**Preject:**

**The task:**

You’re a data analyst freelancer.

A young startup recently launched an innovative online magazine and is investing heavily in online marketing in order to grow its subscriber base. As a part of their growth effort they need to better understand their user base and is contacting you to consider hiring you for the job.. As a part of the selection process, they are sending you a sample of 50k of their users and expect you to answer basic exploratory questions on this dataset.

The company is currently investing heavily in online marketing channels in order to acquire subscribers. The marketing channels are Criteo, Facebook, Instagram, Adwords and some select Partners (grouped under the “Partner” category). Additionally, they offer a referral programme for their existing subscribers to invite their friends.

The core of their business model is offering interesting content so users acquire and retain a paid subscription. At the moment, subscriptions can be 1 Month, 3 Month, 6 Month and 12 Month long.

The business database consisting of a few tables:

|  |  |
| --- | --- |
| **Table/dict** | **Fields** |
| Main (users) | User ID,  Gender,  Date of birth - split into three columns Day Month and Year,  Location |
| Product | Product,  Intervals purchased so far,  Purchase date,  Voucher (percentage of reduction) |
| User attribution channel | User ID  Attribution Channel |
| User article | Datetime  Liked - binary (0-no,1-yes)  Article info - list containing- article category (1-7), article id, reading length in minutes  User\_id |
| Dict 1 - marketing costs | Per channel, budget spent |
| Dict 1 - product prices | Price for every product |

You can find these files in the drive or on github following these four links :

* [Main](https://raw.githubusercontent.com/aaronmcdaid/P2---Data-Analytics-With-Python/master/Berlin/WS%202019/5.%20Datasets/preject_datasets/main.csv)
* [Product](https://raw.githubusercontent.com/aaronmcdaid/P2---Data-Analytics-With-Python/master/Berlin/WS%202019/5.%20Datasets/preject_datasets/product.csv)
* [User attribution](https://raw.githubusercontent.com/aaronmcdaid/P2---Data-Analytics-With-Python/master/Berlin/WS%202019/5.%20Datasets/preject_datasets/attribution.csv)
* [User article](https://raw.githubusercontent.com/aaronmcdaid/P2---Data-Analytics-With-Python/master/Berlin/WS%202019/5.%20Datasets/preject_datasets/user_article.csv)

**In addition:**

marketing\_budget\_split = {'Criteo':185000,'TV':30000,'Facebook':550000,'Adwords':40000,'Instagram':700000,'Partner':'25% Share of first subscription price with minimum of 5 euros'}

prices = {'1M':12.99,'3M':32.99,'6M':53.99,'12M':83.99}

**Questions:**

Every question is an open ended question and phrased in a non precise way on purpose. Part of the analyst's job is to answer ambiguous tasks with analytical answers. There exist multiple ways of answering the same question. We are expecting you to explain which hypothesis you chose when giving your answers.

* Who are our customers? (gender, age, where from?)
* Where and what are we selling the most? (locations, products)
* Can you identify some reading patterns? (times a day, days a week, how many times a user goes back to an article, most liked/read articles)
* What is the correlation between article reading length - liked - times read?
* Summarise your findings in a presentation aimed at the startup where you will tell them what you found and what would be your recommendations for them.

**Bonus** - Which marketing channels are performing well?

\*CAC or Customer Acquisition Cost is a very common metric in subscription services and is defined, for each marketing channel, as the Total cost of Marketing divided by the number of customers attributed to that channel

**Topics we cover:**

* Import files onto pandas
* Data frames & pandas
* Missing values
* Data manipulation (e.g creating bins/buckets eg Age groups)
* Aggregations and joins
* Visualizations (gender, age, location, promotion, product, Correlation between different fields)
* Date-string functions
* Analytics : approach a business dataset, what & how to visualize