

REPORT

Submitted to: Prof. Tareq Jaber

Report Prepared by: Jeeva Mary Loui

Date: 7th April 2019

INTRODUCTION

This report summarized the statistical modeling and analysis results associated with a survey held on Social Influence on shopping. This data was collected by a social survey mobile platform Whatsgoodly. The survey question is ‘What social platform has influenced your online shopping most?’

DATASET

Dataset contains survey response in 1450 rows with six columns. The question is the first column which questions ‘What social platform has influenced your online shopping most?’. The following columns are the description of the segment of people who responds to the survey and their answers. The answers are chosen from 4 popular social platforms and a None option. The count and percentage of corresponding responds within a segment are also recorded.

The participants of the survey are wide and clear, which helps to analyse and understand the influence of social platform in consumerism. The data has reports from school going children to global results.

Data cleaning and organizing .

The dataset is a csv file and it has to be imported to python editor. The required python packages to manipulate and produce reasonable analysis are imported. The packages import are ‘os’, ‘pandas’, ‘NumPy’, ‘matplotlib’,.

The dataset is read with pandas read_csv function. The data has column called ‘Segment Type’ which has unique segments , and this segments are converted to a NumPy series through unique() function of pandas package. Subdataframes are made through for loop .The responses are classified to Facebook, Instagram , Twitter , Snapchat , and None-who claims they are not

influenced by social networking platforms in their shopping. To get more insight about the dataset Subdataframes with Answers as the index is also created out from the main data frame.

Analysis of global responses from mobile users.

To begin with the analysis the global results are always preferred .The code used is shown below along with the bar chart .

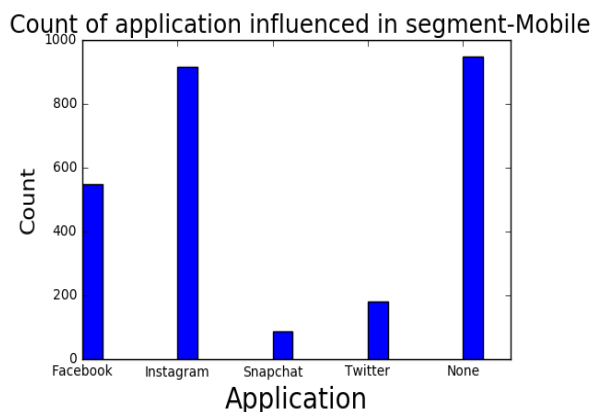
The chart explains that globally most people claims they are not at all influenced by social networks in their shopping. And Instagram is the first influencer globally among the social networks.

```
pos = np.arange(len(app))
count=dfMobile.Count.tolist()
help(plt.bar)
plt.bar(pos,count,color='blue',edgecolor='black',width=.21)
plt.xticks(pos,appl)
plt.xlabel('Application', fontsize=20)
plt.ylabel('Count', fontsize=16)
plt.title(' Count of application influenced in segment-
Mobile',fontsize=18)
plt.show()
```

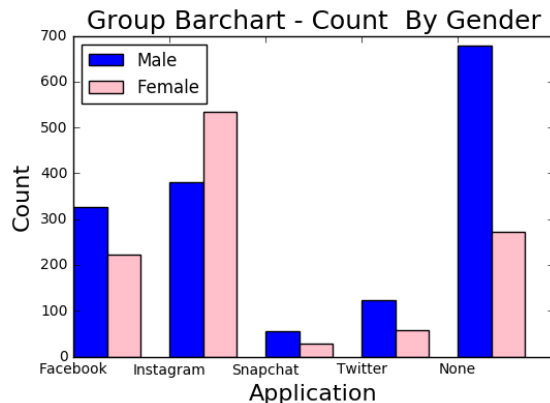
Gender Analysis.

Its is interesting to find the difference in level of influence among men and women are by online platforms and advertisements. The bar chart is best way to find this comparison.The following is well coded to get the analysis result.

```
plt.bar(pos,Count_Male,bar_width,color='blue',edgecolor='black')
```



```
plt.bar(pos+bar_width,Count_Female,bar_width,color='pink',edgecolor='black')
plt.xticks(pos, appl)
plt.xlabel('Application', fontsize=16)
plt.ylabel('Count', fontsize=16)
plt.title('Group Barchart - Count By Gender',fontsize=18)
plt.legend(['Male','Female'],loc=2)
plt.show()
```



The plt.bar is a matplotlib package , the ‘pos’ has the value required for the position of bars in the x axis and its determined by the length of list which contains the answers.Two bars are required in different colours for the comparison and well coded to get the output.

From the Group Bar chart it can be concluded that males are the most whom are not influenced by any applications also the ranks first by level of getting influenced except by Instagram , women stalks Instagram a lot.

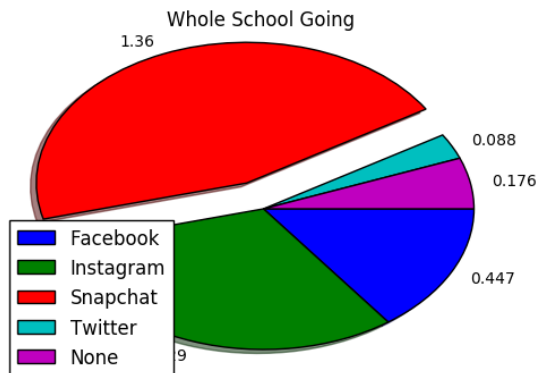
Influence on School Going Section of Society

It is interesting to understand how school going people are influenced by social networks and their least and most influencer online can be found using the pie charts . The following code is run to get the following pie charts which are so simple and easily interpretable.

```
TotalSc=School.groupby(School.Answer).sum()
val=TotalSc.Percentage

colors = ['b', 'g', 'r', 'c', 'm', 'y']
labels = appl
explode = (0, 0, 0.2, 0, 0)
plt.pie(val, colors=colors, labels= val,explode=explode,
        counterclock=False, shadow=True)
```

```
plt.title('Whole School Going ')
plt.legend(labels,loc=3)
plt.show()
```



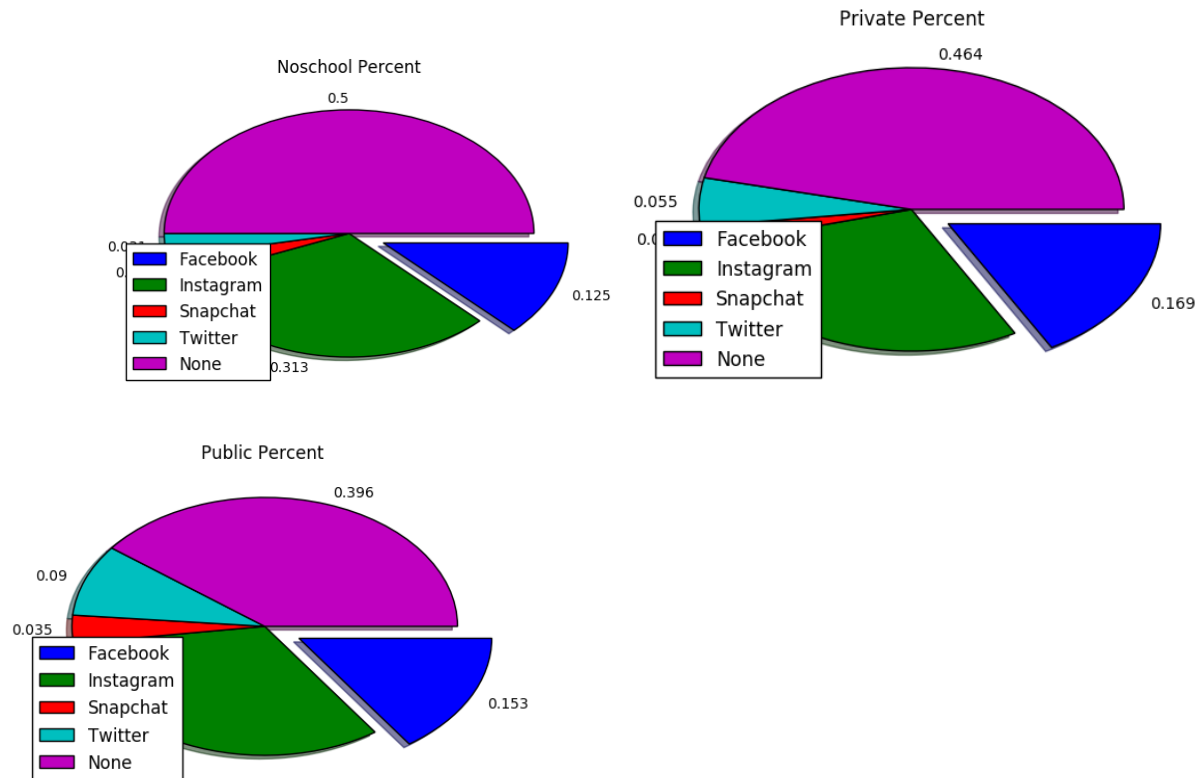
It can be concluded that Snapchat is so popular and biggest influencer to the young who are doing their schooling.

Influence among the different categories in Schooling Section

The schooling section has three different categories: students who do their schooling without going to school, students who are enrolled in public schools and private school.

```
S=School['Segment Description'].unique()

for x in S:
    globals()[('%s' % x).replace("or private school?", "").replace(" ", "")] = pd.DataFrame(data=(School[School['Segment Description']== x]))
    Sc=['Noschool','Private','Public']
    schoolf= {
        'Noschool Count':Noschool.Count.tolist(),'Noschool
        Percent':Noschool.Percentage.tolist(),
        'Private Count':Private.Count.tolist(),'Private
        Percent':Private.Percentage.tolist(),'Public Percent
        ':Public.Percentage.tolist(),}
    schoodf=pd.DataFrame(schoolf,index=appl)
    for x in schoodf:
        if 'Percent' in x:
            j=x
            values=schoodf[j]
            colors = ['b', 'g', 'r', 'c', 'm', 'y']
            labels = appl
            explode = (0.2, 0, 0, 0, 0)
            plt.pie(values, colors=colors, labels= values,explode=explode,
                    counterclock=False, shadow=True)
            plt.title(x)
            plt.legend(labels,loc=3)
            plt.show()
```



For private and No-school categories influence is less as 45 to 50 percent of them chose they are not influenced in their shopping habits by social networks. In all the cases Instagram influence them the most through internet and snapchat the least. This is a contradiction to whole school going category .

Students with and without loan debt

For better social and economic analysis it is the best to know the difference in the level the people with education debt and without it are influenced in their shopping habits and preferences through social platforms.

The following code produces better insight which gives group bar chart.

```
L=Loan['Segment Description'].unique()
for x in L:
    if 'Yes' in x:
        j=x
```

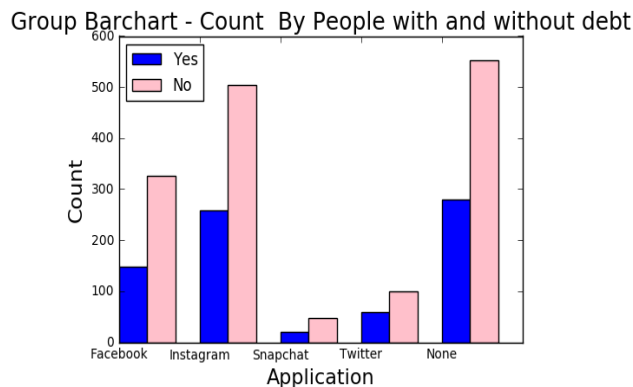
```

globals()['yes' ] = pd.DataFrame(data=(Loan[Loan['Segment
Description']== j]))
else:
    globals()['no' ] = pd.DataFrame(data=(Loan[Loan['Segment
Description']== x]))

yes=yes.set_index('Answer')
yes=yes.drop('Segment Description',axis=1)
no=no.drop('Segment Description',axis=1)
yesc= yes.Count.tolist()
no=no.set_index('Answer')
noc=no.Count.tolist()

pos=pos = np.arange(len(app))
plt.bar(pos,yesc,bar_width,color='blue',edgecolor='black')
plt.bar(pos+bar_width,noc,bar_width,color='pink',edgecolor='black')
plt.xticks(pos, app)
plt.xlabel('Application', fontsize=16)
plt.ylabel('Count', fontsize=16)
plt.title('Group Barchart - Count By Gender',fontsize=18)
plt.legend(['Yes','No'],loc=2)
plt.show()

```



Analysis based on zodiac signs

Finding out how people with different zodiac sign is influenced by zodiac sign is interesting. From the data pie charts are used to get the analysis results

```

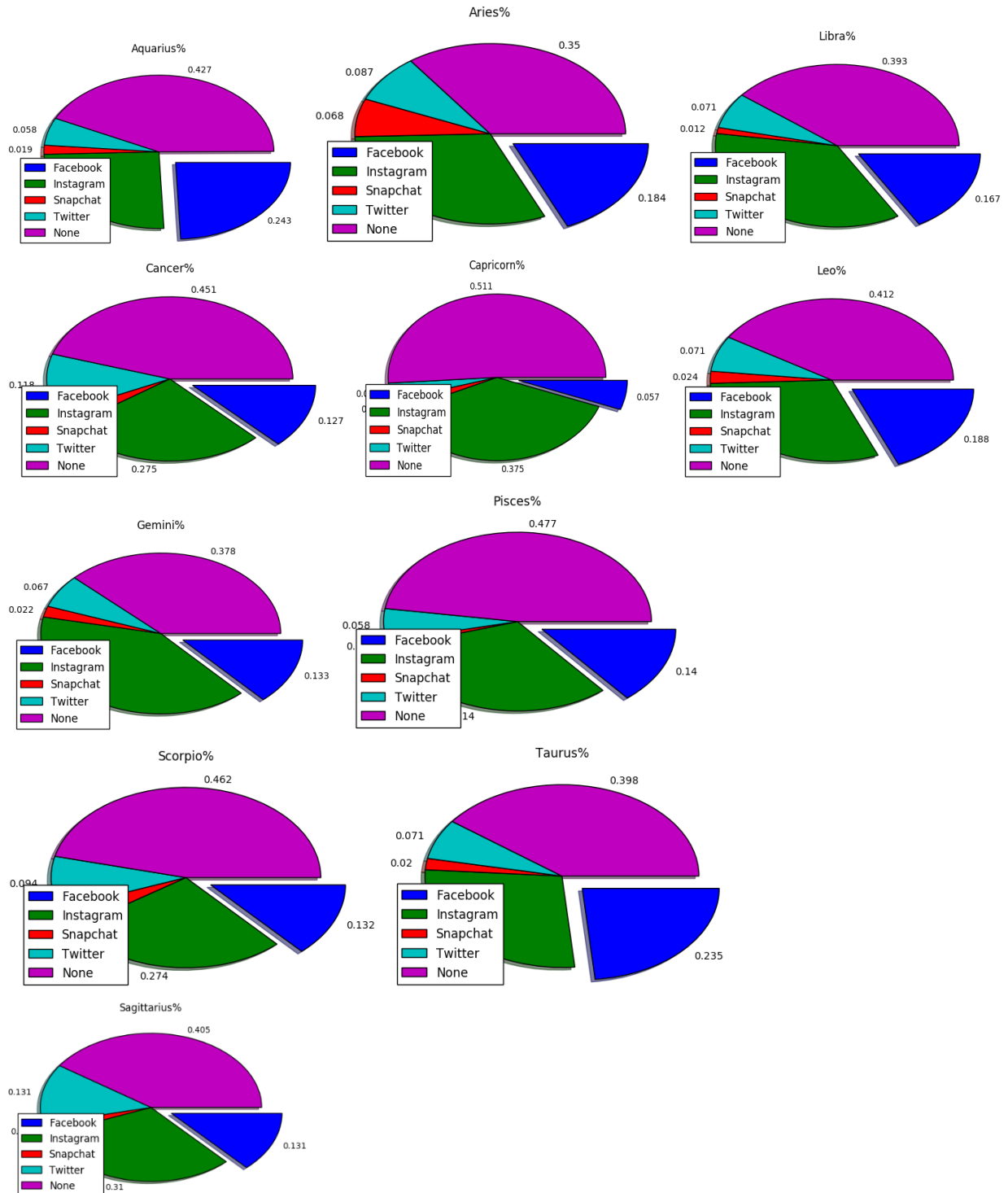
Z=Zodiac['Segment Description'].unique()
for x in Z:
    globals()['df%s' % x] = pd.DataFrame(data=(Zodiac[Zodiac['Segment
Description']== x]))
z=['Capricorn%', 'Sagittarius%', 'Scorpio%', 'Libra%', 'Virgo', 'Leo%', 'Cancer%',
  'Gemini%', 'Taurus%', 'Aries%', 'Pisces%', 'Aquarius%']
Zdf={'Capricorn':Capricorn.Count.tolist(), 'Capricorn%':Capricorn.Percentag
e.tolist(),

```

```

'Sagittarius
Count':Sagittarius.Count.tolist(),'Sagittarius%':Sagittarius.Percentage.to
list(),
'Scorpio
Count':Scorpio.Count.tolist(),'Scorpio%':Scorpio.Percentage.tolist(),
'Libra Count':Libra.Count.tolist(),'Libra%':Libra.Percentage.tolist(),
'Leo':Leo.Count.tolist(),'Leo%':Leo.Percentage.tolist(),
'Cancer
Count':Cancer.Count.tolist(),'Cancer%':Cancer.Percentage.tolist(),
'Gemini
Count':Gemini.Count.tolist(),'Gemini%':Gemini.Percentage.tolist(),
'Taurus
Count':Taurus.Count.tolist(),'Taurus%':Taurus.Percentage.tolist(),
'Aquarius
Count':Aquarius.Count.tolist(),'Aquarius%':Aquarius.Percentage.tolist(),
'Aries Count':Aries.Count.tolist(),'Aries%':Aries.Percentage.tolist(),
'Pisces
Count' :Pisces.Count.tolist(),'Pisces%':Pisces.Percentage.tolist(),}
zdata=pd.DataFrame(Zdf,index=appl)
zdata.info
maxiumcount=zdata.idxmax(axis=1)
values=[]
print('Applications where most used by the zodiac group:\n' ,maxiumcount)
type(x)
type(values)
"""PIE CHART SHOWS HOW PEOPLE WITH DIFFERENT ZODIAC SIGNS GET INFLUENCED
BY DIFFERENT APPLICATION"""
for x in zdata:
    if '%' in x:
        j=x
        values=zdata[j]
        colors = ['b', 'g', 'r', 'c', 'm', 'y']
        labels = appl
        explode = (0.2, 0, 0, 0, 0)
        plt.pie(values, colors=colors, labels= values,explode=explode,
                counterclock=False, shadow=True)
        plt.title(x)
        plt.legend(labels,loc=3)
        plt.show()

```

With different pie chart of zodiac sign broadly it can be concluded that influence in shopping has not much relation with the zodiac sign

Source

<https://data.world/ahalps/social-influence-on-shopping>