

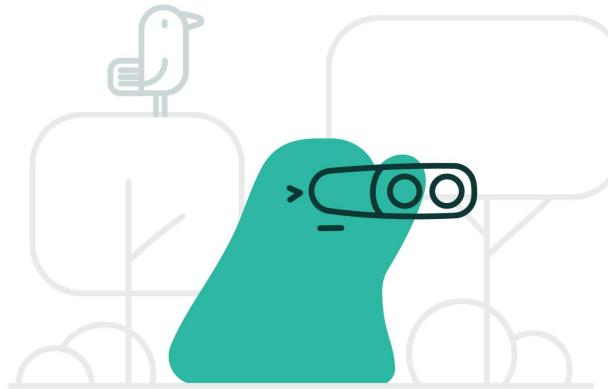
# Understanding Distance: Moving from Local Search to Nationwide Search

Xavier Sánchez Loro

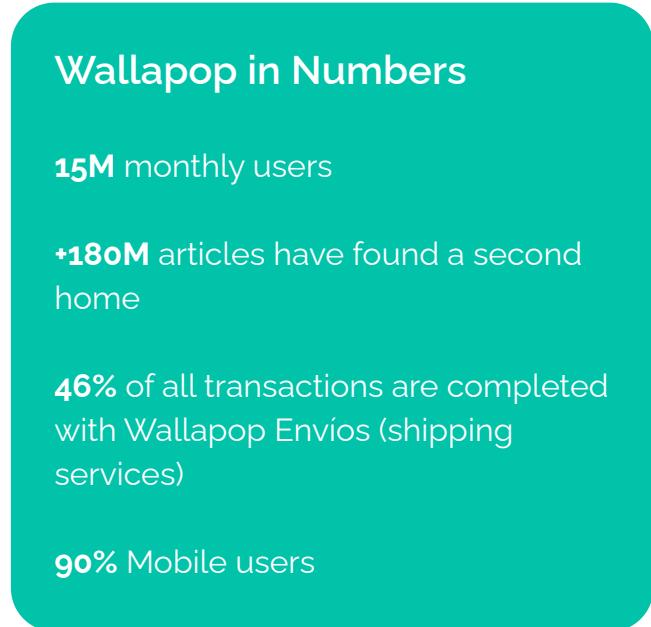
15/6/2022

## Outline

1. What's Wallapop?
2. The beginnings: Local search
3. Next generation: Nationwide search
4. Know your domain: Distance sensitivity
  - a. Context
  - b. Experiments
5. Conclusion



# What's Wallapop?

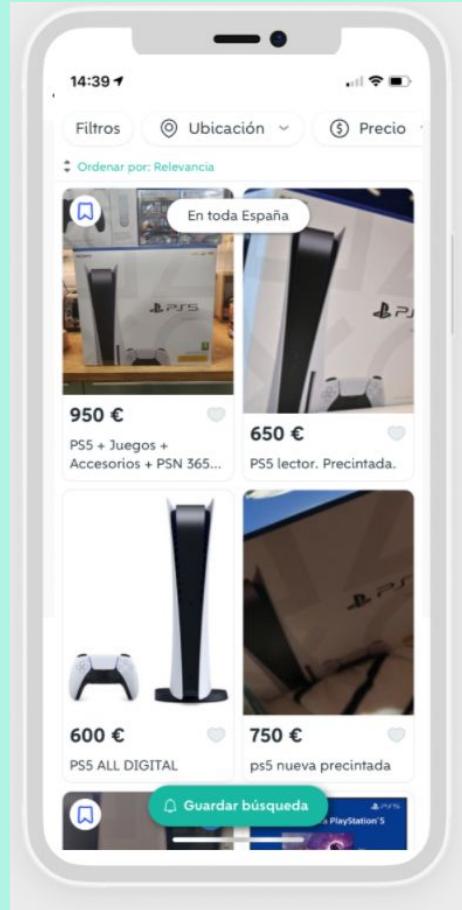


Wallapop started as a **local** marketplace in 2013

- Most transactions happened locally, face 2 face.
- Search was also local (showing nearby items first in search results)

In 2021 we introduced **nationwide search** with more focus on textual matching and less on distance.

- Good for Wallapop (shipping adoption) and for some users (opening more opportunities to discover unique catalog)
- But many users still prefer local sales in some cases.



# Wallapop Catalog Characteristics

## Catalog characteristics

Several tens of millions of products  
Hundreds of thousands of new  
(unique) products uploaded daily

Highly volatile user generated  
content

Geolocated

Wide array of product categories  
Consumer Goods  
Cars  
Real Estate  
Fashion

Support for shipping products  
nationwide and cross-border

## Catalog characteristics

Several tens of millions of products  
Hundreds of thousands of new  
(unique) products uploaded daily

Highly volatile user generated  
content

Geolocated

Wide array of product categories  
Consumer Goods  
Cars  
Real Estate  
Fashion

Support for shipping products  
nationwide and cross-border

## Catalog Volatility

Hundreds of thousands of new  
(unique) listings per day!

**Survival bias:** best items are  
sold quickly (**recency** is an  
important ranking factor)

Very **sparse** feedback per listing

Aggregations by product-id are  
not so useful

## Catalog characteristics

Several tens of millions of products  
Hundreds of thousands of new  
(unique) products uploaded daily

Highly volatile user generated  
content

Geolocated

Wide array of product categories  
Consumer Goods  
Cars  
Real Estate  
Fashion

Support for shipping products  
nationwide and cross-border

## Catalog Volatility

Hundreds of thousands of new  
(unique) listings per day!

**Survival bias:** best items are  
sold quickly (**recency** is an  
important ranking factor)

Very **sparse** feedback per listing

Aggregations by product-id are  
not so useful

## User Generated Content

**Inconsistent quality** of  
images and text (typos,  
ambiguity, keyword stuffing,  
etc..)

**Unstructured** listings (few  
attributes)

In many products, image  
tells more than text

# The beginnings: local search

## Local search characteristics

Each user **sees items around them** based on their location

**Not strict distance** sorting

Applying **relevance per distance areas**

Allows having the feeling of sorting by distance but allows to implement more sophisticated relevance per areas



## Local search characteristics

Visibility of products is **distributed according to geolocation** of the product uploaded and where buyer is searching

Products has **less competition** to be visible in their area for a certain keyword

**Topicality** of the item is **traded-off for locality and recency**



## Local search insights

Works really well for some users and product types

Relevance of the item is **biased towards location**

- Topicality suffers
- Highly relevant (and shippable) content is not shown to the user

Very useful for building a **strong local community**

**Not boosting transactional services adoption**



# Next generation: nationwide (and cross-border) search

## Search in all the country

Increase product visibility range

Increase competition between products

“good” shippable products sell faster

“bad” products are penalised

Reduced number of items sold

Increased number of shipping transactions

Some categories and type of products are better fitted for local search

## Search in all the country

Increase product visibility range

Increase competition between products

“good” shippable products sell faster

“bad” products are penalised

Reduced number of items sold

Increased number of shipping transactions

Some categories and type of products are better fitted for local search

## Several experiments online

Varying the impact of location and distance

With no distance as ranking factor

With distance as ranking factor

Content grouped by location and interleaved

## Search in all the country

Increase product visibility range

Increase competition between products

“good” shippable products sell faster

“bad” products are penalised

Reduced number of items sold

Increased number of shipping transactions

Some categories and type of products are better fitted for local search

## Several experiments online

Varying the impact of location and distance

With no distance as ranking factor

### **With distance as ranking factor**

Content grouped by location and interleaved

Users still don't see the same listing

Less variance on results per user than in local search

The different variants have basically very similar performance, with a better balance for variant with distance as ranking factor

The different variants have basically very similar performance, with a better balance for variant with distance as ranking factor

+24% in shipping  
transactions

The different variants have basically very similar performance, with a better balance for variant with distance as ranking factor

**+24%** in shipping  
transactions

**-17%** in local sales  
**-3 %** items sold

The different variants have basically very similar performance, with a better balance for variant with distance as ranking factor

**+24%** in shipping  
transactions

**+2,2%** CTR

**-17%** in local sales  
**-3 %** items sold

The different variants have basically very similar performance, with a better balance for variant with distance as ranking factor

**+24%** in shipping transactions

**+2,2%** CTR

Lower clicks positions,  
buyers click more on top-ranked items.

**-17%** in local sales  
**-3 %** items sold

The different variants have basically very similar performance, with a better balance for variant with distance as ranking factor

**+24%** in shipping transactions

**+2,2%** CTR

Lower clicks positions,  
buyers click more on  
top-ranked items.

**-17%** in local sales  
**-3 %** items sold

**-5.5%** Total chat  
decreased

The different variants have basically very similar performance, with a better balance for variant with distance as ranking factor

**+24%** in shipping transactions

**+2,2%** CTR

Lower clicks positions, buyers click more on top-ranked items.

**-17%** in local sales  
**-3 %** items sold

**-5.5%** Total chat decreased

**30%** of relevance searches had the distance filter applied or the order changed to distance

**Different behaviour** between Recurrent and New users

**Retention rate is not affected**

# Know your domain: Distance sensitivity

## From local search to relevance

Local search & local transactions



Relevance search & shipped transactions

We want the best of both worlds!

We saw **different level of sensitivity** to distance

We saw **different level of sensitivity** to distance

Some queries convert locally (bicicleta, iphone, sofa),  
some of them don't (records, pokemon cards, books, etc..)

We saw **different level of sensitivity** to distance

Related to product type/Query sensitivity

**Category + Subcategory**



Some queries convert locally (bicicleta, iphone, sofa),  
some of them don't (records, pokemon cards, books, etc..)

We saw **different level of sensitivity** to distance

Related to product type/Query sensitivity

**Category + Subcategory**



Some queries convert locally (bicicleta, iphone, sofa),  
some of them don't (records, pokemon cards, books, etc..)

Users in big cities convert more locally

Users countryside convert at larger distances

We saw **different level of sensitivity** to distance

Related to product type/Query sensitivity

**Category + Subcategory**

Some queries convert locally (bicicleta, iphone, sofa),  
some of them don't (records, pokemon cards, books, etc..)

**Location**

Users in big cities convert more locally

Users countryside convert at larger distances

We saw **different level of sensitivity** to distance

Related to product type/Query sensitivity

**Category + Subcategory**

Some queries convert locally (bicicleta, iphone, sofa),  
some of them don't (records, pokemon cards, books, etc..)

**Location**

Users in big cities convert more locally  
Users countryside convert at larger distances

Non shippable items have very low conversions on long distances

We saw **different level of sensitivity** to distance

Related to product type/Query sensitivity

**Category + Subcategory**

Some queries convert locally (bicicleta, iphone, sofa),  
some of them don't (records, pokemon cards, books, etc..)

**Location**

Users in big cities convert more locally

Users countryside convert at larger distances

**Item shippability**

Non shippable items have very low conversions on long distances

# Distance sensitivity experiments

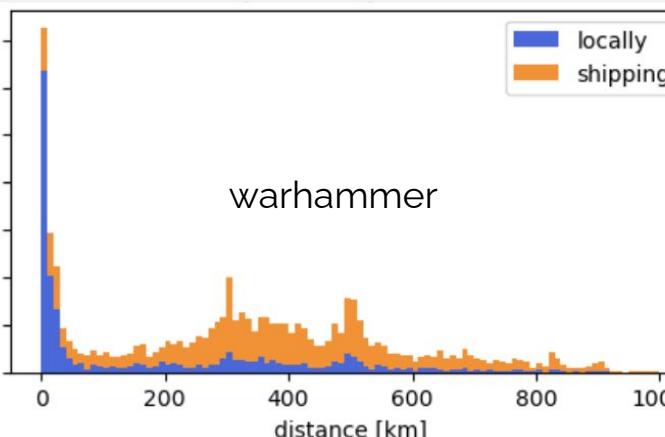
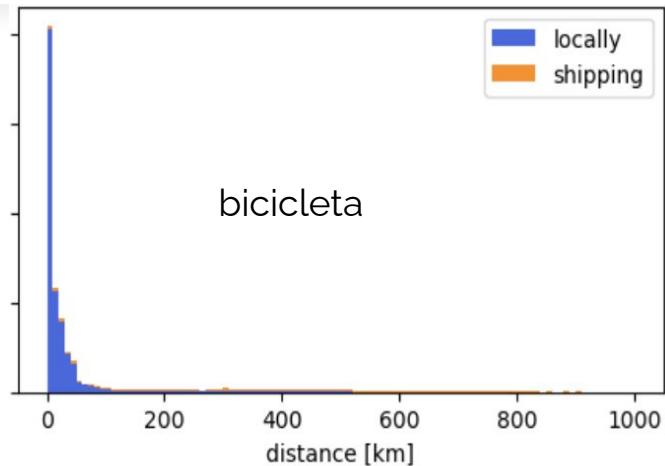
**Nationwide search is the baseline**

**Nationwide search is the baseline**

**Query sensitivity:** we **adapt the importance of distance**  
depending on the query

Nationwide search is the baseline

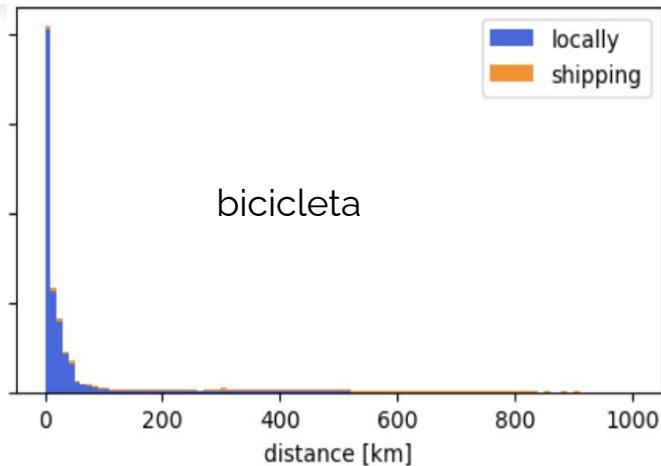
Query sensitivity: we adapt the importance of distance  
depending on the query



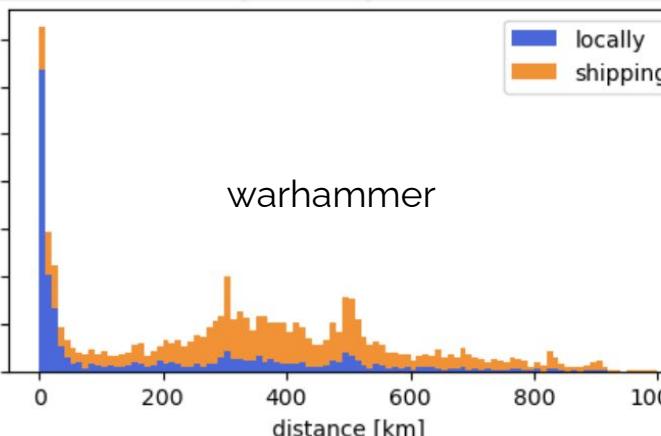
Nationwide search is the baseline

Query sensitivity: we **adapt the importance of distance** depending on the query

**Categories** and **subcategories** that cannot be shipped use **local search**



bicicleta



warhammer

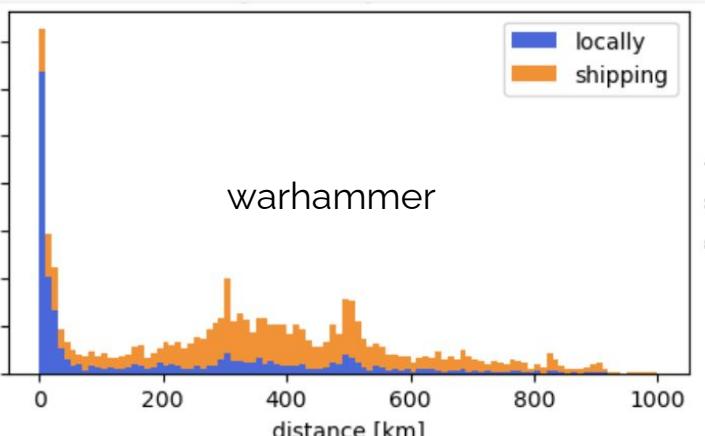
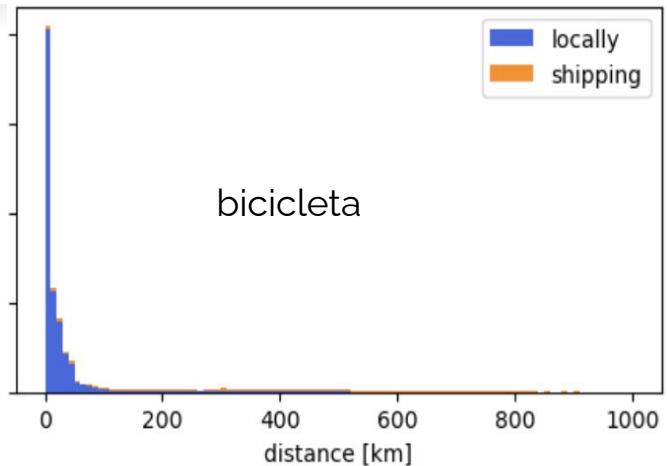
Nationwide search is the baseline

**Query sensitivity:** we **adapt the importance of distance** depending on the query

**Categories** and **subcategories** that cannot be shipped use **local search**

**Detection of keywords that are distance sensitive**

Several methods tested for keyword extraction



**Nationwide search is the baseline**

**Query sensitivity:** we **adapt the importance of distance** depending on the query

**Categories** and **subcategories** that cannot be shipped use **local search**

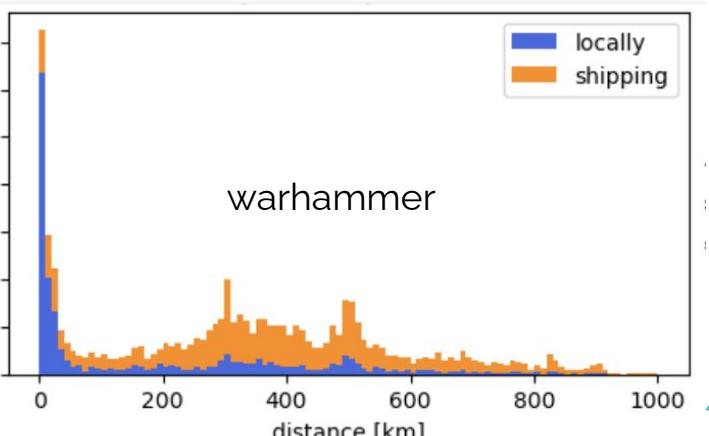
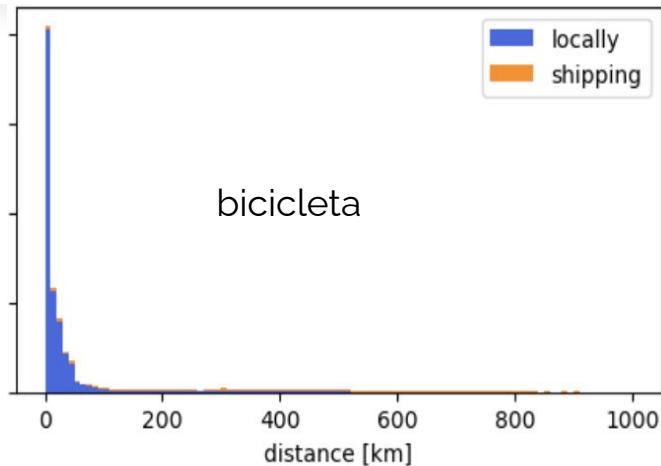
**Detection of keywords that are distance sensitive**

Several methods tested for keyword extraction

**Switch to local search**

Easy to test

Aggressive option



**Nationwide search is the baseline**

**Query sensitivity:** we **adapt the importance of distance** depending on the query

**Categories** and **subcategories** that cannot be shipped use **local search**

**Detection of keywords that are distance sensitive**

Several methods tested for keyword extraction

**Switch to local search**

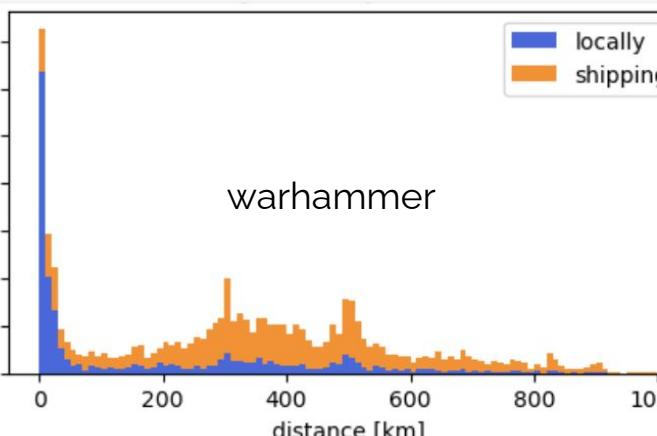
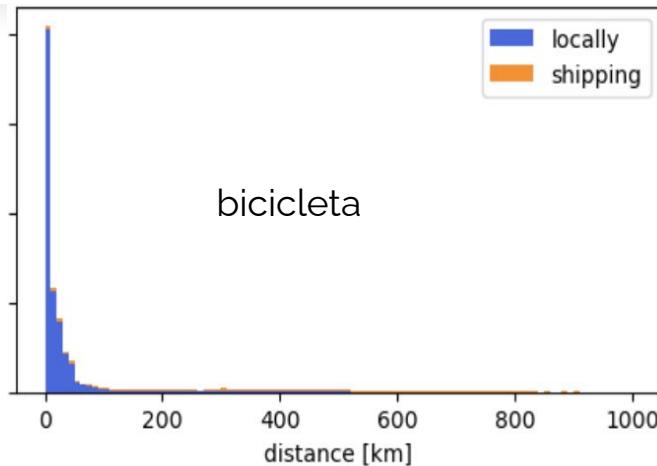
Easy to test

Aggressive option

**Define areas of sensitivity** to distance and adapt ranking for them

Sensible

Not sensible



# Distance sensitivity experiments: switching to local search

## Hypothesis

By changing to local search on a considerable amount of distance-sensitive queries we believe we would be able to have a higher impact on searches that will translate in more total sold items

## Variants description

A - Baseline: current behavior

B - Variant: distance sensitivity  
10k queries

## 10k queries impact

## Change

## General impact

## Change

**Primary Goal**

Total Sold Items

**+6,65 %****Primary Goal**

Total Sold Items

**+1,21 %****Secondary Goal**

Local Sold Items

**+8,80 %****Secondary Goal**

Local Sold Items

**+1,70 %****Health Metric**

TRX with Shipping

**-18,81 %****Health Metric**

TRX with Shipping

**0 %****Health Metric**

Chats

**+7,55 %****Health Metric**

Chats

**+0,87 %**

## 2nd iteration

Increase keyword coverage

Adapt keyword selection

## Variants description

A - Baseline: current behavior

B - Variant: distance sensitivity  
10k queries

C - Variant: distance sensitivity  
110k queries

		Change vs B	Change vs C
<b>Primary Goal</b>	Sold Items	+1,1%	+3,8%
<b>Secondary Goal</b>	Purchase Intention	+0,7%	+2,2%
<b>Health Metric</b>	Local Sold Items	+1,5%	+6,9%
<b>Health Metric</b>	TRX with Shipping	-0,3%	-3,2%

# Distance sensitivity experiments: adapting area size (without location dimension)

## Hypothesis

By fine-tuning results based on query and distance sensitivity area size we believe that users will find more meaningful results and we will increase total transactions

## Variants description

A - Baseline: relevance by default, we don't use any file

B - Variant 1: Distance sensitive queries, based on transaction data in normal space (km)

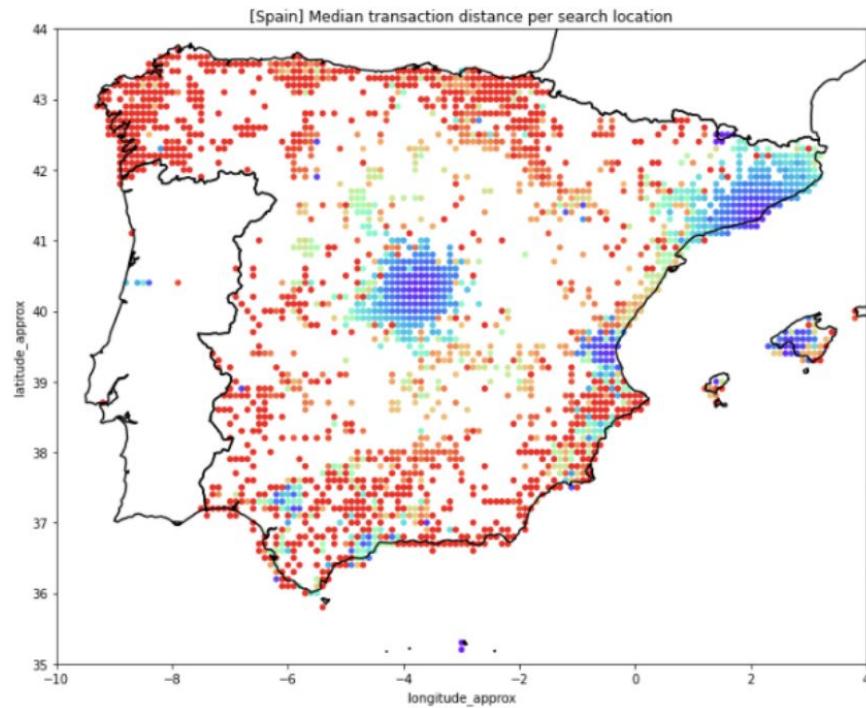
C - Variant 2 : Distance sensitive queries, based on transaction data in logarithm space (km)

D - Variant 3 : Distance sensitive queries, based on chat/ship data in normal space (km)

		Change V1	Change V2	Change V3
<b>Primary Goal</b>	Sold Items	+1.68%	+2.48%	+1.62%
<b>Secondary Goal</b>	Purchase Intention	+1.69%	+2.84%	+1.89%
<b>Health Metric</b>	Local Sold Items	+4.76%	+9.14%	+5.20%
<b>Health Metric</b>	TRX with Shipping	-1.83%	-4.97%	-2.51%

# Distance sensitivity experiments: adapting area size (adding location dimension)

## Understanding Distance: Moving from Local Search to Nationwide Search



## Hypothesis

By considering search query and search location we believe that users will find more meaningful results and we will increase total transactions without heavily penalising shipped transactions

## Variants description

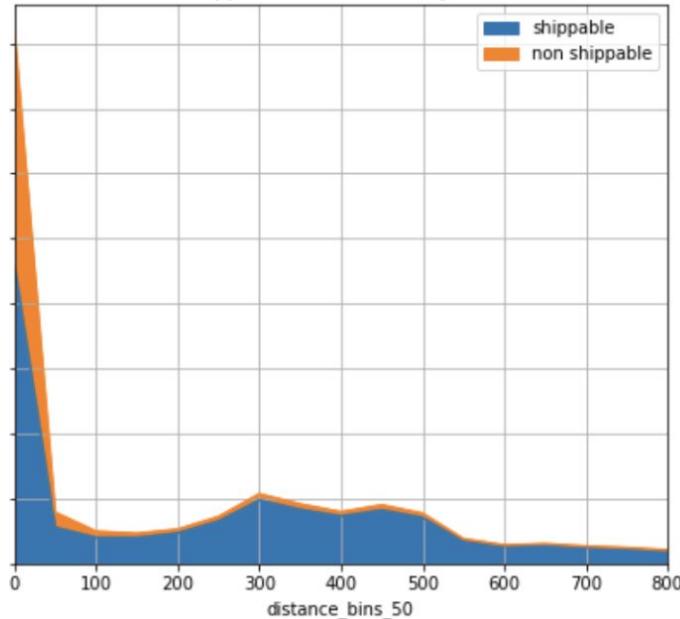
A - Baseline: Distance sensitive queries, based on transaction data in normal space (km)

B - Variant 1: Distance sensitive queries and location sensitivity score

C - Variant 2 : Distance sensitive queries (80 percentile radius) and location sensitivity score

		Change V1	Change V2
<b>Primary Goal</b>	Sold Items	-1,30%	-0,82%
<b>Secondary Goal</b>	Purchase Intention	-0,96%	-0,55%
<b>Health Metric</b>	Local Sold Items	-2,34%	-1,49%
<b>Health Metric</b>	TRX with Shipping	-012%	-0,04%

# Distance sensitivity experiments: item shippability



We see that above certain threshold most of transactions happen on shippable items...

... but with nationwide we show 'a lot' of non-shippable items over long distances

## Hypothesis

By considering item shippability we believe that users will find more meaningful results and we will increase total transactions increasing shipped transactions without damaging local transactions

## Variants description

A - Baseline: Distance sensitive queries, based on transaction data in normal space (km)

B - Variant 1: Penalize non-shippable items above median transaction kms

Change V1		
Primary Goal	Sold Items	+1,14%
Secondary Goal	Purchase Intention	-1,81%
Health Metric	Local Sold Items	-1,05%
Health Metric	TRX with Shipping	+3,51%

# Conclusions

### Know your domain

Which ranking factors are core to your users

Identify and iterate them

### Balancing user and business needs

What is good for the business might not be ideal for the user

### Reduce defects generated by your relevance algorithm

Impossible to address all the possible issues, corner cases

Understand at which level you should deal with your key factors

# Thank you!