**EXECUTIVE SUMMARY**

In response to the evolving landscape of advertising and consumer outreach, our aviation company is embarking on a transformative project in collaboration with a leading social networking platform. The objective is clear: shift from traditional tele-calling methods to data-driven advertising strategies that not only optimize ad spend but also ensure precision targeting.

In a world where mobile and laptop usage is ubiquitous, our project acknowledges the diverse preferences of our customer base. To this end, we are developing two distinct models to cater to mobile and laptop users. These models will enable us to efficiently tailor digital advertisements to individuals with a high likelihood of purchasing our products.

In a world where data is the new currency, Aviation Company is poised to leverage this valuable resource to revolutionize advertising strategies.

**Introduction to the problem statement**

An aviation company that provides domestic as well as international trips to the customers now wants to apply a targeted approach instead of reaching out to each of the customers. This time they want to do it digitally instead of telecalling. Hence they have collaborated with a social networking platform, so they can learn the digital and social behavior of the customers and provide the digital advertisement on the user page of the targeted customers who have a high propensity to take up the product. Propensity of buying tickets is different for different login devices. Hence, you have to create 2 models separately for Laptop and Mobile. Anything which is not a laptop can be considered as mobile phone usage. The advertisements on the digital platform are a bit expensive; hence, you need to be very accurate while creating the models.

* 1. **Understanding and need of Project/Study**

In our day to day life we spend most of the time on social media browsing different websites, online shopping and various pages for entertainment, which simply proves that most of our free time we like to invest on internet. All the brands promote their products on social media as no other platform is as booming as social media as it has become part of nearly everyone's daily routine.

Predictive analytics is going to become increasingly popular - by analyzing the big data from social media, companies will be able to identify the hallmarks of a customer who is about to cease business with said company, and take efforts to correct it. They can also identify common behaviors between different customers to see what makes them more likely to become a power user or be completely on board with the company's ideals. They can then use this information to convert social media users into customers of their business. So here we have problem statement related to aviation industry, we will analyze the given data using python and will perform univariate, bivariate analysis and EDA. In this project we will try to predict and analyze the digital and social behavior of the customers and provide the digital advertisement on the user page of the targeted customers who have a high propensity to take up the product.

* 1. **Understanding business/social opportunity**

The collaboration between our aviation company and a prominent social networking platform presents an exciting business and social opportunity at the intersection of technology and personalized marketing. In a departure from traditional tele-calling methods, we are embarking on a journey to harness the power of digital advertising, offering a more tailored and efficient approach to our customers. By tapping into the digital and social behavior of our customer base, we aim to provide targeted advertisements that cater to their unique preferences and propensities.

This initiative is rooted in the recognition that the propensity for purchasing airline tickets can vary significantly based on the user's login device, distinguishing between laptops and mobile devices. By developing two distinct models for these categories, we can ensure that our advertising strategies are highly precise, delivering the right message to the right audience. This precision is of paramount importance, particularly in a digital advertising landscape where every ad spend must be optimized to yield the highest possible return on investment.

* 1. **Company and industry overview**

Aviation Company is a prominent player in the aviation industry, known for its high-quality services. The aviation sector is increasingly competitive, prompting Just Aviation Company to embrace data-driven advertising to maintain its market position and adapt to evolving customer behavior.

* 1. **Overview of Theoretical Concepts**

The project leverages data science techniques such as machine learning, customer segmentation, and predictive modeling. These methodologies will assist in constructing models that can accurately identify the customer groups with the highest likelihood of conversion, enabling Just Aviation Company to optimize its advertising campaigns.

**RESEARCH METHODOLOGY**

**2.1 Scope of the Study**

The scope of the study encompasses a comprehensive analysis of data-driven advertising strategies within the aviation industry.

**2.2 Methodology**

**2.2.1 Data Collection**

The data collection process is already done by the aviation company. They collaborated with the social media company to get the data of customers reaction to the aviation company.

**2.2.2 EDA**

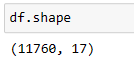
**I . Information of the Data**

A screenshot of a computer program

Description automatically generated

**OBSERVATION:** The dataset has a total of 17 features and missing values in some of the columns.

**II. Data shape:**



**III. Features having Missing values:**

A screen shot of a computer program

Description automatically generated

**OBSERVATION**: Here the features and its missing value count is displayed. We will try to handle this missing value using the right technique in the Exploratory data analysis part.

**IV. Data manipulation:**

A close up of a computer

Description automatically generated

**CONCLUSION:** The feature preferred device has so many values, but we need only mobile and laptop, so we change the values in the feature accordingly to only mobile and laptop.

A close-up of a number

Description automatically generated

**CONCLUSION:**The feature ‘yearly\_avg\_outstation\_checkins’ has some unwanted values such as ‘\*’, so we are removing the unwanted symbols from all the features.

So, likewise we changed the values in the features which had ‘three’ instead of ‘3’.

A screenshot of a computer

Description automatically generated

**CONCLUSION:**The feature ‘preferred\_location\_type’ has Tour and Travel and Tour Travel with two spaces inbetween, so we change this to general formal Tour and Travel.

**V . Outlier Removal:**

Below Boxplot shows the outliers in all of the features

A graph with a line graph

Description automatically generated with medium confidenceA graph of a bar graph

Description automatically generated

A graph with a blue rectangle

Description automatically generatedA graph of a bar graph

Description automatically generated with medium confidence

A graph of a graph

Description automatically generated

**OBSERVATION:**

*‘total\_likes\_on\_outstation\_checkin\_given’,*

*’Yearly\_avg\_comment\_on\_travel\_page’,*

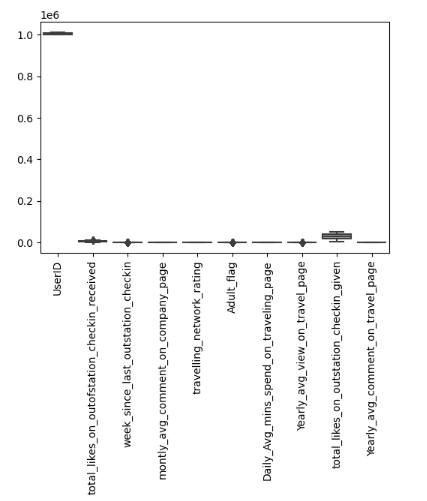
*’total\_likes\_on\_outstation\_checkin\_received’,*

*’monthly\_avg\_comment\_on\_company\_page’,*

*’daily\_avg\_mins\_spend\_on\_travelling\_page’.*

These features has outliers that we can observe from the above visualizations. So, we will remove all the outliers from all the features using Inter Quartile Range .

**After treating outliers:**



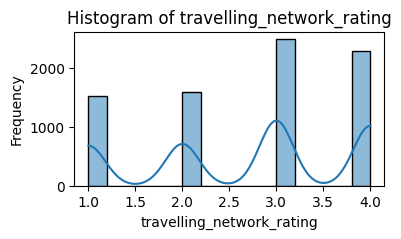
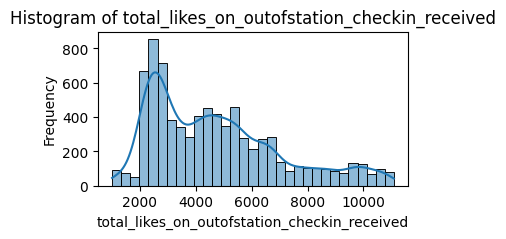
This is how our data looks like after treating outliers. Now, the shape of data is (7138, 17)

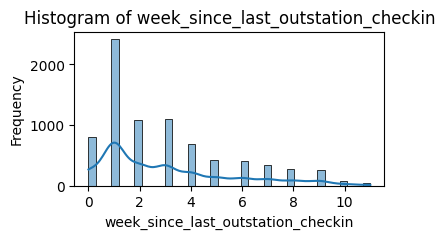
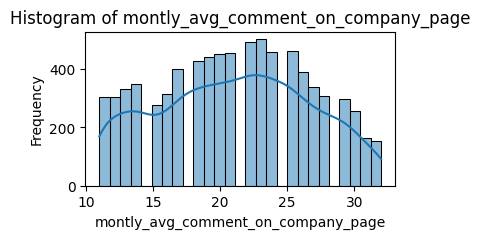
Before treating outlier, our data shape was (11498, 17).

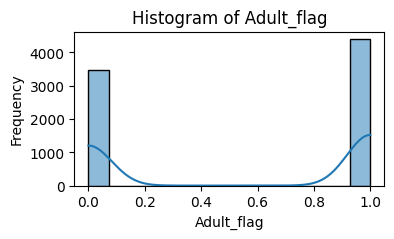
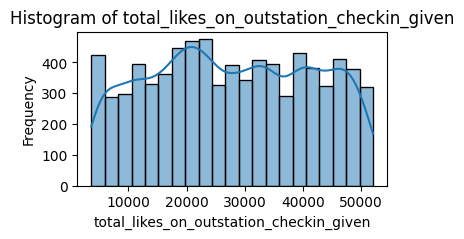
**VI. Data Visualization : Univariate Analysis**

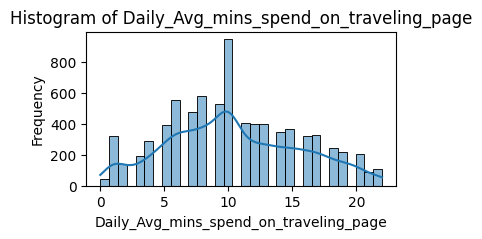
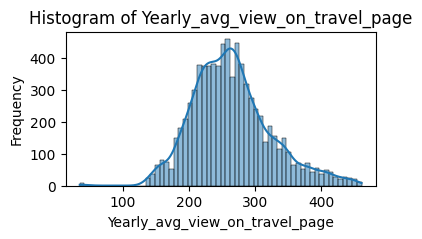
Now, we will try to explore all the features one by one

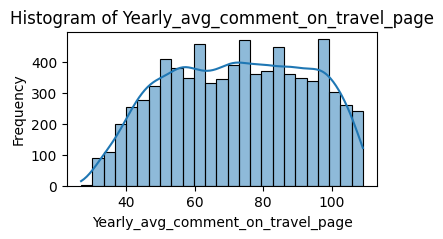
1. **Numerical Data**





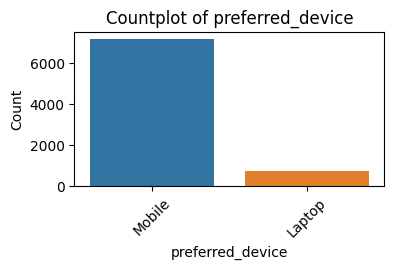
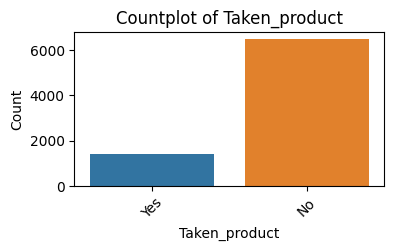


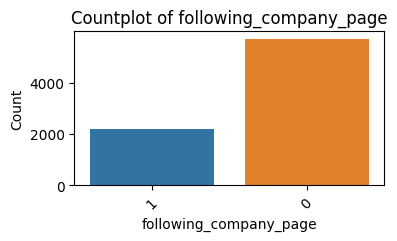
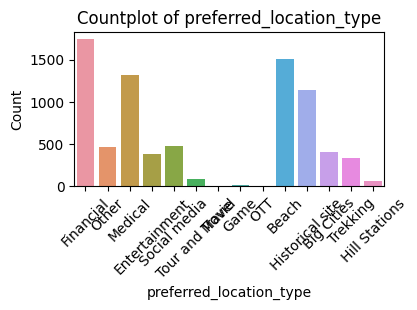


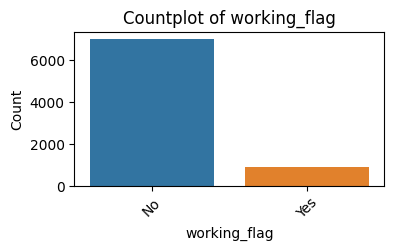


**Categorical Data:**

In case of categorical variable we are interested to know the frequencies of levels .we can observe the frequencies in terms of count plot for categorical variables analyzing categorical variable frequencies levels .





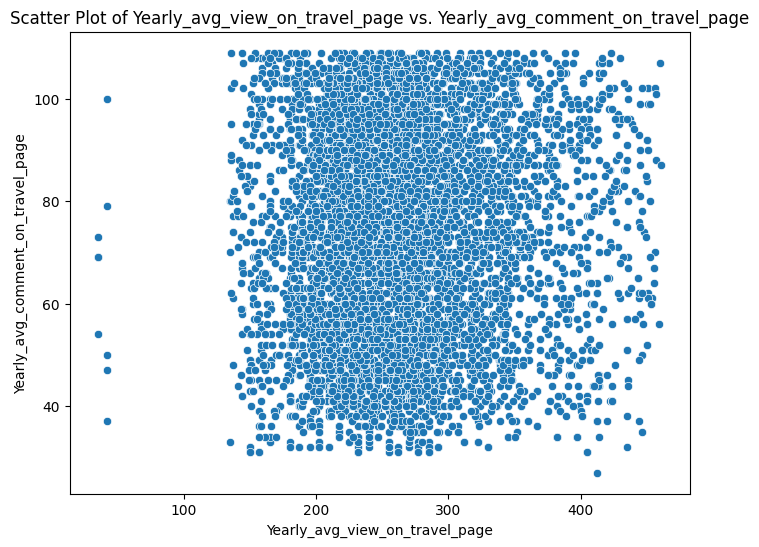


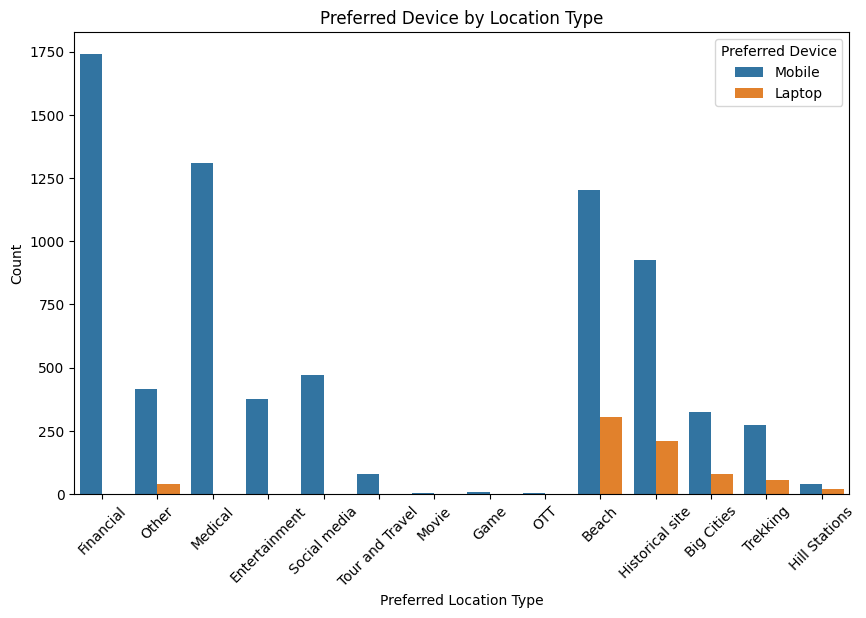
**Observations of Bivariate Analysis:**

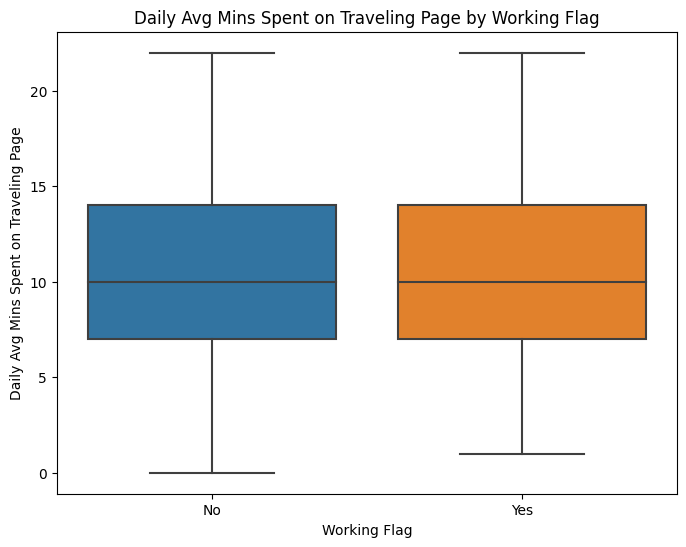
* Most of the users have not taken the product.
* Most user prefer booking product via Smartphone.
* Most preferred locations/trips are Financial or to Beach locations.
* Most Users are not following company page.
* Most of the Users are not working professional.

**VII. Data Visualization: Bivariate Analysis**

With Bivariate Analysis we trying to figure out relativity between available features .





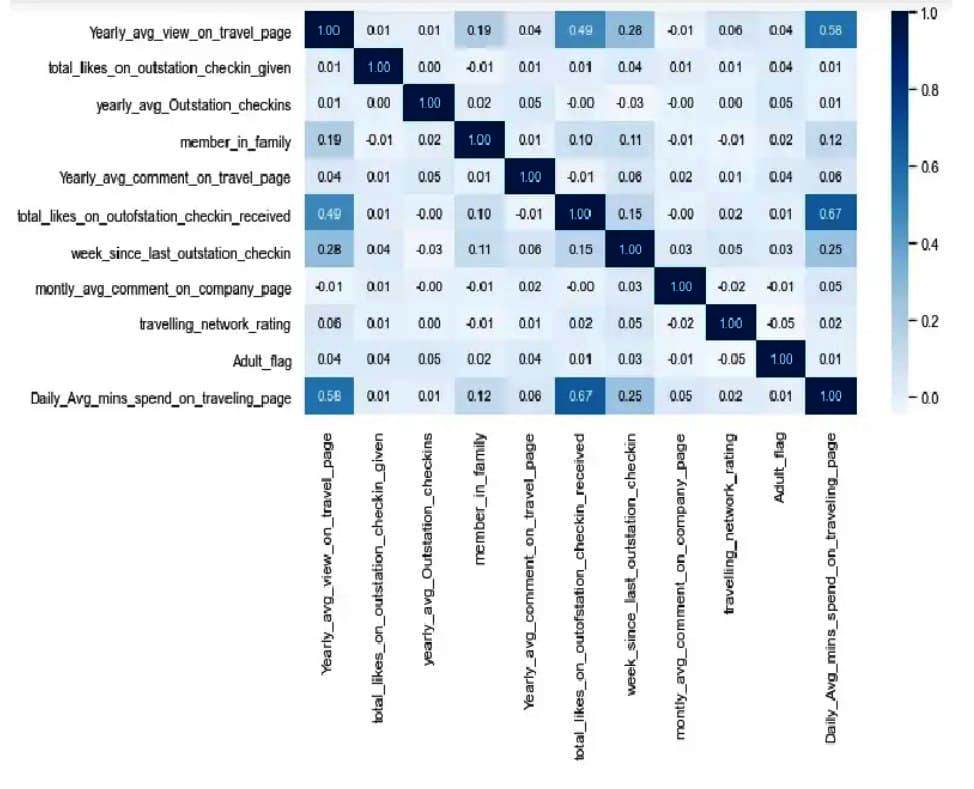


**Observation** : We can see that families with family members count 1 or 3 are more likely to buy the product.

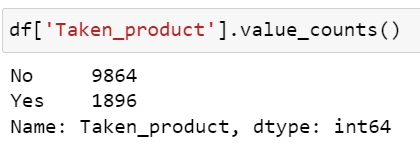
A graph of a group of colored bars

Description automatically generated with medium confidence

**Correlation Heatmap :**



**2.2.3 Sampling Method**



CONCLUSION: Here we can see that the data is imbalanced, so we try to balance the data by using SMOTE technique.

**Before handling imbalance data:**

A screenshot of a computer

Description automatically generated

**After handling imbalance data:**

A screenshot of a computer code

Description automatically generated

CONCLUSION: Here, after separating the dataset the mobile and laptop dataset part has imbalance data and using a sampling technique called SMOTE to handle the imbalance data.

**DATA ANALYSIS AND INTERPRETATION**

**Feature removal:**

In feature selection we only removed two columns after separating the dataset as we only 17 features in total.



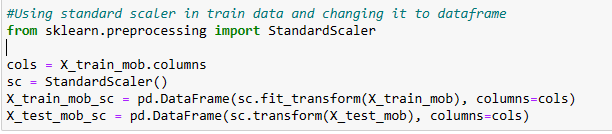
Using the above code we permenantly remove these two columns from out dataset.

**Label Encoding :**

We performed label encoding for object datatype fields on our both datasets(Mobile and Laptop) using Label Encoder

1. **Standard scaler:**

We used scaling technique to make all the X\_train features in same scale using Sklearn’s Standard scaler technique. Below is the code we used to scale the X\_train data.



Likewise, we did the standard scaler for laptop dataset too.

**Machine Learning Model :**

* We first used Logistic Regression, where the accuracy is low as well as other evaluation metrics scores are also low.

precision recall f1-score support

0 0.87 0.98 0.92 1199

1 0.66 0.24 0.35 237

accuracy 0.85 1436

macro avg 0.76 0.61 0.63 1436

weighted avg 0.83 0.85 0.82 1436

* Random forest Classifier model performed well , its accuracy score was greater than Logistic Regression and other performance metrics values also looked good .

precision recall f1-score support

0 0.99 1.00 0.99 1199

1 0.99 0.94 0.96 237

accuracy 0.99 1436

macro avg 0.99 0.97 0.98 1436

weighted avg 0.99 0.99 0.99 1436

* I also trained and tested my data on K- Nearest Neighbor Classifier model which was having high accuracy than that of Random Forest model but its performance metrics values were not as good as Random forest model .

precision recall f1-score support

0 0.96 0.98 0.97 1199

1 0.88 0.78 0.83 237

accuracy 0.95 1436

macro avg 0.92 0.88 0.90 1436

weighted avg 0.94 0.95 0.94 1436

* I trained and test my data on “Bagging with Decision Tree Classifier Model” which gave me best accuracy score and its performance metrics values were also approximately equal to that of Random Forest Classifier model.

precision recall f1-score support

0 0.98 1.00 0.99 1199

1 0.99 0.92 0.95 237

accuracy 0.98 1436

macro avg 0.98 0.96 0.97 1436

weighted avg 0.98 0.98 0.98 1436

**Conclusion :**

By implementing the various models , we have observed that “Bagging with Decision Tree” and “Random Forest Classifier” Models are giving good accuracy and precision but to get more accuracy and precision we will try to balance the data using SMOTE technique and then implementing all the models again with new data set.

***SMOTE*** : Synthetic Minority Over-sampling Technique

**Business Insights:**

* We observed that user mostly travel in the group of 3 or 1 person, so I would recommend the aviation company to provide more offers for the users who are travelling in group of 3 or with 1 person.So, that we can retain most of customers.
* Also, it is observed that users who are not adults are more likely to buy the product. So, it is recommended for the company to show ads to non-Adults.
* We can observe here that the most visited locations are Beach, Financial matter related trips and least visited place is hill station so the company should provide offers and discount to the most preferred locations by users.
* We also observe that the people who don’t follow company page have high average view on company page and people who follow company page has less view this means our social media team is not effective to gain online presence so I would recommend that social media campaigns should be there so that we can grab attention of social media mob as it clearly impact business.
* Improve social media campaigns to enhance online visibility and brand recognition because very less users follow the company page.
* Tailor advertising to mobile users than laptop users for better engagement and ROI as many prefer to buy the product from Mobile.