# SPORTS, SPORTS FANS, & SYNTHETIC DATA

24MM viewers tuned in each day for the Winter Olympics. Olympic viewing is streaming-first and fragmented, unlike the communal Super Bowl experience.

arima

# THE MOST VALUABLE MARKETING DATASET WILL BE DRIVEN BY SPORTS

I'm not a particularly rabid sports fan in any way. However, I've had my fair share of career highlights as a marketer, as a creative director, and as a strategist on projects in the sports world. Whether we're sports fans or not, we are swept up and included in the global media landscape.

This year, because the Super Bowl and the Winter Olympics opened within the same weekend, we had a rare alignment of two of the largest communal viewing events on Earth. The national, ritualized, and culturally saturated Super Bowl contrasts with the multinational, multi-time zone, multi-sport Olympics.

The formats and rhythms of these two events are very different. However, some of the fundamentals remain the same when it comes to synchronized audience attention at scale.

Sports remains the most reliable mechanism for assembling large audiences around a shared moment. By-appointment viewing has migrated almost entirely into live sports. There is no equivalent, certainly not with the consistency of live sports.

News spikes unpredictably. Entertainment launches are staggered across platforms. Social media (that sparked the beginning of the end of mass audiences) disperses attention into hundreds of thousands of micro-communities. Only sports produces recurring, predictable, emotionaly charged mass audiences who show up daily and weekly, seasonally for playoffs, and cyclically every four years.

\\

\\

## It comes as no surprise that we now see a collision of "capital" and "attention".

The global sports economy is estimated to be ~$2–3 trillion in value from media rights, venues, leagues, teams, apparel, betting, sponsorship, and more. The growth projections for the industry are strong. The investor class has leaned in, too.

Private equity firms have created dedicated sports investment funds measured in billions. Sovereign wealth funds are acquiring leagues, teams, and media rights portfolios. Technology platforms are competing aggressively for broadcast rights that were once dominated by cable networks.

Amazon secured NFL Thursday Night Football and Prime Monday Night Hockey. YouTube acquired NFL Sunday Ticket. Netflix entered live sports programming. Live sports rep-resent one of the few remaining forms of programming that can guarantee large, simultaneous audiences. It's also cheaper to produce than a television show, and more reliable in quality.

Last November, I attended a breakfast hosted by The Globe and Mail where Saudi Arabia's Minister of Investment ex-plained how the Kingdom's Public Investment Fund had invested billions across football, golf, esports, and gaming infrastructure. He remarked specifically that their esports investment had often surpassed traditional sports invest-ments. Think of esports as an early signal of where future and younger audiences may congregate.

The esports economy generates its own billions of dollars and attracts hundreds of millions of viewers, particularly younger audiences. For investors and marketers, the implication is pretty clear: start thinking beyond traditional leagues, and think about hybrid physical/digital ecosystems.

# The definition of "sports fan" is **widening.**

Women's sports are experiencing record attendance and sponsorship demand. New leagues are emerging and established leagues are reaching previously underserved fans.

A Buffalo Bills fan might follow the Toronto Raptors, play NBA 2K26, and participate in a fantasy league. But I'm willing to bet that you can't find that fan's profile in any single dataset because modern sports fans are multidimensional.

From a data perspective, that fan is pretty much invisible.

Fan data exists everywhere and nowhere. Ticketing platforms capture one set of transactions, while streaming platforms track engagement, and social media platforms capture sentiment and spark conversation. An individual sports fan might appear as multiple disconnected identities across dozens of databases owned by different organizations, each with their own data structures, privacy constraints, and commercial incentives.

Our opinions about first-party data have been challenged. First-party data is now plagued by gaps, bots, and fraudsters. None of these problems are going to be easily fixed. As a marketer, I'm willing to bet that every one of our clients is going to soon want to build AI-driven predictive models for all kinds of strategic pursuits.

Over the past several years, synthetic data has emerged as a practical solution to fragmented or incomplete datasets across industries. While the use of synthetic data has barely emerged in marketing, it has been widely used in healthcare, finance, and autonomous driving.

\\

I was introduced to Winston Li and Ray Kong at Arima when I was imagining how to build look alike audiences that might use AI. I was working on a high-pressure pitch on a telco account. Kristen Wozniak said "Oh, you should talk to Arima, they're building models around synthetic data."

Didn't win the pitch. But we definitely shook the room, and Arima and their synthetic data have stuck with me ever since.

Synthetic datasets are statistically generated populations designed to mirror real-world characteristics without exposing identifiable personal information. Arima's founder Winston Li describes Arima's algorithm as his "Covid baby." Arima built the Synthetic Society to allow organizations to test scenarios and train predictive models with far greater flexibility than traditional datasets.

The synthetic model of a population enables organizations to explore and query demographic, behavioral, and economic questions and queries at scale.

Winston, Ray, and I got to talking. And the more I understood, the more it became cleatr that we could also build a specialized synthetic population focused entirely on sports fans.

Part way through 2025, we co-authored a submission for funding from the AI Compute Access Fund that offers financial support to startups and small businesses. The goal was to finance and access the compute power we would need to scale and commercialize innovative AI projects.

The Synthetic Sports Fan Dataset is now underway, and the Arima team will spend a significant portion of our time on it this year. Presentations are already on deck for Canada, USA, and EMEA.

# $8MM USD

Average cost of a 30 sec. Super Bowl TV commercial

\\

+ Sports remains the most reliable mechanism for assembling large audiences around a shared moment

+ Super Bowls, Olympics, World Cups, and major league seasons concentrate audience attention.

+ By-appointment viewing has migrated almost entirely to live sports.

**a**

When we can model fan identities across sports, geographies, media behaviors, consumption patterns, and emerging categories like esports, the dataset will allow marketers to explore questions that that traditional data cannot answer:

+ Which cross-sport fan segments are growing fastest?

+ How does engagement shift across life stages?

+ What combinations of sports properties overlap within households?

+ Where do sponsorship opportunities create incremental reach rather than redundancy?

+ What behaviors signal future fandom before it becomes visible in ticket sales or merch sales?

These questions all have material economic consequences.

\\

*Why* is this important right now? Three forces converging:

+ **Mass communal events remain economically dominant.**
Super Bowls, Olympics, World Cups, and major league seasons concentrate audience attention

+ **AI adoption is accelerating across marketing and business strategy.**
Organizations increasingly expect predictive modeling, scenario planning, and better decision-making support. First-party data limitations are becoming unavoidable.

+ **Privacy regulation and platform fragmentation reduce visibility into consumer behavior.**

\\

# Oh...

And did I mention bots? Or the automated, AI-driven fraudsters who are guaranteed to mess with the integrity of your data?

Even though direct observation is incomplete, you are still going to need to build models. You will use synthetic data to solve some of your problems...

**a**

## For the investors to think about:

Infrastructure layers tend to generate disproportionate value over time. I survived the dotcom crash and have the benefit of hindsight (insight gained the hard way, unfortunately). I have given several keynotes about what survives a crash and the long economic and creative impact that follows.

Media and media rights are infrastructure. Betting platforms have infrastructure. Streaming has reinvented distribution and has its own infrastructure.

Sports fan intelligence will need infrastructure too.

Organizations that understand audience behavior with greater precision will allocate capital more effectively. Whether you are trying to price sponsorship more accurately or capture emerging markets earlier, you're going to need a better foundation to work with.

Datasets that become foundational to industry decision-making will compound in value (and some of you already know my fintech obsession with business models that compound in value…) There are smarter paths to startup success than chasing unicorns.

As we refine the sports fan dataset, we need partners who will help expand our continuous data within a shared modeling environment.

The Synthetic Sports Fan Dataset is being built as an eco-system with coherent, world-building marketing logic in mind.

The Synthetic Sports Fan Dataset is being built as an ecosystem with coherent, worldbuilding marketing logic in mind.

The Super Bowl and the Winter Olympics remind us that  people still gather around shared narratives. But the fragmentation of digital media has made it difficult to understand those moments.

\\

\\

## THE NEXT EVOLUTION IS PREDICTIVE

We want to anticipate how fan communities evolve before they fully emerge.

Sports offers one of the best possible environments for that exploration because emotional engagement and economic activity intersect so visibly and consistently. It's digital and measurable, and so it will eventually be "simulatable" (if that's a real word).

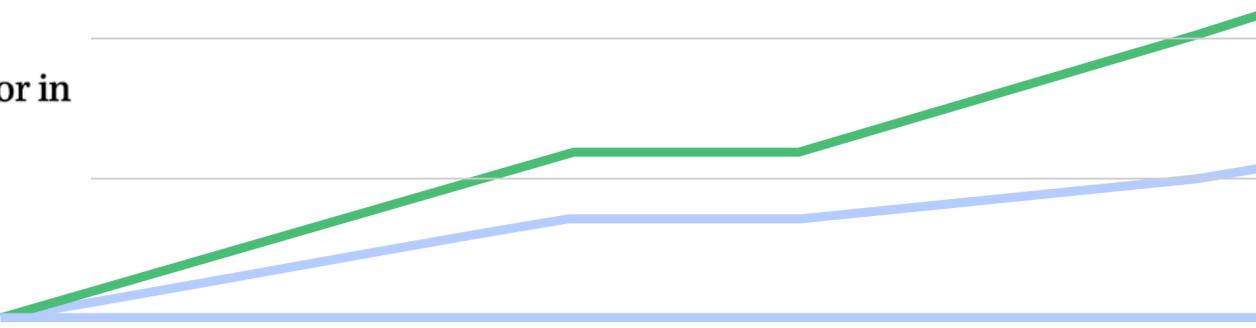**Amazon and Netflix have become big investors in streaming sports rights.**

\\

# Follow the money...

Private equity and sovereign wealth funds have also dedicated billions to acquire teams, leagues, and media rights portfolios.
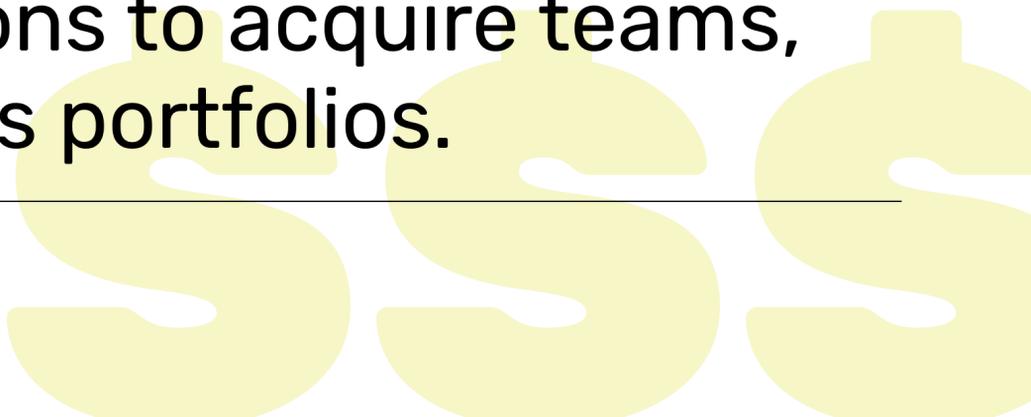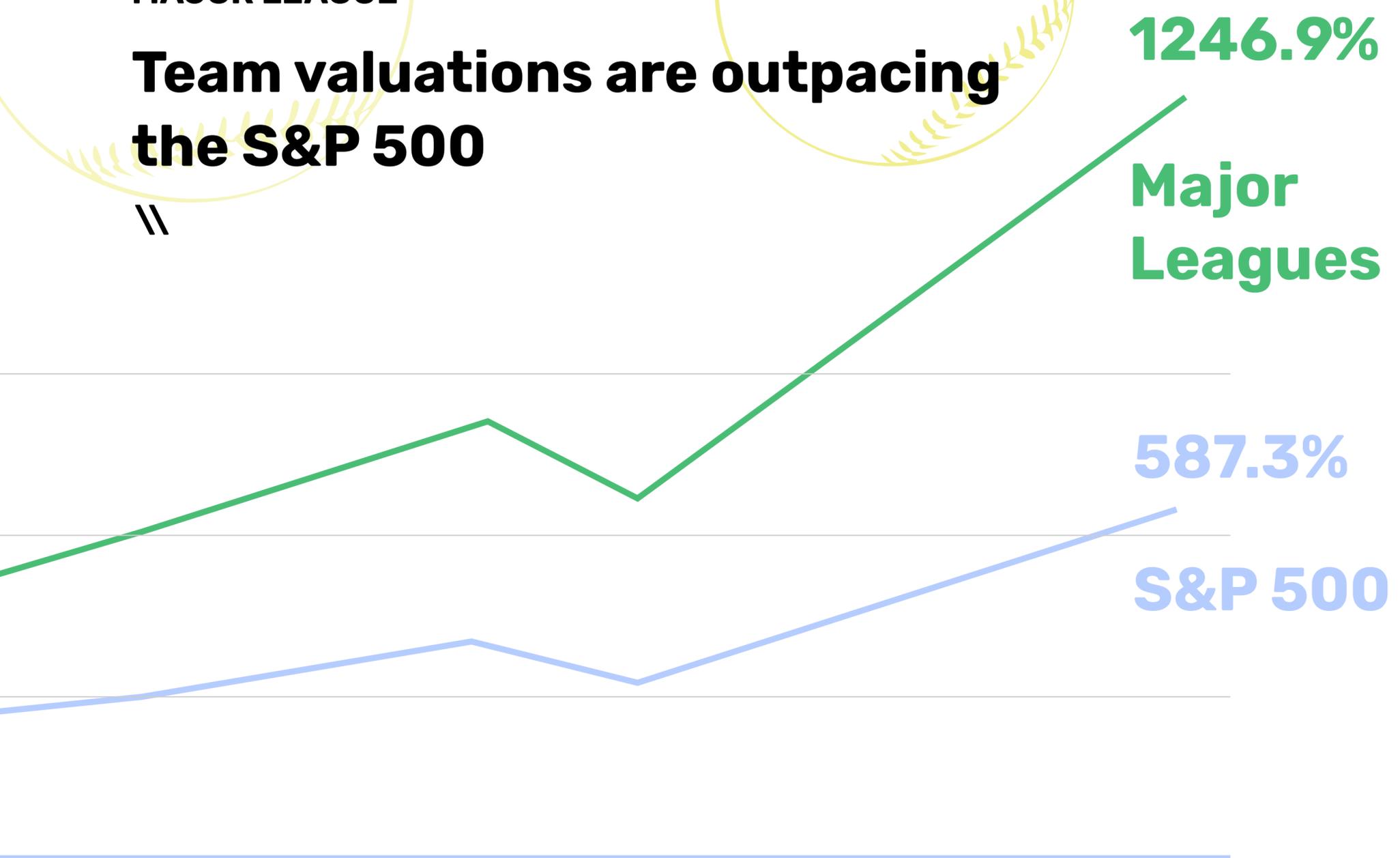
# Team valuations are outpacing the S&P 500

\\

**1246.9%**

**Major Leagues**

**587.3%**

**S&P 500**

The cumulative total returns of each of the four major sports leagues, based on average team value, have surpassed the returns of the S&P 500

NFL, MLB, NBA, NHL vs. S&P 500 from 2002 to 2024
SOURCE: Pitchbook, J.P. Morgan Asset Management

**a**

## For the investors to think about:

Infrastructure layers tend to generate disproportionate value over time. I survived the dotcom crash and have the benefit of hindsight (insight gained the hard way, unfortunately). I have given several keynotes about what survives a crash and the long economic and creative impact that follows.

Media and media rights are infrastructure. Betting platforms have infrastructure. Streaming has reinvented distribution and has its own infrastructure.

Sports fan intelligence will need infrastructure too.

Organizations that understand audience behavior with greater precision will allocate capital more effectively. Whether you are trying to price sponsorship more accurately or capture emerging markets earlier, you're going to need a better foundation to work with.

Datasets that become foundational to industry decision-making will compound in value (and some of you already know my fintech obsession with business models that compound in value…) There are smarter paths to startup success than chasing unicorns.

As we refine the sports fan dataset, we need partners who will help expand our continuous data within a shared modeling environment.

The Synthetic Sports Fan Dataset is being built as an ecosystem with coherent, worldbuilding marketing logic in mind.

The Super Bowl and the Winter Olympics remind us that  people still gather around shared narratives. But the fragmentation of digital media has made it difficult to understand those moments.

\\

\\

## THE NEXT EVOLUTION
## IS PREDICTIVE

We want to anticipate how fan communities evolve before they fully emerge.

Sports offers one of the best possible environments for that exploration because emotional engagement and economic activity intersect so visibly and consistently. It's digital and measurable, and so it will eventually be "simulatable" (if that's a real word).

# WHAT COMES NEXT

My next dispatch in collaboration with Arima will detail initial findings, modeling approaches using their tools, and opportunities for collaboration as the dataset expands.

\\

🔍 arimadata.com

🔍 Winter Olympic sports

🔍 eSports NBA2K

🔍 Raptors win championship

🔍 creative destruction

🔍 synthetic data, automotive

🔍 Gartner report, Arima

## GUEST WRITER AND CONSULTANT TO ARIMA: SABAA QUAO

**Sabaa** is a creative leader, career entrepreneur, and business strategist. As a digital media veteran, he has built and grown several platforms and companies across content, tech, and culture, collaborating with marketing and technology teams in over a dozen countries.

He was previously the Chief Creative and Innovation Officer at Cossette and went on to become the founding President of PlusCo Venture Studio, the corporate venture and innovation offshoot of Cossette and Plus Company Canada.

The venture studio's first cohort of AI and fintech companies spanned solutions from synthetic data to knowledge management, expertise-as-software, and embedded finance. Each company offers early insights into where creative destruction will impact the marketing and advertising ecosystem. Collectively, the first cohort grew at a rate consistent with companies born in the AI era.

For 2026, Sabaa's venture studio approach will extend to collaborators and startups based in Austin, London, Lagos, Dubai, Nairobi, and Toronto.

Sabaa is a frequent keynote speaker and an industry advisor on creativity, entrepreneurship, and business model innovation. He has led, co-chaired, and served as a member of over 20 national and international marketing and short film juries.

In Canada, Sabaa sits on the Cannes Lions Advisory Board with The Globe and Mail and serves as the current Digital Young Lions co-chair.

## https://www.linkedin.com/in/sabaa/

# FUTURE OUTLOOK AND ACCESS

**Early insights from Arima's Synthetic Sports Fan Dataset will be released in an upcoming briefing.**

🔍 synthetic data

Register to receive the first release and participation details as an Arima customer, strategic data partner, or investor.

Scan to register

arima