MARKETING DATA EXPLORATORY DATA ANALYSIS

TITLE: HNG STAGE 2 TASK

SLACK ID: COBIRD

INTRODUCTION

ABOUT THE TASK

- Marketing data on different campaign methods and strategy were collected
- Objective is to carry out exploratory data analysis
- To identify useful insights from the data
- Make data-supported recommendations to improve various aspects of the marketing campaign
- Python programing language was used for the analysis

ABOUT THE DATA

- Data used for the task was from here
- Using python, data contained 20,005 rows and 15 columns
- 6 of the columns were numerical and useful while 8 were categorical
- The columns contained no null values

```
df.shape
```

```
(200005, 15)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200005 entries, 0 to 200004
Data columns (total 15 columns):
 #   Column            Non-Null Count   Dtype  
---  ------            --------------   -----  
 0   Campaign_ID       200005 non-null  int64  
 1   Company           200005 non-null  object 
 2   Campaign_Type     200005 non-null  object 
 3   Target_Audience   200005 non-null  object 
 4   Duration          200005 non-null  object 
 5   Channel_Used      200005 non-null  object 
 6   Conversion_Rate   200005 non-null  float64
 7   Acquisition_Cost  200005 non-null  object 
 8   ROI               200005 non-null  float64
 9   Location          200005 non-null  object 
 10  Date              200005 non-null  object 
 11  Clicks            200005 non-null  int64  
 12  Impressions       200005 non-null  int64  
 13  Engagement_Score  200005 non-null  int64  
 14  Customer_Segment  200005 non-null  object 
dtypes: float64(2), int64(4), object(9)
memory usage: 22.9+ MB
```

FINDINGS & INSIGHTS

FINDINGS 1

- Five types of audience, six categories of channel types, and five types of campaign were deplored for this project

	Channel	Counts	Percentage_Count(%)
0	Email	33599	16.80
1	Google Ads	33440	16.72
2	YouTube	33393	16.70
3	Instagram	33392	16.70
4	Website	33361	16.68

Campaign_Type	Counts	Percentage (%)
Influencer	40170	20.08
Search	40157	20.08
Display	39988	19.99
Email	39871	19.94
Social Media	39819	19.91

Target_Audience	Counts	Percentage (%)
Men 18-24	40259	20.13
Men 25-34	40024	20.01
All Ages	40021	20.01
Women 25-34	40013	20.01
Women 35-44	39688	19.84

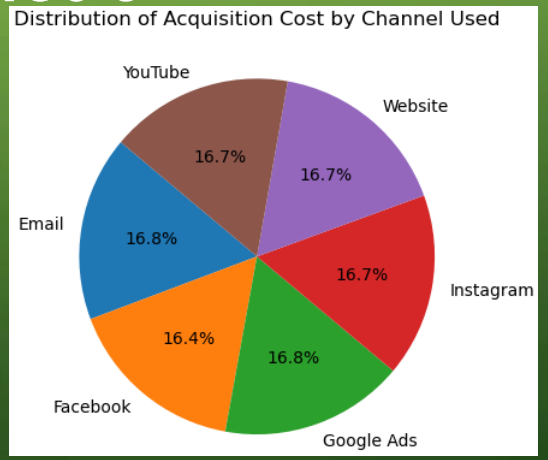
FINDINGS 2

- Email and Website media channels were the most effective channel while YouTube and Instagram were least effective.
- Facebook and Website gave highest ROI while Instagram and YouTube yielded least ROI

	Channel	Mean_Conversion_Rate	Conversion Rank	Mean_ROI_Rate	ROI Rank
0	Email	0.080282	1	4.996487	4
1	Website	0.080182	2	5.014114	2
2	Google Ads	0.080181	3	5.003126	3
3	Facebook	0.079990	4	5.018672	1
4	YouTube	0.079890	5	4.993720	5
5	Instagram	0.079886	6	4.988706	6

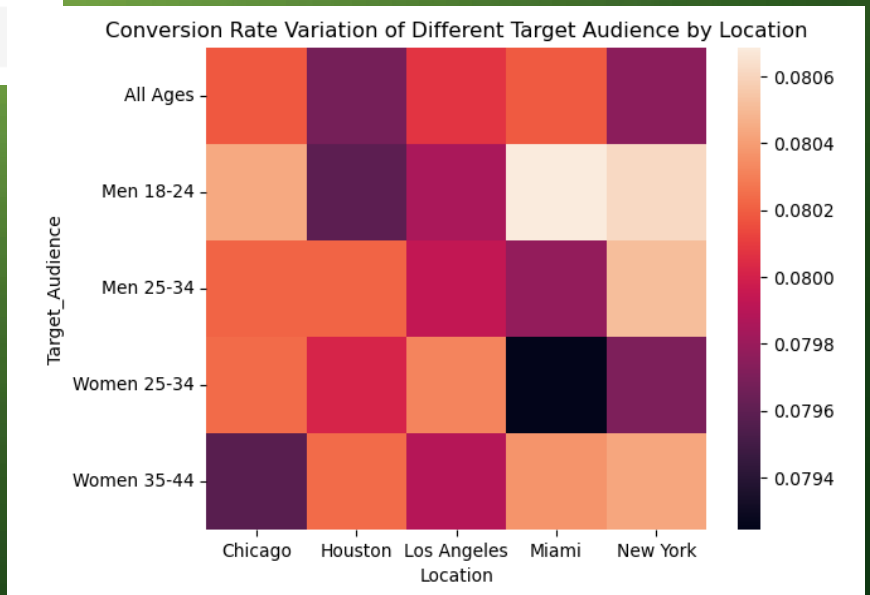
FINDINGS 3

- No significant difference in marketing cost across the different channel (see pie chart)



FINDINGS 4

- The preference of a campaign method does not differ significantly by location or targeted audience
- Campaigns to men aged 18 – 24 years gave the highest conversion rate in New York, Miami, and Chicago), while women aged 25-34 years had the highest conversion rate in Los Angeles



CODE & METHODOLOGY

METHODOLOGY

- Python programming language was used for this data exploration
- Pandas and NumPy libraries were used for the data manipulation
- Seaborn and Matplotlib were used for visualizations
- Jupyter Notebook provided the IDE for code development

CODES

- To find the number of unique channels and their percentages

```
Channel = df['Channel_Used'].value_counts()
# Find the percentages of the result in a percent seriers and round it to 2 decimals
Percent = 100*Channel/sum(Channel)
Percent = Percent.round(2)
Percent

# Combine both results into a dataframe via a dictyionary and check to see the output
Unique_Channels = pd.DataFrame({'Channel': Channel.index,
                    'Counts': Channel.values, 'Percentage_Count(%)': Percent.values})
Unique_Channels.head()
```

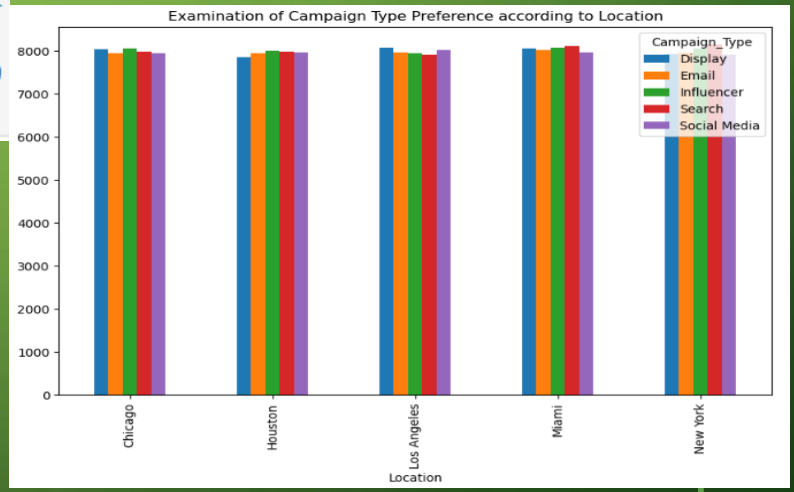
- Explore how channel type influences campaign effectiveness and ROI

```
import numpy as np
bb = (df.groupby('Channel_Used')['Conversion_Rate']
      .mean()).sort_values(ascending = False)
ConverRate_Channel = pd.DataFrame({'Channel': bb.index,
        'Mean_Conversion_Rate': bb.values,'Conversion Rank':[1,2,3,4,5,6]})
ConverRate_Channel

bROI = (df.groupby('Channel_Used')['ROI'].
        mean()).sort_values(ascending = False)
ROI_Channel = pd.DataFrame({'Channel': bROI.index,
            'Mean_ROI_Rate': bROI.values,'ROI Rank':[1,2,3,4,5,6]})
ROI_Channel

Channel_Explored = pd.merge(ConverRate_Channel, ROI_Channel,
                            on = 'Channel', how = 'inner')
Channel_Explored
```

CODES ON VISUALS

- To examine location influence on campaign type preference

```
import matplotlib.pyplot as plt
pd.crosstab(df['Location'], df['Campaign_Type']).
        plot(kind = 'bar', figsize = (10,6)
            )
plt.title ('Examination of Campaign Type
            Preference according to Location'
            )
plt.show
```



- Conversion Rate Varies with Target Audience by Location

```
pivot = df.pivot_table(index = 'Target_Audience', columns = 'Location',
                       values = 'Conversion_Rate', aggfunc = 'mean')
sns.heatmap(pivot, annot= False)
plt.title('Conversion Rate Variation of Different Target Audience by Location')
plt.show()
```

Code for Pie Chart

```
Pie_values = df.groupby('Channel_Used')['Acqu_Cost_Value'].sum()
plt.pie(Pie_values, labels =Pie_values.index, autopct = '%1.1f%%',
        startangle = 140, colors = plt.cm.tab10.colors)
plt.title ('Distribution of Acquisition Cost by Channel Used')
plt.show()
```

RECOMMENDATIONS

- Management should consider scrapping the use of YouTube and Instagram since both give both the lowest Conversion Rate and ROI of all channel types
- Rather still, marketing teams handling this two channels can be re-trained for more effectiveness
- This is especially so since the acquisition / marketing cost spent on all the channels are almost equal as seen from the pie chart
- Management should prioritize use of website due to having 2nd high Conversion Rate, 2nd highest ROI, and 2nd lowest acquisition cost
- Marketing materials and team should target audiences aged 18 – 24 more since they generally provide more ROI across different location
- Women aged 25-34 generally have the lowest conversion rate, especially in Miami. Since this market is huge, the team handling this population segment should be retrained and the marketing materials reviewed for improved marketing effectiveness

CONCLUSIONS

- The data mining and gathering process of this dataset is very commendable
- Further analysis needs to be done to explore how other variables influences the effectiveness of the campaign
- This analysis did not consider the duration of the campaign. This is a very vital variable that can influence marketing campaign's effectiveness and generated engagements