Data Analysis in Python – Project Proposal

Teammates:

Amit Korenberg, 300092178

Mor Stone, 200735447

1. Introduction

yellowHEAD is a digital marketing company, that provides both paid and organic services to promote apps and web products. Currently, the company’s portfolio mostly consists of gaming apps and ecommerce apps/websites.

Our data: all performance data from the last 100 days of one social casino game campaigns.

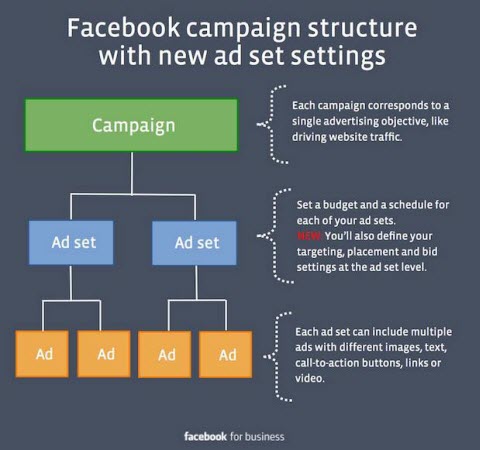
The targeted Facebook users go through the following flow:

1. A user is scrolling his Facebook news feed and stops when he sees our ad, that presents the game with a “Play Now” call-to-action.
2. Since he finds our game interesting, he clicks on the ad which leads him to the app store, there he can find more information about the game.
3. The user can decide whether to install the game for free or leave the store.
4. After installing the game, the user enjoys playing and can purchase in-app features, such as: free coins/tokens, level-up, etc.
5. Our client, a big game developer, has set performance targets for this app:

* Reach 200$ revenue per campaign after 7 days.
* Reach 50 installs per campaign after 7 days.

1. Attributes are the following:

* Campaign\_ID: Following company restrictions, we removed campaign names and replaced with unique IDs.
* Num\_Creative: the number of ads (videos) under each campaign, according to the following campaign structure by Facebook:



\* We skipped ad-set level in order to simplify the data, since the structure is mostly to create order.

* GEO: the targeted countries – Worldwide, Tier 1, US.
* Audience: the targeted group of people, all above 1M users.
* Device: the targeted OS and device - Android, iPhone, iPad.
* Opt\_Goal: what Facebook optimizes for, meaning we ask Facebook to show our ads to people who are most likely to perform a certain desired action:
  + Purchases - Most likely to purchase in the app.
  + Value - Most likely to deposit high amount.
* Amount\_Spent\_D3: The campaign spend in the first 3 days.
* Impressions\_D3: Number of impressions in the first 3 days.
* Clicks\_D3: Number of clicks in the first 3 days.
* Installs\_D3: Number of installs in the first 3 days.
* Installs\_D7: Number of installs in the first 7 days.
* Installs\_D7\_target: Marked as “1” if Installs\_D7 is above 50, and “0” otherwise.
* FTD\_D3: Number of FTDs (First Time Depositors) in the first 3 days.
* Rev\_D3: The revenue after 3 days.
* Rev\_D7: The revenue after 7 days.
* Rev\_D7\_target: Marked as “1” if Rev\_D7 is above 200$, and “0” otherwise.

1. Research Questions:

* 1st Question: Predict if Revenue\_D7 result is above 200$.
* 2nd Question: Predict if Installs\_D7 results is above 50 installs.

Our 1st question is highly beneficial, since that’s the main KPI from our client.

Our 2nd question is also a client KPI and could help evaluating a certain campaign, especially when Rev\_D7 prediction is lower than 200$. In this case, driving enough D7 installs could also drive more paying users.