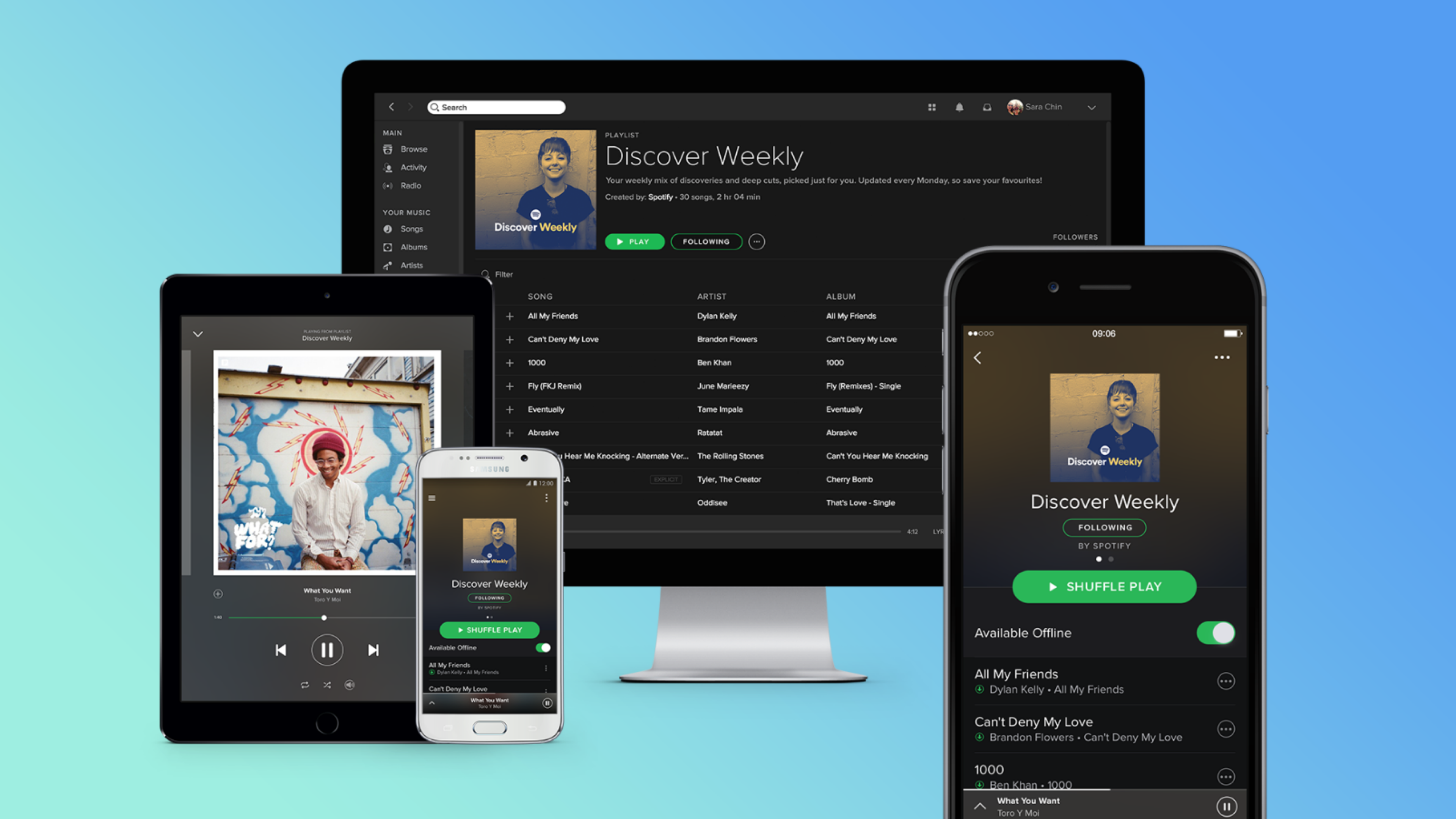
Forecasting
@Spotify

Yorgos Askalidis
Sr. Data Scientist, Growth Forecasting

rev 2
May 23rd, 2019, NYC


Search
Sara Chin
MAIN
Browse
Activity
Radio
YOUR MUSIC
Songs
Albums
Artists
Discover Weekly
PLAYLIST
Discover Weekly
Your weekly mix of discoveries and deep cuts, picked just for you. Updated every Monday, so save your favourites!
Created by: Spotify • 30 songs, 2 hr 04 min
PLAY
FOLLOWING
FOLLOWERS
Filter
SONG
ARTIST
ALBUM
All My Friends
Dylan Kelly
All My Friends
Can't Deny My Love
Brandon Flowers
Can't Deny My Love
1000
Ben Khan
1000
Fly (FKJ Remix)
June Marieezy
Fly (Remixes) - Single
Eventually
Tame Impala
Eventually
Abrasive
Ratatat
Abrasive
u Hear Me Knocking - Alternate Ver...
The Rolling Stones
Can't You Hear Me Knocking
Tyler, The Creator
Cherry Bomb
Oddisee
That's Love - Single
Discover Weekly
What You Want
Toro Y Moi
Discover Weekly
FOLLOWING
SHUFFLE PLAY
Available Offline
All My Friends
Can't Deny My Love
What You Want
09:06
Discover Weekly
Discover Weekly
FOLLOWING
BY SPOTIFY
SHUFFLE PLAY
Available Offline
All My Friends
Dylan Kelly • All My Friends
Can't Deny My Love
Brandon Flowers • Can't Deny My Love
1000
Ben Khan • 1000
What You Want
Toro Y Moi

Forecasting is an important part of a business

Full Year 2019 Guidance:

- **Total MAUs**: 245-265 million, up 18-28% Y/Y
- **Total Premium Subscribers:** 117-127 million, up 21-32% Y/Y
- **Total Revenue:** €6.35-€6.8 billion, up 21-29% Y/Y
- **Gross Margin:** 22.0-25.0%
- **Operating Profit/Loss:** €(180)-(€340) million

investors.spotify.com

Forecasting is an important part of a business

Full Year 2019 Guidance:

Monthly Active User

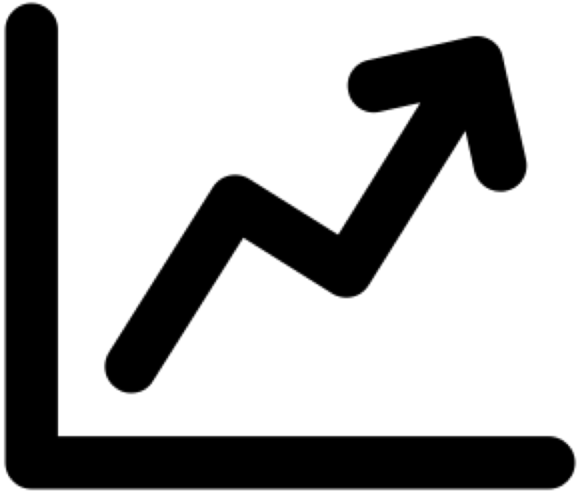
- **Total MAUs**: 245-265 million, up 18-28% Y/Y
- **Total Premium Subscribers:** 117-127 million, up 21-32% Y/Y
- **Total Revenue:** €6.35-€6.8 billion, up 21-29% Y/Y
- **Gross Margin:** 22.0-25.0%
- **Operating Profit/Loss:** €(180)-(€340) million

investors.spotify.com

Quick facts about forecasting at Spotify

Happens at the end of each quarter

Forecast two years ahead

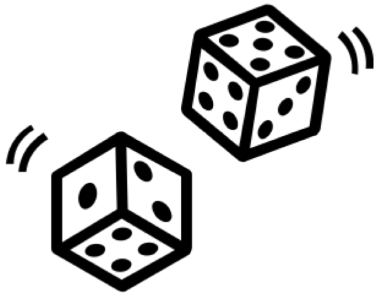


Uses past daily time-series data

Produces daily forecast

All data and graphs shown in this talk are generated by **random processes.**

No Spotify data is used.



1

Forecasting MAU for **existing** markets

Generally speaking, a market is considered "existing" if Spotify launched there at least one year ago

Prophet is a time-series forecasting package

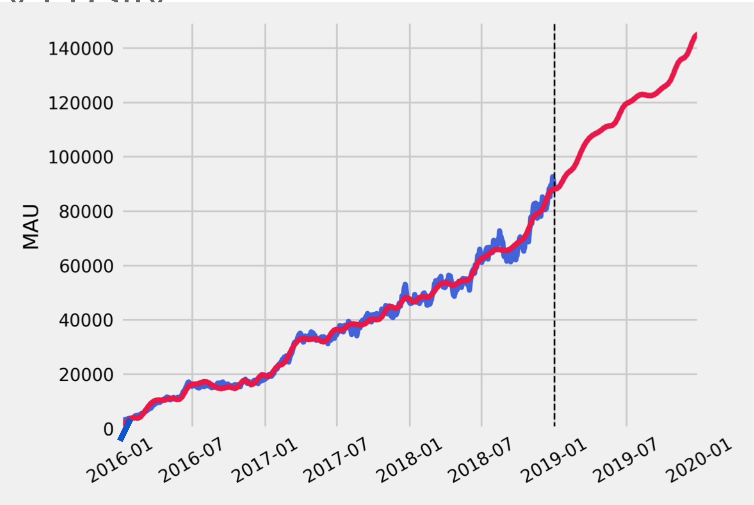
Prophet explained with gifs   Detailed documentation of prophet here.

Available for Python and R

Prophet can produce **reliable** forecast very easily

```
m = Prophet()
m.fit(df)

future = m.make_future_dataframe(365)
forecast = m.predict(future)
```



Artificially created data

Prophet is a time-series forecasting package

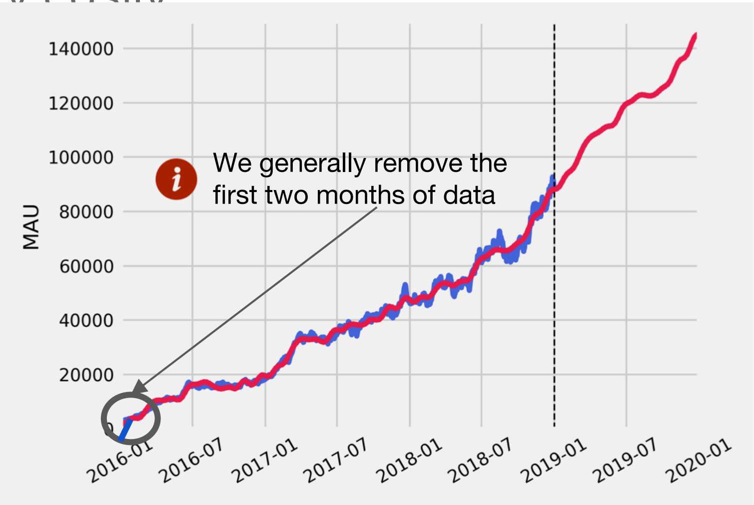
Prophet explained with gifs   Detailed documentation of prophet here.

Available for Python and R

Prophet can produce **reliable** forecast very easily

```
m = Prophet()
m.fit(df)

future = m.make_future_dataframe(365)
forecast = m.predict(future)
```



Artificially created data

Prophet is highly optimizable

Prophet has many parameters that can be optimized via grid search to produce an even more accurate forecast

Detailed documentation of prophet here
Prophet explained with gifs

We want a more granular model

Forecasting on a lower level than MAU allows us to better understand the underlying drivers of users growth.

- Are we **retaining** users better?
- Are we **activating** users better?
- Are we **reactivating** users better?

We divide MAU into cohorts based on date

Let’s consider MAU that became MAU by **activating** upon registration

We divide MAU into cohorts based on date

Let’s consider MAU that became MAU by **activating** upon registration

For each day we track the **number** of such activated registrations

We use prophet to forecast that number in the future

We divide MAU into cohorts based on date

Let's consider MAU that became MAU by **activating** upon registration

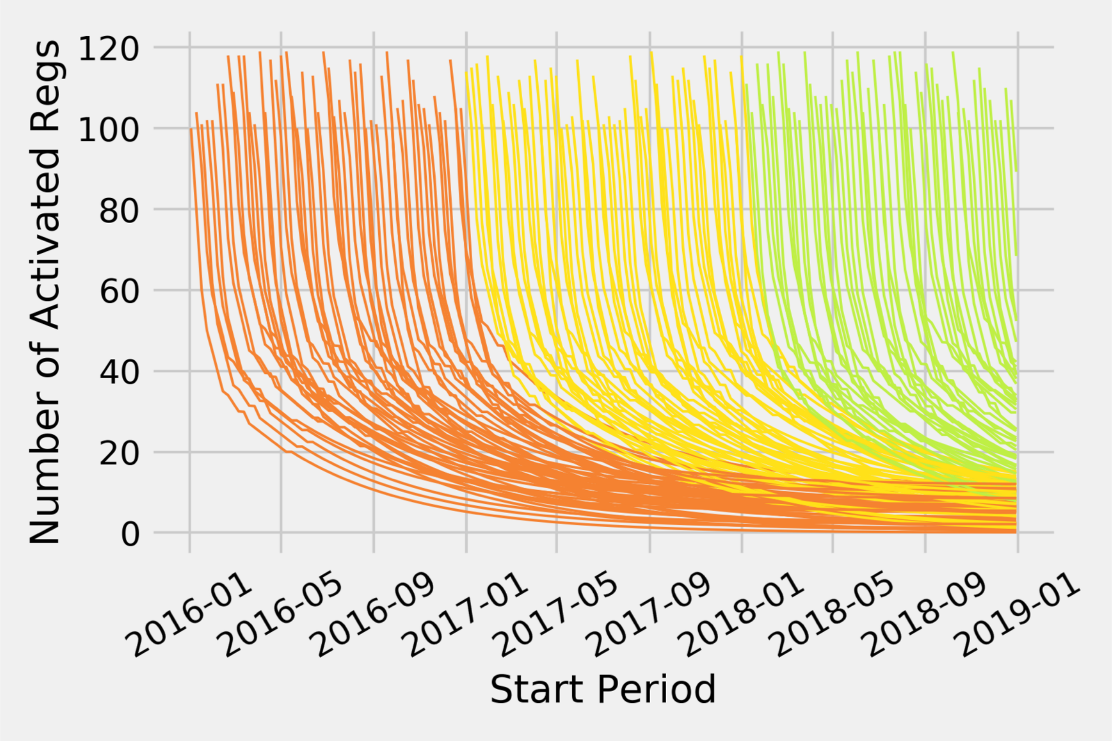
For each day we track the **number** of such activated registrations

We use prophet to forecast that number in the future

We track how many of that original cohort **remain MAU** on the following days

Note that this number can only decrease. If a user from that cohort churns and becomes MAU in a future date, that user will be counted as a reactivated user of that future date.

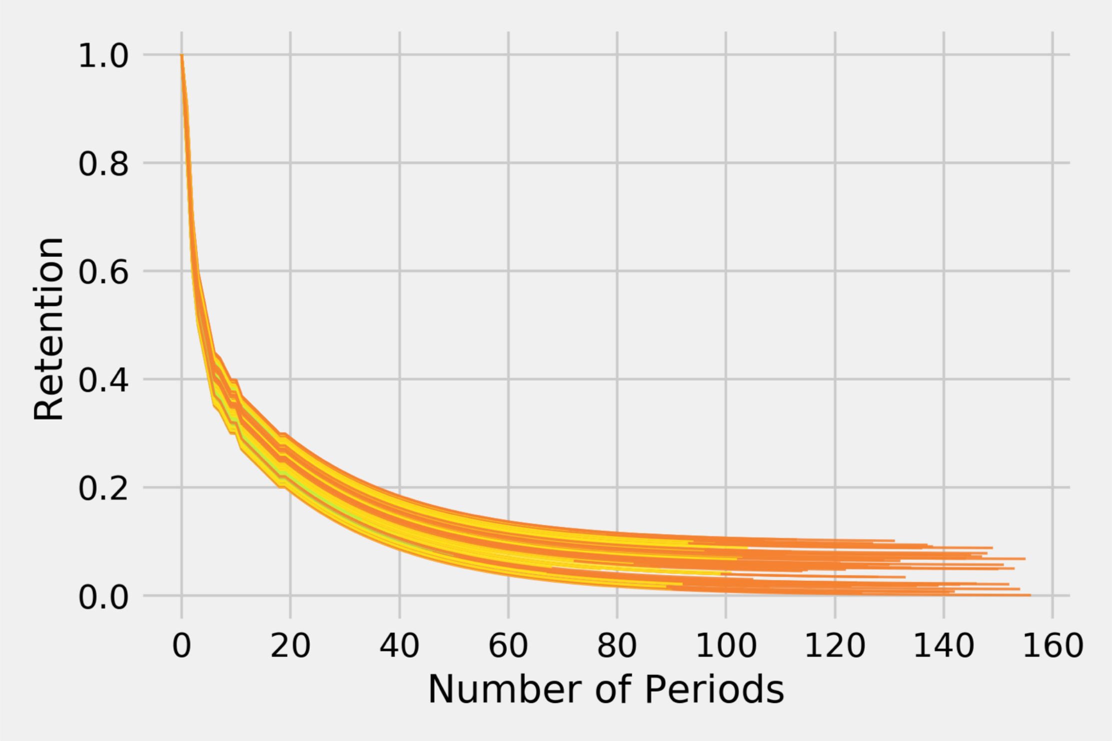
This is what retention curves can look like



Artificially generated data

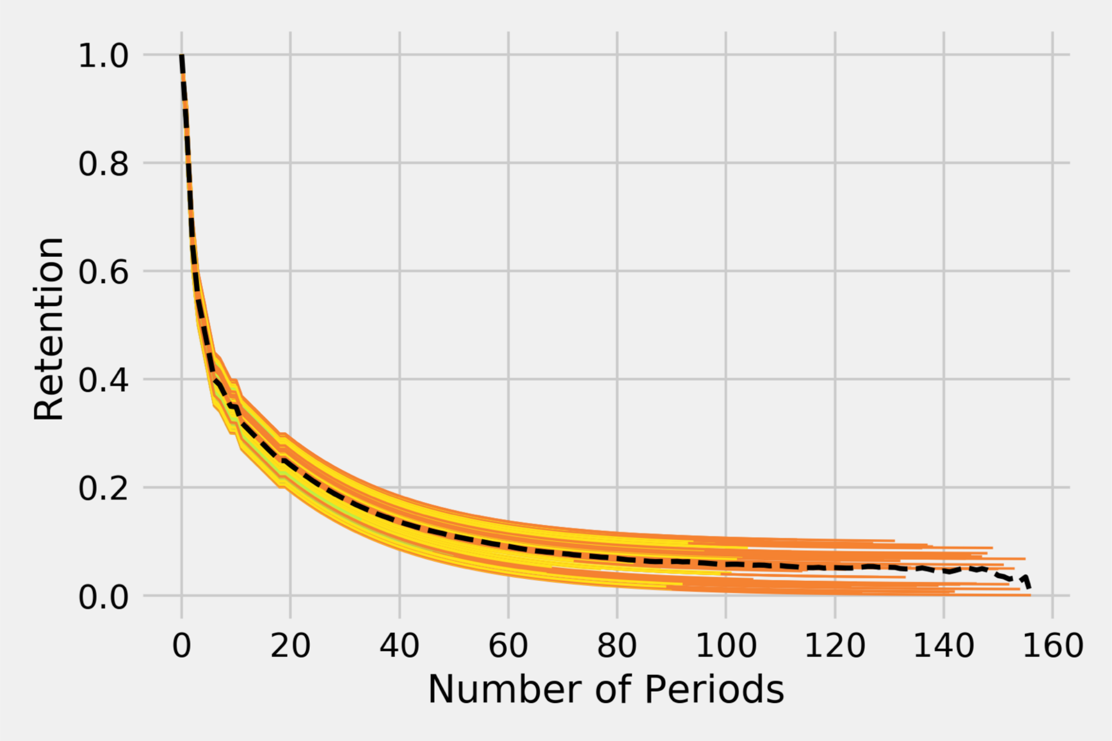
**also known as survival curves*

We normalize & study the retention curves



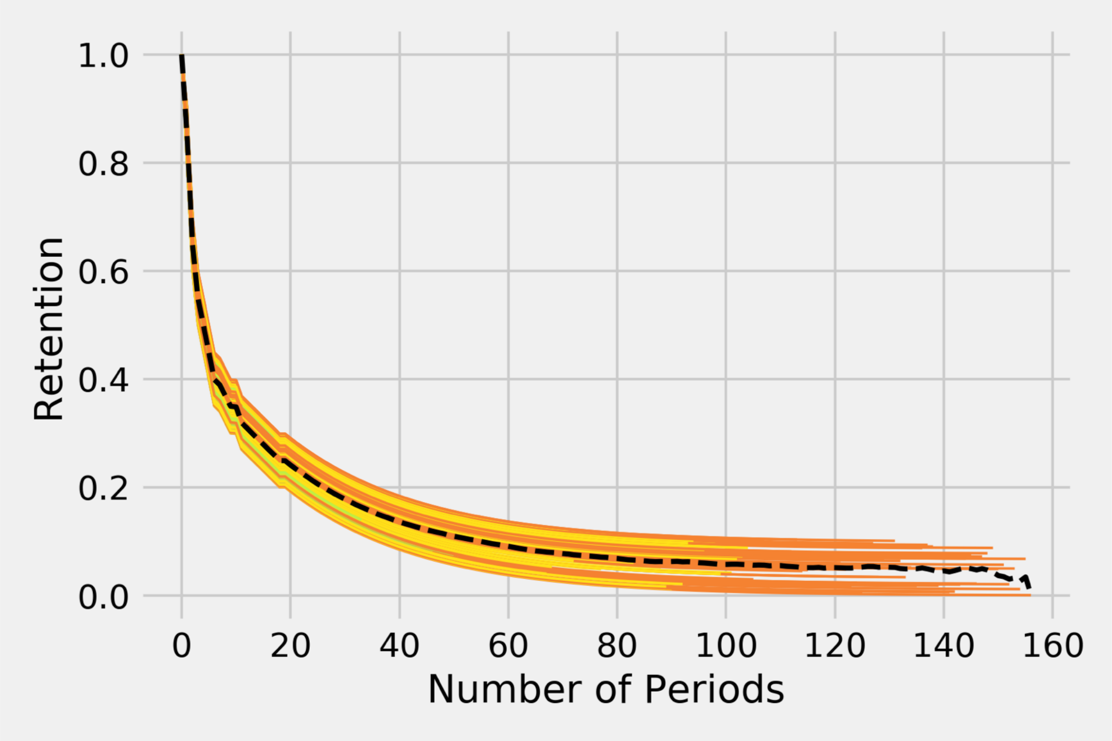
Artificially generated data

We calculate the average retention curve



Artificially generated data

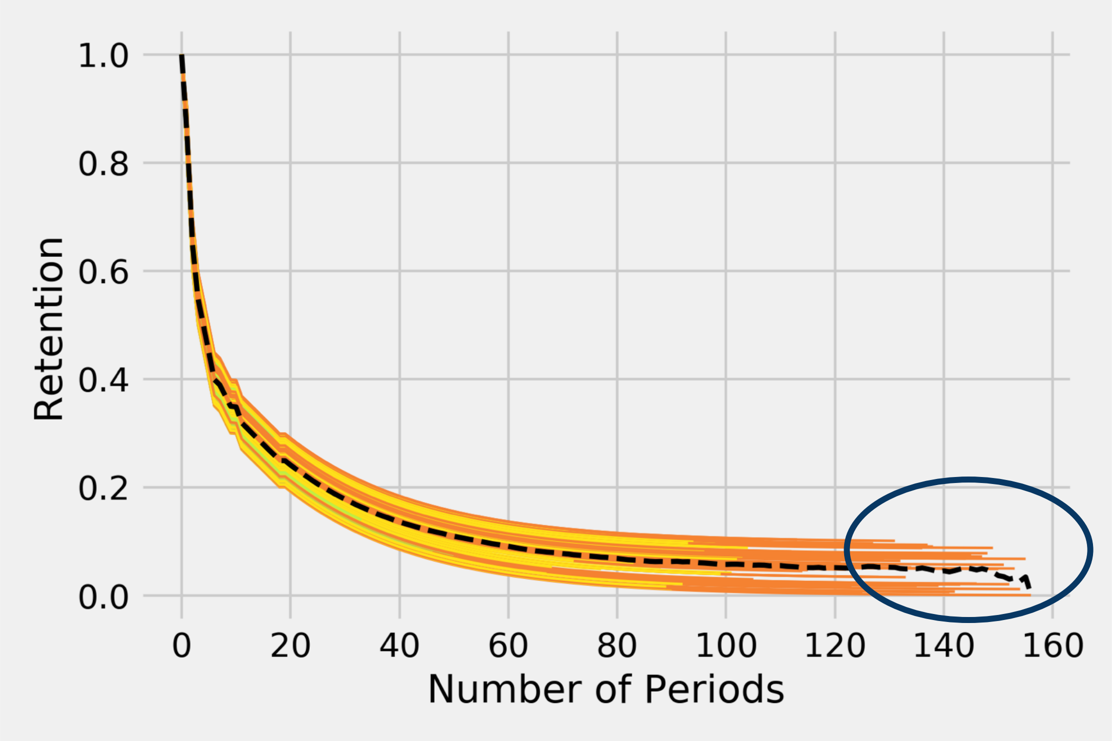
We calculate the average retention curve



When calculating the average, we can weigh various curves differently

Artificially generated data

We calculate the average retention curve

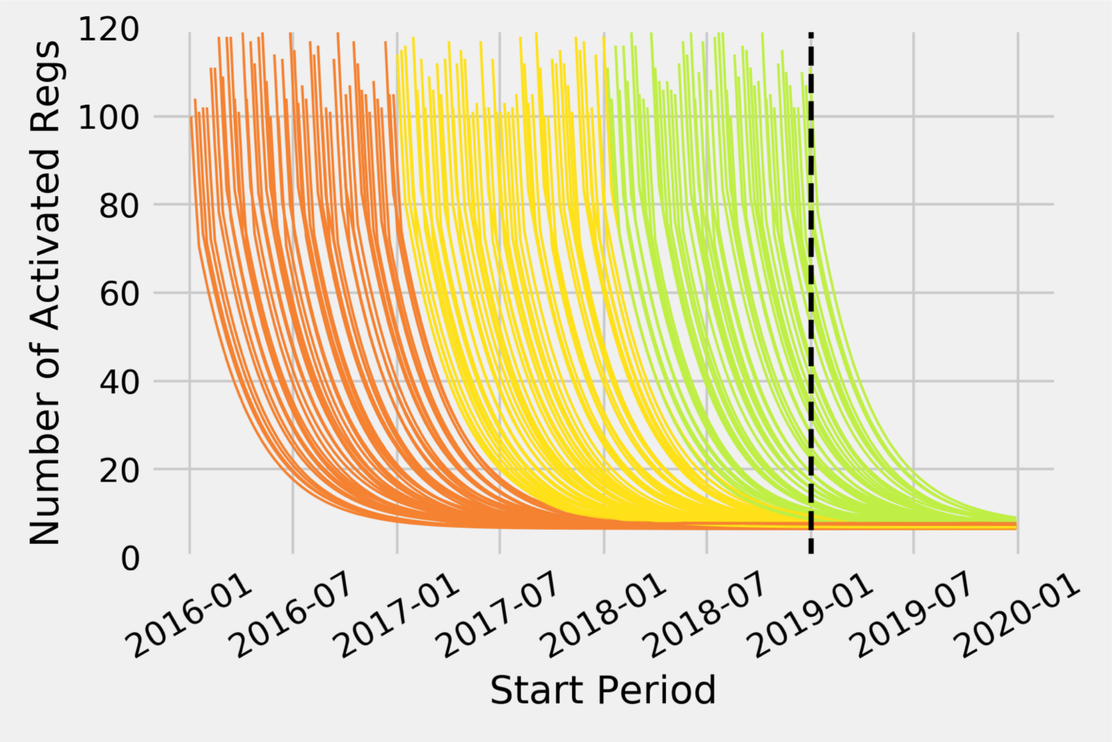


When calculating the average, we can weigh various curves differently

Each of these retention curves is non-increasing. The average should be too.

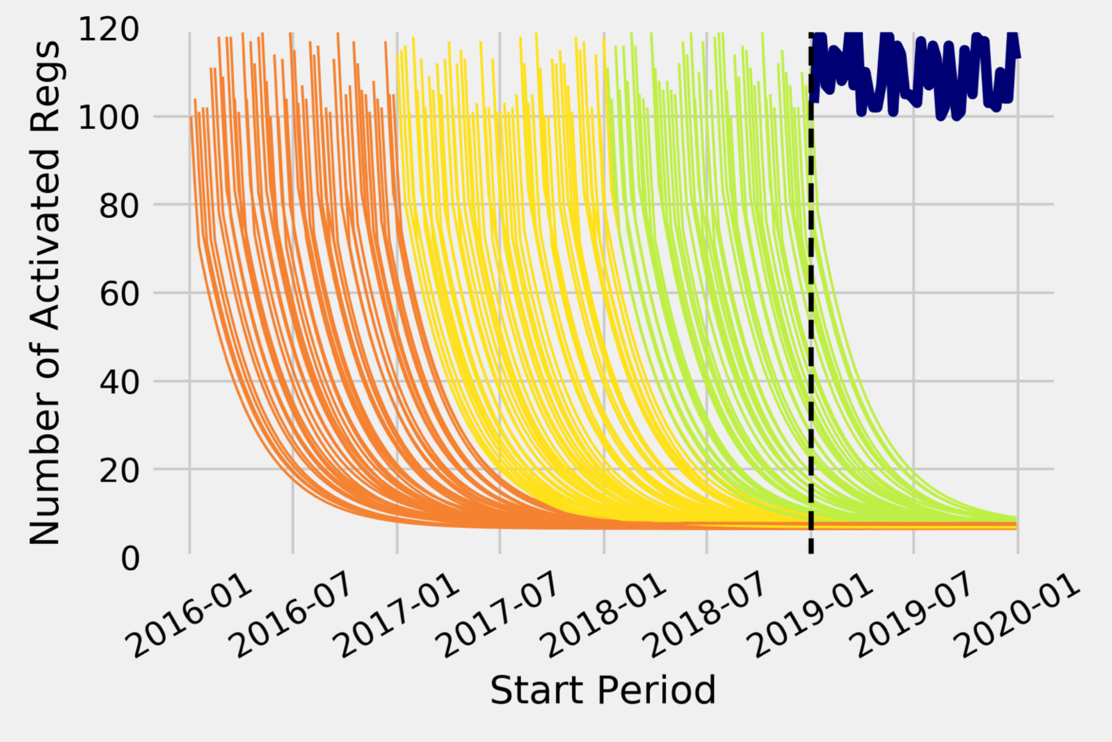
Artificially generated data

We extrapolate the historical curves



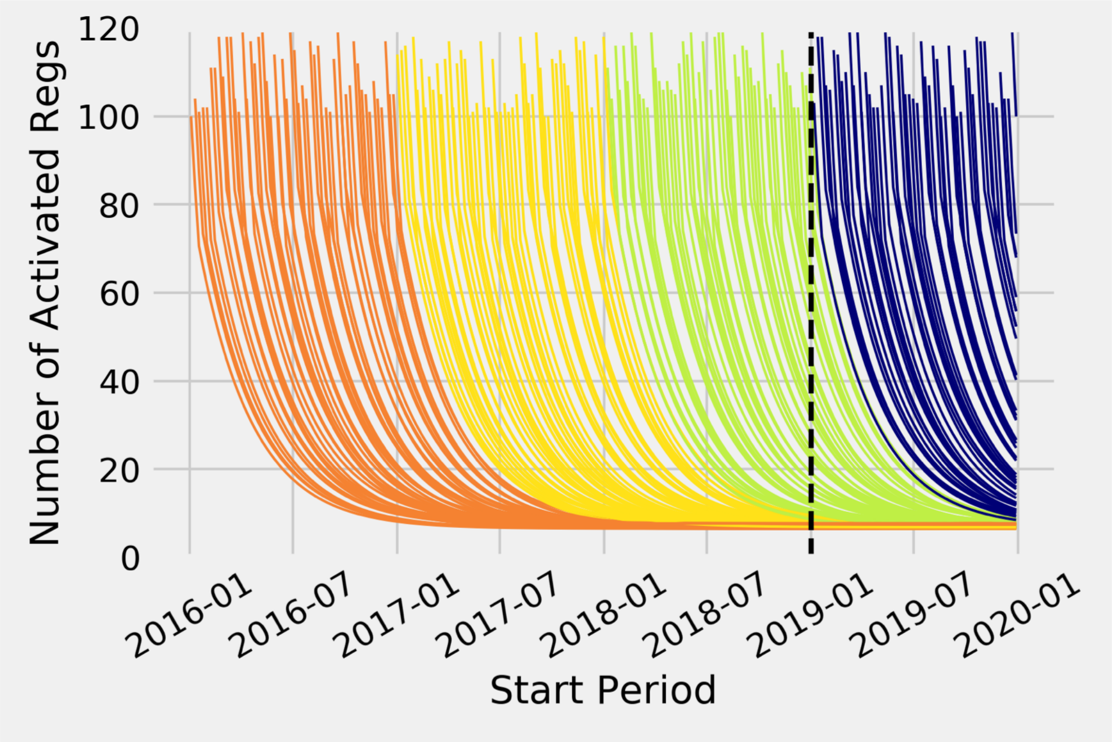
Artificially generated data

Prophet forecasts future cohort sizes



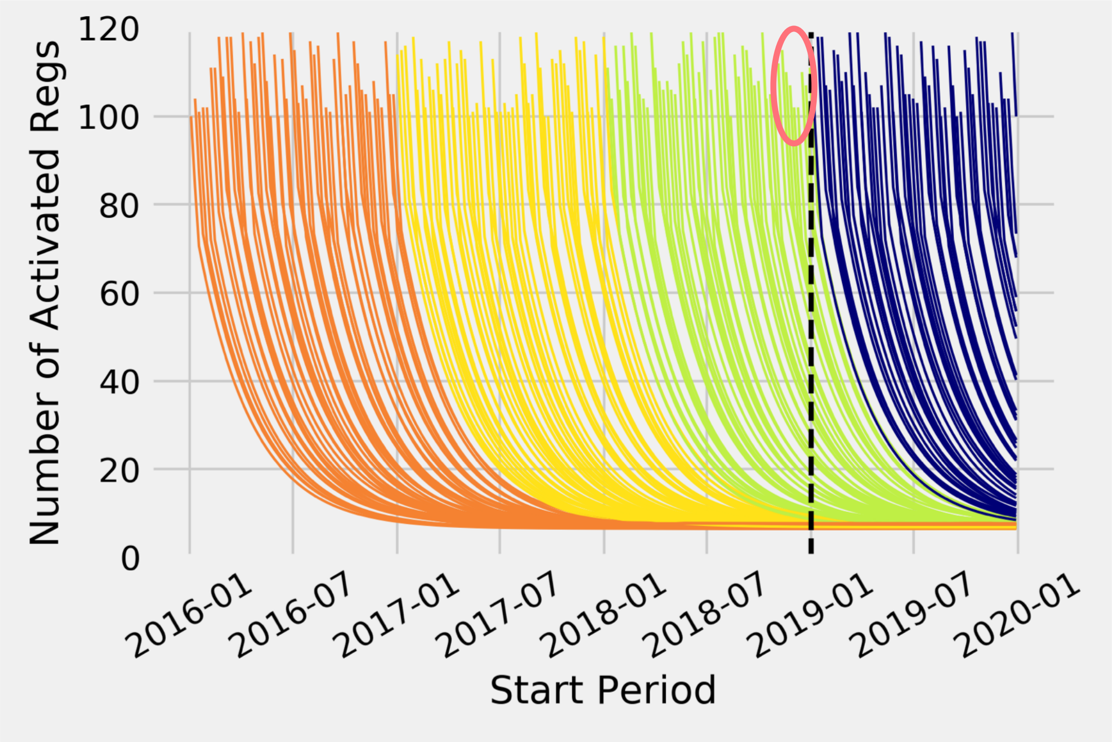
Artificially generated data

We apply the average retention curve



Artificially generated data

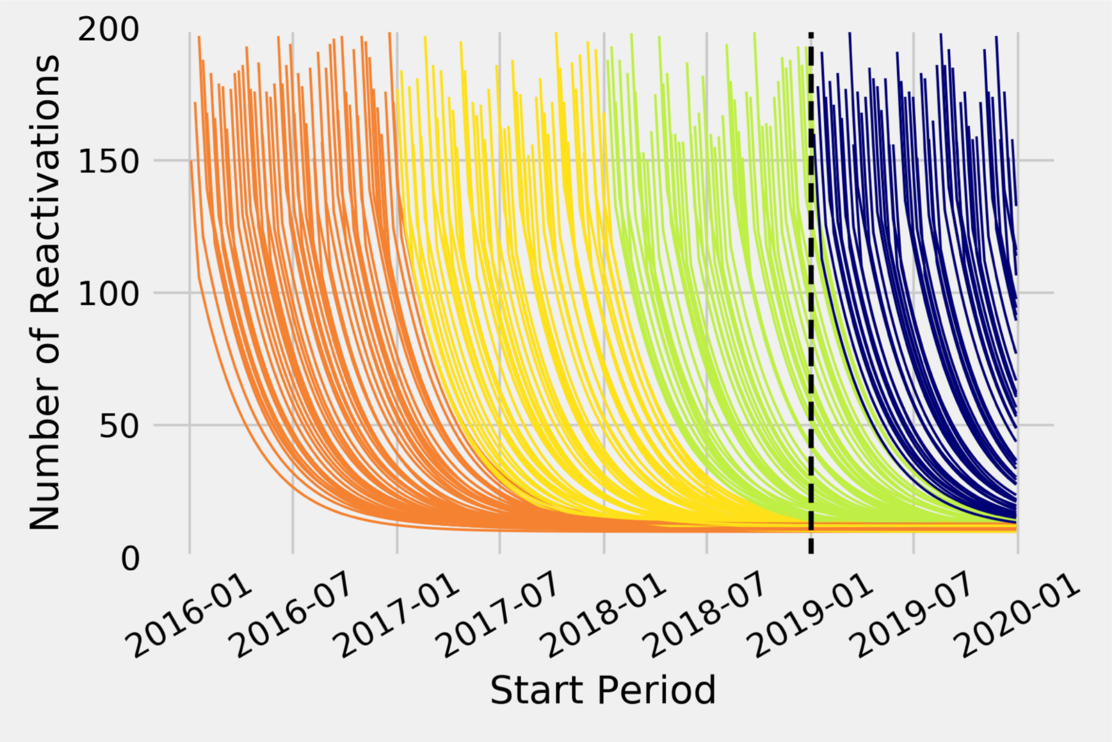
We apply the average retention curve



We use the average curve on the last few of the historical cohorts as well.

Artificially generated data

We <u>repeat</u> the process for reactivated users



Artificially generated data

Pros & Cons of the cohort model

More granular forecast that produces a bottoms up forecast

Better understanding of driving factors

Complex data engineering

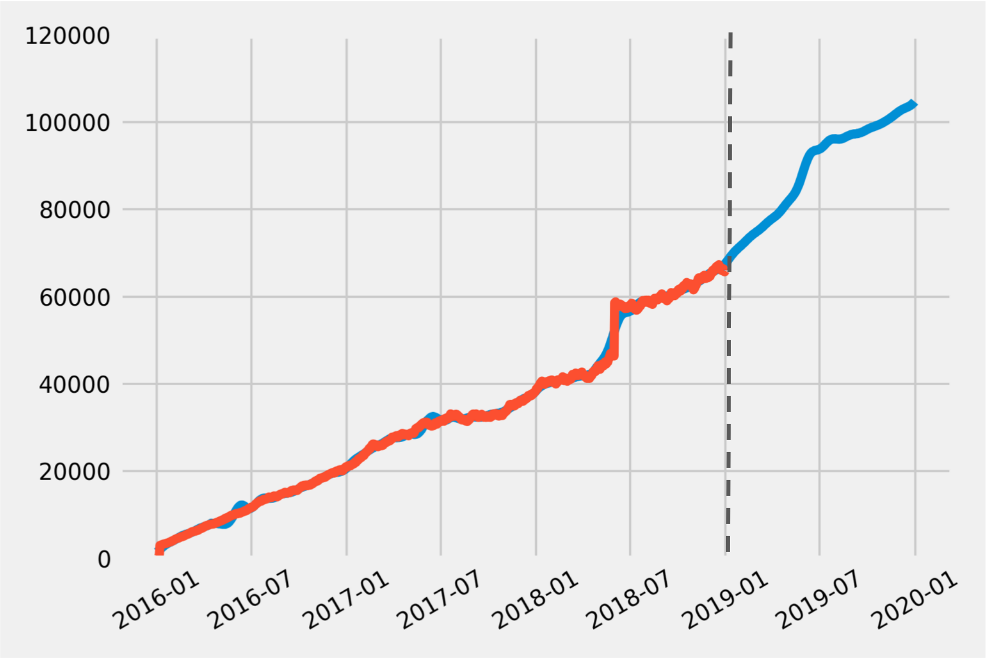
2

Forecasting in a fast-changing world

2.1

How we handle irregularities

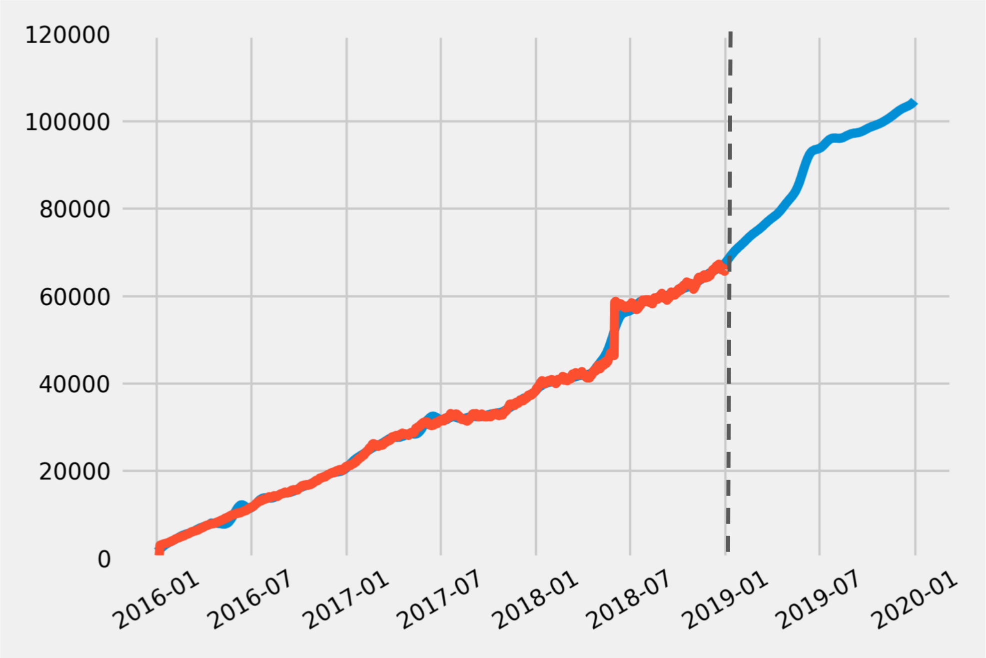
Imagine that we undergo a step change



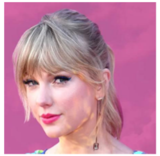
Maybe a marketing campaign, maybe a new product.

Artificially generated data

Imagine that we undergo a step change

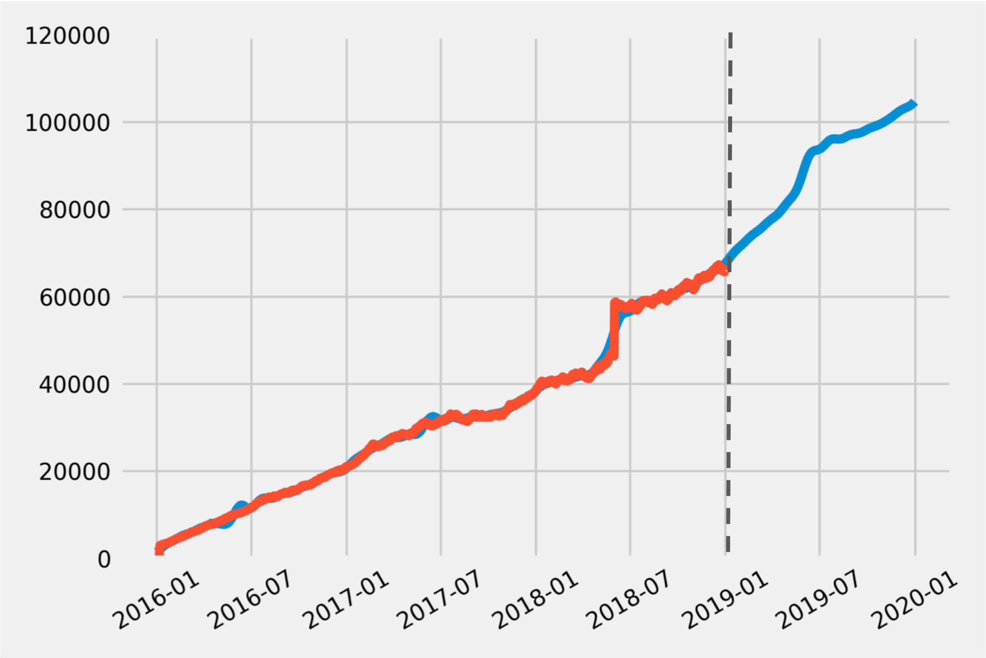


Maybe a marketing campaign, maybe a new product.

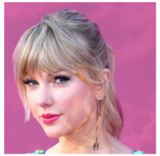



Artificially generated data

Imagine that we undergo a step change



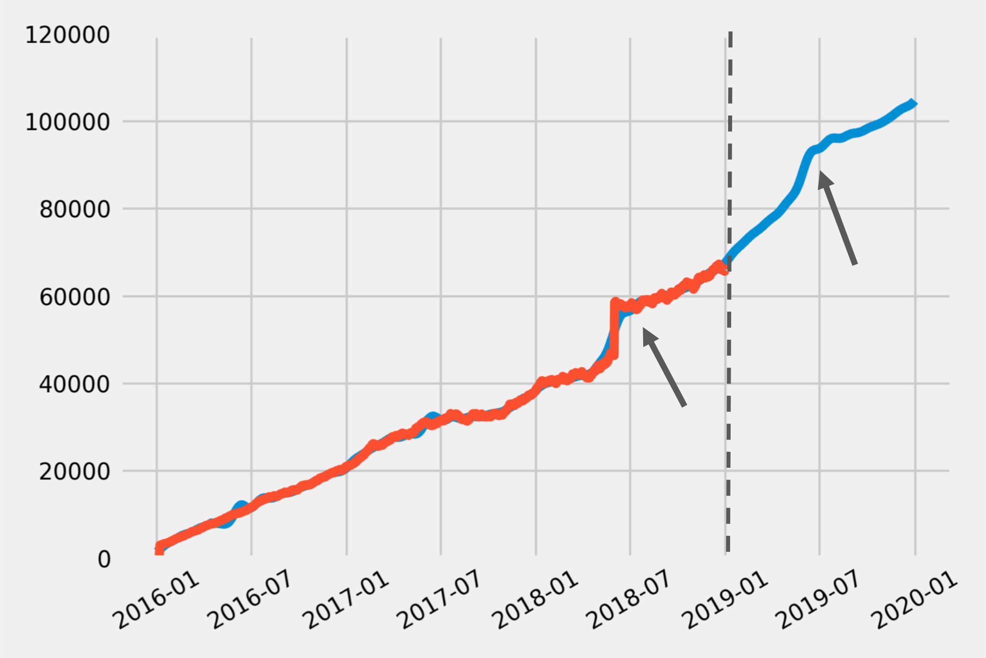
Maybe a marketing campaign, maybe a new product.



The team needs to be aware of global and local major events

Artificially generated data

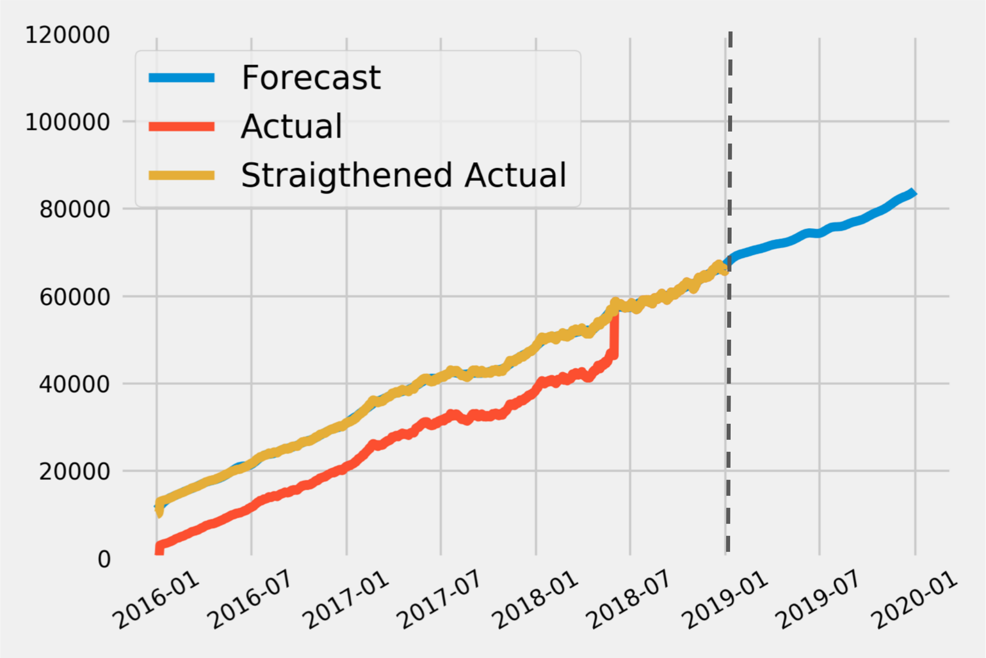
Prophet will pick up this “trend”



Prophet will forecast this step change in the future, even though we humans know that it was a one-time thing.

Artificially generated data

We can make the step change go away



Can we assume that the **rate of growth** remained somehow steady?

Artificially generated data

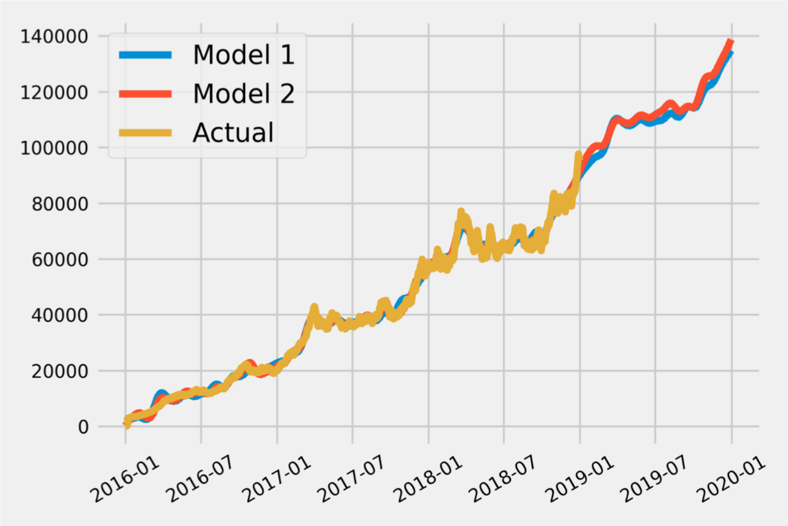
2.2

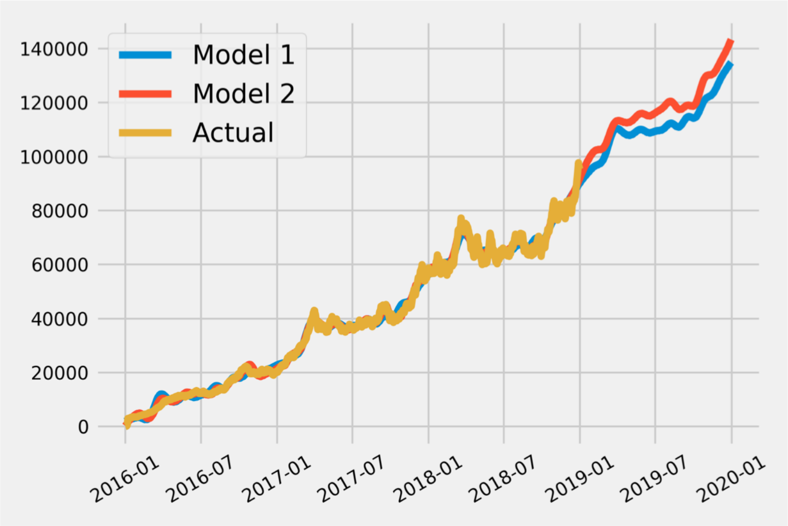
Additional **robustness** tests

We compare against other models

Once we have the first version of our forecast, we compare it against other models, such as prophet, holt-winters and even special forecasts that the local teams have given us.

We adjust as needed.



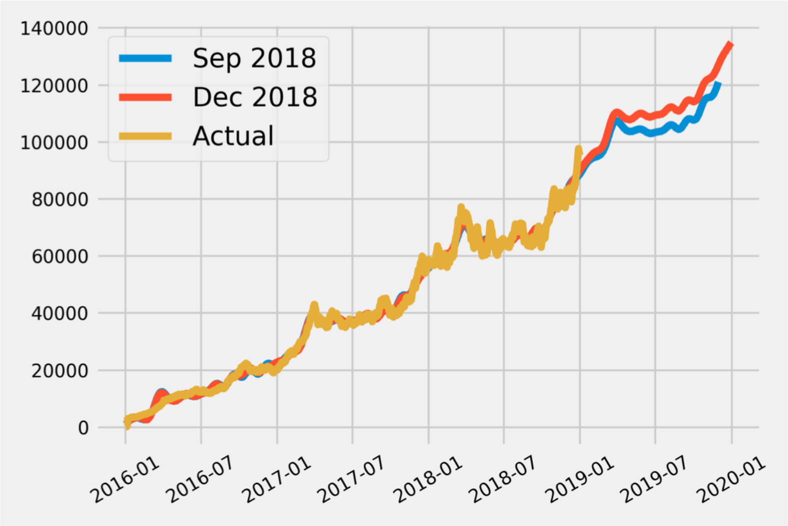


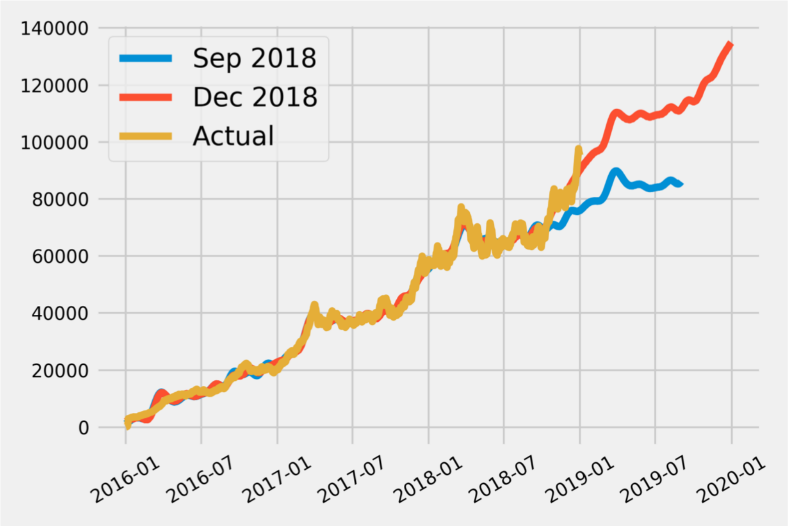
Artificially generated data

We check against the previous forecast

We aim to understand the differences between the current and previous forecast.

We check actual data and communicate with local market leaders





Artificially generated data

Learnings

1. **Explore weighing recent observations heavier**

1. **Be aware of the environment you're operating in**
    a. Competitor moves
    b. Big marketing campaigns
    c. New markets, new products

1. **Compare multiple models**

1. **Find ways to QA all results with experts & stakeholders**

3

Future Work

Future work

We use prophet with its default parameters and its performing very well, but we want to explore **grid search.**

Use seasonal survival curves for predicting the retention of future cohorts

Better exploration and alerting on our forecast.

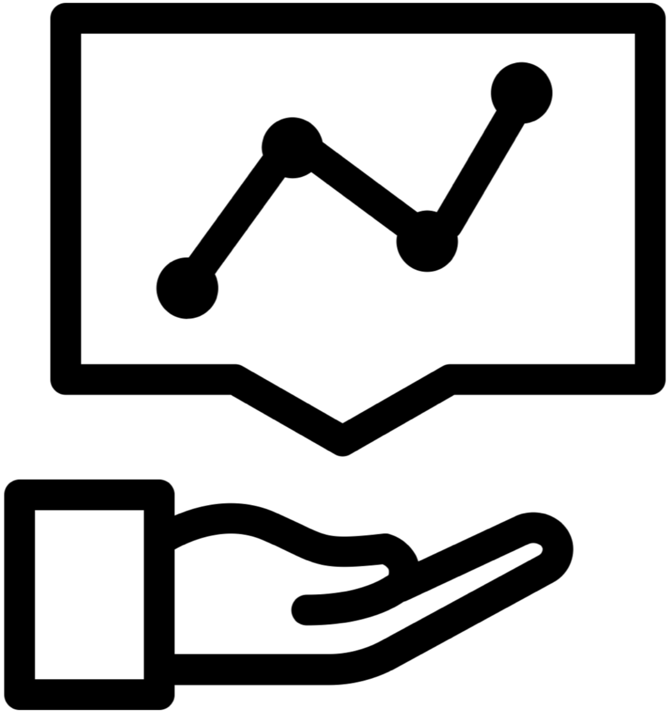
Anomaly detection on the cohort level as well as on the country level.

Forecasting as a Service

Time series
data file

OR SQL query

→



→

Best available
forecast

Forecasting as a Service

Best available forecast

Across many train/test splits: sliding and expanding sizes

Across many possible models: Holt Winters, Prophet, custom internal

Across many possible parameters: grid search

Further external reading:
Uber's Omphalos
Uber Tech Day, ML and forecasting at Uber

GET IN TOUCH

Warren Wertheim
Head of Forecast

Michael Donnelly
Sr. Data Scientist

Keerti Agrawal
Data Scientist

Mahan Hosseinzadeh
Data Scientist

Yorgos Askalidis
Sr. Data Scientist

GET IN TOUCH

Warren Wertheim
Head of Forecast

Michael Donnelly
Sr. Data Scientist

Keerti Agrawal
Data Scientist

Mahan Hosseinzadeh
Data Engineer

Yorgos Askalidis
Sr. Data Scientist

THANK YOU!

yorgos@spotify.com