

# Marketing Data Scientist

## Technical Test

The dataset `technical_test_data.csv` contains activity data of 100K distinct users that installed the game during a given marketing campaign. The dataset includes data of the first 60 days since the users first install. You can find more details on the description of the data in the Appendix section.

We are interested in modelling the revenue that a campaign will make. To do so, we model the *LTV Day X*: Here by LTV Day X we mean the Average User Lifetime Revenue where X is the number of days since the installation day.

Your task is to develop a predictive model to estimate the LTV in order to answer the following questions:

- Considering that the marketing campaign costs equal to \$400, will the campaign be *profitable* after one year? By profitable here we mean that the revenue made by the users brought by this campaign, *i.e.*  $(LTV_{D365}) * (\text{Number of Users})$ , will exceed the total campaign cost, *i.e.* \$400.
- Assuming that the campaign will not be profitable after one year: Will the campaign ever be profitable?
  - If yes, how much time will it pass until the campaign becomes profitable?
  - Regardless of the profitability of the campaign, how much money will the campaign generate?
- Given the obtained results, what would you recommend regarding this type of investment? Which confidence can you provide on your recommendation? Which is the maximum amount that would you recommend to pay for every new game user?
- Create a presentation including the methodology and the results.
- Please share your code. And please prepare it in such a way that any other member of your team can continue working on it with minimum help from your side. Assume that right after you complete this project you are asked to work on a completely different new topic and the colleague being handed over your project must rely fully on your code and documentation.

# Appendix

## Data Set Description

The dataset `technical_test_data.csv` contains activity data of 100K distinct users that installed the game during a given marketing campaign. The dataset includes data of the first 60 days since the install.

Each row of the dataset corresponds to a day of activity for a given user. The rest of columns correspond to dimensions and metrics regarding the user and the activity day. Find a detailed description of each field below:

Field	Description
<code>user_id</code>	User ID
<code>days_from_install</code>	difference between install date and activity date
<code>platform</code>	register platform where the user installed the app (iOS or Android)
<code>device_type</code>	type of device used for the first install (unknown, iphone or ipad)
<code>os_version</code>	operative system version of the first install device
<code>day_of_week_install</code>	day of the week of the first install date. Integer from 0-6, starting with Sunday.
<code>iap_revenue</code>	IAP (In App Purchases) revenue for the given user and day from install (\$). Note that this metric is normalized to [0,1].
<code>ad_revenue</code>	Ad Revenue for the given user and day from install (\$)
<code>num_transactions</code>	number of IAP transactions for the given user and day from install
<code>num_sessions</code>	number of game sessions started for the given user and day from install
<code>time_played</code>	total time played for the given user and day from install (counting all sessions). Note that a user may have time played in a day with no sessions if the session was started in the previous day
<code>end_level</code>	the max game level for the given user and day from install

## Particular Example

The following rows correspond to the data of a single user.

User: 3353038463122941451

days_from_install	day_of_week_install	iap_revenue	ad_revenue	num_transactions	num_sessions	time_played	end_level
0	1	0.0142	0.0000	1	3	5230	1
1	1	0.0071	0.0001	2	5	7423	9
2	1	0.0042	0.0000	1	3	4725	13
3	1	0.0000	0.0000	0	2	7175	13
7	1	0.0000	0.0000	0	3	2333	15
8	1	0.0000	0.0000	0	2	965	15
14	1	0.0057	0.0000	2	4	10133	15
17	1	0.0000	0.0000	0	2	2134	15
21	1	0.0071	0.0003	1	3	8251	15
22	1	0.0000	0.0001	0	2	2115	16
23	1	0.0042	0.0003	1	6	5685	16
24	1	0.0000	0.0004	0	2	4070	16
25	1	0.0000	0.0000	0	1	145	16
33	1	0.0000	0.0002	0	1	1343	16
42	1	0.0071	0.0000	1	3	3337	16
43	1	0.0000	0.0001	0	3	3591	16
44	1	0.0000	0.0001	0	1	961	16
50	1	0.0000	0.0001	0	3	4207	17
57	1	0.0000	0.0000	0	1	0	17

Each row corresponds to a distinct day of activity of the user before day 60. If a day is not in the dataset, (for example, day 4) it means that the user did not have activity on that day.

On day 0, the first day since installation, the user spent \$0.01419 with 1 transaction and had 3 sessions.

To compute the LTV DX,

$$\text{LTV DX} = \text{SUM}(\text{iap\_revenue} + \text{ad\_revenue}) \text{ where days\_from\_install} \leq X$$

In particular, LTV D3 = \$0.0256 and LTV D60 = \$0.0512.