**Introduction**

Welcome to the assignment on Facebook advertising strategy analysis!

Facebook is betting a lot on its e-commerce ambition, slowly bringing the businesses and consumers to coalesce on Instagram, WhatsApp, and Messenger (F8 refresh 2021 conference).

Users of Facebook and its family apps provide a massive amount of information which are used to develop a detailed profile of users' personal choices, habits, and interests. Advertisers of the Fb platform use the ad-targeting function to pinpoint relevant users who are most likely to buy. This helps businesses see through a microscope and define an audience ready to click and convert.

For example, A travel agency might promote a visit to a tourist place as a honeymoon destination package to newly engaged couples. GoPro manufacturers might target photographers or people who love to hike or ski. A baby product company would like to target pregnant women and women with a new baby. Let's imagine a company wants to advertise its new car. They may wish to promote one set of features, performance and the 2 kW stereo, to women in their early twenties. They might decide that they want to talk about its fuel efficiency and reduced emissions to men in their thirties, and they might want to push the spacious interior and safety rating to men and women in their thirties and early forties who are interested in Families magazine and who like pages of nappy and baby clothes manufacturers.

The ultimate aim of businesses analysing these data and optimizing their advertising strategy is to

* improve their brand awareness,
* improve future sales,
* increase customer base,
* maximize the amount of revenue returned for the minimum expenditure.

**Dataset**

**Dataset**

**Download**

**Data Descriptions (fb\_ad\_data.csv)**:

1. ad\_id: a unique ID for each ad.
2. xyz\_campaign\_id: an ID associated with each ad campaign of xyz company.
3. fb\_campaign\_id: an ID associated with how Facebook tracks each campaign.
4. age: age of the person to whom the ad is shown.
5. gender: gender of the person to whom the ad is shown
6. interest: a code specifying the category to which the person’s interest belongs (interests areas mentioned in the person’s Facebook public profile).
7. Impressions: the number of times the ad was shown.
8. Clicks: number of clicks on for that ad.
9. Spent: Amount paid by xyz company to Facebook, to show that ad.
10. Total\_Conversion: Total number of people who enquired about the product after seeing the ad.
11. Approved\_Conversion: Total number of people who bought the product after seeing the ad.

**Part A. Statistical Inferencing Tasks: (75 points)**

1. Compute the following:
   1. Probability of an ad having one click (4 points)
   2. Probability of an ad showed to Male (4 points)
   3. Probability of an ad showed to Female (4 points)
   4. Probability of an ad having one-click given that it is shown to Male. (4 points)
   5. Probability of an ad having one click given that it is shown to Female. (4 points)
   6. Probability of an ad having one click and it is shown to Male. (4 points)
   7. Based on the computation of the above-mentioned probabilities, draw inference regarding the independence of events, if any. In particular, do you think that the event of “an ad having one click” is independent of the events that “the ad is shown to Male” or “the ad is shown to Female”. (4 points)
2. Compute the mean value of a click on an ad. (4 points)
3. Compute 95% and 99% confidence intervals for the true value of a mean click. Note that a 100(1−α)% confidence interval for the mean can be computed as (for n ≥ 30)¯x−Zα/2S√n≤μ≤¯x+Zα/2S√n

where n=number of observations, ¯x is sample average, s is sample standard deviation, α is the significance level (=0.05 for 95% confidence interval, and 0.01 for 99% confidence interval), and Zα2 is that value from a standard normal distribution such that area to the right is α/2. (4 + 4 = 8 points)

1. Suppose a data scientist is designing a recommendation system based on the available data. One of the crucial parameters in the algorithm is the mean value of a click on an ad. She believes that the algorithm might perform arbitrarily bad for low or high values of mean click. Based on past experience, she believes that the mean number of clicks on an ad is 31. She wants to formally test this hypothesis based on the given advertising data. What should be the null and alternative hypotheses? What statistical conclusion can she draw? (11 points)
2. Suppose the data scientist, in the future, wants to predict several probabilities regarding a number of clicks on an ad. So, she wants to fit a probability density function to the given click data. Help her by doing the following:
   1. Fit an exponential pdf. [One way to do this will be to use the following function in python: scipy.stats.expon.fit()]  (4 points)
   2. Compute the probability of a number of click exceeding 100 for an ad using,
      * clicks data (4 points)
      * pdf fitted in part I above (4 points)
   3. Analyse the difference between the computed probabilities in 2, if any. (2 points)
   4. Repeat the exercise mentioned in 2 and 3 for a number of clicks exceeding 400 for an ad. (4 + 4 + 2 = 10 points)

**Part B: Exploratory Data Analysis (75 Points)**

To analyse the business performances, one can construct various Key Performance Indicators (KPI), such as, Click through rate, Conversion rate, Return on advertising spend, etc. The choice will mainly depend on the impactful solution to a relevant question which business organisations are seeking. For the current study, we will focus on the "Cost per conversion (CPC)". It is defined as follows:

CPC = (Amount spent by the company on an ad) / (Total number of people who enquired about the product + Total number of people who bought the product)

Exploratory Data Analysis Tasks:

Carry out detailed Exploratory Data Analysis on the assigned data set where the target KPI is CPC and draw meaningful insights from various data displays, pictorial representations, measures, and present your findings. It must include the following (but don’t necessarily restrict yourself to this):

1. Scatter plot and correlation: CPC vs Click, CPC vs Spent, CPC vs Total\_Conversion
2. CPC analysis by age, gender, and interest. Then draw conclusions about the groups to target (and not to target) in these genres, so that the business performance can be improved.