# Digital Marketing Firm
# PA - Case Study

**Analysis brief**

Your CPO wants to understand the drivers of user retention on Digital Marketing Firm. You need to investigate several hypotheses and present your diagnosis in 10 to 15 slides: where should Digital Marketing Firm invest or what should Digital Marketing Firm fix in order to improve user retention?

Dataset : SF ID, region, tier, vertical, revenue, legacy Digital Marketing Firm

**Questions**

1. Is there any specific data cleansing action(s) you would implement before starting the analysis ?
2. Is user retention steady over time?
3. How would you formulate your first hypotheses and how to confirm/infirm them ?
4. Based on your findings, create a deck of 10-15 slides maximum for your CPO & product managers. Pay attention to the story telling.
5. According to you, are there any specific missing attributes that are worth monitoring?
6. What  actions would you recommend after doing the analysis?

# User retention

What is user retention ?

User retention or cohort retention is a key metric to measure the growth of SaaS (Software as a Service) and digital products.

It can depend on global trends, product type, client, geography and metadata

# Summary

1. Available Data
2 Data Cleansing
2. Analysis
3. Propositions

# Available Data

Tables:

- definition: stocks all the definitions for the fields
- retention_overall: overall retention figures and total users per month
- retention_by_product: depending on the type of product
- user_retention_by_account: corresponding to specific clients
- geography + service: geography of the firms and service type
- account_metadata: type of the clients
- Digital Marketing Firm_exportfolio: client of Digital Marketing Firm(ex) or Digital Marketing Firm

# Available Data

Fields in the different tables

| | Field | Definition |
|---|---|---|
| 0 | product | e product on which the user connected. Please note if the same user connects on several module in the same month. We count him/her several times. Example: if user X connected on CS Digital on March 8th and CS Apps on March 12th, then we have count this user as 1 active user for CS Digital and 1 active user for CS Apps |
| 1 | first_connection | First connection month |
| 2 | month | Month |
| 3 | relative_month | Rank of the month compared to first_connection |
| 4 | total_users_per_month | Total users who connected each "month" |
| 5 | new_users_onboarded_on_first_connection_month | Total number of new users onboarded on "first_connection" month |
| 6 | account_id | Identifier of the client for Contentsquare (a client = for example |
| 7 | owner_entity | Office that manages the account |
| 8 | service_category | % of total contract value dedicated to professional service. Professional service is a team that delivers hands-on website analysis for our clients (example: "a product page analysis"), with data extracted fro |
| 9 | industry high level | Industry of the account |
| 10 | owner_entity | ffice that manages the account (geography) |
| 11 | ARR_total | Annual Recurring Revenue: price paid by the account each year to have access oftware |
| 12 | High level revenue tier | Company category of the account, depending on its turnover |
| 13 | Employee tier | Company category of the account, depending on its number of employees |
| 14 | is_legacy_clicktale_account | Boolean that shows if the account was originally a client signed |

# Data Cleansing

Total at first connection

retention_overall

| | first_connection | month | relative_month | total_users_per_month | new_users_onboarded_on_first_connection_month |
|---|---|---|---|---|---|
| **310** | 2021-02-01 00:00:00 | 2021-11-01 | 9.0 | 532.0 | 2006.0 |
| **311** | 2021-02-01 00:00:00 | 2021-12-01 | 10.0 | 436.0 | 2006.0 |
| **312** | 2021-02-01 00:00:00 | 2022-01-01 | 11.0 | 438.0 | 2006.0 |
| **313** | 2021-02-01 00:00:00 | 2022-02-01 | 12.0 | 418.0 | 2006.0 |
| **314** | 2021-02-01 00:00:00 | 2022-03-01 | 13.0 | 400.0 | 2006.0 |
| **315** | CS Digital | 2019-10-01 | 2019-10-01 00:00:00 | 0.0 | 806.0 |
| **316** | CS Digital | 2019-10-01 | 2019-11-01 00:00:00 | 2.0 | 492.0 |
| **317** | CS Digital | 2019-10-01 | 2019-12-01 00:00:00 | 4.0 | 412.0 |
| **318** | CS Digital | 2019-10-01 | 2020-01-01 00:00:00 | 6.0 | 362.0 |

# Data Cleansing

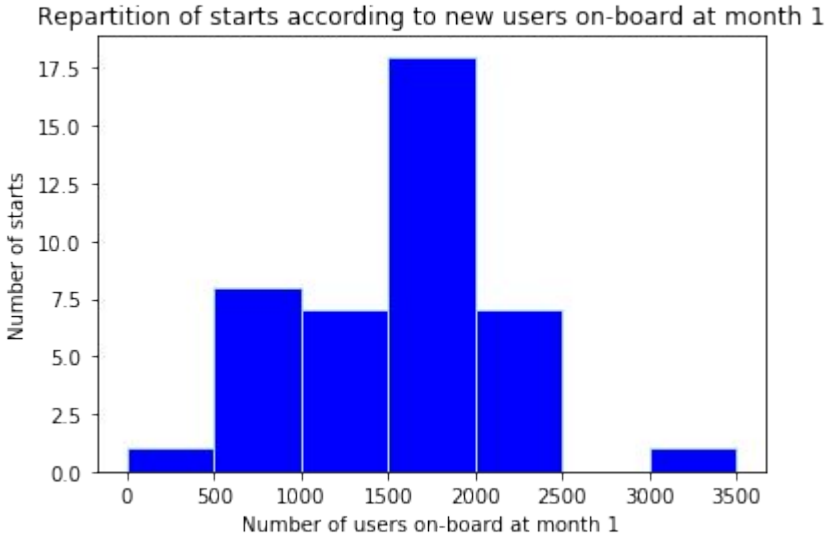
retention_overall

- Change of type timestamp to datetime
- Shift of data and addition of a column from 315
- Division by 2 relative_month
- This table and the next are not connected directly to clients ids

- geography and account_metadata and user_retention_by_account don't have the same account_id: it looks like the three last letters of the id have to be removed in order to find correlation between the first and the two next tables
- joining account_metada, geography and user_retention_by_account removes some accounts about 400 over 1300. We'll use the cohort present in all the tables for country, service category and ARR analysis

# Analysis

- Repartition of number of new users on-boarded all first_connections mingled

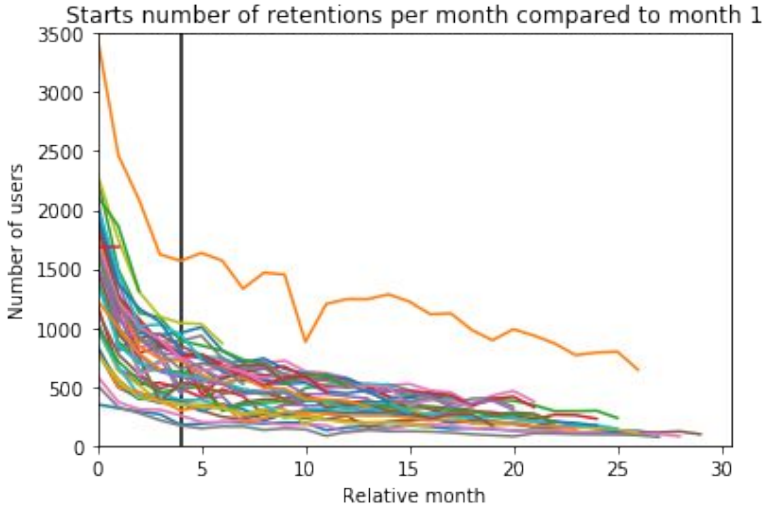


# Analysis

Retention overall

44 starts characterized by their on-boarded users at first connection

- drop in the retention by 25% in the first months for 39 starts
- drop in the retention by 50% in the first 3 months for 37 starts with small rebounds in frequentation




**Global Trend: Rapid decrease over time**

# Analysis

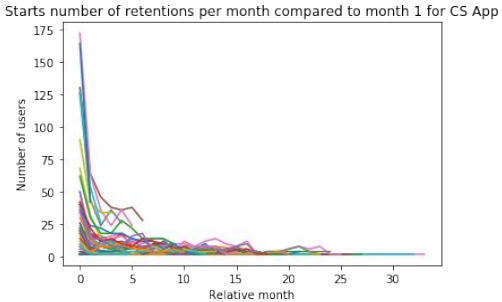
## Retention_by_product

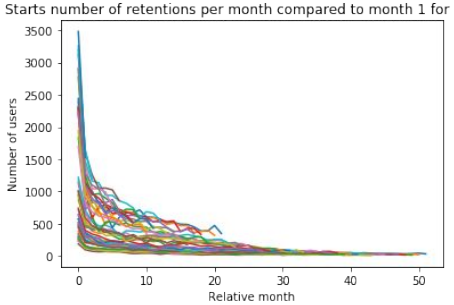
| | product | first_connection | month | relative_month | total_users_per_month | new_users_onboarded_on_first_connection_month |
|---|---|---|---|---|---|---|
| **0** | Digital | 2017-12-01 | 2017-12-01 | 0.0 | 326.0 | 326.0 |
| **1** | Digital | 2017-12-01 | 2018-01-01 | 1.0 | 220.0 | 326.0 |
| **2** | Digital | 2017-12-01 | 2018-02-01 | 2.0 | 196.0 | 326.0 |
| **3** | Digital | 2017-12-01 | 2018-03-01 | 3.0 | 170.0 | 326.0 |
| **4** | Digital | 2017-12-01 | 2018-04-01 | 4.0 | 168.0 | 326.0 |
| **...** | ... | ... | ... | ... | ... | ... |
| **2562** | CS App | 2022-01-01 | 2022-02-01 | 1.0 | 54.0 | 126.0 |
| **2563** | CS App | 2022-01-01 | 2022-03-01 | 2.0 | 26.0 | 126.0 |
| **2564** | CS App | 2022-02-01 | 2022-02-01 | 0.0 | 164.0 | 164.0 |
| **2565** | CS App | 2022-02-01 | 2022-03-01 | 1.0 | 42.0 | 164.0 |
| **2566** | CS App | 2022-03-01 | 2022-03-01 | 0.0 | 138.0 | 138.0 |

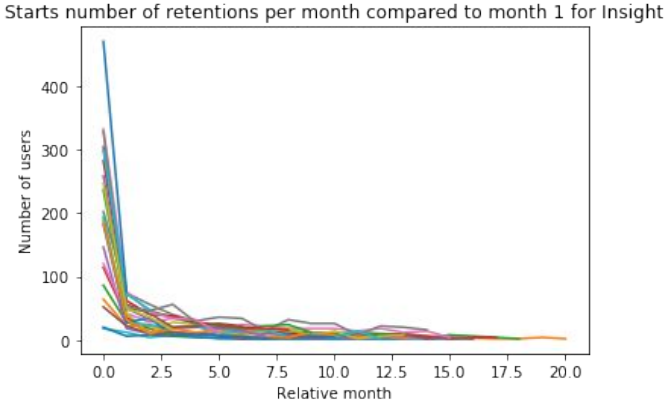
| | product | number_of_starts |
|---|---|---|
| **0** | Digital | 52 |
| **0** | Emerch | 22 |
| **0** | Find & Fix | 21 |
| **0** | Insight | 23 |
| **0** | CS App | 45 |

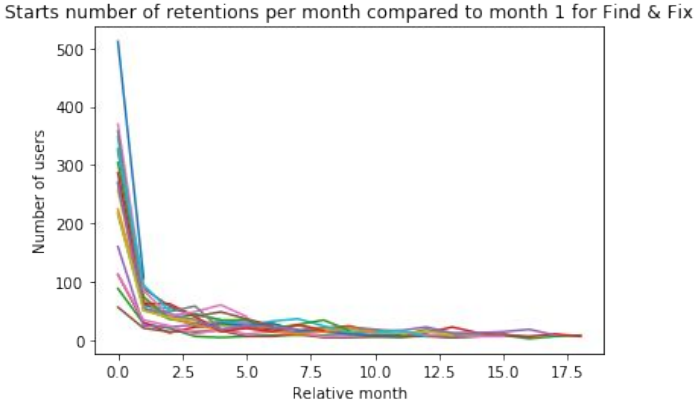
# Analysis

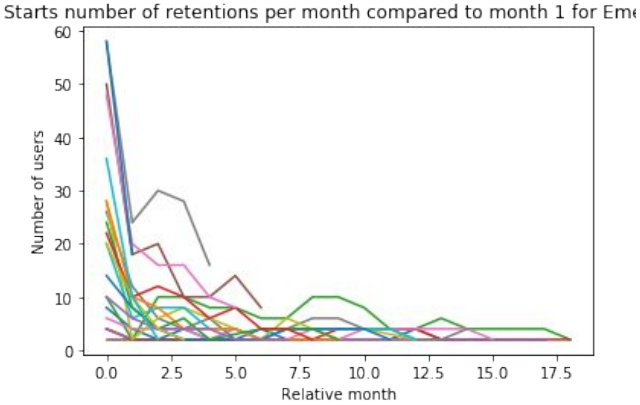
## Retention_by_product



# Analysis

## Retention_by_product

| | product | month_50_decrease | month_50_decrease_n_starts | total_number_starts_product |
|---|---|---|---|---|
| **0** | Digital | [2294.0] | 1 | 52 |
| **1** | Digital | [268.0, 296.0, 442.0, 940.0, 1012.0, 1700.0, 1838.0, 1942.0, 2192.0, 2228.0, 2290.0, 2304.0, 2314.0, 2388.0, 2442.0, 2778.0, 2898.0, 2916.0, 3262.0, 3482.0] | 20 | 52 |
| **2** | Digital | [186.0, 232.0, 236.0, 266.0, 290.0, 292.0, 446.0, 474.0, 514.0, 520.0, 618.0, 644.0, 710.0, 738.0, 866.0, 958.0, 998.0, 1134.0, 1176.0, 1218.0, 2214.0, 2290.0, 3126.0, 3200.0] | 24 | 52 |
| **3** | Digital | [386.0, 490.0, 572.0] | 3 | 52 |
| **4** | Digital | [352.0] | 1 | 52 |
| **5** | Digital | [326.0] | 1 | 52 |
| **0** | Emerch | [2.0, 4.0, 46.0] | 3 | 22 |
| **1** | Emerch | [10.0, 20.0, 22.0, 24.0, 26.0, 28.0, 36.0, 48.0, 50.0, 58.0] | 10 | 22 |
| **2** | Emerch | [8.0, 10.0, 14.0, 20.0] | 4 | 22 |
| **4** | Emerch | [6.0] | 1 | 22 |
| **0** | Find & Fix | [2.0, 300.0] | 2 | 21 |
| **1** | Find & Fix | [56.0, 88.0, 112.0, 160.0, 216.0, 218.0, 224.0, 256.0, 268.0, 270.0, 286.0, 304.0, 324.0, 328.0, 348.0, 358.0, 370.0, 512.0] | 18 | 21 |
| **0** | Insight | [254.0] | 1 | 23 |
| **1** | Insight | [20.0, 52.0, 64.0, 86.0, 114.0, 120.0, 146.0, 182.0, 186.0, 190.0, 194.0, 202.0, 236.0, 246.0, 258.0, 282.0, 298.0, 304.0, 328.0, 332.0, 470.0] | 21 | 23 |
| **2** | Insight | [18.0] | 1 | 23 |
| **0** | CS App | [2.0, 4.0, 138.0] | 3 | 45 |
| **1** | CS App | [8.0, 10.0, 12.0, 14.0, 16.0, 18.0, 20.0, 22.0, 24.0, 26.0, 30.0, 34.0, 36.0, 38.0, 40.0, 42.0, 50.0, 62.0, 68.0, 90.0, 126.0, 130.0, 164.0, 172.0] | 24 | 45 |
| **2** | CS App | [6.0, 8.0, 18.0, 22.0, 42.0] | 5 | 45 |
| **5** | CS App | [8.0, 20.0, 36.0] | 3 | 45 |

Digital 2 months

Emerch 1 month

Find & Fix 1 month

Insight 1 month

CS App 1 month

# Analysis

Retention_by_product

50% decrease in the first 1 or 2 months, whatever the start of recording, which might be coherent with the global trend. Some rebounds between month 2 and 5.

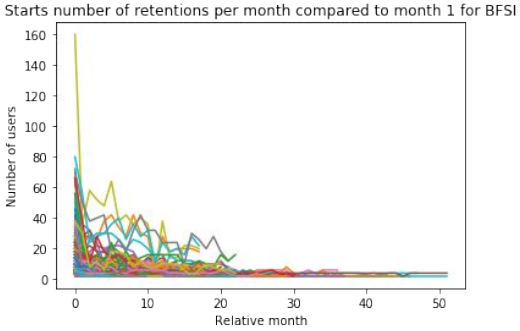
# Analysis

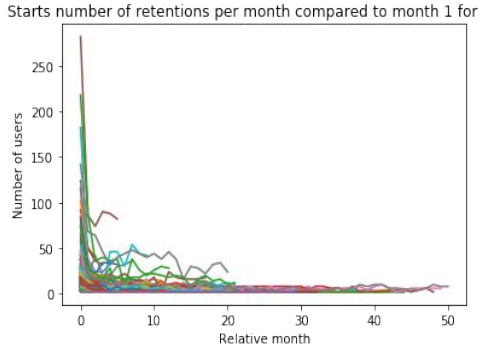
user_retention_by_account

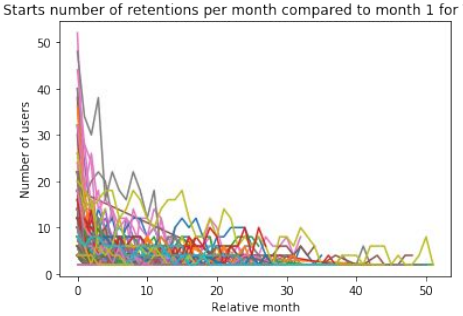
50% decrease in the first 1 or 2 months, whatever the start of recording, which might be coherent with the global trend

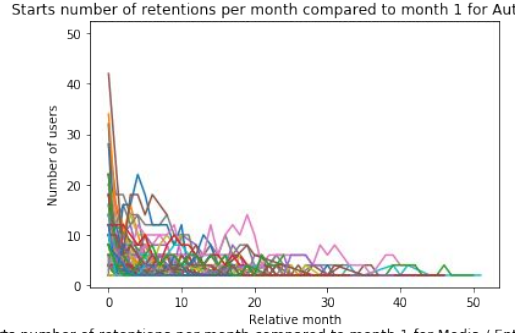
# Analysis

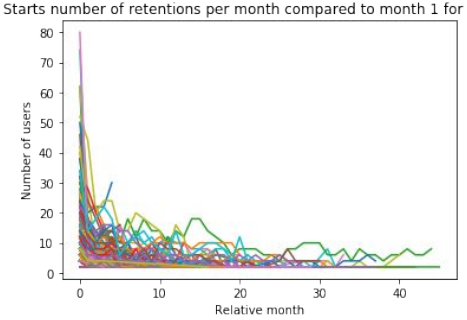
retention by industry

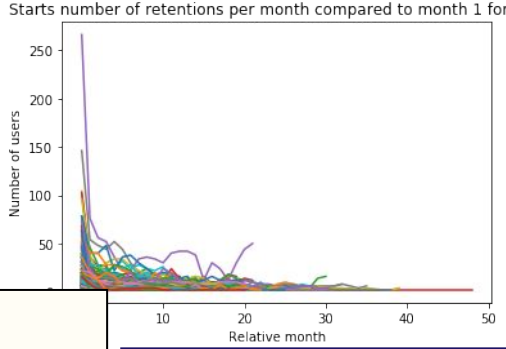


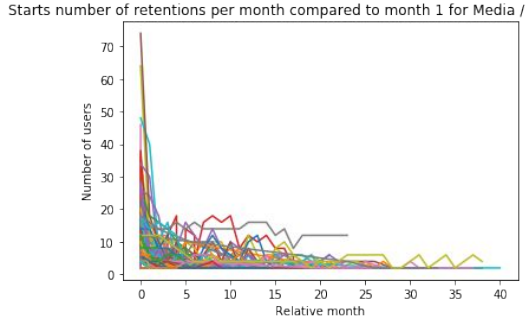



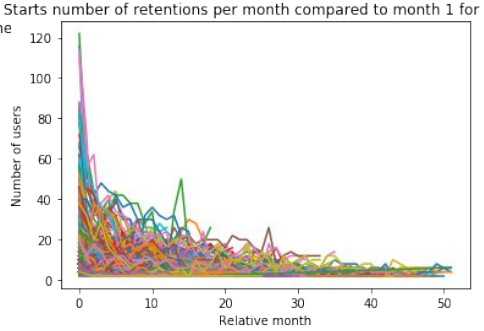



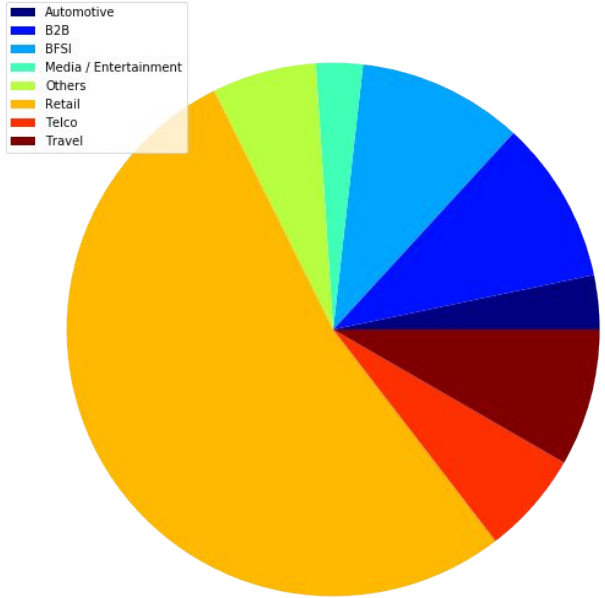



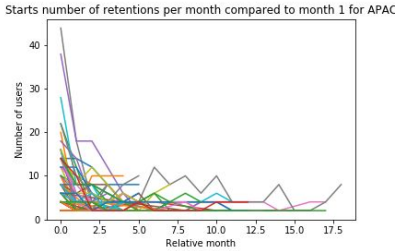
# Analysis

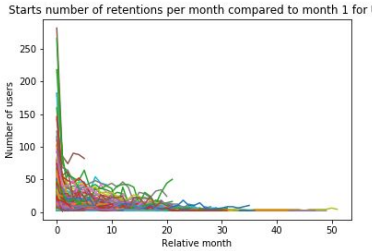
repartition of starts by industry

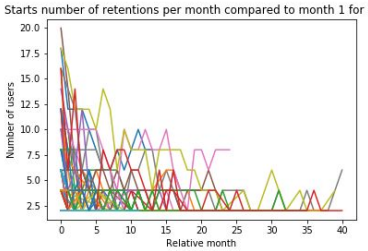


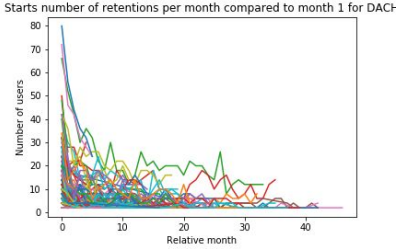
# Analysis

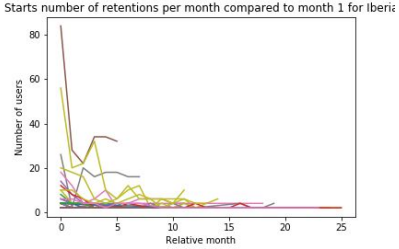
Retention by countries

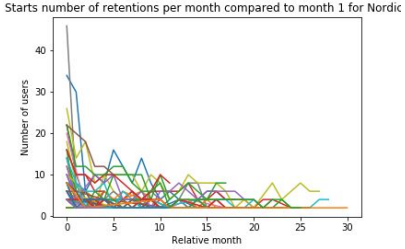


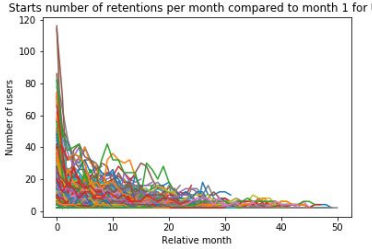



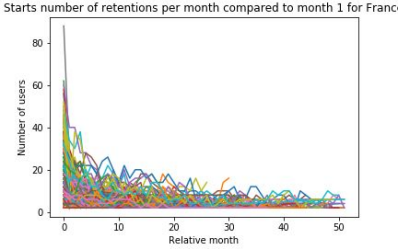



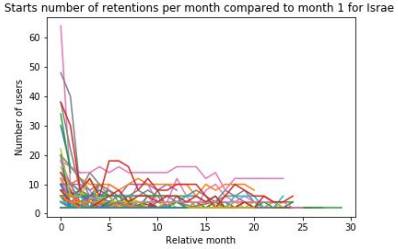



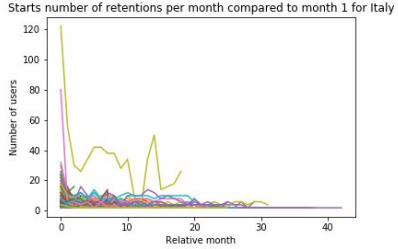



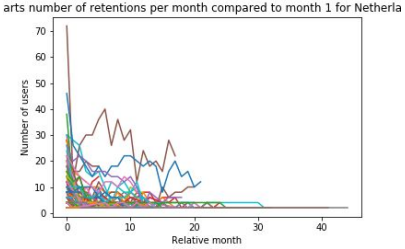



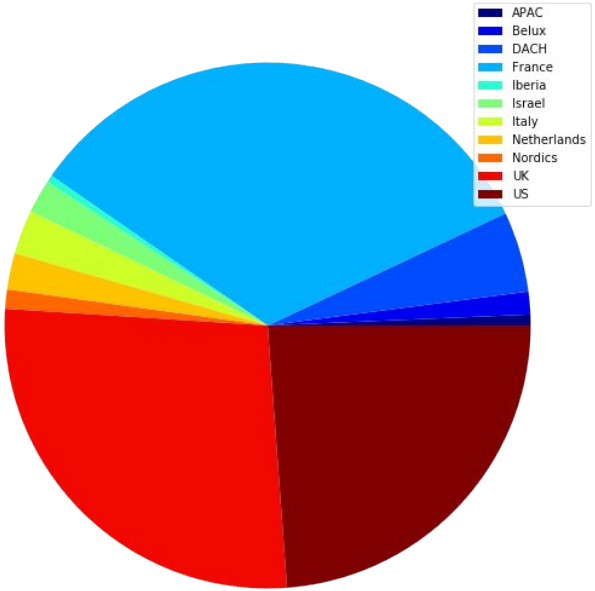



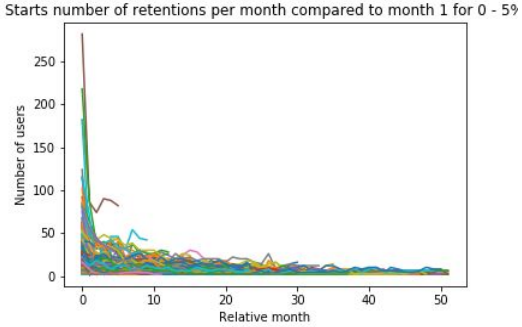
# Analysis

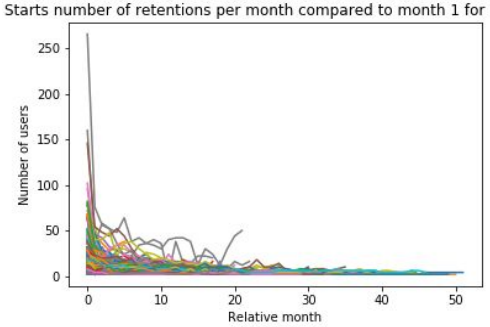
repartition of clients per country

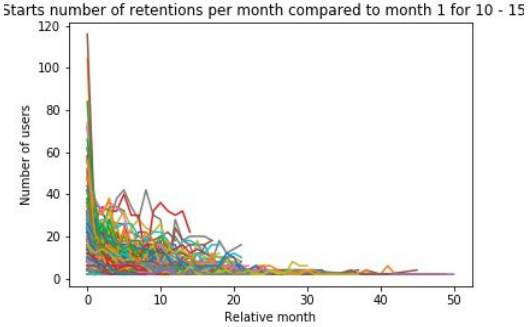


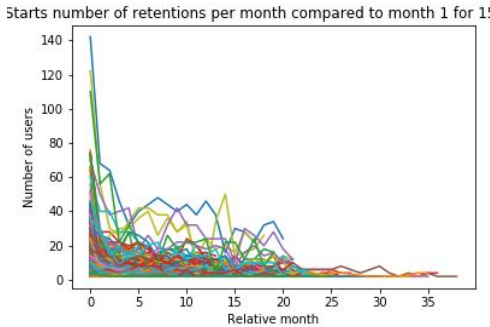
# Analysis

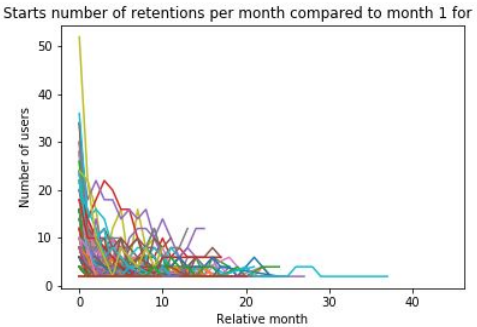
retention per service category



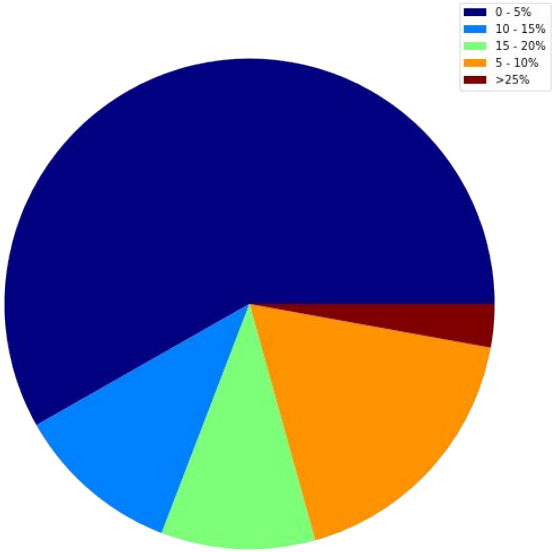



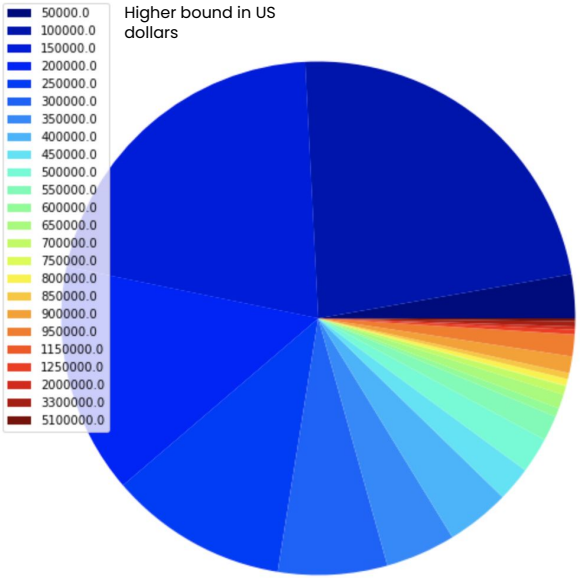
# Analysis

Repartition of clients per service category



# Analysis

Total ARR
Drop mostly in the first month



# Analysis

account_metadata/geography/Digital Marketing Firm

- When joined with user_retention_by_account, the table gives indications about the retention by industry, ARR and revenue. Exception for retail for which it is about 5 months and represents half of the clients.
- Rebounds often present after first drop. Useful: have knowledge of advertising campaigns or changes in the versions
- Overall, the same high drop in the first month or two is observed. Except for retail and automotive which take about 5 months for the same 50% smoothing.
- Identically for the geography and service table.
- On the last table there are only few firms previously Digital Marketing Firm client.

# Propositions

Valuable attributes to record:

- For more accurate analysis, we should take into account the ids which don't belong to the full join of the last three tables
- finer grained data in the first 2 months to see periodicity and where the drop actually occurs
- segmentation of the users or personal data in order to cluster
- we could try to find pattern in the users that don't check the website every month
- Try to understand what causes the rebounds

Solutions proposed to the clients:

- focus on the first two months
- design campaigns aimed at several uses - retaining first users (as it is a rolling measure)
- experimental surveys on reasons for not coming back to the website (if it is an advertisement campaign on a site, for instance)