

# **Facebook Digdata Challenge - Jasper's Designer Boutique**

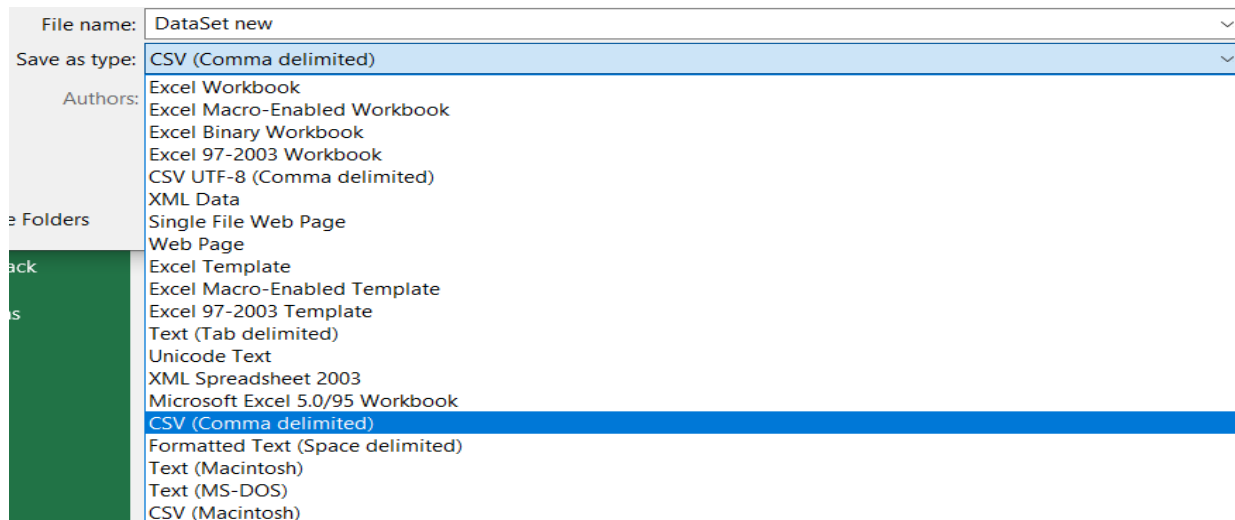
Submitted by- Jaden Clive Pinto

Before we put forward the observations and conclusions of the studies conducted it is important to define two terms, namely Test group and Control group. The test group consists of the participants of the study who are exposed to changes, in this case being those people who have seen the ads and the control group refers to the participants who have not seen the ads.

Now the significance of splitting the participants into two such groups allows us to measure how effective the ads are. This means that an ad in which the positive responses from the Control group are equal to or even greater than those from the Test group is an unsuccessful one.

The observations of the study are presented to us in an Excel File. This Excel file contains a large amount of data since multiple studies were conducted. We could manually look through this data and check and compare values from each and every record, but this would not be very efficient. Instead what we could do is to create a program which will automate these tasks for us. There are several ways through which this can be achieved but the one I will be using is CSV (Comma Separated Values) File Handling in Python. And the “csv” module which can be imported in python gives us the tools required to work with the Excel file.

Since we only need the observations of the survey we can modify the given “Data Set” file. We can get rid of all sheets other than “Dig Dataset” since this sheet contains the observations of the study and then save the file as a .csv file.



Now we create a new folder and save this .csv file and create a new python file in that folder. We can import the csv module in this python file and create functions to deal with and study data.

## The better advertising strategy- Standard Ads or influencers Ads

First to verify if the file is being read properly we can write a function called `printContents()` which basically lists out the contents of the file.

```
def printContents():  
    with open ('DataSet.csv','r',newline='') as f:  
        print("Contents of file")  
        ReadObj= csv.reader(f)  
        for line in ReadObj:  
            print(line)
```

The output upon calling `printContents()`:

```
Contents of file  
['Study_id', 'Cell_id', 'Brand within 3s', 'Sound Off', 'Format', 'Platform', 'Question Type', 'Test  
Positive %', 'Control Positive %']  
['1', 'Ads_only', '1', '1', 'Video', 'FB+IG', 'ad_recall', '15.4%', '8.7%']  
['2', 'Influencer', '1', '1', 'Video', 'FB+IG', 'ad_recall', '12.5%', '3.0%']  
['2', 'Ads_only', '1', '1', 'Video', 'FB+IG', 'ad_recall', '17.0%', '2.5%']  
['3', 'Influencer', '0', '0', 'Mix', 'FB+IG', 'ad_recall', '13.5%', '10.1%']  
['3', 'Ads_only', '0', '0', 'Mix', 'FB+IG', 'ad_recall', '18.4%', '10.0%']  
['4', 'Influencer', '1', '0', 'Video', 'FB+IG', 'ad_recall', '15.4%', '5.4%']  
['4', 'Ads_only', '1', '0', 'Video', 'FB+IG', 'ad_recall', '9.7%', '3.3%']  
['5', 'Influencer', '0', '1', 'Image', 'FB+IG', 'ad_recall', '14.5%', '3.5%']  
['5', 'Ads_only', '0', '1', 'Image', 'FB+IG', 'ad_recall', '17.0%', '4.7%']  
['6', 'Influencer', '1', '0', 'Image', 'FB+IG', 'ad_recall', '20.3%', '4.2%']  
['6', 'Ads_only', '1', '0', 'Image', 'FB+IG', 'ad_recall', '16.2%', '9.8%']  
['7', 'Influencer', '0', '1', 'Image', 'FB+IG', 'ad_recall', '14.9%', '11.2%']  
['7', 'Ads_only', '0', '1', 'Image', 'FB+IG', 'ad_recall', '17.1%', '9.8%']  
['8', 'Influencer', '1', '1', 'Mix', 'IG', 'ad_recall', '14.3%', '8.0%']  
['8', 'Ads_only', '1', '1', 'Mix', 'IG', 'ad_recall', '15.7%', '9.6%']  
['9', 'Influencer', '1', '1', 'Video', 'FB', 'ad_recall', '27.0%', '23.2%']  
['9', 'Ads_only', '1', '1', 'Video', 'FB', 'ad_recall', '27.2%', '23.3%']  
['10', 'Influencer', '0', '1', 'Video', 'FB+IG', 'ad_recall', '21.3%', '16.2%']  
['10', 'Ads_only', '0', '1', 'Video', 'FB+IG', 'ad_recall', '23.0%', '18.1%']  
['11', 'Influencer', '0', '1', 'Video', 'FB+IG', 'ad_recall', '24.8%', '20.2%']  
['11', 'Ads_only', '0', '1', 'Video', 'FB+IG', 'ad_recall', '24.1%', '19.4%']  
['12', 'Influencer', '1', '1', 'Image', 'FB+IG', 'ad_recall', '23.3%', '16.3%']
```

```
['12', 'Ads_only', '1', '1', 'Image', 'FB+IG', 'ad_recall', '22.9%', '14.7%']  
['13', 'Influencer', '1', '1', 'Image', 'FB+IG', 'ad_recall', '23.8%', '17.6%']  
['13', 'Ads_only', '1', '1', 'Image', 'FB+IG', 'ad_recall', '23.5%', '17.0%']  
['14', 'Influencer', '1', '1', 'Image', 'FB', 'ad_recall', '65.5%', '69.8%']  
['14', 'Ads_only', '1', '1', 'Image', 'FB', 'ad_recall', '49.1%', '65.2%']  
['15', 'Influencer', '1', '1', 'Image', 'FB+IG', 'ad_recall', '24.8%', '12.7%']  
['15', 'Ads_only', '1', '1', 'Image', 'FB+IG', 'ad_recall', '21.3%', '13.0%']  
['1', 'Ads_only', '1', '1', 'Video', 'FB+IG', 'brand_awareness', '59.6%', '40.2%']  
['2', 'Influencer', '1', '1', 'Video', 'FB+IG', 'brand_awareness', '6.4%', '4.1%']  
['2', 'Ads_only', '1', '1', 'Video', 'FB+IG', 'brand_awareness', '6.7%', '7.8%']  
['3', 'Influencer', '0', '0', 'Mix', 'FB+IG', 'brand_awareness', '9.4%', '5.5%']  
['3', 'Ads_only', '0', '0', 'Mix', 'FB+IG', 'brand_awareness', '11.8%', '11.2%']  
['4', 'Influencer', '1', '0', 'Video', 'FB+IG', 'brand_awareness', '7.8%', '7.2%']  
['4', 'Ads_only', '1', '0', 'Video', 'FB+IG', 'brand_awareness', '9.5%', '6.3%']  
['5', 'Influencer', '0', '1', 'Image', 'FB+IG', 'brand_awareness', '20.2%', '10.0%']  
['5', 'Ads_only', '0', '1', 'Image', 'FB+IG', 'brand_awareness', '20.4%', '14.0%']  
['6', 'Influencer', '1', '0', 'Image', 'FB+IG', 'brand_awareness', '30.4%', '27.0%']  
['6', 'Ads_only', '1', '0', 'Image', 'FB+IG', 'brand_awareness', '34.9%', '23.6%']  
['7', 'Influencer', '0', '1', 'Image', 'FB+IG', 'brand_awareness', '26.5%', '4.6%']  
['7', 'Ads_only', '0', '1', 'Image', 'FB+IG', 'brand_awareness', '6.3%', '3.0%']  
['8', 'Influencer', '1', '1', 'Mix', 'IG', 'brand_awareness', '21.0%', '21.7%']  
['8', 'Ads_only', '1', '1', 'Mix', 'IG', 'brand_awareness', '22.2%', '13.2%']  
['9', 'Influencer', '1', '1', 'Video', 'FB', 'brand_awareness', '19.0%', '10.6%']  
['9', 'Ads_only', '1', '1', 'Video', 'FB', 'brand_awareness', '17.1%', '12.1%']  
['10', 'Influencer', '0', '1', 'Video', 'FB+IG', 'brand_awareness', '51.0%', '75.2%']  
['10', 'Ads_only', '0', '1', 'Video', 'FB+IG', 'brand_awareness', '73.6%', '63.1%']  
['11', 'Influencer', '0', '1', 'Video', 'FB+IG', 'brand_awareness', '40.2%', '40.3%']
```

And so on....

This implies that the file is being read with no errors.

And before we start checking which strategy was better overall it is important to mention how many times each type of advertising strategy was used to ensure fairness of comparison. In order to do this we create a function called **CountEach()** which lets us know how many times the survey for Ads only or for ads by influencers was conducted.

```
def CountEach():  
    with open ('DataSet.csv','r',newline='') as f:  
        AdsCount=0  
        InfluencerCount=0  
        ReadObj= csv.reader(f)  
        for line in ReadObj:  
            if line[1]=='Ads_only':  
                AdsCount=AdsCount+1  
            elif line[1]=='Influencer':  
                InfluencerCount=InfluencerCount+1  
        print("Ads Count=",AdsCount)  
        print("Influencer Count=", InfluencerCount)
```

The output upon calling **CountEach()** –

```
Ads Count= 45  
Influencer Count= 42
```

Since these values are not the same it is therefore essential to use percentages relative to the total studies conducted in each case instead of simply using the total count to compare which was more successful.

Next, we create functions to check how many ads of each type are actually effective. As discussed before, one way of considering an ad to be successful is if the percentage of positive responses from the people who saw the social media ad is higher than for those who did not. To count the number of effect ads, we use functions **EffectiveAds()** and **EffectiveInfluencers()**. It is important to note that since the positive responses are stored as percentages, we can compare them by converting them into float data type (decimal numbers) and removing the percentage (%) sign.

```
def EffectiveAds():
    with open ('DataSet.csv','r',newline='') as f:
        EffectiveAdsCount=0
        ReadObj= csv.reader(f)
        for line in ReadObj:
            if line[1]=='Ads_only':
                if float(line[7].replace("%","")) > float(line[8].replace("%","")):
                    EffectiveAdsCount = EffectiveAdsCount+1
        print("The successful number of Ads only=",EffectiveAdsCount)

def EffectiveInfluencers():
    with open ('DataSet.csv','r',newline='') as f:
        EffectiveInfluencersCount=0
        ReadObj= csv.reader(f)
        for line in ReadObj:
            if line[1]=='Influencer':
                if float(line[7].replace("%","")) > float(line[8].replace("%","")):
                    EffectiveInfluencersCount = EffectiveInfluencersCount+1
        print("The successful number of ads by influencers=",EffectiveInfluencersCount)
```

Running the above functions, gives us output-

```
The successful number of Ads only= 33
The successful number of ads by influencers= 25
```

Finally, in order to check which advertising strategy is better, we can compare the percentage of effective advertisements relative to the total advertisements in each case. This is achieved using the function **ComparePercentage()**. Note that before creating the function, we need to make the previously defined variables global, which can be done using the keyword `global` followed by the name of the variable.

```
def ComparePercentage():  
    percent_ads_only = (EffectiveAdsCount/AdsCount) * 100  
    percent_influencers = (EffectiveInfluencersCount/InfluencerCount) * 100  
    print("Percentage of Effective Ads=", percent_ads_only, "%")  
    print("Percentage of Effective Ads by influencers=", percent_influencers, "%")  
    if percent_ads_only > percent_influencers:  
        print("Therefore Ads only is better and more effective")  
    else:  
        print("Therefore Ads by influencers are better and more effective")
```

Upon calling the function we get output-

```
Percentage of Effective Ads= 73.33333333333333 %  
Percentage of Effective Ads by influencers= 59.523809523809526 %  
Therefore Ads only is better and more effective
```

And hence we can conclude by saying that overall, the standard ads are more effective than ads by influencers.

## Brand Awareness- Standard Ads or influencers Ads

We can create out a new function called **BrandAwareness()** in order to compare the effective of Standard Ads and Influencer Ads and to see which one is more effective. This function is similar to those defined previously but it takes into consideration only those surveys which where conducted for Brand Awareness. This can be done if checking that the 7<sup>th</sup> element (having index value of 6) of the list which tells us the purpose of the survey is in fact Brand Awareness.

```
def BrandAwareness():
    with open ('DataSet.csv','r',newline='') as f:
        BrandAdsCount=0
        BrandInfluencerCount=0
        EffectiveBrandAdsCount=0
        EffectiveBrandInfluencerCount=0
        ReadObj= csv.reader(f)
        for line in ReadObj:
            if line[6] == 'brand_awareness':
                if line[1]=='Ads_only':
                    BrandAdsCount=BrandAdsCount+1
                    if float(line[7].replace("%","")) > float(line[8].replace("%","")):
                        EffectiveBrandAdsCount= EffectiveBrandAdsCount+1
                elif line[1]=='Influencer':
                    BrandInfluencerCount=BrandInfluencerCount+1
                    if float(line[7].replace("%","")) > float(line[8].replace("%","")):
                        EffectiveBrandInfluencerCount= EffectiveBrandInfluencerCount+1
        print("Ads Count=",BrandAdsCount)
        print("Influencer Count=", BrandInfluencerCount)
        print("Number of effective standard ads=", EffectiveBrandAdsCount)
        print("Number of effective influencer ads=", EffectiveBrandInfluencerCount)
        percent_brand_ads_only= (EffectiveBrandAdsCount/BrandAdsCount) * 100
        percent_brand_influencer= (EffectiveBrandInfluencerCount/BrandInfluencerCount) * 100
        print("Percentage of effective standard ads=", percent_brand_ads_only,"%")
        print("Percentage of effective influencer ads=", percent_brand_influencer,"%")
        if percent_brand_ads_only > percent_brand_influencer:
            print("Therefore Standard Ads better are better for Brand Awareness")
        else:
            print("Therefore Influencer Ads better are better for Brand Awareness")

BrandAwareness()
```

Running the above code gives us output-

```
Ads Count= 15  
Influencer Count= 14  
Number of effective standard ads= 12  
Number of effective influencer ads= 7  
Percentage of effective standard ads= 80.0 %  
Percentage of effective influencer ads= 50.0 %  
Therefore Standard Ads better are better for Brand Awareness
```

And hence we can conclude by saying that in terms of ads whose purpose is Brand Awareness, the standard ads are more effective than ads by influencers.



To look into Brand Awareness more, we can check which format of the ad is more effective- Images or Videos

In order to do this we create the function **BrandAwarenessMedia()** which essential goes through the entries of all surveys and compares ads that were presented in the form of a video or an image to check which one of the two was more effective for brand awareness.

```
def BrandAwarenessMedia():
    with open ('DataSet.csv','r',newline='') as f:
        Image=0
        Video=0
        EffectiveImage=0
        EffectiveVideo=0
        ReadObj= csv.reader(f)
        for line in ReadObj:
            if line[6] == 'brand_awareness':
                if line[4]=='Image':
                    Image=Image+1
                    if float(line[7].replace("%","")) > float(line[8].replace("%","")):
                        EffectiveImage= EffectiveImage+1
                elif line[4]=='Video':
                    Video=Video+1
                    if float(line[7].replace("%","")) > float(line[8].replace("%","")):
                        EffectiveVideo= EffectiveVideo+1
        print("Number of image ads=",Image)
        print("Number of video ads=", Video)
        print("Number of effective iamge ads=", EffectiveImage)
        print("Number of effective video ads=", EffectiveVideo)
        percent_image= (EffectiveImage/Image) * 100
        percent_video= (EffectiveVideo/Video) * 100
        print("Percentage of effective image ads=", percent_image,"%")
        print("Percentage of effective influencer ads=", percent_video,"%")
        if percent_image > percent_video:
            print("Therefore Ads with images are better for Brand Awareness")
        else:
            print("Therefore Ads with videos are better for Brand Awareness")

BrandAwarenessMedia()
```

Running the above code gives us output-

```
Number of image ads= 14  
Number of video ads= 11  
Number of effective iamage ads= 9  
Number of effective video ads= 7  
Percentage of effective image ads= 64.28571428571429 %  
Percentage of effective influencer ads= 63.63636363636363 %  
Therefore Ads with images are better for Brand Awareness
```

Technically, the ads presented in the form of an image were more effective than ones presented as videos but the difference is by a very small margin and hence is negligible.

In conclusion, as far as Brand Awareness goes, the success of advertisements was independent of their format.

Furthermore, we can now explore how the brand of the advertising company being presented in the first 3 seconds affected the success of the advertisement in terms of brand awareness.

This is done using the function **BrandWithinThree()** which iterates through all the records in the CSV file containing the observations of the survey in which the participants were asked about Brand Awareness. The function informs us how brand placement within the first few seconds on an advertisement can affect the effectiveness of the advertisement.

```
def BrandWithinThree():
    with open ('DataSet.csv','r',newline='') as f:
        InThree=0
        NotInThree=0
        EffectiveInThree=0
        EffectiveNotInThree=0
        ReadObj= csv.reader(f)
        for line in ReadObj:
            if line[6] == 'brand_awareness':
                if line[2]=='1':
                    InThree=InThree+1
                    if float(line[7].replace("%","")) > float(line[8].replace("%","")):
                        EffectiveInThree= EffectiveInThree+1
                elif line[2]=='0':
                    NotInThree=NotInThree+1
                    if float(line[7].replace("%","")) > float(line[8].replace("%","")):
                        EffectiveNotInThree= EffectiveNotInThree+1
        print("Number of ads in which brand was shown within 3s=",InThree)
        print("Number of ads in which brand was shown after 3s=", NotInThree)
        print("Number of effective ads in which brand was shown within 3s=", EffectiveInThree)
        print("Number of effective ads in which brand was shown after 3s=", EffectiveNotInThree)
        percent_in_three= (EffectiveInThree/InThree) * 100
        percent_after_three= (EffectiveNotInThree/NotInThree) * 100
        print("Percentage of effective ads in which brand was shown within 3s=", percent_in_three,"%")
        print("Percentage of effective ads in which brand was shown after 3s=", percent_after_three,"%")
        if percent_in_three > percent_after_three:
            print("Therefore Ads in which brand was shown within 3s are better for Brand Awareness")
        else:
            print("Therefore Ads in which brand was shown after 3s are better for Brand Awareness")

BrandWithinThree()
```

Upon running the above code, we get the following output-

```
Number of ads in which brand was shown within 3s= 19  
Number of ads in which brand was shown after 3s= 10  
Number of effective ads in which brand was shown within 3s= 12  
Number of effective ads in which brand was shown after 3s= 7  
Percentage of effective ads in which brand was shown within 3s= 63.1578947368421 %  
Percentage of effective ads in which brand was shown after 3s= 70.0 %  
Therefore Ads in which brand was shown after 3s are better for Brand Awareness
```

Hence, we see that advertisements in which the brand of the company was presented to the viewer after the first three seconds was more effective for brand awareness compared to those in which the brand was presented within the first three seconds.

## What can be done to improve advertising in the future?

We could conduct more surveys to take more factors that affect the effectivity of the advertisement into consideration, some of which could be-

- Presence of text and numbers on the ads – do they distract the viewer from the branding or make them more engaged with the ad?
- Colours used in the ad- did users prefer lighter or darker theme of colours for in the ad? Did any one particular colour capture the attention of users more than the others?
- Containing the contact details of the brand such as a website, phone number, email address etc – did the users find these details of any use?

Aside from conducting surveys, another method of advertisements would be through email campaigns. Facebook could start an email campaign for brands, where the customers would have the option to join the mailing list for a company through Facebook. The marketing emails are sent to the email account which the customer uses for Facebook.

An email campaign is the easiest way to engage with an audience since there is no algorithm we have to deal with. Anybody who signs up to the mailing list gets an email every time.