DRAFT KINGS | THE CROWN IS YOURS

June 13th

How to (Maybe) Make Money Betting on the NBA with Rocket Science

By Rob Barnes and Siu Fai Hsu

DRAFT KINGS | THE CROWN IS YOURS

AGENDA

1. What We Do
2. The “Rocket Science”
3. Our NBA Kalman Filters
4. How We’re Utilising Databricks
5. How You Make Money
6. Summary

DRAFT KINGS | THE CROWN IS YOURS

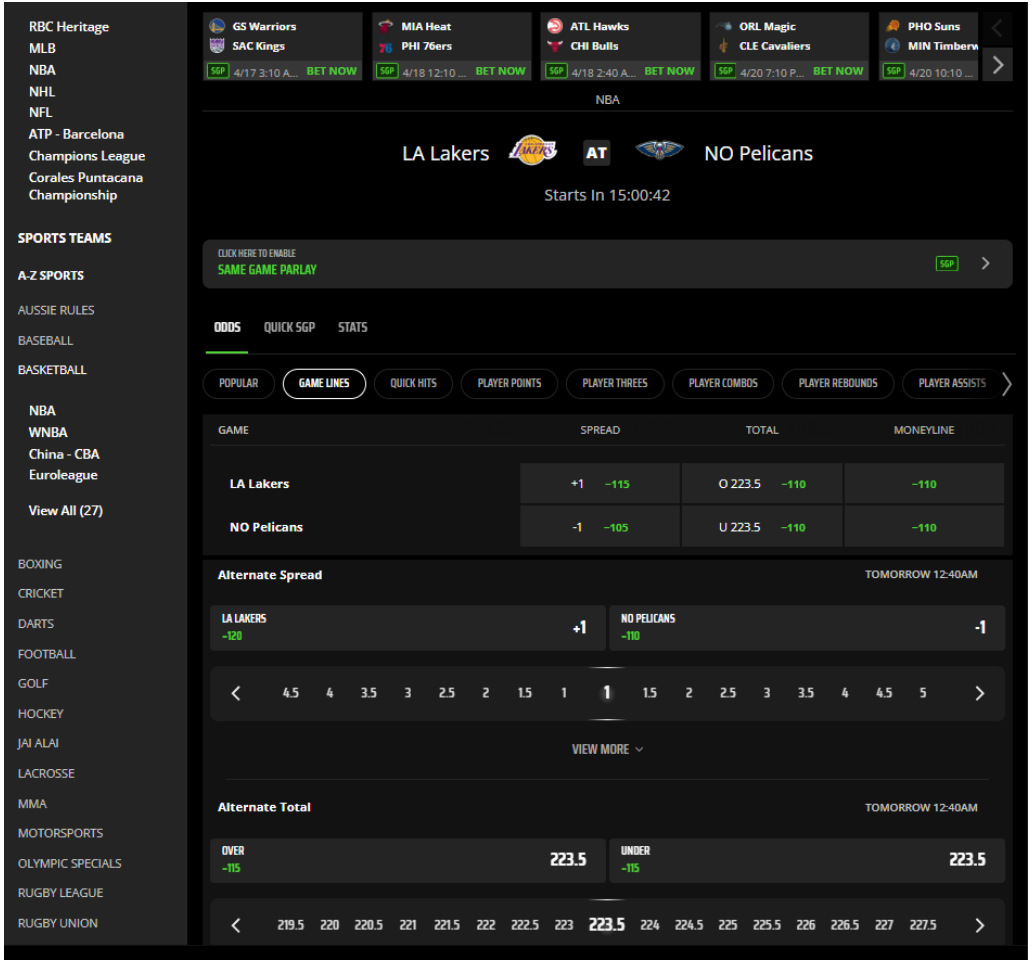
01 What We Do

What We Do

We are part of the Sports Data Science Team at DraftKings, responsible for setting the prices customers see on our sportsbook.

Our complex monte carlo simulation based pricing engines are powered by multiple ML models and generate probabilities for various sports events, such as:

- Final game scores
- Timing of the next touchdown
- The next team to score a 3-pointer



Business Problem

WHAT

At the end of 2022 we were asked to produce models that are capable of producing NBA player props markets (over/under player gets x points/assists etc)

HOW

- Produce player centric NBA models that are able to understand the value of players
- Produce several new player based features and models that to help us produce props markets
- Step change our MLOps practices by utilizing Databricks to produce robust feature and model pipelines and monitor our models.

DRAFT KINGS | THE CROWN IS YOURS

20 The “Rocket Science”

KALMAN FILTERS

Kalman Filters are an **algorithm** that uses a series of measurements to produce an estimate of an unknown variable.

They are designed to work in scenarios where high degrees of **uncertainty** are present, both in the variable you're attempting to estimate and the the method being used to measure.

Kalman Filters also account for **uncertainty in the estimate** which gets updated alongside the estimate after every measurement.

Estimate Update

$$\hat{\boldsymbol{x}}_{n,n} = \hat{\boldsymbol{x}}_{n,n-1} + \boldsymbol{K}_n(\boldsymbol{z}_n - \boldsymbol{H}\hat{\boldsymbol{x}}_{n,n-1})$$

Kalman Gain

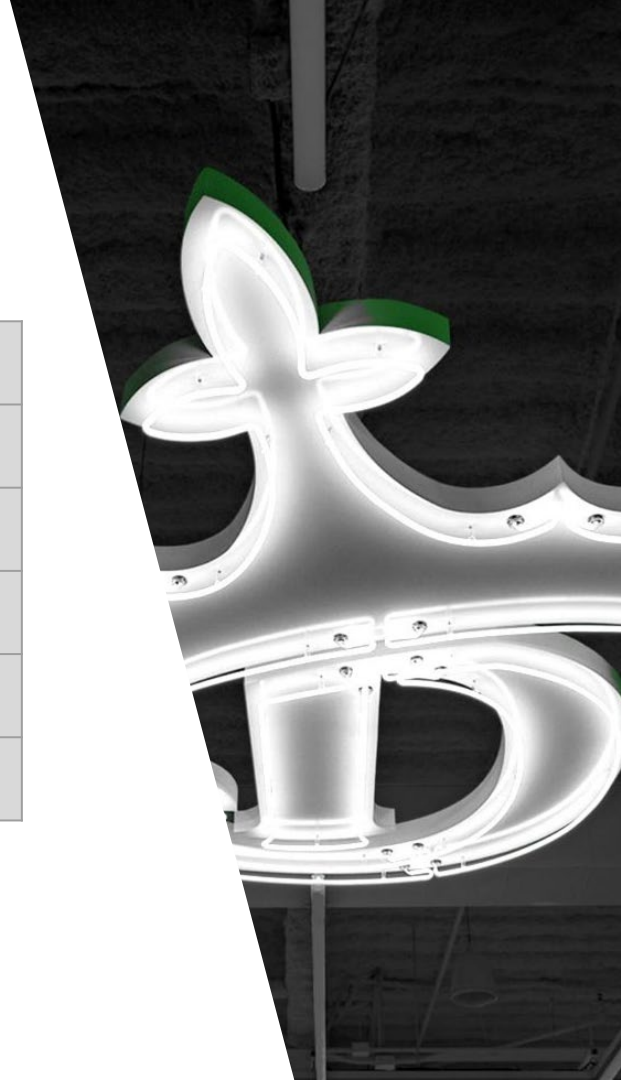
$$\boldsymbol{K}_n = \boldsymbol{P}_{n,n-1}\boldsymbol{H}^T\left(\boldsymbol{H}\boldsymbol{P}_{n,n-1}\boldsymbol{H}^T + \boldsymbol{R}_n\right)^{-1}$$

Covariance Update

$$\boldsymbol{P}_{n,n} = (\boldsymbol{I} - \boldsymbol{K}_n\boldsymbol{H})\,\boldsymbol{P}_{n,n-1}(\boldsymbol{I} - \boldsymbol{K}_n\boldsymbol{H})^T + \boldsymbol{K}_n\boldsymbol{R}_n\boldsymbol{K}_n^T$$

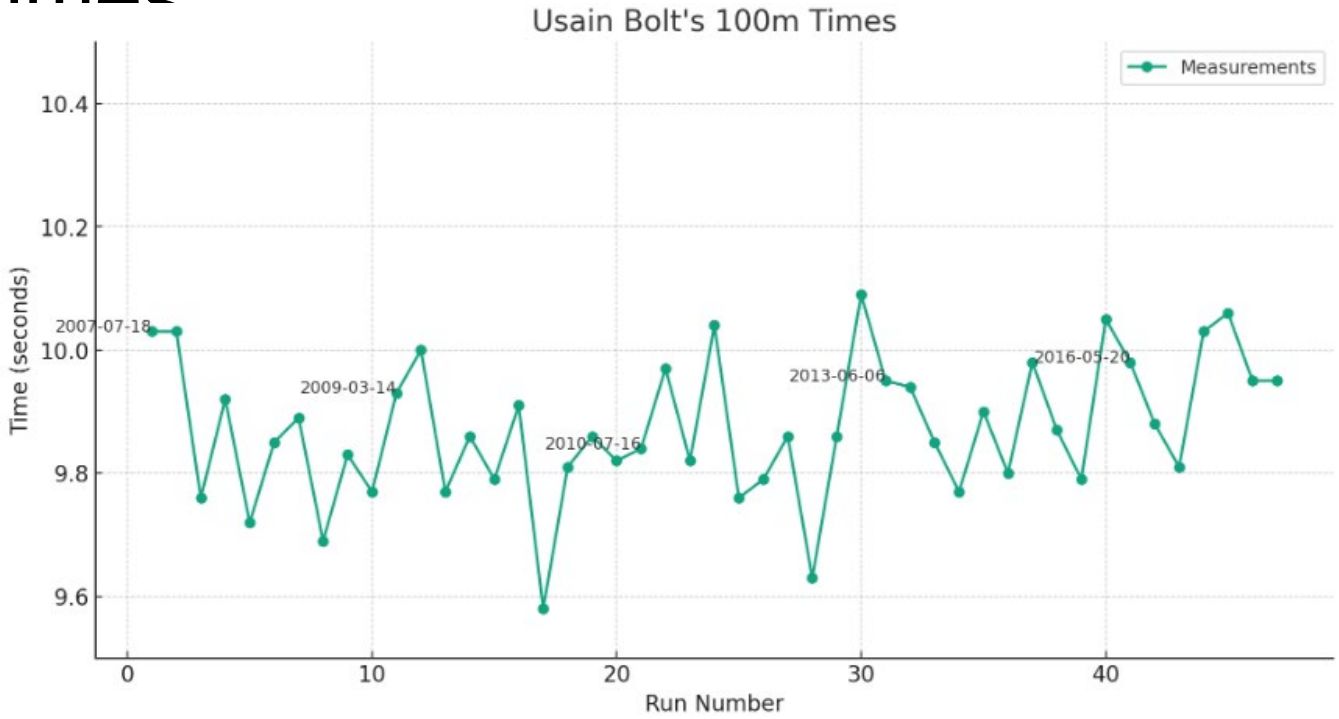
Applications

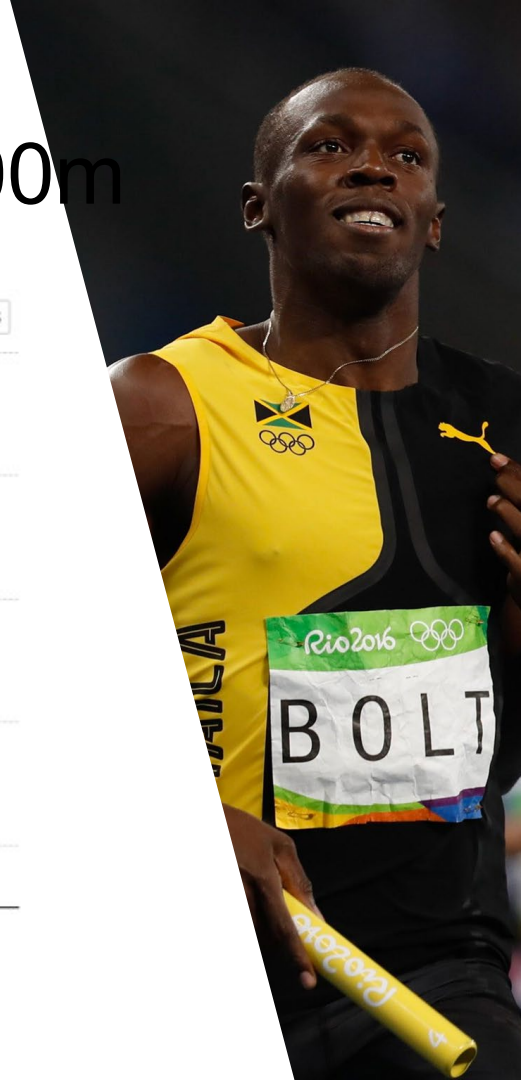
	Auto Pilot
	Portfolio Optimization
	Financial Time Series Prediction
	GPS
	Rocket Science
	And Many More!!!



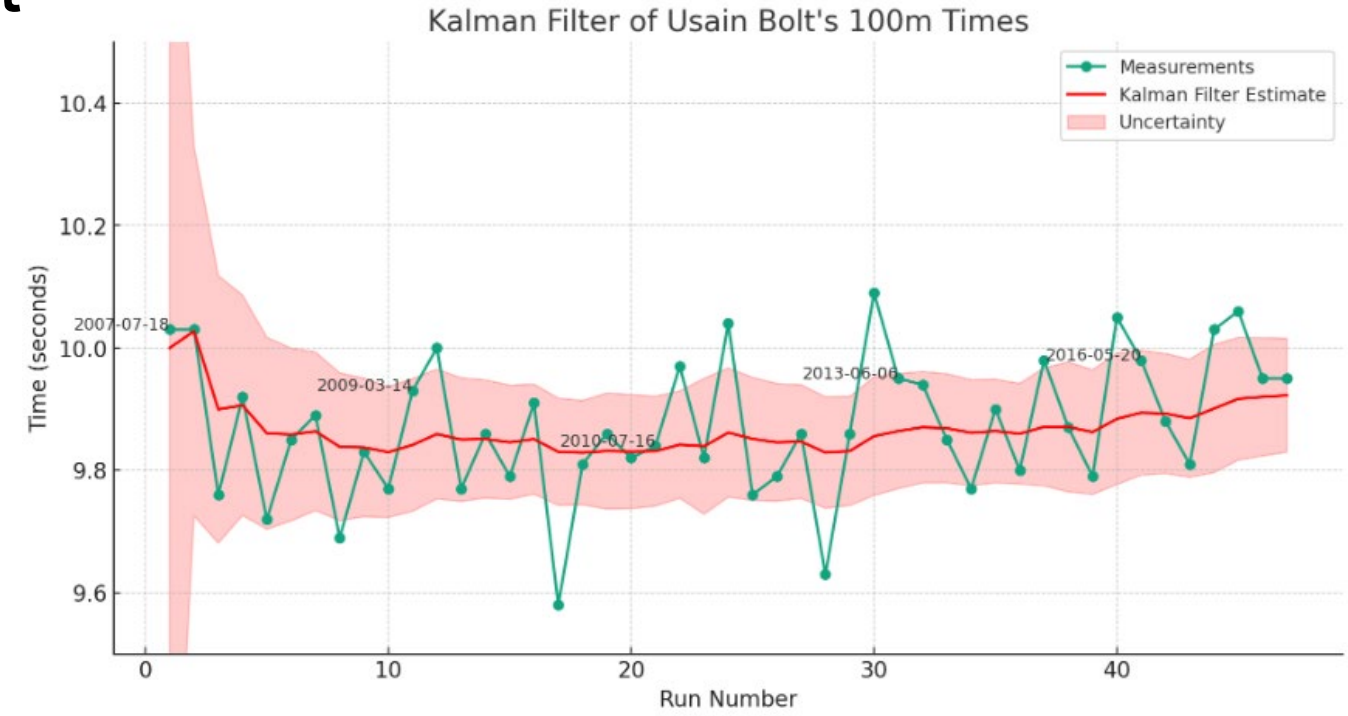
Doesn’t Use Kalman Filters

A Simple Example - Usain Bolt's 100m times



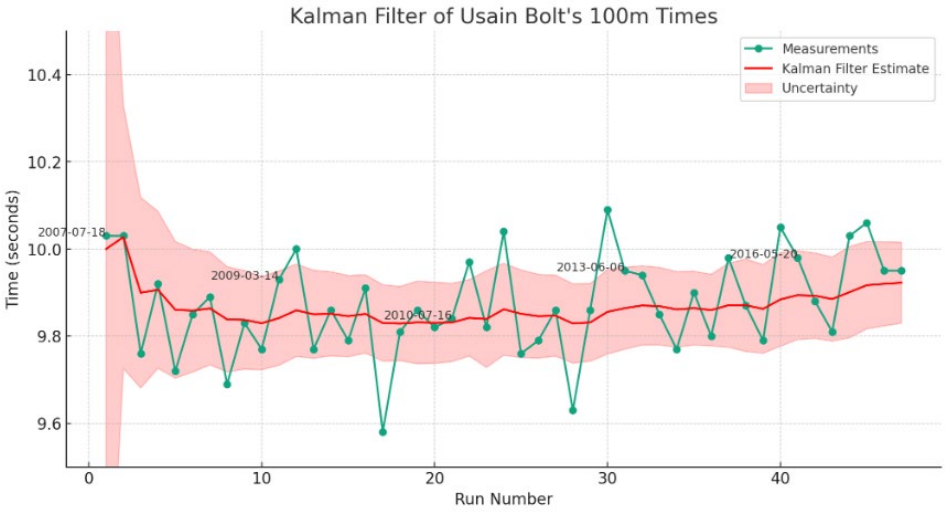


A Simple Example - Usain Bolt's 100m times



A Simple Example - The Parameters

- Starting Expectation
- Starting Uncertainty
- Measurement Noise
- Drift



Kalman Filters as a Rating System

- Kalman Filters can be used for situations where there are multiple players/teams participating, and only a subset of players participate in any given observations
-

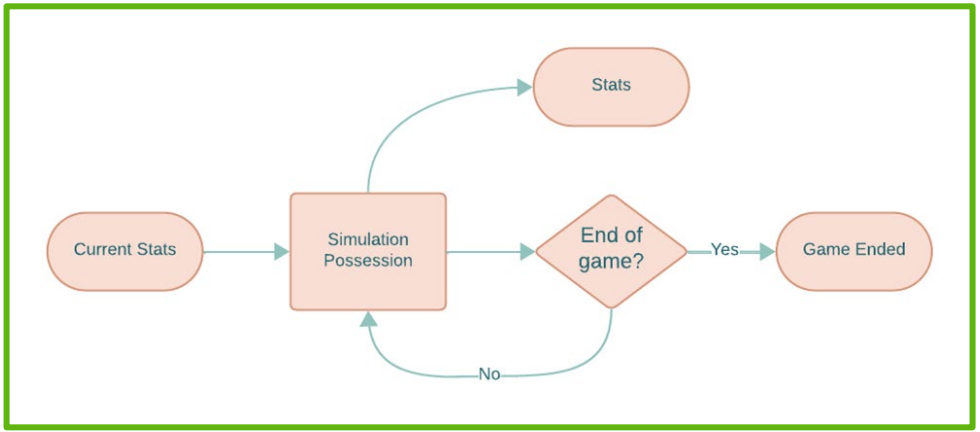
Rating System	Elo	Glicko	Trueskill	Kalman Filters
Uncertainty Value	No	Yes	Yes	Yes
Measurement Noise	No	No	Yes	Yes
Updates	Only involved players	Only involved players	Only involved players	All entities/players

DRAFT KINGS | THE CROWN IS YOURS

3.0 Our NBA Kalman Filters

Simulation Engine

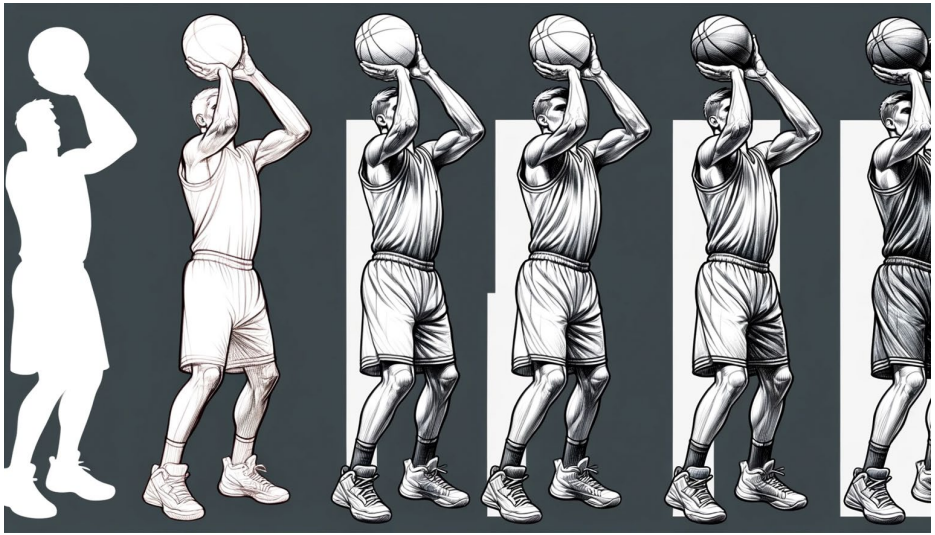
At DraftKings, our sports engines are built using **Monte Carlo Simulation**



Enrichment Probability Models

During possession simulation, details are enriched sequentially based on probability models.

- Given a three pointer made, who has shoot the three?
- Is the three pointer being assisted?
- If yes, who has passed the ball?



Kalman Filters as Probability Model

Simple

$$P(X) = \frac{1}{1 + e^{-\mu}}$$

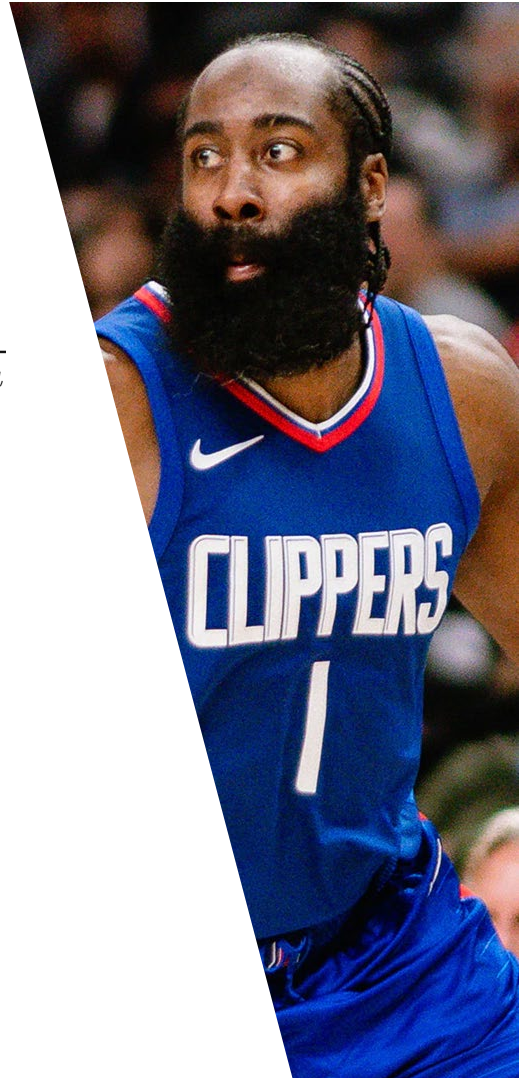
- A rating for every team/player
- Simple transformation from ratings to probabilities

Team / Player Specific

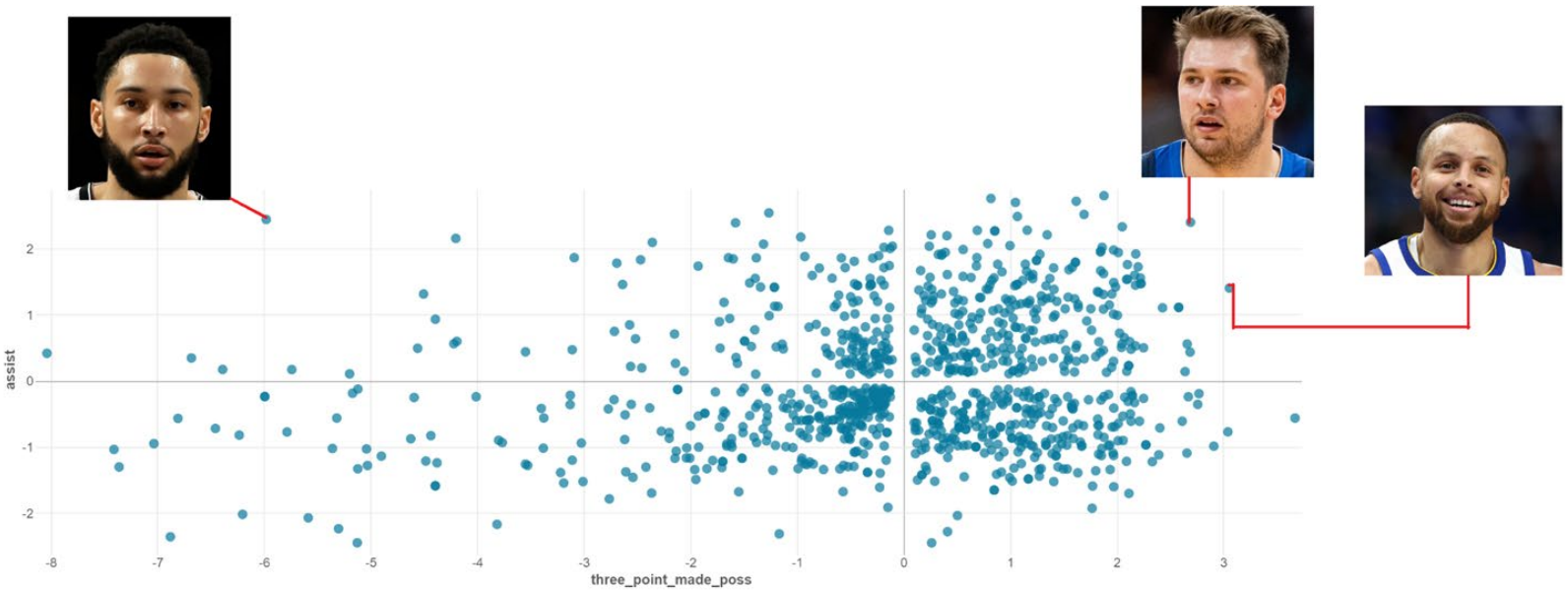
- Do not need additional features

Easy to update

- Simple matrix operations to update the ratings after the game



Three Pointer Ratings vs Assists Ratings



DRAFT KINGS | THE CROWN IS YOURS

04 How We're Utilizing Databricks

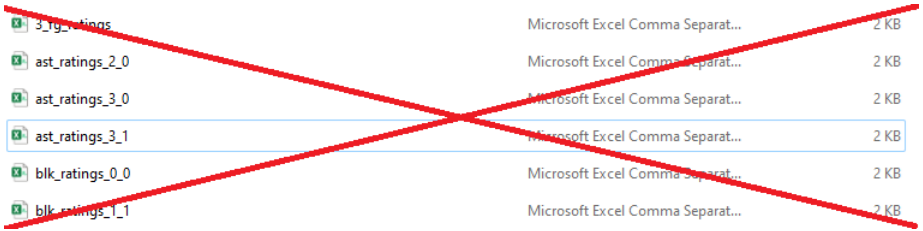
“The most valuable commodity I know of is information”

– A quote from 1987 film Wall Street

Before & After Databricks

Before

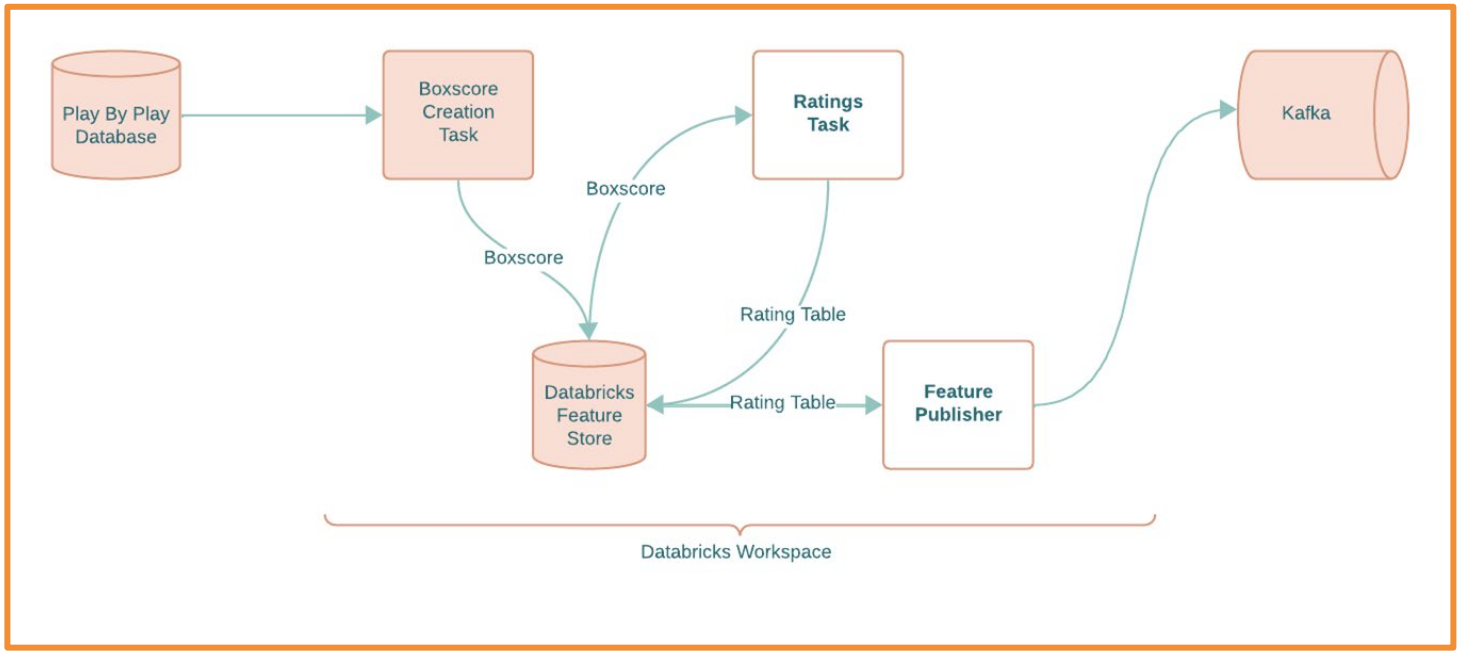
- Ratings are trained locally and stored in a csv
- Need to do a deployment to update the ratings
- Basically only updates every year
- Not feasible for fast moving player ratings



After!

- Workflow are set up in the databricks to update dataset and ratings daily
- Dataset and ratings can be reproduced easily
- The ratings for calculation is sent downstream automatically
- Able to backtest the model performance using historical ratings

Architecture



Rating Tasks

Feature Publisher

The Feature Publisher task reads the CDC (change data capture) table

- Figure out which rows have been changed

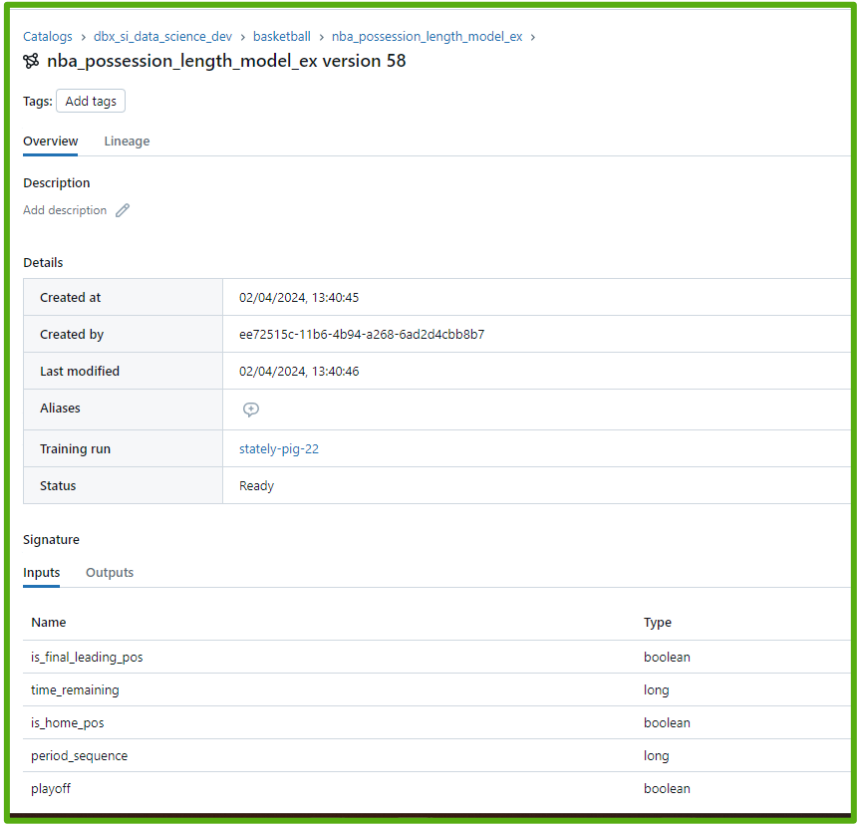
The change is then sent downstream to kafka

- The updated team/player ratings are mapped to different games
- Form calculation requests by combing ratings with other information

Model Injection

The models are trained on the updated dataset by scheduled workflows.

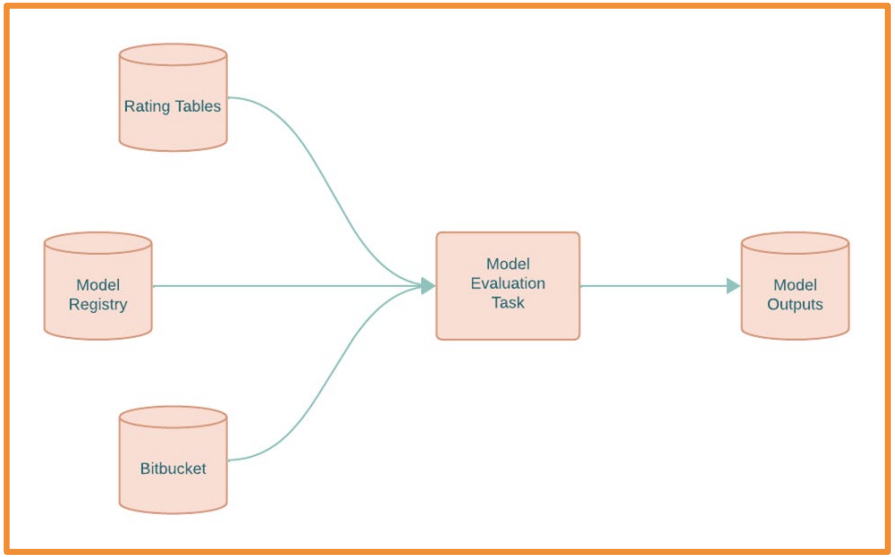
When the engine pod is started up, it will grab the latest version of the model from the model registry.



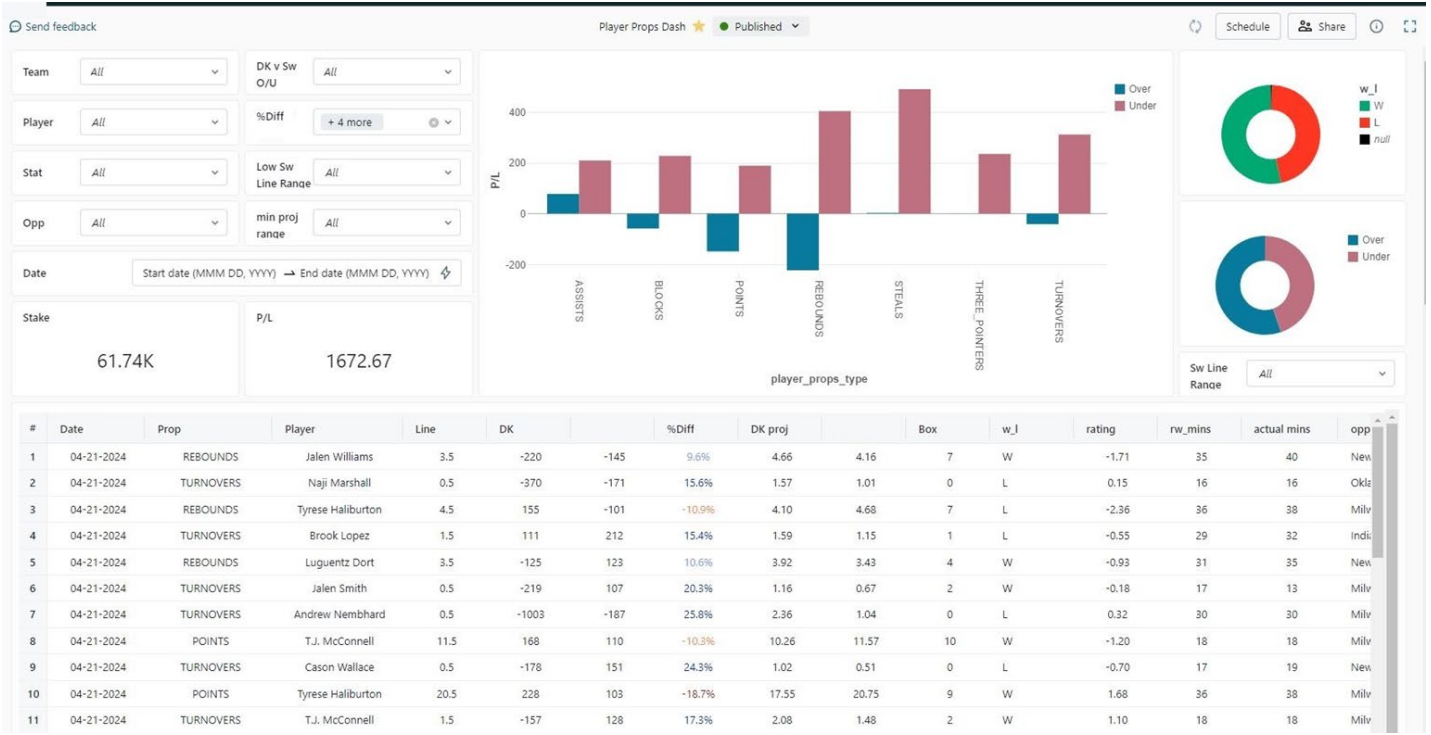
Model Backtesting

Historical KF ratings

- **Historical kf team/player ratings right before the game starts**
- **Avoid data leakage**



Historical Dashboard



#	Date	Prop	Player	Line	DK		%Diff	DK proj		Box	w_l	rating	rw_mins	actual mins	opp
1	04-21-2024	REBOUNDS	Jalen Williams	3.5	-220	-145	9.6%	4.66	4.16	7	W	-1.71	35	40	New
2	04-21-2024	TURNOVERS	Naji Marshall	0.5	-370	-171	15.6%	1.57	1.01	0	L	0.15	16	16	Okla
3	04-21-2024	REBOUNDS	Tyrese Haliburton	4.5	155	-101	-10.9%	4.10	4.68	7	L	-2.36	36	38	Milv
4	04-21-2024	TURNOVERS	Brook Lopez	1.5	111	212	15.4%	1.59	1.15	1	L	-0.55	29	32	Indi
5	04-21-2024	REBOUNDS	Luguentz Dort	3.5	-125	123	10.6%	3.92	3.43	4	W	-0.93	31	35	New
6	04-21-2024	TURNOVERS	Jalen Smith	0.5	-219	107	20.3%	1.16	0.67	2	W	-0.18	17	13	Milv
7	04-21-2024	TURNOVERS	Andrew Nembhard	0.5	-1003	-187	25.8%	2.36	1.04	0	L	0.32	30	30	Milv
8	04-21-2024	POINTS	T.J. McConnell	11.5	168	110	-10.3%	10.26	11.57	10	W	-1.20	18	18	Milv
9	04-21-2024	TURNOVERS	Cason Wallace	0.5	-178	151	24.3%	1.02	0.51	0	L	-0.70	17	19	New
10	04-21-2024	POINTS	Tyrese Haliburton	20.5	228	103	-18.7%	17.55	20.75	9	W	1.68	36	38	Milv
11	04-21-2024	TURNOVERS	T.J. McConnell	1.5	-157	128	17.3%	2.08	1.48	2	W	1.10	18	18	Milv

DRAFT KINGS | THE CROWN IS YOURS

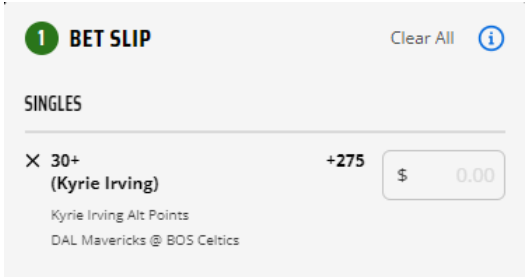
50 How You Make Money

Betting Strategy

NOW LET'S MAKE SOME MONEY

WINNING INGREDIENTS

- Bookmaker offers to return **3.75** times your stake.
- Model suggests **30%** of the time you will win.
- Expected return is 30% x 3.75 = **1.125**

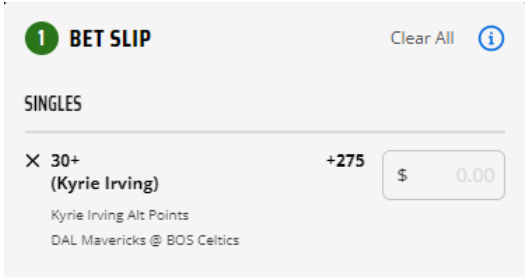


Betting Strategy

NOW LET'S MAKE SOME MONEY

KELLY'S CRITERION

- Maximize the expected growth rate.
- Everytime we only bet a proportion of the bankroll
- ***b = (pO - 1) / (O - 1)***

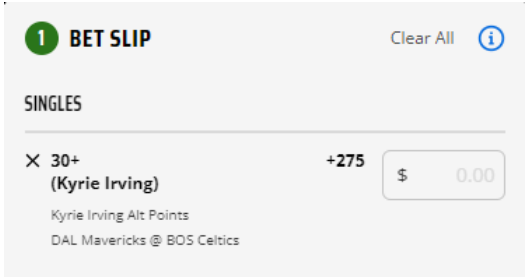


Betting Strategy

NOW LET'S MAKE SOME MONEY

EVALUATION IS KEY

- Analyse on the betting performance to optimise the betting strategy
- You only need to bet when you are confident with your prices



Sportsbook Strategy

We don't have the luxury of choosing when we bet

- Need to be confident for **ALL** of the markets that we offer

Apply Margin

- We can control our prices

Learn from customer's bet

- Identify professional customers
- Adjust the probability of the model based on their bets

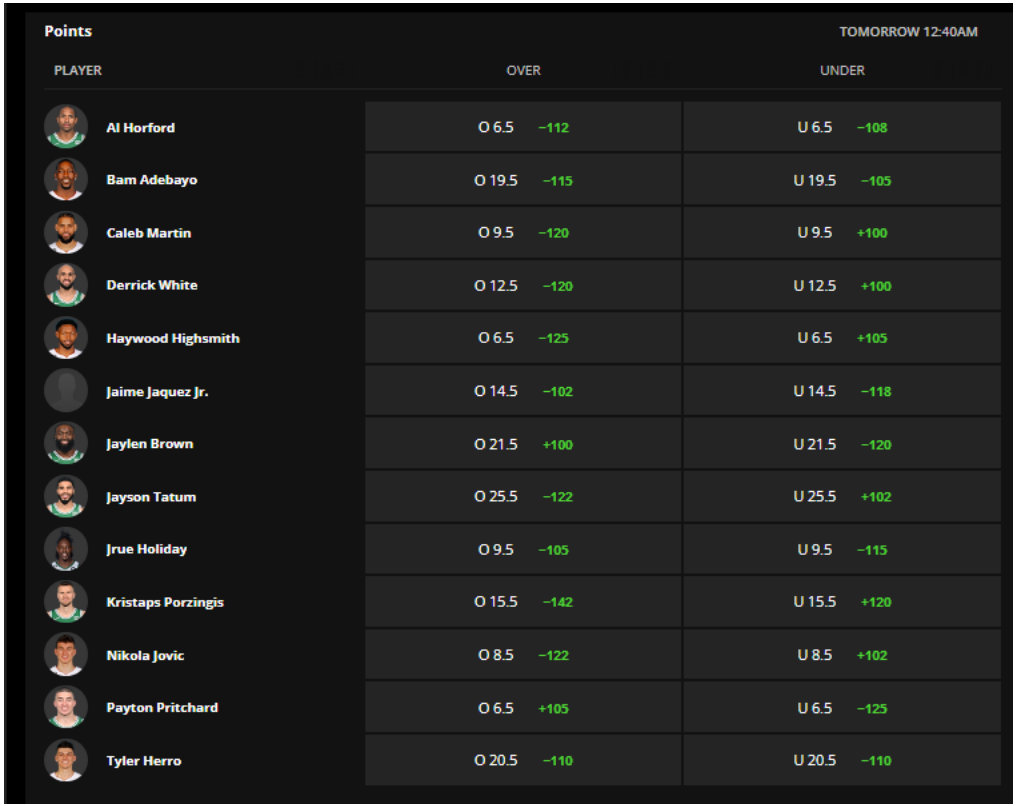
DRAFT KINGS | THE CROWN IS YOURS

60 Summary

Summary

- We achieved our goal to build a model that is able to produce NBA player props
- We produced 200+ separate Kalman FIlter based models all producing ratings for various different player and team attributes.
- All of the above models use features engineered through Databricks workflows
- All models are registered on Databricks and retrained and optimized through workflows running on a daily cadence
- The same practices applied here are now being adopted by other sports as they look to do similar.

https://medium.com/draftkings-engineering



DRAFT KINGS
THE CROWN IS YOURS

THANK YOU!
Any Questions?