



worten

CUSTOMER DATA

The KEY to unlock Customer VALUE

Paula Figueiredo
8 Abril 2021

pcmestre@worten.pt

AGENDA

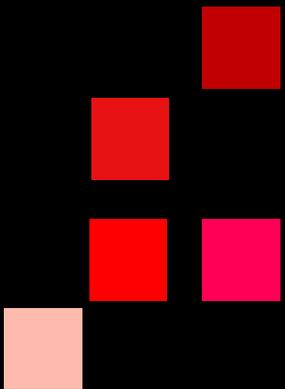
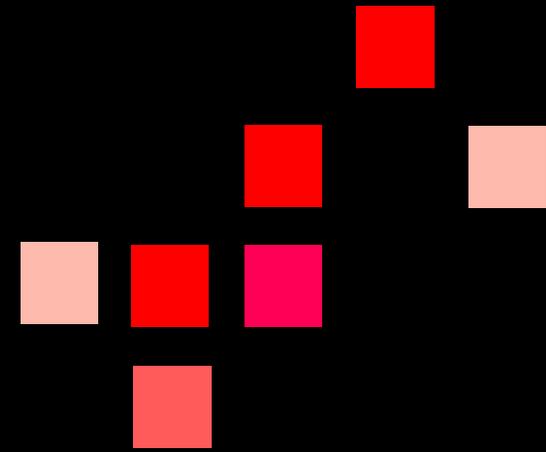
Introduction

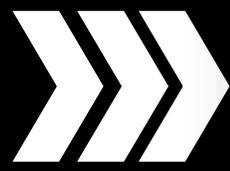
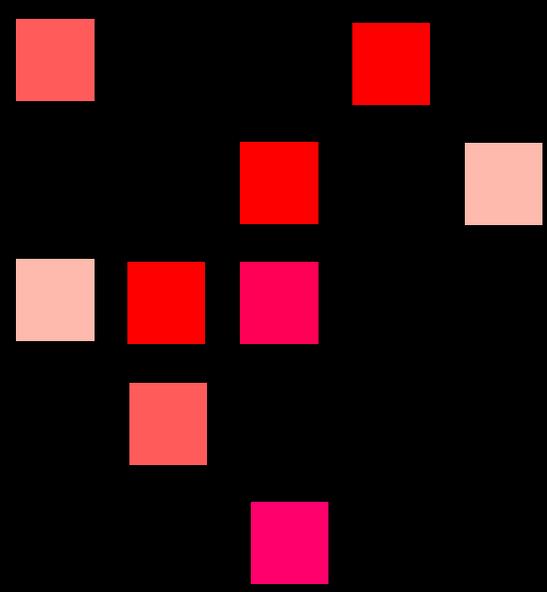
How do we collect data?

What do we do?

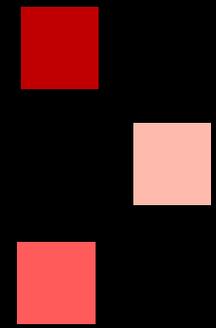
How do we use data?

Online Data

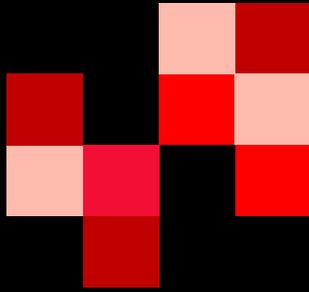




INTRODUCTION

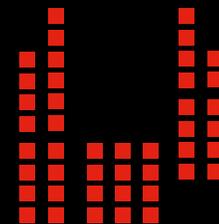
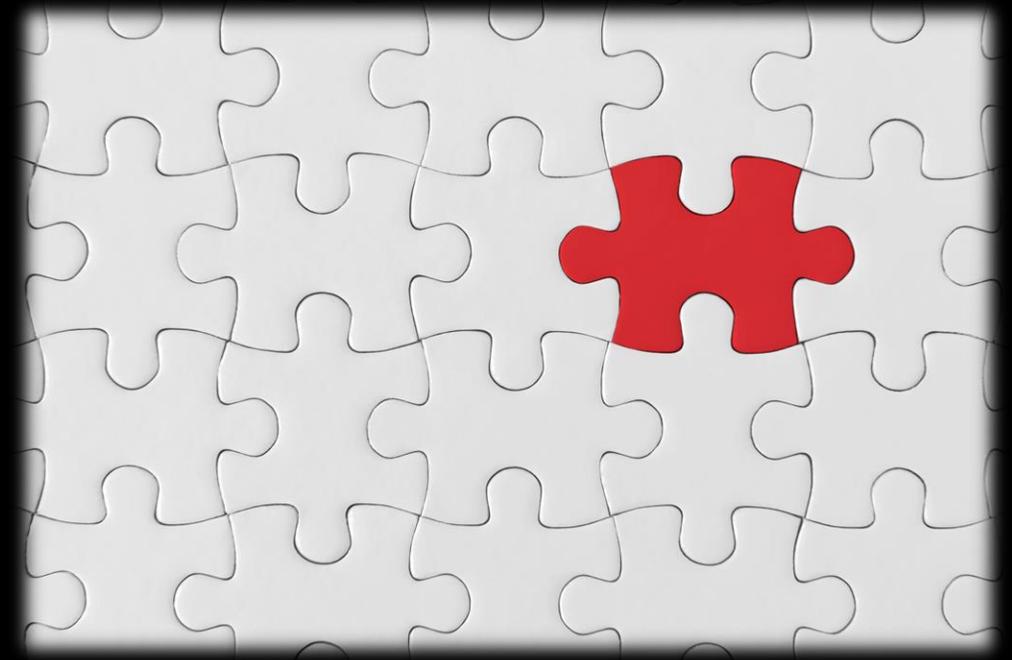


KNOWING OUR CUSTOMERS



In other contexts, where businesses are smaller and local, business owners know their customers:

- »»» Their product/brand preferences
- »»» Financial availability
- »»» Family structure
- »»» Other aspects from their lives that could influence their choices



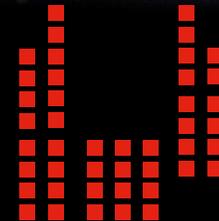
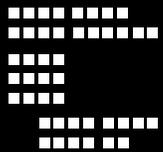
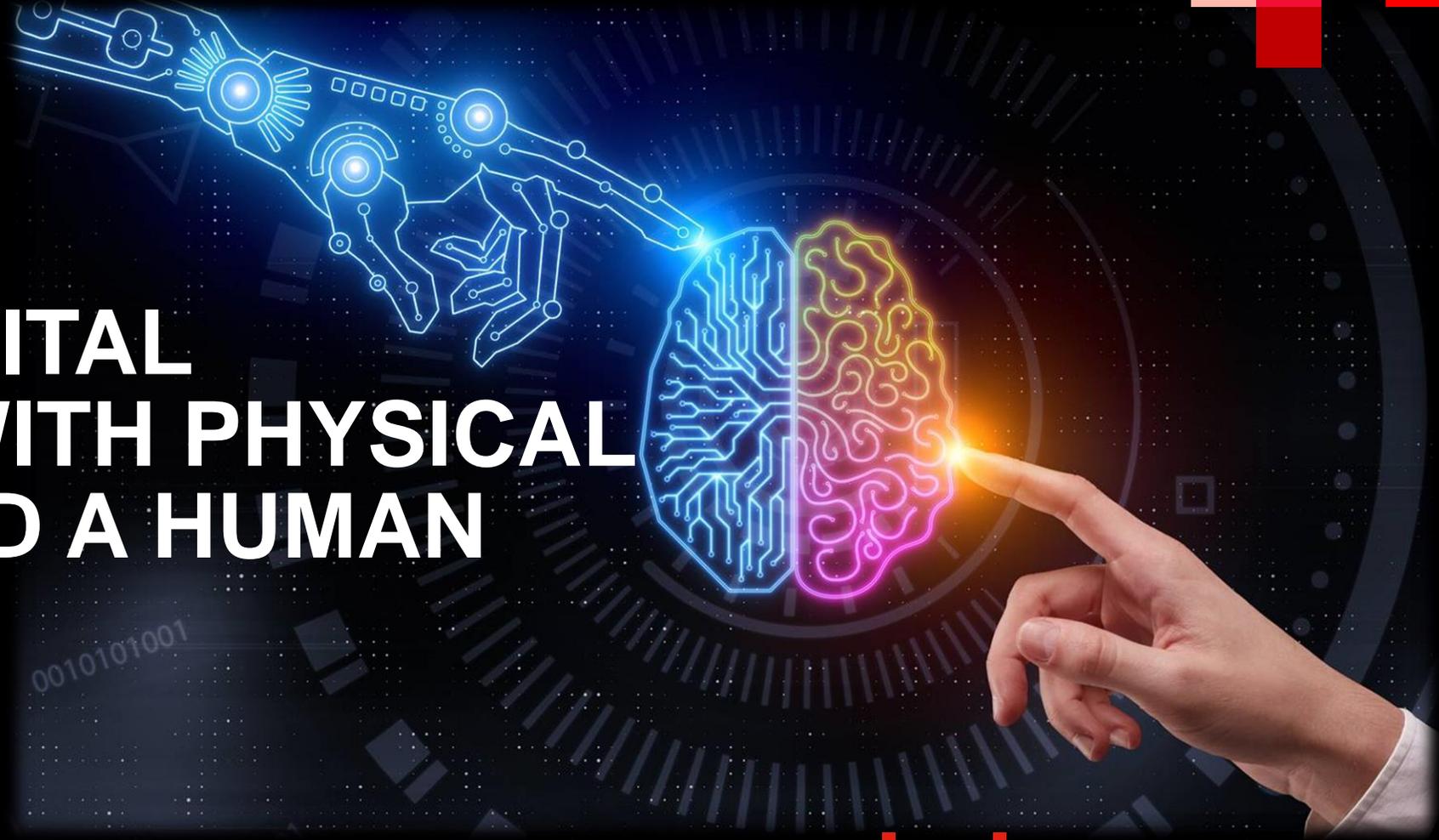
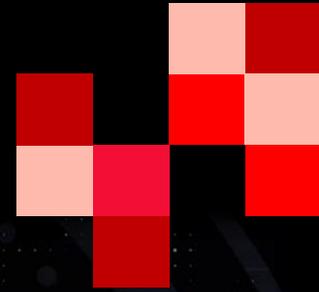
WHAT DO CUSTOMERS

WANT?

- »»» Simplicity
- »»» Convenience
- »»» Save Time and Money
- »»» Feel unique

OUR VISION

»» TO BE A DIGITAL
RETAILER WITH PHYSICAL
STORES AND A HUMAN
TOUCH



ONE ONGOING CONVERSATION

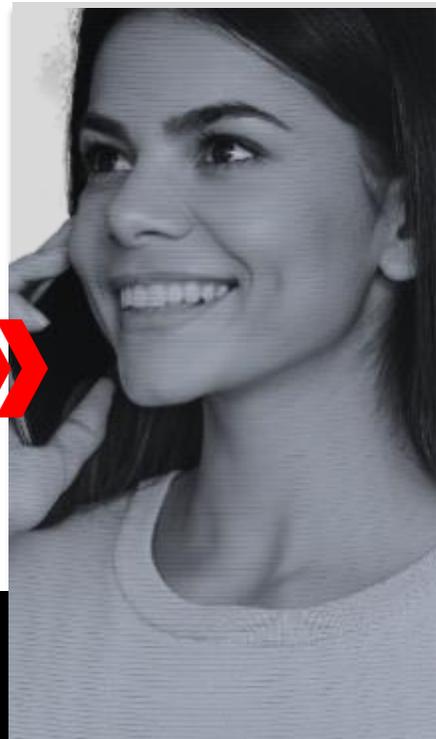
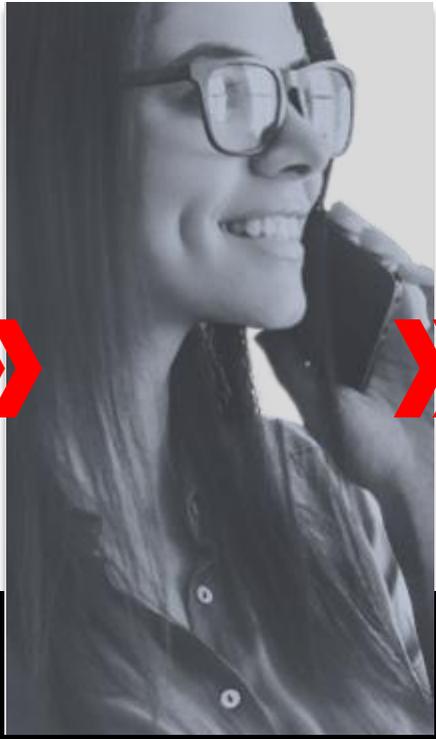


Effortless interactions, which may flow over different moments and different channels



Each interaction must flow from the last as if it was an **uninterrupted** conversation





CONSISTENCY



AGILITY

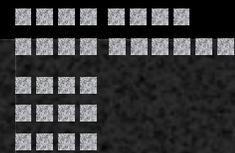
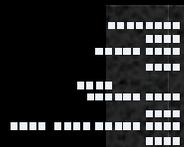


CONVENIENCE



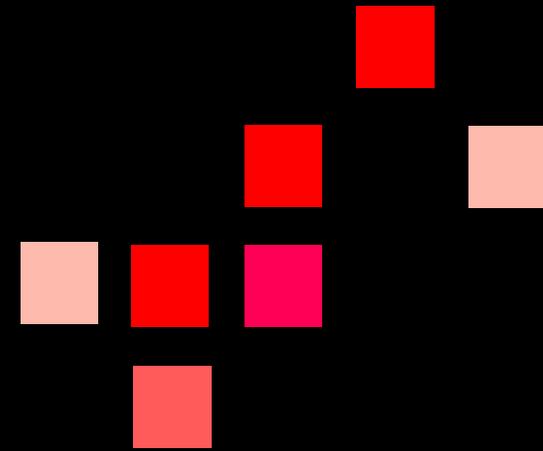
PRO-ACTIVITY

**ONE LONG
UNINTERRUPTED CONVERSATION**



worten

DATA ANALYSIS

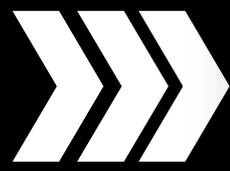
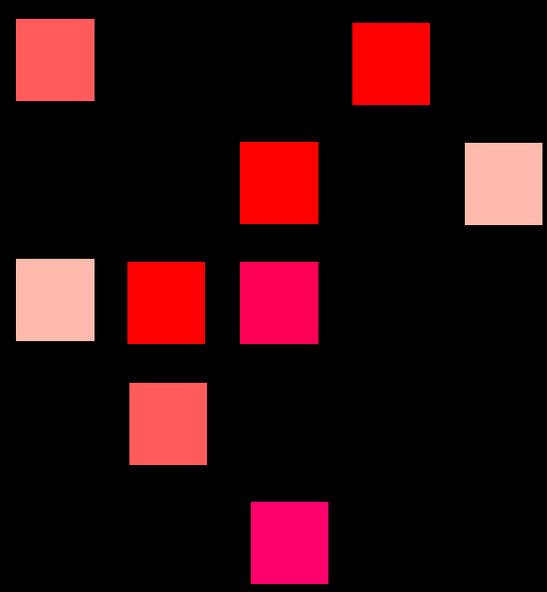


Transforming data into
knowledge

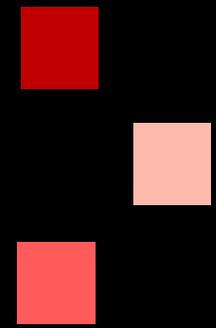
To

Create a 360° view of
customers.

CUSTOMER - CENTRIC
CULTURE



HOW DO WE **COLLECT** DATA?



CUSTOMER DATA FROM DIFFERENT SOURCES



What they tell us
(NPS/NSS and other surveys)



Their purchase habits
(On&Offline)



Browsing, Reviews
and Favourites

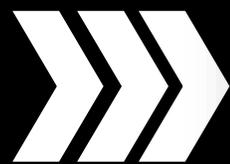
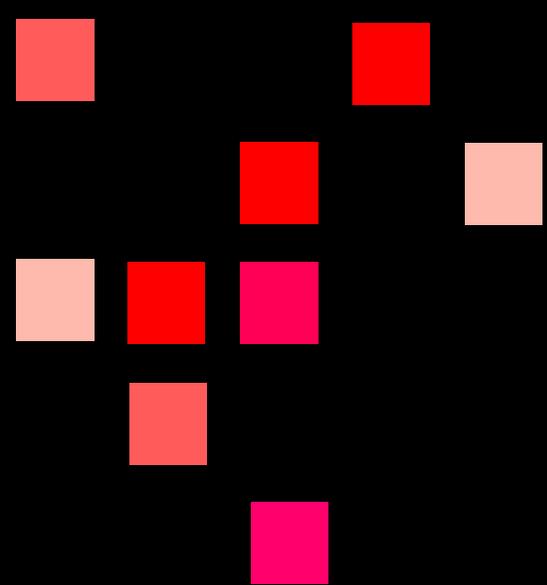


Other interactions
(repairs, complaints,
installations...)

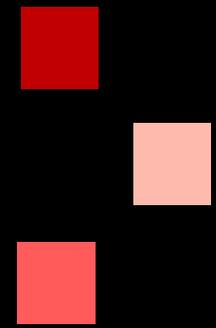


THERE ARE **MANY** OPPORTUNITIES

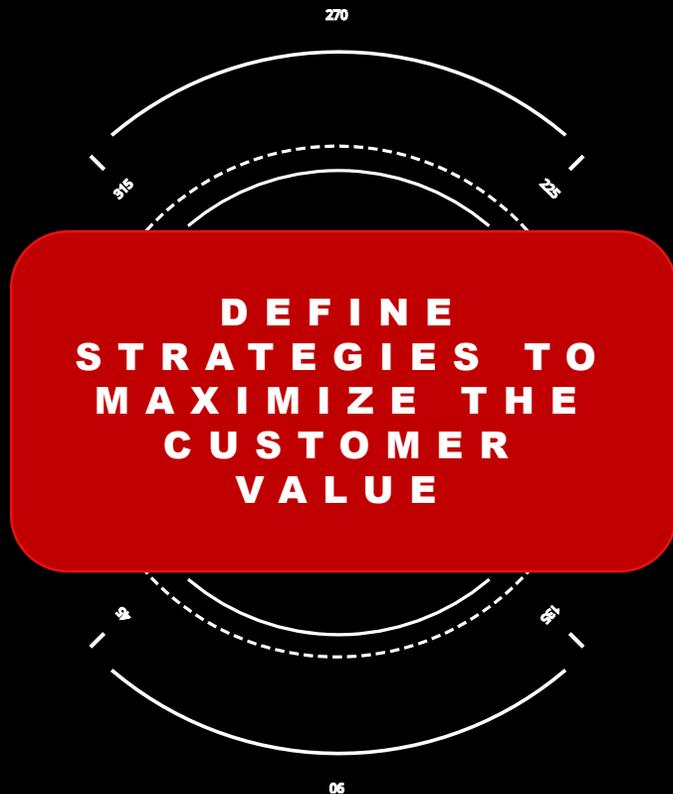
TO COLLECT RELEVANT DATA



WHAT DO WE **DO**?



CUSTOMER VALUE MANAGEMENT



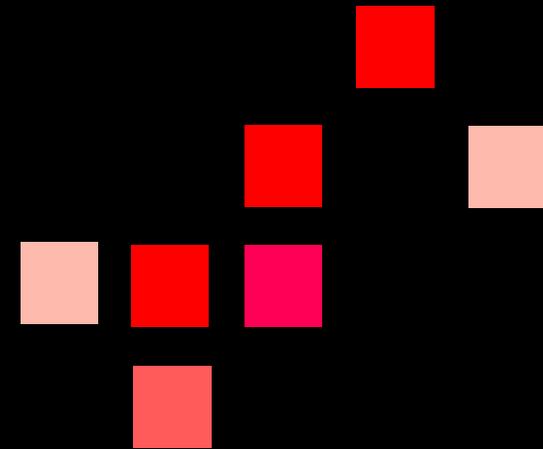
- »»» Analyze and understand customer segments
- »»» Test targeted communications along the customer's lifecycle
- »»» Share gathered insights with the company



CAMPAIGN MANAGEMENT

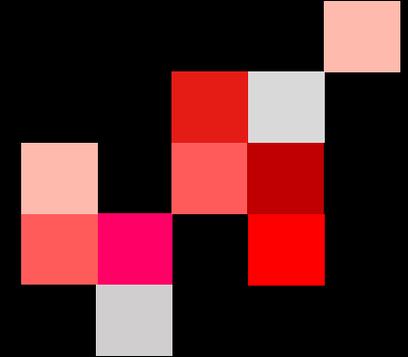
DIRECT MARKETING

SMS AND NEWSLETTERS COMMUNICATION



CAMPAIGN MANAGEMENT

DIRECT MARKETING



SMS AND NEWSLETTERS COMMUNICATION



**CUSTOMER
SEGMENTATION**

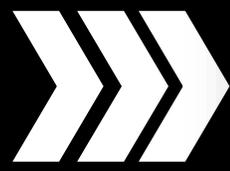
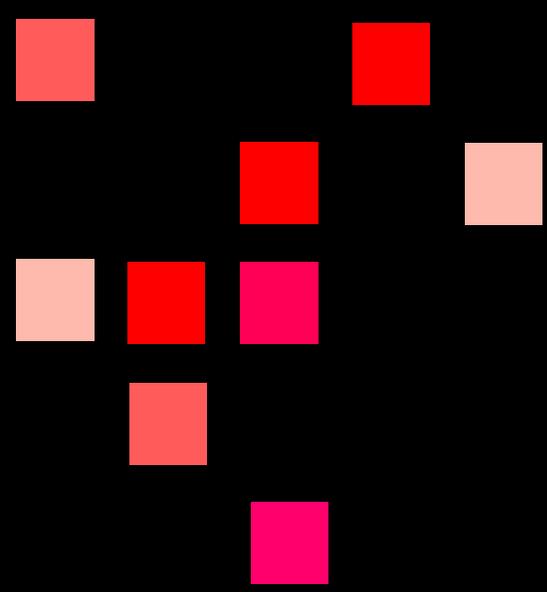


VALUE

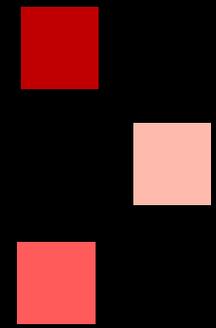
PRODUCT

LIFECYCLE

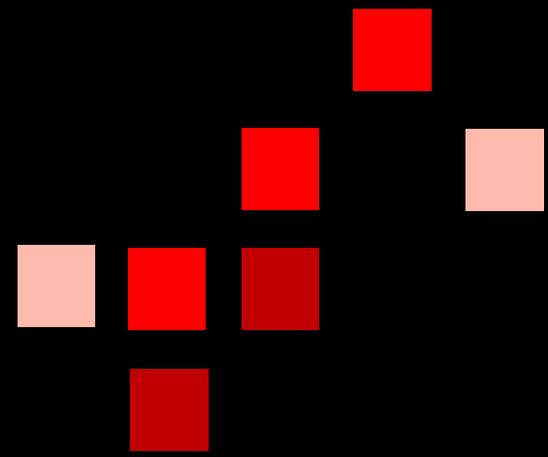
PURCHASE CHANNEL



HOW DO WE **USE** DATA?



DATA ANALYSIS



»» They should be integrated so that they maximize the return (value) they bring to a company

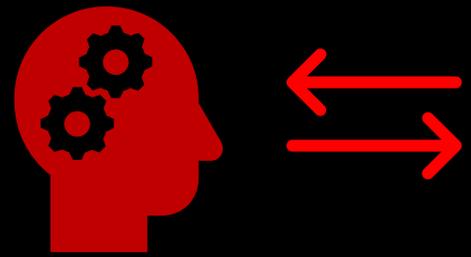
- Descriptive**
What happened?
- Diagnosis**
Why did it happen?

4 TYPES OF DATA ANALYSIS

Predictive
What will happen?

Prescriptive
How can we make it happen?

BIG DATA

The background of this section is a blue-toned image with binary code (0s and 1s) and data streams, giving it a digital, high-tech appearance.

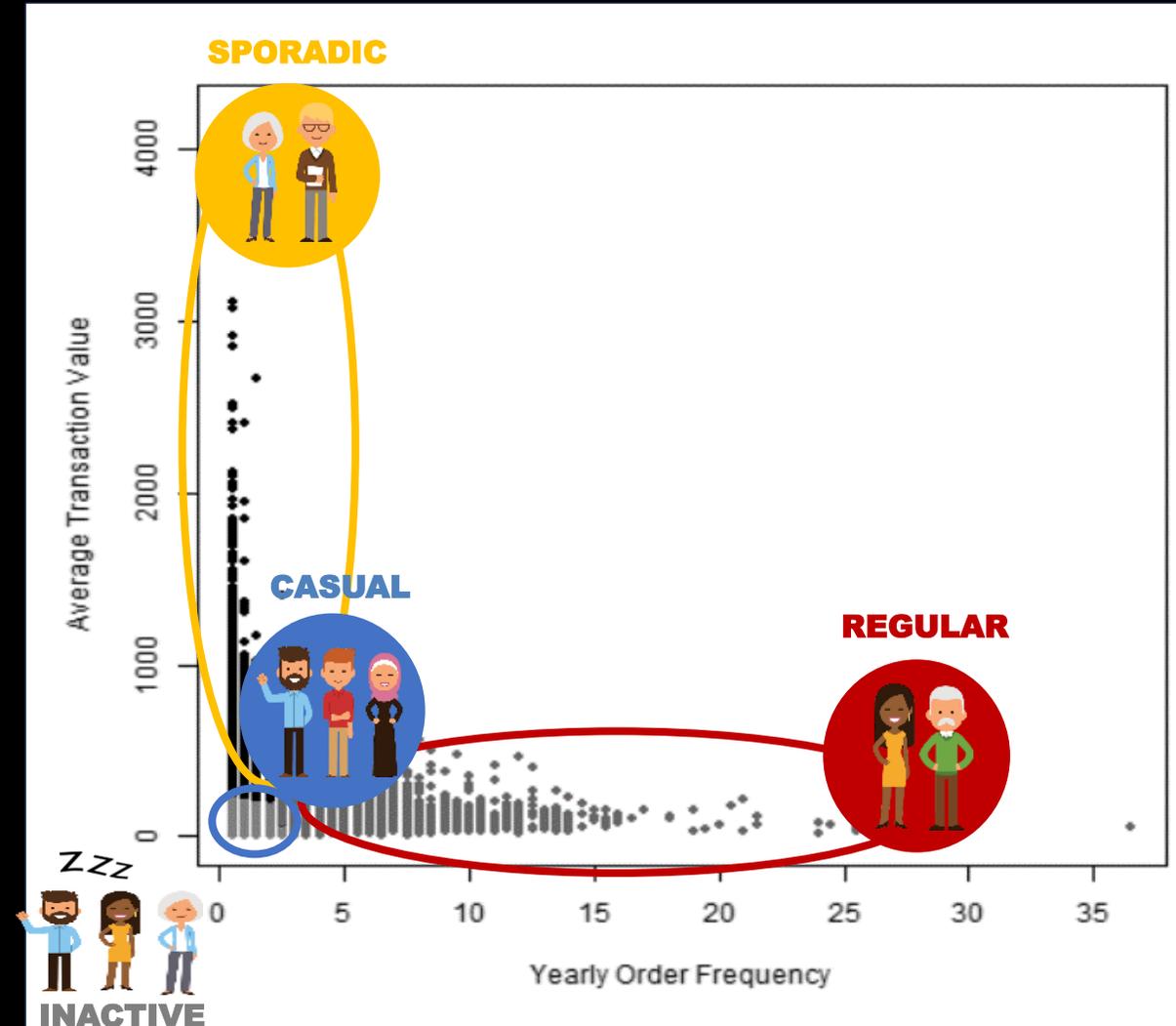
AI & MACHINE LEARNING

FREQUENCY & BASKET VALUE PREDICTION

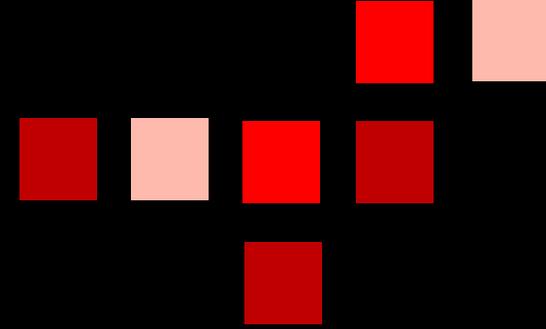
Identifying customer moves between purchases allows us to:

Predict **how much** and **what** the customer will purchase in the future

Identify and incentivize customer behavior



CUSTOMER SEGMENTATION



To have a consistent and simple customer description we defined several approaches to customer segmentation. These are 3 of them:

- **Value/Frequency Segments:** based on order frequency and average basket value;
- **Product Interest Segments:** based on the types of product the customers buy;
- **Trend Segments:** a classification of customer activity/recency;

CUSTOMER SEGMENTATION

Value/Frequency

The Value/Frequency segment is computed by a **k-medoids** algorithm – a ‘k-means’ but where the cluster centroids have to be part of the sample.

This process runs on these variables:

- Average Yearly Order Frequency (# Orders)
- Average Basket Value (€)

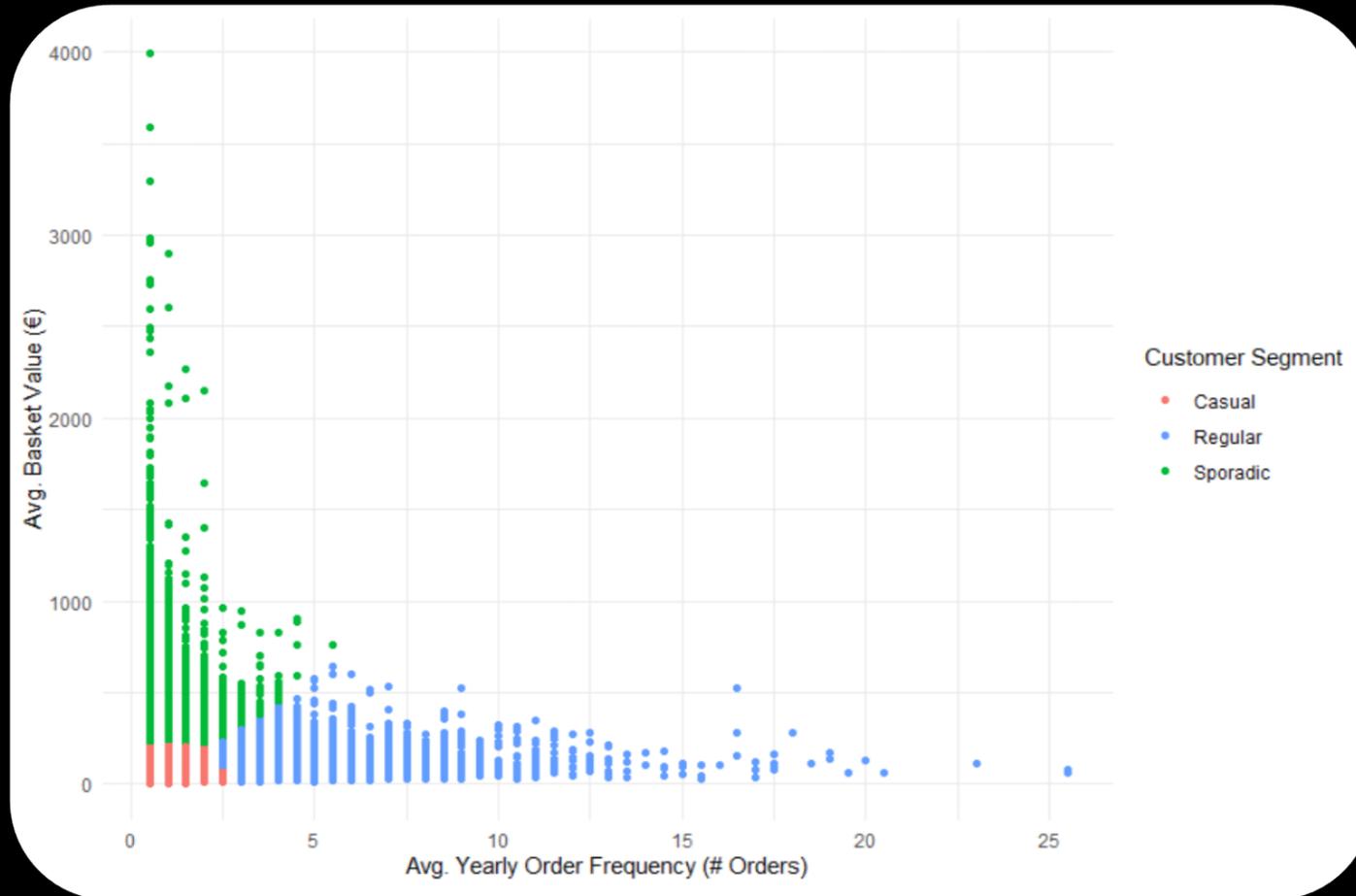
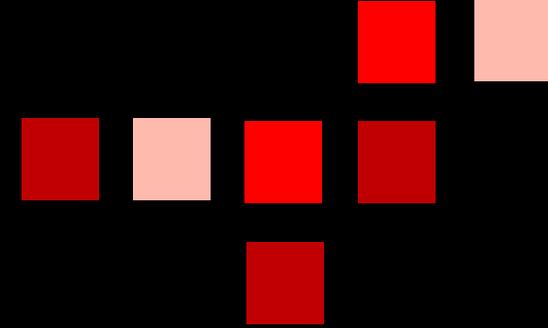
The resulting clusters are the following:

- Casual: a customer with low frequency and basket value
- Sporadic: a customer with high basket value
- Regular: a customer with high frequency

A high frequency and high basket value cluster didn't show up because the basket value tends to decrease with frequency since it essentially is computed as
$$\text{Avg Basket Value} = \frac{\text{Total Value}}{\text{Order Frequency}}$$

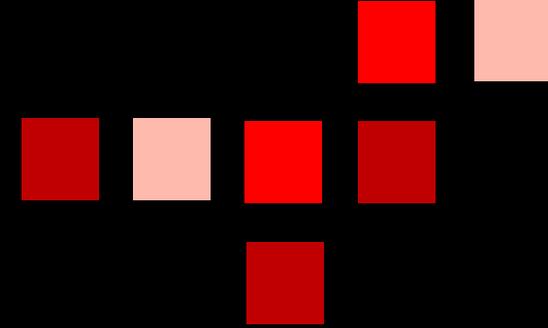
CUSTOMER SEGMENTATION

Value/Frequency



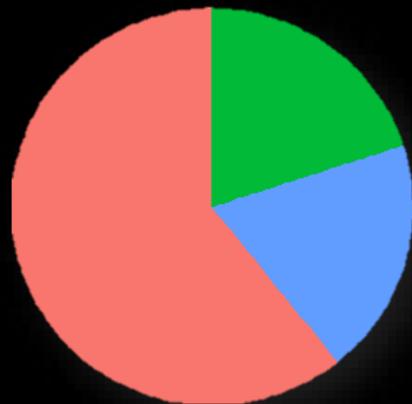
A look into the three segments for a sample of 20.000 customers

CUSTOMER SEGMENTATION

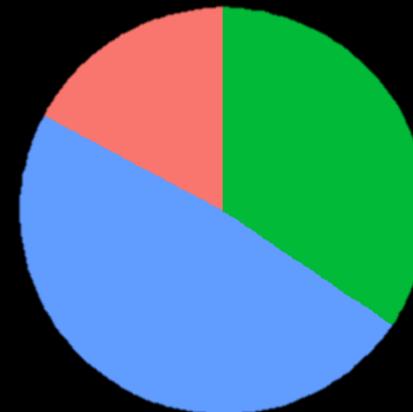


Value/Frequency

Migrating **Casual** customers to higher segments is very much of interest as the smaller **Sporadic** and **Regular** segments make up more than 80% of the value spent by all customers. This means creating a relationship where the customers trust us with their high value purchases and frequently visit our stores.



Customers per segment



Total value spent (€) per segment

CUSTOMER SEGMENTATION

Product Interest

To measure customer interest for product categories we divided our products into 14 categories:

- Big appliances
- Small appliances
- Telecom
- TV
- IT
- Audio
- Services
- Accessories
- Tickets
- Gaming
- Entertainment
- Photography
- Outdoors
- Marketplace Only

We found that interests would be skewed to be higher for expensive items like big appliances if we focused on value and higher for cheap things like accessories if we focused on number of items bought.

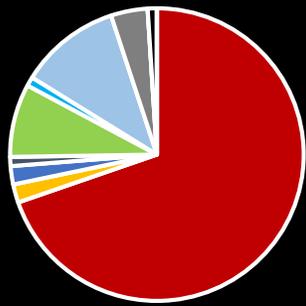
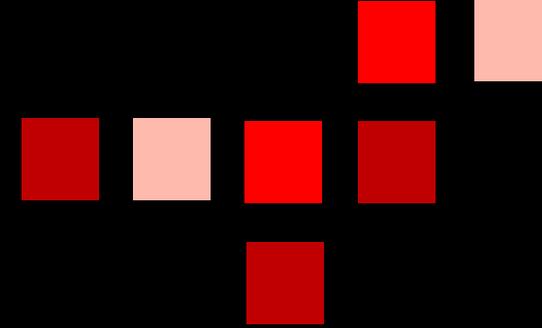
To balance this, for customer U we compute the interest for a certain category C as follows:

Interest =
(% of items U bought that belong to C) × (% of value spent by U that belongs to C)

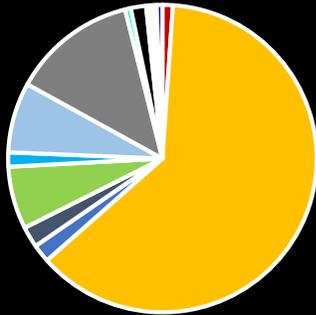
Then we apply a k-means algorithm on this 14-dimensional space.

CUSTOMER SEGMENTATION

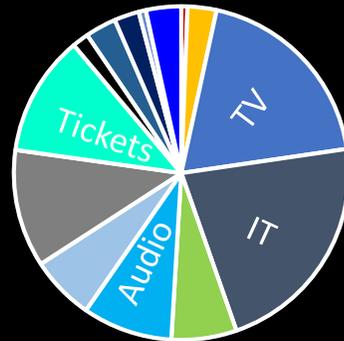
Product Interest – Cluster Centroids



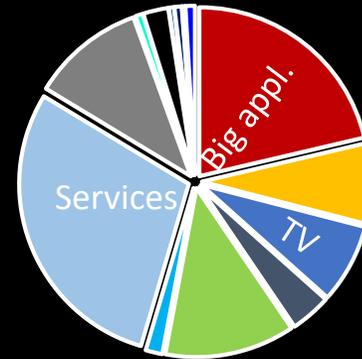
Big appliances



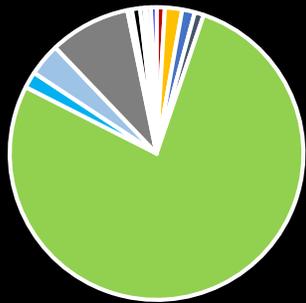
Telecom



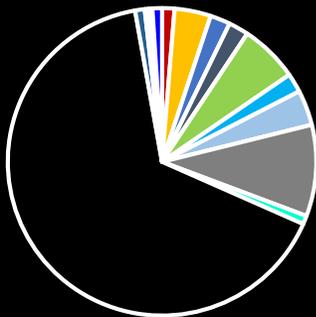
General Tech



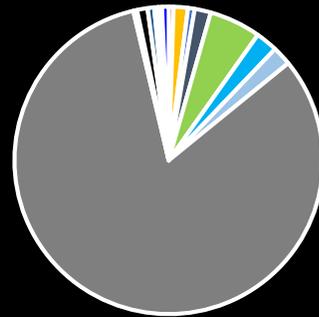
Home



Small appliances



Gaming



Accessories

- Big appliances
- Telecom
- TV
- IT
- Small appliances
- Audio
- Services
- Accessories
- Tickets
- Gaming
- Entertainment
- Photography
- Marketplace Only
- Outdoors

CUSTOMER SEGMENTATION

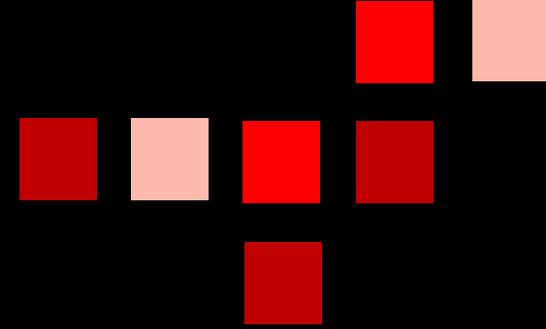
Trend

To add to the information given by the value/frequency segments and the product interest segments, we classify the customers according to their activity based on the last 24 months with these labels:

Segment	12 to 24 months ago	Last 12 months
New	Unregistered	First Purchase!
Recent	First purchase!	
Escaping	Active (made purchases)	Inactive (no purchases)
Regained	Inactive	Active
Lost	Inactive	Inactive
Stable	Active	Active (same Value/Frequency Segment as the other period)
Rising	Active	Active (stronger V/F segment)*
Declining	Active	Active (weaker V/F segment)*

* Casual < Sporadic < Regular

CUSTOMER SEGMENTATION



In the end these approaches together form a simple but accurate description of the customer's history

Customer	Value/Frequency	Product Interest	Trend
#001	Sporadic	Gaming	New
#002	Regular	Home	Recent
#003	Casual	Accessories	Stable
#004	Sporadic	Big Appliances	Regained
#005	Inactive	Telecom	Lost

NEXT BASKET

PREDICTION

Frequent Itemsets

- **Simplest question:** Find sets of items that appear together “frequently” in baskets
- **Support** for itemset I : Number of baskets containing all items in I
 - (Often expressed as a fraction of the total number of baskets)
- Given a **support threshold s** , then sets of items that appear in at least s baskets are called **frequent itemsets**

<i>TID</i>	<i>Items</i>
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Support of
{Beer, Bread} = 2

NEXT BASKET

PREDICTION

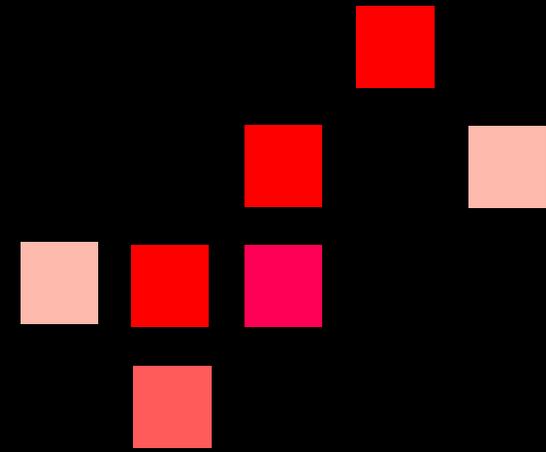
Association Rules

- **Association Rules:**
If-then rules about the contents of baskets
- $\{i_1, i_2, \dots, i_k\} \rightarrow j$ means: “if a basket contains all of i_1, \dots, i_k then it is *likely* to contain j ”
- **In practice there are many rules, want to find significant/interesting ones!**
- **Confidence** of this association rule is the probability of j given $I = \{i_1, \dots, i_k\}$

$$\text{conf}(I \rightarrow j) = \frac{\text{support}(I \cup j)}{\text{support}(I)}$$

NEXT BASKET

PREDICTION₂₇₀



»»» Average Yearly Order Frequency - 1.71

»»» Average Number of Distinct Products per Transaction – 1.45

**PREDICTIONS USING STANDARD
METHODS ARE VERY DIFFICULT TO
OBTAIN**

PROJECTS & ANALYTICS

STATISTICAL MODELS
DEVELOPMENT
AND AI

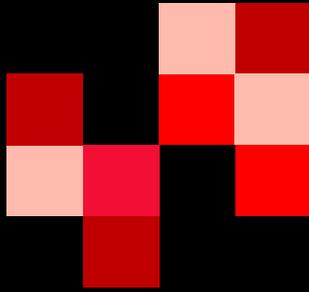
PURCHASE PROPENSITY
PROFILING
RECOMMENDATION



TAKE ADVISORY

AND PERSONALIZATION
TO THE NEXT LEVEL

RECOMMENDER SYSTEMS

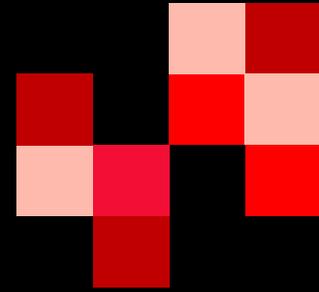


Some implemented/studied approaches:

- Collaborative Filtering
- Sequential Collaborative Filtering (several variations)
- TransRec (He, Kang, McAuley – 2017)
- Item2vec (Barkan, Koenigstein – 2017)
- FPMC (Rendle, Freudenthaler, Schmidt-Thieme – 2010)

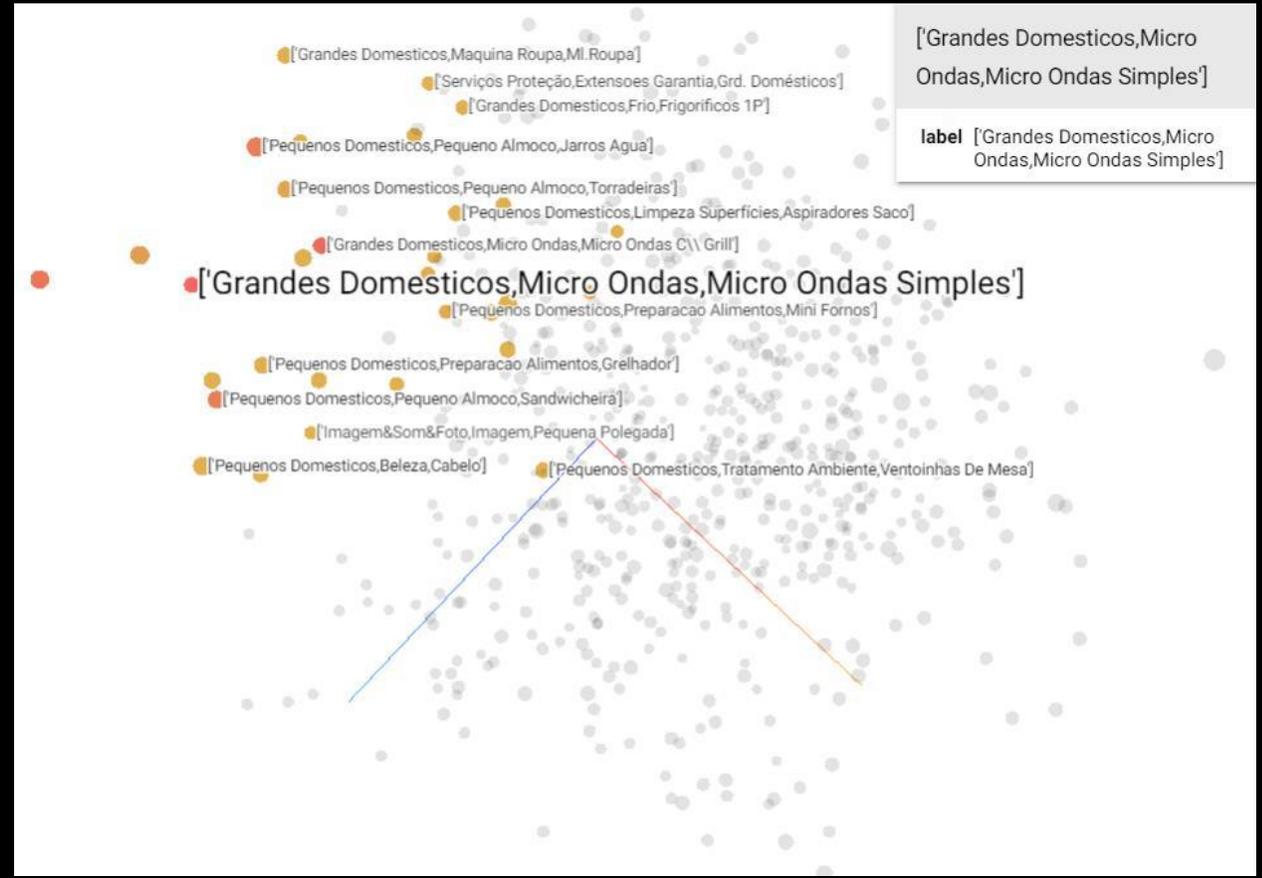
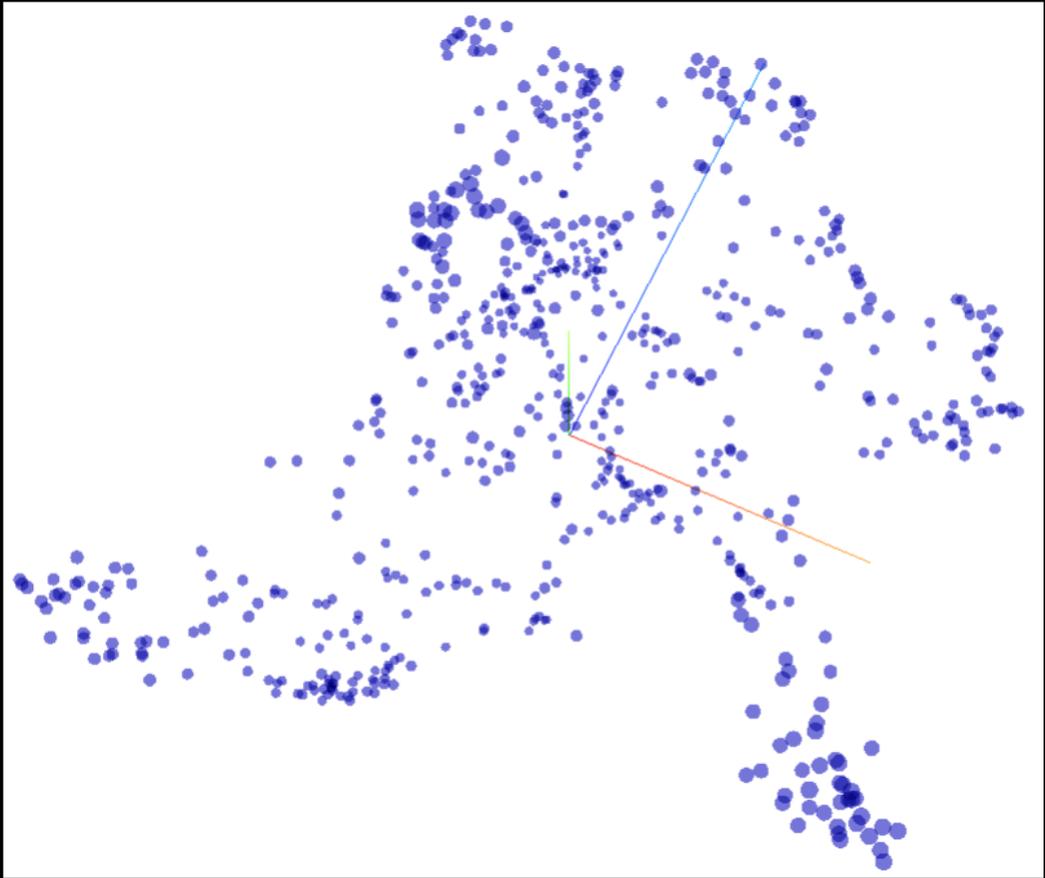
Software used: SAS, R, Python.

RECOMMENDER SYSTEMS



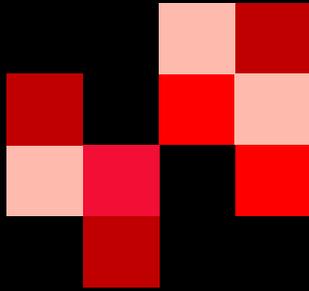
Item2Vec

Embeddings Visualization in Tensorflow



RECOMMENDER SYSTEMS

Sequential Collaborative Filtering (item-to-item)



Purchase History

User	Date	Item
#001	1	A
#001	2	B
#001	2	C
#001	3	D
#002	1	C
#002	2	A
#002	2	D

$$\text{Similarity}(i1,i2) = \frac{\#users \text{ that bought } i2 \text{ after buying } i1}{\sqrt{\#users \text{ that bought } i1} \times \sqrt{\#users \text{ that bought } i2}}$$

Similarity Matrix

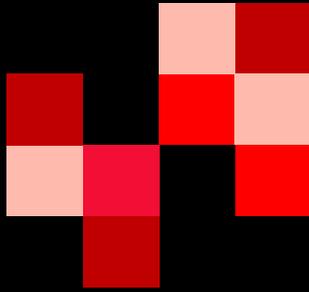
	A	B	C	D
A	-	0.707	0.5	0.5
B	0	-	0	0.707
C	0.5	0	-	1
D	0	0	0	-

This approach was found to capture the customers' **general interests** relatively better than others.

It generates recommended items based on the customer's **full purchase history**.

RECOMMENDER SYSTEMS

Consecutive Collaborative Filtering (item-to-item)



Purchase History

User	Date	Item
#001	1	A
#001	2	B
#001	2	C
#001	3	D
#002	1	C
#002	2	A
#002	2	D

$$\text{Similarity}(i1,i2) = \frac{\#users \text{ that bought } i1 \text{ and } i2 \text{ in consecutive transactions (} i1 \text{ first)}}{\sqrt{\#users \text{ that bought } i1} \times \sqrt{\#users \text{ that bought } i2}}$$

Similarity Matrix

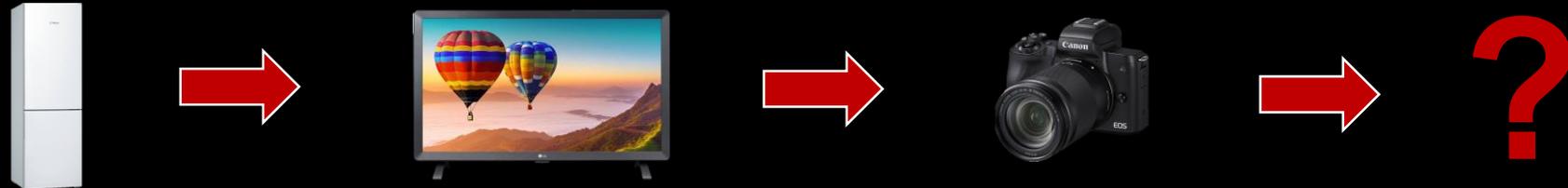
	A	B	C	D
A	-	0.707	0.5	0
B	0	-	0	0.707
C	0.5	0	-	1
D	0	0	0	-

This approach was found to capture the customers' **recent trends** relatively better than others.

It generates recommended items based on the customer's **latest order**.

RECOMMENDER SYSTEMS

BEST RESULTS SO FAR



Sequential CF (to capture general interest)

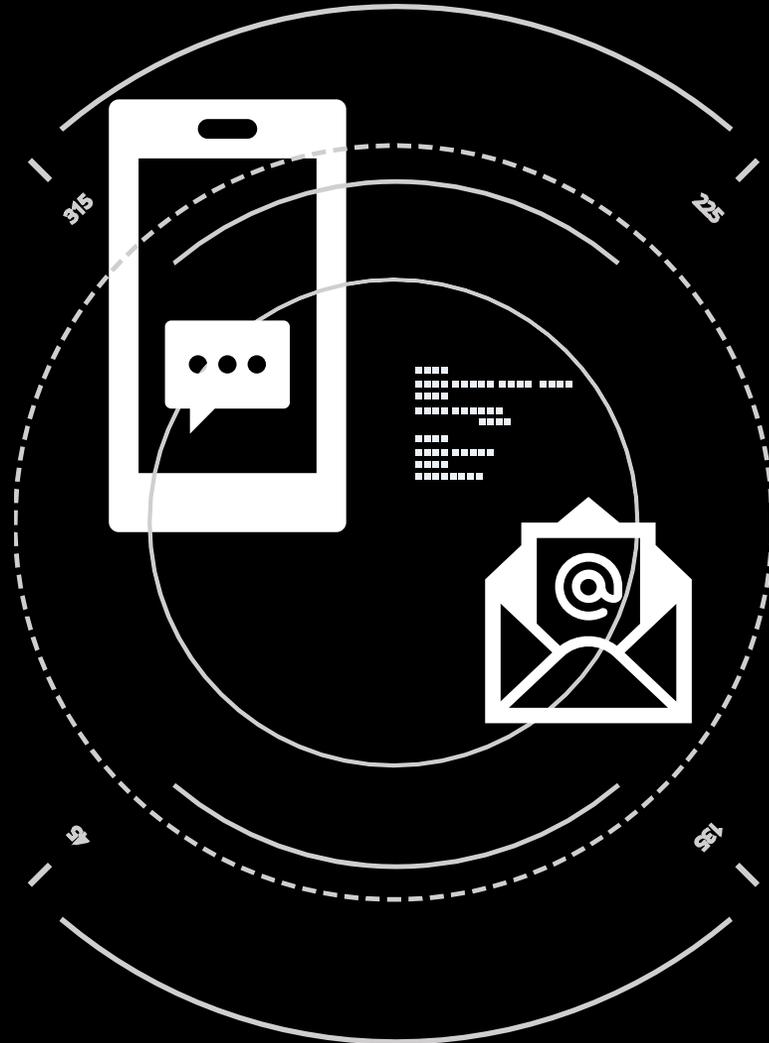
Consecutive CF (to capture recent trend)

Ensemble Algorithm: Combined Similarities between Customers and Items from both models.
hit@10 = 34.8% for Regular Customers

CAMPAIGN RESPONSE

PREDICTION

270



STUDYING CUSTOMERS'
HISTORICAL RESPONSE TO
CAMPAIGNS



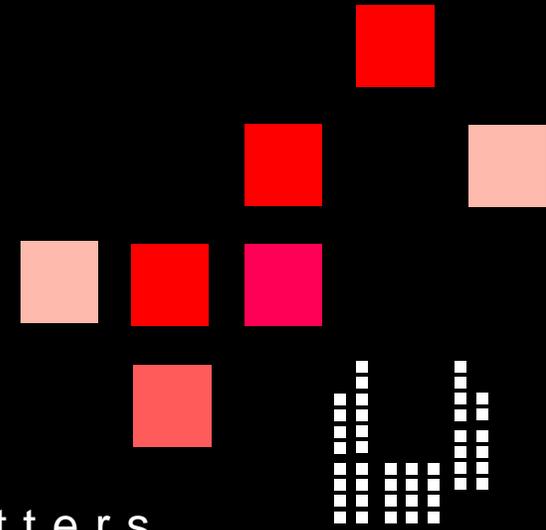
Predict future response results
to target the right customers

Maximize incremental sales

Maximize customer satisfaction

worten

PERSONALIZATION



Personalized coupons

worten

SAMSUNG
Galaxy S21 Series 5G

Temos uma oferta para ti!
Damos-te 50€ ou 100€ de desconto direto para comprares o teu Galaxy S21 ou S21+
Só até 21 de fevereiro

Galaxy S21

50€ de desconto com o código

CÓDIGO

COMPRA JÁ

Galaxy S21+

100€ de desconto com o código

CÓDIGO

COMPRA JÁ

Cross-sell newsletters

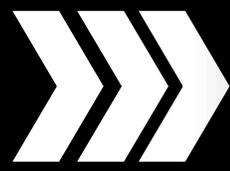
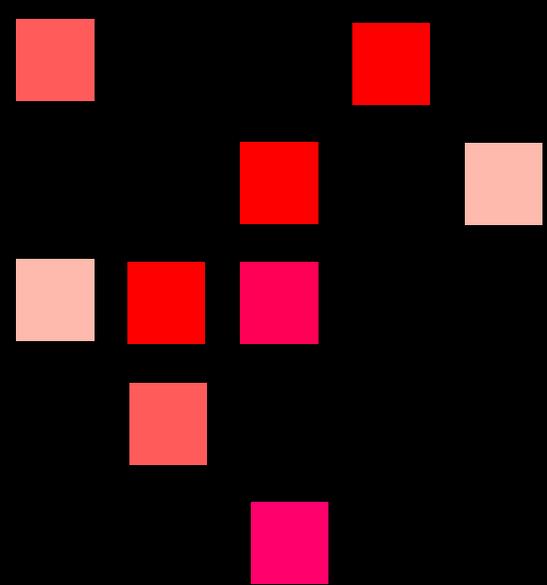
worten

Olá %%First Name%%

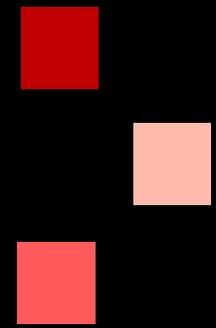
Obrigado por comprares o teu iPhone na Worten.
Descobre agora como podes tirar o melhor proveito dele!

Desfruta ao máximo do teu novo iPhone

worten



ONLINE DATA



SEARCH ENGINE

MARKETING

SEARCH ENGINE OPTIMIZATION (SEO)

+

SEARCH ENGINE ADVERTISING (SEA)

=

SEARCH ENGINE MARKETING (SEM)

SEO

Search Engine Optimization

SERP

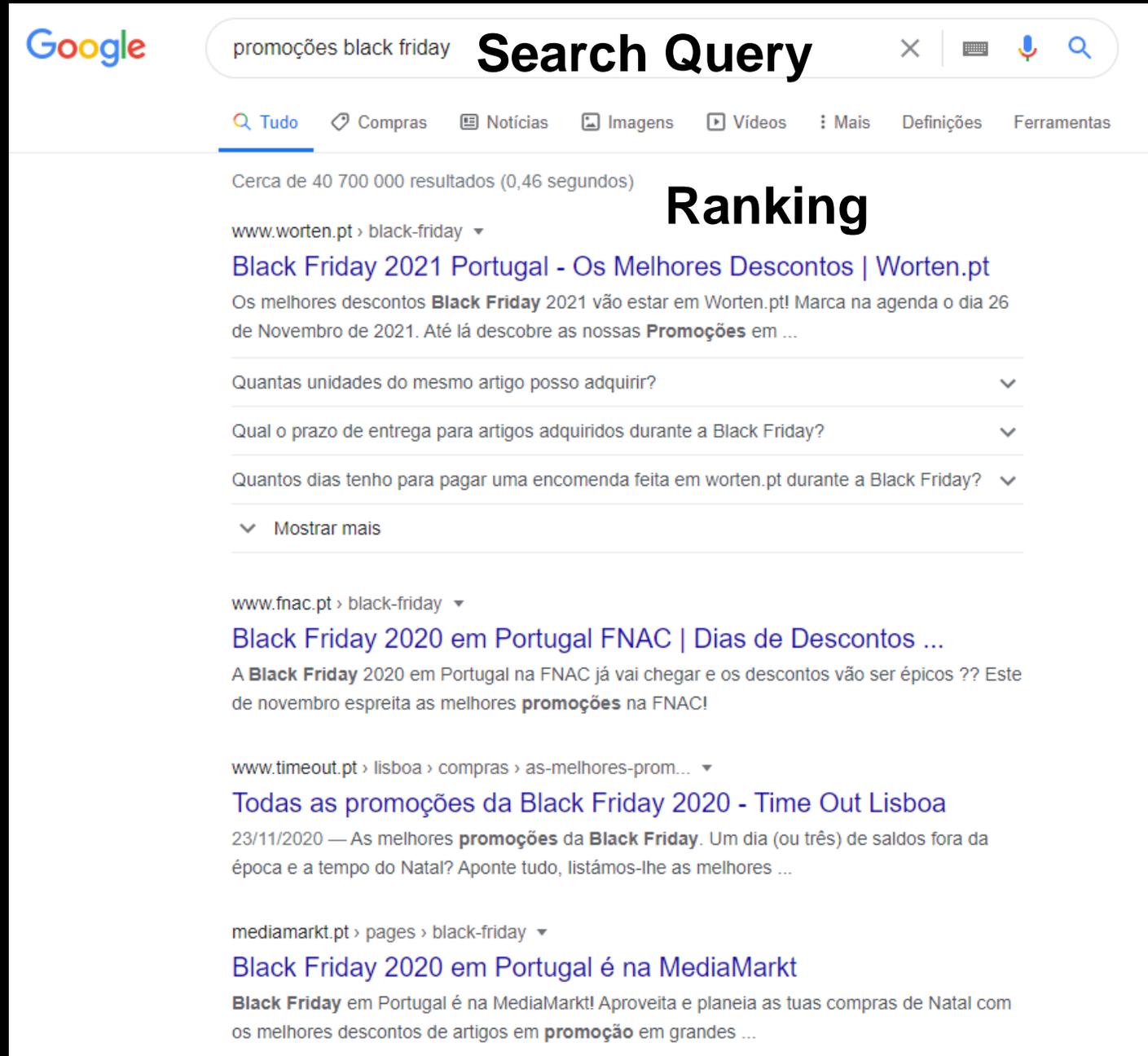
Search

Engine

Result

Page

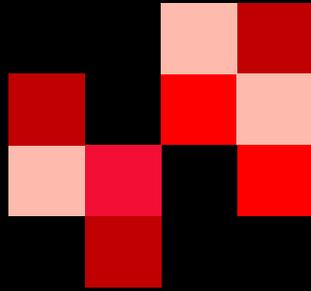
Based on KeyWords



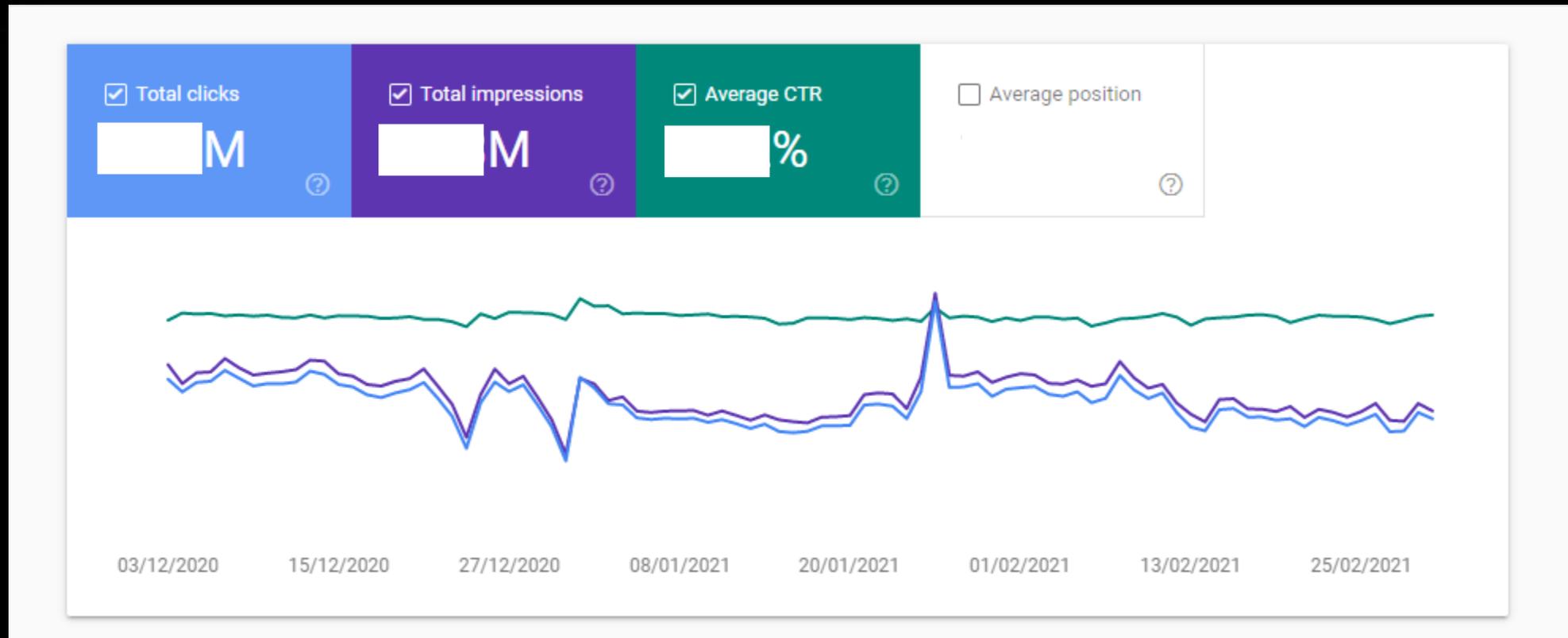
The image shows a Google search results page for the query "promoções black friday". The search bar at the top contains the text "promoções black friday" and "Search Query". Below the search bar, there are navigation links for "Tudo", "Compras", "Notícias", "Imagens", "Vídeos", "Mais", "Definições", and "Ferramentas". The search results show approximately 40,700,000 results in 0.46 seconds. The word "Ranking" is prominently displayed in large black text. The first result is from "www.worten.pt" titled "Black Friday 2021 Portugal - Os Melhores Descontos | Worten.pt". The snippet for this result reads: "Os melhores descontos **Black Friday** 2021 vão estar em Worten.pt! Marca na agenda o dia 26 de Novembro de 2021. Até lá descobre as nossas **Promoções** em ...". Below the snippet are three expandable sections with dropdown arrows: "Quantas unidades do mesmo artigo posso adquirir?", "Qual o prazo de entrega para artigos adquiridos durante a Black Friday?", and "Quantos dias tenho para pagar uma encomenda feita em Worten.pt durante a Black Friday?". A "Mostrar mais" link is also present. The second result is from "www.fnac.pt" titled "Black Friday 2020 em Portugal FNAC | Dias de Descontos ...". The snippet reads: "A **Black Friday** 2020 em Portugal na FNAC já vai chegar e os descontos vão ser épicos ?? Este de novembro espreita as melhores **promoções** na FNAC!". The third result is from "www.timeout.pt" titled "Todas as promoções da Black Friday 2020 - Time Out Lisboa". The snippet reads: "23/11/2020 — As melhores **promoções** da **Black Friday**. Um dia (ou três) de saldos fora da época e a tempo do Natal? Aponte tudo, listámos-lhe as melhores ...". The fourth result is from "mediamarkt.pt" titled "Black Friday 2020 em Portugal é na MediaMarkt". The snippet reads: "**Black Friday** em Portugal é na MediaMarkt! Aproveita e planeia as tuas compras de Natal com os melhores descontos de artigos em **promoção** em grandes ...".

SEARCH ENGINE

MARKETING

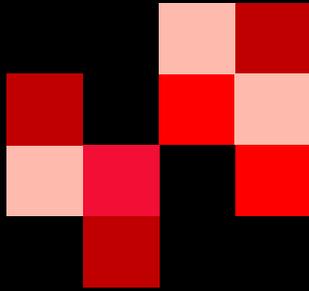


Google Search Console

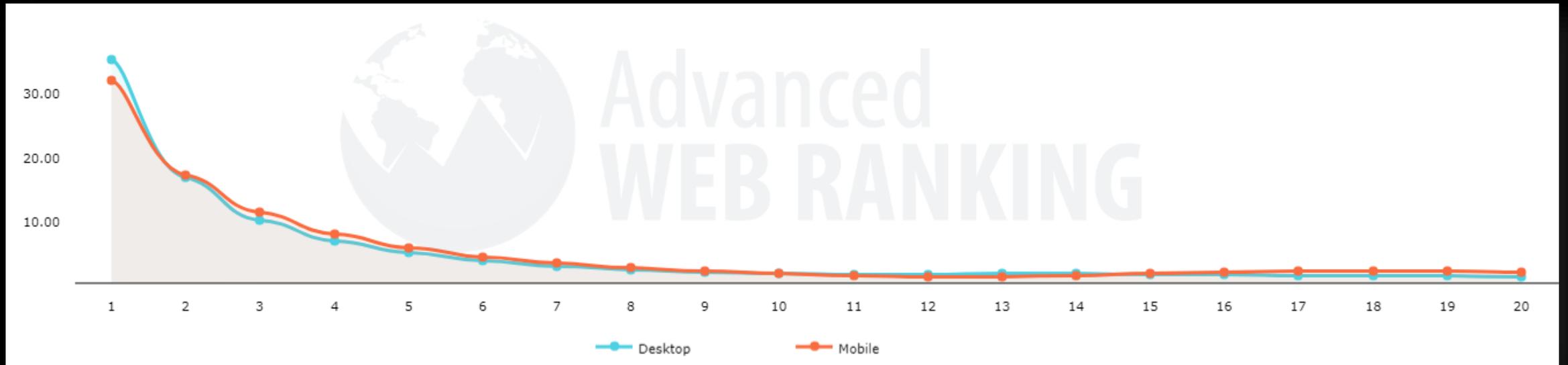


SEARCH ENGINE

MARKETING



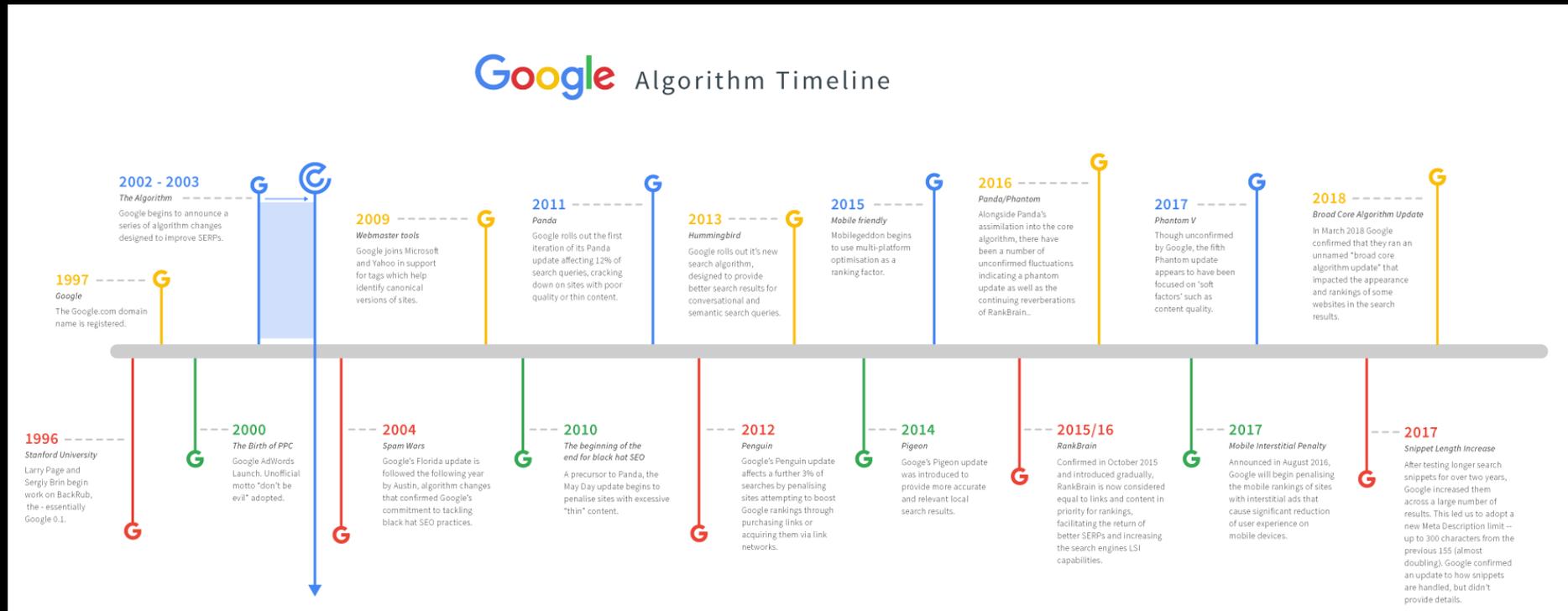
CTR – Click Through Rate



SEARCH ENGINE

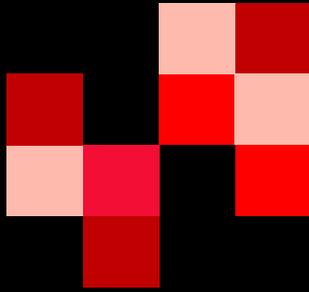
MARKETING

Algorithms are frequently changed and not completely known, so constant optimization is required.

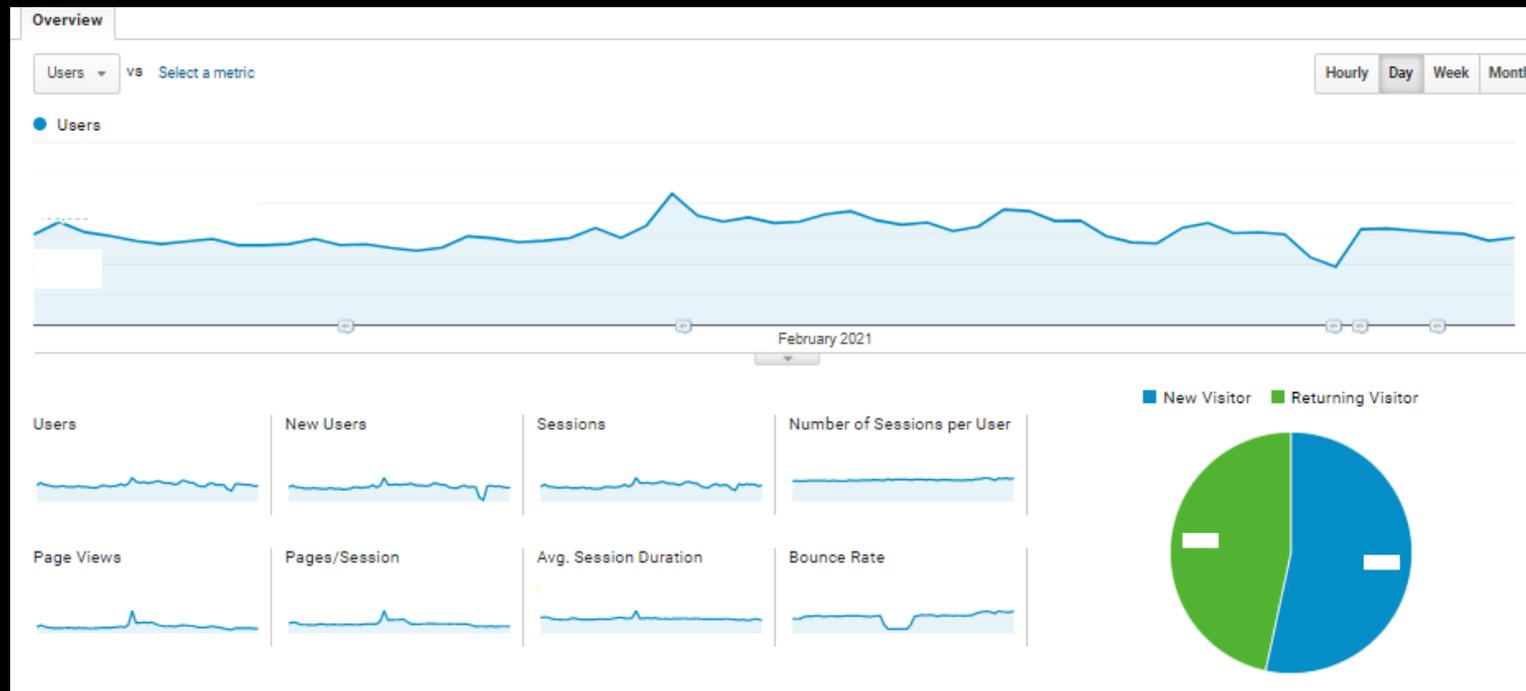


SEARCH ENGINE

MARKETING



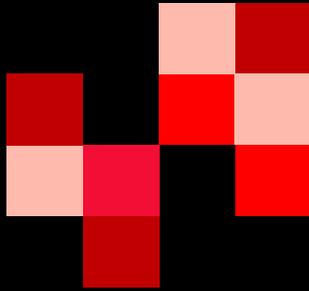
Google Analytics



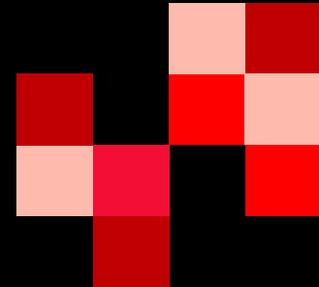
The continuous optimization strategy aims at increasing relevant KPIs, such as PageViews, PageSessions and Avg Sessions

SEA

Search Engine Advertising



SEARCH ENGINE MARKETING

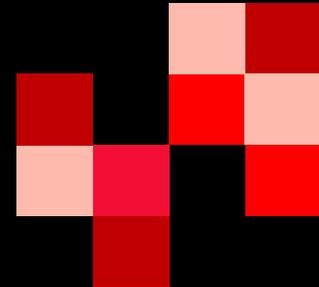


A look on the Web Advertising providers' side of the problem...

Performance-based Advertising

- **Introduced by Overture around 2000**
 - Advertisers **bid** on **search keywords**
 - When someone searches for that keyword, the **highest bidder's ad is shown**
 - Advertiser is charged only if the ad is clicked on
- Similar model adopted by Google with some changes around 2002
 - Called **Adwords**

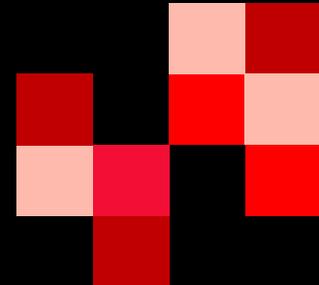
SEARCH ENGINE MARKETING



A look on the Web Advertising providers' side of the problem...

Web 2.0

- **Performance-based advertising works!**
 - Multi-billion-dollar industry
- **Interesting problem:**
What ads to show for a given query?



Ads vs. Search Results

Web

Results 1 - 10 of about 2,230,000 for **geico**. (0.04 sec)

[GEICO Car Insurance. Get an auto insurance quote and save today ...](#)

GEICO auto insurance, online car insurance quote, motorcycle insurance quote, online insurance sales and service from a leading insurance company.

[www.geico.com/](#) - 21k - Sep 22, 2005 - [Cached](#) - [Similar pages](#)

[Auto Insurance](#) - [Buy Auto Insurance](#)

[Contact Us](#) - [Make a Payment](#)

[More results from www.geico.com »](#)

[Geico, Google Settle Trademark Dispute](#)

The case was resolved out of court, so advertisers are still left without legal guidance on use of trademarks within ads or as keywords.

[www.clickz.com/news/article.php/3547356](#) - 44k - [Cached](#) - [Similar pages](#)

[Google and GEICO settle AdWords dispute | The Register](#)

Google and car insurance firm **GEICO** have settled a trade mark dispute over ... Car insurance firm **GEICO** sued both Google and Yahoo! subsidiary Overture in ...

[www.theregister.co.uk/2005/09/09/google_geico_settlement/](#) - 21k - [Cached](#) - [Similar pages](#)

[GEICO v. Google](#)

... involving a lawsuit filed by Government Employees Insurance Company (GEICO). GEICO has filed suit against two major Internet search engine operators, ...

[www.consumeraffairs.com/news04/geico_google.html](#) - 19k - [Cached](#) - [Similar pages](#)

Sponsored Links

[Great Car Insurance Rates](#)

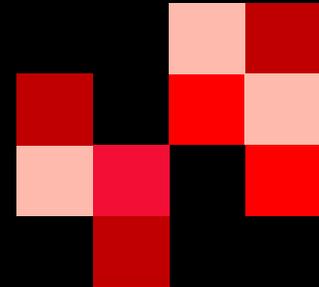
Simplify Buying Insurance at Safeco
See Your Rate with an Instant Quote
[www.Safeco.com](#)

[Free Insurance Quotes](#)

Fill out one simple form to get
multiple quotes from local agents.
[www.HometownQuotes.com](#)

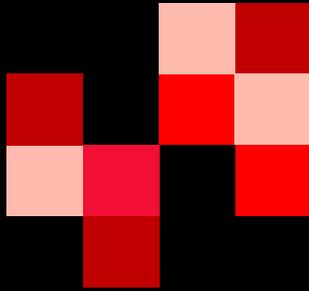
[5 Free Quotes. 1 Form.](#)

Get 5 Free Quotes In Minutes!
You Have Nothing To Lose. It's Free
[sayyessoftware.com/Insurance](#)
Missouri



Adwords Problem

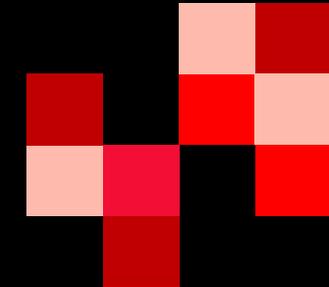
- **Given:**
 - 1. A set of bids by advertisers for search queries
 - 2. A click-through rate for each advertiser-query pair
 - 3. A budget for each advertiser (say for 1 month)
 - 4. A limit on the number of ads to be displayed with each search query
- **Respond to each search query with a set of advertisers such that:**
 - 1. The size of the set is no larger than the limit on the number of ads per query
 - 2. Each advertiser has bid on the search query
 - 3. Each advertiser has enough budget left to pay for the ad if it is clicked upon



Adwords Problem

- A stream of queries arrives at the search engine: q_1, q_2, \dots
- Several advertisers bid on each query
- When query q_i arrives, search engine must pick a subset of advertisers whose ads are shown
- **Goal:** Maximize search engine's revenues
 - **Simple solution:** Instead of raw bids, use the "expected revenue per click" (i.e., Bid*CTR)

SEARCH ENGINE MARKETING

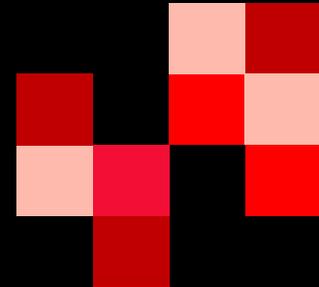


Adwords Problem

Advertiser	Bid	CTR	Bid * CTR
A	\$1.00	1%	1 cent
B	\$0.75	2%	1.5 cents
C	\$0.50	2.5%	1.125 cents

Click through rate Expected revenue

Advertiser	Bid	CTR	Bid * CTR
B	\$0.75	2%	1.5 cents
C	\$0.50	2.5%	1.125 cents
A	\$1.00	1%	1 cent



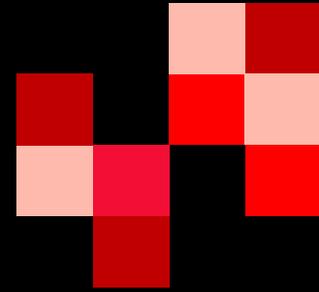
Adwords Problem Complications: CTR

- **CTR: Each ad has a different likelihood of being clicked**
 - **Advertiser 1** bids \$2, click probability = 0.1
 - **Advertiser 2** bids \$1, click probability = 0.5
 - **Clickthrough rate (CTR)** is measured **historically**
 - **Very hard problem: Exploration vs. exploitation**
 - Exploit:** Should we keep showing an ad for which we have good estimates of click-through rate
 - or**
 - Explore:** Shall we show a brand new ad to get a better sense of its click-through rate

There are several algorithms used to solve this problem:

- Greedy Algorithm
- Balanced Algorithm
- ...

ATTRIBUTION MODELS



 In the **Last Interaction** attribution model, the last touchpoint—in this case, the *Direct channel*—would receive 100% of the credit for the sale.

 In the **Last Non-Direct Click** attribution model, all direct traffic is ignored, and 100% of the credit for the sale goes to the last channel that the customer clicked through from before converting—in this case, the *Email* channel.

 In the **Last AdWords Click** attribution model, the last AdWords click—in this case, the first and only click to the *Paid Search* channel—would receive 100% of the credit for the sale.

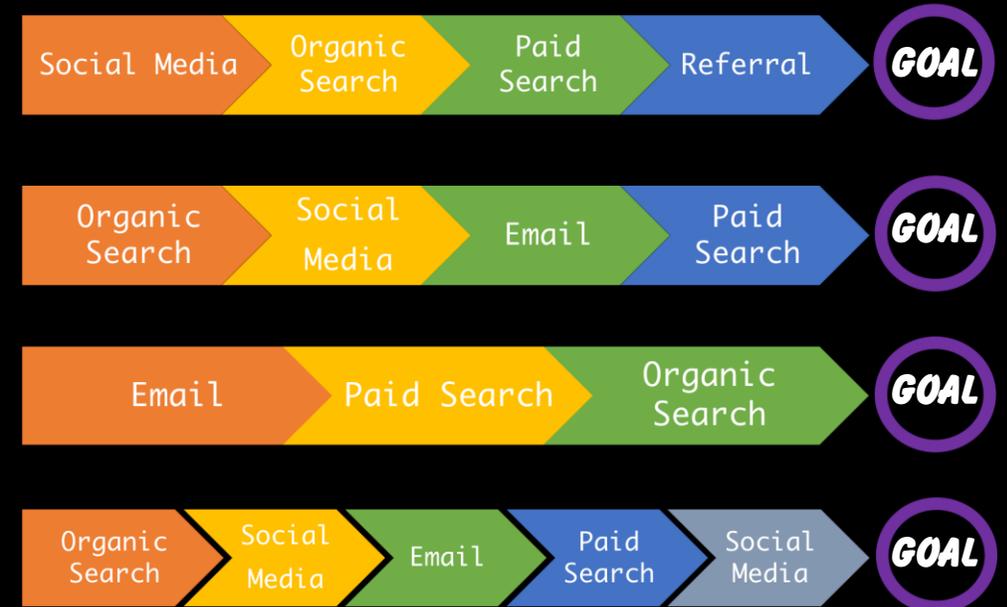
 In the **First Interaction** attribution model, the first touchpoint—in this case, the *Paid Search* channel—would receive 100% of the credit for the sale.

 In the **Linear** attribution model, each touchpoint in the conversion path—in this case the *Paid Search, Social Network, Email, and Direct* channels—would share equal credit (25% each) for the sale.

 In the **Time Decay** attribution model, the touchpoints closest in time to the sale or conversion get most of the credit. In this particular sale, the *Direct* and *Email* channels would receive the most credit because the customer interacted with them within a few hours of conversion. The *Social Network* channel would receive less credit than either the *Direct* or *Email* channels. Since the *Paid Search* interaction occurred one week earlier, this channel would receive significantly less credit.

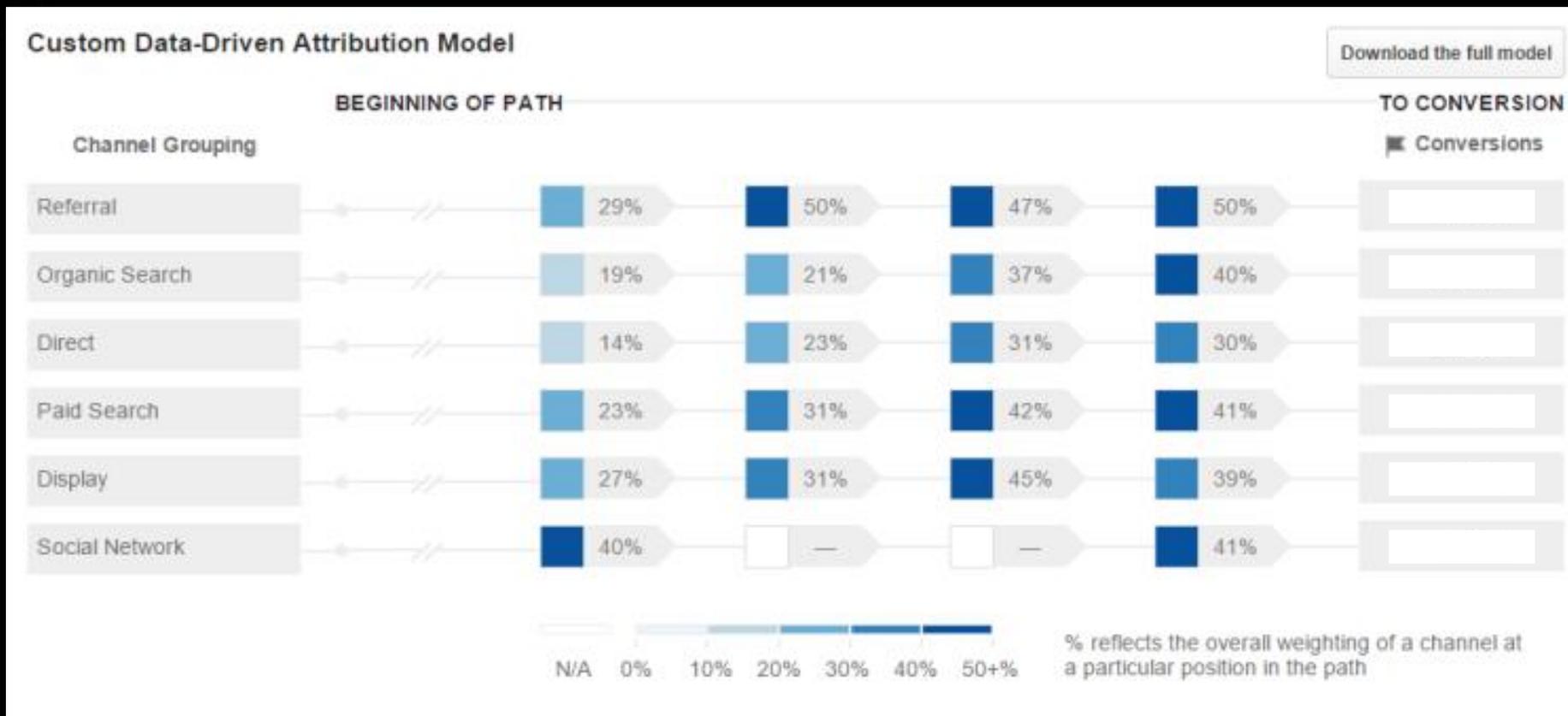
 In the **Position Based** attribution model, 40% credit is assigned to each the first and last interaction, and the remaining 20% credit is distributed evenly to the middle interactions. In this example, the *Paid Search* and *Direct* channels would each receive 40% credit, while the *Social Network* and *Email* channels would each receive 10% credit.

Used to model customer behaviour and understand what drives visits and conversions.



ATTRIBUTION MODELS

A continuous optimization problem...



BIG DATA

Online customers' information



PAGEVIEWS

FAVOURITES &
BANNER CLICKS

BEHAVIOURAL
SHIFT

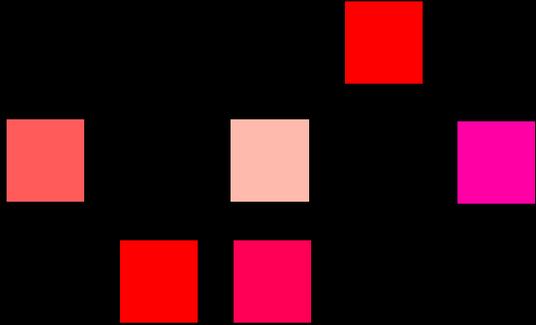
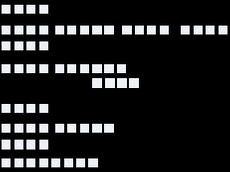
270
225
135
51.6
51.6

PADEMIC CONTEXT
ACCELERATED
eCOMMERCE GROWTH

NEW DIGITAL
CHANNELS
&
AI TOOLS

LARGER AMOUNT OF DATA
TO PROFILE THE CUSTOMER

QUESTIONS?



worten