

# SISTEMAS DE RECOMENDACIÓN (SEMANA 2)

**Cristian Cardellino - Luis Biedma**

# SPOILERS

Viernes:

1. Obtención de Datos
2. Evaluación y Validación

Sábado:

1. Algoritmos híbridos
2. Learning to Rank
3. Más casos de estudio (con algoritmos)
4. Segundo Práctico

# CONOCIENDO USUARIOS...

- Evidencia: Datos que revelan gustos de usuarios.
- Boom or Bust.
- ALTAMENTE DEPENDIENTE DEL DOMINIO.



# EXPLÍCITO VS. IMPLÍCITO (AGAIN)

Jim



Explicit Rating:  
match.com bio:

I am a vegan. I enjoy a fine Cabernet Sauvignon, long walks in the woods, reading Chekov by the fire, French Films, Saturdays at the art museum, and Schumann piano works.

Implicit Ratings:



what we found in  
Jim's pocket

Receipts for:

12 pack of Pabst Blue Ribbon beer, Whataburger, Ben and Jerry's ice cream, pizza & donuts  
DVD rental receipts: Marvel's The Avengers, Resident Evil: Retribution, Ong Bak 3

# RATING EXPLÍCITO

## Contras:

- Vagancia
- Información Parcial
- Falta de Actualización
- Outliers vs. Diversidad

## Pros:

- It's something
- Mínima información para tomar decisiones

# PULGARES VS. ESTRELLAS

<https://www.appcues.com/blog/rating-system-ux-star-thumbs>

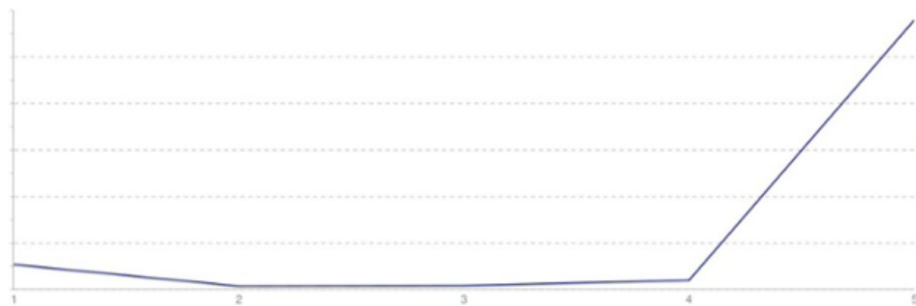
## Pulgares:

- Sencillez
- Más parecido a implícito
- Si dependemos mucho de la personalización, es lo mejor



## Estrellas:

- Decisiones complejas
- + engagement -> - key actions
- Mejora personalización... si las usás bastante



# RATING IMPLÍCITO

## Nuestra selección para Luis



## N SERIE OZARK

91 % para ti 2018 16+ 2 temporadas

Es oficial: se estrenará otra temporada

Un asesor financiero arrastra a su familia desde Chicago hasta los lagos de Ozark para lavar 500 millones de dólares en cinco años y tranquilizar a un jefe narco.

▶ REPRODUCIR

✓ MI LISTA



Protagonistas: Jason Bateman, Laura Linney, Sofia Hublitz

Géneros: Dramas de TV sobre crimen, Thrillers de TV, Dramas de TV

Este programa es: Inquietante, Siniestro

*Paciencia. Austeridad. Sacrificio.*

DESCRIPCIÓN GENERAL

EPISODIOS

TRÁILERS Y MÁS

MÁS SIMILARES

DETALLES

# RATING IMPLÍCITO

- Enfocado en comportamiento

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



# ENCONTRAR COMPORTAMIENTO ÚTIL

- Mucho involucramiento del usuario -> + data relevante
- Muchos usuarios de una vez -> + relaciones entre items

## Propósito

The Netflix logo is displayed in a bold, red, sans-serif font. It is centered within a light gray rectangular box that has a subtle gradient and a soft drop shadow, giving it a three-dimensional appearance.

**NETFLIX**

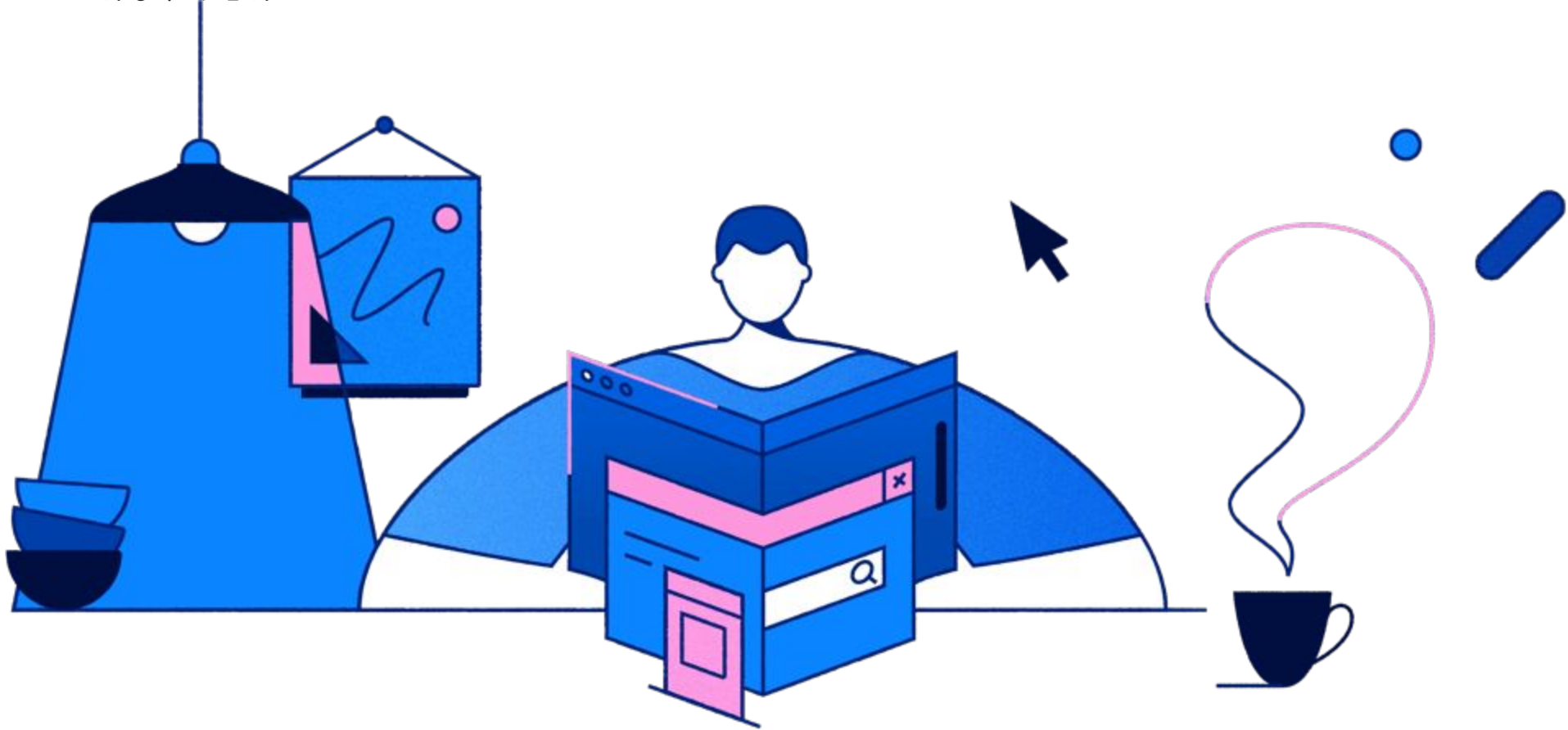
The Amazon logo consists of the word "amazon" in a black, lowercase, sans-serif font. Below the text is a curved orange arrow that starts under the letter 'a' and points towards the letter 'n', representing a smile.

**amazon**

# ENCONTRAR COMPORTAMIENTO ÚTIL

1. “Está mirando”: Sin meta específica - toma pausas y muestra interés.
2. Se interesa: Ya sabía qué buscaba?
3. Añade producto a un carrito o lista para comprar.
4. Compra producto.
5. Consume producto (película, canción, viaje...).
6. Califica producto (a veces) <- **EXPLÍCITO**
7. Revende producto o se deshace - Se termina ciclo de vida de producto para el usuario (puede reiniciarse).

BROWSER



# RECOPILANDO INFORMACIÓN

Page View: Se interesa o se pierde? **Clicks/Conversiones**

Page Duration: Depende del dominio

Menos de 5 segundos	->	No hay interés
Más de 5 segundos	->	Hay interés
Más de un minuto	->	Mucho interés
Más de 5 minutos	->	Se fue a buscar café
Más de 10 minutos	->	<b>ABORT</b>

# RECOPILANDO INFORMACIÓN

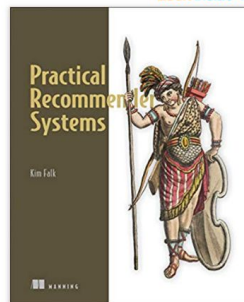
## Expansion Clicks:

### Practical Recommender Systems 1st Edition

by Kim Falk (Author)

★★★★☆ 4 ratings

[Look inside](#)



ISBN-13: 978-1617292705

ISBN-10: 1617292702

[Why is ISBN important?](#)

Have one to sell?

[Sell on Amazon](#)

[Add to List](#)

Share [Email](#) [Facebook](#) [Twitter](#) [Pinterest](#)

**Paperback**  
\$35.49

**Other Sellers**  
from \$29.34

Buy new

**Only 15 left in stock (more on the way).**

Ships from and sold by Amazon.com.

**\$35.49**

List Price: \$49.99

Save: \$14.50 (29%)

**19 New from \$29.34**

This item ships to **Nice, France**. **Want it Thursday, Nov. 14?** Order within **7 hrs 51 mins** and choose **AmazonGlobal Priority Shipping** at checkout. [Learn more](#)

[Deliver to Johanna - Nice 06000](#)

Qty: 1

[Add to Cart](#)

[Buy Now](#)

#### More Buying Choices

**19 New from \$29.34** | **13 Used from \$34.00**

**32 used & new from \$29.34**

[See All Buying Options](#)

#### Summary

Online recommender systems help users find movies, jobs, restaurants-even romance! There's an art in combining statistics, demographics, and query terms to achieve results that will delight them. Learn to build a recommender system the right way: it can make or break your application!

[Read more](#)

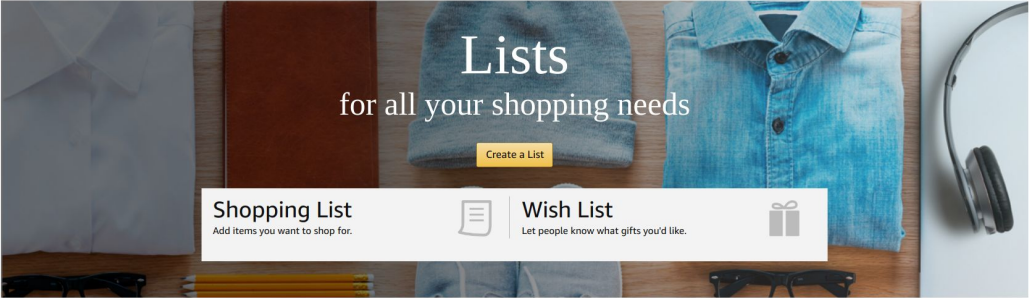
# RECOPILANDO INFORMACIÓN

Social Media



Save for Later

**Your Lists**   Your Idea Lists   Your Friends



The banner features a background image of folded clothes and a pair of headphones. The text 'Lists for all your shopping needs' is centered. Below it is a 'Create a List' button. At the bottom, there are two tabs: 'Shopping List' (with a document icon) and 'Wish List' (with a gift icon).

**Lists**  
for all your shopping needs

Create a List

**Shopping List**  
Add items you want to shop for.

**Wish List**  
Let people know what gifts you'd like.

**NETFLIX**

Conheça mais o Nubank

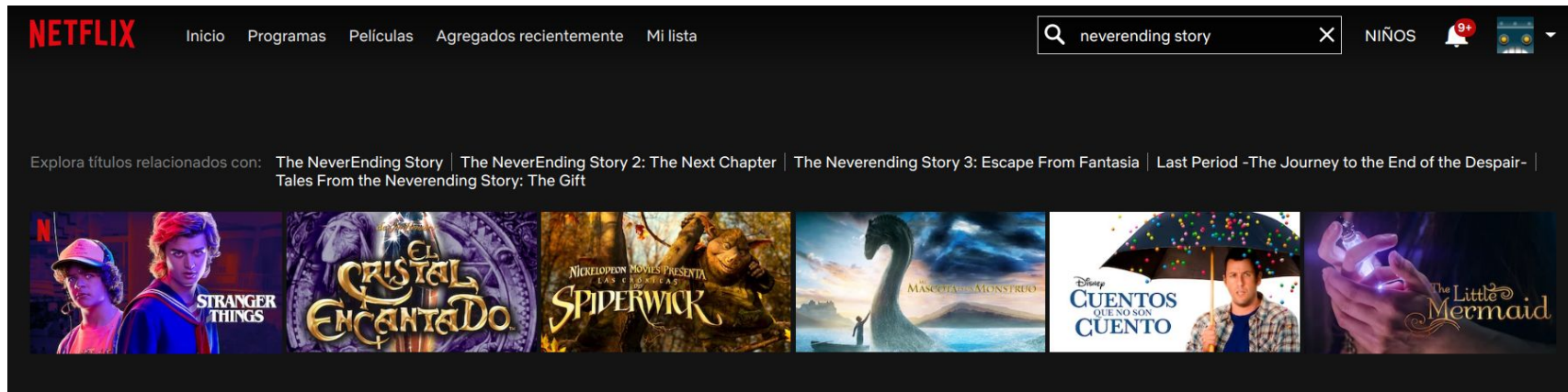
Receba nossa newsletter

**Mi lista**

Favoritos ▾   🔔   🛒

# RECOPILANDO INFORMACIÓN

Search Terms:

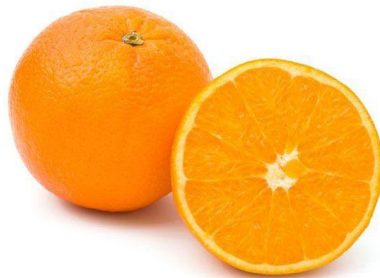


Conectar término de búsqueda con consumo resultante!

# RECOPILANDO INFORMACIÓN

## Act of Buying

- Mayor conversión!
- Para mí vs. Regalo (Outliers)
- Implica buena presentación
- No necesariamente implica gusto de usuario (al menos en una primera compra)





# RECOPILANDO INFORMACIÓN

## Consuming Products

- Se puede perder contacto con compradores
- 0 no:
  - Empezar a reproducir: **positivo**
  - Dejar de reproducir: **ups... depende cuándo**
  - Resumir reproducción: **me olvido de lo negativo**
  - Acelerar: **puede ser malo, si es la primera vez...**
  - Reproducir hasta el final: **suenan las campanas! qué es final?**
  - Re-reproducir: **es contenido muy bueno... o no se entendió**



# RECOPILANDO INFORMACIÓN

## Visitor Ratings

- Netflix: *The more you rate, the better your suggestions*
- Siempre usado, pero debe contrastarse con **comportamiento**
- Se puede ayudar al usuario
- Si además deja una reseña, es muy bueno
- Consumir y no dejar una reseña significa algo?

Deep Learning (Adaptive Computation and...

Ian Goodfellow



**It's okay**

# RECOPILANDO INFORMACIÓN

## A Sense of Control

- Te dejo agregar preferencias y...
- ...probablemente no las uso (manipulación?)
- Documentales vs. Comedias
- Cuidado con los datasets!
- **Guardar ratings en el producto**
- Negative Rating vs. No Rating
- Up-voting (Reputation Systems)



# RECOPILANDO INFORMACIÓN

**NETFLIX**

Angela, take a minute to create a *personalized* Netflix website.

How often do you watch ...	Never	Sometimes	Often
Action & Adventure	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Children & Family	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Comedies	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Documentaries	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dramas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Foreign Movies	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Horror	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sci-Fi & Fantasy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
TV Shows	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

# IDENTIFICAR USUARIOS

Recolectar datos de usuarios solamente funciona si tenés una forma de identificarlos unívocamente.



# COOKIES



cookie

*nombre femenino*

## 1. INFORMÁTICA

Archivo pequeño que envía un servidor web al disco duro del internauta que lo visita con información sobre sus preferencias y pautas de navegación.

"un estudio descubre que hay empresas que están utilizando cookies imborrables"

- Cookie -> User (cuidado!)
- Cross-device?
- **Intentar que la persona siempre cree un user**



# OTROS RECURSOS

Tu sitio es único, igual que tus datos... o no?

- Social Media
- Información Extra (cuestionarios? otros sitios?)
- En algunos casos se depende mucho de usuarios similares (data redundante?)

# THE COLLECTOR

- Server side: API que esté en contacto con lo que hace el usuario.



- Client side: Función en JavaScript que postee la evidencia hacia el collector.



Default

+ New

All Sources

Search all logs

Last hour

Search

Field Explorer

- LogglyNotifications
- Apache
- Derived Fields
- ELB
- HTTP
- Java
- JSON
- Syslog
- System
- Other

Event Timeline

Chart



Event View

View as: List | Grid

Sort: Descending | Ascending

Drill down: Expand Events

More Options

Create

2017-09-19 15:15:20.688	{ json: { level: "ERROR", logger: "com.loggly.lab.RequestHandler", context: "default", thread: "http-bio-8080-exec-7" }, http: { clientHost: "54.173.157.5", contentType: "application/json" } }
2017-09-19 15:15:13.538	{ system: { user: "service" }, derived: { syslogAppName: "sshd", CombinedHost: "ip-172-31-25-95", syslogAppId: "19722:15:13.538163+00:00" } }
2017-09-19 15:15:09.831	{ json: { level: "ERROR", logger: "com.loggly.lab.RequestHandler", context: "default", thread: "http-bio-8080-exec-12" }, http: { clientHost: "54.173.157.5", contentType: "application/json" } }
2017-09-19 15:15:02.230	[{"values": [0], "dstypes": ["gauge"], "dsnames": ["value"], "time": 1505859288.142, "interval": 10.000, "host": "ip-172-31-0-16.ec2.internal", "plugin": "processes", "plugin_instance": "", "type": "ps_state", "type_instance": "blocked"}, {"values": [27797748], "dstypes": ["derive"], "dsnames": ["value"], "time": 1505859288.143, "interval": 10.000, "host": "ip-172-31-0-16.ec2.internal", "plugin": "processes", "plugin_instance": "", "type": "fork_rate", "type_instance": ""}, {"values": [15918096], "dstypes": ["derive"], "dsnames": ["value"], "time": 1505859298.137, "interval": 10.000, "host": "ip-172-31-0-16.ec2.internal", "plugin": "cpu", "plugin_instance": "0", "type": "cpu", "type_instance": "user"}, {"values": [14914], "dstypes": ["derive"], "dsnames": ["value"], "time": 1505859298.137, "interval": 10.000, "host": "ip-172-31-0-16.ec2.internal", "plugin": "cpu", "plugin_instance": "0", "type": "cpu", "type_instance": "nice"}, {"values": [1868736], "dstypes": ["derive"], "dsnames": ["value"], "time": 1505859298.137, "interval": 10.000, "host": "ip-172-31-0-16.ec2.internal", "plugin": "cpu", "plugin_instance": "0", "type": "cpu", "type_instance": "system"}, {"values": [265337637], "dstypes": ["derive"], "dsnames": ["value"], "time": 1505859298.137, "interval": 10.000, "host": "ip-172-31-0-16.ec2.internal", "plugin": "cpu", "plugin_instance": "0", "type": "cpu", "type_instance": "idle"}, {"values": [83376], "dstypes": ["derive"], "dsnames": ["value"], "time": 1505859298.137, "interval": 10.000, "host": "ip-172-31-0-16.ec2.internal", "plugin": "cpu", "plugin_instance": "0", "type": "cpu", "type_instance": "idle"}]

Save this search as...

Manage Saved Searches »

Recent Searches »

All Searches

My Searches

★ Java NPE in last 24 hours

☆ 404 Errors in last 1 hour

☆ All my events

☆ My Favorite Search - New ...

☆ my test search

☆ Nginx - Client Errors

☆ Nginx - Error Logs

☆ Nginx - Server Errors

☆ Renamed this search

☆ Responses Over SLA

☆ Saved Search D1

☆ Search GETs

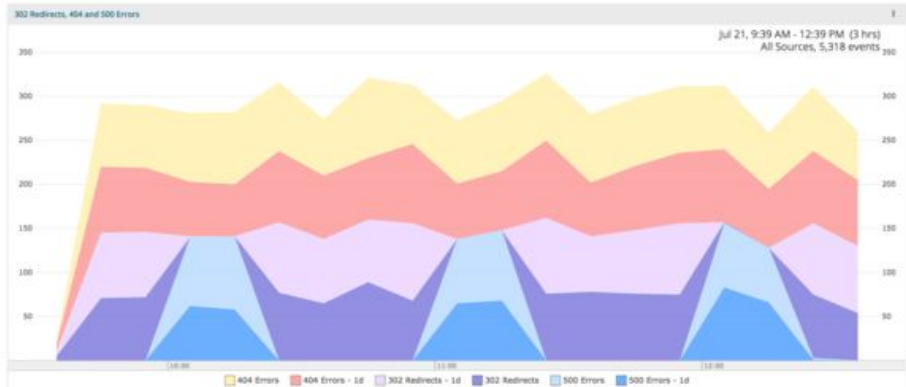
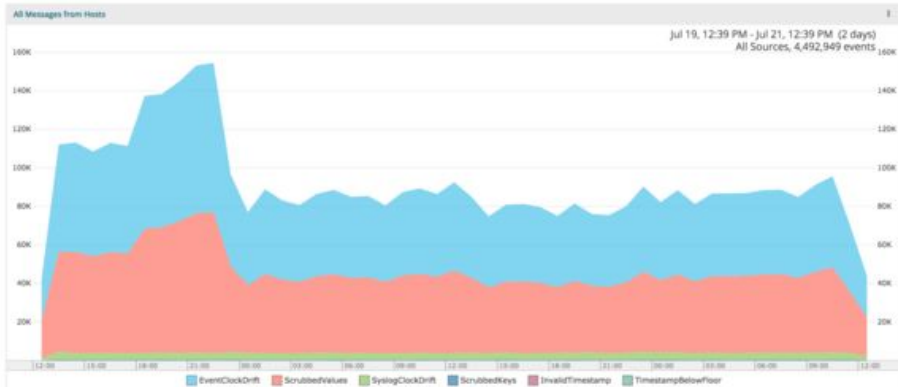
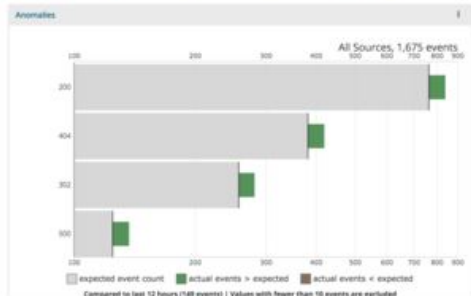
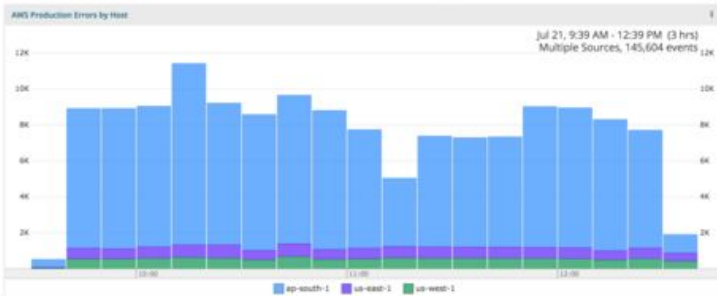
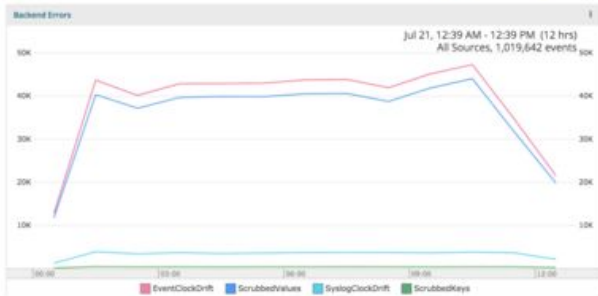
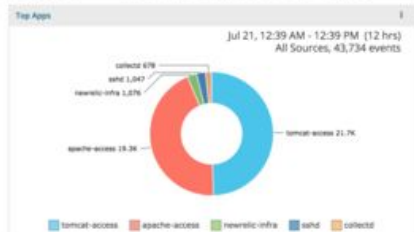
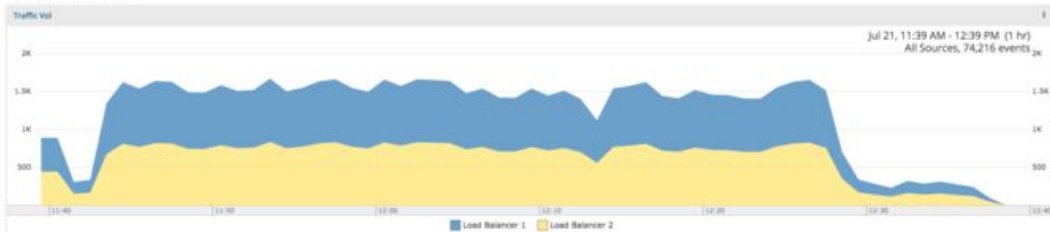
☆ Test Search 234

## System Issues Dashboard

Monitoring production issues

Refresh: 15 minutes | 5 minutes | Off Last Updated: July 21, 12:39:26 PM

Time range: Chart default Source: Chart default Legend: Visible



# MODELO DE DATOS

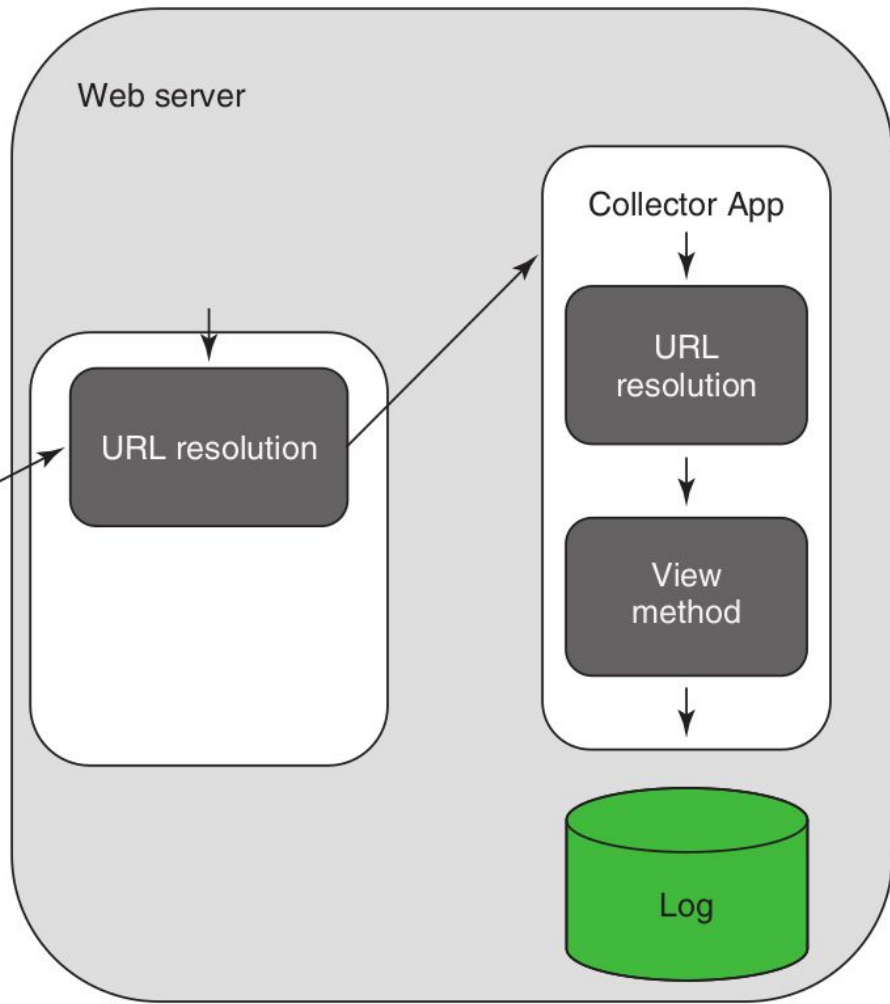
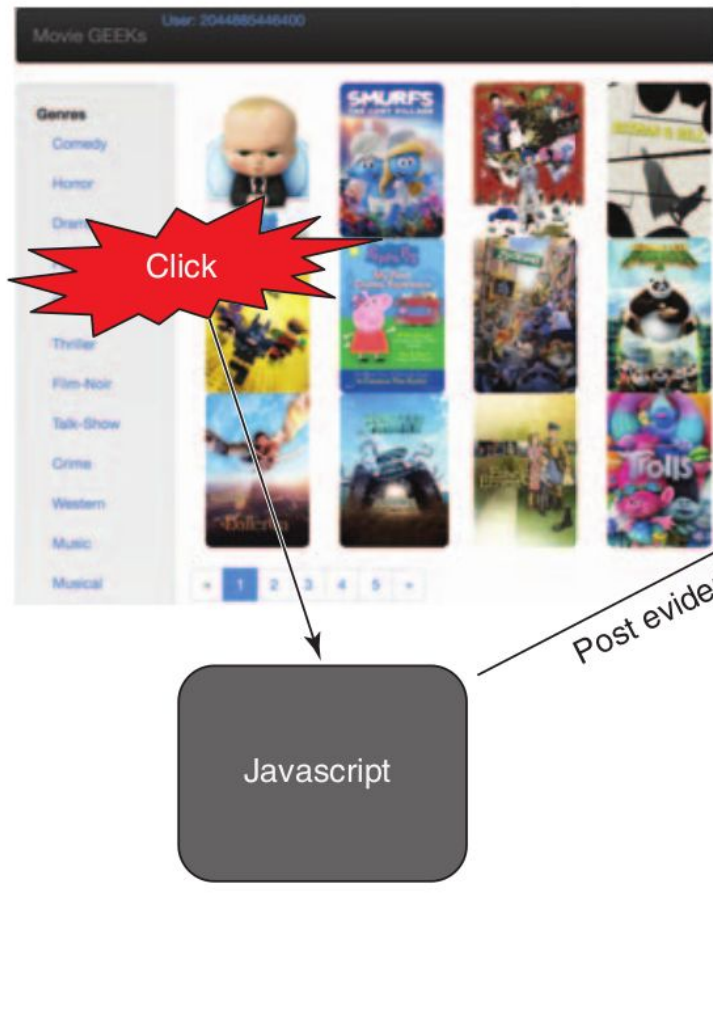
- Momento de acción



- User ID
- Content ID
- Session ID
- Evento

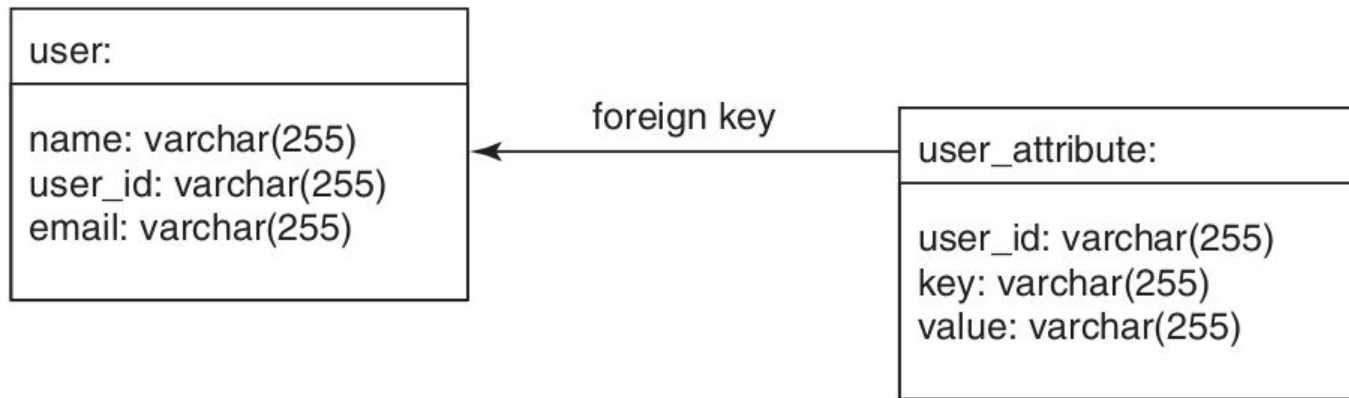
# SNITCH





# MODELAR USUARIOS

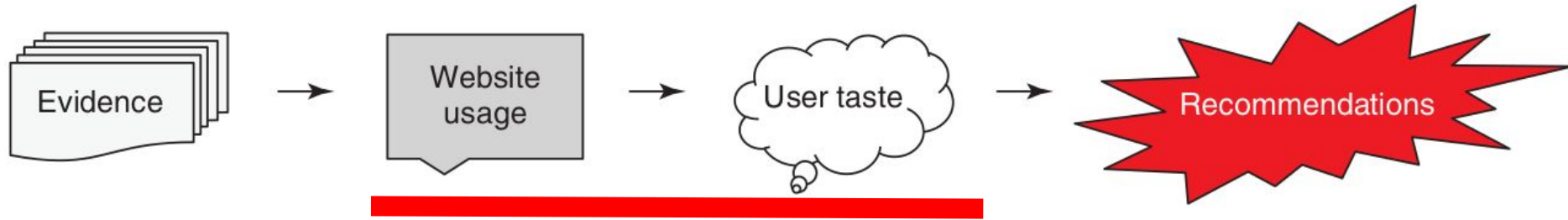
- LinkedIn: Cargo actual, Educación, Años de experiencia...
- Libros: Gustos, hábitos de compra...
- Marketplace: Direcciones de envío
- **Flexibilidad + Simplicidad**



# EN RESUMEN

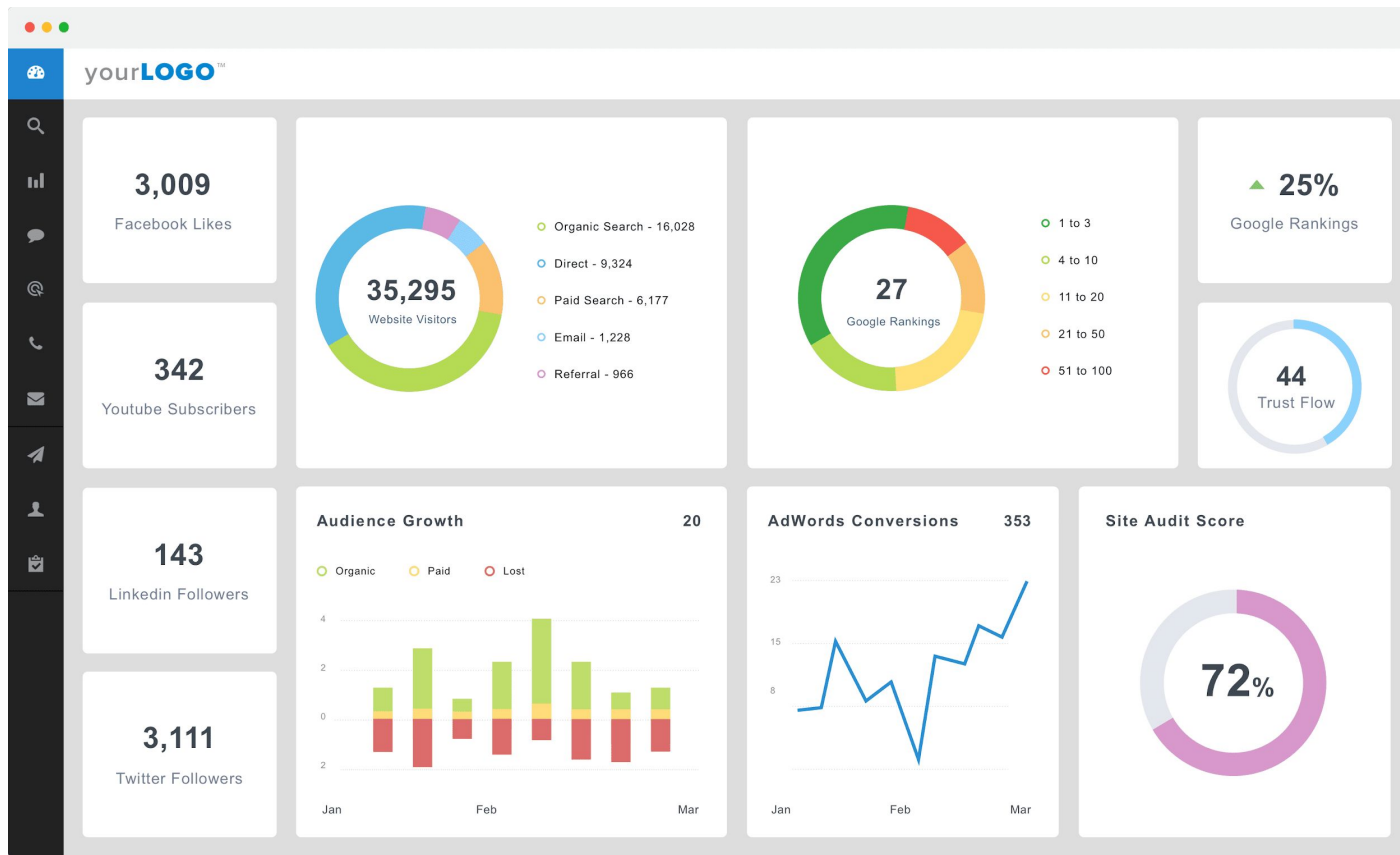
- Loggear users!!!
- Snitch...
- Es bueno guardar todos los eventos que podamos
- Eventos/Logs -> Ratings Implícitos
- Ratings Implícitos son más útiles... Sólo si sabés que significan
- Ratings Explícitos pueden estar sesgados

# ANALYTICS





# DASHBOARD



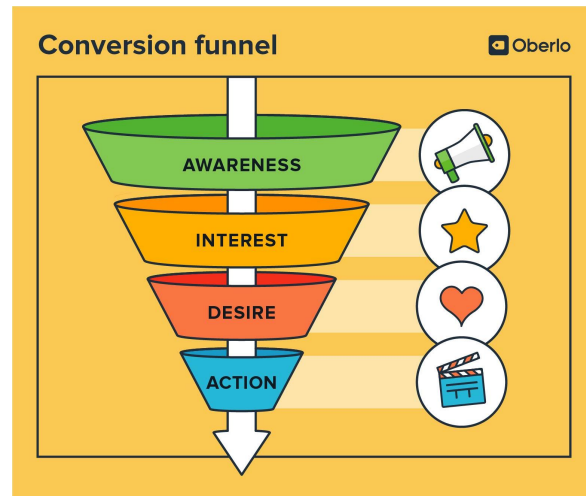
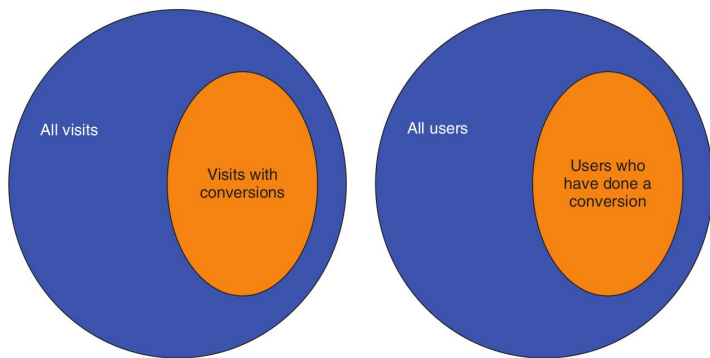
# "CÓMO VAMOS?"

Qué esperamos mejorar agregando un sistema de recomendación?

- Tener un dashboard puede ayudar mucho con esta pregunta
- Analytics -> RecSys
- Benchmarks
- Normalmente no se comparte la forma de un sitio de trackear la información del mismo (Competencia - Seguridad)
- Éxito del sitio/recomendador <- KPIs

# KPIs

- Definido por el **propósito**
- Número de visitantes (PRIMERO)
- *Conversion rate*: # Acciones Positivas / # Visitas o Users
  - Lectura de contenido
  - Suscripción
  - Descargas
  - Ventas (o ventas de los productos más caros)



# THE CONVERSION PATH

## THE VISITOR'S CONVERSION PATH

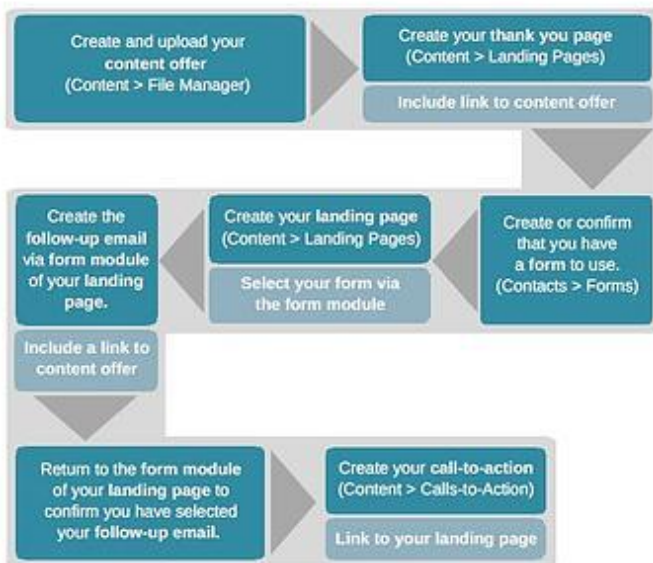
This is the path a visitor of your site would follow to download an offer via a form. This is also a model you can use to verify that all 5 connection points have been correctly made.

To verify the 5 connections within your conversion path, the rule of thumb is to reference the tool that is higher in the flow chart. For example when verifying that your Landing Page Form is appropriately linked to the Thank You Page, you check the form module of your Landing Page.



## THE MARKETER'S ORDER OF OPERATIONS

The order that you, the marketer, builds a conversion path is not actually in the same order that the visitor goes through the conversion path. For example, you can't create a call to action without having a landing page to link it to. By using the order of operations below you will save time and have the right assets when you need them.



## ADDITIONAL STEPS

- Verify all 5 connections using the figure to the left.
- Promote your CTA on your blog, in your emails, and your website pages.
- Schedule a date in the future to review your landing page's performance.

# PERSONAS

## THE CASUAL USER



Pete

Uses most phone features

Uses phone to make, use contacts send texts and take pictures

Always has mobile device with him

## THE BUSINESS USER



Jennifer

Whats a simple phone, but functions as an integrated device

Wants to easily read email and call back the sender

Needs "Popular" mail sever integration

## THE POWER USER









Brad

Will use almost all built-in mobile functionality

Will extend phone functionality with additional software

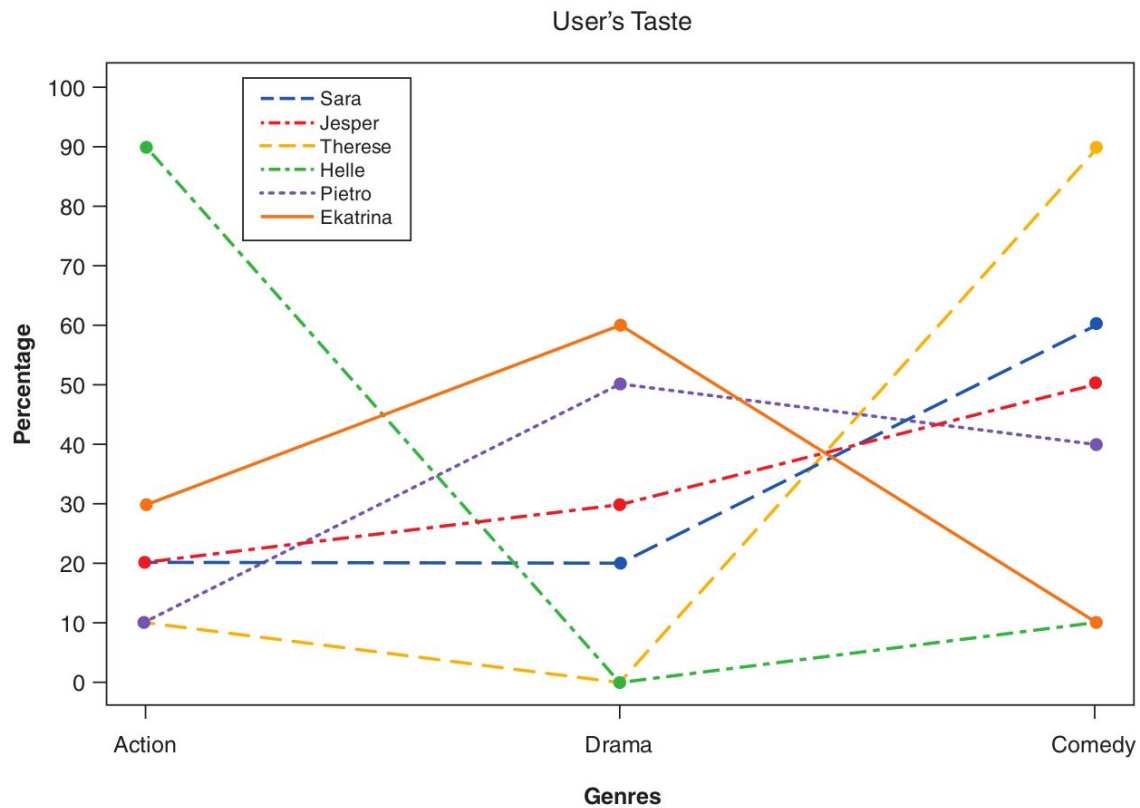
Will look through and change change every menu option

# PERSONAS: EJEMPLO

<div>Sara</div> <div>Comedy, action, drama</div> <div></div> <div>There's always room for another romantic comedy, except when I want to see a CSI-style series.</div> <div>User ID: 400001</div>	<div>Jesper</div> <div>Comedy, drama, action</div> <div></div> <div>I'm up for laughs and will choose a comedy most days but will watch drama and rarely an action movie.</div> <div>User ID: 400002</div>
<div>Therese</div> <div>Comedy</div> <div></div> <div>Anything that makes me laugh is a hit.</div> <div>User ID: 400003</div>	<div>Helle</div> <div>Action</div> <div></div> <div>Anything with superheroes and anything that blows up</div> <div>User ID: 400004</div>
<div>Pietro</div> <div>Drama</div> <div></div> <div>The more complicated the drama, the better it is.</div> <div>User ID: 400005</div>	<div>Ekaterina</div> <div>Drama, action, comedy</div> <div></div> <div>Nothing beats drama, but I sometimes watch action, and rarely a comedy.</div> <div>User ID: 400006</div>

	Action	Drama	Comedy
Sara	20	20	60
Jesper	20	30	50
Therese	10	0	90
Helle	90	0	10
Pietro	10	50	40
Ekaterina	30	60	10

# PERSONAS: EJEMPLO

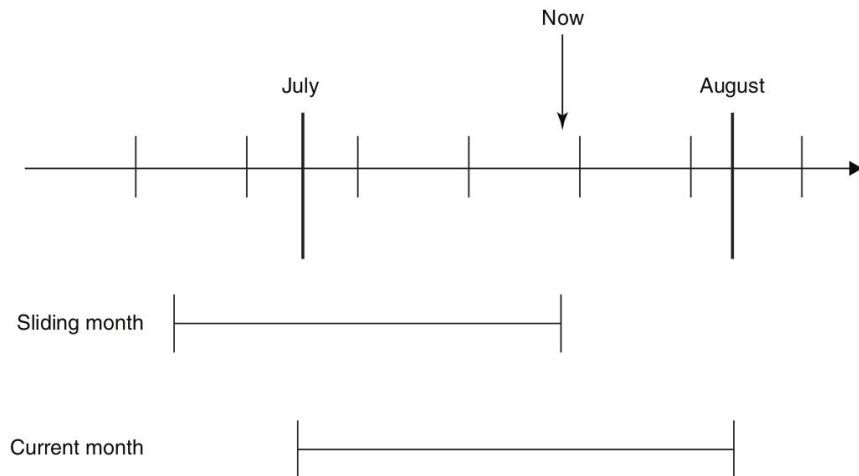
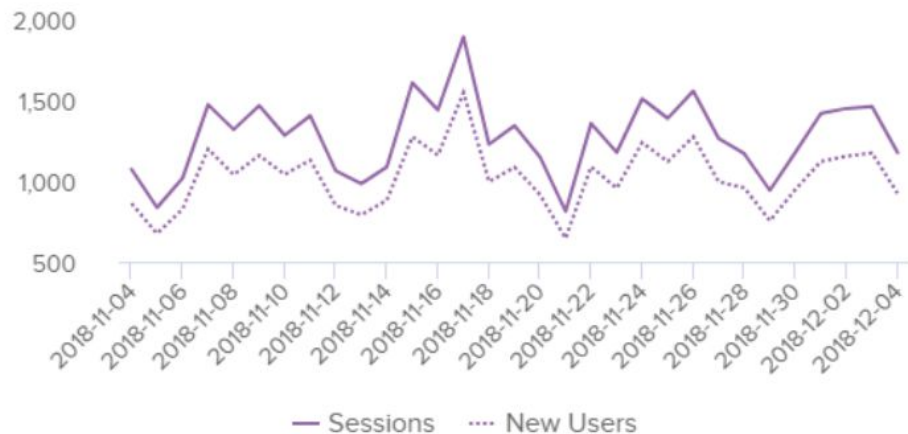


# UNA NOTA SOBRE KPIS

Inevitablemente van a desembocar en SERIES DE TIEMPO

## Sessions and Users

last 30 days





# OTRO RESUMEN...

- Los KPIs son buenos :)
- Visitante -> Conversor
- Es importante ubicar Funnel y Path de Conversión
- Importante generar analytics (quizás antes que TODO)
- Un dashboard logra amistades!