**CAPSTONE Project**

**Business analytics**

**Table of Contents**

[1. Introduction of the business problem 3](#_Toc145106266)

[1.1 Problem statement 3](#_Toc145106267)

[1.2 Requirements of research study 3](#_Toc145106268)

[2. Data report 4](#_Toc145106269)

[2.1 Data collection based on methodology, frequency and time 4](#_Toc145106270)

[2.2 Visual representation of data 4](#_Toc145106271)

[2.3 Understanding of the attributes 5](#_Toc145106272)

[3. Exploratory data analysis 6](#_Toc145106273)

[Univariate analysis for understanding the dataset properly 6](#_Toc145106274)

[Variables 7](#_Toc145106275)

[1.Yearly avg view on travel page 7](#_Toc145106276)

[2. Preferred device 9](#_Toc145106277)

[3. Total likes on outstation check in given 9](#_Toc145106278)

[4. Yearly avg outstation checkin 10](#_Toc145106279)

[5. Members of family 11](#_Toc145106280)

[6. Preferred location type 12](#_Toc145106281)

[7. Yearly avg comment on travel page 15](#_Toc145106282)

[8. Week since last outstation check in 16](#_Toc145106283)

[9. Following company page 16](#_Toc145106284)

[10. Total likes on out of station check ins 17](#_Toc145106285)

[11. Monthly avg comment on company page 18](#_Toc145106286)

[12. Working flag 18](#_Toc145106287)

[13. Traveling network rating 19](#_Toc145106288)

[14. Adult flag 19](#_Toc145106289)

[15. Daily avg mins spend on traveling page 20](#_Toc145106290)

[16. Taken product 21](#_Toc145106291)

[Findings from the data analysis 22](#_Toc145106292)

[Multivariate and Bivariate Analysis for deriving correlation between the variables 23](#_Toc145106293)

[Heatmap 23](#_Toc145106294)

[Model interpretation and model building 24](#_Toc145106295)

[Models 25](#_Toc145106296)

[Logistic regression model 27](#_Toc145106297)

[LDA 29](#_Toc145106298)

[Naive Bayes model 31](#_Toc145106299)

[Random forest model 34](#_Toc145106300)

[KNN 36](#_Toc145106301)

[Gradient boosting 39](#_Toc145106302)

[Performance media for mobile 42](#_Toc145106303)

[Performance media for the laptop 42](#_Toc145106304)

[Ensemble modeling 43](#_Toc145106305)

[Model tuning 43](#_Toc145106306)

[Interpretation 44](#_Toc145106307)

[Recommendation 44](#_Toc145106308)

[Reference list 45](#_Toc145106309)

# 1. Introduction of the business problem

## 1.1 Problem statement

The aviation company wants to improve the effectiveness of their digital advertising campaigns by collaborating with social media platforms. To target the clients who would like international and domestic flights, the organization has decided to plan the Conventional telemarketing with the data driven digital advertising. The business plans to create a distinct prediction model for individual categories since it understands the customer's inclination to buy tickets differs across the mobile devices and laptops (Drummond et al. 2023). Also, precision is essential during this effort due to the higher expense of digital advertisement. The business wants to discover and engage customers who are highly likely to buy airline tickets by analyzing digital and social behaviours. By doing this, they can make sure that their advertising campaigns are efficient and profitable. This programme aims to maximize advertising spending while enhancing the company's marketing strategy and consumer interaction. The investigation's success depends on its capacity to optimize accuracy while lowering costs in the setting of pricey digital advertising. This strategy ultimately seeks to increase marketing effectiveness, increase ticket sales, and optimize utilization of resources for the aviation industry.

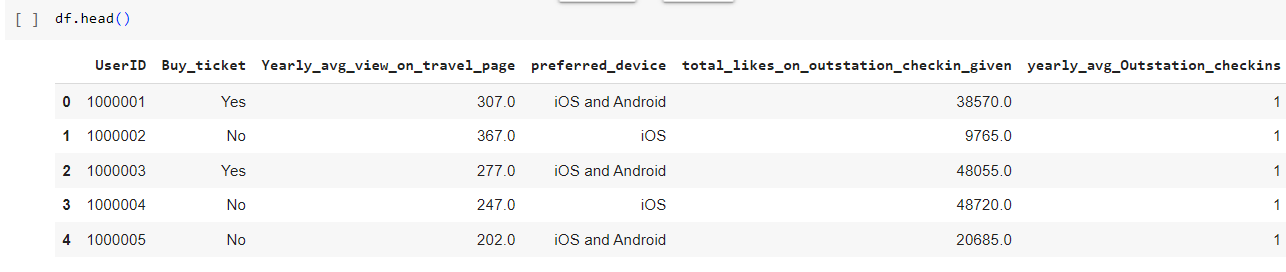
## 1.2 Requirements of research study

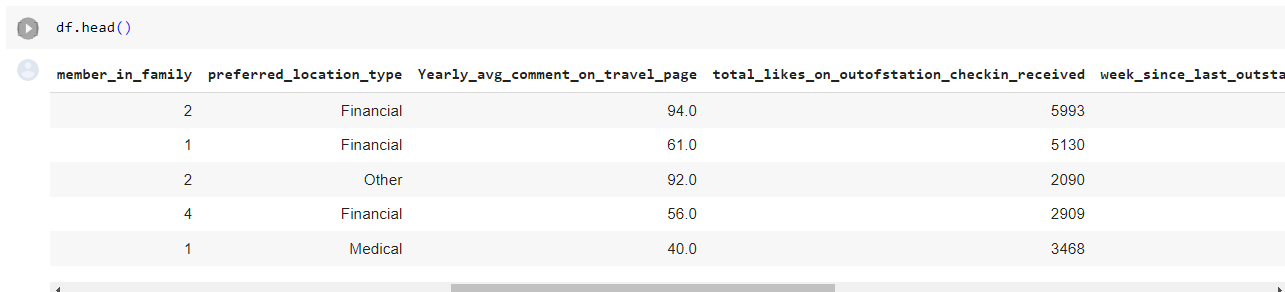
The ubiquity of digital media in the tourism sector has resulted in significant changes to how travelers plan, enjoy, evaluate, and market their travels. Tourism-related webpages and social media networks provide a rich source of information, including user opinions, likes, comments, trip check-ins, and destination recommendations. Social media marketing, in particular, has shown to be an effective strategy for increasing business awareness, increasing website traffic, improving search engine results, generating prospects, and driving sales at a low cost (He, 2022). Marketing professionals in the hospitality and tourism industries recognise the importance of digital platforms in this dynamic environment, marked by more than 2 million evaluations that are updated frequently, especially given the intangible nature of their services. The airline company is examining its clients' digital behaviour as part of this project. They hope to identify the precise client groups most likely to interact with their product offerings by doing this. It is crucial to do a thorough study of the available data in order to understand the "what" and "why" underlying these technological dynamics. In order to ensure that the business's long-term strategies remain in line with the changing demands and preferences of its intended consumer base, this study will be used as the basis for developing predictive products.

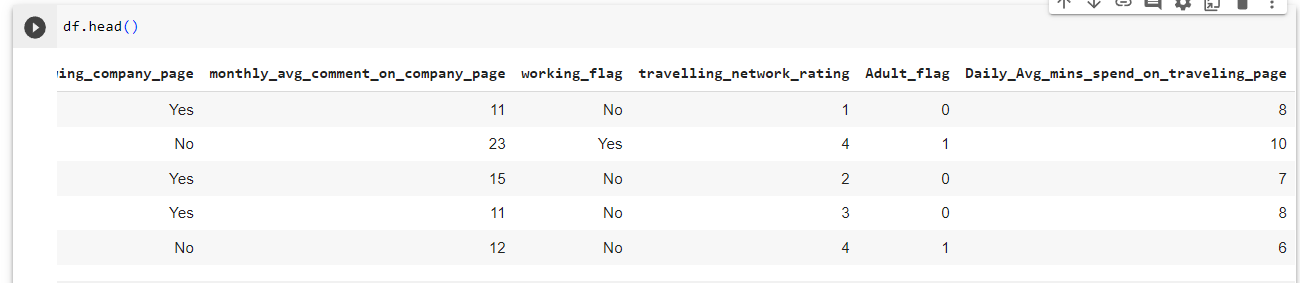
# 2. Data report

## 2.1 Data collection based on methodology, frequency and time

Data is collected based on the customers’ records. The below section has shown the structure of data.

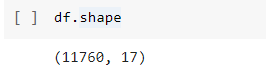




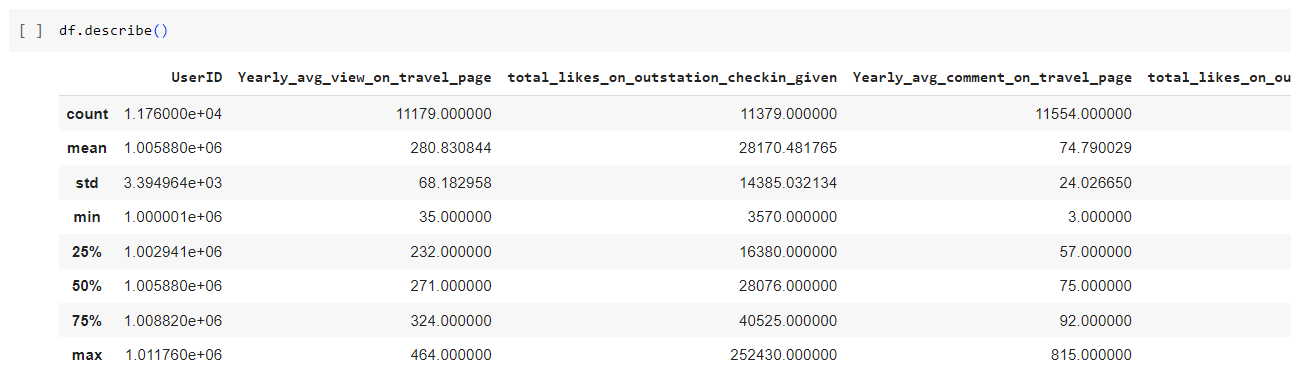


## 2.2 Visual representation of data

Rows and columns are discussed in the below section.



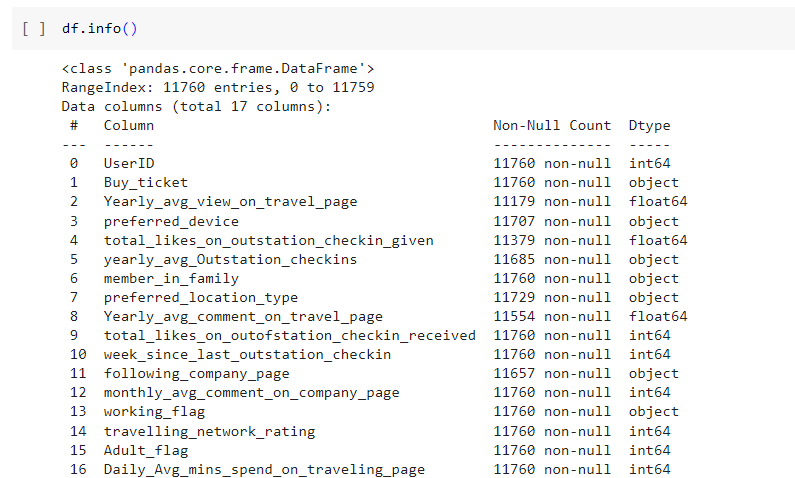
describe() function has used to define the unique values based on the summary of the dataset.



Based on the above implementation, the figure has given mean value, min value, max value, percentile, and standard deviation of these datasets.

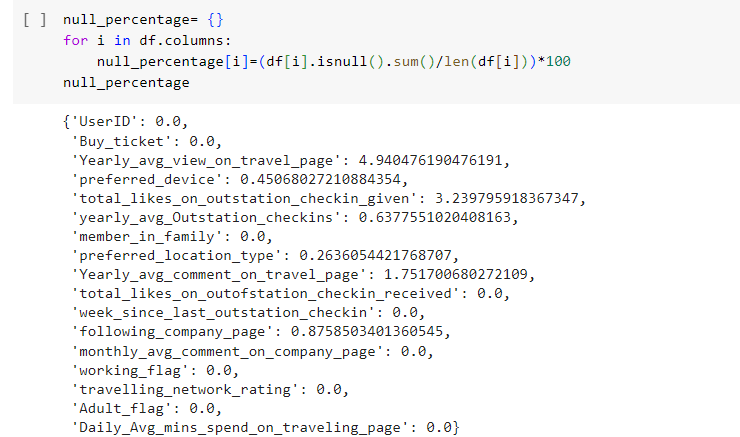
## 2.3 Understanding of the attributes

info() has used for getting the output of the missing values and data types based on the provided datasets.



# 3. Exploratory data analysis

## Univariate analysis for understanding the dataset properly

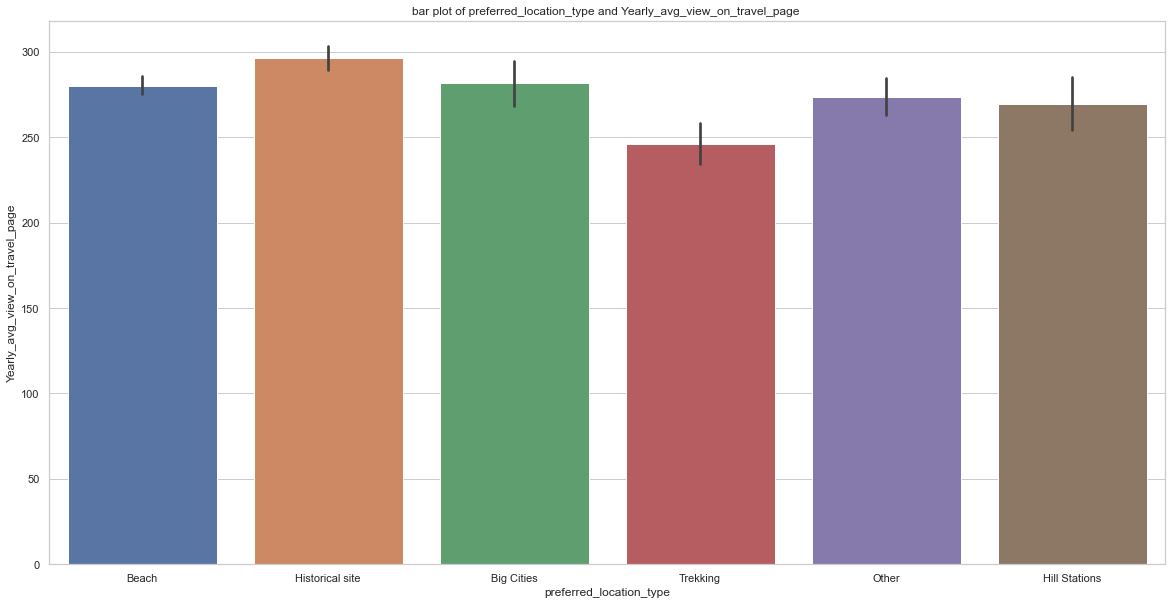


The large percentage of missing values in the ***“yearly average views on travel page”*** column raises the possibility that there may be difficulties in routinely gathering or recording this data. This lack of information may limit our ability to understand user engagement with and interest in travel-related content, which may have an impact on decisions about marketing tactics and content planning. The lack of information in the ***“total likes on outstation check-ins and yearly average comments on travel page”*** columns further highlights a gap in our knowledge of user interactions and feelings related to travel experiences. The evaluation of user-generated content and the measurement of interest in and interaction with the destination depend heavily on these measures. Since these variables reflect consumer preferences and significantly impact marketing initiatives in the tourism sector, addressing these missing data concerns is essential for making well-informed decisions. The reliability and completeness of these important insights should be improved by implementing data gathering and quality control strategies.

# Variables

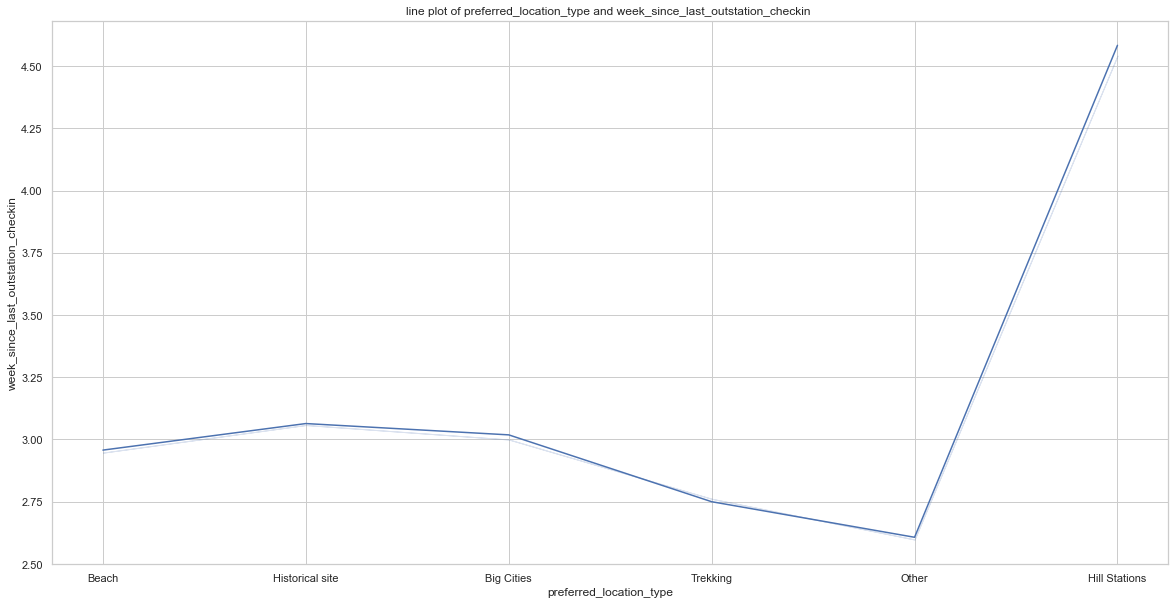
## 1.Yearly avg view on travel page

The below figure has provided several insights into the passengers’ performance by displaying the yearly average outstation checking for different categories. According to the research implementation, most popular locations are defined as the ***“beach” destination with 2424 hotel check-ins***, demonstrating a strong preference for the coastal gateways. Also, the ***financial site*** has indicated a mix of ***leisure and work travel with 2376 hotel check-ins***. Moreover, specific hobbies are also visible, as seen by the low check-in rates of the ***“Game” along with “Movie” classifications***, which suggest ***professional travel preferences***. ***“Trekking” and “hill stations”*** are popular apps that capture a lot of check-ins (Keke, 2023). On the other hand, ***“medical” and “historical sites” travel locations*** highlight the health and cultural related travel trends. The research results emphasize a wide range of the traveler interest highlighting the requirements for specialized marketing plans to accommodate such various tastes. Furthermore, it highlights the importance of the ***“Financial” and “beach”*** segments through the travel business, necessitating research for focused advertisement and services.

****

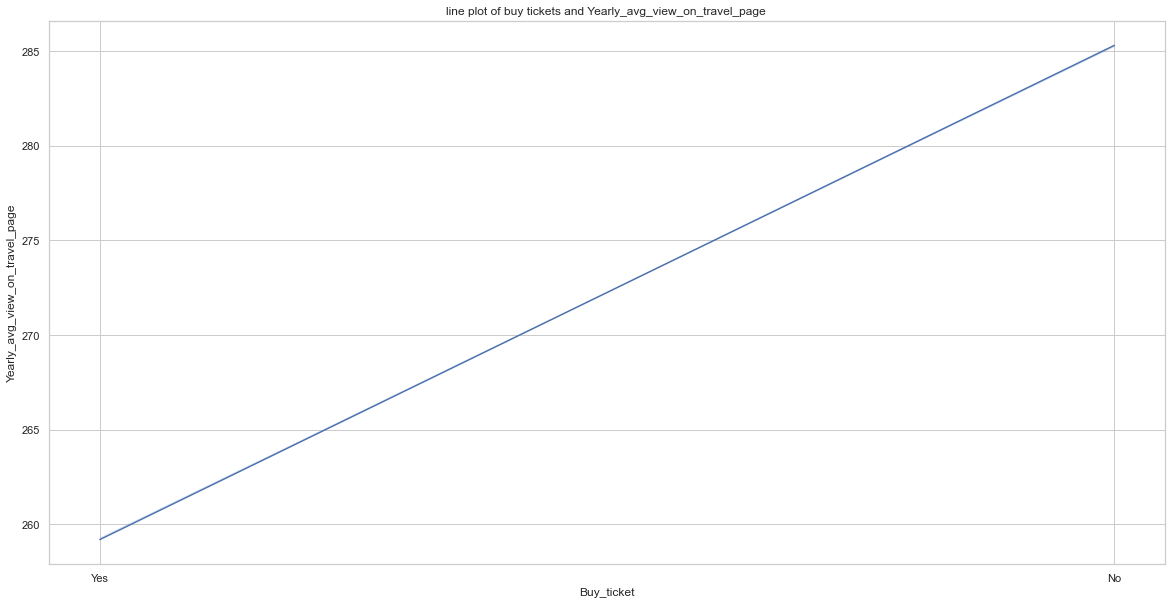
**Distribution plot**

The below Graph has been used for the observation set and visualized yearly average view on the travel pages through the line chart.

****

**Probability plot**

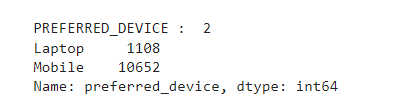
The below Graph has represented the probability plot to access the data which has allowed a particular distribution. According to the research findings, the data are significantly plotted based on the theoretical distribution in an organized way which has been pointed through the straight line. The straight line has indicated the departures according to the specific distribution.



Here, Missing values have been imputed according to the variable of the data sets and plotted the probability plot and new distribution.

## 2. Preferred device

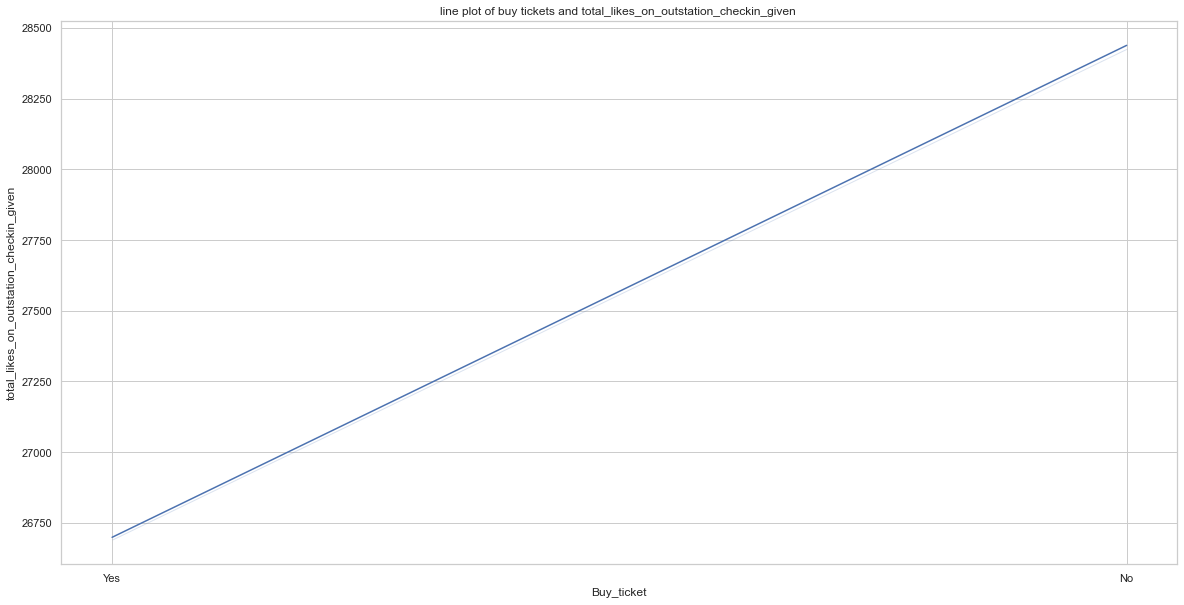
The below figure Has calculated the preferred devices according to the mobile and laptop. Based on the above findings, laptop has 1108 users while mobile has 10,652 users. According to the major inequalities, it has been stated that most customers receive Aviation Services through mobile devices. This result has highlighted the significance of the mobile platform Optimisation for digital advertising along with the organisation's User experience. Different strategies including mobile app development and web design, would be privatized for developing the marketing audience. However, it is essential to make sure that User experience is consistent through the platform to draw the customer's attention and improve the general accessibility of its Aviation Services (Ozuem and Willis, 2022).

****

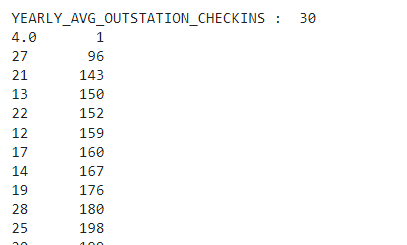
Group by category has created for counting the total number of mobile and laptops used. Based on above implementation, mobile is significantly used by 90% of the consumers. Also, mobile device is preferred devices as count. Tab is maximum used products with the count of value 4172.

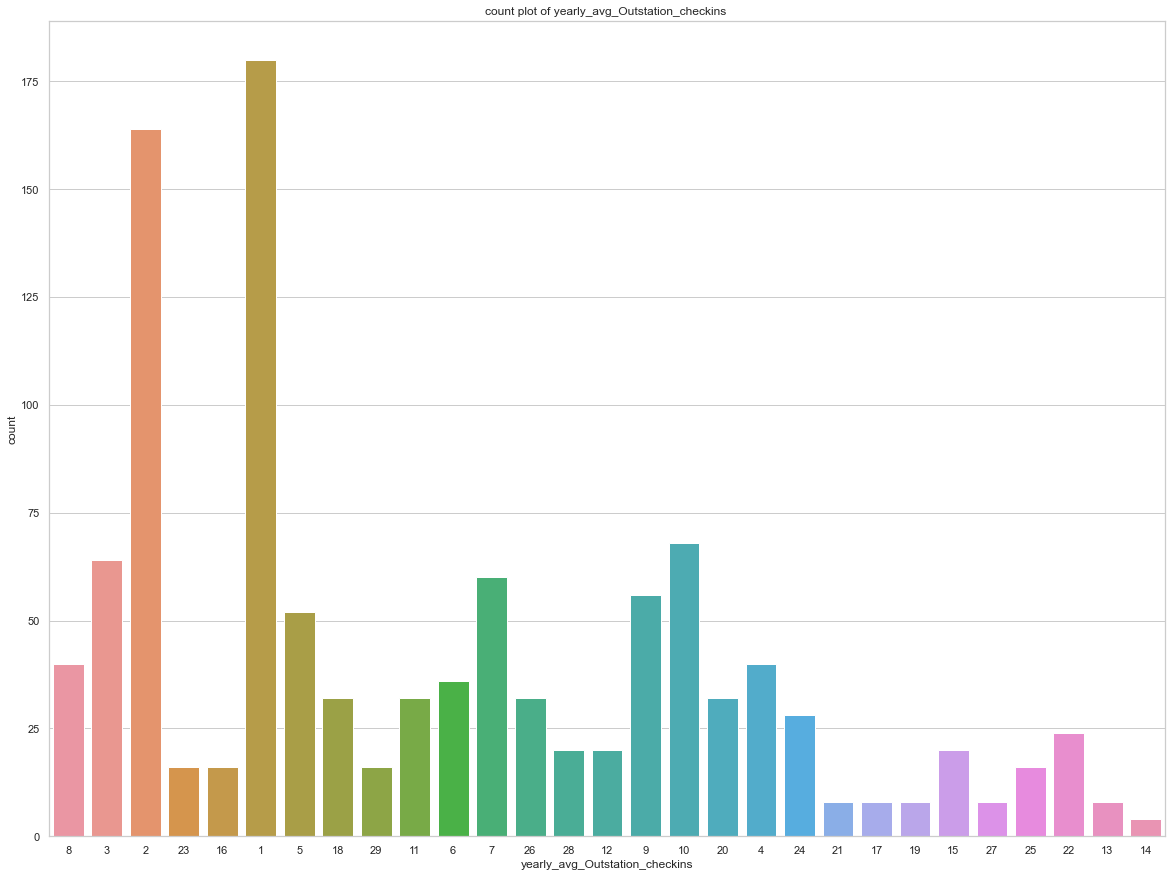
## 3. Total likes on outstation check in given

Distribution plot has plotted in the below section based on variable after imputing the missing value. Groupby function has used for discussing the total likes on the outstation check of the airport services. The below image has shown the distribution of the yearly average outstation seconds with the maximum value.

****

## 4. Yearly avg outstation checkin

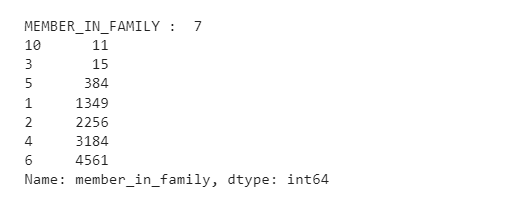
****

****

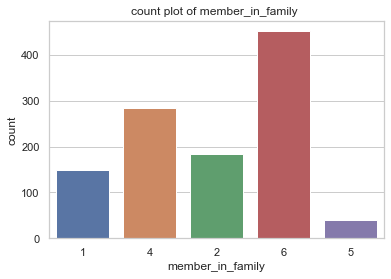
The above figure has shown the yearly average outstation checking based on the airport services.

## 5. Members of family

The counterplot has counted the different categories and returned the count of its occurrences. Also, this plot has been provided to display the members of the family by the Seaborn Python library.

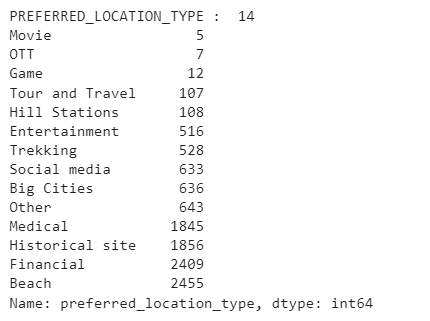
****

Most people have effectively three members in a family, followed by two and four members. Also, ten members in the family have a minimum value based on the result findings.

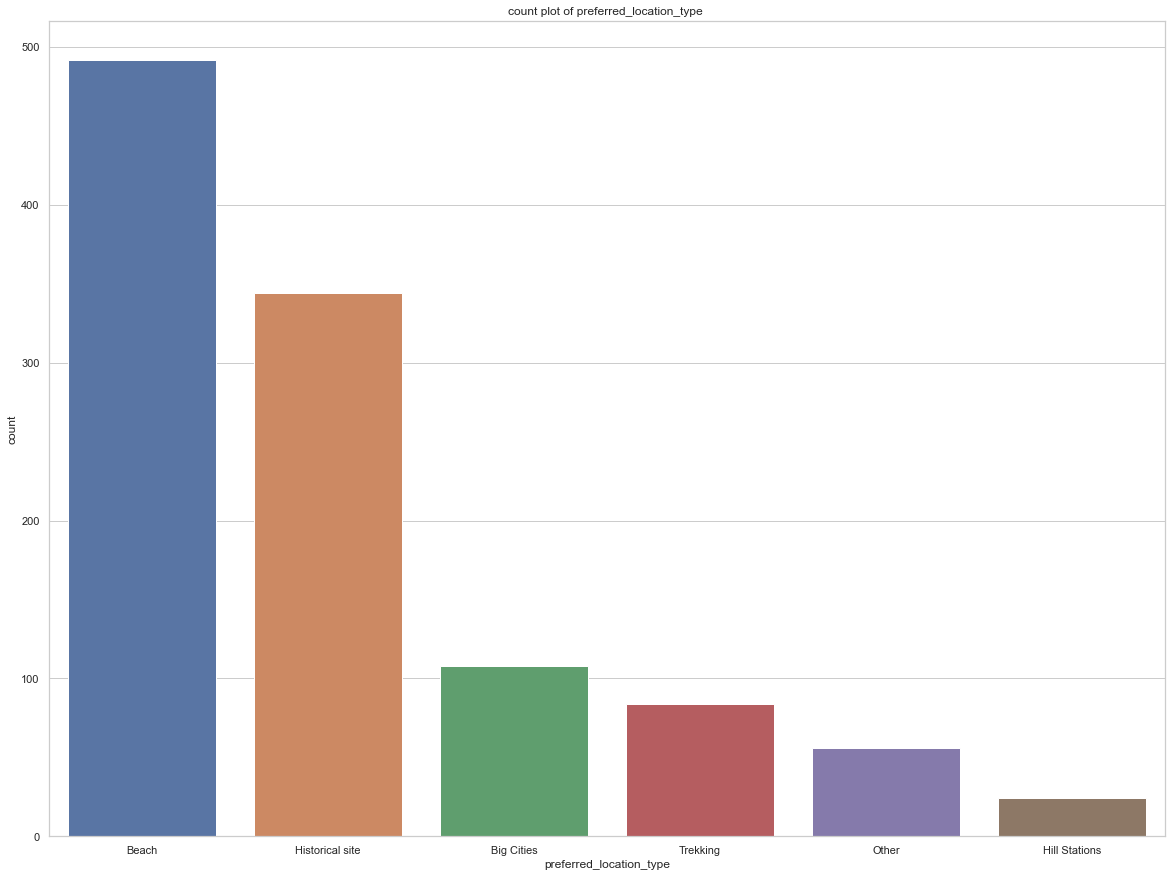
****

## 6. Preferred location type

According to the result implementation, maximum customers are preferred for financial and beach locations which are followed by the historical site.

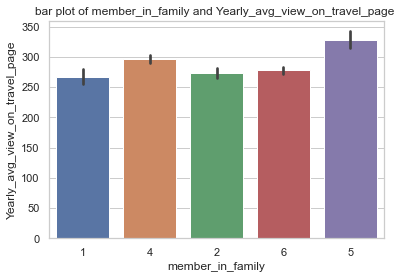
****

With the help of the preferred data, the diagram below explains the client's preferred location type and travel preferences. According to the analysis, most popular destinations are defined as beach destinations, demonstrating a significant Tendency for the coastal gateways. The financial preferences are listed as second for indicating leisure and business travel. Also, ***OTT movies and game***s have specialized travel preferences and highlighted particular interest. Based on the above implementation, historical sites, trekking and hill stations are defined as the interest of the nature and history enthusiasts (Ren et al. 2023).

****

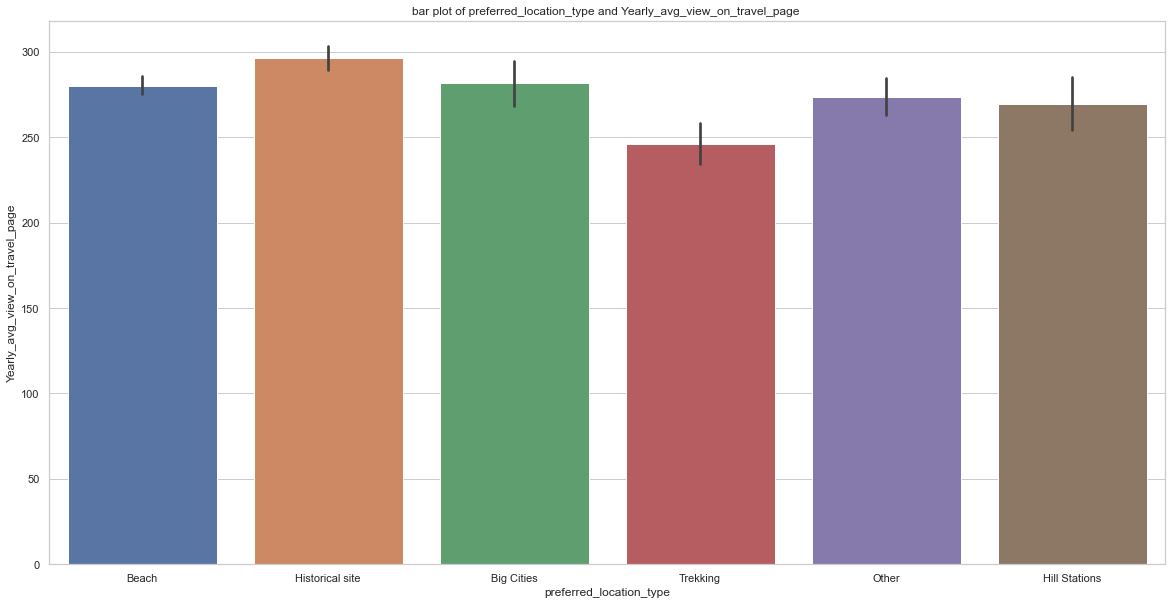
Such results are Illustrated the wide range of the traveller interest and also significant of the customized marketing approaches to satisfy such various preferences. The result findings are held to create the personalized travel experience and also optimize the marketing initiative for developing with the popular sections, increasing the consumer happiness and engagement

## 7. Yearly avg comment on travel page

****

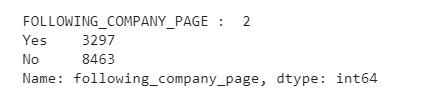
The information provides light on how family size and the yearly average number of views on a travel page relate to one another. Mainly, there is a strong association between the size of the family and page involvement on travel. Families with three members exhibit the most significant yearly average views, indicating their keen interest in vacation preparation (Saboune, 2022). This might be ascribed to the harmony between small-group characteristics and the freedom to go to different places. In contrast, individuals or couples (families of one or two) display comparatively lower view counts, indicating that they could value various forms of transportation differently. It's interesting to note that families with ten members have a surprisingly low number of views, which may be due to the rarity of such large families in the dataset. These results highlight the need to adapt travel marketing plans to particular family sizes in order to maximize engagement and satisfy a range of preferences within the travel sector.

## 8. Week since last outstation check in

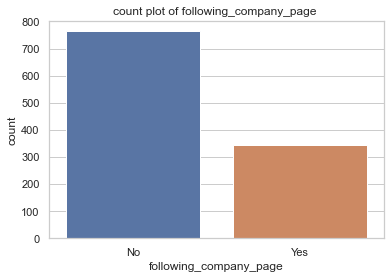
****

Based on the above implementation, it has been summarized that most consumers (3065) last checked in 1 week ago.

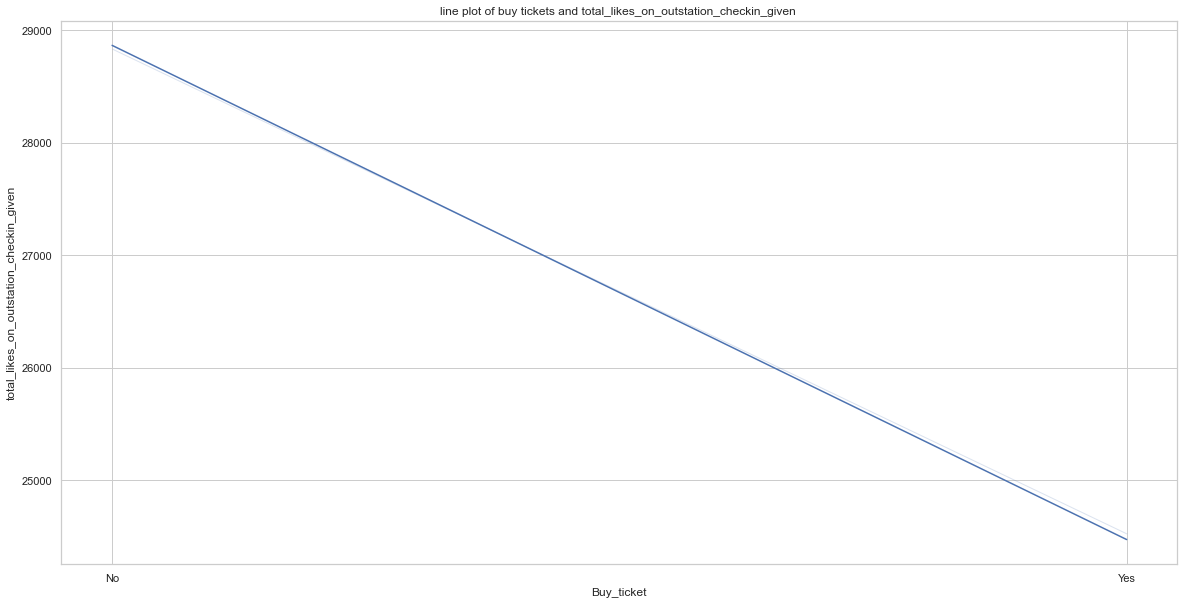
## 9. Following company page

****

The figure below has discussed that the Majority of the customers who are more than 80000 are not allowed to follow the company page. Also, approximately 3500 consumers are followed the company page.

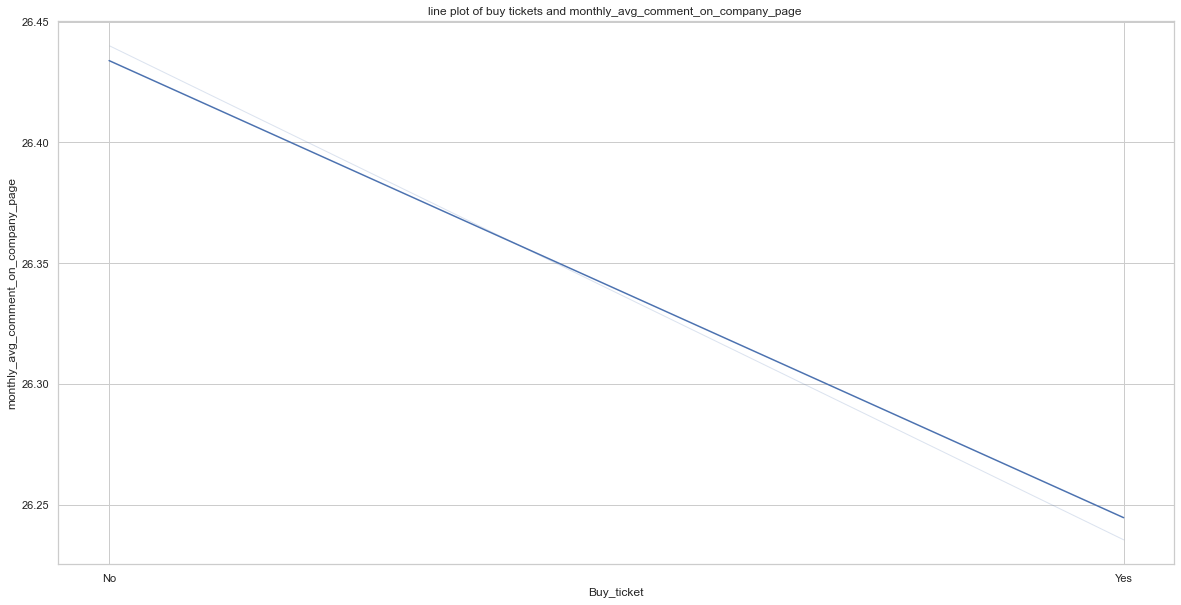
****

## 10. Total likes on out of station check ins

****

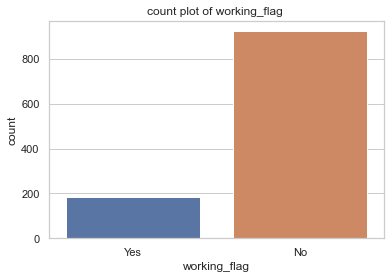
The above figure has shown the total likes based on the station check in of the airport services.

## 11. Monthly avg comment on company page

****

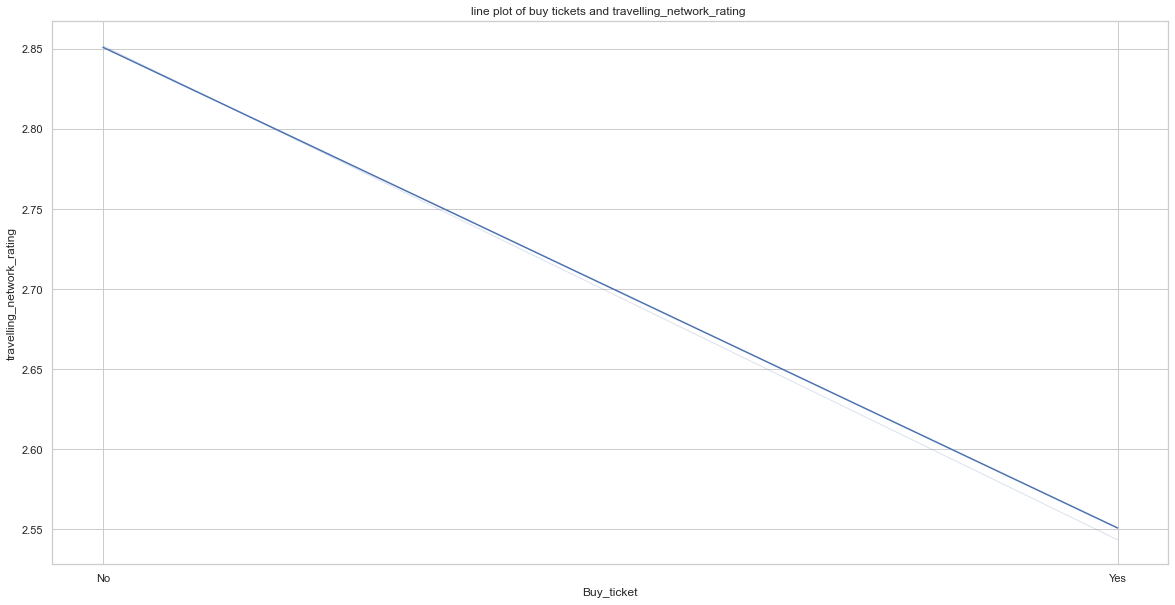
Based on the above implementation, It has been assumed that maximum monthly average has defined as value 11 that are counted to value 405.

## 12. Working flag

****

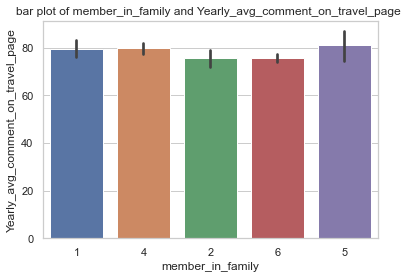
Based on the above implementation, it has been stated that the Majority of the customers are defined as 84%, and they are not working in this airport services.

## 13. Traveling network rating

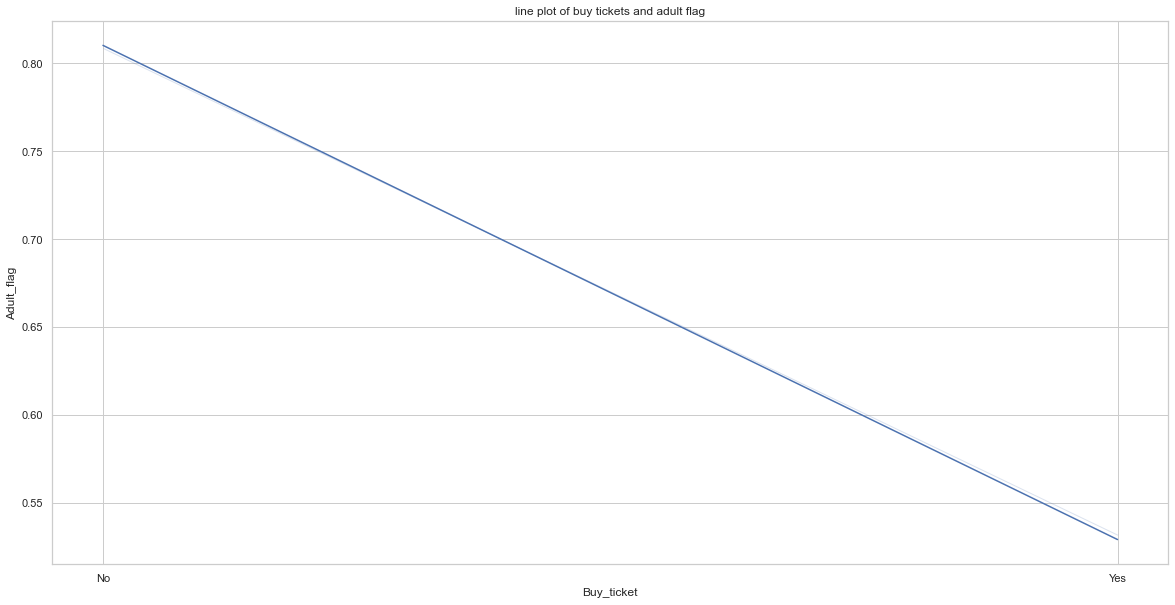
****

In the above figure, Traveling network rating has been discussed. Here, ***maximum rating*** is defined as ***value 3*** and that is ***followed by value 4***. The Count of the ***traveling network rating 3*** is defined as ***3617*** and count of the ***traveling network rating 1*** is defined as ***2181*** which is comparatively low based on above data.

## 14. Adult flag

****

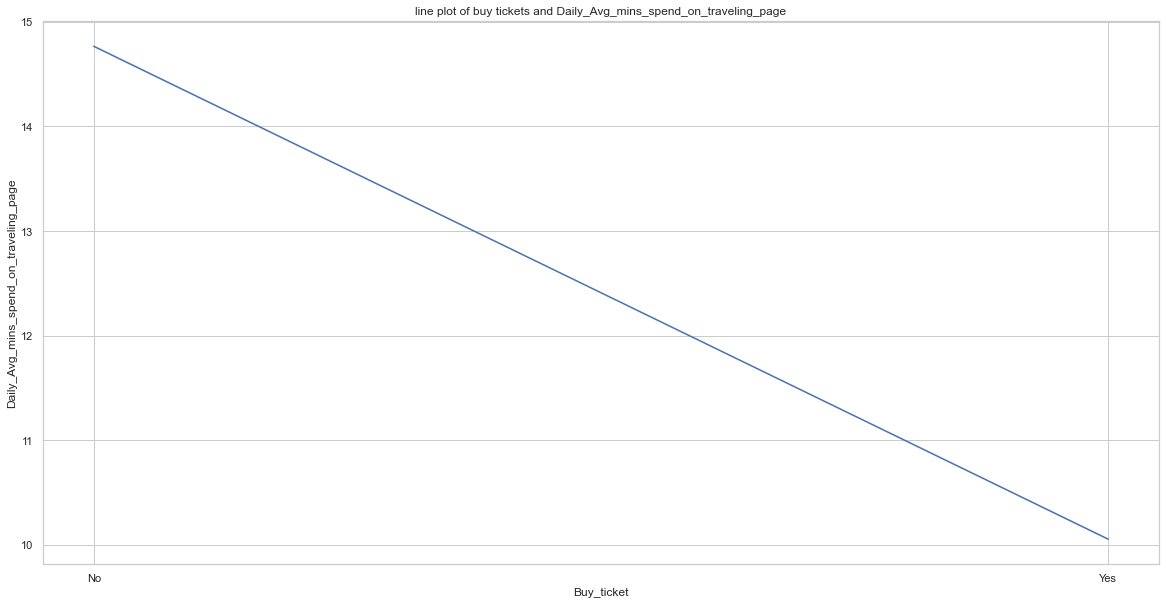
In the above figure, Four different categories, such as 0,1,2 and 3 are used to define the adult flag. Based on the above implementation, it has been assumed that data rollup problems have occurred defining 1 for being written. The ***count of the*** ***adult flag 1*** is defined as ***6633*** and ***distribution*** is defined as ***57.2***. Also, the ***count of the adult flag 0*** is defined as ***4963*** and this ***distribution*** is defined as ***42.8***.

****

Based on the above implementation, it has been summarized that 57% of the responses are included as adults and others percent are included as customers.

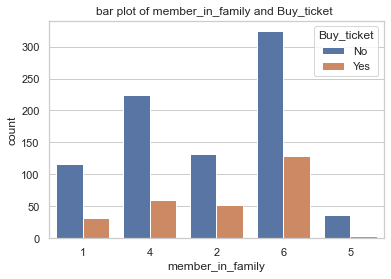
## 15. Daily avg mins spend on traveling page

The distribution plot has been defined in the below figure. The outliers are defined as the extreme value according to the data set. in the figure below, the average minimum value is discussed according to the travelling page.

****

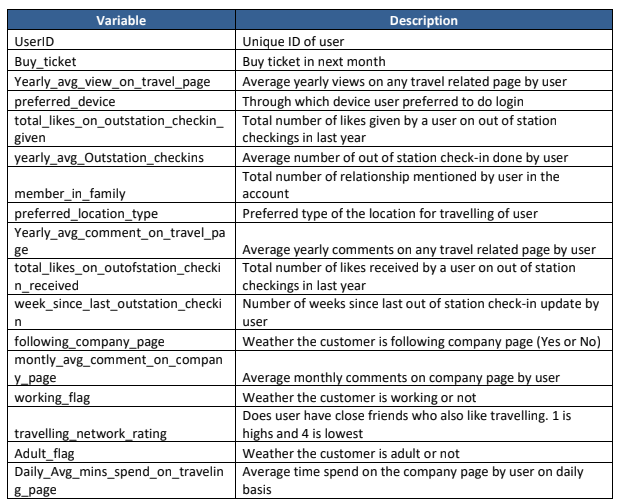
The line chart has represented the distribution of different quantitative data to compare the different variables. The above Graph has shown the data set quartiles and also defines the distribution which has represented the outlier presence.

## 16. Taken product

****

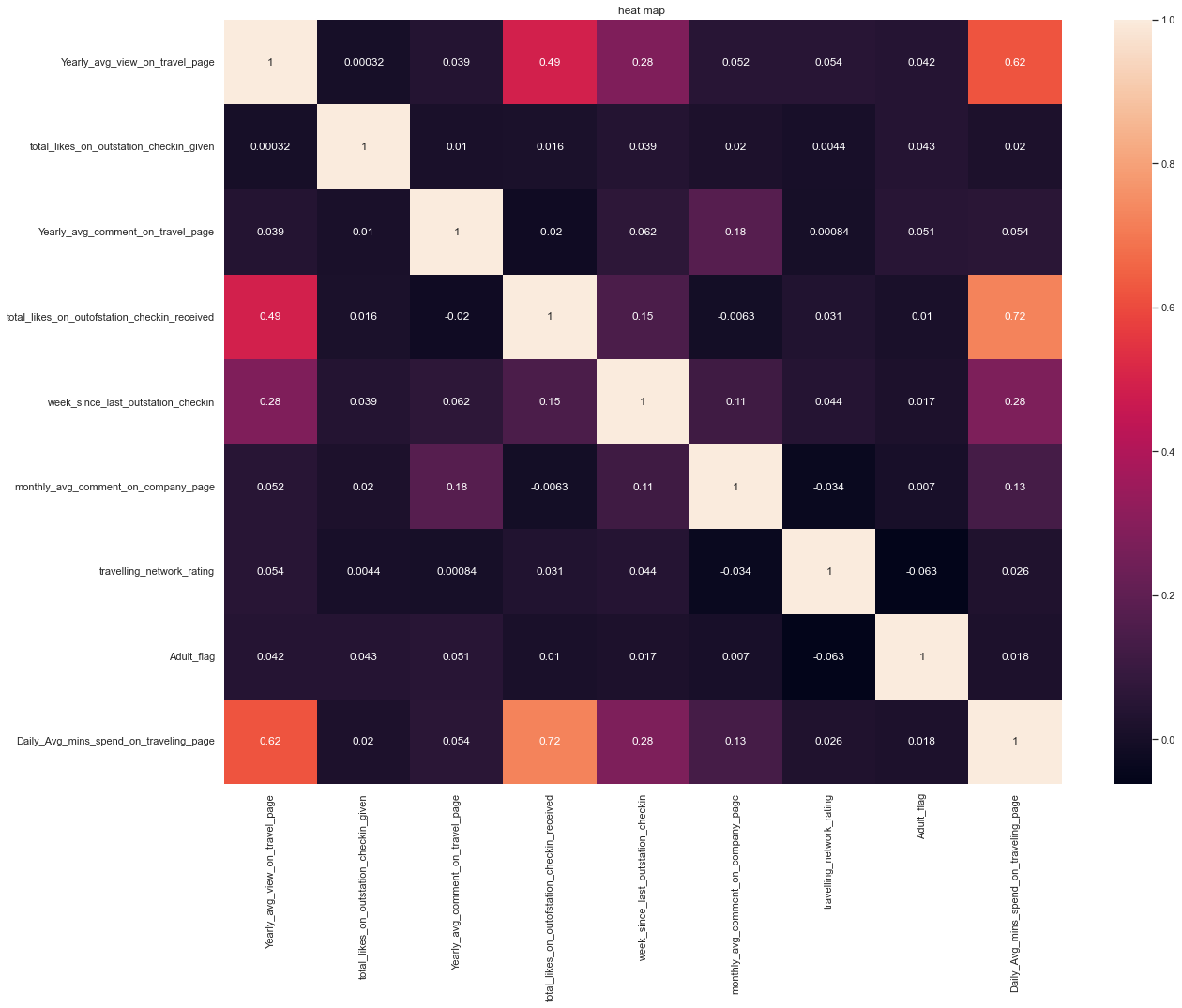
According to the above implementation, the Majority of the customers are defined as 85% and they have not taken the products in the past.

# Findings from the data analysis



## Multivariate and Bivariate Analysis for deriving correlation between the variables

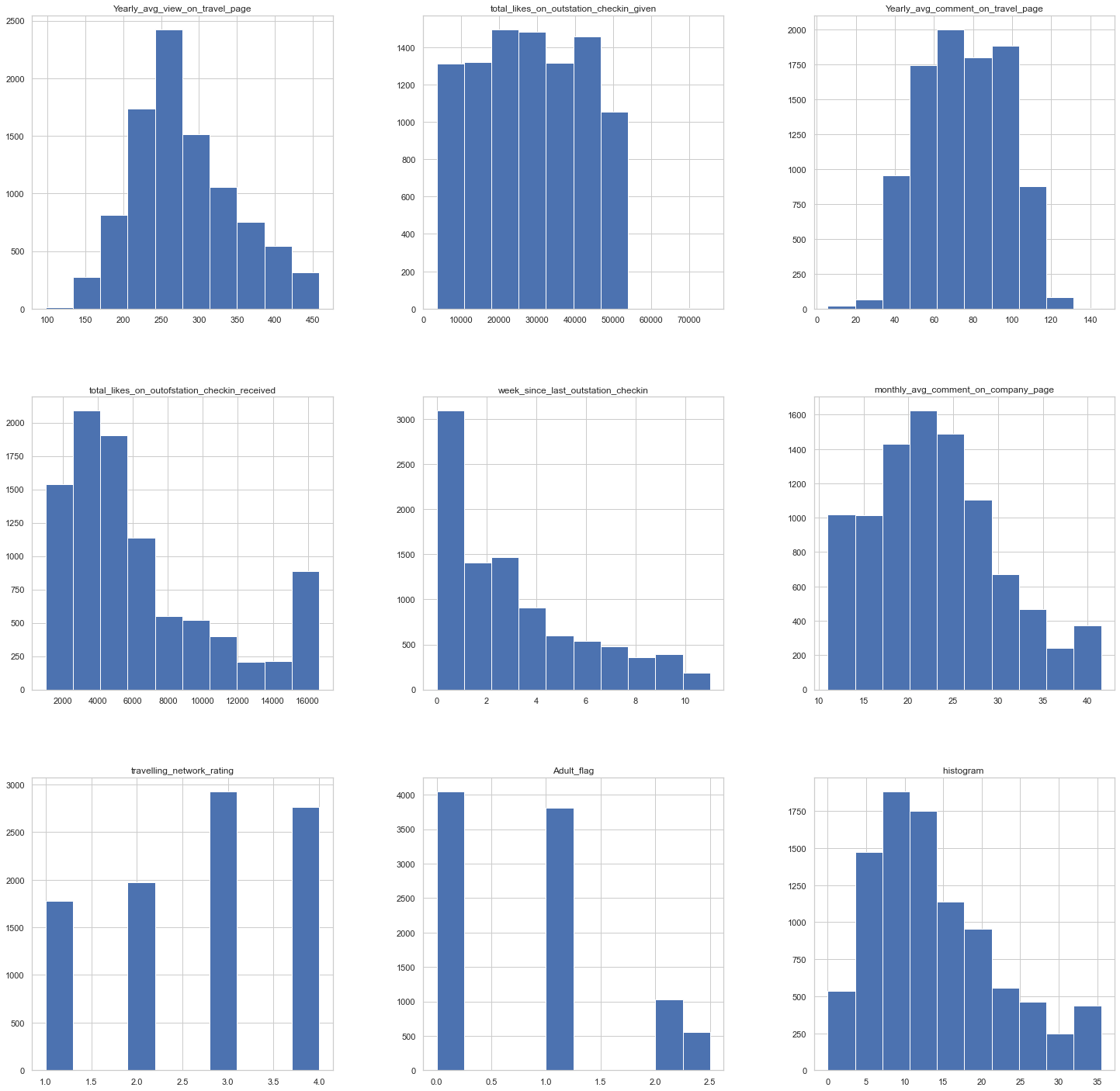
## Heatmap



The heat map has presented the above discussions based on the total like on the out of station checking receive and the yearly average view on the travel page. According to the above implementation, it has been discussed that ***yearly average view on the travel page*** and the ***total likes on the out of station taking received*** may have the ***highest correlation*** with the value ***daily avg mins spend on the traveling page*** for increasing the value proportionally. Also, the data percentage has retained for the model evaluation process which is counted as 92.2%.

# Model interpretation and model building

The model is trained according to the data and also prepared to apply different algorithms and techniques. Also, the data modelling technique has made it easier for integrating the high level business approaches with the data structures, rules, and also technical implementation of the physical data (Shrivastava et al. 2023).



Here, modelling process has meanted to trend the machine learning algorithm for predicting the label according to the different features and also tuned it to represent the business requirements which has validated it based on the holdout data. Moreover, the modelling output has defined as the tranined model for using the Predicted value based on various data points. The EDA analysis has represented the above figure to define different variables of the data.

# Models

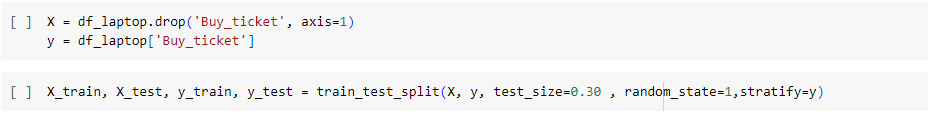
The dependent variable is the categorical binary, and the implementation below has defined different models for predictions and improperrances.

* Logistic regression
* Random forest
* K nearest neighbors

The classification model has predicted the conclusion according to the input data which is given for the training value. it will predict the category and class for the provided data. The main objective is to predict whether the product will be significantly received or not. For the model-building process, data is divided by a 70:30 ratio, and the size of the training data is defined as 70% whereas the size of the testing data is defined as 30%. Also, the model will be evaluated To define the classification report, roc curve, and accuracy score. the accuracy score of the model has used Different techniques for observing the data according to the case study.

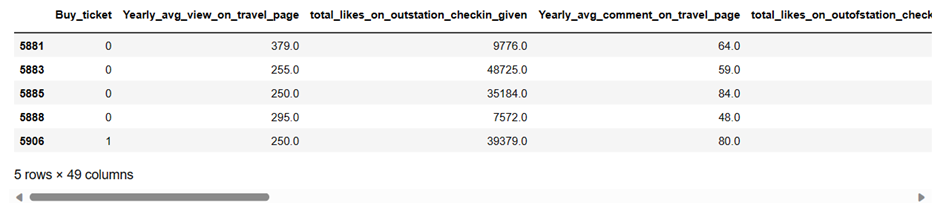
**Data splitting in the testing and training process**

The data is splitted into different categories, suggesting and training datasets. Here, training data has been defined as 70% of the observations and the testing data is defined as 30% of the observations. For the model evaluation, the training data set will be significantly used to define the accuracy score of the model and the testing data set will be used for defining the testing set.



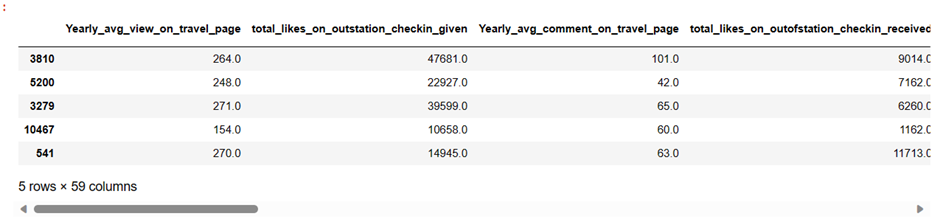
**df\_laptop**

To drop preferred data, data is encoded and defined in the figure below.

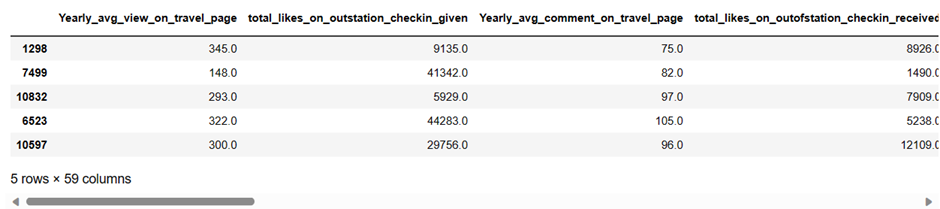


In thai figure, data is not properly scaled and variable values are defined in the same range. To scale with the data, the dataset is splitted for training and testing the data based on the above ratio.

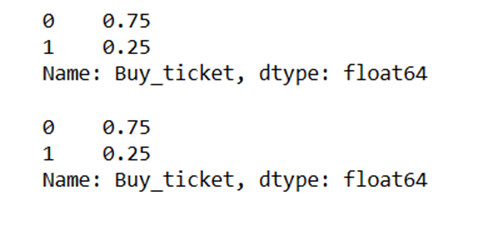
**Training head of the split dataset**



**Testing head of the split dataset**



The training and testing set's value count is defined as the target variable's value count.



## Logistic regression model

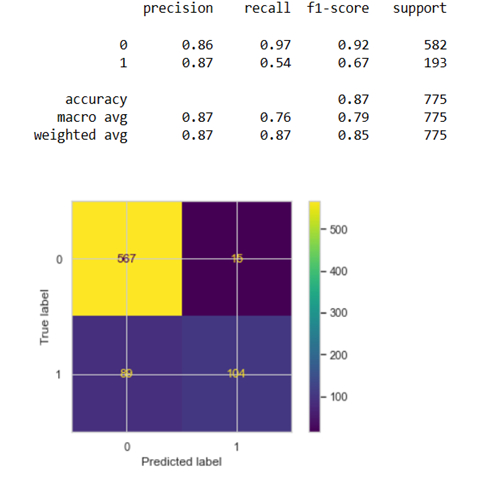
By estimating the likelihood that an instance belongs to a given class, the statistical technique of logistic regression is utilized for binary classification. Despite its name, the generalized linear model (GLM) type of logistic regression is used to model the association between a binary outcome variable and one or more predictor variables. As a result of its ease of use and interpretability, logistic regression is frequently used. It's crucial to remember that it works best when the connection between the predictors and the outcome is roughly linear and the outcome variable is binary or has the potential to become binary.

**Building the models to train the datasets**

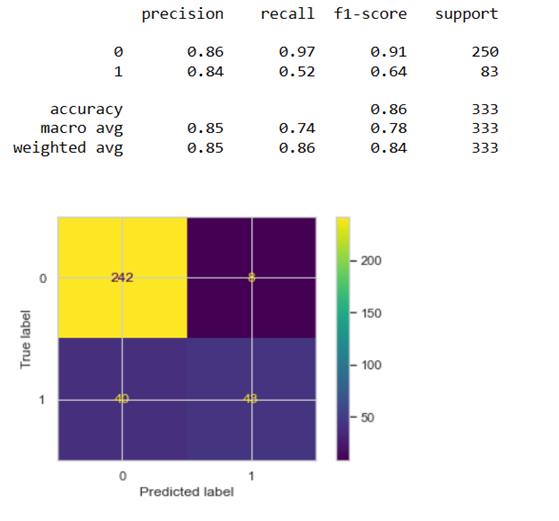
**Logistic regression**

***Parameters of the model evaluation (confusion matrix, classification report, accuracy)***

**Confusion matrix and classification report for the training data**



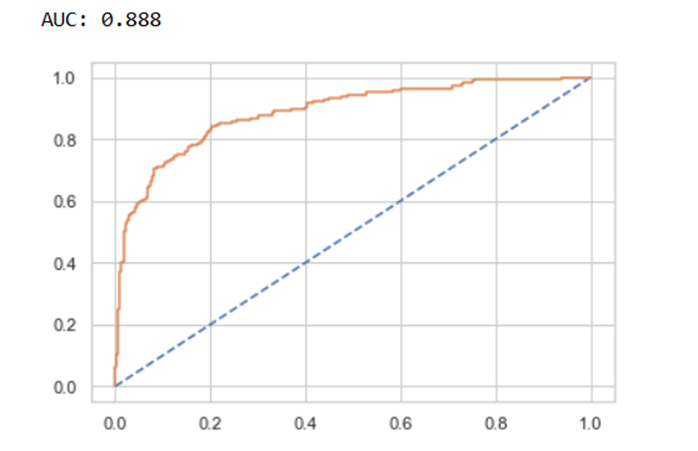
**Confusion matrix and classification report for the testing data**

****

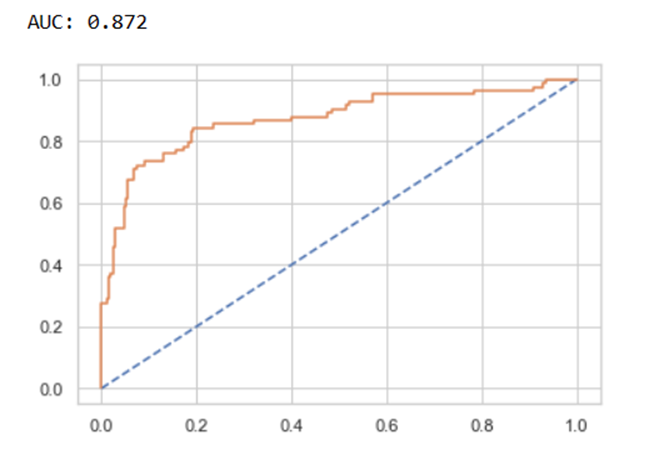
**ROC Curve**

The ROC curve has been plotted in the below section and AUC has been calculated based on the evaluation of the performance measurements. Also, the ROC curve has been generated to plot the TPR based on the FPR at different threshold settings and this AUC has been defined as the area based on the ROC curve. The model has discussed the predictive ability which has AUC closer towards value 1.

**ROC curve for the training data**



**ROC curve for the testing data**

****

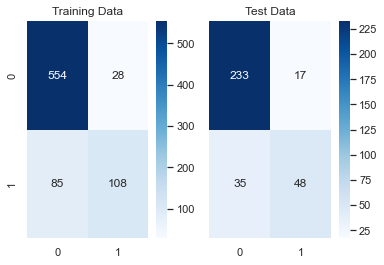
* 87% accuracy rate for train data
* Test data accuracy rating: 86%
* The F1 scores, which were 67% and 64% for the Train and Test data, respectively, were comparable to the default model. We can conclude that while our model appears to be a right fit model and is avoiding the underfit and overfit scenarios, its scores are not particularly strong. The train data appears to be doing slightly better when compared with the test data, however the difference could be more significant..

## LDA

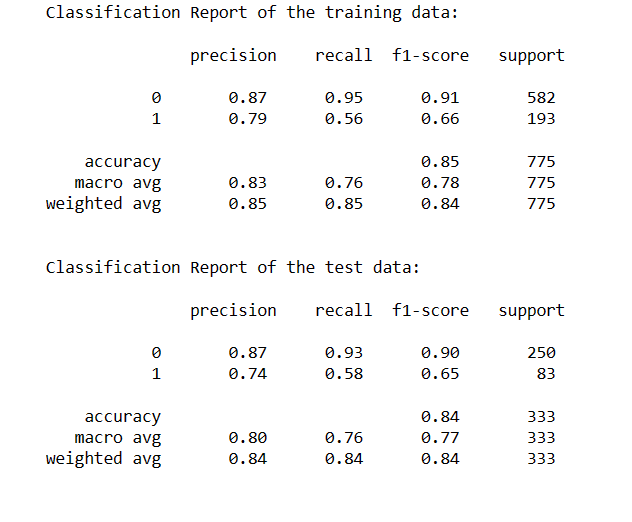
LDA presupposes that the data has a Gaussian distribution and that the covariance matrices of the classes are comparable. It is especially helpful when you want to keep the discriminatory information between classes while reducing the dimensionality of the data and you have numerous classes. In conclusion, LDA is a method for identifying linear combinations of features that best distinguish between classes. This makes it effective for dimensionality reduction and classification tasks in situations where class reparability is a key consideration.

**Building the models to train the datasets**

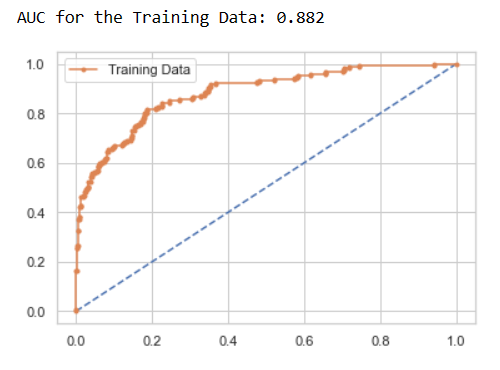
**Confusion matrix of the LDA training set and testing set**



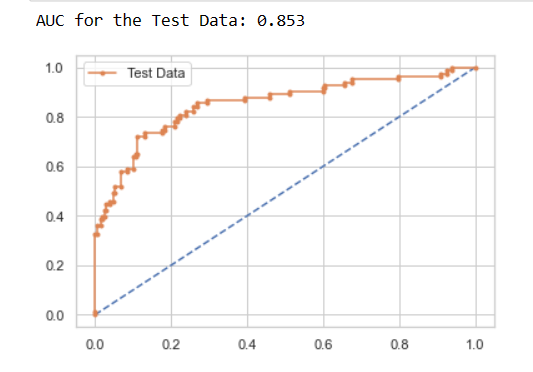
**Classification report of the LDA training set and testing set**



**ROC curve of the training set**



**ROC curve of the training set**



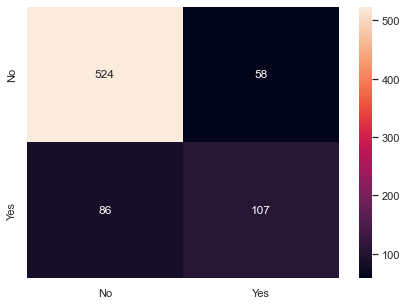
Based on the above implementation, the Train data accuracy rate is 85% and the Test data accuracy score is 84%. The F1 scores were comparable to the default model, which had test data accuracy ratings of 66% and 65%, respectively. This LDA model appears to be a right fit model as well.

## Naive Bayes model

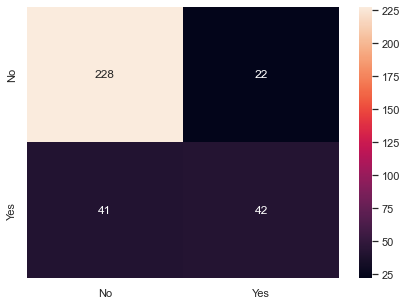
This theorem has discussed the classification tasks based on feature independence.

**Building the models to train the datasets**

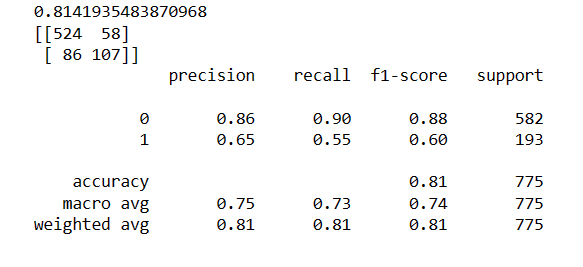
**Confusion matrix of the training set**



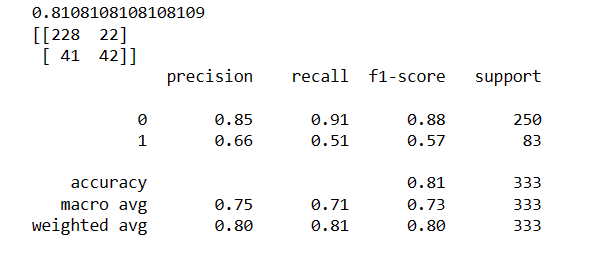
**Confusion matrix of the testing set**



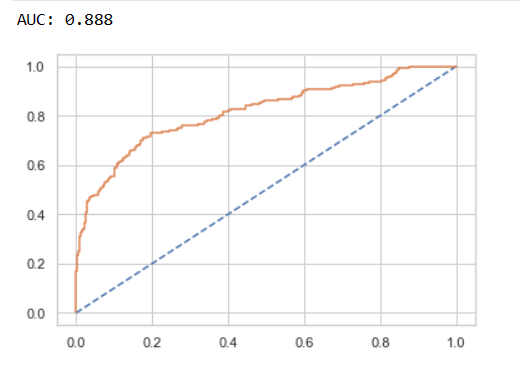
**Classification report of the training dataset**



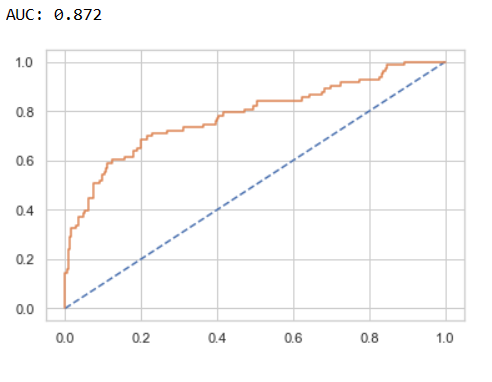
**Classification report of the testing dataset**



**ROC curve for the training dataset**



**ROC curve for the testing dataset**



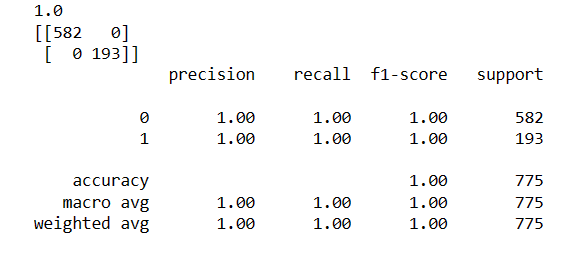
Based on above implementation, it has been summarised that Score of accuracy for train data is calculated as 81% and Test data accuracy rate is calculated as 81%. While compared to the default model, the F1 scores were similar, with 60% and 57% for the Train and Test data, respectively. The accuracy of the data from the test and the train is identical. This model, we can say, is a right fit model.

## Random forest model

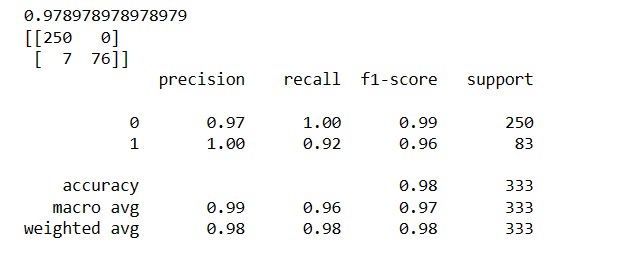
The Random forest has used for predictive analysis when considering family size and the yearly average views on a vacation page. Random Forest may find patterns and relationships in the data by using parameters like family size as input variables and yearly average views as the objective variable. Based on the size of the family, it can assist in predicting the anticipated level of interaction for individuals or families, providing insightful information for travel marketing plans.

**Building the models to train the datasets**

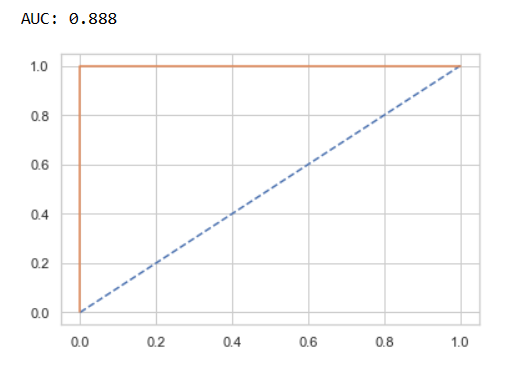
**Classification report of the training set**



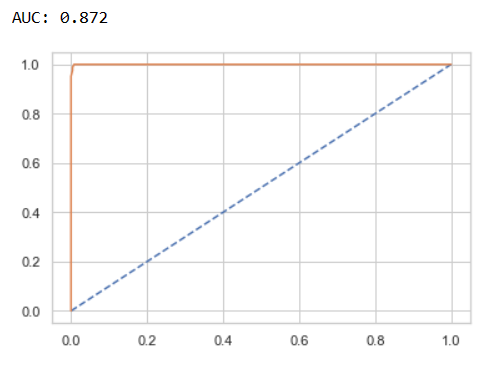
**Classification report of the testing set**



**ROC curve of the training dataset**



**ROC curve of the testing dataset**



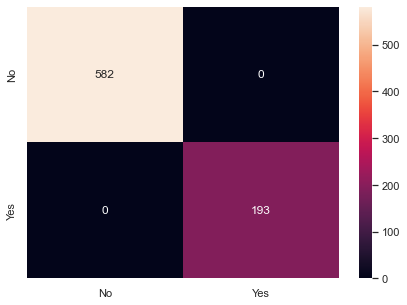
Based on above implementation, it has been summarised that 100% accuracy rate are defined as the train data. The test data accuracy score is defined as 98%. Compared to the default model's accuracy ratings of 100% and 96% for the train and test datasets, respectively, the F1 values were comparable. The accuracy of the data from the test and the train is identical. This model, we can say, is a right fit model.

## KNN

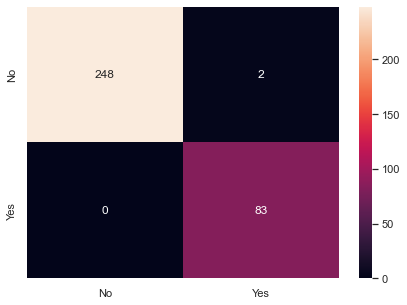
KNN is simple to comprehend and use, making it a fantastic place to start when learning about machine learning algorithms. However, it has certain drawbacks, including sensitivity to noisy data and poor performance on high-dimensional data. It can also be computationally expensive for massive datasets because it needs to determine how far the new point is from every training data point. KNN is a flexible algorithm that can work well for some datasets and issues, especially when the decision boundaries are simple.

**Building the models to train the datasets**

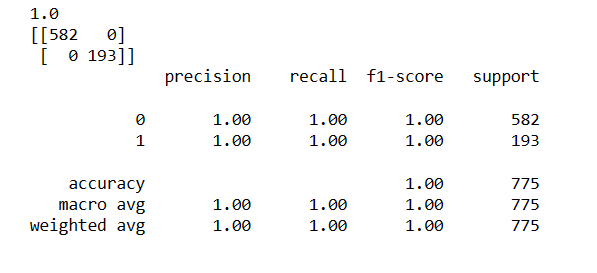
**Confusion matrix of the training set and testing set**



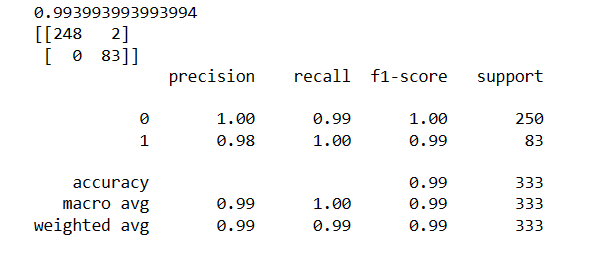
**Confusion matrix of the testing set**



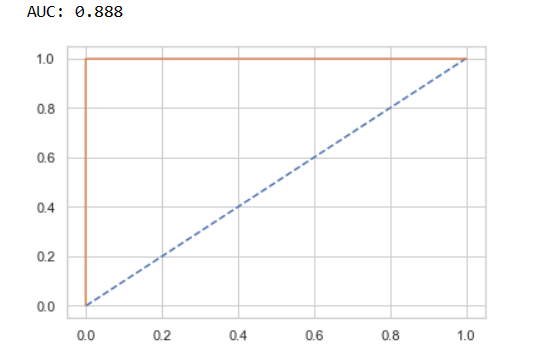
**Classification report of the training dataset**



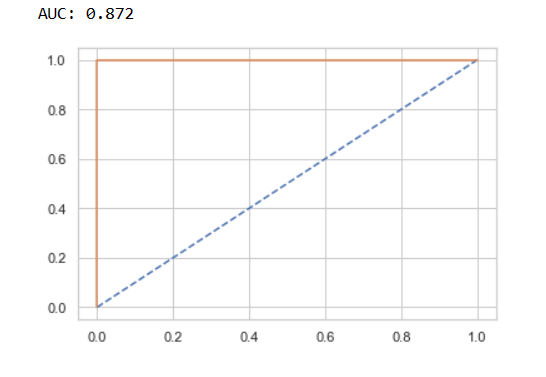
**Classification report of the testing dataset**



**ROC curve based on the training dataset**



**ROC curve based on the testing dataset**



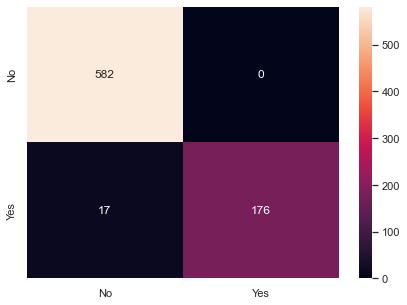
Based on above implementation, 100% accuracy rate is defined as the train data. The test data accuracy score is defined as 99%. Also, The F1 scores were comparable to the default model's 100% and 99% accuracy values for the Train and Test datasets, respectively. The precision of train and test data hardly differ from one another. Therefore, it has been concluded that this is a good fit model.

## Gradient boosting

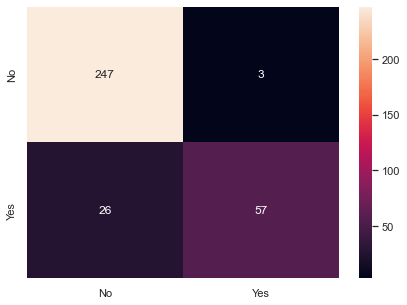
This machine learning technique has offered different insights based on yearly average views and family size on the vacation page.

**Building the models to train the datasets**

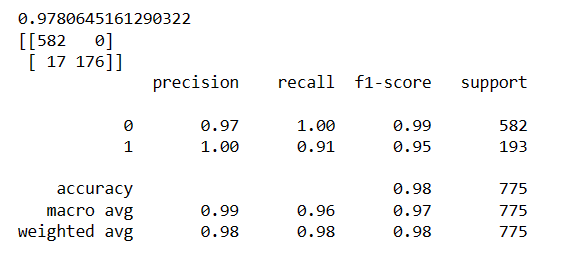
**Confusion matrix of the training set**



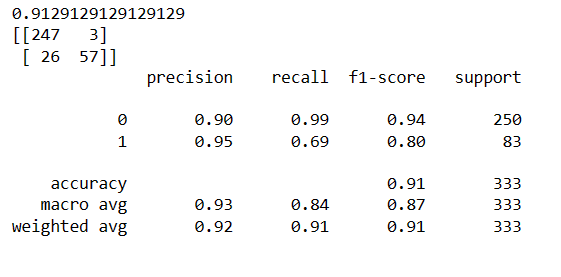
**Confusion matrix of the testing set**



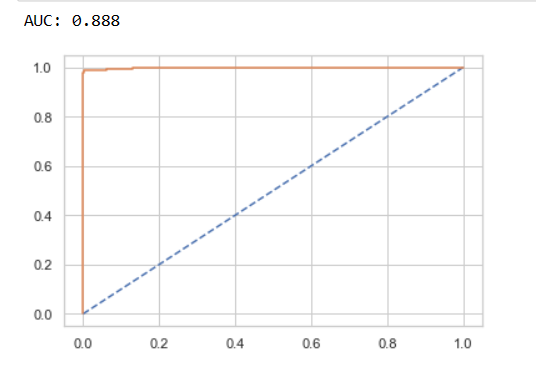
**Classification report of the training dataset**



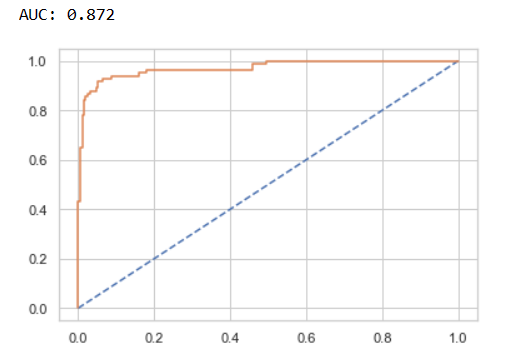
**Classification report of the testing dataset**



**ROC curve of the training dataset**



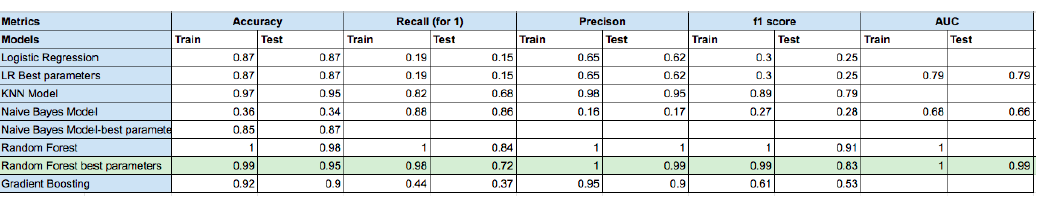
**ROC curve of the testing dataset**



Based on the above implementation, the accuracy score for train and test data is impressive; there is a slight discrepancy between the two. We may state that this model is a suitable fit model because the accuracy scores are high.

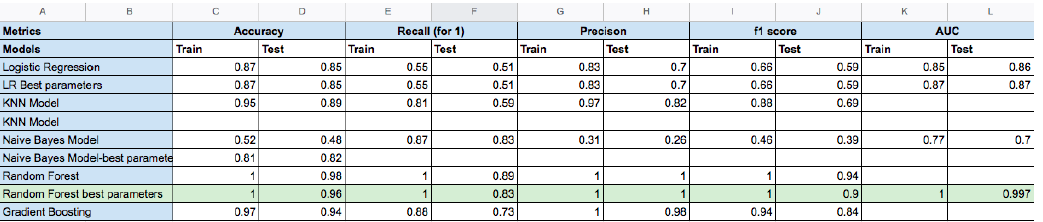
# Performance media for mobile

The performance has been tested based on testing and training datasets to compare the individual models.



# Performance media for the laptop

The performance has checked the performance of the predictions based on testing and training datasets to compare the individual models.



# Ensemble modeling

This technique has combined different best models for producing the predictive model. This model has been assembled to define the average of two different models. The bagging methods are discussed in several of the model performances. The ability of boosting algorithms to increase forecast accuracy by iteratively emphasizing misclassified cases in the training data, as demonstrated by Adaboost and XGBoost, is well known. A possible explanation for the modest drop in accuracy compared to Random Forest is that boosting is more likely to overfit the training data, mainly when the model grows excessively complicated during iterations. The unique problem, dataset, and tolerance level for overfitting influence the decision between the Random Forest and boosting methods in practice. While boosting techniques like Adaboost and XGBoost are practical choices when optimizing performance models and are especially helpful when prioritising accuracy over interpretability, Random Forest still excels at processing various data.

# Model tuning

Modifying hyperparameters optimise a model's performance during model tuning, a critical stage in machine learning. Hyperparameters differ from model parameters because they affect how an algorithm acts during training but are not learned from data. One example of a hyperparameter is the depth parameter in decision trees. Even while some models might perform well without tuning, testing hypotheses through experimentation is critical. This suggests that the model's basic architecture may not require further fine-tuning, saving computational resources and demonstrating that the model's design is appropriate for your particular dataset.

# Interpretation

Based on the above implementation, Model-building process has defined as the optimum model. The curiosis code has been used for making the predictive value. Also, random for his model has used as the optimum model where that you this is code is defined as a 95% according to the testing data. According to the accuracy value of 95%, the product value has been predicted according to the customers. The data's findings provide up important prospects for the aviation company. First off, the significant association between the overall number of likes on outstation check-ins, the yearly average views on the travel page, and the daily average minutes spent on the page shows how important page activity and social engagement are for boosting user engagement. Also, even though most customers don't follow the business page, there is still a huge opportunity to increase user engagement and loyalty. To turn these people into followers, engaging content and marketing methods could be developed. Understanding user check-in habits and geographical choices can also help direct targeted advertising campaigns. There is a demand for mobile-optimized and location-specific advertisements given the prevalence of mobile devices and the appeal of beach and financial destinations. Also, Random Forest stands out as the most accurate and best predictive model. By maximizing advertising expenditure and enhancing the effectiveness of marketing initiatives, Random Forest implementation for forecasts can result in cost savings.

# Recommendation

The airline company should use a data-driven strategy to maximize advertising performance. First, since outstation check-ins are a crucial aspect, they should focus on clients who haven't checked in in a while. Also, targeting customers who stay on the travel page longer, particularly those using mobile devices, can increase engagement and rate of conversion. Adults play a significant role in the purchasing decision, thus careful targeting should concentrate on this group. Digital advertisements that are based on the online behaviour of these clients with high purchase propensity and informed by data from the networking platform can be quite successful. Also, the Random Forest model, which has shown great predictive accuracy, should be utilized to discover prospective customers, optimizing ad expenditure and producing results at a reasonable cost.

# Reference list

Drummond, C., McGrath, H. and O'Toole, T., 2023. Beyond the platform: Social media as a multi-faceted resource in value creation for entrepreneurial firms in a collaborative network. Journal of Business Research, 158, p.113669.

He, H., 2022. A comprehensive review on the role of online media in sustainable business development and decision making. Soft Computing, 26(20), pp.10789-10803.

Keke, M.E., 2023. Impact of digital marketing on the aviation industry. In E3S Web of Conferences (Vol. 402, p. 02009). EDP Sciences.

Ozuem, W. and Willis, M., 2022. Influencer Marketing. In Digital Marketing Strategies for Value Co-creation: Models and Approaches for Online Brand Communities (pp. 209-242). Cham: Springer International Publishing.

Ren, S., Karimi, S., Velázquez, A.B. and Cai, J., 2023. Endorsement effectiveness of different social media influencers: The moderating effect of brand competence and warmth. Journal of Business Research, 156, p.113476.

Saboune, F.M.F., 2022, November. Virtual Reality in Social media marketing will be the new model of advertising and monetization. In 2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 1-7). IEEE.

Shrivastava, A., Jodhana, L.S., Chourasia, S. and Verma, B.L., Analyzing the Effect of Social Media Influencers on the Purchase Decision of Consumers. GURUGRAM UNIVERSITY BUSINESS REVIEW (GUBR), p.34.