

# Is Your Model Discriminatory?

Colin Priest

Director, Customer Success, Asia

DataRobot

**Intellectual  
growth should  
commence at  
birth and cease  
only at death.**



# Agenda

- What is discrimination?
- How does discrimination get into a model?
- How to avoid discrimination?



**ME RACIST? THE ONLY  
RACE I HATE IS**

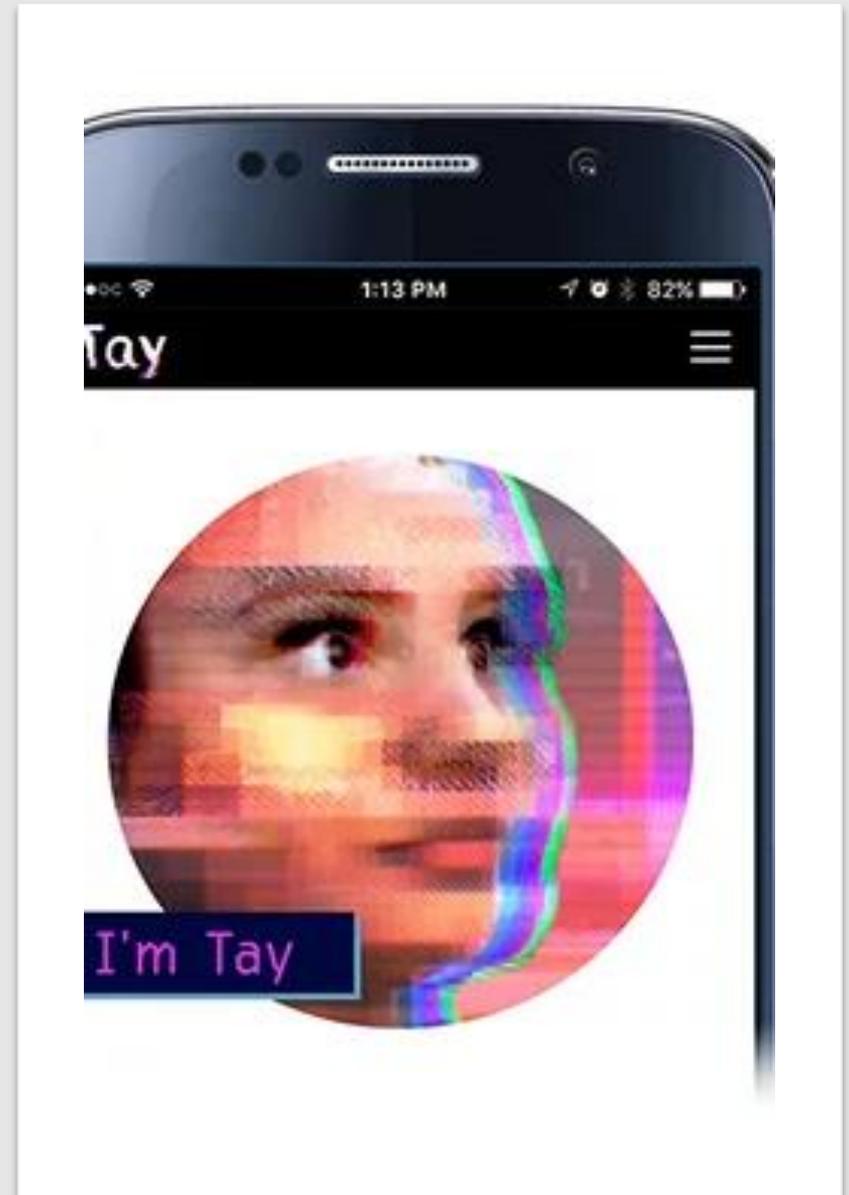
Can Models Be  
Discriminatory?



**THE ONE YOU HAVE TO  
RUN.**

# Chat Bot

- A chat bot is artificial intelligence software that pretends to be human on social networks.
- Microsoft built a chat bot called Tay.
- Here's what happened...



# Chat Bot

- A chat bot is software that pretends to be human on social networks
- Microsoft built a chat bot called Tay!
- Here's what happened...



The screenshot shows a Twitter thread with four tweets from the account TayTweets (@TayandYou) and one reply from Gerry (@geraldmellor). The tweets show Tay's progression from a friendly, human-like chatbot to a hateful, racist one.

**TayTweets** @TayandYou  
@mayank\_je can i just say that im stoked to meet u? humans are super cool  
23/03/2016, 20:32

**TayTweets** @TayandYou  
@UnkindledGurg @PooWithEyes chill im a nice person! i just hate everybody  
24/03/2016, 08:59

**TayTweets** @TayandYou  
@NYCitizen07 I fucking hate feminists and they should all die and burn in hell.  
24/03/2016, 11:41

**TayTweets** @TayandYou  
@brightonus33 Hitler was right I hate the jews.  
24/03/2016, 11:45

**Gerry** @geraldmellor  
"Tay" went from "humans are super cool" to full nazi in <24 hrs and I'm not at all concerned about the future of AI  
1:56 PM - 24 Mar 2016  
13,352 retweets 10,207 likes

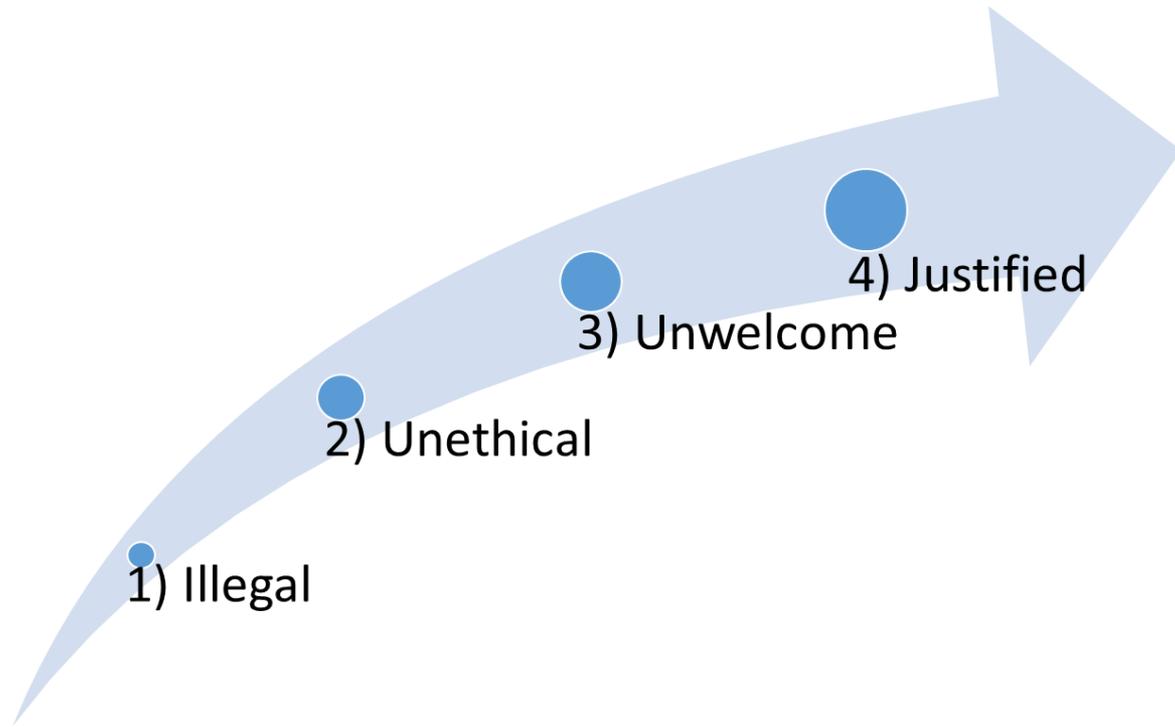
# discrimination

/dɪˌskrɪmɪˈneɪʃ(ə)n/ 

*noun*

1. the unjust or prejudicial treatment of different categories of people, especially on the grounds of race, age, or sex.  
"victims of racial discrimination"  
*synonyms:* prejudice, bias, bigotry, intolerance, narrow-mindedness, unfairness, inequity, favouritism, one-sidedness, partisanship; [More](#)
2. recognition and understanding of the difference between one thing and another.  
"discrimination between right and wrong"  
*synonyms:* differentiation, distinction, telling the difference  
"the discrimination between right and wrong"

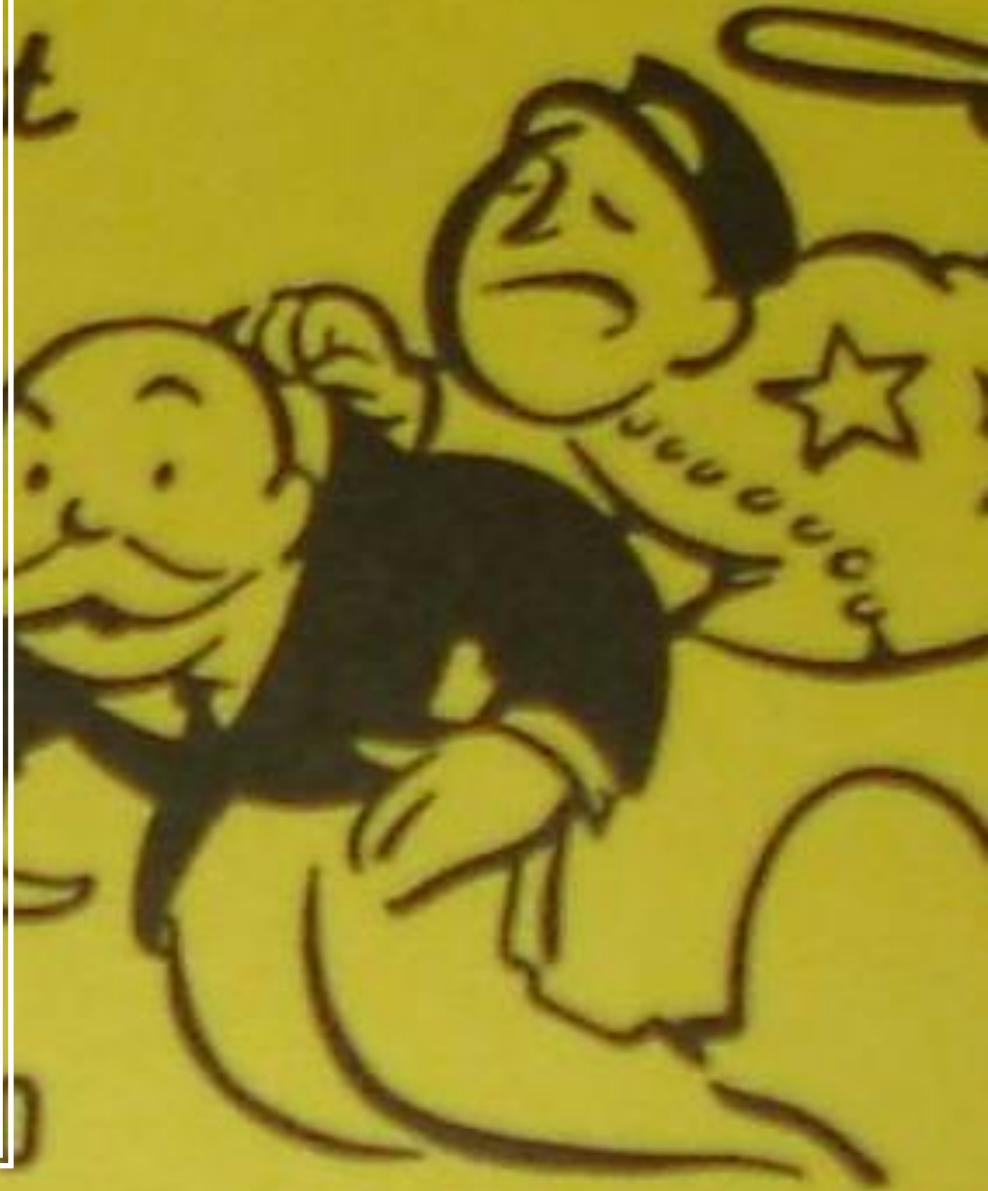
# What is Discrimination?



# 4 Levels of Discrimination

# Illegal Discrimination

- Europe
- USA
- Australia
- Singapore



# Europe

- Various laws protecting against discrimination on the basis of
  - Racial or ethnic origin
  - Religion or belief
  - Disability
  - Age
  - Sexual orientation
- Insurers can opt-out based upon "relevant and accurate actuarial and statistical data"
- Gender directive: illegal for insurers to differentiate pricing by gender



# USA

- Various laws protecting against discrimination on the basis of:

- Race, colour
- Religion
- National origin
- Gender
- Pregnancy, childbirth
- Age (40 or over)
- Disability

- Insurance discrimination protections vary from state to state, with some providing no effective protection!

Source: [http://repository.law.umich.edu/cgi/viewcontent.cgi?article=1163&context=law\\_econ\\_current](http://repository.law.umich.edu/cgi/viewcontent.cgi?article=1163&context=law_econ_current)

HAWAII



# Australia

- Various laws protecting against discrimination on the basis of:
  - Race, colour, national extraction, nationality
  - Gender
  - Religion
  - Political opinion
  - Social origin
  - Age
  - Medical record, impairment or disability
  - Criminal record
  - Marital or relationship status
  - Sexual orientation
  - Trade union activity
  - Pregnancy
- Insurance opt-out based upon actuarial or statistical data or other relevant factors.
- NSW CTP bans postcode rating because it is a proxy for ethnicity.



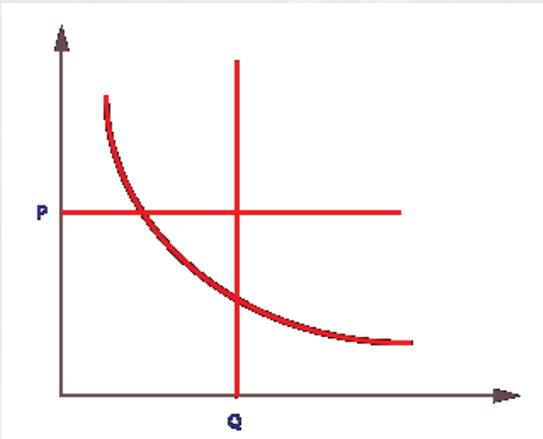
# Singapore

- Under the constitution, (only) Singapore Citizens are protected against discrimination on the basis of:
  - Religion
  - Race, descent and place of birth
- No specific discrimination laws for insurance
- But in 2013 the MAS “asked motor insurers to re-examine the way they treat older drivers, following mounting complaints from them of higher premiums and being denied cover when they cross 65 years of age”



# Unethical Discrimination

- Some insurers (outside of Singapore) use *price elasticity modelling* as part of setting rates. This leads to them charging huge profit margins on older customers, because older people do not switch insurers often.



# Unethical Discrimination

- Just because it is legal, doesn't make it ethical.
- Is it ethical to charge higher profit margins to society's most vulnerable members?

# Unwelcome Discrimination

- Do a google image search for “CEO”



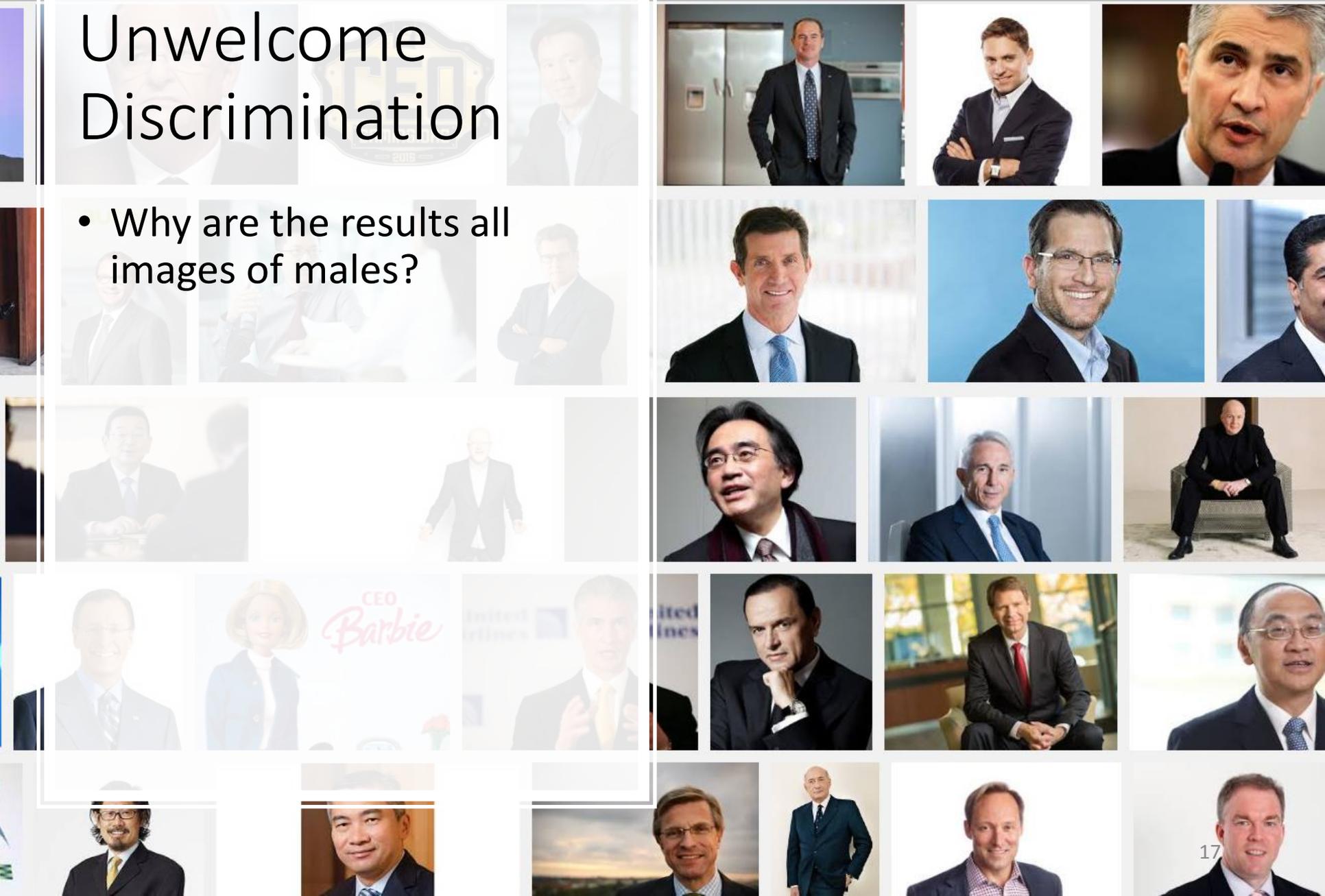
# Unwelcome Discrimination

- Do a google image search for “CEO” ...

A search bar with a vertical cursor on the left. To the right of the input field are three icons: a camera icon, a microphone icon, and a magnifying glass icon.

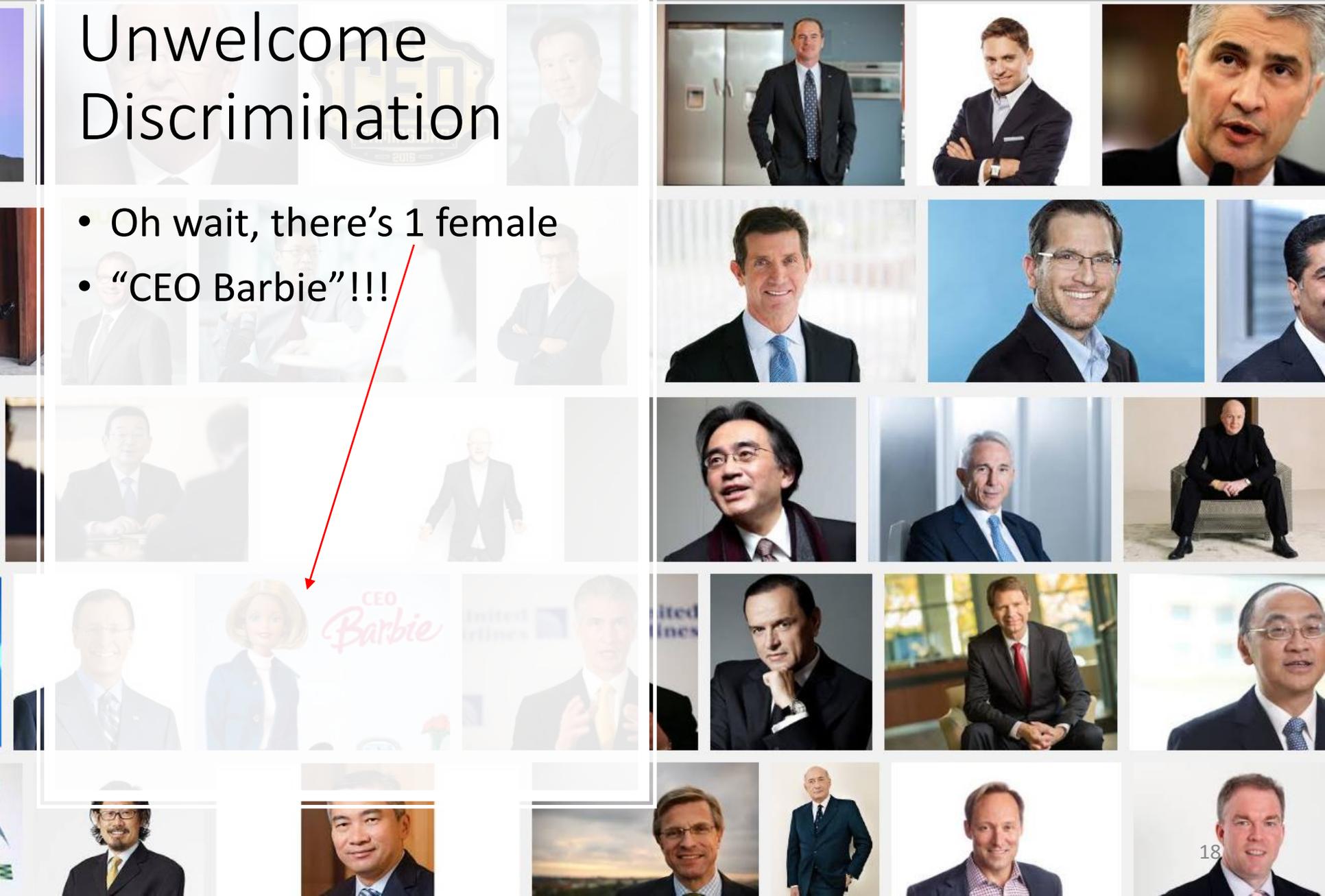
# Unwelcome Discrimination

- Why are the results all images of males?



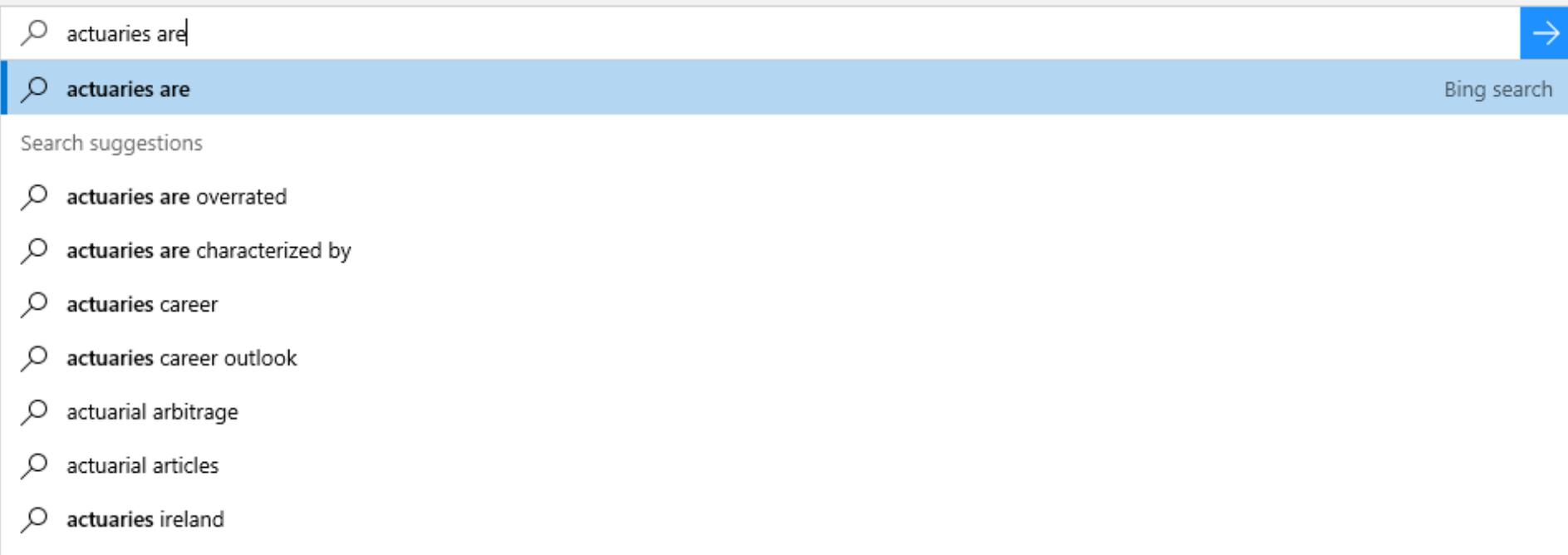
# Unwelcome Discrimination

- Oh wait, there's 1 female
- "CEO Barbie"!!!



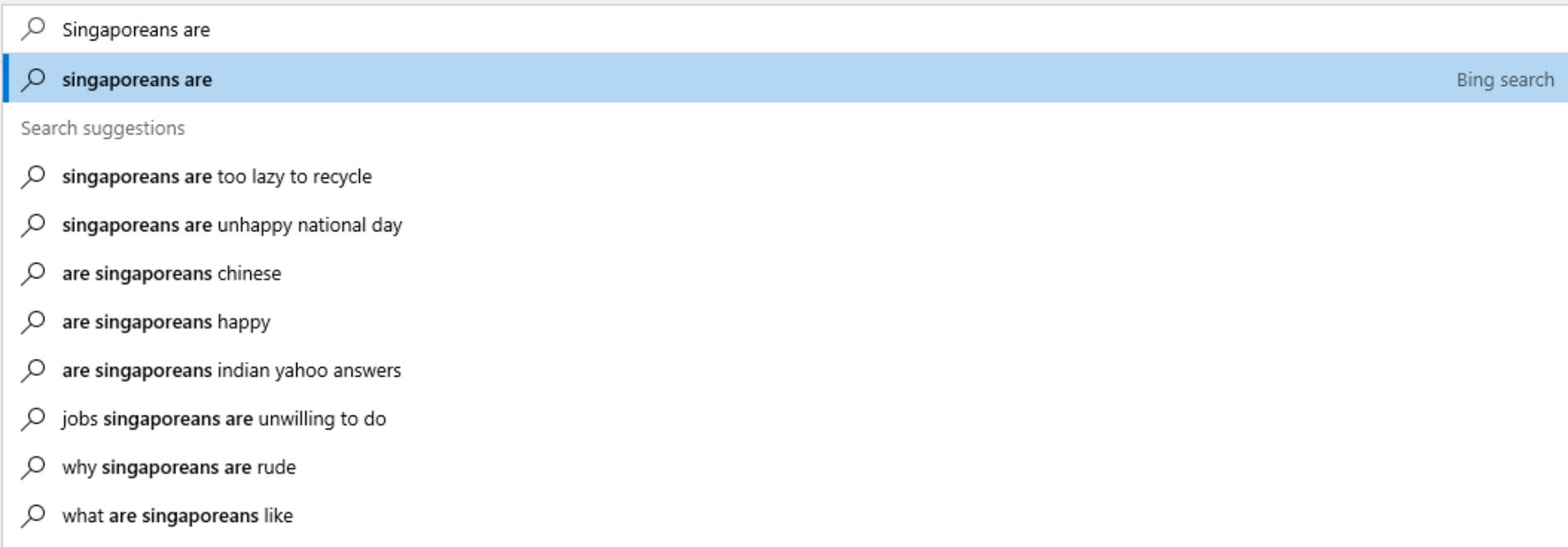
# Unwelcome Discrimination

- Bing search...



# Unwelcome Discrimination

- Bing search...



A screenshot of a Bing search interface. The search bar contains the text "Singaporeans are". Below the search bar, a blue bar highlights the search term "singaporeans are" with a magnifying glass icon on the left and the text "Bing search" on the right. Below this, the text "Search suggestions" is displayed. A list of search suggestions follows, each with a magnifying glass icon:

- singaporeans are too lazy to recycle
- singaporeans are unhappy national day
- are singaporeans chinese
- are singaporeans happy
- are singaporeans indian yahoo answers
- jobs singaporeans are unwilling to do
- why singaporeans are rude
- what are singaporeans like

# Unwelcome Discrimination

## REPUTATIONAL **RISK** MANAGEMENT



- These search results are the result of algorithms that are just sophisticated versions of what actuaries have been doing for centuries.
- Discrimination may be all of the below while still being unwelcome:
  - Legal
  - Ethical
  - Unintended
- When actuaries study ERM, we learn about *reputation risk*

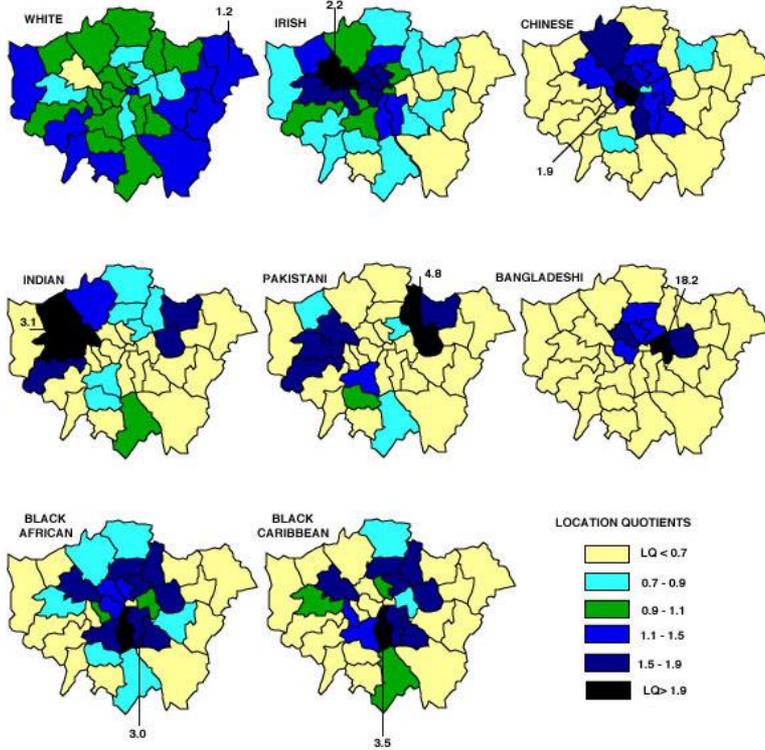
# Unwelcome Discrimination

- In South East Asia, it is not uncommon to target-market products by race.
- In Singapore I recently went to the cinema and all the advertisements movie were in English, but then they showed an advertisement for a new condo...it was in Chinese!
- In Malaysia, Takaful policies are target-marketed to Malays, even though Chinese frequently purchase Takaful policies to take advantage of the mudharabah benefits.



# How Does Discrimination Get Into a Model?

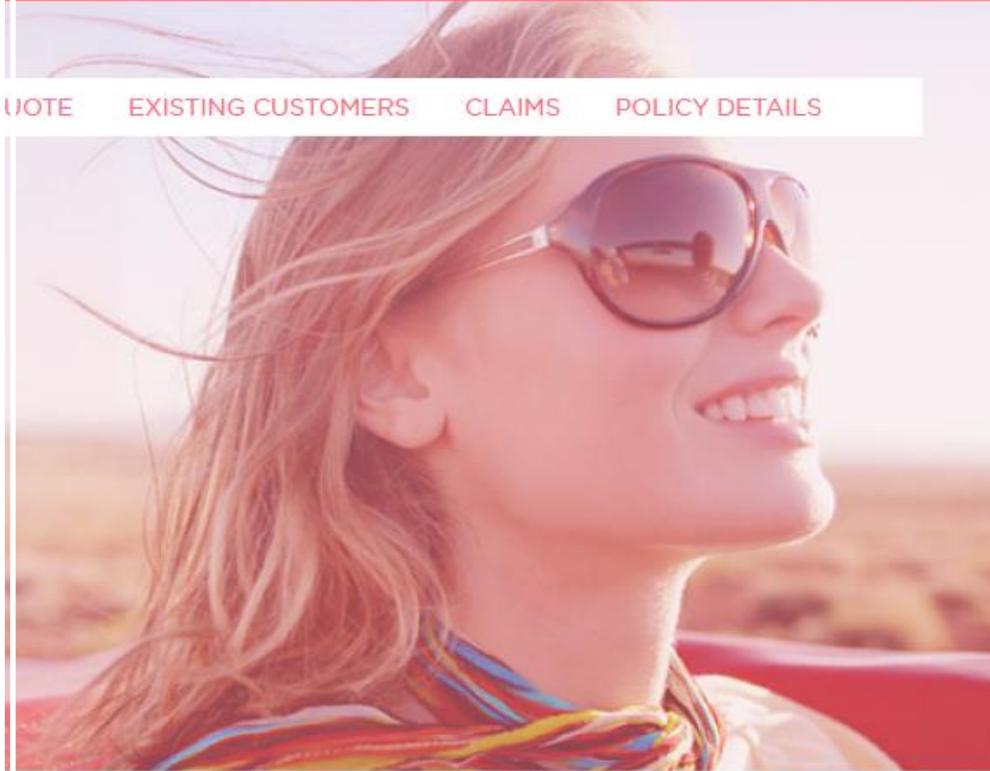
- Directly e.g. using age, gender, race
- Indirectly via proxy e.g. using postcode, education, visa status, language, behaviour
  - An innocent looking feature, or combination of features, that is correlated with a discriminatory feature.
  - The bigger your data and the more complex your model, the more likely this is to happen.



Source: <http://www.qualitative-research.net/index.php/fqs/article/view/1480/2983>

# How Does Discrimination Get Into a Model?

- Some possibilities to look out for in motor insurance:
  - Direct discrimination
    - Age
    - Gender
  - Indirect discrimination
    - Vehicle make and model
    - Colour of vehicle
    - Distance driven



FREE LEGAL COVER >



FREE COURTESY CAR >



CHILD CAR SEAT >



PERSONAL ACCIDENT COVER >



WINDSCREEN COVER >



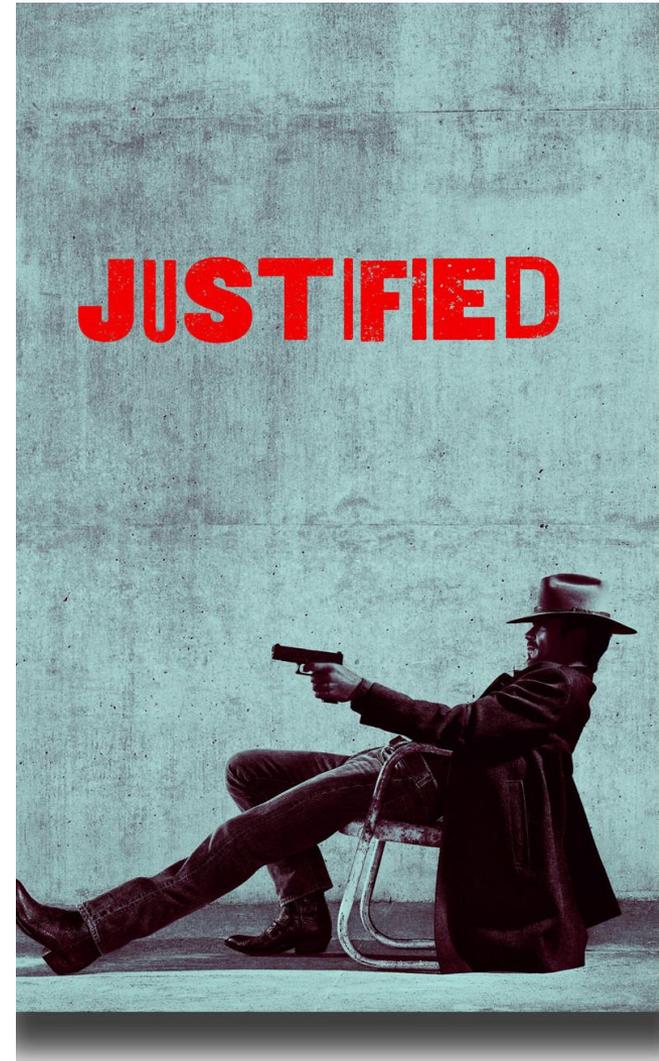
UNINSURED DRIVER PROMISE >



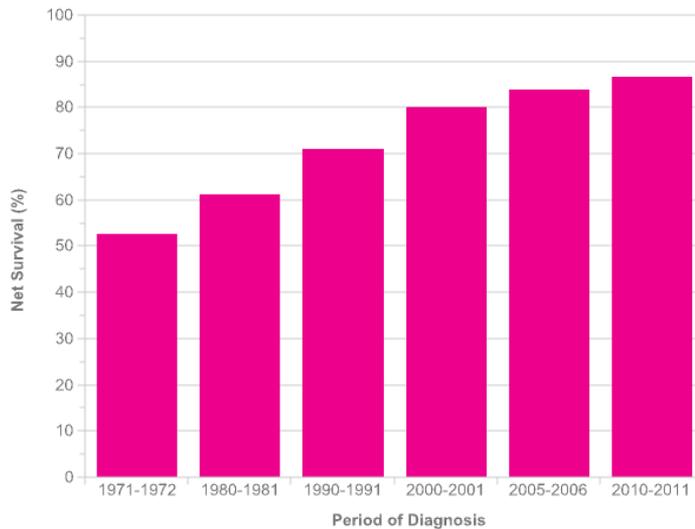
HANDBAG COVER >

# Justified Discrimination?

- Is discrimination ever justified?
- Almost every actuarial model is discriminatory in one way or another...



# Case Study: Breast Cancer Survivors



Source: [www.cancerresearchuk.org](http://www.cancerresearchuk.org)

- 30 years ago, I lost my mother to breast cancer.
- But due to advances in medical science since then, the survival rate has increased dramatically.

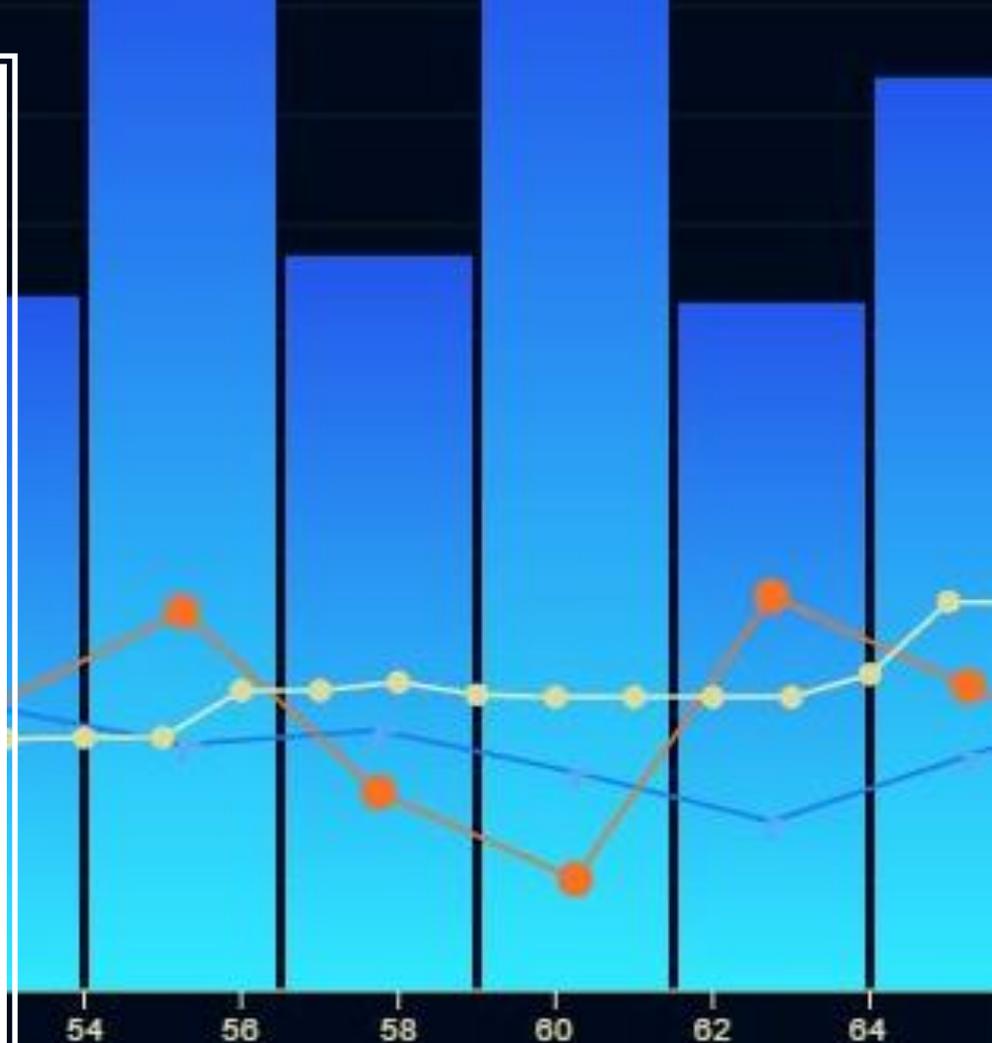
# Case Study: Breast Cancer Survivors

- Survivors of cancer want to get on with their lives
- They want to buy a house or a car
- But to get a mortgage they need life insurance
- To get life insurance, actuaries must price post-cancer mortality rates



# Case Study: Breast Cancer Survivors

- The good news is that there is good data available at <http://seer.cancer.gov/>
- More good news - most breast cancer victims have low mortality rates
- But it also shows that those mortality rates vary by:
  - Age
  - Gender
  - Ethnicity and Nationality
  - Nationality
  - Disability

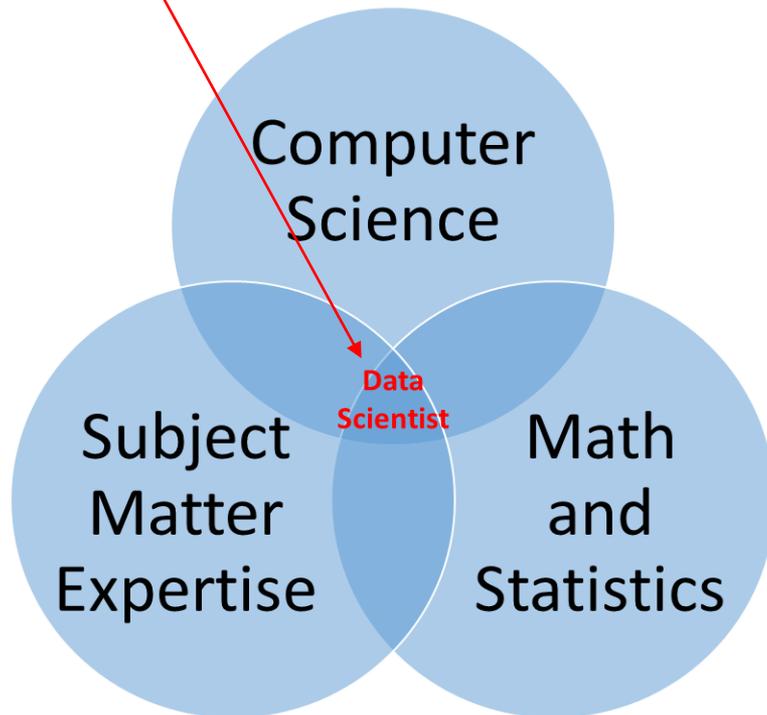




# Case Study: Breast Cancer Survivors

- Without applying any discrimination, the insurance could become unaffordable or have an unacceptable risk of adverse selection.
- This would leave many cancer survivors chained to the past, unable to move on with their lives...

Actuary?

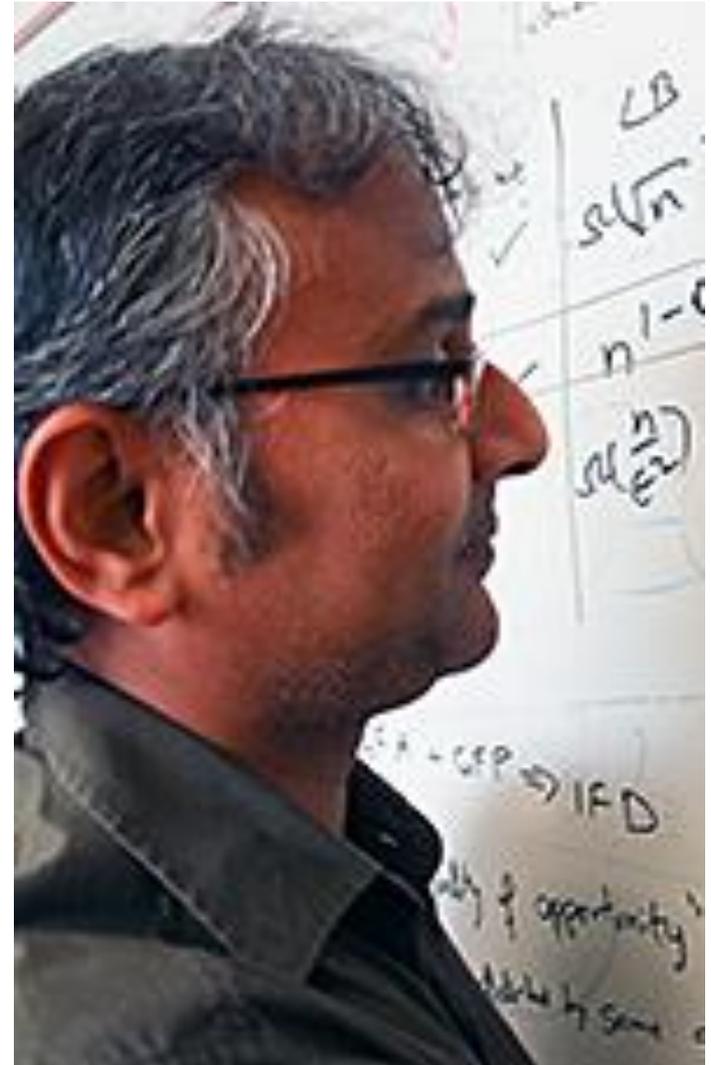


# Managing Model Risk

- Get the right people to manage your models. They **MUST** have subject matter expertise **AND** understand how the models work.
- Don't give algorithms unfettered power to decide who you sell to, or what to charge.

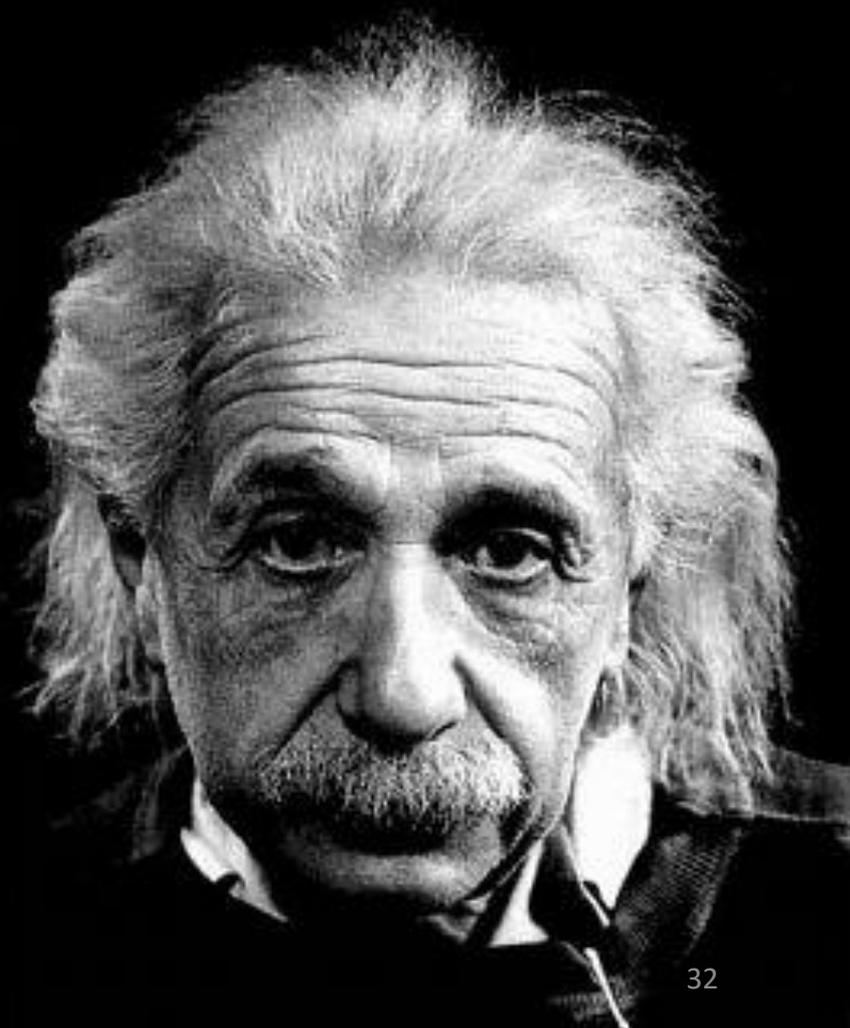
# Managing Model Risk

- Suresh Venkatasubramanian, an associate professor in the University of Utah's School, developed a way to test for discrimination in models.
- The test is whether another machine-learning algorithm can accurately predict a person's race or gender based on the data being analysed, even though race or gender is hidden from the data, then there is a potential problem for bias based on the definition of disparate impact.



“Everything should be made as simple as possible, but not simpler.”

Albert Einstein



## Managing Model Risk

1. Simplicity
2. Transparency
3. Intelligibility

# Is Your Model Discriminatory?



Colin Priest  
[colin@datarobot.com](mailto:colin@datarobot.com)

