

# **Data for Building RS**

---

**What data do you  
think is required for  
building RS?**

---



# Data required for Building RS

- Users
- Assets
- Interactions between users and assets

# Users & their metadata

- Demographic data
- Taste preferences
- Domain related Segmentation (RFM)
- Registration Information
- 2nd party or 3rd party data



# Collecting User Data

- During registration or subscription
- Feature engineering
- Partnerships

# Assets & their metadata

Categorisation of  
the asset

Details of asset  
creators

Domain related  
Segmentation

---

# Collecting Item Data

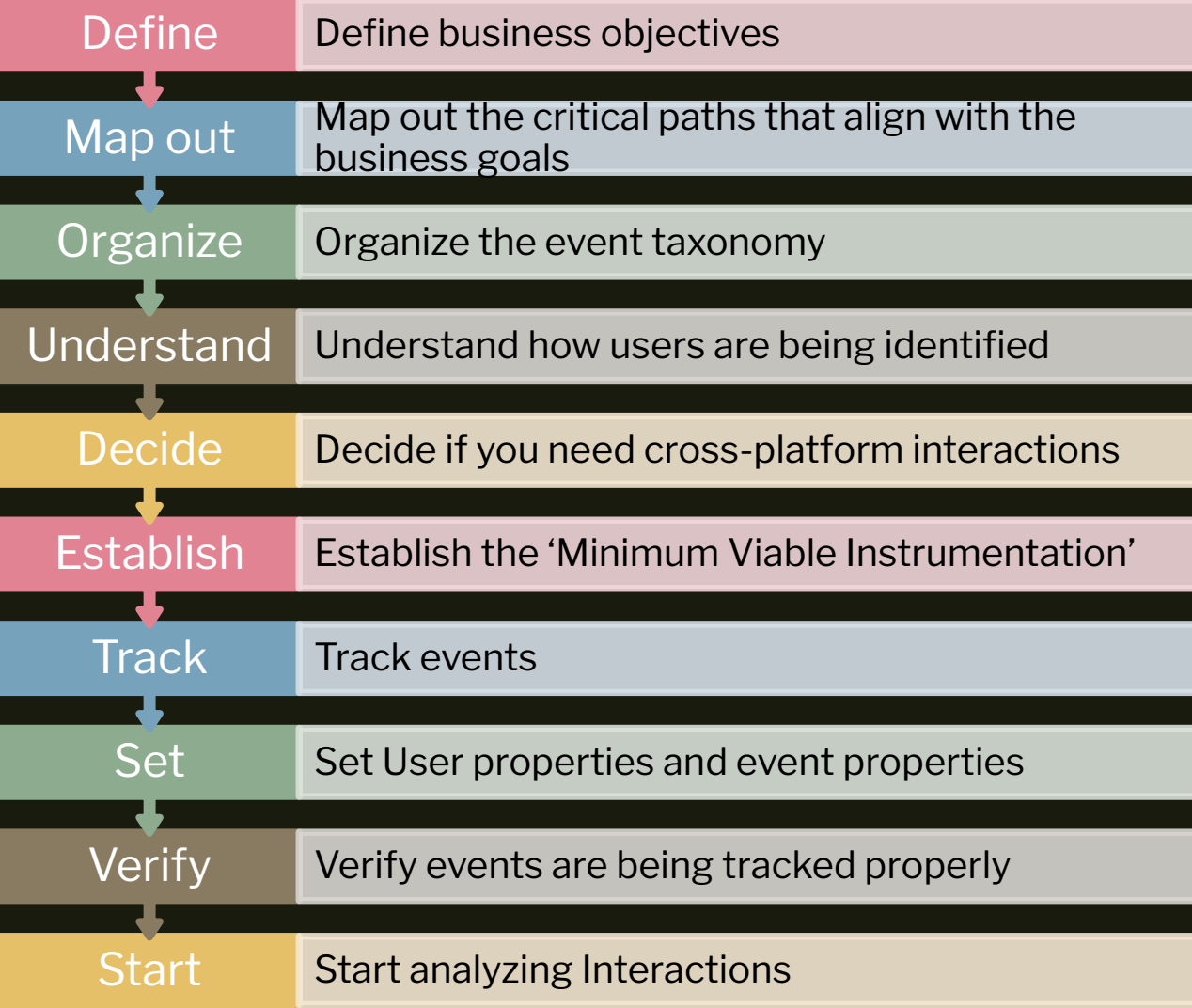
- 
- Gathering information prior to onboarding the content
  - Feature engineering
  - Using ML models for categorization / extracting features

---

## Interaction data

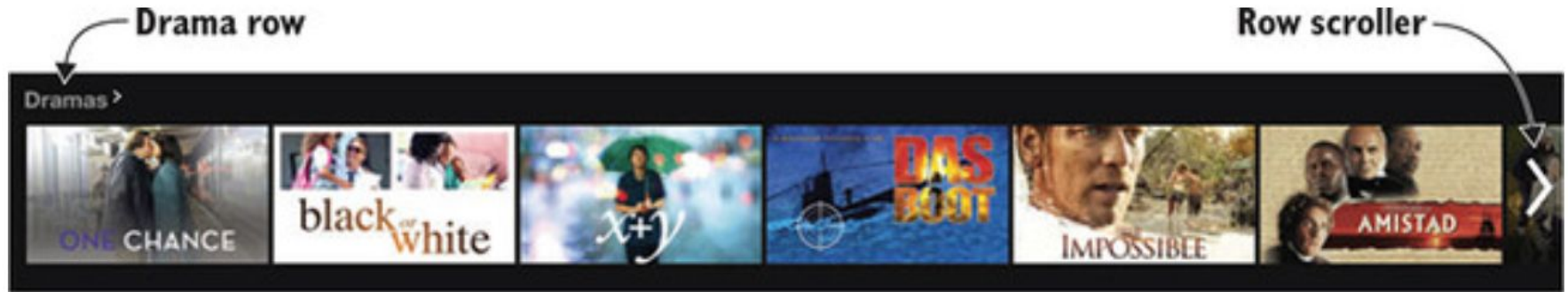
- 
- Ratings
  - Likes
  - Comments
  - Shares
  - Favourites
  - Shares
  - Timestamp of interaction
  - Domain related information





# Steps in Capturing Interactions

# Netflix - An Example of Data Collection



# Netflix - An Example of Event logging

userId	contentId	Event	Date
1234	2	genreView	2017-06-07 20:01:00
1234	41335	details	2017-06-07 20:02:21
1234	41335	moreDetails	2017-06-07 20:02:30
1234	41335	addToList	2017-06-07 20:02:55
1234	41335	playStart	2017-06-07 20:03:01

# Netflix - An Example of Event definition

Event	Meaning	Event name
Scrolling a themed row	User is interested in the theme, here Dramas.	genreView
Placing the mouse over a film to request an overview of content	User is interested in the movie (a drama), thereby showing interest in this category.	details
Clicking the film to request the details of content	User is more interested in the movie.	moreDetails
Adding the film to My List	User intends to watch the movie later.	addToList
Starting to watch the film	User “purchases” the movie.	playStart

---

# Things to keep in mind

- 
- Explore what you have
  - Capture what you want
  - Data quality
  - Start small
  - Keep adding features for improvement
  - Think about edge cases
  - Think about the impact of the system on various communities & regions
  - Monitor the results