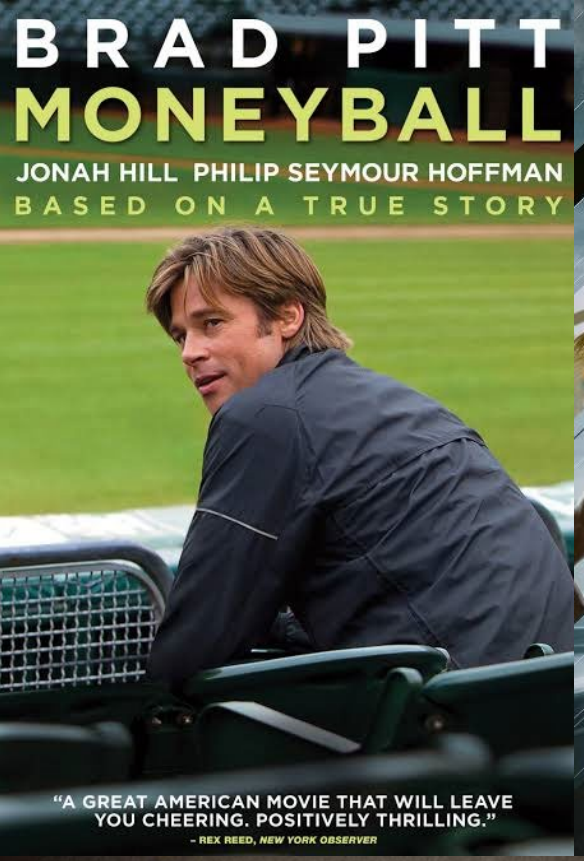
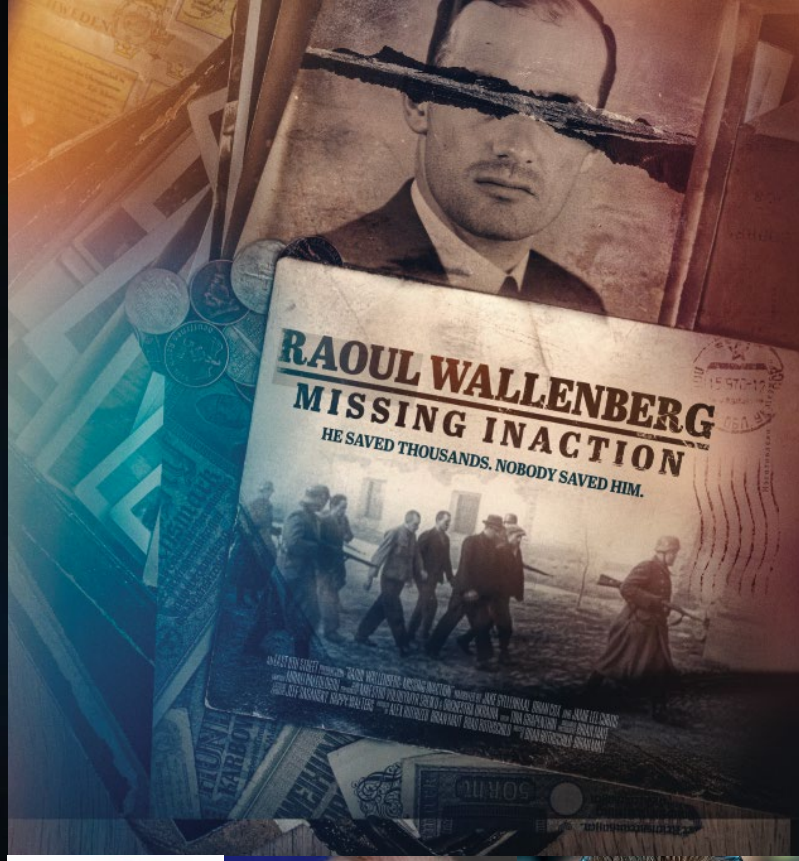
The Moneyball Mindset: Using Data and AI to Drive Smarter Decisions and Real Results

Ari Kaplan
Data + AI Leader of the Year 2025

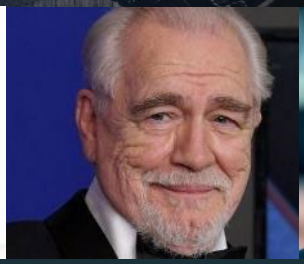

BRAD PITT
MONEYBALL
JONAH HILL PHILIP SEYMOUR HOFFMAN
BASED ON A TRUE STORY
"A GREAT AMERICAN MOVIE THAT WILL LEAVE YOU CHEERING. POSITIVELY THRILLING."
– REX REED, NEW YORK OBSERVER


FROM THE DIRECTOR OF TOP GUN: MAVERICK
BRAD PITT
DAMSON IDRIS
KERRY CONDON
JAVIER BARDEM
F1
THE MOVIE
ONLY IN THEATERS JUNE 27
FILMED FOR IMAX

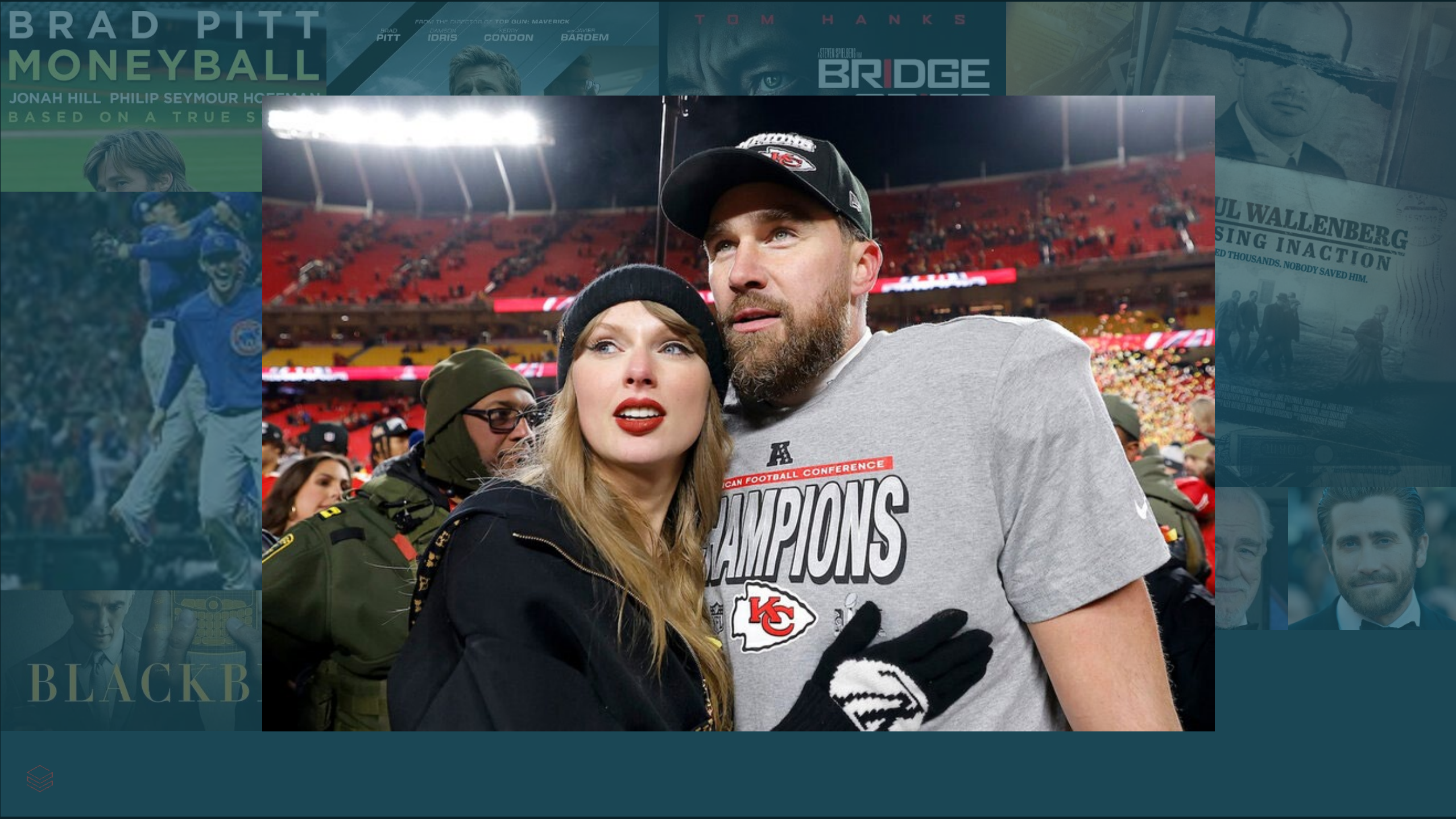

TOM HANKS
BRIDGE OF SPIES
INSPIRED BY TRUE EVENTS
6 OSCAR NOMINATIONS INCLUDING BEST PICTURE
"MAGNIFICENT COLD WAR THRILLER"
THE GUARDIAN
TOTAL FILM
HENRY FITZHERBERT - SUNDAY EXPRESS
12
12

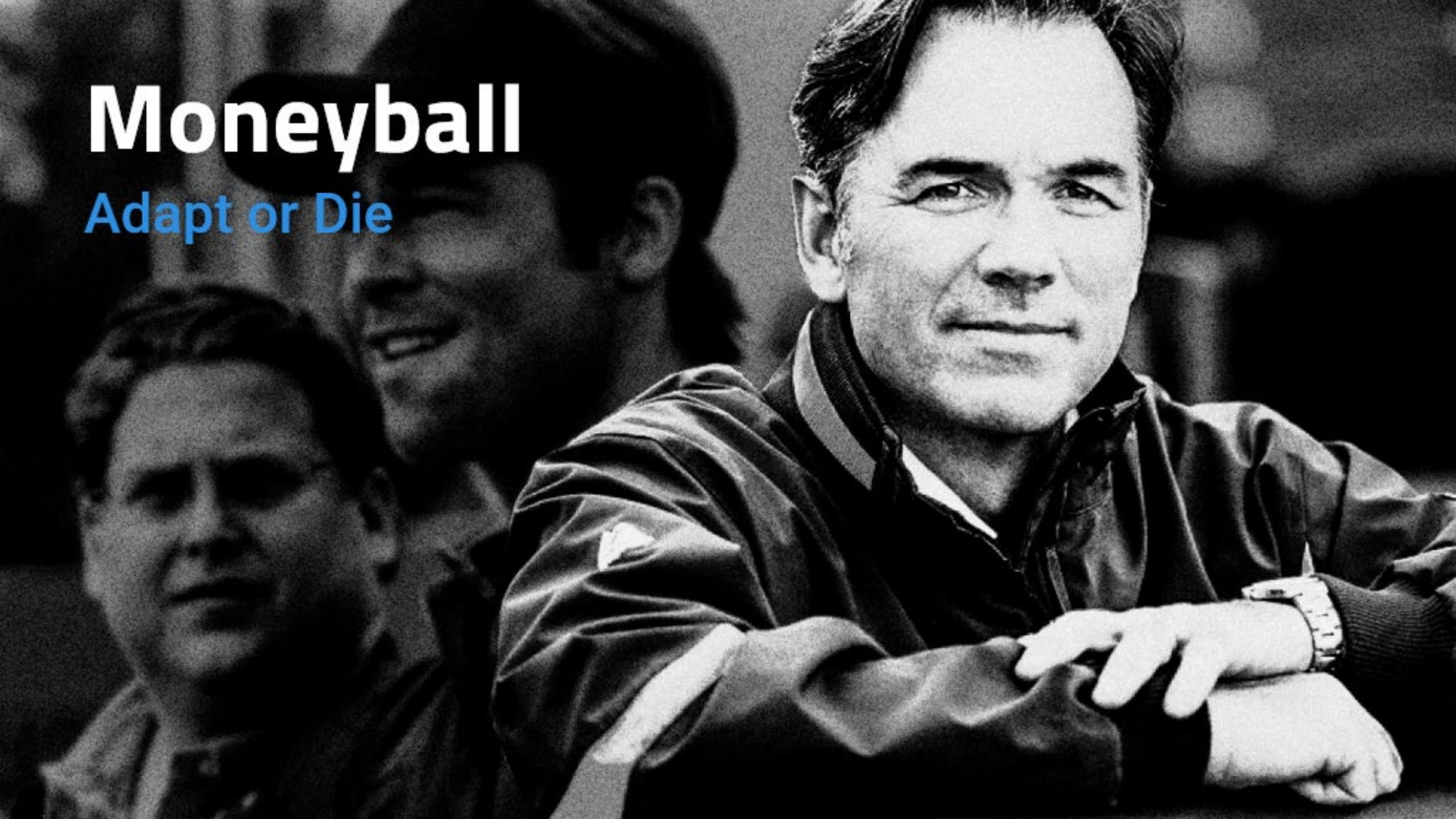

RAOUL WALLENBERG
MISSING INACTION
HE SAVED THOUSANDS. NOBODY SAVED HIM.


BLACKBERRY



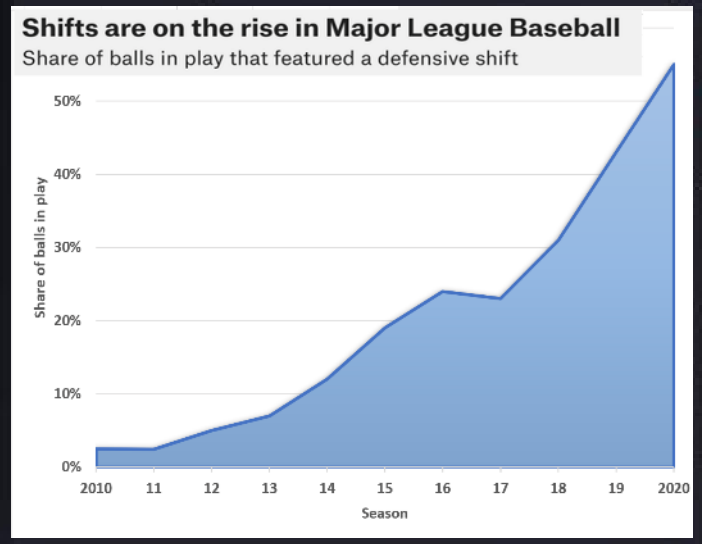

44

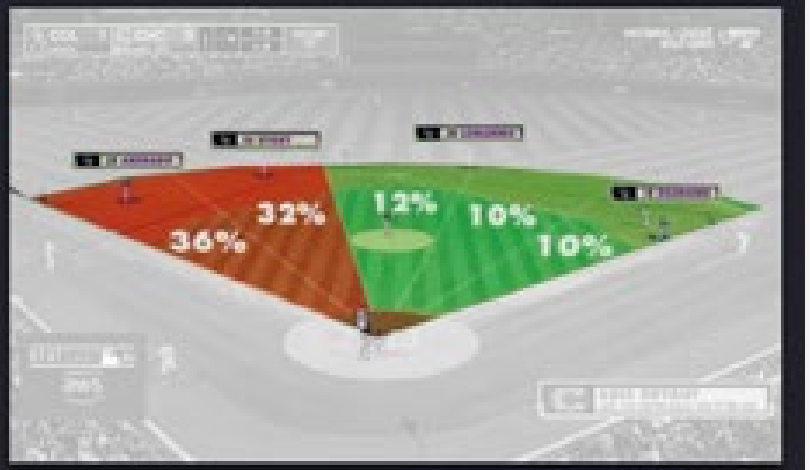

BRAD PITT
MONEYBALL
JONAH HILL PHILIP SEYMOUR HOFFMAN
BASED ON A TRUE S
FROM THE DIRECTOR OF TOP GUN: MAVERICK
BRAD PITT
DAMSON IDRIS
KERRY CONDON
JAVIER BARDEM
TOM HANKS
BRIDGE
UL WALLENBERG
SING INACTION
ED THOUSANDS. NOBODY SAVED HIM.
BLACKB
AMPIONS

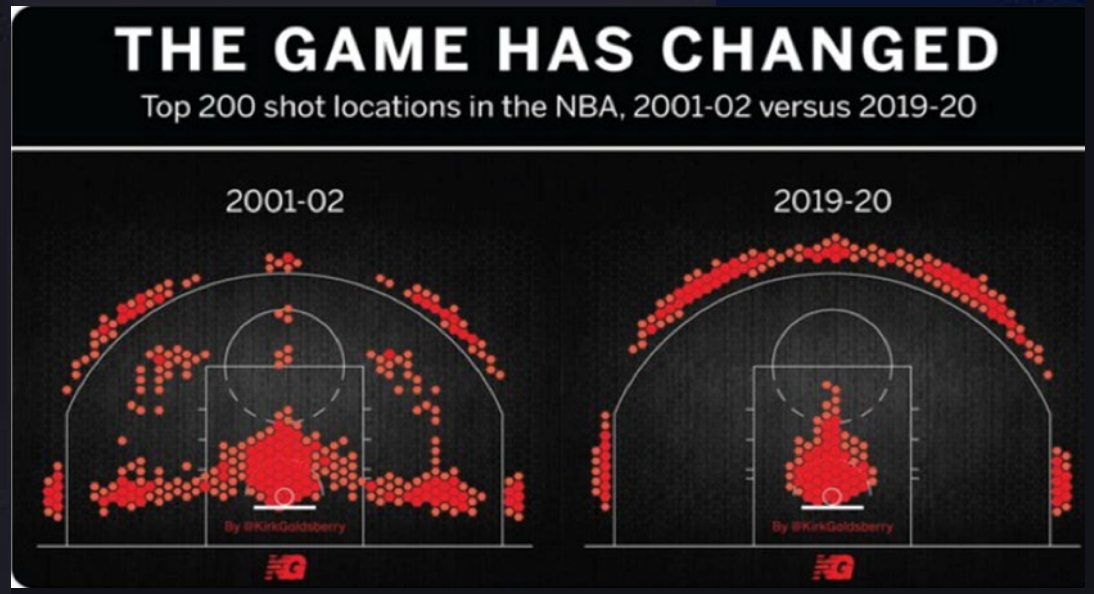

Moneyball
Adapt or Die

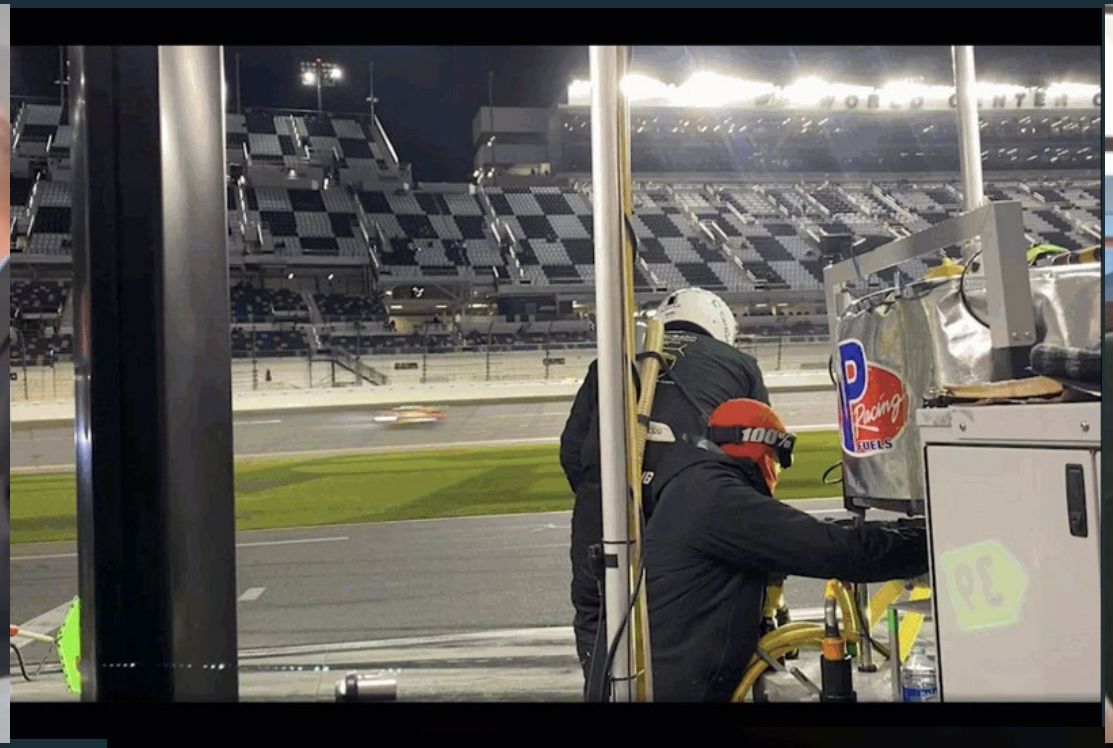
AI rapidly changed how the game is played

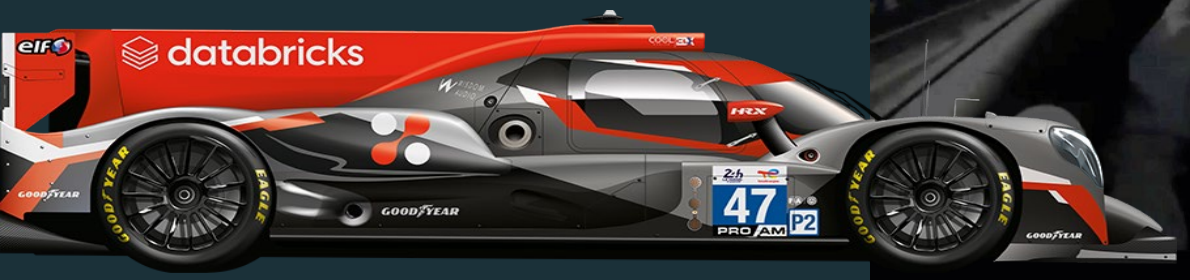
Highest strikeout rate in history, **Longer** games
Most defensive shifts










databricks
47
PRO AM
P2


HOT LAPS

Formula 1

- 12 billion data points / season
- 80,000 car components
- 300 sensors, 1000x per second
- 378 million car simulations
- 2 second pit stops

Data & AI Use Cases

Above the Field	On the Field	Business Operations
Player value forecasting	In-Game Strategy	Fan/customer 360
Economics of contracts	Amateur Scouting	Ticket Pricing and retention
Roster modeling	International Scouting	Marketing Strategy
Trade-deadline analysis	Pro-Scouting	Customer Segmentation
Amateur Drafts	Player Development	Concession / management
	Injury mitigation	

Apply analytical models again and again to get answers

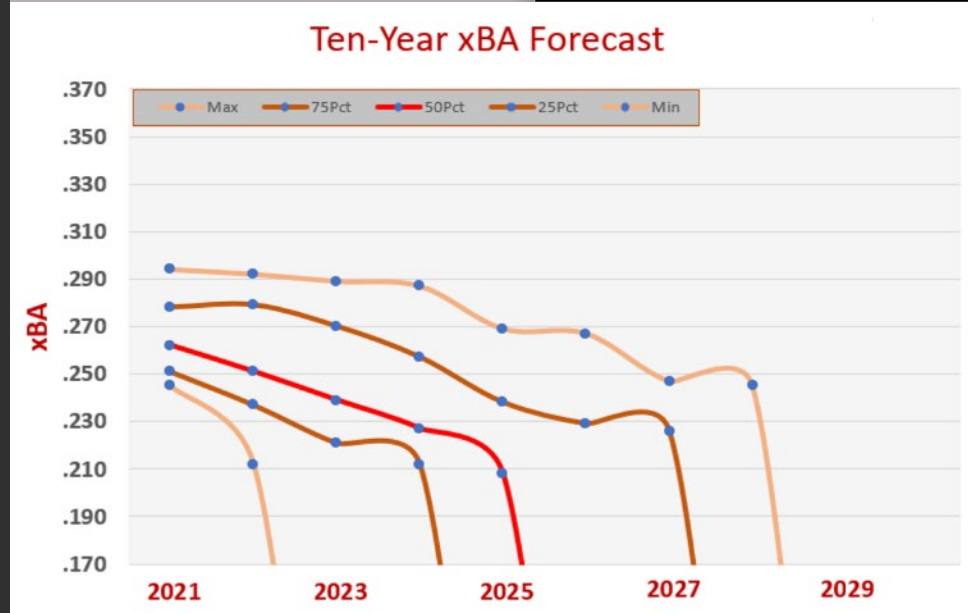
Understand and manage risks

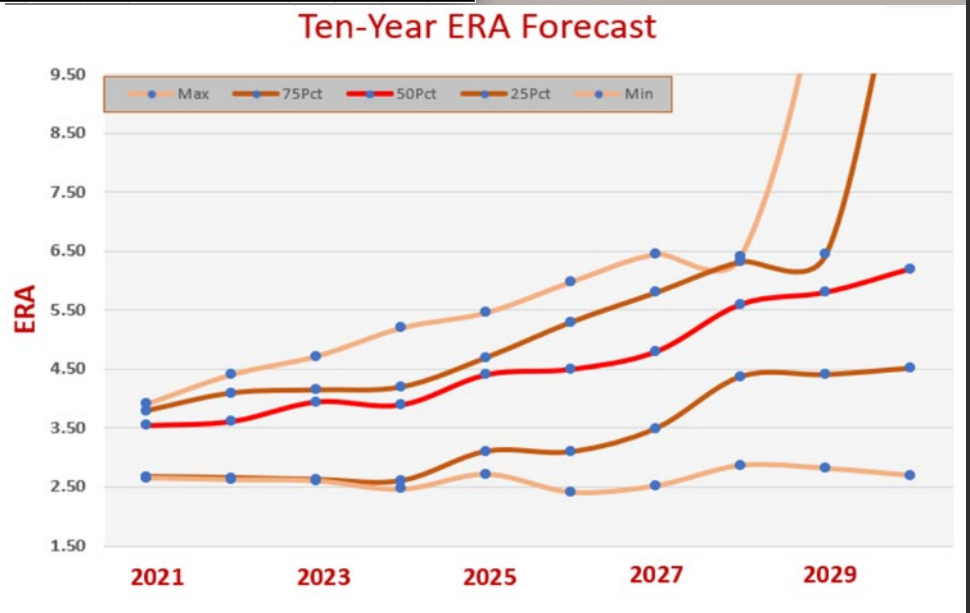
MGR [redacted]

Pos	Player	Age	AB	OBP	SLG	Bat RC
1-SS	[redacted]	21	575	0.359	0.428	85
2-CF	[redacted]	33	475	0.339	0.428	67
3-1B	[redacted]	33	475	0.354	0.473	76
4-3B	[redacted]	33	475	0.342	0.498	78
5-LF	[redacted]	35	450	0.317	0.474	67
6-RF	[redacted]	25	425	0.307	0.471	61
7-C	[redacted]	28	425	0.373	0.486	72
8-2B	[redacted]	25	425	0.327	0.372	51
9-P	pitchers		300	0.170	0.159	1
OF	[redacted]	34	400	0.367	0.431	61
UT	[redacted]	29	325	0.336	0.448	48
C	[redacted]	31	200	0.287	0.343	20
IF	[redacted]	25	175	0.294	0.291	15
OF	[redacted]	29	125	0.368	0.344	16
	others		250	0.285	0.325	23
	totals		5500			740

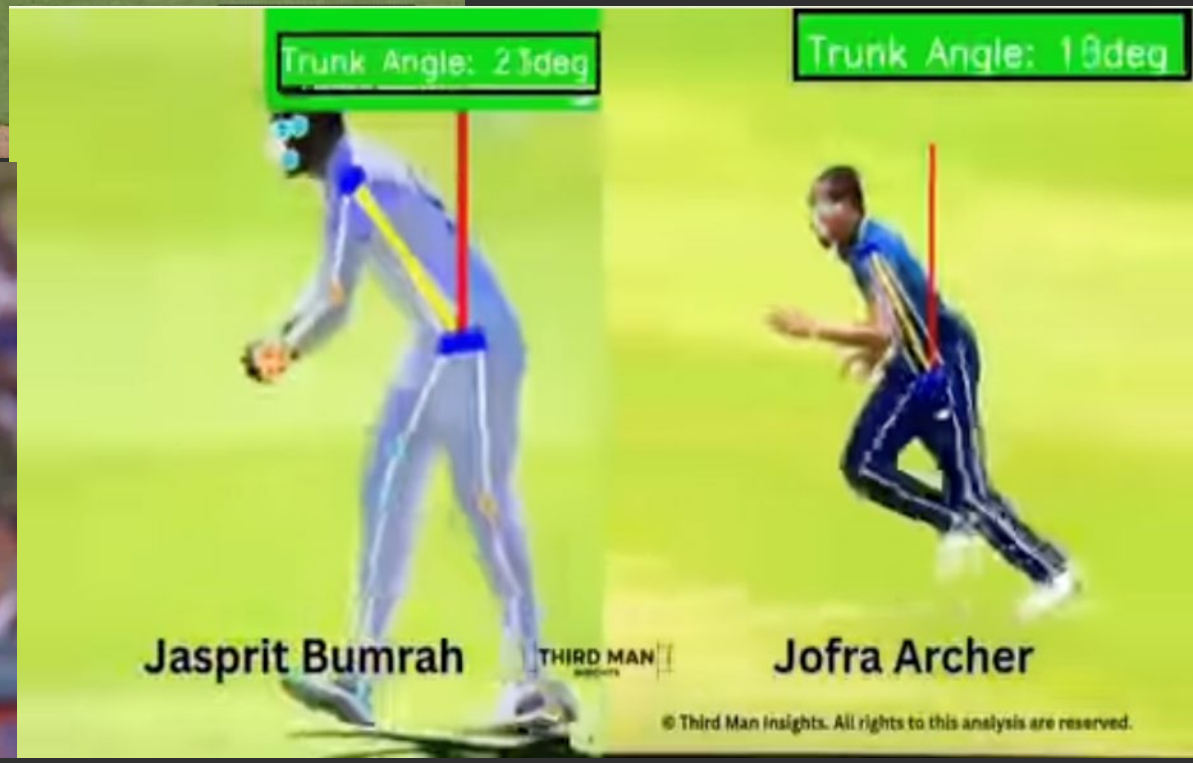
Pos	Pitcher	Age	IP	ERA	RA
SP	[redacted]	30	221	3.67	**98**
SP	[redacted]	34	214	3.83	**99**
SP	[redacted]	28	201	4.12	**100**
SP	[redacted]	32	133	4.80	**77**
SP	[redacted]	28	133	4.06	**65**
S/R	[redacted]	24	114	5.03	**69**
CL	[redacted]	28	83	3.14	**31**
RP	[redacted]	28	68	3.84	**31**
RP	[redacted]	34	62	3.48	**26**
RP	[redacted]	32	50	4.28	**26**
RP	[redacted]	27	45	5.00	**27**
RP	[redacted]	26	34	5.03	**21**
RP	[redacted]	25	22	5.71	**15**
	others		60	5.63	**41**
	totals		1440		**726**

Forecast WPctg:	0.51
Forecast Wins:	83
Forecast Losses:	79

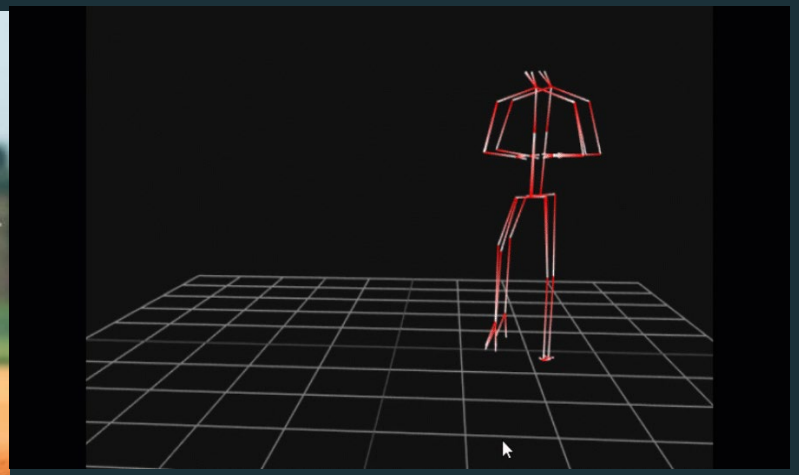


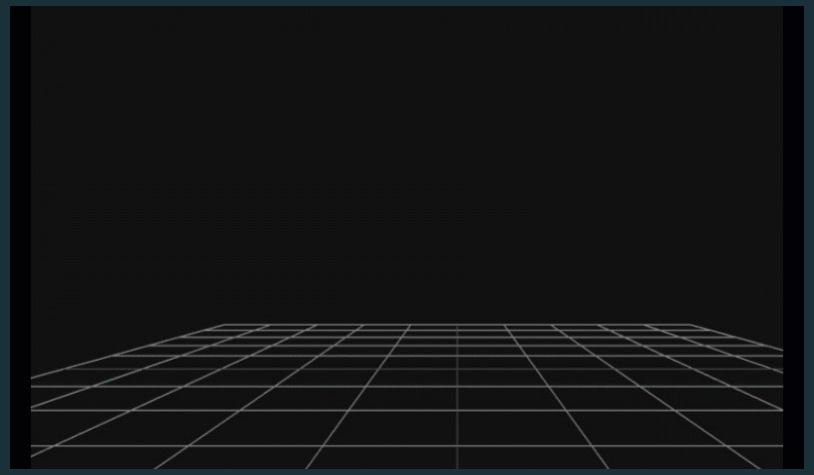


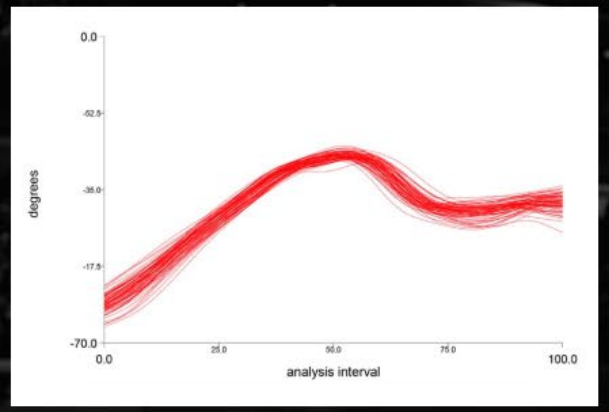
Raw Data to Insights in Any Industry

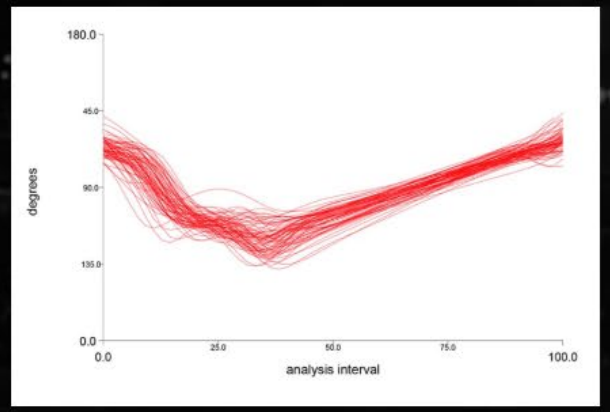


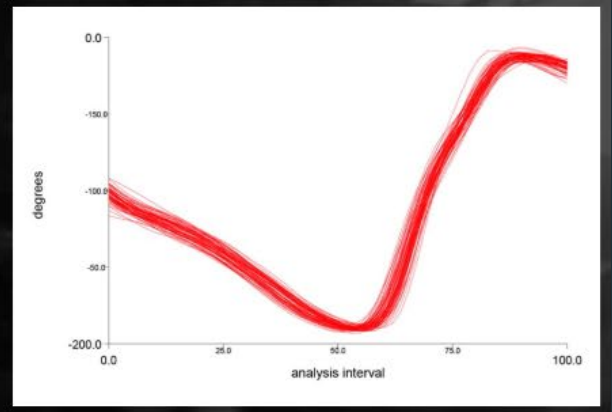
KINATRAX











Leading Leg Flexion

Throwing Forearm Pronation

Internal/External Shoulder Rotation

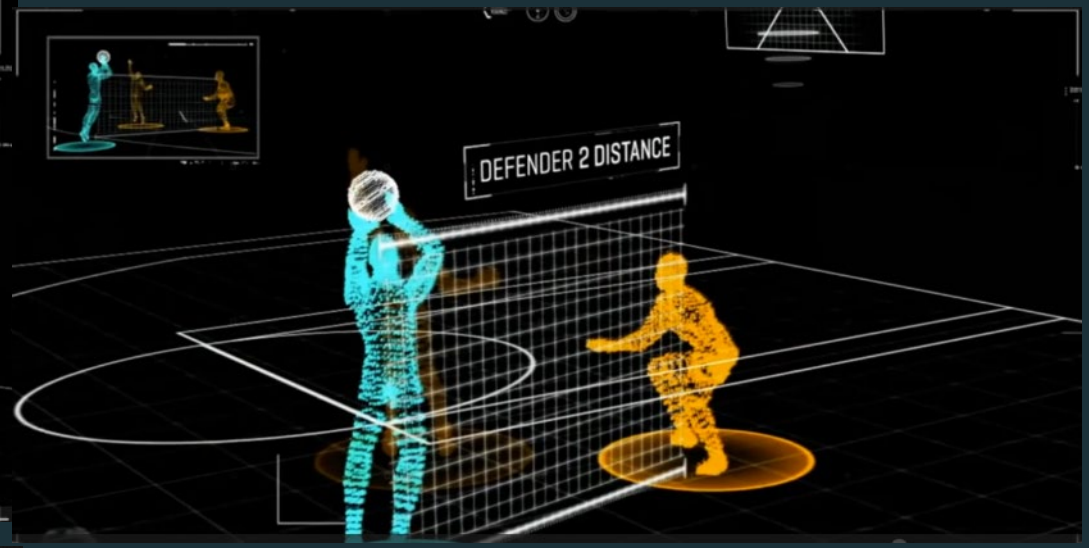
Opportunity: Feature Engineering

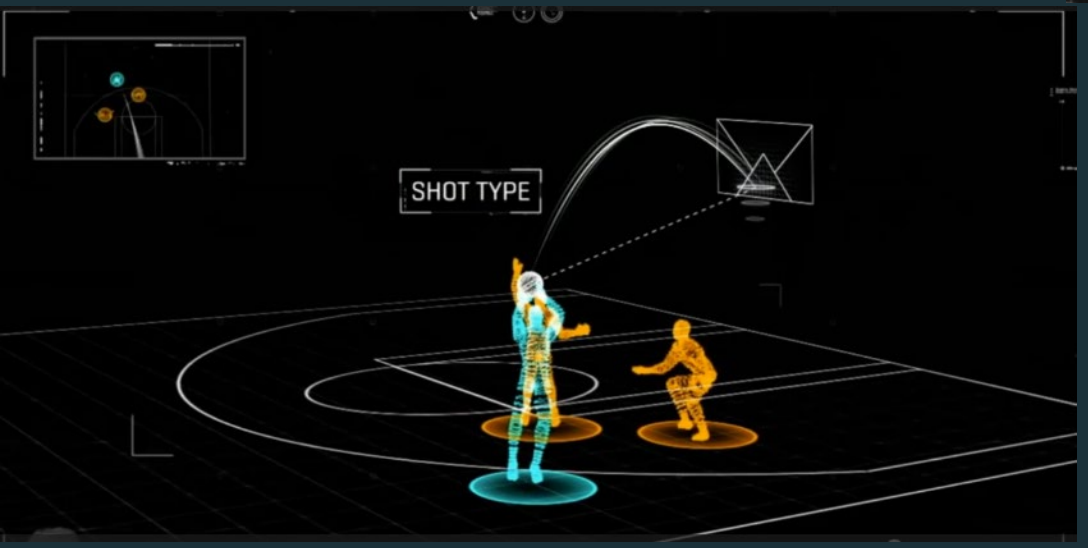
- **Primitives**: dribble, shot, rebound, possession, turnover
- **Handoffs offense:** middle, wing, step_up, dribble_handoff, handback
- **Simple field goal attempts**: heave, angle_layup, driving_layup, jumper, post
- **Drives start:** handoff, pick, isolation, closeout

Images: Second Spectrum

Opportunity: Feature Engineering

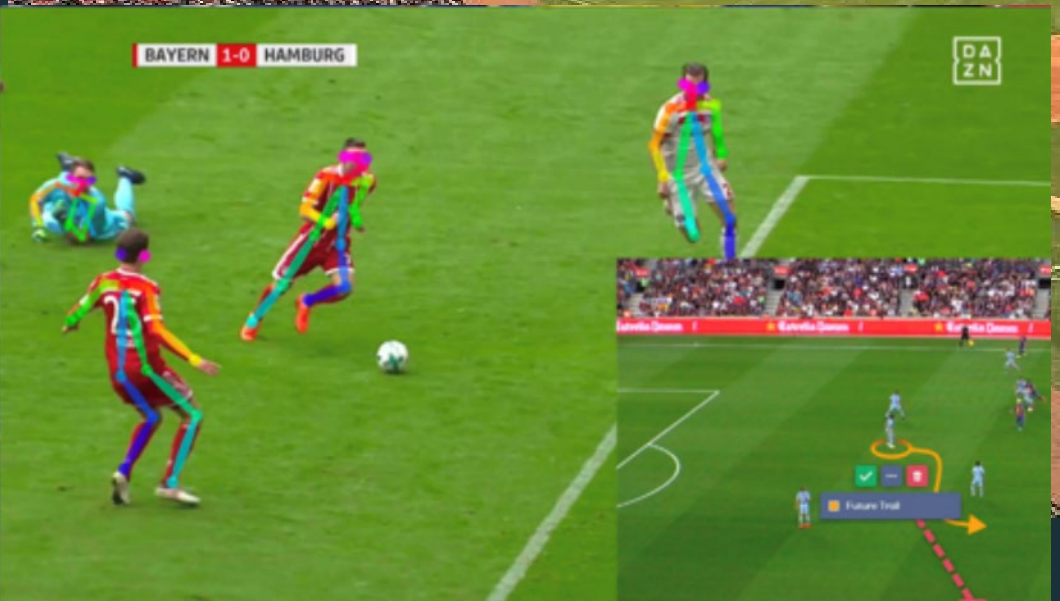
- **Complex field goal attempts:** lob, tip, cut_layup, standstill_layup, driving_layup, catch_shoot, shake_raise, over_Screen, pull_up, step_back, post
- **Picks offense**: pick, reject, slip, pop, roll, wing, middle
- **Drives result**: shot_near_basket, pullup, interior pass, kickout, pullout, turnover, stoppage, blowby

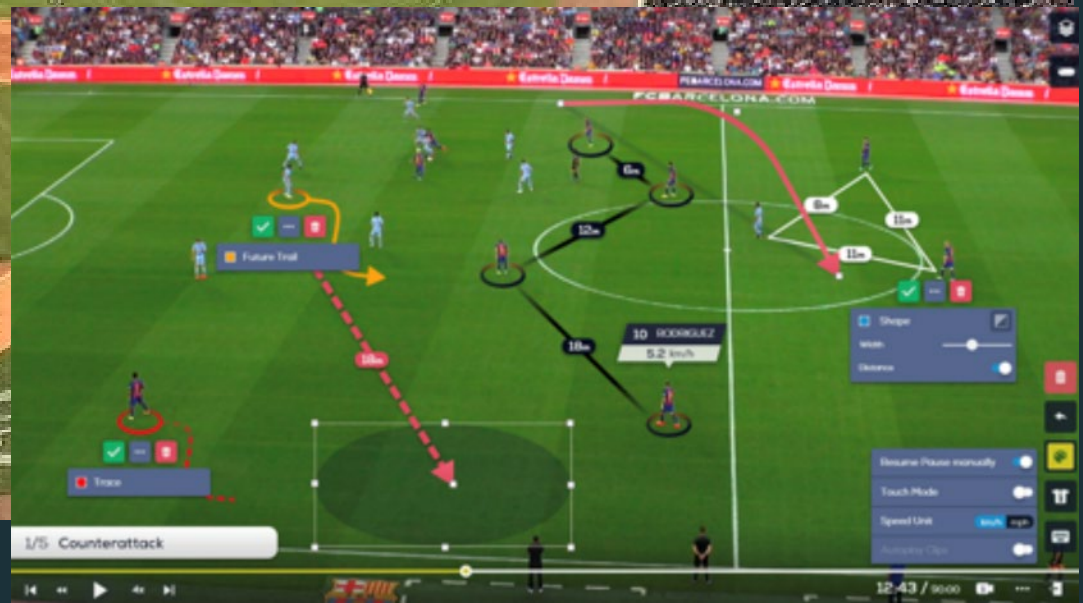




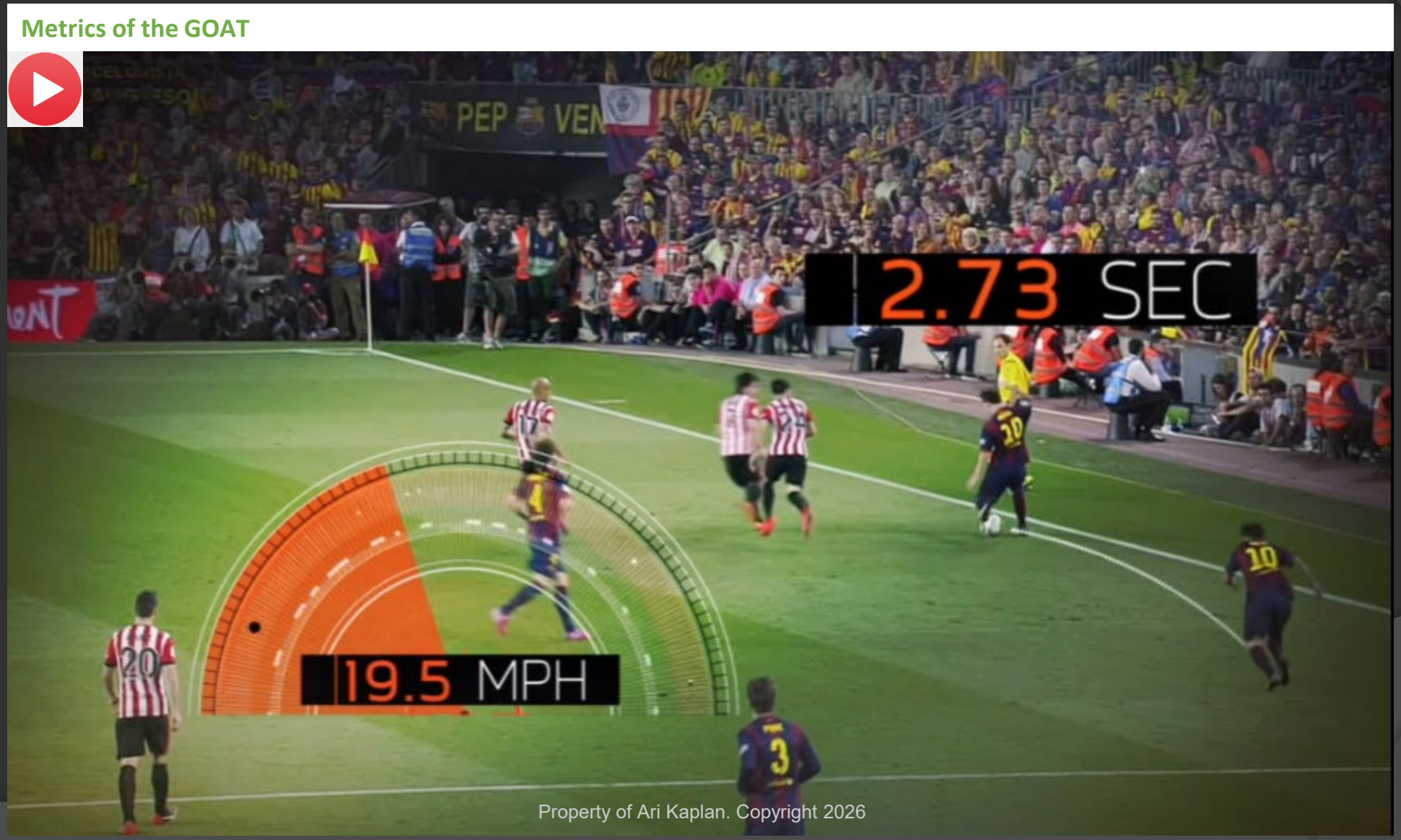
Images: Second Spectrum

Opportunity: Feature Engineering

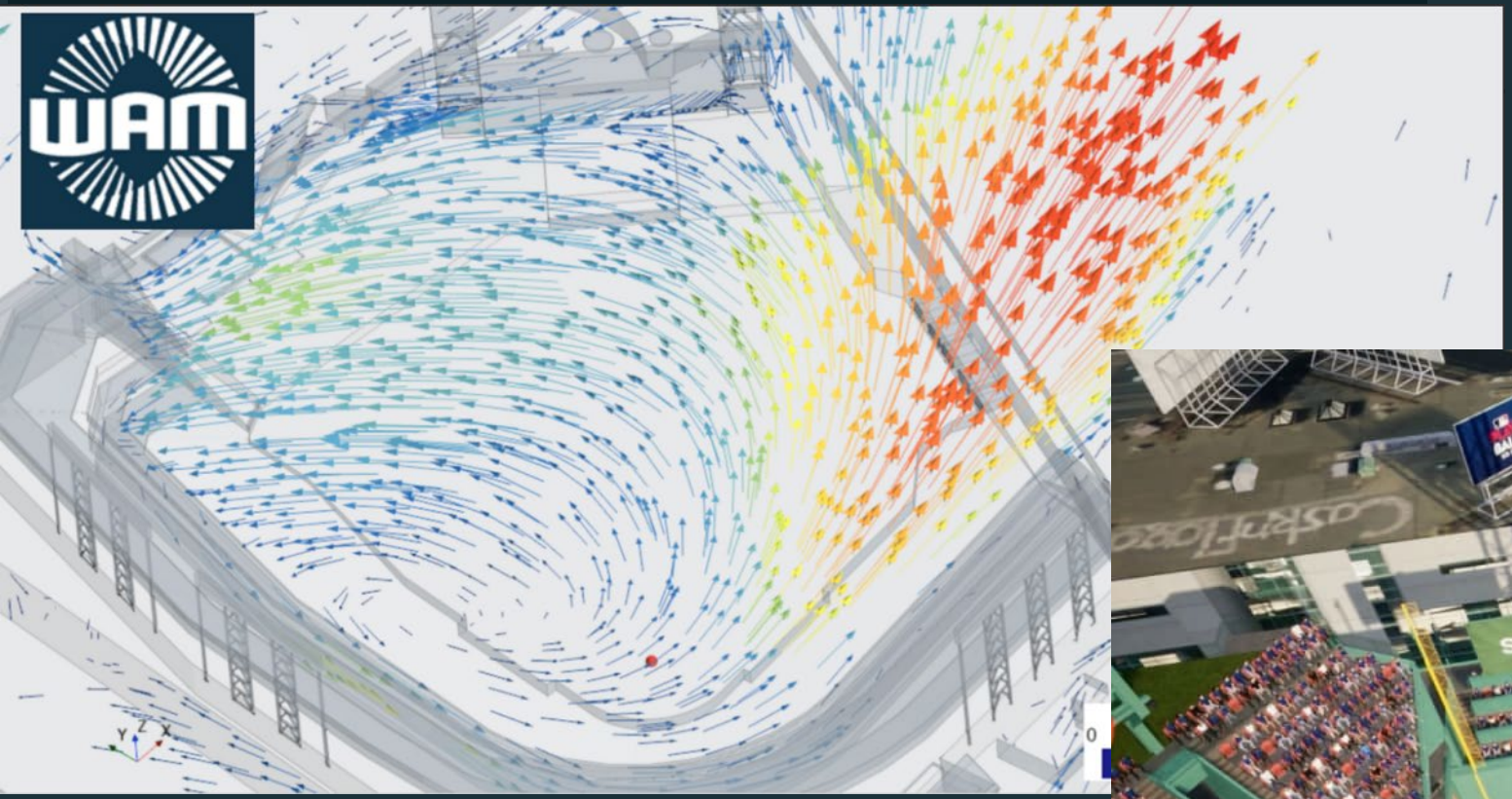




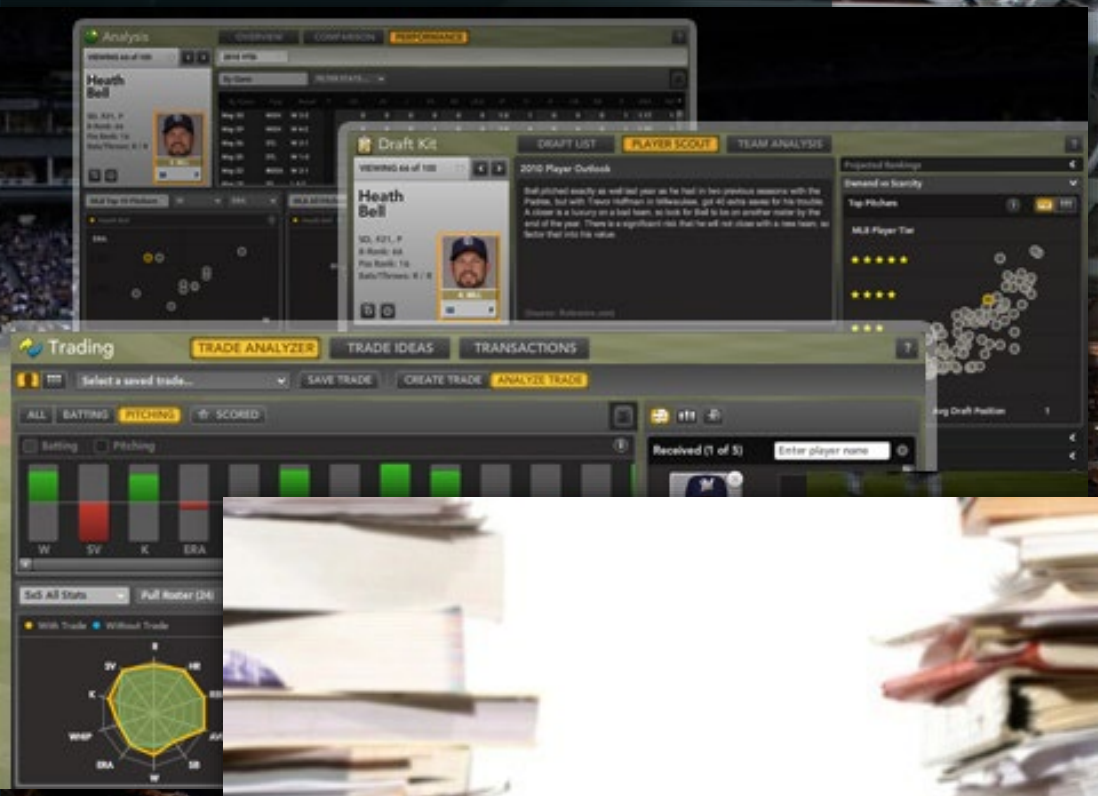
Messi: Science of Soccer

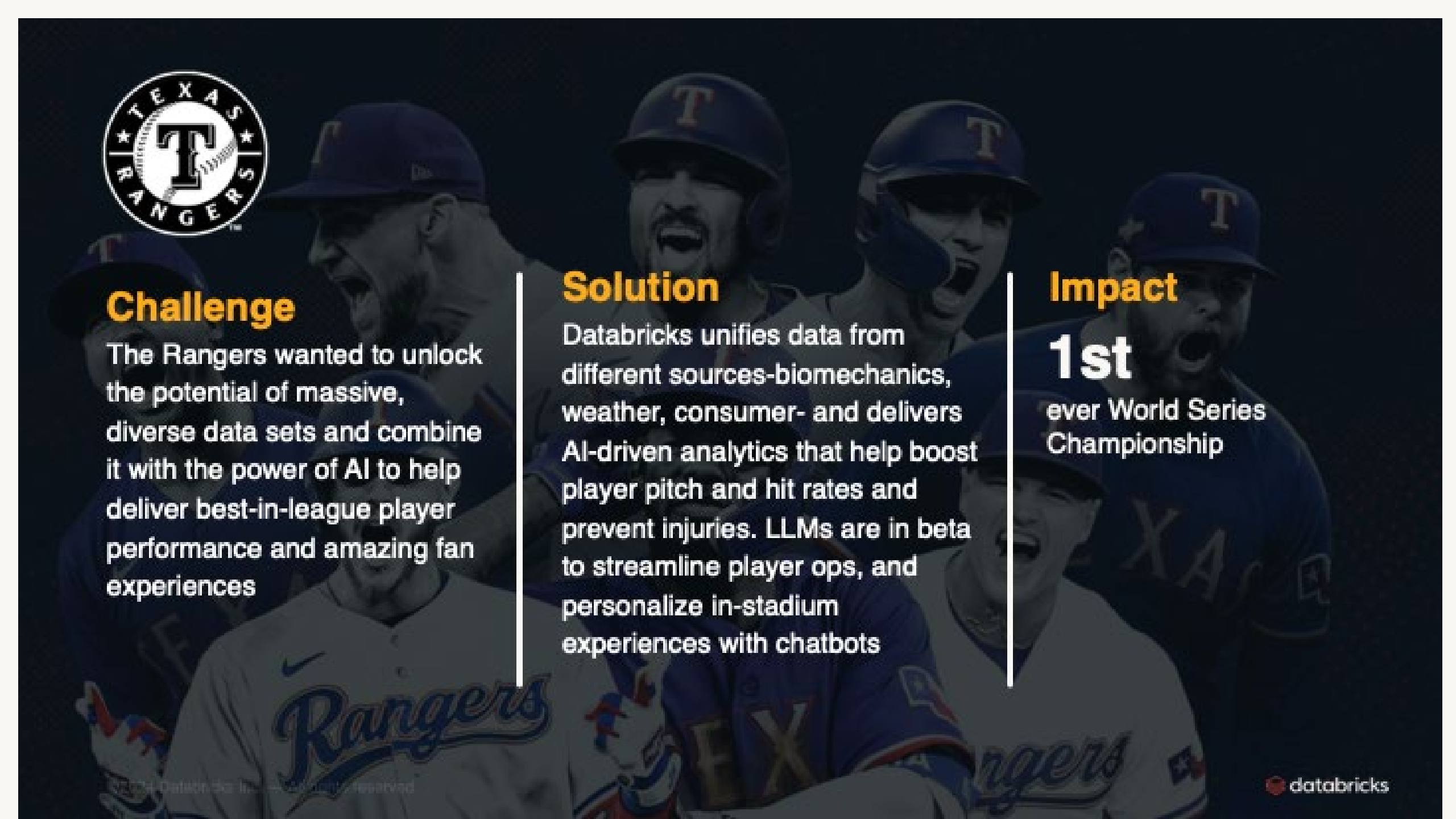


External Data for better insights



Prepare to Win!





Challenge

The Rangers wanted to unlock the potential of massive, diverse data sets and combine it with the power of AI to help deliver best-in-league player performance and amazing fan experiences

Solution

Databricks unifies data from different sources-biomechanics, weather, consumer- and delivers AI-driven analytics that help boost player pitch and hit rates and prevent injuries. LLMs are in beta to streamline player ops, and personalize in-stadium experiences with chatbots

Impact

1st

ever World Series Championship

Databricks is delivering the data intelligence that helps the Rangers to win

4X

more data at the same cost as other multi-cloud data warehouses

300+

users accessing data securely under a unified governance model

ANALYSTS/PLAYERS / COACHES / SCOUTS / BUSINESS OPS

10X

faster democra- tization of insights across the organization

1ST

ever world series championship

President - Oracle Users Group

during Java, MySQL, Peoplesoft acquisitions

IOUG

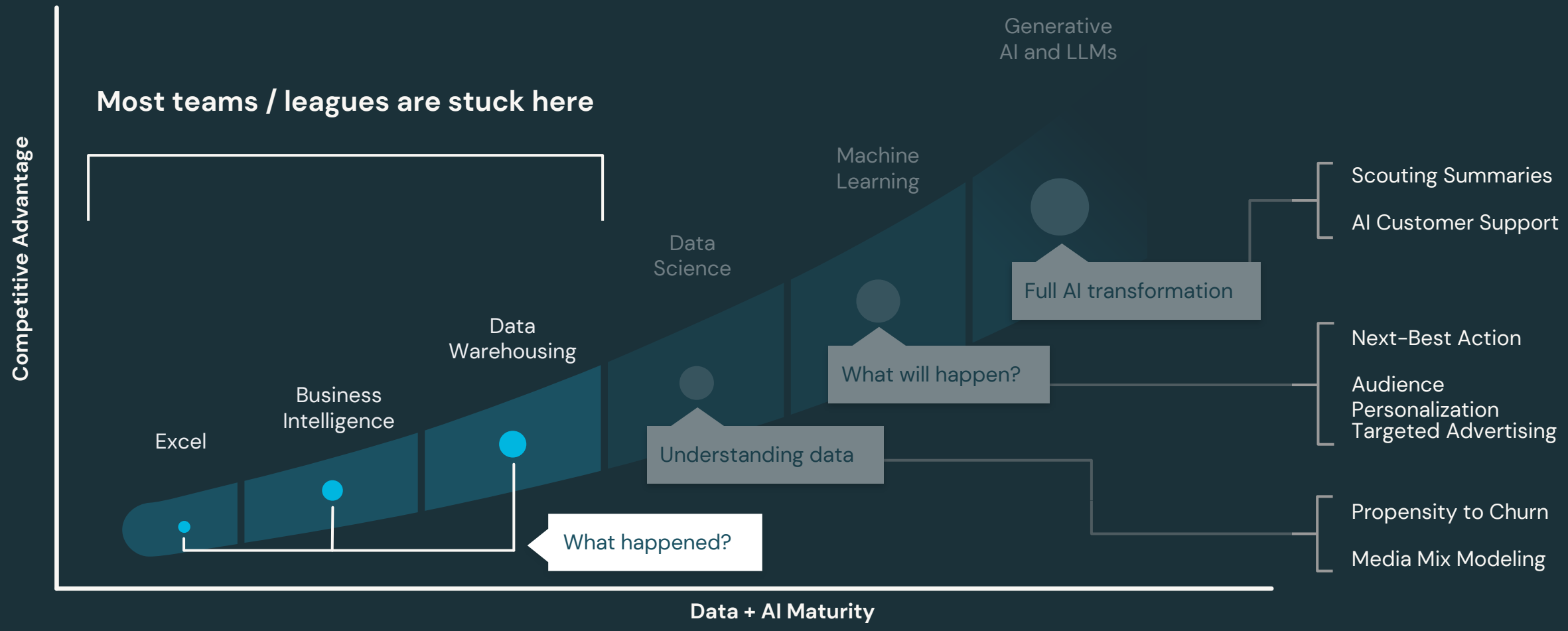
"The Father of Mobile Business Software"

Expand Beyond

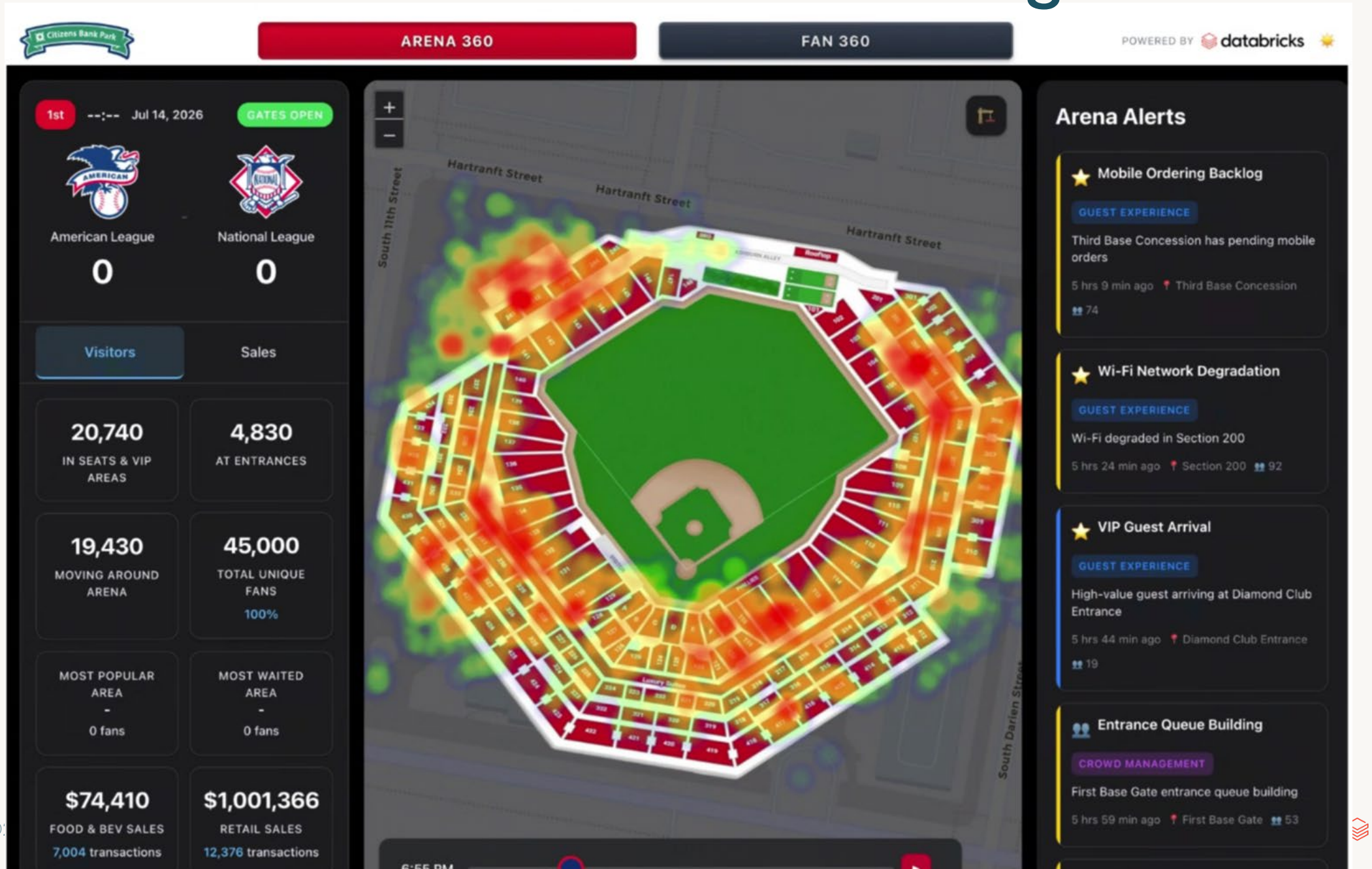
Wireless Enterprise Management Now!™

PocketDBA: pioneering technology on handhelds

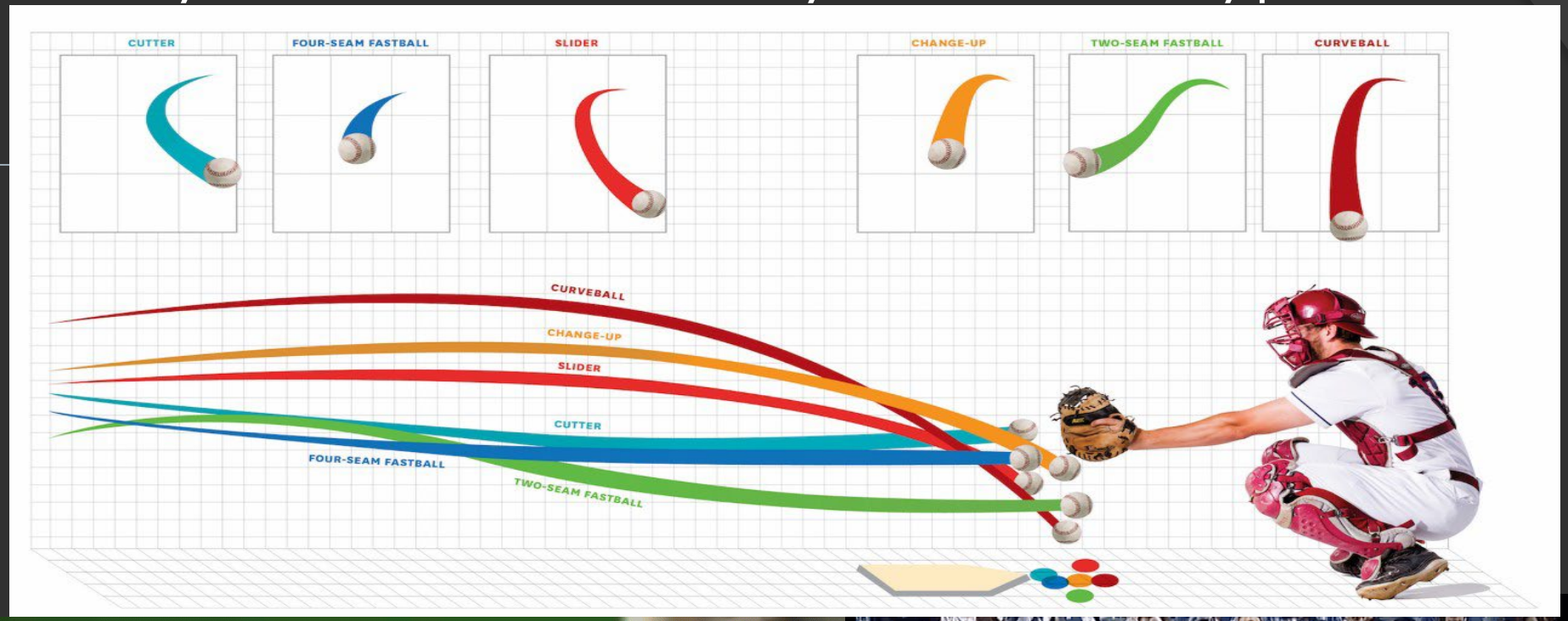
Data and AI maturity curve



Real-time stadium insights



Key to Baseball Analytics: Pitch Types



https://s3-us-west-2.amazonaws.com/edgar-images-production/users%2F4098%2Fwxcf36s2apo-baseball_no-pitchers.jpg

Work Load Analysis

Lovelytics

Player Fatigue Analysis: Workload Tracker

Select Date








	Bogaerts X	Devers R	Guerrero V	Hernandez T	Judge A	Perez S	Semien M	Trout M
Position	SS	3B	1B	LF	RF	C	2B	CF
Total Games	68	54	62	48	60	124	72	66
Work	41	26	41	34	39	27	41	31
Last Day Off	5/12/2021	5/12/2021	5/12/2021	5/12/2021	5/14/2021	5/14/2021	5/12/2021	5/12/2021
Days Since	3	3	3	3	1	1	3	3

EFFORT LAST 8 DAYS

40%> 20%> Avg 20%< 40%<

Player Name	Avg Effort	-1	-2	-3	-4	-5	-6	-7
Bogaerts X	64							
Devers R	26							
Guerrero V	61							
Hernandez T	46							
Judge A	49							
Perez S	71							
Semien M	70							
Trout M	55							

MINUTES ON FIELD

Base Mins on Field

Player & Season Avg.	Base Mins on Field
Judge A 6	12
Perez S 4	10
Semien M 6	10
Trout M 4	3
Bogaerts X 6	1
Guerrero V 5	1
Devers R 4	0
Hernandez T 4	0

RUNNING

Sprints

Player & Season..	Sprints
Semien M 4.3	6
Devers R 1.8	2
Guerrero V 2.2	2
Judge A 1.7	1
Bogaerts X 3.1	0
Hernandez T 1.7	0
Perez S 0.1	0
Trout M 2.7	0

BATTING

Pa

Player & Season..	Pa
Guerrero V 4.3	5
Hernandez T 3.8	5
Semien M 4.3	5
Bogaerts X 4.3	4
Devers R 3.5	4
Judge A 4.2	4
Trout M 4.1	4
Perez S 3.7	3

ACUTE TO CHRONIC WORKLOAD

Player Name	Ratio	Last 7	Last 28
Bogaerts X	1.04	72	69
Devers R	1.00	38	38
Guerrero V	1.08	69	64
Hernandez T	.77	30	39
Judge A	1.00	56	56
Perez S	.82	58	71
Semien M	1.10	80	73
Trout M	.98	63	64

THROWING

Hard Throws

Player & Season Avg.	Hard Throws
Perez S 0	0
Semien M 0	0
Guerrero V 1	0
Judge A 0	1
Hernandez T 0	1
Trout M 0	1
Devers R 1	0
Bogaerts X 1	0

CATCHING

Blocked Balls

Player & Season..	Blocked Balls
Perez S 3.6	2
Semien M 0	0
Guerrero V 0	0
Judge A 0	0
Hernandez T 0	0
Trout M 0	0
Devers R 0	0
Bogaerts X 0	0

Statistical Analysis: “Be Careful What Variables to Include”

Be careful what set of variables you base your conclusions on

The Uncanny Similarities Between Celine Dion and Napoleon

Canadian songstress Celine Dion

St Helena resident Napoleon Bonaparte

Have you ever noticed how similar Celine Dion and Napoleon are? No? Well then, you obviously haven't examined the evidence:

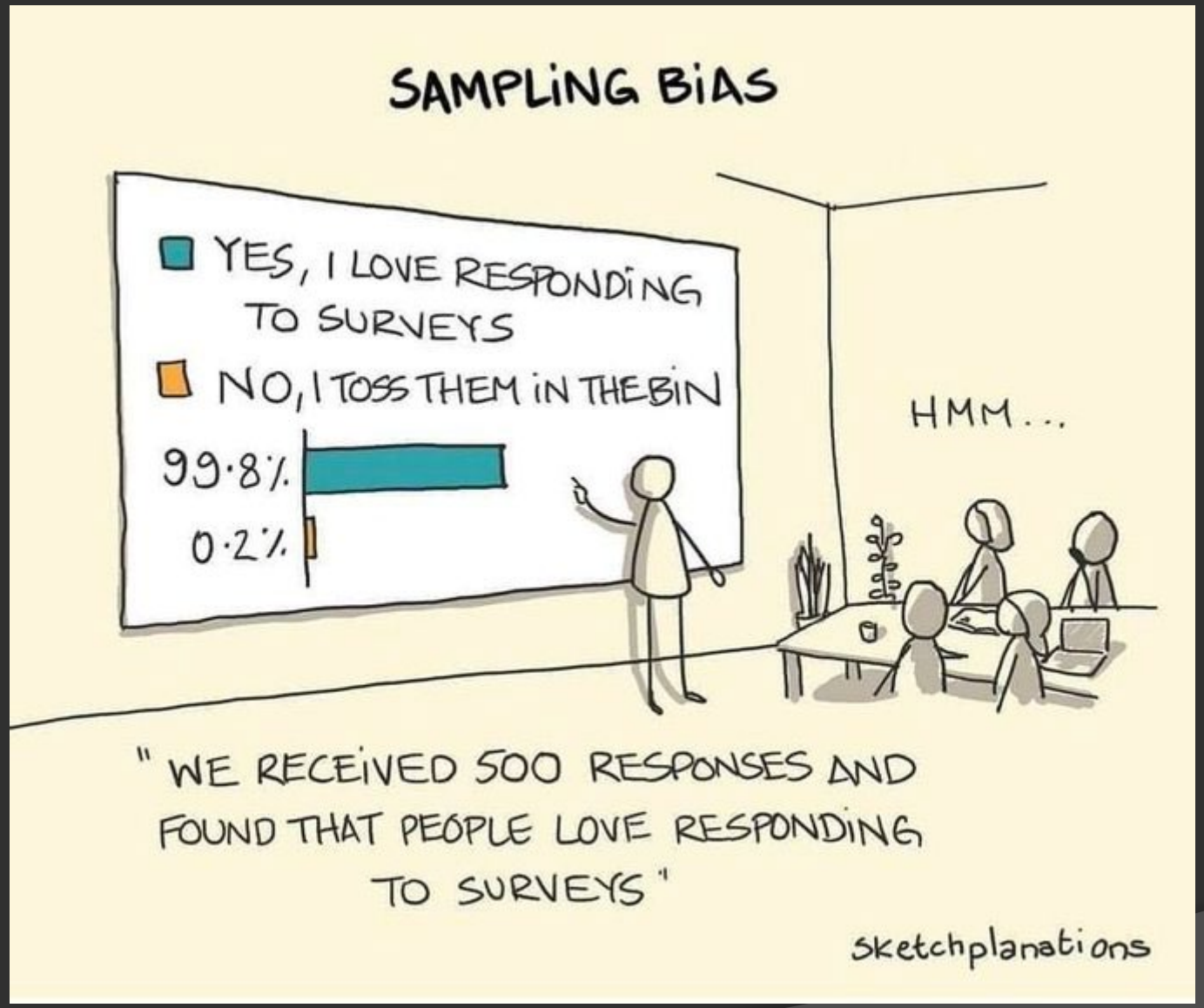
	Celine Dion	Napoleon
Carbon-based life form:	Yes	Yes
Capable of speaking French:	Yes	Yes
Number of fences they climbed over as children:	Unknown	Unknown
Born before or after the Jurassic era:	After	After
Faster or slower than a cheetah:	Slower	Slower
Qualifications in Dentistry:	None	None
Number of successful invasions of Russia:	Zero	Zero

Conclusion

There are probably one or two differences between them as well (haven't had time to check) but by examining the above evidence, we can safely conclude that Celine Dion and Napoleon are two of the most similar people in history.

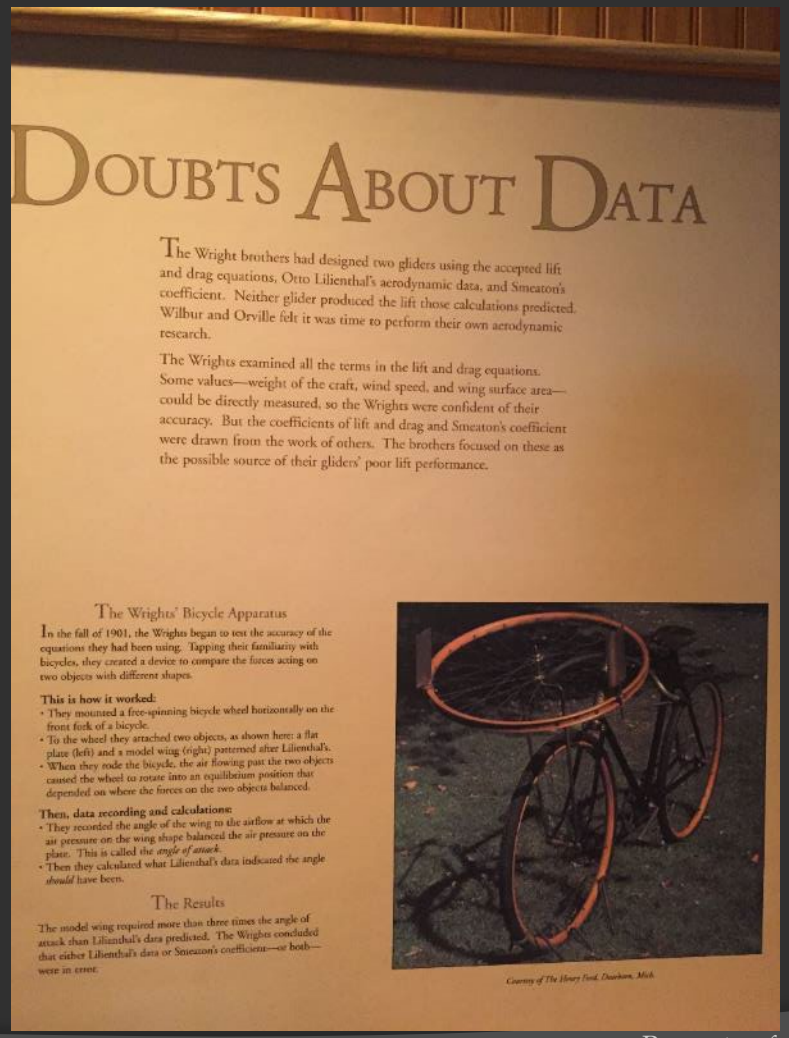
@Michael1979

Statistical Analysis: Sampling Bias



Creativity, be the first

- Wright Brothers: did not take prior work as fact – and their careers flew!



Doubts About Data

The Wright brothers had designed two gliders using the accepted lift and drag equations, Otto Lilienthal's aerodynamic data, and Smeaton's coefficient. Neither glider produced the lift those calculations predicted. Wilbur and Orville felt it was time to perform their own aerodynamic research.

The Wrights examined all the terms in the lift and drag equations. Some values—weight of the craft, wind speed, and wing surface area—could be directly measured, so the Wrights were confident of their accuracy. But the coefficients of lift and drag and Smeaton's coefficient were drawn from the work of others. The brothers focused on these as the possible source of their gliders' poor lift performance.

The Wrights' Bicycle Apparatus

In the fall of 1901, the Wrights began to test the accuracy of the equations they had been using. Tapping their familiarity with bicycles, they created a device to compare the forces acting on two objects with different shapes.

This is how it worked:

- They mounted a free-spinning bicycle wheel horizontally on the front fork of a bicycle.
- To the wheel they attached two objects, as shown here: a flat plate (left) and a model wing (right) patterned after Lilienthal's.
- When they rode the bicycle, the air flowing past the two objects caused the wheel to rotate into an equilibrium position that depended on where the forces on the two objects balanced.

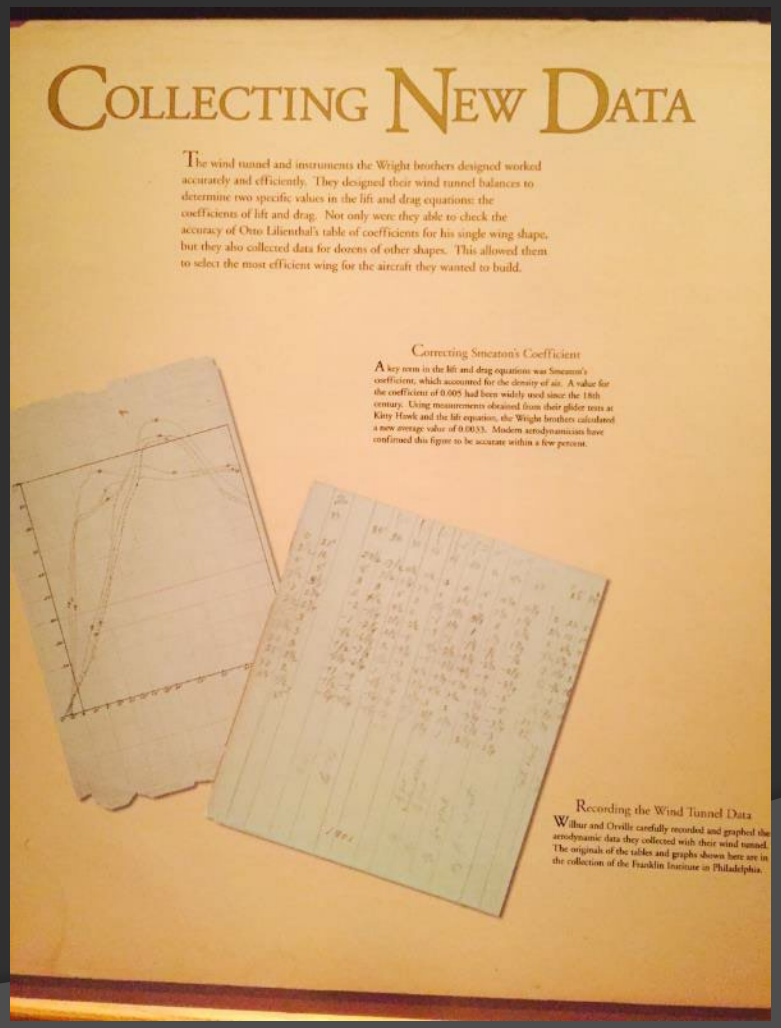
Then, data recording and calculations:

- They recorded the angle of the wing to the airflow at which the air pressure on the wing shape balanced the air pressure on the plate. This is called *the angle of attack*.
- Then they calculated what Lilienthal's data indicated the angle *should* have been.

The Results

The model wing required more than three times the angle of attack than Lilienthal's data predicted. The Wrights concluded that either Lilienthal's data or Smeaton's coefficient—or both—were in error.

Courtesy of The Henry Ford, Dearborn, Mich.



Collecting New Data

The wind tunnel and instruments the Wright brothers designed worked accurately and efficiently. They designed their wind tunnel balances to determine two specific values in the lift and drag equations: the coefficients of lift and drag. Not only were they able to check the accuracy of Otto Lilienthal's table of coefficients for his single wing shape, but they also collected data for dozens of other shapes. This allowed them to select the most efficient wing for the aircraft they wanted to build.

Correcting Smeaton's Coefficient

A key term in the lift and drag equations was Smeaton's coefficient, which accounted for the density of air. A value for the coefficient of 0.005 had been widely used since the 18th century. Using measurements obtained from their glider tests at Kitty Hawk and the lift equation, the Wright brothers calculated a new average value of 0.0033. Modern aerodynamicists have confirmed this figure to be accurate within a few percent.

Recording the Wind Tunnel Data

Wilbur and Orville carefully recorded and graphed the aerodynamic data they collected with their wind tunnel. The originals of the tables and graphs shown here are in the collection of the Franklin Institute in Philadelphia.

Communicate Actions Simply

Example: Goalie Jordan Pickford

Hitting Mechanics: what's the optimal stance?





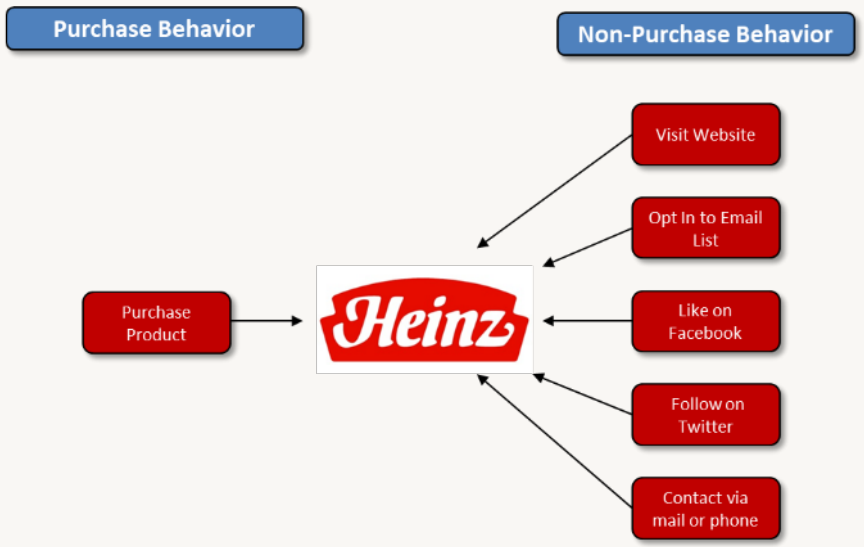


Experience Transfer...

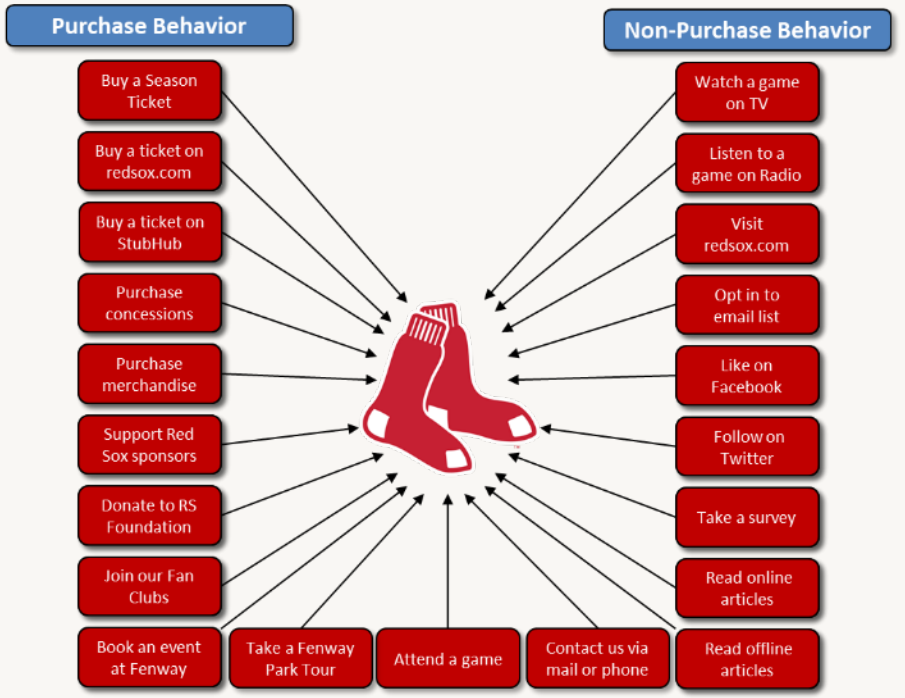
APPENDIX

Use Case #1: Customer Data Aggregation

Ways Customers Engage with Most Brands



Ways Fans Engage with the Red Sox

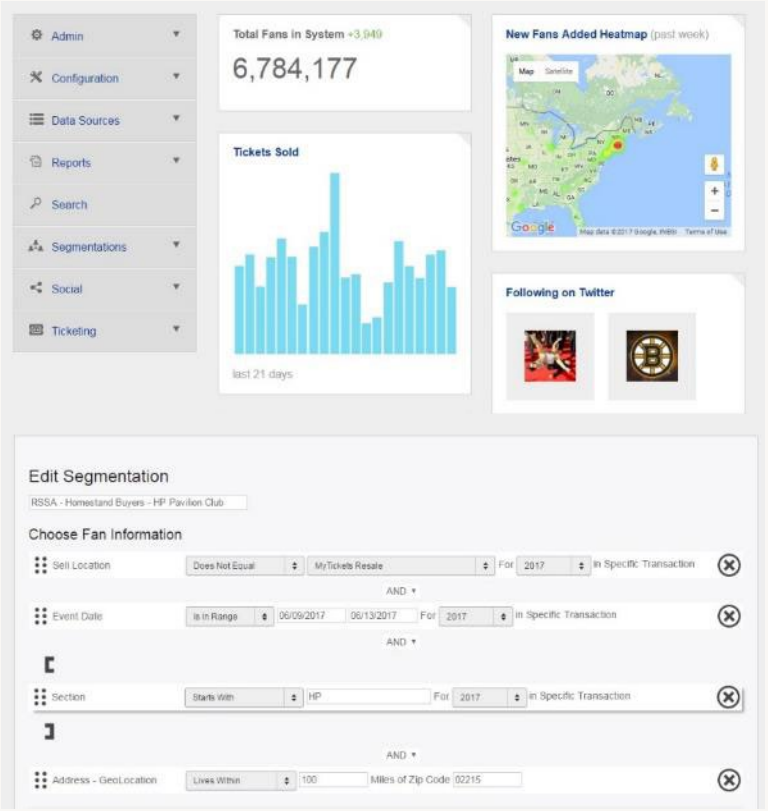


Use Case #1: Customer Data Aggregation

Red Sox Data Environment Sources

acxiom
MLB.com
facebook
StubHub!
TICKETS.COM
Microsoft Dynamics
TURNKEY INTELLIGENCE
RED SOX REPLAY
FortressGB
twitter
micros
Appetize
ABI master mind
hostanalytics
CONCUR
regonline
retailpro
satisfi
ADP

Customer Profile Dashboard



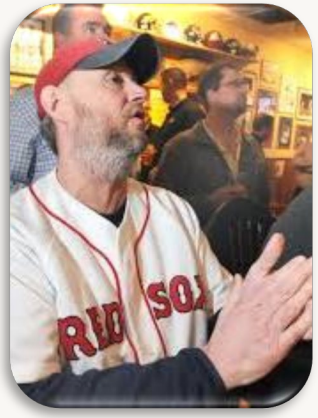
Customer Segmentation (Prior to 2018)

Devoted Dave

Family Focused Fran

Social Stan

Pink Hat Pam



Social Scott

Pink Hat Paula

Clustering-Derived Customer Segments

NESN Nelson

Student 9s Sam

Millennial Mel

One Night Stan

Boston Bargain Bob

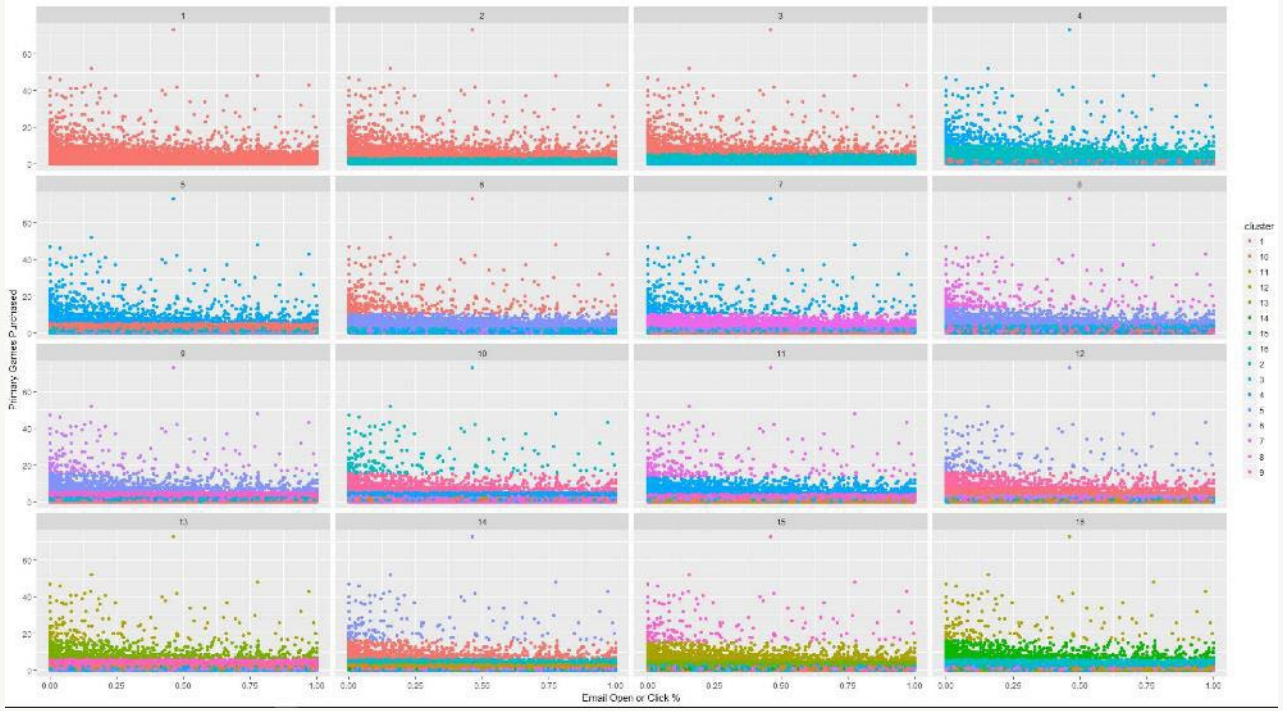
Clark Griswold

Richie Rich
(But Don't Buy Tix)

Lead List Leslie

STH Superfan Sal

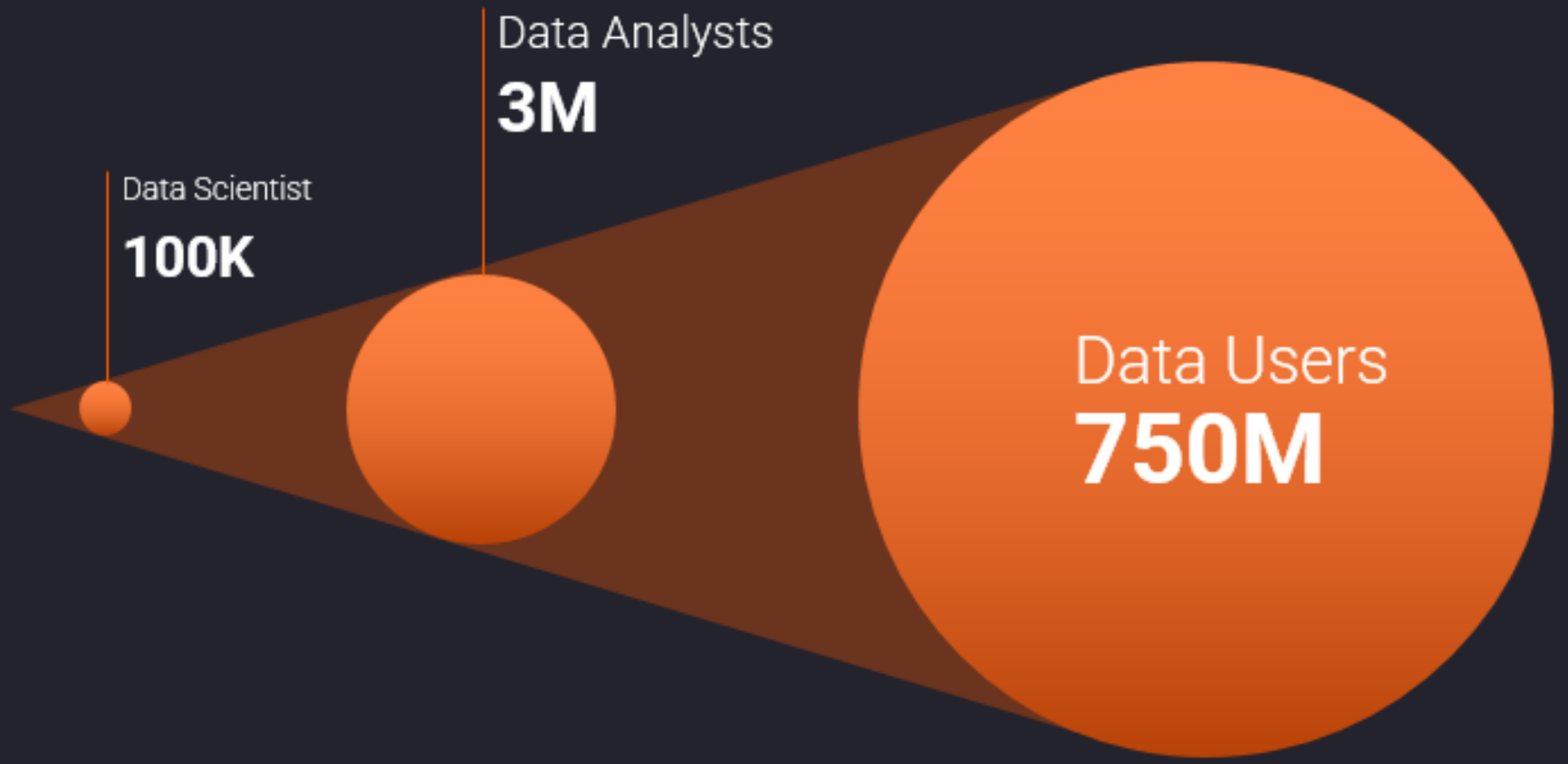
Customer Segmentation via Clustering Models



Data Inputs

Age
Children in Home
Marital Status
Gender
Distance from Stadium
Ticket Purchase History
Game Attendance History
Ticket Resale History
Household Income
Plan Purchase History
Digital Footprint

Democratization: Creating Citizen Data Scientists



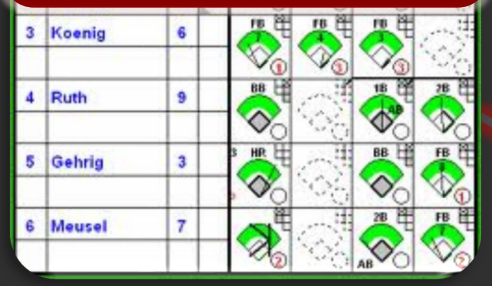
Extending AI to a Wider Audience

7500x gap between people who describe themselves as Data Scientist on LinkedIn and Excel's user base

Multistructural Data Sources (externally collected)

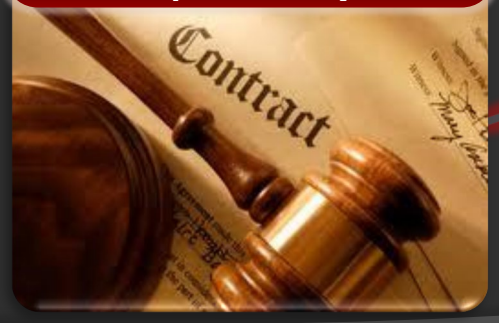
Statistics (MLB, STATS, AriBall, Inside Edge)

Play-by-play (MLB Advanced Media)



Defense (BIS)

Contracts & Financials (eBIS, Stadium operations)



Pitch & hit mechanics (Hawkeye, SportVision, Trackman)

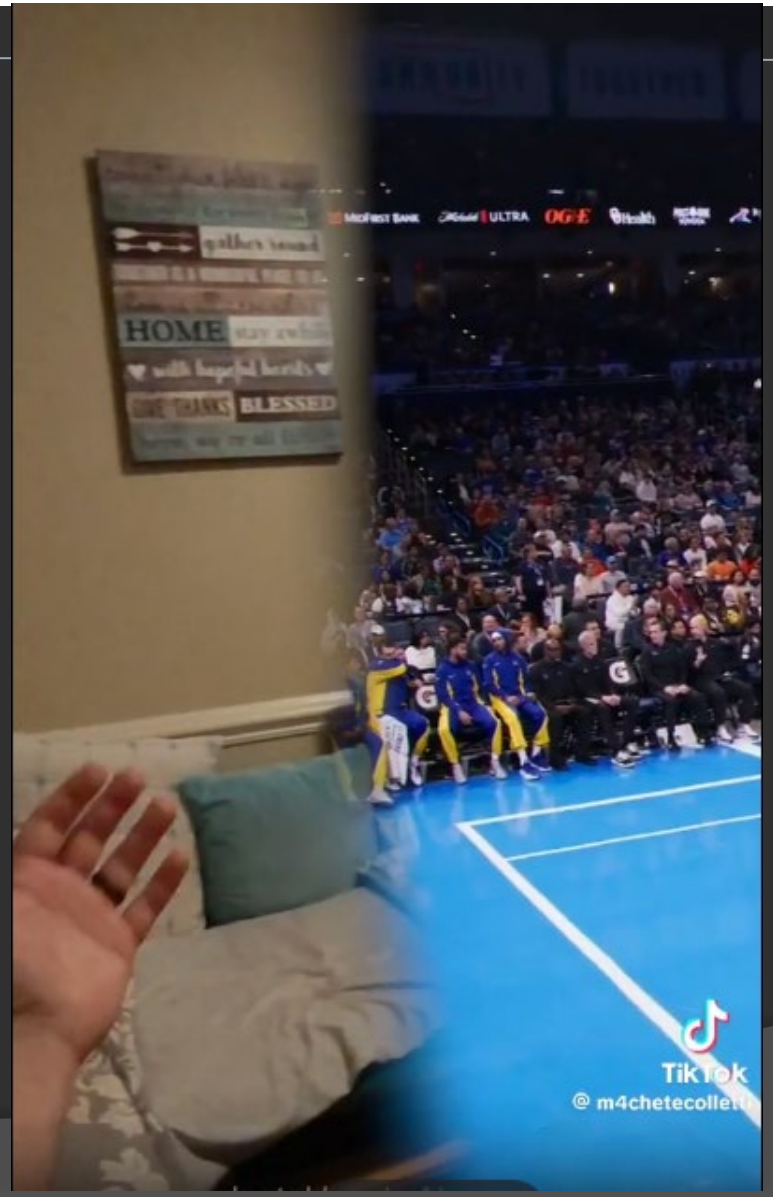
Multistructural data sources (internally collected)

Future of Stadiums

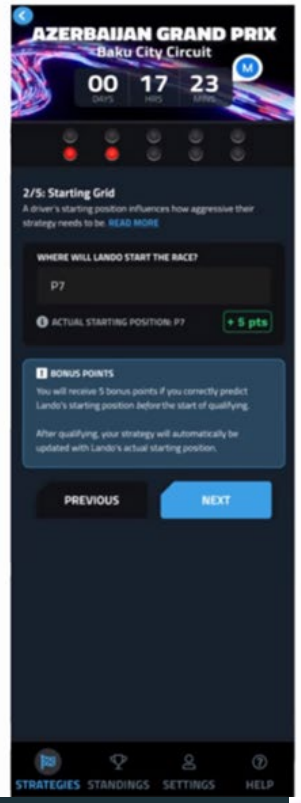
FC Bayern basketball floor now an LED screen

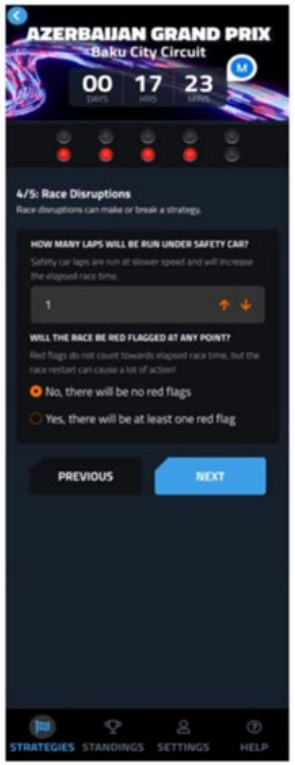
Future of Fan Experience: VR

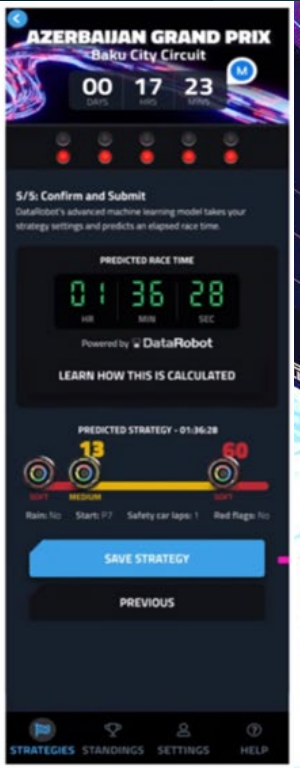
Available now!



Formula 1







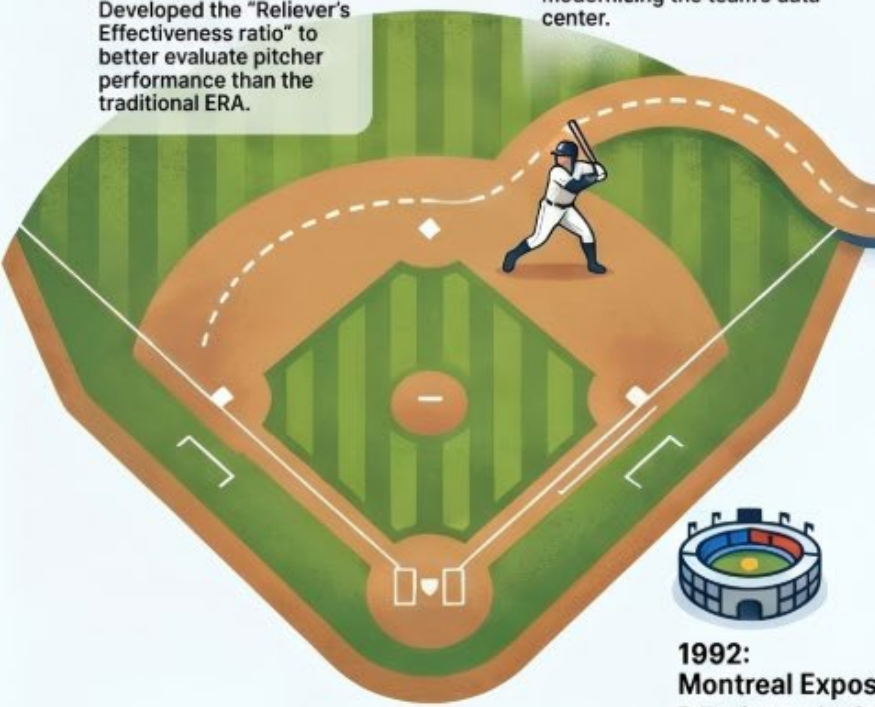
From the Diamond to the Lakehouse: The Career Pipeline of Ari Kaplan

1989: Caltech SURF Research

Developed the "Reliever's Effectiveness ratio" to better evaluate pitcher performance than the traditional ERA.

1990: Baltimore Orioles Internship

Hired on the spot by the Orioles' owner after his SURF presentation, Kaplan began creating scouting software and modernizing the team's data center.



1991: San Diego Padres

Kaplan continued his summer consultancy work, providing analytical insights for the Padres' front office.

1992: Montreal Expos

Following graduation, Kaplan joined the Expos to develop their scouting and player development database systems from the ground up.

2000s–2010: Shaping the Modern Front Office

2009: Founding of AriBall (Scoutables)

Partnered with former Dodgers GM Fred Claire to launch a SaaS platform that uses risk models to forecast and prevent player injuries.

2010: Leading the Chicago Cubs Analytics

Recruited by owner Tom Ricketts to create and lead the organization's first dedicated analytics department, speerheading their technical transformation.

2021–2023: McLaren Formula 1 Team

Applied AI models to race strategy for drivers Lando Norris and Daniel Ricciardo, where millisecond decisions are made during race lapa.

Current: Head of Technical Evangelism at Databricks

Kapian now leads global evangelism for the $480+ data and AI company, focusing on generative AI and the democratization of data intelligence.

The "Moneyball" Inspiration:

Kaplan's ploneering work in shifting baseball from "gut teel" to data-driven decision-making was part of the composite inspiration for the book and film Moneyball.

"Data Intelligence Democratizes Analytics":

Kaplan advocates for platforms where anyone can engage with data in their native language, reducing the need for specialized data scientists for every insight.