seo.do

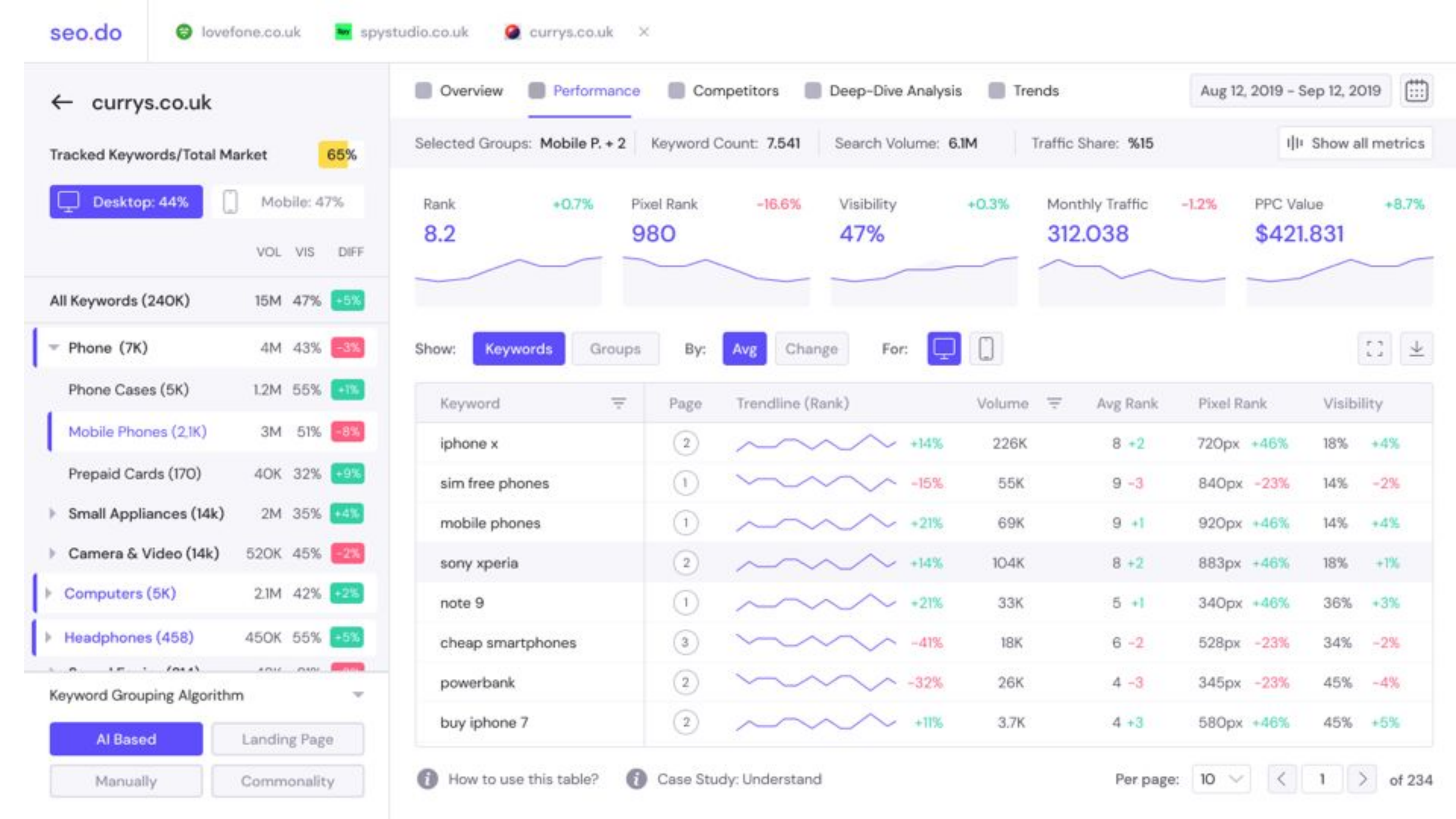
# Our Story w/ Clickhouse @seo.do

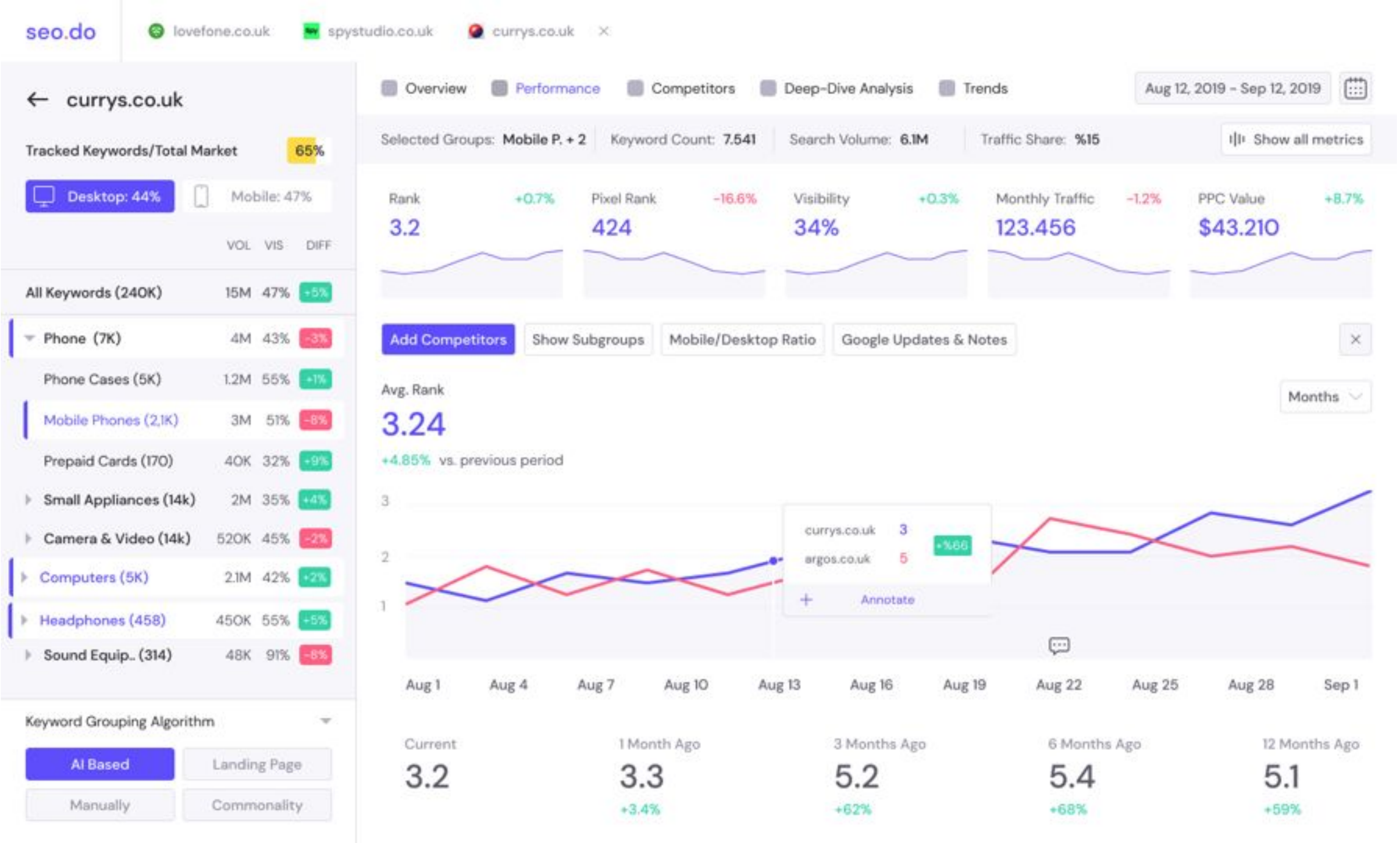
Metehan Çetinkaya

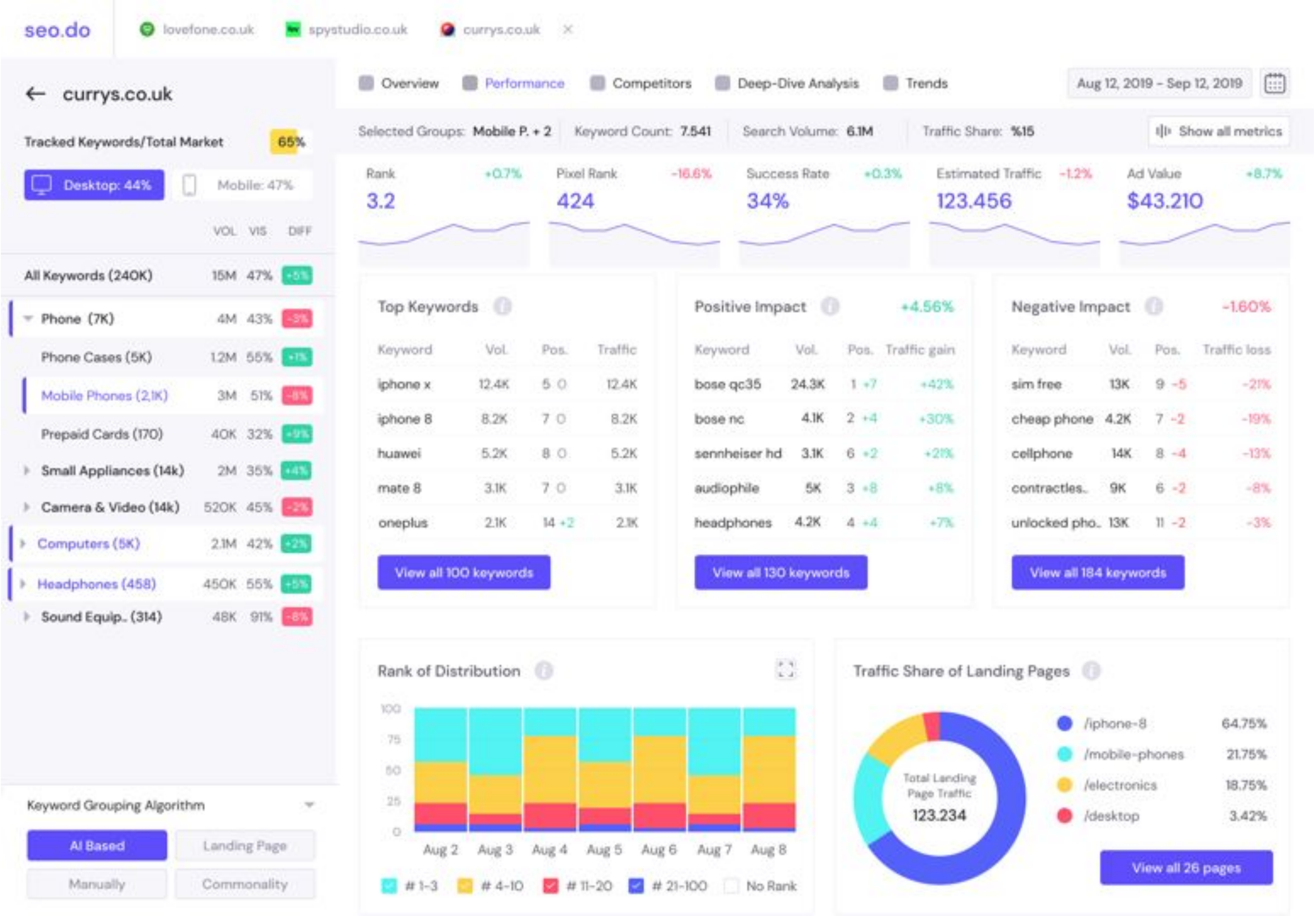
# What is seo.do?

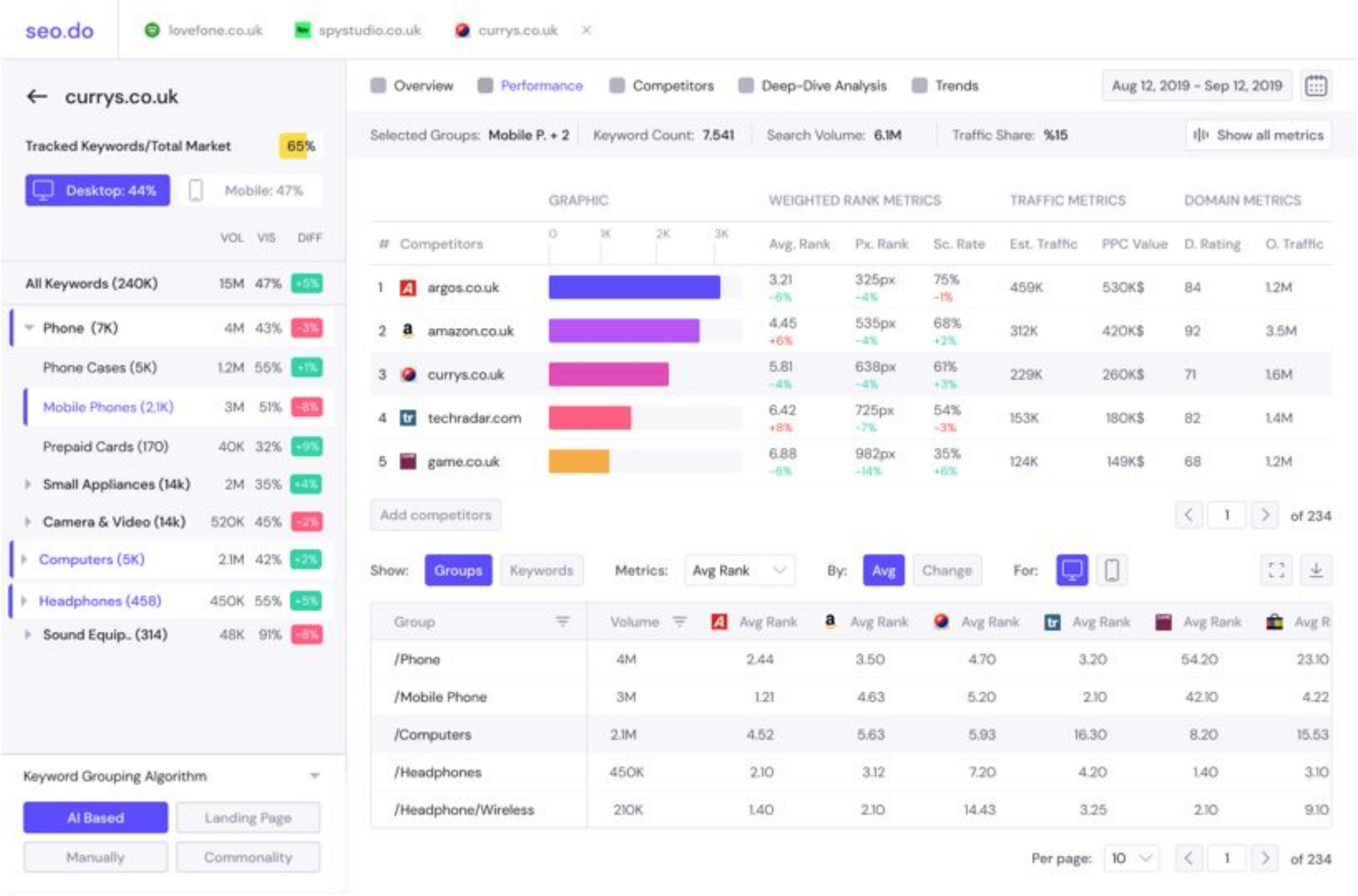
What is seo.do?

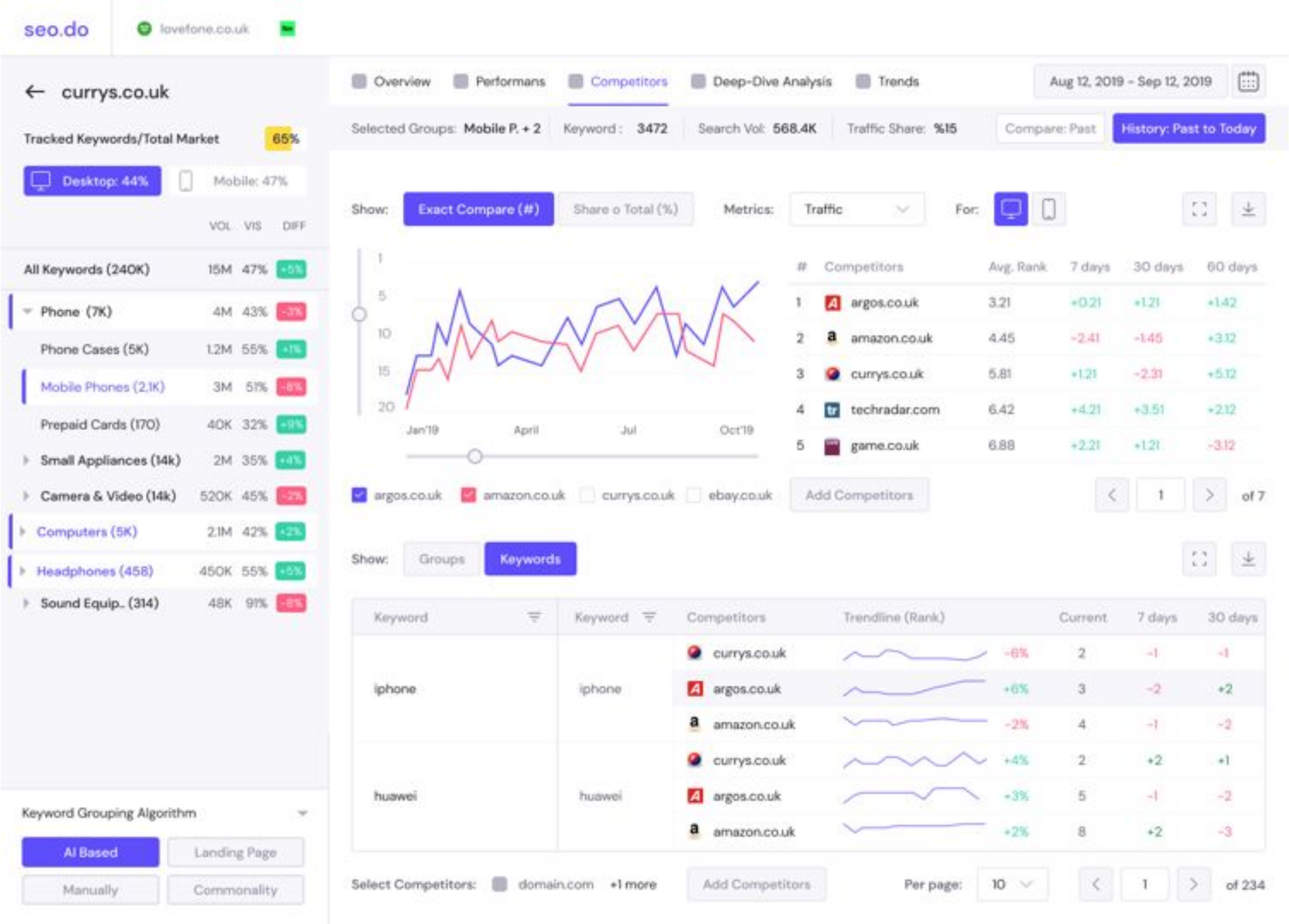
*Yandex Metrica for SEO professionals.*

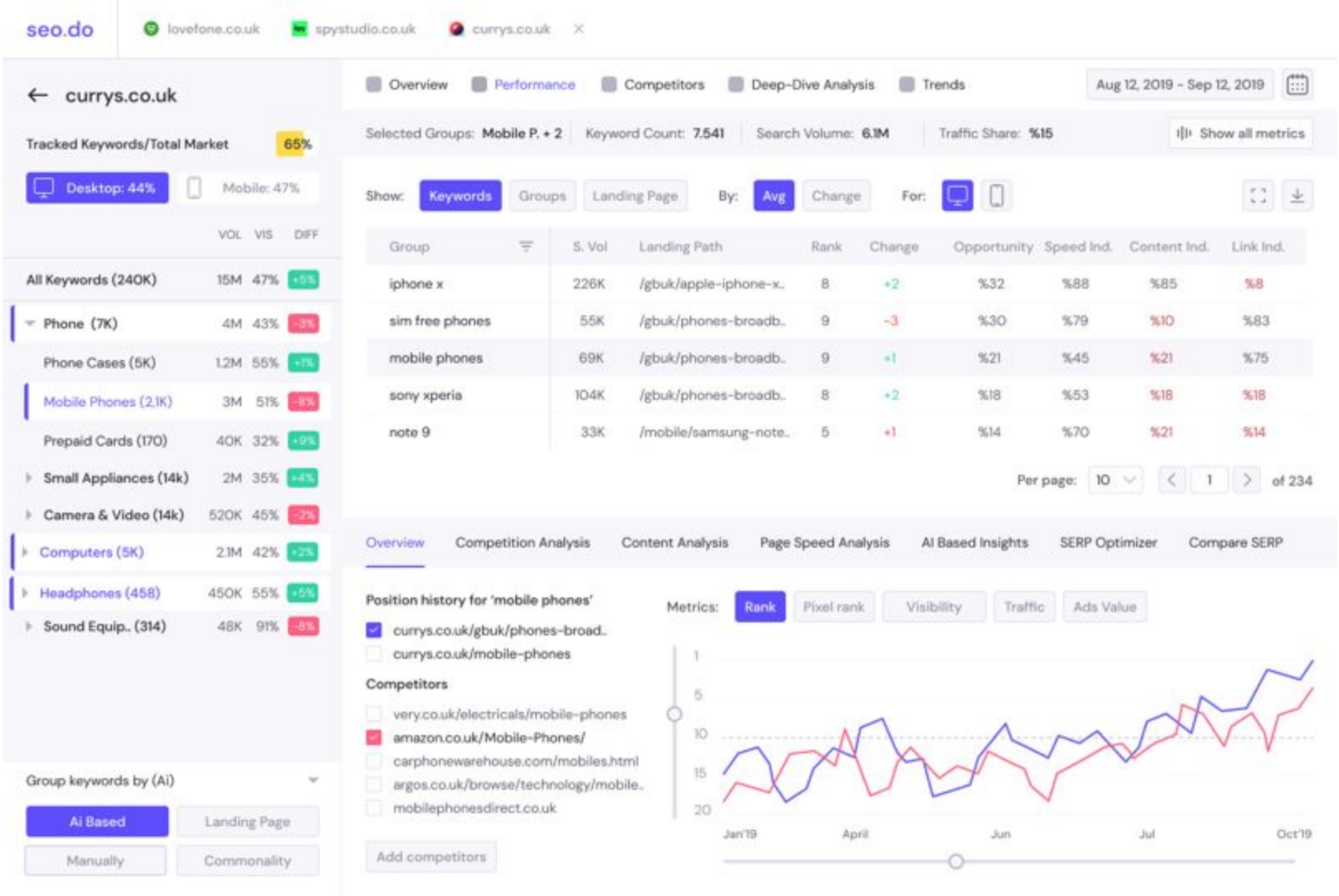

seo.do
lovefone.co.uk
spystudio.co.uk
currys.co.uk
currys.co.uk
Tracked Keywords/Total Market 65%
Desktop: 44%
Mobile: 47%
VOL VIS DIFF
All Keywords (240K) 15M 47% +5%
Phone (7K) 4M 43% -3%
Phone Cases (5K) 1.2M 55% +1%
Mobile Phones (2.1K) 3M 51% -8%
Prepaid Cards (170) 40K 32% +9%
Small Appliances (14k) 2M 35% +4%
Camera & Video (14k) 520K 45% -2%
Computers (5K) 2.1M 42% +2%
Headphones (458) 450K 55% +5%
Keyword Grouping Algorithm
AI Based
Landing Page
Manually
Commonality
Overview
Performance
Competitors
Deep-Dive Analysis
Trends
Aug 12, 2019 - Sep 12, 2019
Selected Groups: Mobile P. + 2
Keyword Count: 7.541
Search Volume: 6.1M
Traffic Share: %15
Show all metrics
Rank +0.7% 8.2
Pixel Rank -16.6% 980
Visibility +0.3% 47%
Monthly Traffic -1.2% 312.038
PPC Value +8.7% $421.831
Show: Keywords Groups
By: Avg Change
For:
Keyword Page Trendline (Rank) Volume Avg Rank Pixel Rank Visibility
iphone x 2 +14% 226K 8 +2 720px +46% 18% +4%
sim free phones 1 -15% 55K 9 -3 840px -23% 14% -2%
mobile phones 1 +21% 69K 9 +1 920px +46% 14% +4%
sony xperia 2 +14% 104K 8 +2 883px +46% 18% +1%
note 9 1 +21% 33K 5 +1 340px +46% 36% +3%
cheap smartphones 3 -41% 18K 6 -2 528px -23% 34% -2%
powerbank 2 -32% 26K 4 -3 345px -23% 45% -4%
buy iphone 7 2 +11% 3.7K 4 +3 580px +46% 45% +5%
How to use this table?
Case Study: Understand
Per page: 10
1
of 234

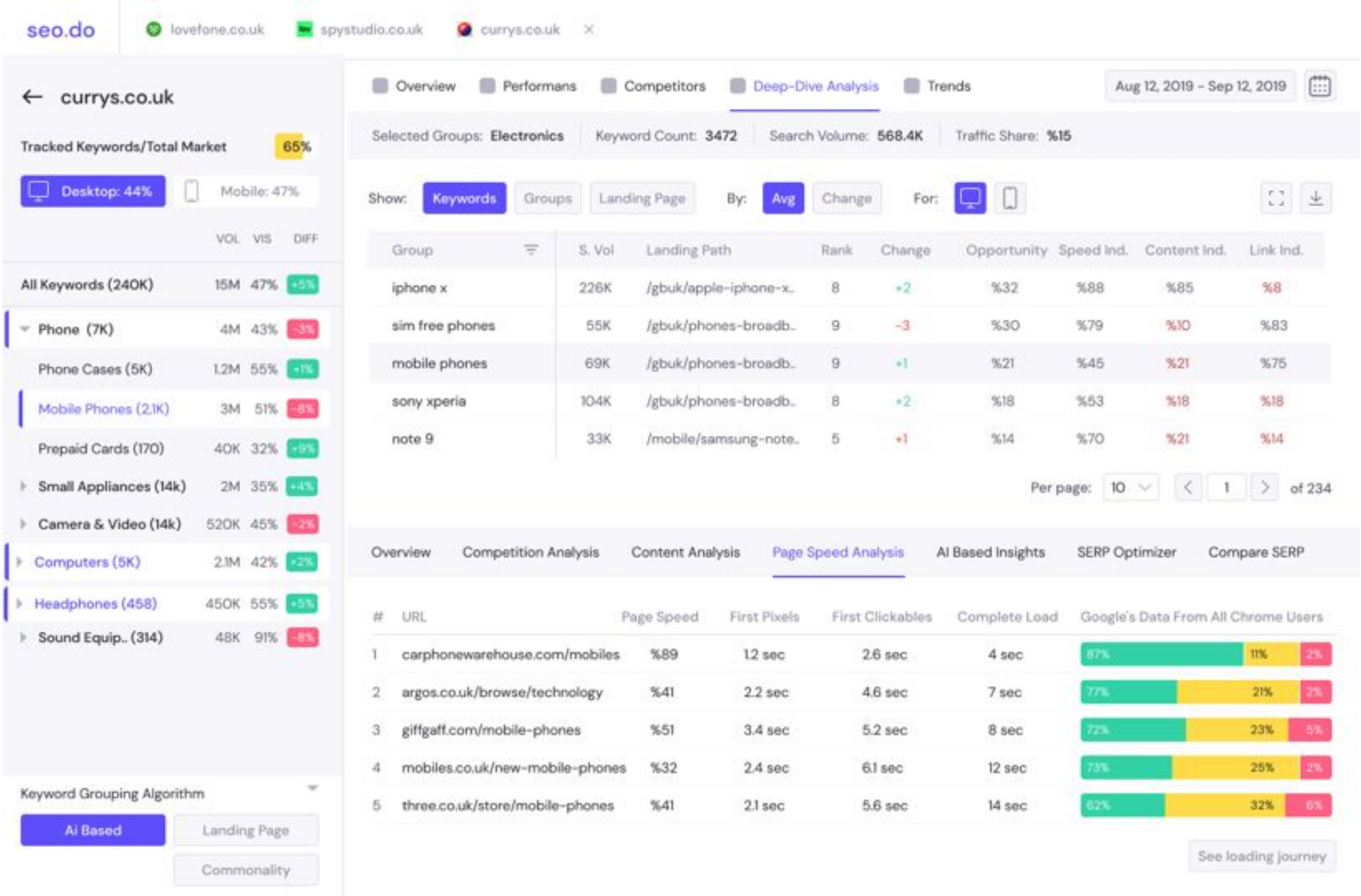

seo.do
lovefone.co.uk
spystudio.co.uk
currys.co.uk
currys.co.uk
Tracked Keywords/Total Market 65%
Desktop: 44%
Mobile: 47%
VOL VIS DIFF
All Keywords (240K) 15M 47% +5%
Phone (7K) 4M 43% -3%
Phone Cases (5K) 1.2M 55% +1%
Mobile Phones (2,1K) 3M 51% -8%
Prepaid Cards (170) 40K 32% +9%
Small Appliances (14k) 2M 35% +4%
Camera & Video (14k) 520K 45% -2%
Computers (5K) 2.1M 42% +2%
Headphones (458) 450K 55% +5%
Sound Equip.. (314) 48K 91% -8%
Keyword Grouping Algorithm
AI Based
Landing Page
Manually
Commonality
Overview
Performance
Competitors
Deep-Dive Analysis
Trends
Aug 12, 2019 - Sep 12, 2019
Selected Groups: Mobile P. + 2
Keyword Count: 7.541
Search Volume: 6.1M
Traffic Share: %15
Show all metrics
Rank +0.7% 3.2
Pixel Rank -16.6% 424
Visibility +0.3% 34%
Monthly Traffic -1.2% 123.456
PPC Value +8.7% $43.210
Add Competitors
Show Subgroups
Mobile/Desktop Ratio
Google Updates & Notes
Avg. Rank
3.24
+4.85% vs. previous period
Months
currys.co.uk 3
argos.co.uk 5
+%66
Annotate
3
2
1
Aug 1
Aug 4
Aug 7
Aug 10
Aug 13
Aug 16
Aug 19
Aug 22
Aug 25
Aug 28
Sep 1
Current 3.2
1 Month Ago 3.3 +3.4%
3 Months Ago 5.2 +62%
6 Months Ago 5.4 +68%
12 Months Ago 5.1 +59%

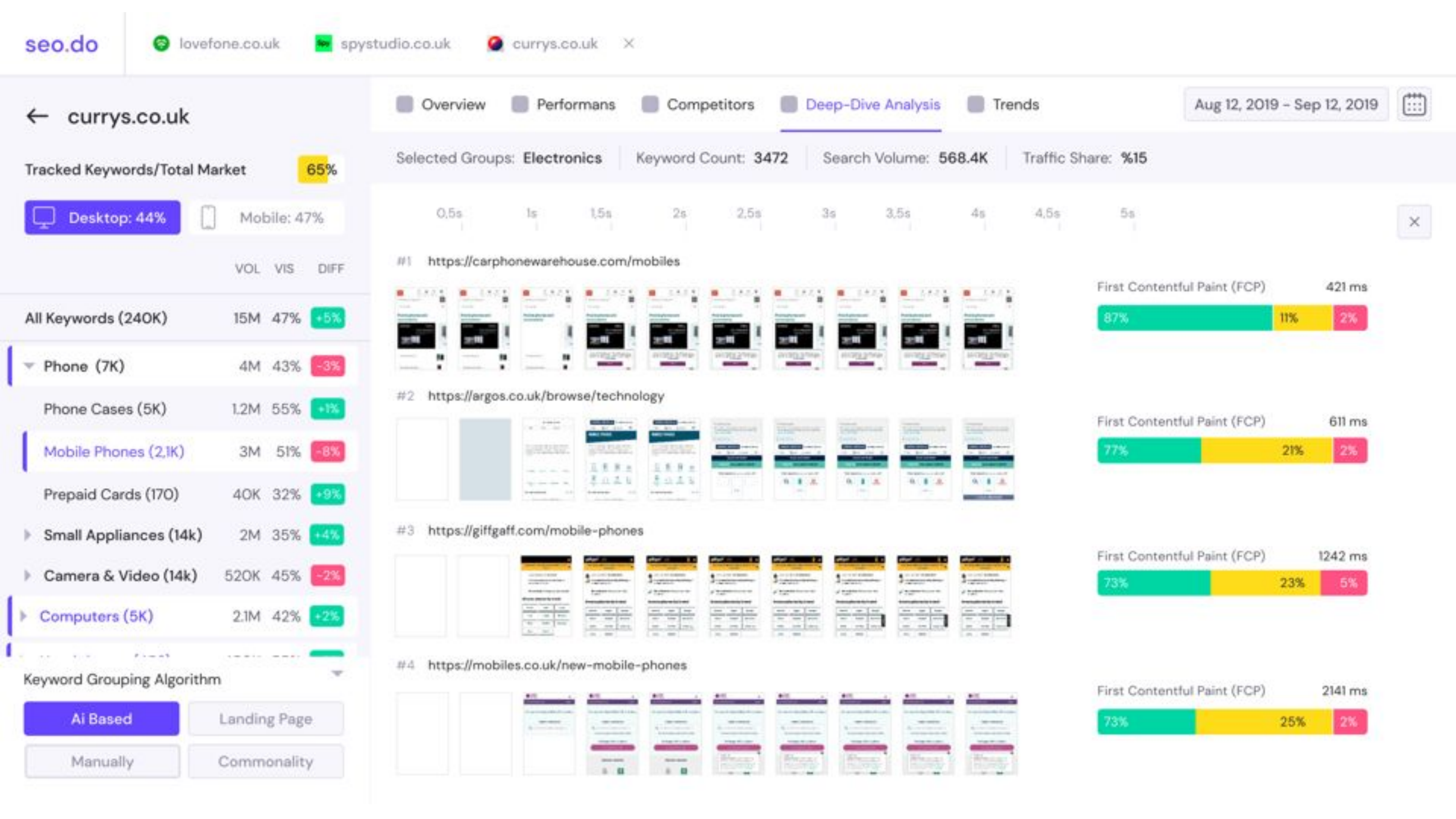

seo.do
lovefone.co.uk
spystudio.co.uk
currys.co.uk
currys.co.uk
Tracked Keywords/Total Market 65%
Desktop: 44%
Mobile: 47%
VOL VIS DIFF
All Keywords (240K) 15M 47% +5%
Phone (7K) 4M 43% -3%
Phone Cases (5K) 1.2M 55% +1%
Mobile Phones (2,1K) 3M 51% -8%
Prepaid Cards (170) 40K 32% +9%
Small Appliances (14k) 2M 35% +4%
Camera & Video (14k) 520K 45% -2%
Computers (5K) 2.1M 42% +2%
Headphones (458) 450K 55% +5%
Sound Equip.. (314) 48K 91% -8%
Keyword Grouping Algorithm
AI Based
Landing Page
Manually
Commonality
Overview
Performance
Competitors
Deep-Dive Analysis
Trends
Aug 12, 2019 - Sep 12, 2019
Selected Groups: Mobile P. + 2
Keyword Count: 7.541
Search Volume: 6.1M
Traffic Share: %15
Show all metrics
Rank +0.7% 3.2
Pixel Rank -16.6% 424
Success Rate +0.3% 34%
Estimated Traffic -1.2% 123.456
Ad Value +8.7% $43.210
Top Keywords
Keyword Vol. Pos. Traffic
iphone x 12.4K 5 0 12.4K
iphone 8 8.2K 7 0 8.2K
huawei 5.2K 8 0 5.2K
mate 8 3.1K 7 0 3.1K
oneplus 2.1K 14 +2 2.1K
View all 100 keywords
Positive Impact +4.56%
Keyword Vol. Pos. Traffic gain
bose qc35 24.3K 1 +7 +42%
bose nc 4.1K 2 +4 +30%
sennheiser hd 3.1K 6 +2 +21%
audiophile 5K 3 +8 +8%
headphones 4.2K 4 +4 +7%
View all 130 keywords
Negative Impact -1.60%
Keyword Vol. Pos. Traffic loss
sim free 13K 9 -5 -21%
cheap phone 4.2K 7 -2 -19%
cellphone 14K 8 -4 -13%
contractles.. 9K 6 -2 -8%
unlocked pho.. 13K 11 -2 -3%
View all 184 keywords
Rank of Distribution
100
75
50
25
0
Aug 2 Aug 3 Aug 4 Aug 5 Aug 6 Aug 7 Aug 8
# 1-3 # 4-10 # 11-20 # 21-100 No Rank
Traffic Share of Landing Pages
Total Landing Page Traffic
123.234
/iphone-8 64.75%
/mobile-phones 21.75%
/electronics 18.75%
/desktop 3.42%
View all 26 pages

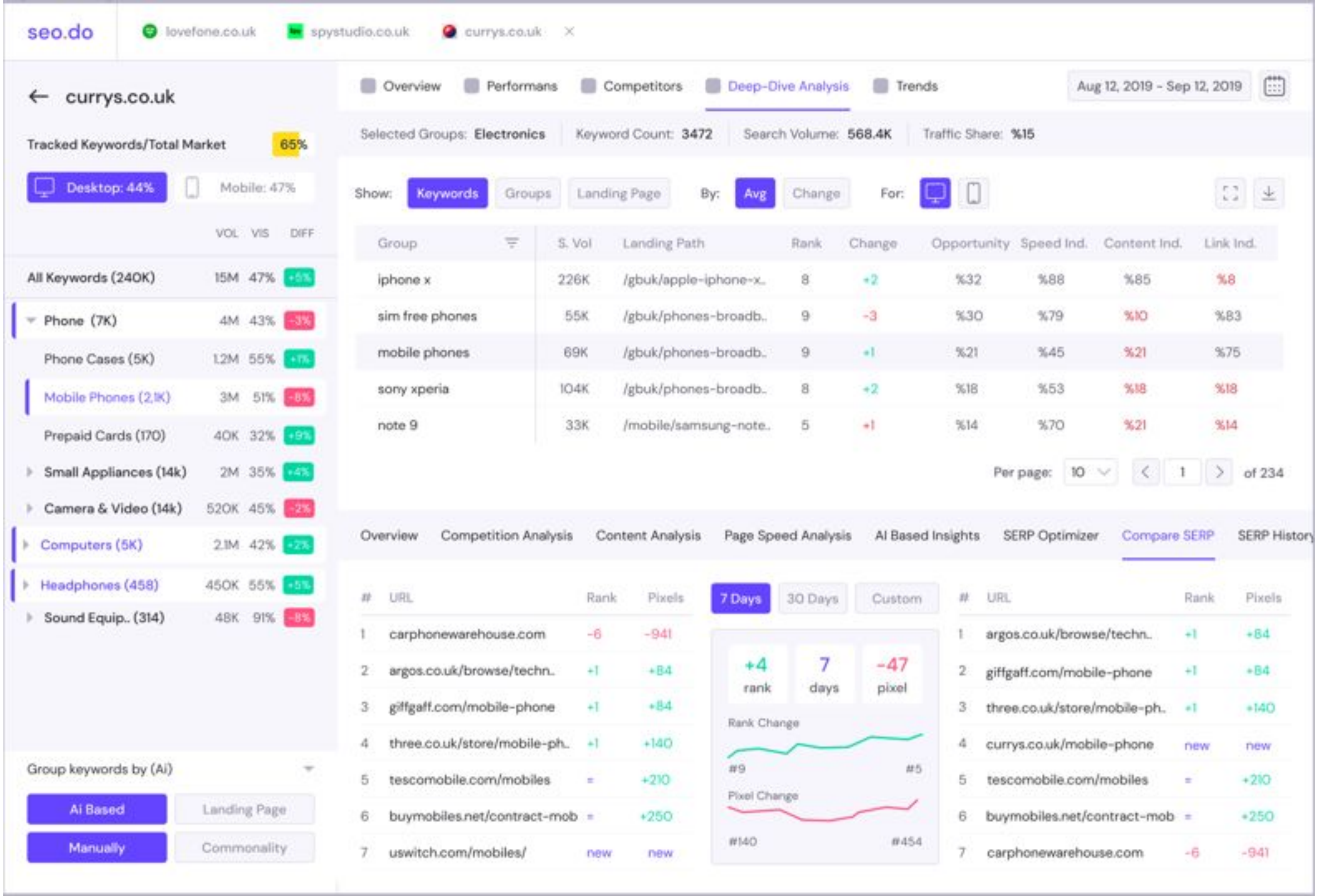

seo.do
lovefone.co.uk
spystudio.co.uk
currys.co.uk
currys.co.uk
Tracked Keywords/Total Market 65%
Desktop: 44%
Mobile: 47%
VOL VIS DIFF
All Keywords (240K) 15M 47% +5%
Phone (7K) 4M 43% -3%
Phone Cases (5K) 1.2M 55% +1%
Mobile Phones (2,1K) 3M 51% -8%
Prepaid Cards (170) 40K 32% +9%
Small Appliances (14k) 2M 35% +4%
Camera & Video (14k) 520K 45% -2%
Computers (5K) 2.1M 42% +2%
Headphones (458) 450K 55% +5%
Sound Equip.. (314) 48K 91% -8%
Keyword Grouping Algorithm
AI Based
Landing Page
Manually
Commonality
Overview
Performance
Competitors
Deep-Dive Analysis
Trends
Aug 12, 2019 - Sep 12, 2019
Selected Groups: Mobile P. + 2
Keyword Count: 7.541
Search Volume: 6.1M
Traffic Share: %15
Show all metrics
GRAPHIC
WEIGHTED RANK METRICS
TRAFFIC METRICS
DOMAIN METRICS
# Competitors 0 1K 2K 3K Avg. Rank Px. Rank Sc. Rate Est. Traffic PPC Value D. Rating O. Traffic
1 argos.co.uk 3.21 -6% 325px -4% 75% -1% 459K 530K$ 84 1.2M
2 amazon.co.uk 4.45 +6% 535px -4% 68% +2% 312K 420K$ 92 3.5M
3 currys.co.uk 5.81 -4% 638px -4% 61% +3% 229K 260K$ 71 1.6M
4 techradar.com 6.42 +8% 725px -7% 54% -3% 153K 180K$ 82 1.4M
5 game.co.uk 6.88 -6% 982px -14% 35% +6% 124K 149K$ 68 1.2M
Add competitors
1 of 234
Show: Groups Keywords
Metrics: Avg Rank
By: Avg Change
For:
Group Volume Avg Rank Avg Rank Avg Rank Avg Rank Avg Rank Avg R
/Phone 4M 2.44 3.50 4.70 3.20 54.20 23.10
/Mobile Phone 3M 1.21 4.63 5.20 2.10 42.10 4.22
/Computers 2.1M 4.52 5.63 5.93 16.30 8.20 15.53
/Headphones 450K 2.10 3.12 7.20 4.20 1.40 3.10
/Headphone/Wireless 210K 1.40 2.10 14.43 3.25 2.10 9.10
Per page: 10
1 of 234

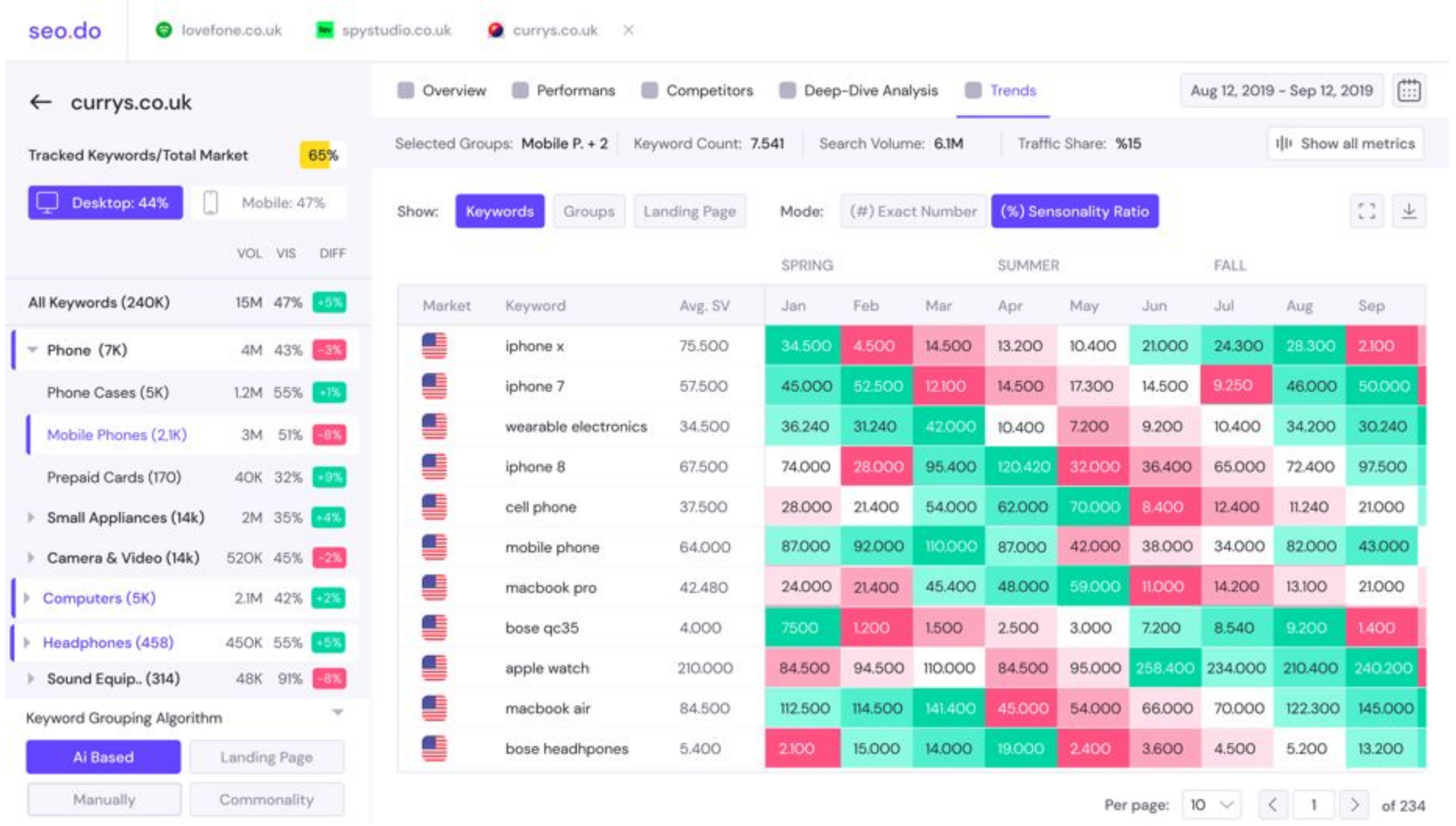

seo.do
lovefone.co.uk
currys.co.uk
Tracked Keywords/Total Market 65%
Desktop: 44%
Mobile: 47%
VOL VIS DIFF
All Keywords (240K) 15M 47% +5%
Phone (7K) 4M 43% -3%
Phone Cases (5K) 1.2M 55% +1%
Mobile Phones (2,1K) 3M 51% -8%
Prepaid Cards (170) 40K 32% +9%
Small Appliances (14k) 2M 35% +4%
Camera & Video (14k) 520K 45% -2%
Computers (5K) 2.1M 42% +2%
Headphones (458) 450K 55% +5%
Sound Equip.. (314) 48K 91% -8%
Keyword Grouping Algorithm
AI Based
Landing Page
Manually
Commonality
Overview
Performans
Competitors
Deep-Dive Analysis
Trends
Aug 12, 2019 - Sep 12, 2019
Selected Groups: Mobile P. + 2
Keyword : 3472
Search Vol: 568.4K
Traffic Share: %15
Compare: Past
History: Past to Today
Show:
Exact Compare (#)
Share o Total (%)
Metrics:
Traffic
For:
1
5
10
15
20
Jan'19
April
Jul
Oct'19
# Competitors Avg. Rank 7 days 30 days 60 days
1 argos.co.uk 3.21 +0.21 +1.21 +1.42
2 amazon.co.uk 4.45 -2.41 -1.45 +3.12
3 currys.co.uk 5.81 +1.21 -2.31 +5.12
4 techradar.com 6.42 +4.21 +3.51 +2.12
5 game.co.uk 6.88 +2.21 +1.21 -3.12
argos.co.uk
amazon.co.uk
currys.co.uk
ebay.co.uk
Add Competitors
1
of 7
Show:
Groups
Keywords
Keyword
Keyword
Competitors
Trendline (Rank)
Current
7 days
30 days
iphone
iphone
currys.co.uk -6% 2 -1 -1
argos.co.uk +6% 3 -2 +2
amazon.co.uk -2% 4 -1 -2
huawei
huawei
currys.co.uk +4% 2 +2 +1
argos.co.uk +3% 5 -1 -2
amazon.co.uk +2% 8 +2 -3
Select Competitors:
domain.com
+1 more
Add Competitors
Per page:
10
1
of 234


seo.do
lovefone.co.uk
spystudio.co.uk
currys.co.uk
currys.co.uk
Tracked Keywords/Total Market 65%
Desktop: 44%
Mobile: 47%
VOL VIS DIFF
All Keywords (240K) 15M 47% +5%
Phone (7K) 4M 43% -3%
Phone Cases (5K) 1.2M 55% +1%
Mobile Phones (2.1K) 3M 51% -8%
Prepaid Cards (170) 40K 32% +9%
Small Appliances (14k) 2M 35% +4%
Camera & Video (14k) 520K 45% -2%
Computers (5K) 2.1M 42% +2%
Headphones (458) 450K 55% +5%
Sound Equip.. (314) 48K 91% -8%
Group keywords by (Ai)
Ai Based
Landing Page
Manually
Commonality
Overview
Performance
Competitors
Deep-Dive Analysis
Trends
Aug 12, 2019 - Sep 12, 2019
Selected Groups: Mobile P. + 2
Keyword Count: 7.541
Search Volume: 6.1M
Traffic Share: %15
Show all metrics
Show: Keywords Groups Landing Page
By: Avg Change
For:
Group S. Vol Landing Path Rank Change Opportunity Speed Ind. Content Ind. Link Ind.
iphone x 226K /gbuk/apple-iphone-x.. 8 +2 %32 %88 %85 %8
sim free phones 55K /gbuk/phones-broadb.. 9 -3 %30 %79 %10 %83
mobile phones 69K /gbuk/phones-broadb.. 9 +1 %21 %45 %21 %75
sony xperia 104K /gbuk/phones-broadb.. 8 +2 %18 %53 %18 %18
note 9 33K /mobile/samsung-note.. 5 +1 %14 %70 %21 %14
Per page: 10
1
of 234
Overview
Competition Analysis
Content Analysis
Page Speed Analysis
AI Based Insights
SERP Optimizer
Compare SERP
Position history for 'mobile phones'
currys.co.uk/gbuk/phones-broad..
currys.co.uk/mobile-phones
Competitors
very.co.uk/electricals/mobile-phones
amazon.co.uk/Mobile-Phones/
carphonewarehouse.com/mobiles.html
argos.co.uk/browse/technology/mobile..
mobilephonesdirect.co.uk
Add competitors
Metrics: Rank Pixel rank Visibility Traffic Ads Value
1
5
10
15
20
Jan'19
April
Jun
Jul
Oct'19


seo.do
lovefone.co.uk
spystudio.co.uk
currys.co.uk
currys.co.uk
Tracked Keywords/Total Market 65%
Desktop: 44%
Mobile: 47%
VOL VIS DIFF
All Keywords (240K) 15M 47% +5%
Phone (7K) 4M 43% -3%
Phone Cases (5K) 1.2M 55% +1%
Mobile Phones (2,1K) 3M 51% -8%
Prepaid Cards (170) 40K 32% +9%
Small Appliances (14k) 2M 35% +4%
Camera & Video (14k) 520K 45% -2%
Computers (5K) 2.1M 42% +2%
Headphones (458) 450K 55% +5%
Sound Equip.. (314) 48K 91% -8%
Keyword Grouping Algorithm
Ai Based
Landing Page
Commonality
Overview
Performans
Competitors
Deep-Dive Analysis
Trends
Aug 12, 2019 - Sep 12, 2019
Selected Groups: Electronics
Keyword Count: 3472
Search Volume: 568.4K
Traffic Share: %15
Show: Keywords Groups Landing Page
By: Avg Change
For:
Group S. Vol Landing Path Rank Change Opportunity Speed Ind. Content Ind. Link Ind.
iphone x 226K /gbuk/apple-iphone-x.. 8 +2 %32 %88 %85 %8
sim free phones 55K /gbuk/phones-broadb.. 9 -3 %30 %79 %10 %83
mobile phones 69K /gbuk/phones-broadb.. 9 +1 %21 %45 %21 %75
sony xperia 104K /gbuk/phones-broadb.. 8 +2 %18 %53 %18 %18
note 9 33K /mobile/samsung-note.. 5 +1 %14 %70 %21 %14
Per page: 10
1
of 234
Overview
Competition Analysis
Content Analysis
Page Speed Analysis
AI Based Insights
SERP Optimizer
Compare SERP
# URL Page Speed First Pixels First Clickables Complete Load Google's Data From All Chrome Users
1 carphonewarehouse.com/mobiles %89 1.2 sec 2.6 sec 4 sec 87% 11% 2%
2 argos.co.uk/browse/technology %41 2.2 sec 4.6 sec 7 sec 77% 21% 2%
3 giffgaff.com/mobile-phones %51 3.4 sec 5.2 sec 8 sec 72% 23% 5%
4 mobiles.co.uk/new-mobile-phones %32 2.4 sec 6.1 sec 12 sec 73% 25% 2%
5 three.co.uk/store/mobile-phones %41 2.1 sec 5.6 sec 14 sec 62% 32% 6%
See loading journey


seo.do
lovefone.co.uk
spystudio.co.uk
currys.co.uk
currys.co.uk
Tracked Keywords/Total Market 65%
Desktop: 44%
Mobile: 47%
VOL VIS DIFF
All Keywords (240K) 15M 47% +5%
Phone (7K) 4M 43% -3%
Phone Cases (5K) 1.2M 55% +1%
Mobile Phones (2,1K) 3M 51% -8%
Prepaid Cards (170) 40K 32% +9%
Small Appliances (14k) 2M 35% +4%
Camera & Video (14k) 520K 45% -2%
Computers (5K) 2.1M 42% +2%
Keyword Grouping Algorithm
Ai Based
Landing Page
Manually
Commonality
Overview
Performans
Competitors
Deep-Dive Analysis
Trends
Aug 12, 2019 - Sep 12, 2019
Selected Groups: Electronics
Keyword Count: 3472
Search Volume: 568.4K
Traffic Share: %15
0,5s 1s 1,5s 2s 2,5s 3s 3,5s 4s 4,5s 5s
#1 https://carphonewarehouse.com/mobiles
First Contentful Paint (FCP) 421 ms
87% 11% 2%
#2 https://argos.co.uk/browse/technology
First Contentful Paint (FCP) 611 ms
77% 21% 2%
#3 https://giffgaff.com/mobile-phones
First Contentful Paint (FCP) 1242 ms
73% 23% 5%
#4 https://mobiles.co.uk/new-mobile-phones
First Contentful Paint (FCP) 2141 ms
73% 25% 2%


seo.do
lovefone.co.uk
spystudio.co.uk
currys.co.uk
currys.co.uk
Tracked Keywords/Total Market 65%
Desktop: 44%
Mobile: 47%
VOL VIS DIFF
All Keywords (240K) 15M 47% +5%
Phone (7K) 4M 43% -3%
Phone Cases (5K) 1.2M 55% +1%
Mobile Phones (2,1K) 3M 51% -8%
Prepaid Cards (170) 40K 32% +9%
Small Appliances (14k) 2M 35% +4%
Camera & Video (14k) 520K 45% -2%
Computers (5K) 2.1M 42% +2%
Headphones (458) 450K 55% +5%
Sound Equip.. (314) 48K 91% -8%
Group keywords by (Ai)
Ai Based
Landing Page
Manually
Commonality
Overview
Performans
Competitors
Deep-Dive Analysis
Trends
Aug 12, 2019 - Sep 12, 2019
Selected Groups: Electronics
Keyword Count: 3472
Search Volume: 568.4K
Traffic Share: %15
Show: Keywords Groups Landing Page
By: Avg Change
For:
Group S. Vol Landing Path Rank Change Opportunity Speed Ind. Content Ind. Link Ind.
iphone x 226K /gbuk/apple-iphone-x.. 8 +2 %32 %88 %85 %8
sim free phones 55K /gbuk/phones-broadb.. 9 -3 %30 %79 %10 %83
mobile phones 69K /gbuk/phones-broadb.. 9 +1 %21 %45 %21 %75
sony xperia 104K /gbuk/phones-broadb.. 8 +2 %18 %53 %18 %18
note 9 33K /mobile/samsung-note.. 5 +1 %14 %70 %21 %14
Per page: 10
1
of 234
Overview
Competition Analysis
Content Analysis
Page Speed Analysis
AI Based Insights
SERP Optimizer
Compare SERP
SERP History
# URL Rank Pixels
1 carphonewarehouse.com -6 -941
2 argos.co.uk/browse/techn.. +1 +84
3 giffgaff.com/mobile-phone +1 +84
4 three.co.uk/store/mobile-ph.. +1 +140
5 tescomobile.com/mobiles = +210
6 buymobiles.net/contract-mob = +250
7 uswitch.com/mobiles/ new new
7 Days
30 Days
Custom
+4 rank
7 days
-47 pixel
Rank Change
#9
#5
Pixel Change
#140
#454
# URL Rank Pixels
1 argos.co.uk/browse/techn.. +1 +84
2 giffgaff.com/mobile-phone +1 +84
3 three.co.uk/store/mobile-ph.. +1 +140
4 currys.co.uk/mobile-phone new new
5 tescomobile.com/mobiles = +210
6 buymobiles.net/contract-mob = +250
7 carphonewarehouse.com -6 -941




| Market | Keyword | Avg. SV | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | iphone x | 75.500 | 34.500 | 4.500 | 14.500 | 13.200 | 10.400 | 21.000 | 24.300 | 28.300 | 2.100 |
| | iphone 7 | 57.500 | 45.000 | 52.500 | 12.100 | 14.500 | 17.300 | 14.500 | 9.250 | 46.000 | 50.000 |
| | wearable electronics | 34.500 | 36.240 | 31.240 | 42.000 | 10.400 | 7.200 | 9.200 | 10.400 | 34.200 | 30.240 |
| | iphone 8 | 67.500 | 74.000 | 28.000 | 95.400 | 120.420 | 32.000 | 36.400 | 65.000 | 72.400 | 97.500 |
| | cell phone | 37.500 | 28.000 | 21.400 | 54.000 | 62.000 | 70.000 | 8.400 | 12.400 | 11.240 | 21.000 |
| | mobile phone | 64.000 | 87.000 | 92.000 | 110.000 | 87.000 | 42.000 | 38.000 | 34.000 | 82.000 | 43.000 |
| | macbook pro | 42.480 | 24.000 | 21.400 | 45.400 | 48.000 | 59.000 | 11.000 | 14.200 | 13.100 | 21.000 |
| | bose qc35 | 4.000 | 7500 | 1.200 | 1.500 | 2.500 | 3.000 | 7.200 | 8.540 | 9.200 | 1.400 |
| | apple watch | 210.000 | 84.500 | 94.500 | 110.000 | 84.500 | 95.000 | 258.400 | 234.000 | 210.400 | 240.200 |
| | macbook air | 84.500 | 112.500 | 114.500 | 141.400 | 45.000 | 54.000 | 66.000 | 70.000 | 122.300 | 145.000 |
| | bose headhpones | 5.400 | 2.100 | 15.000 | 14.000 | 19.000 | 2.400 | 3.600 | 4.500 | 5.200 | 13.200 |

# 50.000 keywords

# 50.000 keywords

$$50.000 \times 2 \times 100 \times 365$$

50.000 keywords

# 50.000 x 2 x 100 x 365

3.6 billion rows per client

# How did our use case comply with ClickHouse?

- Continuous data insertion
  - Since there is no lock implementation in ClickHouse, inserting data to the database does not affect query performance
- No updates / deletes necessary
  - We don't have a use case for UPDATE & DELETE operations, to delete obsolete data we simply use PARTITION and it's operations
    - You can still update the data with using ALTER mutations
- Billions of rows to be processed
  - Time series data which will be aggregated constantly to create reports etc.
- SQL dialect close to Standard SQL

seo.do

# How did our use case comply with ClickHouse?

- Very handy built-in functions
  - avgIf, sumIf, countIf, order by if, date conversion functions (toWeek, toMonth etc.), splitByChar
    - *‘ SELECT fullName, age,  avgIf(wage, age BETWEEN  18 and 25) as youthWageAvg from people where country_id = 745261’*
- High compression ratio
  - Compression ratio totally depends on the data and since our data is timeseries and there are repeated strings, compression ratio for us is around 1:15
- Continuous development and great community

seo.do

# How did we design our table schemas?

## Understanding ClickHouse Index Structure

- Index structure is not similar to traditional RDBMS's index structure, no B+Tree, does not create unique constraint
- Data is physically sorted on disk
  - There is a background job to sort and merge the data and it will take place eventually
- Need to choose primary key / sorting key by considering query conditions
  - To keep the reads at minimum, need to consider all possible queries with conditions and then choose the primary key

# How did we design our table schemas?

## What Did We Do?

- Created initial table schemas on single server by considering our data structure and possible queries
- Inserted billions of dummy data which completely simulated our actual case
  - Since the data we will have is time series, we created the actual data for one day, and replicate it for 180 days
- Wrote sample queries, few queries for each dimension we would query the data

seo.do

# Refactoring The Schemas & Writing Queries

- Ran single query each time and checked the results
    - Tip : Use ' tail -f /var/log/clickhouse-server/clickhouse-server.log ' to see execution logs of the query (Peak MEM Usage, Threads executed, Marks read) that was executed. Or simply add 'send_logs_level=trace' when connecting to the cli ' clickhouse-client --send_logs_level=trace '
- Refactored the table schemas or even created new table schemas with considering query results
- After satisfying results, we executed bunch of load tests with the sample queries

# Using JOIN in ClickHouse

- Since ClickHouse's compression ratio is very high, compromised from storage to boost the performance
    - How so? Did not force the tables to be atomic, to keep the relations & JOINs at minimum, let there be duplicated data. It's a trade off.
- Tried to avoid JOINs however with our data structure it was not possible, so we kept JOINs at minimum
- Avoided using raw JOINs all the time
- Used JOINs with subqueries

# Raw JOIN

```
select keyword, group_id from keyword_data as ss

INNER JOIN

keyword_info as kw on ss.keyword = kw.keyword

and ss.cid = kw.cid

and ss.cid = 149315

and ss.position = 5

and toDate(ss.timestamp) = toDate('2019-10-20')
```

# Raw JOIN

- Took 5 seconds to complete
  - Slow and keeping system resources busy for long (This is important because ClickHouse can fully utilize system resources under load, average query response time will be high)
- Processed 369.74 million rows
- Executed the query with 8 threads
  - Executed the query on a server which had 16 cores. By default ClickHouse sets max_threads setting to half the number of physical core count, so this query utilized all the cores available to ClickHouse
- Peak Memory Usage : 90 Megabytes

# JOIN With Sub Query

select keyword, group_id from

**(select keyword from keyword_data PREWHERE position = 5 where cid = 149315**

**and toDate(timestamp) = '2019-10-20') as ss**

INNER JOIN

**(select keyword, group_id from keyword_info where cid = 149315) as kw**

on ss.keyword = kw.keyword

# JOIN With Sub Query

- Took 100 milliseconds
  - 50 times faster than raw join and releases resources quickly
- Processed 65.54 thousand rows
- Executed the query with  2 threads
  - Was able to execute the query with 6 less threads. It is very important for us to keep thread count at minimum for each query since our QPS rate will be high. If we were to ignore this, we would have to solve our performance problems with new replicas in the future which means new servers and constant money spend
- Peak Memory Usage : 5 Megabytes

# Creating ClickHouse Cluster

- Replication
    - High Availability
    - Load Scaling
- Sharding
    - Data size
    - Split data into smaller parts

# Creating ClickHouse Cluster

- Created a cluster with 6 servers (2 shards and 3 replicas)
- Set up ZooKeeper with 3 additional servers
  - Since latency is a critical point for ZooKeeper and ClickHouse can utilize all available system resources we don’t run ZooKeeper on the same server with ClickHouse (ZooKeeper Cluster with 3 servers can handle failure of 1 server)
- Used Clickhouse's internal load balancer mechanism to distribute queries over the replicas
  - HAProxy or CHProxy could be used as separate load balancer

# Chaos Testing

- Cluster is set up & running
- Developed a dummy API which will execute random queries with Golang Gin
    - This is not a performance test so we just went with the fastest way for us
- Created a basic load test which will make requests to our Go API continuously, with this way we will be able to monitor ClickHouse behavior with failures

seo.do

# Without Data Loss

- Initiated the load test
- Killed the ClickHouse in a random node during the load test
  - We are actually trying to simulate ClickHouse failure & temporary server crash in this scenario
- Monitored ClickHouse's load balancer's behavior and system's load
- Restarted the ClickHouse in that server
  - Load test is still being executed in the meantime

# What Happened?

- Since the ClickHouse Server in chosen shard was unreachable, ClickHouse load balancer stopped making requests to that node
- Queries started to be distributed over the remaining 2 replicas
- After restarting the Clickhouse Server, chosen node did not get any requests for another 10 minutes then it started to receive requests
    - Because ClickHouse load balancer distributes queries with considering error counts
    - Error count is halved each minute
    - Maximum error count is 1000 by default

# With Data Loss

- Chose an another shard randomly
- This time we didn’t just kill the ClickHouse Server, we formatted disks as well.
- CH configuration lost, all the data is lost
- Chosen node is still in ZooKeeper config
- Reconfigured the server and Reinstalled the CH
- Copied metadata from a running replica
- Monitored ClickHouse behavior and system's load

# What Happened?

- Like in the previous case, chosen node did not get any requests
- After configuration of the server is completed, replication took place
- Like in the other scenario, chosen node started getting requests after 10 minutes

# Deciding the API Framework

- Developed API endpoints which execute sample queries with Flask, Golang Gin and FastAPI
- Initiated the load tests for three of them separately (10K requests per minute, ran for 20 minutes)
- Monitored the results

# Deciding the API Framework

- Flask
  - Was not able to handle all the requests so after some time errors started to raise
  - Average response time was 3 seconds and %10 of the incoming requests resulted with errors
- Gin
  - Was able to handle all the requests without errors
  - Average response time was 350 milliseconds and there was no error at all
- FastAPI
  - Was able to handle all the requests without errors as well
  - Average response time was 300 milliseconds without errors