



---

# ASSIGNMENT 2

---

Kaustubh Jagtap – A0168820B



## Understanding and Pre-Processing the data

**Fans** – This file contains information about the number growth in fans from December 2014, broken down by channel and region. There are two columns, *New Fans* and *Fans*, which contain the same data – the additional number of fans that joined on that date. Presumably, the *Fans* column is supposed to show the cumulative number of fans as of that date. I wrote a Python script to make that happen:

```
1  ## Name: Kaustubh Jagtap
2  import csv
3
4  def read_csv(csvfilename):  ## function to read a csv file
5      rows = []
6      with open(csvfilename) as csvfile:
7          file_reader = csv.reader(csvfile)
8          for row in file_reader:
9              rows.append(row)
10     return rows
11
12     data = read_csv("Fans.csv")      ## read in the file
13     first_row = data[0]              ## isolate the column headers
14     data = data[1:]
15
16     totals = {}                      ## hash table to store running totals
17     def getFans(data, first_row):
18         for rownum in range(len(data)):
19             channel, region = data[rownum][0], data[rownum][3]
20             ## concatenate the channel and region to get unique key
21             concat_val = channel + " " + region
22             ## do the appropriate insertion into dictionary
23             if(concat_val) not in totals:
24                 totals[concat_val] = int(data[rownum][4])
25                 continue
26             else:
27                 totals[concat_val] += int(data[rownum][2])
28                 data[rownum][4] = totals[concat_val]
29
30     data = [first_row] + data
31     with open("newFans.csv", 'w', newline='') as f:
32         writer = csv.writer(f, delimiter=',')
33         writer.writerows(data)        ## write to csv
34     return f
35
36     getFans(data, first_row)
```

**Facebook** – This data file consists of engagement levels amongst Facebook followers, for 120 posts. The data is broken down by region, age group and gender, where each post contains one of each permutation. To derive engagement statistics for Facebook, we will have to divide the aggregated data by 120 to get a 'per-post' value.

**Brand Posts** – These are social media posts by Sid's Cookies as part of their marketing/ business development initiatives to engage the audience and create a fan base. The two key equations from this file are:

$$\text{Consumptions} = \text{Clicks} + \text{Link Clicks} + \text{Photo Clicks} + \text{Video Clicks}$$

$$\text{Engagements} = \text{Comments/Replies} + \text{Favourites} + \text{Shares/Retweets}$$

Initial eyeballing of the data shows that the company is disproportionately more active on twitter than on Facebook or Instagram. It also posts more heavily in India, Indonesia and Australia.

**User Posts** – Shows user posts on Facebook and Twitter, broken down by their region and type of post. Additional data is given for the engagement that post achieves, which tells us how widely-received the post is. For Twitter users, we have data on their number of followers and 'klout score', which can potentially convey the popularity and reach of our customers. Posts can be updated too, and each update receives a unique update ID. However, each update simply creates a new post and there is no way to tell from the data file how many of the posts are actually updates.

## Assumptions and Clarifications

- The first reading for each channel-region combination in *Fans* was the cumulative reading from inception to-date. This makes sense, since the numbers were disproportionately higher than the subsequent fan growth metrics.
- All engagement statistics in *Facebook* are present and recorded (i.e. each post contains information from every country and for all age groups). This assumption leads to the calculation that there are 120 posts in that file.
- Null Values for 'Gender' in *Facebook* were removed for when reporting a breakdown by gender, since it is confusing and adds no valuable insights.
- The equations for *Brand Posts* holds true for all data points, and if the data is correct.
- The date-time under *Brand Posts* are all according to the time zone of that specific country, and not normalised.
- All posts in *User Posts* are considered unique. This is because there is no way to tell which post is an update and which is new. There might hence be some double counting, but it would seem to be negligible.
- Null values across all data files either mean 0, or non-existent, depending on the context. For example, in *User Posts*, null values in Followers and Klout Score for Facebook entries mean that that metric does not exist on Facebook. Conversely, null values in *Brand Posts* are taken to mean 0.

## Dashboards

Taking a holistic view of the provided data, I decided to split my dashboards into 3 separate screens, each visualising closely related data that answers similar questions. The target audience of these dashboards are the marketing/ business development teams, with intention to report to higher management. The unifying question that I seek to answer is, “What does this data mean for the business, and what actions should we take based on the data?”. Essentially, it is the “so what” for the business. The dashboards are explained below, in relation to these questions.

### Fans & Followers Insights:

This dashboard aims to convey the overall Fans and Followers statistics. The top left line chart depicts the growth in fans from December 2014, broken down by channel. The further question that arises from this is – what is our total fan base from inception to-date? This question is answered by the stacked bar chart on the right, which shows the total follower base, broken down by region and channel.

Our biggest platform is Facebook, and some of the data looks worrying (given the recent plunge in followers). Therefore, I decided that the bottom-left chart should focus solely on Facebook, showing the breakdown of engagements (roughly a proxy for active fans), by region and gender. The filters for this dashboard are simple and elegant – date (locked from the left to provide meaningful insights), region (for in-depth analysis of any particular country) and channel (similar to region).

Overall, this dashboard gives us a ‘big picture’ view of our global reach and influence across all 3 social media channels. We can dig deeper by region, gender and age group. The decisions that can be made from this data include:

- 1) **Make decisions on where to put in more marketing effort.** For example, we can tell at a glance that people above 34 are not being reached out to by our campaign. However, this customer segment has purchasing power and does eat cookies (we can reasonably assume). Perhaps we need to spend our marketing budget to target this group by another channel (say, television).
- 2) **Investigate anomalies in social media following.** We can immediately tell from the top left line chart that there has been a sharp decline in Facebook fan followings during March 2015. Is this due to a macro-change (such as people deleting their Facebook accounts for some reason), or was it our content during that time that caused people to boycott us? This needs to be looked into.
- 3) **Explore possible content changes.** As of today, around more than 75% of our Facebook engagement hits are amongst females. Could it be that our content is too female-oriented? Our dashboard seems to suggest that we need to investigate this and find a way to engage more males.

### Brand Outreach Insights:

This dashboard aims to convey all insights related to brand posts (i.e. posts by us). It depicts consumption and engagement patterns across all our channels. The top left line chart gives an overall view of consumption and engagement patterns. It is not a time series; rather, it is an aggregation of the measures, taken over a time window (by day or by hour). The top right table shows the breakdown of engagements and consumptions, by type. This is aggregated data and to be interpreted as the latest snapshot of the figures.

The bottom left stacked bar chart is the pattern of engagement, broken down by channel. The bottom right stacked bar chart is the pattern of consumption, also broken down by channel. These charts, too, can be analysed by day of the week or by time of the day, using the 'Time Parameter' filter. The filters also include specific types of consumption or engagement, where in-depth analysis can be done on each type. I have also included filters for channel and region, for detailed analysis.

This dashboard allows for some unique insights to be drawn, and for key decisions to be made surrounding those insights. These include:

- 1) **Analyse which time of the day posts are more likely to be consumed.** This can be done by toggling the Time Parameter filter. As an example, engagement ramps up in the evening period between 17:00 to 19:00, so we should post our most impactful content at this time to reach as broad an audience as possible and to make our posts go viral.
- 2) **Analyse what day of the week posts are more likely to be consumed.** This is also controlled through the 'Time Parameter' toggle. From the given data, Wednesdays seem to have the highest traffic. Again, we should choose to reserve our most impactful content for this day.
- 3) **Tell what type of post our audience is most receptive to, in every region.** Once we know this, we can customise our content by region for a more targeted marketing effort. For example, by filtering for comments, we can glean that Instagram has a disproportionately high number of comments – we could leverage on this, perhaps by starting discussion threads or posting content that will draw favourable comments.
- 4) **Decide on what channel to use to maximize user engagement and consumption.** The entire data on the dashboard is broken down by channel. This allows us to compare which social-media platform attracts higher levels of user throughput. Further, we can also tell which platform-specific type of consumption or engagement our fans are involved in. This would allow us to tailor our posts by channel, such that they achieve higher levels of activity.

### User Insights:

This dashboard aims to convey all data related to user activity on our Facebook and Twitter platforms. Our fan following on these 2 social media platforms is significantly higher than that on Instagram. Also, we have more data available for user activity on these sites than on Instagram (perhaps because Instagram consumption tends to be more passive). For these reasons, I decided to have a dashboard focused solely on *User Posts* data on Facebook and Twitter.

The top left tables show our number of active users across all regions, and the breakdown of the type of content our users are posting. Both these tables are broken down by channel. Here, 'active user' is taken to mean anyone who has posted on our social media platforms from December to date. At a glance, this gives us a sensing of the size and activeness of our community.

Questions that will arise after looking at the table would most probably pertain to the profiles of our community members – how popular are they and how widely received is *their* content on our platforms? The other three charts show just this – the top right bar chart depicts the engagement levels of user posts on Facebook; the bottom left chart shows the number of twitter followers broken down by their popularity, and the bottom right heat map digs deeper by further breaking this down into 'klout score'.

This dashboard doesn't have the filters tab; this is to allow for a cleaner look. However, the countries in the tables are clickables and act as filters. This dashboard is great for giving us a sense of the profiles of our followers. We can draw some interesting insights from this information:

- 1) **Identify and target repeat customers.** Repeat customers are a big source of revenue, and in a cookie market where consumer tastes and preferences are a big factor, repeat customers signify a segment who are absolutely in love with our cookies specifically. Using the data on this dashboard, we can identify and target these customers with offers, or even test out new flavours with a select group.
- 2) **Explore engaging a social media influencer.** We know how well received our user posts on Facebook are. We also know the follower base and klout score of our twitter followers. We can leverage on this information to expand our market, by engaging key individuals with a wide reach, to promote our brand.
- 3) **Design country-specific social media strategies.** An active community on social media boosts our brand presence and also allows us to gain valuable feedback on our cookies – we need to get our user base talking! From this dashboard, we know how active our users are in each region. Having this information, we can design specific social media strategies to get our user base talking on our platforms.