**1. Overview**

In this project, we have to identify what is the best time to post in social networks. The objective is to find the best times for a user to post on social networks in order to maximize the probability of audience responses. We will use dataset for Facebook Pages. A Facebook page is a public profile that is created for businesses, brands, celebrities, causes, and other organizations to share the information with the large audience. Admin of the pages share the information, known as posts, to broadcast the information to the audience. Lifetime of a post is very short, typically few hours. Facebook uses a ranking algorithm where the rank depends on posting timestamp. Therefore, if it is not posted at right time, it may not reach to the maximum audience. There are different categories of Facebook Pages, such as e-commerce, telecommunication, hospitals, politicians, traffic control etc. Our goal is to see if there is any significant difference in users’ interaction pattern across these different categories of Facebook pages.

We are providing you a small dataset extracted from Facebook Pages in the form of Excel. This excel file contains information about pages such as category of a post message, no of likes, no of comments, posting timestamp, comments, time when comments were created, etc.

Following are the notations used in questions:

tk = 15 minute timespan(one bucket size). Four buckets represent one hour ck(p): number of posts created in time bucket k in a day

Ck(p): estimated number of posts created by page p, in each time bucket tk rk(u): reactions done by users of page p in each time bucket tk

Qk(p): estimated number of reactions received by page p in time bucket tk Sk(p): probability of receiving reactions on page in each time bucket tk Ak(C): aggregated no of posts created by category C in each time bucket tk W(p): fraction of reaction received by a page p in its category

RG(p,k): reaction gain of a page p at time bucket k

The database will have two tables---**PostSummary** and **Comments***---*with following schema:

**PostSummary** (pid, category, postedBy, createdTime, message, likesCount)

**Comments** (pid, commentsText)

You can join the two tables using the pid attribute.

In **Comments** table there are many comments (in first row of Comments table itself) and comments are separated by special symbols **?#+@** (and comment time is there at the end of every comment)

# Exercise 1:

You need to analyze the posting behavior of Facebook Pages using the **PostSummary** table.

Divide 24 hours of a day into 96 buckets each of 15 minute and create a vector Ck(p) for post profile as Ck (p) = [C1 (p),….. ,C96 (p)] ; k= 1, 2, 3, …,96. For example, C1(p) contains information about posts created between 12:00 AM to 12:15 AM.

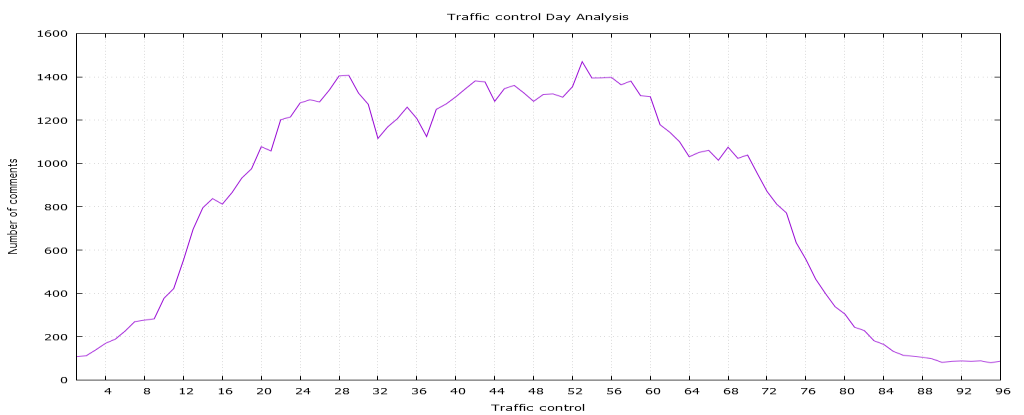
Ck (p) is computed as follows:

**Ck(p) = ∑m =0 ck(p)**

**j**

Here, m is the time period of a year.

Plot the graphs of aggregated no of posts Ck(p) for daily analysis of the following four Traffic Pages namely Bengaluru Traffic Police, Kolkata traffic police and Hyderabad Traffic Police.



# Exercise 2:

Create an excel chart to calculate the average number of likes per post for each category.  
Create a word cloud (you can use any tool) on post data for each organization (PostedBy).

# Exercise 3:

Create a presentation to summarize your findings for all the above exercises in a PowerPoint.  
Please feel free to add any other insights that you may draw from the data.

# Exercise 4:

Create an approximately 500 words write-up to explain the social posting behavior adopted by various categories and any other insights you may have drawn from the data.

# Exercise 5 (optional):

# The objective of this exercise is to evaluate your logical/analytical approach in solving the below problem.

# If you are unable to solve this in excel, please send your thoughts/methodology you would take in solving this.

Similar to Exercise 1, you need to analyze the reactions behavior of users on Facebook Pages using the **Comments** table.

Iterate through reaction profile r(u) of users on each page p, compute Qk (p) as follows:

**Qk(p) = ∑m =0 rk(u)**

**j**

Here, m is the time period of one year

Plot the graphs of estimated number of reactions Qk(p) for daily analysis for the following four e- commerce Pages namely Flipkart, Amazon, Snapdeal and Myntra.

**Hint:** Each column contains multiple comments separated by ?#+@ followed by its time stamp.

You have to extract the time stamp for each comment.