

# The shift to data

Industry trends in finance



# Virtual webinar best practices

- Find a quiet place, free of as many distractions as possible and use headphones if possible
- Stay on mute unless you are speaking
- Remove or silence alerts from cell phones, e-mail pop-ups, etc.
- Participate in activities and ask questions - **this will be interactive!**



# Who we are

We provide customized, industry-tailored data science training and AI solutions — partnering with organizations to educate, equip, and empower their workforce with the skills to achieve their goals and expand their impact.

We deliver:

- High-quality data science training programs
- Customized executive workshops
- Custom software solutions and consulting services

Since 2014, we've worked with thousands of professionals to make their data work for them.



# What you'll walk away with today

1. Examples of how data is used in finance
2. Trends of data usage in the finance industry
3. Steps that you can take to integrate data into your operations

# Case study: using NLP to measure risk

- The IDB needed a dashboard to automatically identify infrastructure projects they fund that might require a modification
- Scores risk using several dozen factors, weighted by Machine Learning algorithms
- The model provides detailed risk reporting, both query-based/searchable and dash-boarded

Infrastructure Project Risk Analysis

This dashboard is powered by a set of machine learning algorithms to highlight transportation and other infrastructure projects that are most at risk of requiring modification or a cancellation. The table below highlights the risk levels of various projects. You can inspect additional project details on the "Contract Details" tab on the left hand side. Additional tabs provide you with the ability to inspect the data and model results further.

Select type of project

- All Construction projects
- Only MOPC projects

Select target variable to see model champion results

Risk of modifying amount

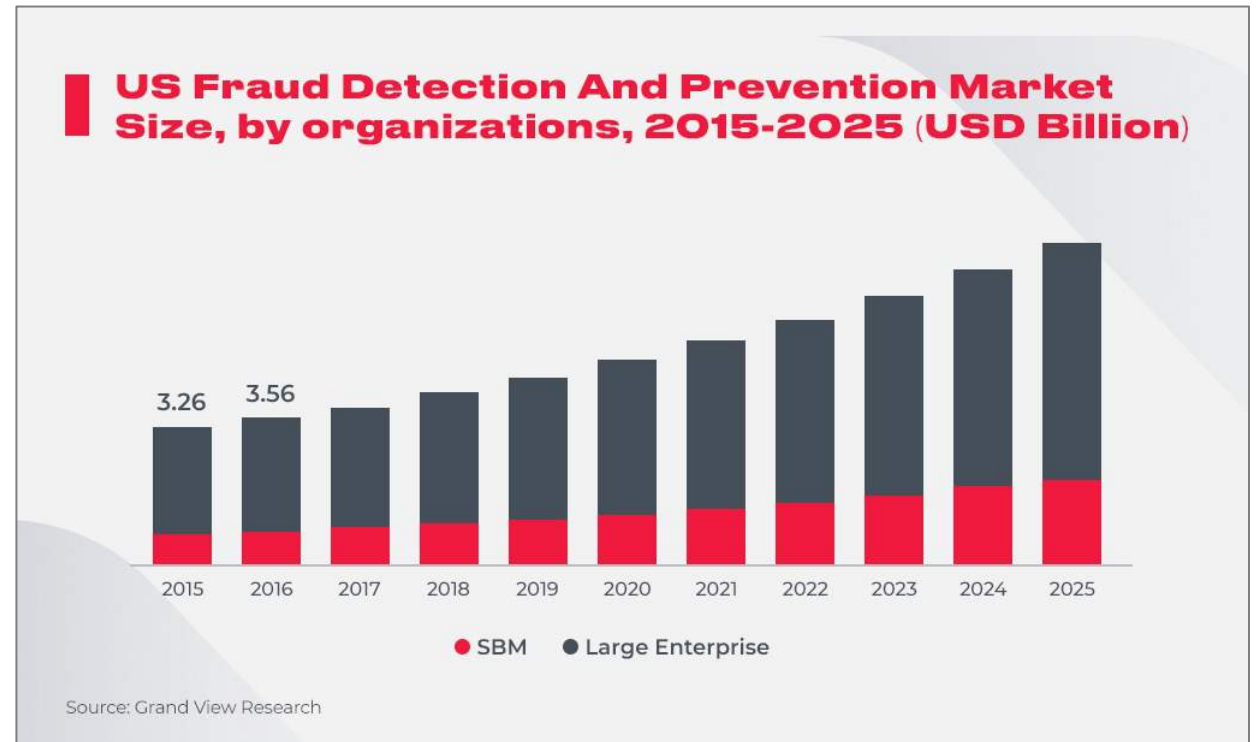
Predictions for contracts

Show 10 entries

Contracts ID	Predicted Level of Risk	Probability of High Risk	Probability of Medium Risk	Probability of Low Risk
1 270774-techno-plus-ingenieria-a-a-2	Medium	11.8%	49.2%	39%
2 270796-rieder-cla-saci-3	Medium	20.2%	48.2%	31.6%
3 270799-consorcio-yguazu-1-2	Medium	20.8%	49.4%	29.8%
4 270799-rieder-cla-saci-3	Medium	21%	49.2%	29.8%
5 272604-consorcio-empresarios-sun-3	High	65.6%	16%	18.4%
6 272604-viaticos-a-3	High	65.8%	17.2%	17%
7 272608-consorcio-vial-paraguari-5	Medium	31.8%	66.2%	2%
8 276640-consorcio-2g-s-1-ing-alejandro-pereira-heras-5	High	69.2%	4.4%	26.4%
9 287158-juan-manuel-palva-zawadzki-1	Medium	1%	90.4%	8.6%
10 287178-ricardo-hermer-saldivar-benitez-4	Low	36.4%	24.4%	39.2%

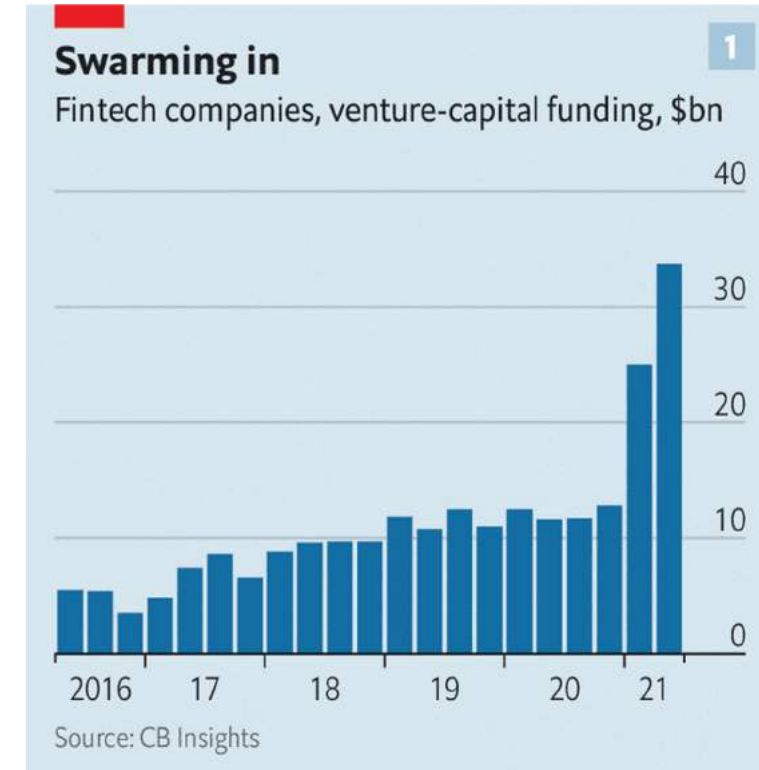
# Leveraging data in finance

1. Risk analytics
2. Identifying consumer behaviors and trends
3. Operational risk mitigation
4. Fraud detection and identity validation
5. Payment and transaction processing



# Key trends

- Enhanced trading & transaction technologies
  - Financial services was the leading sector for venture investment in 2021 with \$134 billion invested, marking a whopping 177 percent year-over-year growth
- NFTs and crypto currencies
  - Alternative stores of value are increasing diversification and trading options



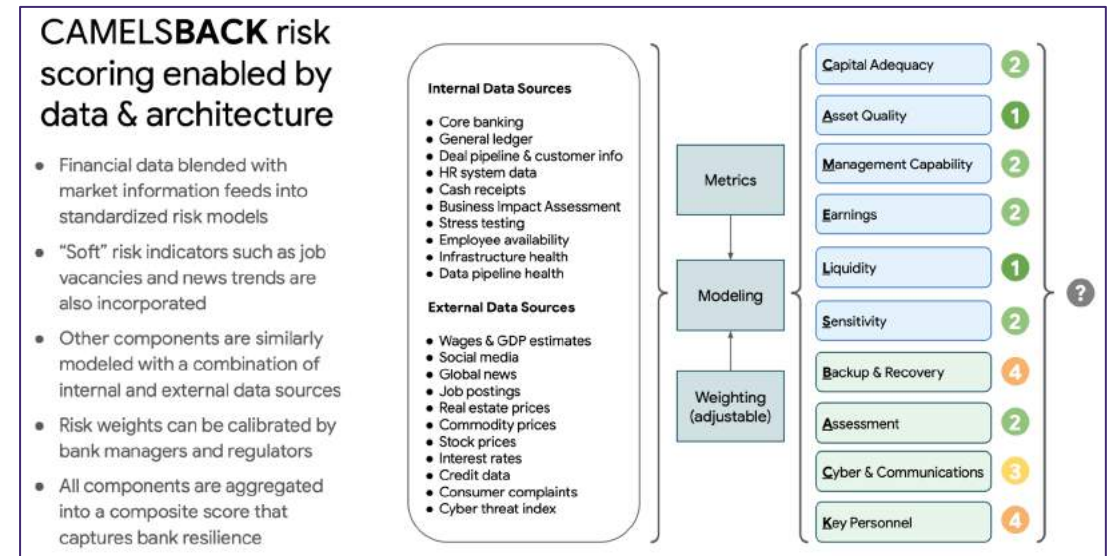
# Key trends

- Skills required of professionals are evolving
  - In 2021, the largest uptick in technical skill demand is for programming in R and Python (up by 33 per cent) and the process development framework Scrum (up by 2.6 per cent) - according to Robert Half, a recruiting firm
  - The talent to fill this demand is in short supply, which is pushing employers to increase salary offers for skilled professionals by 10 to 20 per cent
  - Skills requirements include:
    - Knowledge of databases is increasingly important
    - Knowledge of data structures and processes is important
    - Understanding of how to find patterns, and detect statistical significance
- A growing array of tools available for leveraging AI and machine learning in finance
  - Over 160 R packages for financial use cases
  - <https://cran.r-project.org/web/views/Finance.html>



# Case study: combining disparate data sources

- Enhanced risk models that utilize non-traditional data sources
- FDIC Tech Sprints:  
<https://www.fdic.gov/fditech/techsprints.html>

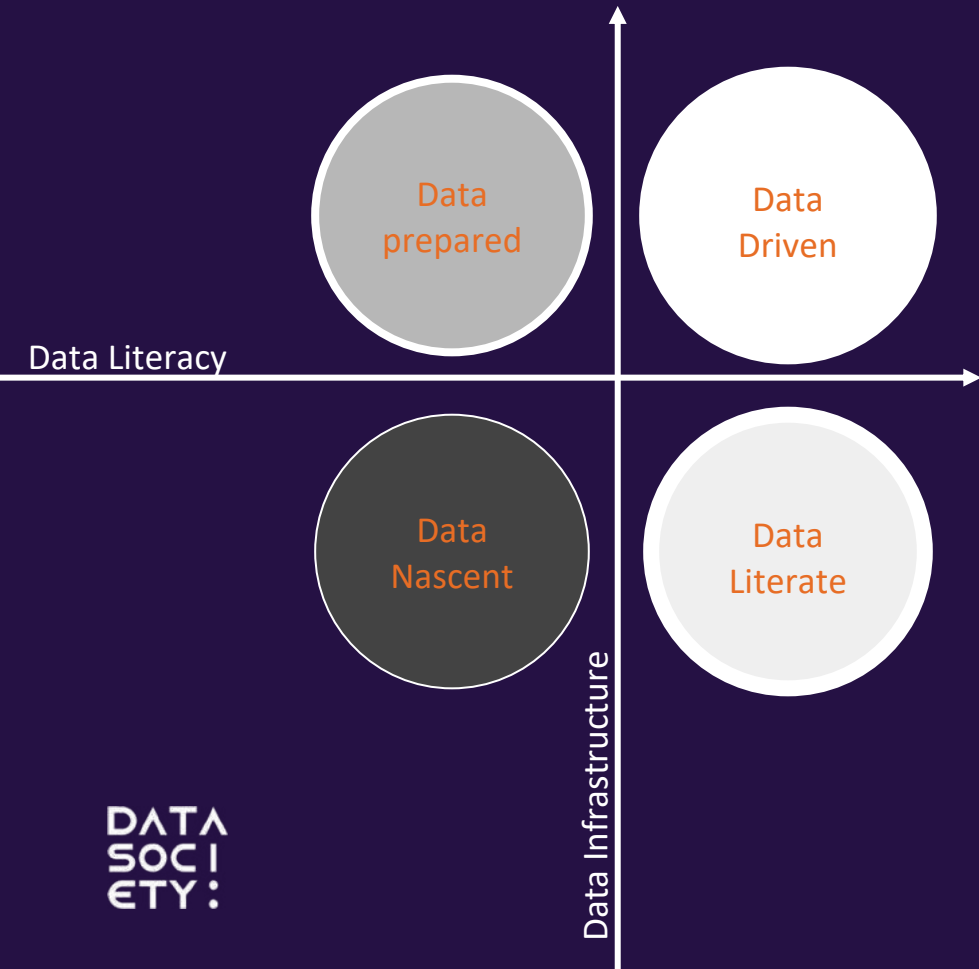


# Discussion

What kinds of data are you using today?



# Ask yourself: where do you sit?



- A data-driven culture incorporates data and analysis into its policy decisions, systems, and processes
- It can be separated into two main categories:
  - Data **infrastructure**
  - Data **literacy**

## Data Infrastructure



**Data Collection**  
Is data collected in a  
timely and clean way?



**Data Storage**  
Is the data stored  
securely with a  
backup?



**Data Access**  
Can staff access data  
easily and in a timely  
manner



**Data Leadership**  
Do executives  
champion data  
utilization?



**Data Governance**  
Are staff aware of data  
standards and  
practices?



**Data Knowledge**  
Does staff understand  
how to ask questions of  
data?

# How can you start with data analytics?



## Ask

Make it specific and measurable



## Inventory

Find the information that you have access to



## Collaborate

Talk to your data team about how they can help

# How can you support data literacy?

- 1. Bring in external / internal experts for “lunch and learns”
- 1. Attend and send team members / colleagues to data conferences
- 1. Set up training opportunities based on identified skills gaps
- 1. Plan an event, such as a data competition



# Discussion

**What question do you want to explore with data analytics?**





# Predicting pregnancy: ethical considerations

## How Companies Learn Your Secrets

By CHARLES DUHIGG FEB. 16, 2012



Antonio Bolfo/Reportage for The New York Times

# How can we mitigate these risks?

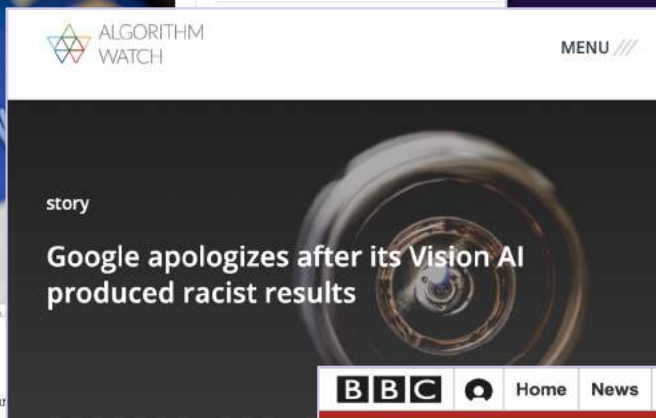


The Dept. of Housing and Urban Development charged Facebook March 28 with violating the Fair Housing Act.

By Tracy Jan and Elizabeth Dwoskin

March 28, 2019 at 6:59 p.m. EDT

The Trump administration delivered its first sanction of a tech giant Thursday, accusing Facebook of housing discrimination in a move that could threaten the way the industry

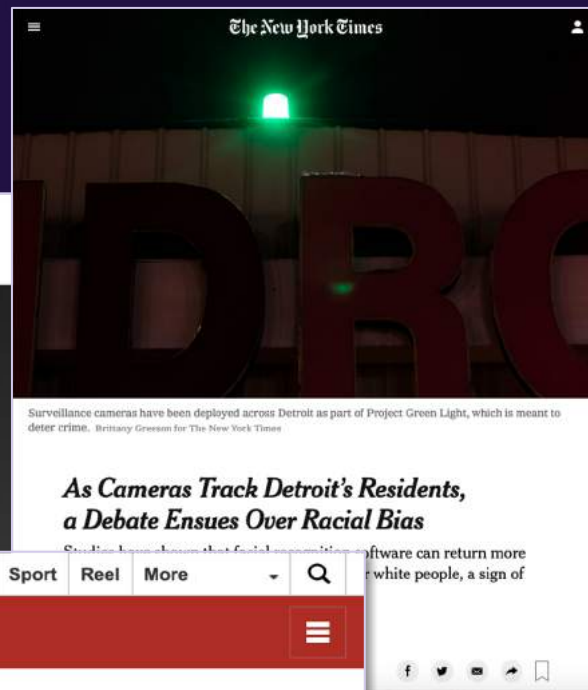


Published: April 7, 2020

Category: [story](#)

By Nicolas Kayser-Bril • [nkb@](#)

A Google service that automates image recognition produced starkly different results depending on the user's image. The company fixed the problem, but the issue is much broader.

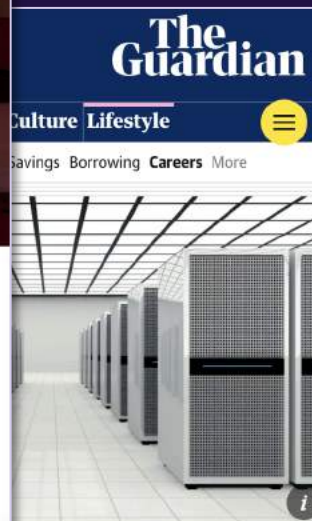


Software can return more accurate results for white people, a sign of

inequality

## Rise of the racist robots - how AI is learning all our worst impulses

There is a saying in computer science: garbage in, garbage out. When we feed machines data that reflects our prejudices, they mimic them - from antisemitic chatbots to racially biased software. Does a horrifying future await people forced to live at the mercy of algorithms?



# Case study: beyond infrastructure

- Discover Financial Services (DFS) needed to onboard new employees so they could effectively navigate the company's systems and data tools
- 40-hour data science training program for employees using R, SQL, and Snowflake, aimed at upskilling employees and working to fulfill the data-centered vision
- By implementing this program to improve data literacy, their new hires were able to increase the impact of Discover's revamped infrastructure



# Discussion

**What is one step you will implement to become more data-driven?**







**Accelerate your  
innovation today**

**DATASOCIETY:**

<https://datasociety.com>

[merav@datasociety.com](mailto:merav@datasociety.com)

[dmitri@datasociety.com](mailto:dmitri@datasociety.com)