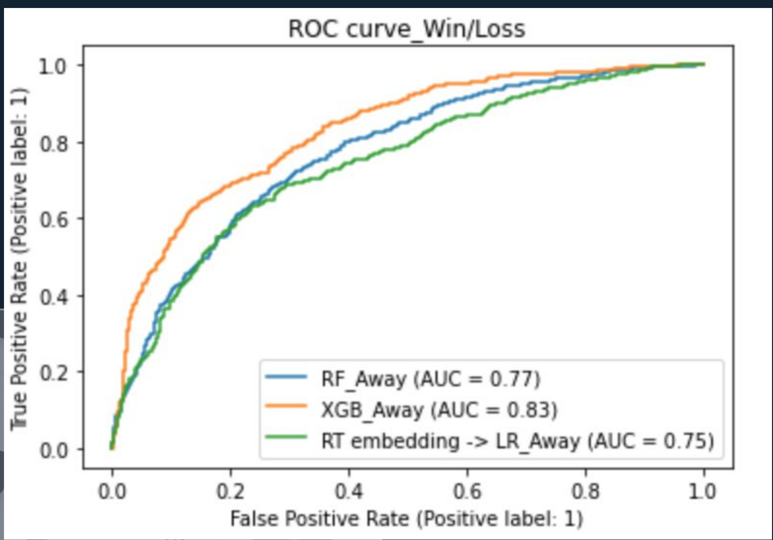






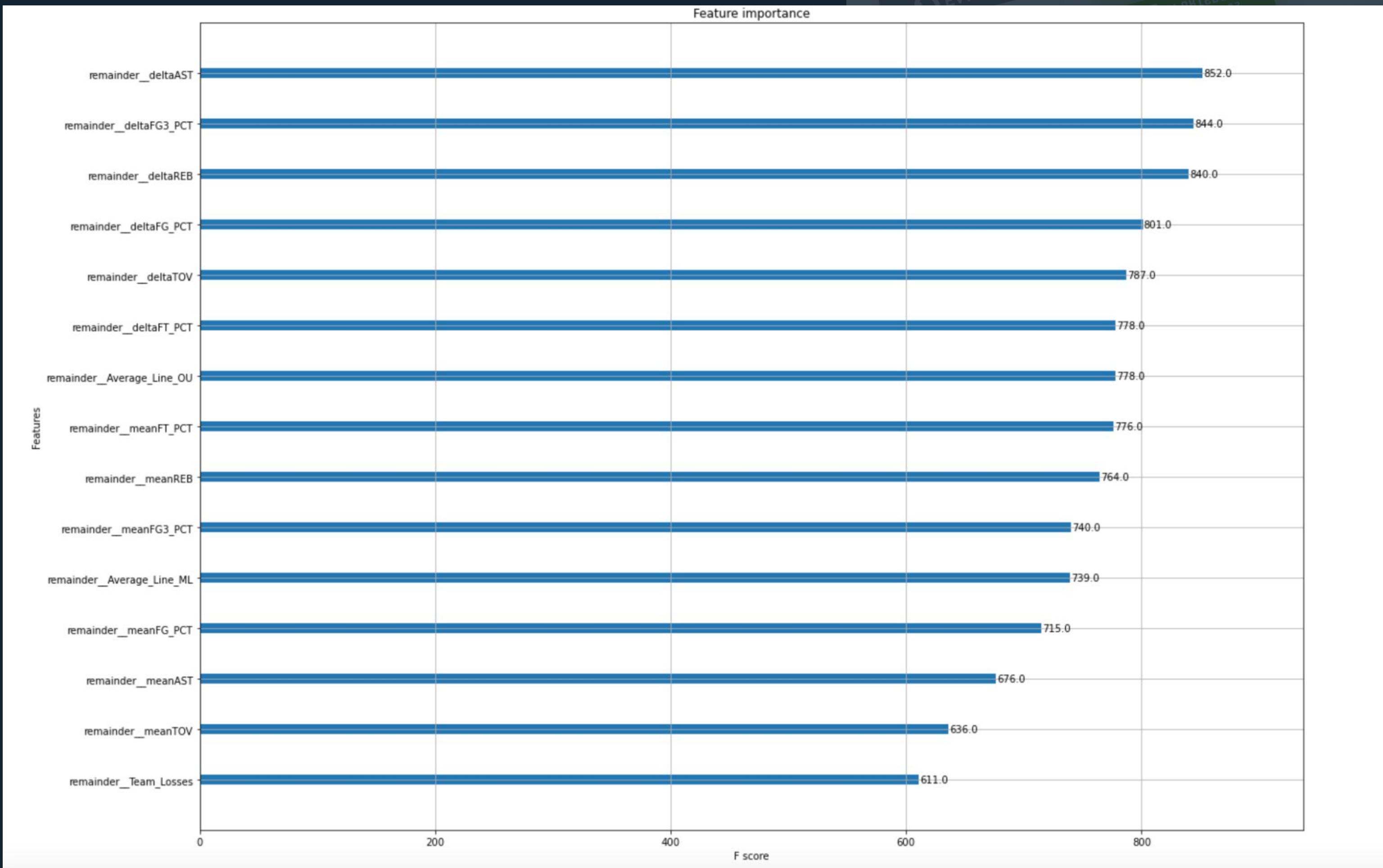


- ROC Curve (Away Teams Data)



Sports Lines Prediction

- Feature importance (XGB on away teams data)



Top 15 features in weight

Dominated by matchup deltas, led to highest accuracy of all models.

Sports Lines Prediction

Approach 2

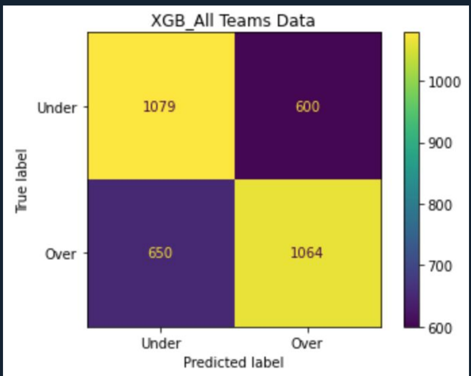
Classification models
Target: Over/Under

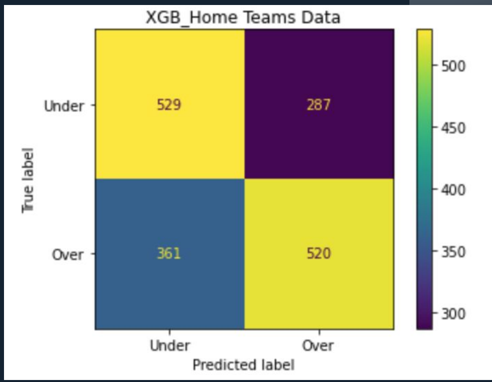
- Result summary

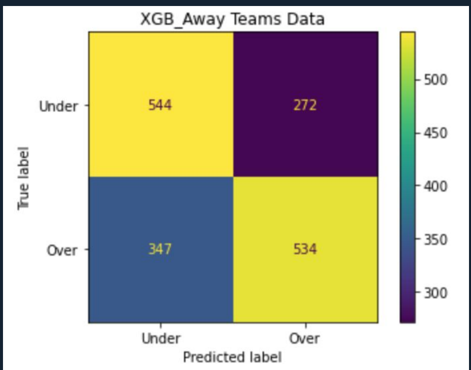
Target	Dataset	Model	Metrics	Score
Over/Under	All Teams	XGB	Accuracy	63.16%
Over/Under	Home Teams	XGB	Accuracy	61.81%
Over/Under	Away Teams	XGB	Accuracy AUC	63.52% 0.71
Over/Under	Away Teams	RF	Accuracy AUC	55.51% 0.61
Over/Under	Away Teams	RT + LR	Accuracy AUC	59.52% 0.62

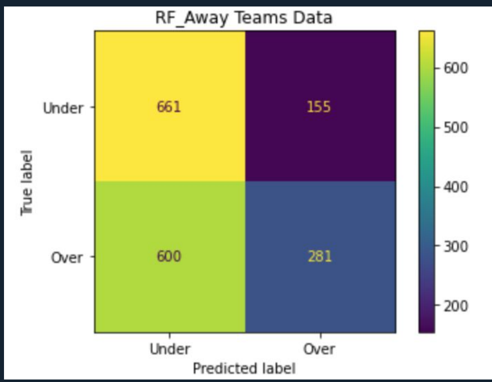
Sports Lines Prediction

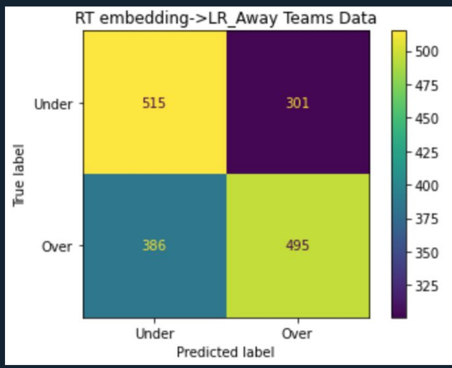
- Confusion matrices



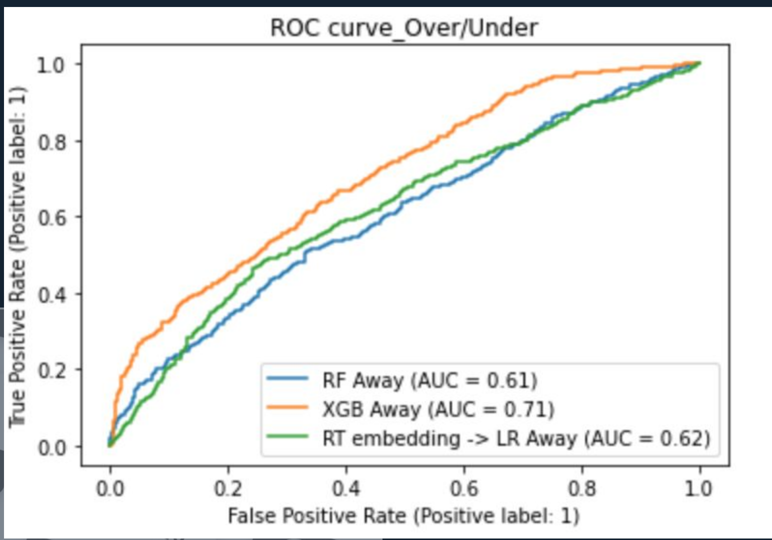






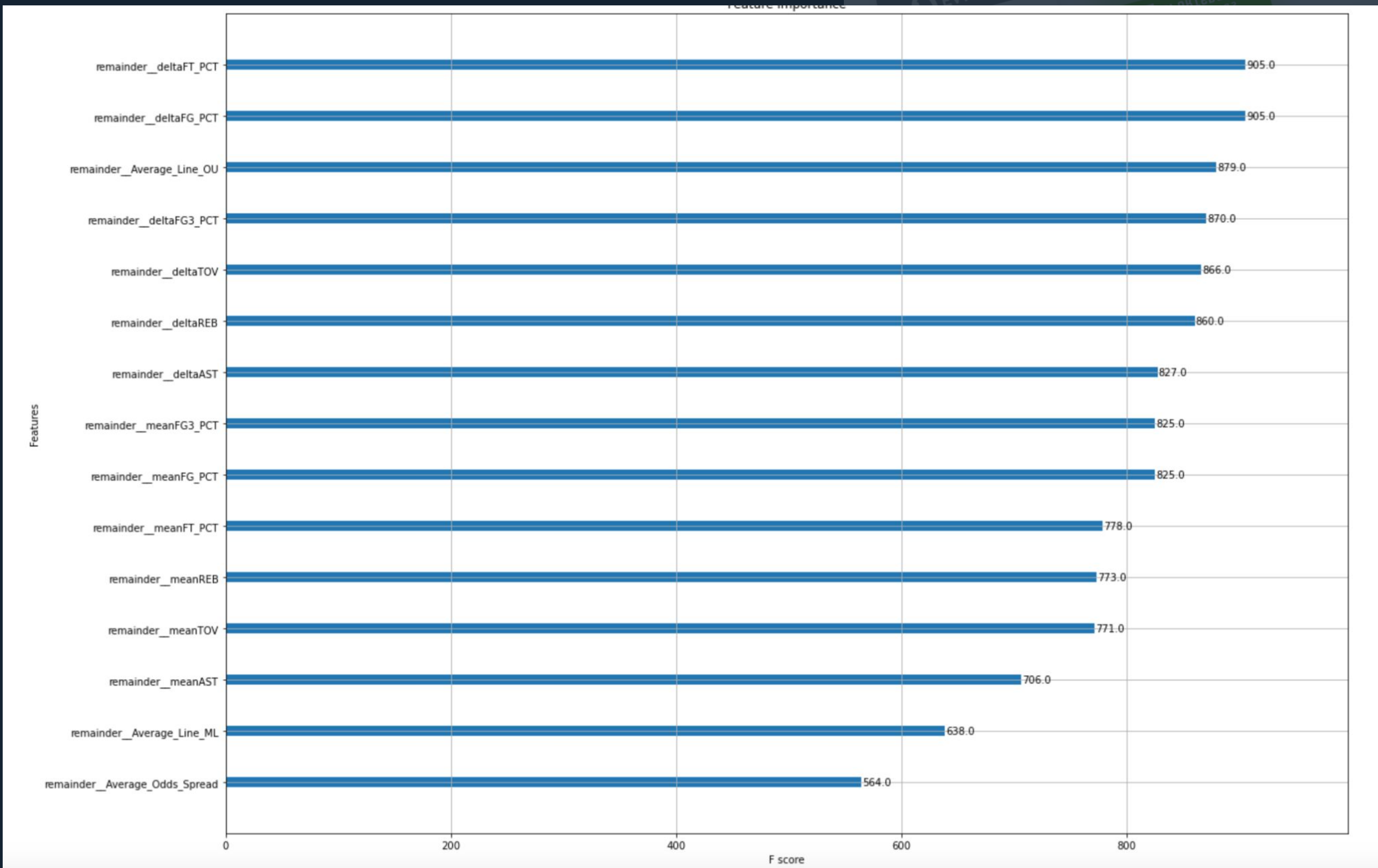


- ROC Curve (Away Teams Data)



Sports Lines Prediction

- Feature importance (XGB on away teams data)



Top 15 features in weight

Similar feature map as the "Cover spread" XGB model

Sports Lines Prediction

Key findings

- Approach 1 works better as an active approach. The model is designed and suited to predict point differentials and point totals by accounting for in-game statistics as the games progress, therefore it will synergize well with a live lines and odds tracking application.

- Approach 2 works better as a passive approach. The classification models are designed with pre-game data in mind as the training data are offset by one game for each specific team.

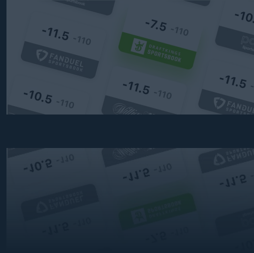
- Among the classification models, XGB has outperformed the other tested methods. Furthermore, using the partitioned dataset with “away teams only” as input has consistently led to a higher accuracy upwards of 60-75%.

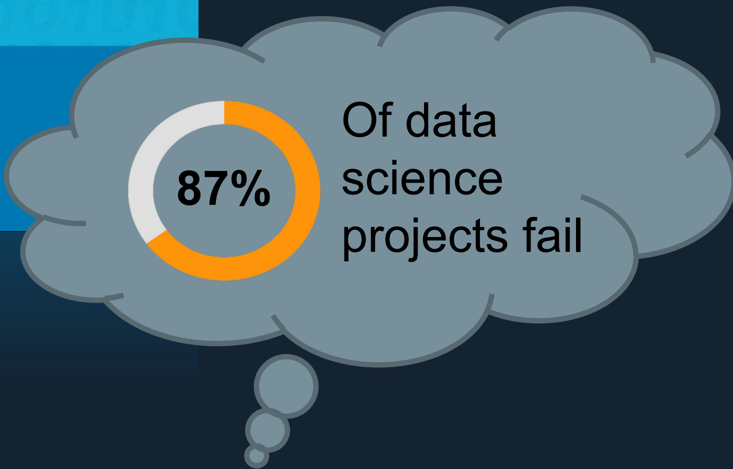
- Both approaches present a viable opportunity to scale up for long term profits in NBA online pre-game and live wagering

Scope Reduction

Selecting the Pilot – Minimum Viable Product (MVP)

Avoiding Failure

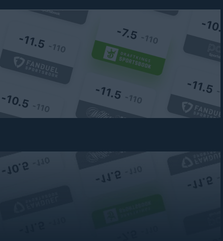




Venture Beat, 2019

Data science projects have high failure rates

- **Exceed Budget** – falsely believe throwing money can solve a problem
- **Lack Collaboration** – these projects tend to span across silos, making it difficult to gain access to required data



The 3 Ss to success

- **Specificity** – focused target product to NBA win probabilities based on moving Point Spread
- **Scope** – narrowed scope to help limit data sources to center efforts on a single model
- **Simplicity** – reduced complexity to deliver a working prototype as part of the MVP

Architecture Overview

Functional Requirements

The first step for developing the pilot was to define our functional requirements.

INPUTS	Must be able to ingest the output from our Python predictive model
	Must be able to fetch live, in-game odds
	Must be able to work with NBA Points Spread
	Must be able to access the live odds from MGMSports
OUTPUTS	Must be able to dynamically refresh and update into Domo
	Must have a live connection into Domo
	Must have a Domo dashboard
	Must have a triggered alerting mechanism for SMS, app pushes, and emails

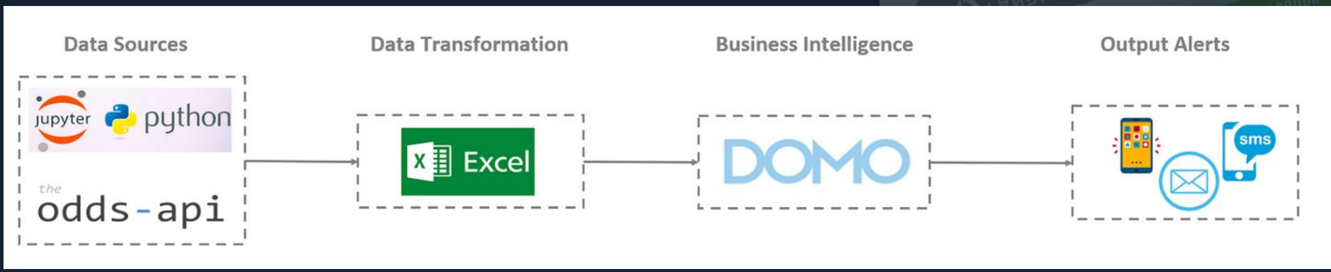
Design Considerations

The table below outlines the selected tools and reasons.

Tool	Selected Vendor	Competitors Considered	Reasons for Selection
Live in-game Odds	the odds-api	•OddsJam •APILayer •SportRadar	•They offered a free plan that allowed 500 calls per month (however a Starter plan was eventually purchased) •They provided better support through online tutorials •They had recently released both a Python and R library, so they offered better scalability •They had a working Excel add-in which could be leveraged for the prototype
Data Extraction and Transformation	Excel	•R Studio •Python •PowerBI	•Considering the Odd API used MS Excel, it made more sense to stay in Excel •Future state will migrate into R/Python
Cloud Data Repository	DOMO	•Tableau •Azure	•Free version of Domo was more accessible •Smaller learning curve •Direct data connectors for MS Excel
Alerting Mechanism	DOMO	•R/Twilio •Python/SMS	•Competitors were very limited •Solutions through R and Python had limited volumes per month •Domo's mobile app also allowed for push notifications, in addition to SMS Texts and Emails

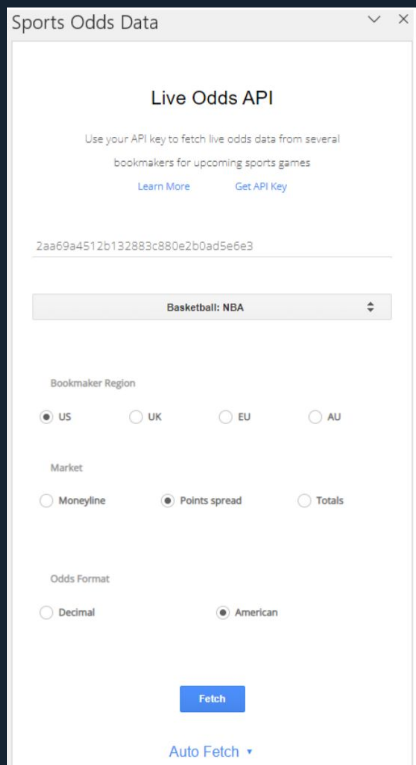
Systems Architecture

For the pilot, we needed 4 distinct components - Live odds, a way to translate the data, a cloud data repository, and an alerting mechanism.



1. Data Sources

For the pilot, we needed 4 distinct components - Live odds, a way to translate the data, a cloud data repository, and an alerting mechanism.



2. Data Transformation

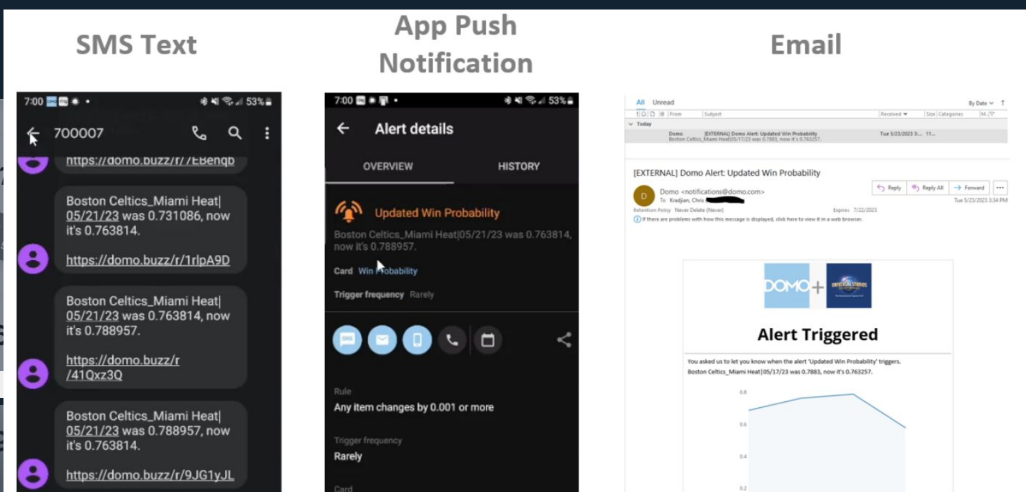
Download and configure the Live-Odds-API excel add-on. At the right, is a screenshot of its UI

3. Cloud-based BI Platform

Connect our MS Excel tool to our SaaS BI platform, Domo

4. DOMO Dashboard w/ Output Alerts

Once uploaded to DOMO custom dashboard, it would trigger automated alerts anytime the opening lines changed.



Data Sources

The process starts with our Python predictive model, where we evaluate the following 9 scenarios:

- all_spread_bet
- home_dogs
- home_fav_between_zero_and_five
- home_fav_between_five_and_ten
- home_fav_ten_plus
- road_favorites
- road_dog_zero_to_five
- road_dog_five_to_ten
- road_dog_ten_plus

Then these scenarios are evaluated across the following 10 changes:

- standard_spread
- plus_one_live_points
- plus_two_live_points
- plus_three_live_points
- plus_four_live_points
- plus_five_live_points
- plus_six_live_points
- plus_seven_live_points
- plus_eight_live_points
- plus_nine_live_points

Sample output from Python

Criteria	standard_spread	plus_one_live_points	plus_two_live_points	plus_three_live_points	plus_four_live_points	plus_five_live_points	plus_six_live_points	plus_seven_live_points	plus_eight_live_points	plus_nine_live_points
2012_all_spread_bet	0.50	0.53	0.58	0.61	0.64	0.67	0.69	0.73	0.75	0.78
2012_home_dogs	0.48	0.50	0.56	0.63	0.68	0.70	0.73	0.75	0.78	0.80
2012_home_fav_between_zero_and_five	0.48	0.52	0.57	0.57	0.59	0.61	0.63	0.68	0.72	0.75
2012_home_fav_between_five_and_ten	0.51	0.59	0.61	0.61	0.63	0.73	0.76	0.80	0.83	0.88
2012_home_fav_ten_plus	0.51	0.59	0.61	0.61	0.63	0.73	0.76	0.80	0.83	0.88
2012_road_favorites	0.51	0.53	0.57	0.59	0.60	0.61	0.63	0.65	0.69	0.71
2012_road_dog_zero_to_five	0.53	0.57	0.62	0.63	0.67	0.71	0.74	0.80	0.81	0.81
2012_road_dog_five_to_ten	0.48	0.50	0.55	0.59	0.61	0.66	0.69	0.72	0.75	0.79
2012_road_dog_ten_plus	0.52	0.55	0.59	0.68	0.71	0.71	0.74	0.77	0.78	0.81
2013_all_spread_bet	0.51	0.54	0.58	0.61	0.65	0.68	0.72	0.74	0.77	0.79
2013_home_dogs	0.53	0.55	0.57	0.60	0.66	0.69	0.73	0.77	0.80	0.82
2013_home_fav_between_zero_and_five	0.45	0.48	0.51	0.53	0.55	0.57	0.59	0.63	0.67	0.70
2013_home_fav_between_five_and_ten	0.44	0.45	0.49	0.56	0.62	0.68	0.73	0.75	0.78	0.81
2013_home_fav_ten_plus	0.44	0.45	0.49	0.56	0.62	0.68	0.73	0.75	0.78	0.81
2013_road_favorites	0.57	0.61	0.63	0.65	0.68	0.70	0.73	0.75	0.78	0.79
2013 [illegible]d_dog zero [illegible]	[illegible]	0.57	[illegible]	0.66	0.71	0.74	0.77	0.79	[illegible]	0.86
[illegible]	[illegible]	[illegible]	0.52	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2017_home_fav_between_zero_and_five	0.54	0.59	0.63	0.64	0.66	0.66	0.69	0.71	[illegible]	0.77
2017_home_fav_between_five_and_ten	0.39	0.45	0.51	0.54	0.62	0.68	0.71	0.77	0.82	0.84
2017_home_fav_ten_plus	0.39	0.45	0.51	0.54	0.62	0.68	0.71	0.77	0.82	0.84
2017_road_favorites	0.49	0.53	0.56	0.60	0.62	0.66	0.69	0.72	0.75	0.78
2017_road_dog_zero_to_five	0.47	0.54	0.58	0.63	0.65	0.67	0.69	0.72	0.76	0.76
2017_road_dog_five_to_ten	0.53	0.56	0.58	0.61	0.67	0.71	0.72	0.74	0.75	0.77
2017_road_dog_ten_plus	0.57	0.64	0.68	0.70	0.77	0.80	0.83	0.85	0.86	0.86
2018_all_spread_bet	0.48	0.50	0.54	0.57	0.60	0.63	0.66	0.69	0.72	0.75
2018_home_dogs	0.48	0.50	0.53	0.56	0.59	0.63	0.68	0.69	0.72	0.76
2018_home_fav_between_zero_and_five	0.50	0.51	0.54	0.55	0.57	0.58	0.64	0.66	0.69	0.74
2018_home_fav_between_five_and_ten	0.49	0.50	0.57	0.60	0.63	0.66	0.71	0.76	0.76	0.77
2018_home_fav_ten_plus	0.49	0.50	0.57	0.60	0.63	0.66	0.71	0.76	0.76	0.77
2018_road_favorites	0.48	0.52	0.52	0.56	0.58	0.62	0.64	0.69	0.72	0.75
2018_road_dog_zero_to_five	0.46	0.48	0.52	0.56	0.60	0.64	0.67	0.70	0.75	0.78
2018_road_dog_five_to_ten	0.48	0.51	0.56	0.60	0.62	0.63	0.66	0.69	0.71	0.74
2018_road_dog_ten_plus	0.51	0.52	0.53	0.54	0.60	0.64	0.68	0.70	0.72	0.74

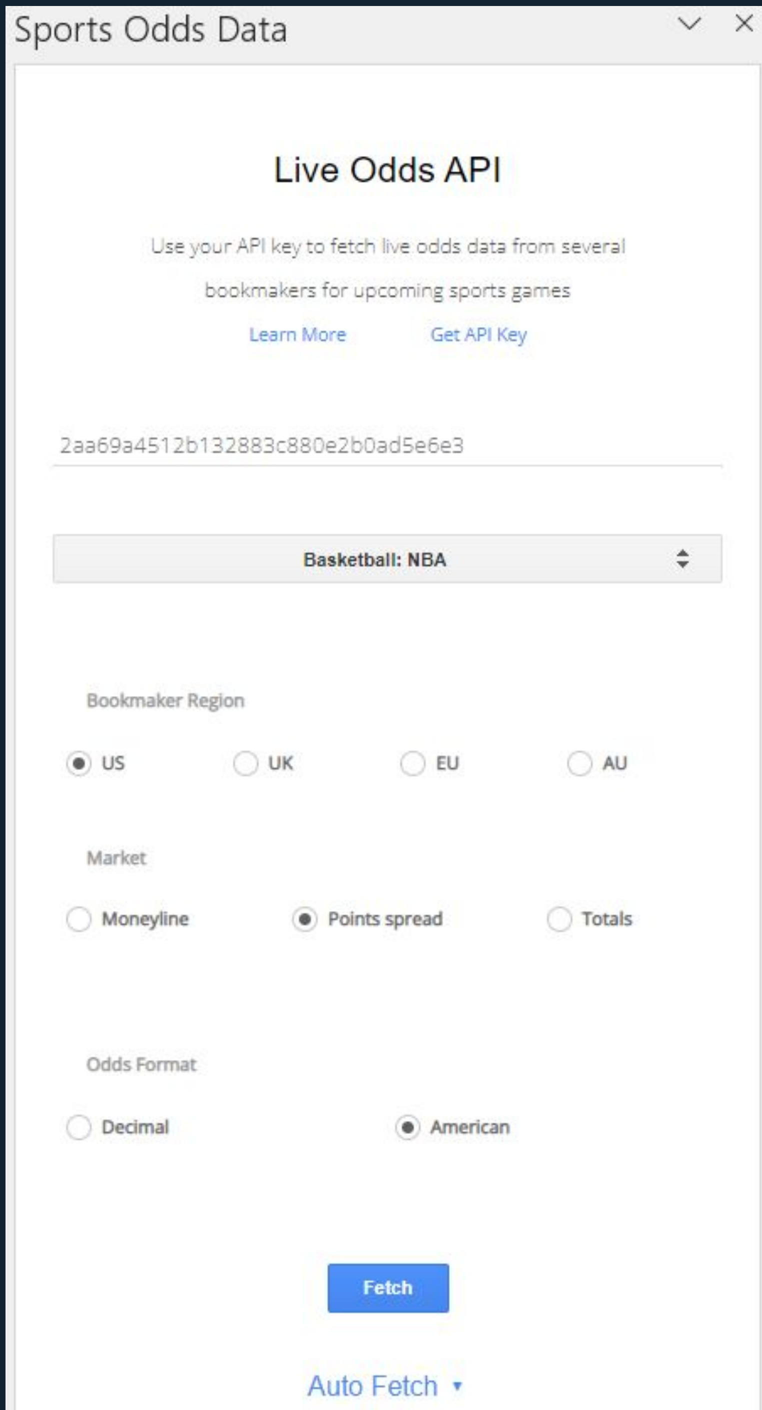
Data Transformation

The transformation layer starts with MS Excel, with the Live-Odds-API add-on installed and configured

Once a unique API Key has been obtained, the add-on allows you to fetch odds for dozens of sports, including all 4 major professional US sports (NBA, NFL, MLB, NHL), as well Golf, MLS, and every soccer league in the world.

You can then select various parameters, Bookmaker Region and then specify the specific game type.

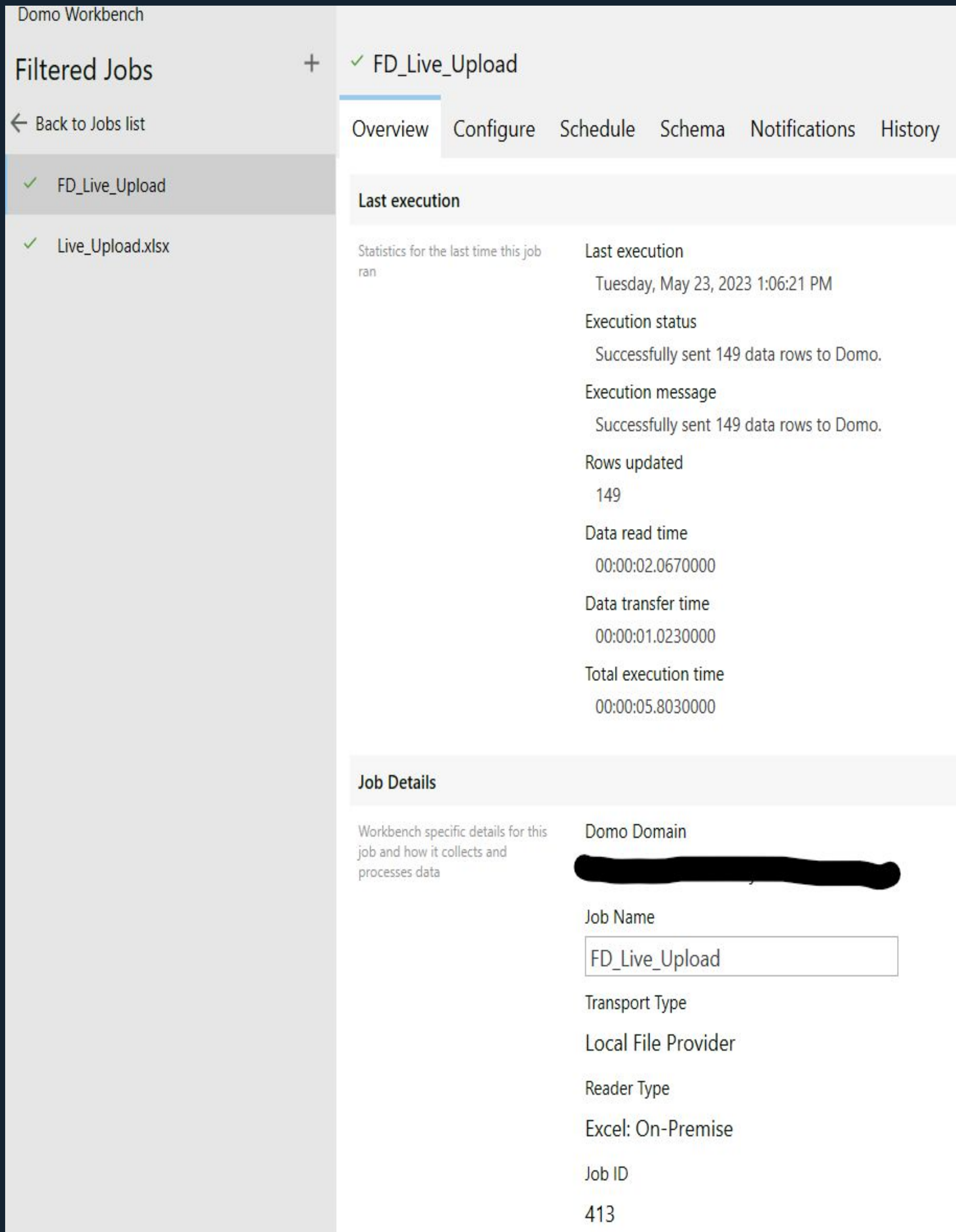
This add-on also allows you to set-up an automated data fetching interval, however the free license only allows 500 monthly pings.



Cloud-Based BI Platform

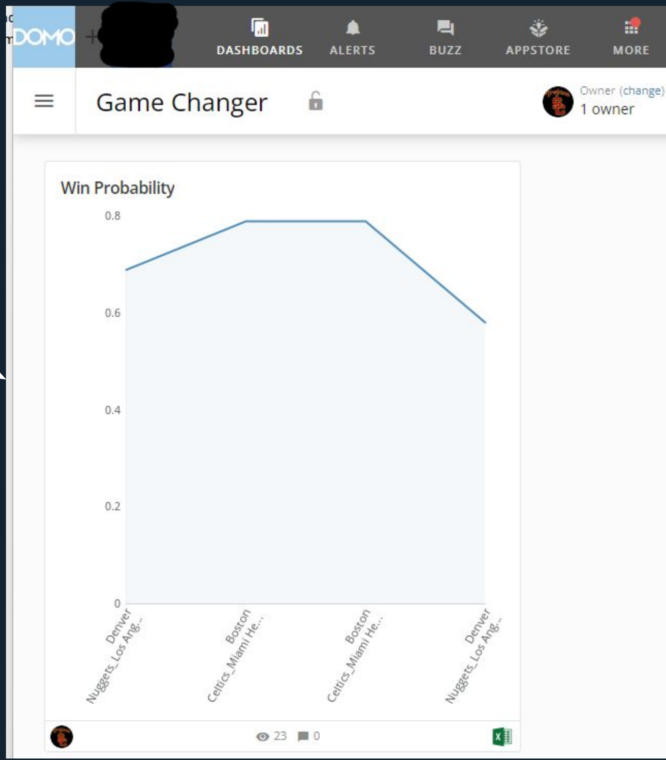
In this step, we established a live connection between our MS Excel tool and our SaaS BI platform, Domo, using an application called Domo Workbench.

This Domo Workbench job would effectively monitor the saved MS Excel model. As soon as its ‘Last Modified’ state was updated, Workbench would automatically start its upload job - normally within seconds.

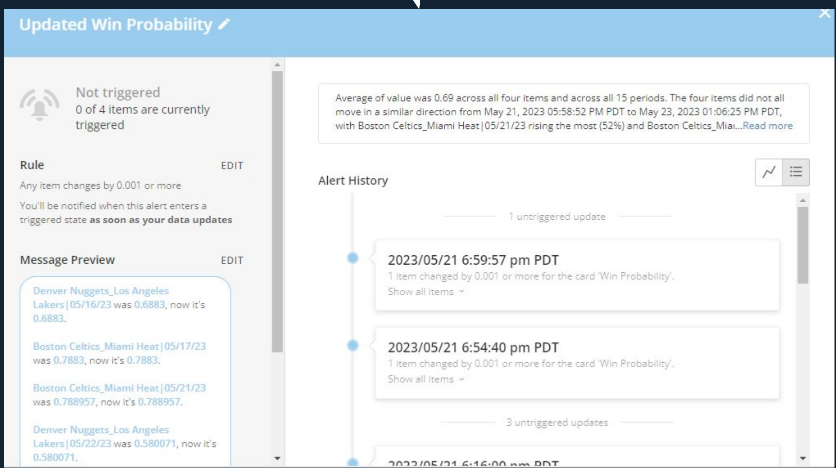


Output Alerts

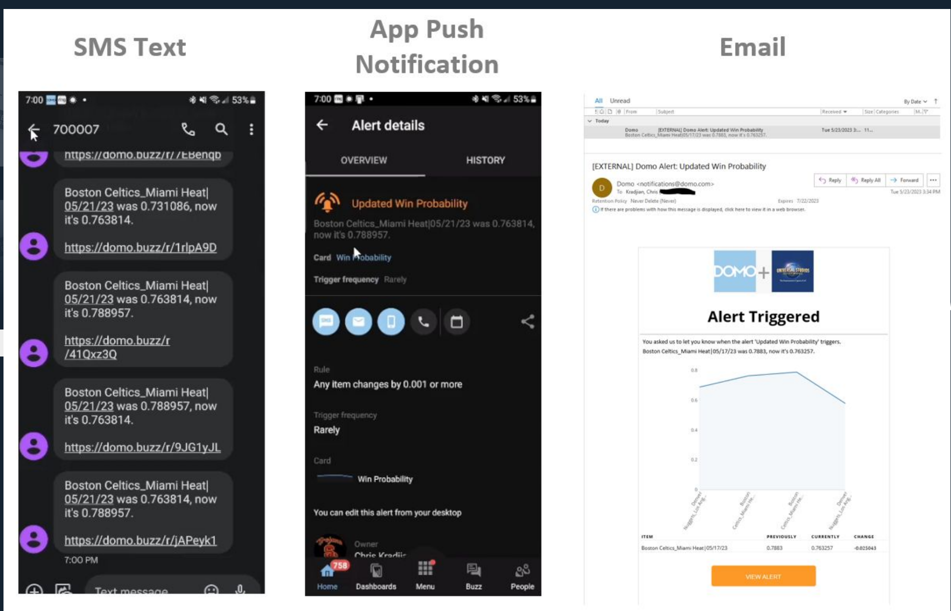
Once the source data was uploaded into the Domo UI, it automatically updates into a custom dashboard



This dashboard was configured with the following set of alerts.



Any time the opening lines would change by 0.1%, it would trigger an alert. - SMS texts, mobile app push notifications, and emails.



Live End-to-End Video Demonstration

Conclusions

Our initial analysis of the data validates that our strategy to profitability is supported by the historical data.

Analytics to Prototype

We have demonstrated the prototype can be deployed into production for NBA point spreads. We will continue building out this proof of of concept, which connects the back-end (data analysis/models) to the front-end (user interface).

Recommendations

Expand Further on Modeling

Linear programming optimization methods

Source more recent data than 2018 season, and include playoffs data

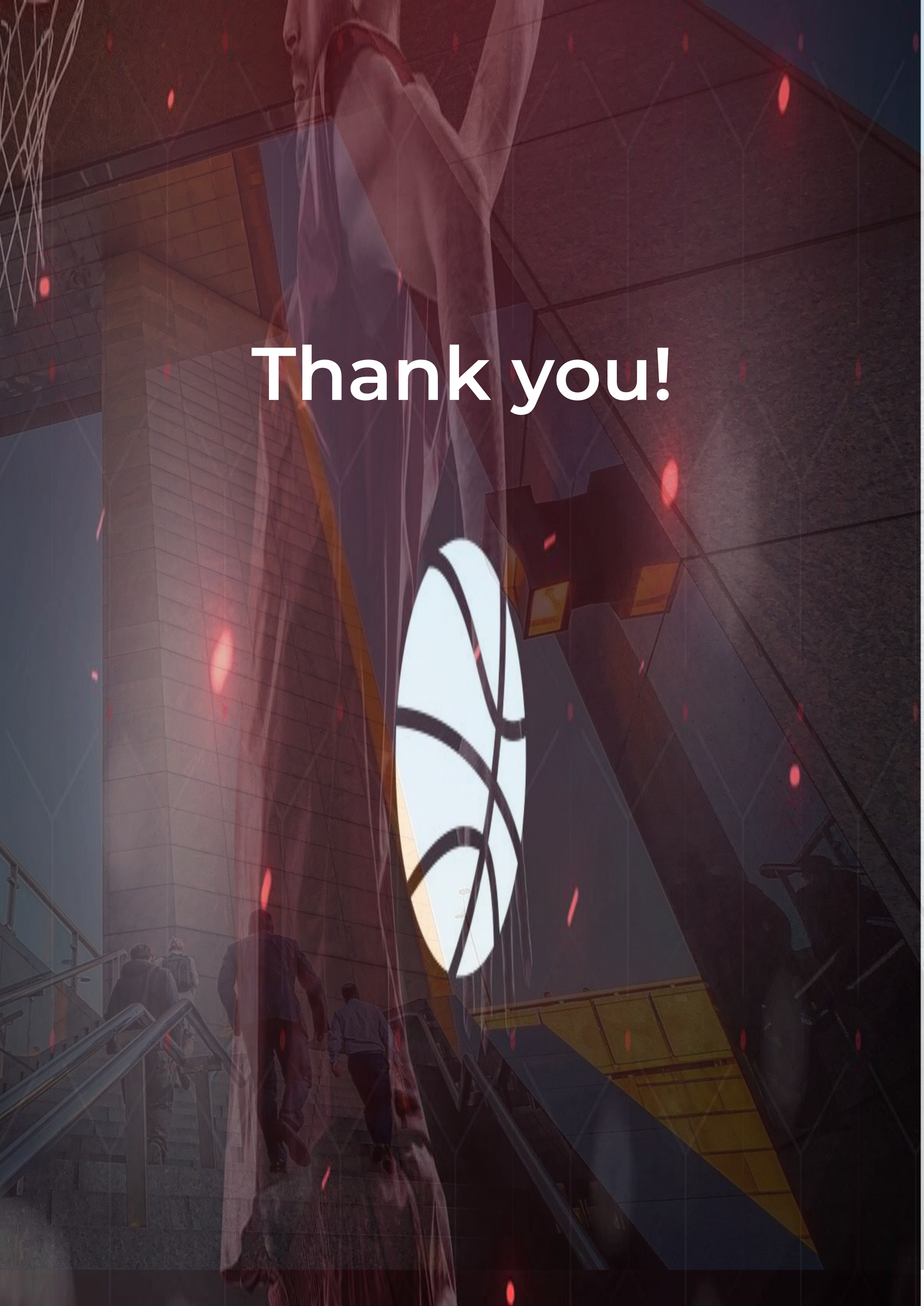
Explore additional features: Offense vs Defense metrics, player metrics etc.

Scale

Translate methodologies to other sports and/or bet types

Automating Live In-Game Odds

Live connections with API


Thank you!