TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NO.:T010/075

International Air Passenger Demand Modeling and Airport Choice Modeling in Nepal

by

Nishub Luitel (PUL075MSTrE010)

A THESIS

SUBMITTED TO THE DEPARTMENT OF CIVIL ENGINEERING

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE IN TRANSPORTATION ENGINEERING

DEPARTMENT OF CIVIL ENGINEERING

LALITPUR, NEPAL

December, 2023

COPYRIGHT

Head

Department of Civil Engineering

Pulchowk Campus, Institute of Engineering

Lalitpur, Nepal

TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

DEPARTMENT OF CIVIL ENGINEERING

The undersigned certify that they have read, and recommended to the Institute of Engineering for acceptance, a thesis entitled **"International air passenger demand modeling and airport choice modeling in Nepal"** submitted by **Nishub Luitel** in partial fulfilment of the requirements for the degree of Master of Science in Transportation Engineering.

Supervisor, Dr. Pradeep Kumar Shrestha
Assistant Professor
Department of Civil Engineering

Co supervisor, Mr Ashok Kumar Shrestha
Deputy Team Leader/Aerodrome Engineer,CSC
Air Transport Capacity Enhancement Project

External Examiner: Saroj Kumar Pradhan
Technical Advisor
National Road Safety Council

Committee Chairperson, Anil Marsani
Coordinator, M.Sc. in Transportation Engineering
Department of Civil Engineering

Date..

ABSTRACT

Air travel demand forecasting is an important task for the concerned civil aviation authority as well as airlines. This paper intends to analyze and forecast international air travel demand in Nepal and apportion the demand between TIA, the only existing international airport of Nepal and the other proposed international airports. Econometric variables like GDP, CPI, remittance, employment migration, tourist arrivals, exchange rates, GDP per capita, net national income per capita, world GDP etc were taken as an explanatory variable for the demand generated in aviation industry. A regression model was developed using above mentioned variables as explanatory variable. The statistical result showed almost perfect correlation between themselves, as suggested by variance inflation factor (VIF) value which made the model biased in terms of coefficients. Furthermore, several models considering different combinations of independent variables were developed. Finally a regression model considering exchange rate, number of labour permit, number of tourist arrival as explanatory variable showed satisfactory result and was taken as demand model for the forecast.

Airport choice model between the existing Tribhuwan international airport and the proposed Nijgadh international airport for international air passengers was developed. Airport access time, airport access cost and announced delay at the airport was taken as attributes to define the characteristics of airport. Binary logistic regression was performed to the data collected from passenger survey taken at the check in of Tribhuwan international airport. Announced delay was found to be the most significant variable while other variables airport access time and airport access cost were statistically significant but had lesser impact.

Keywords: Demand modelling, SP survey, Airport choice model, econometric variables, binary logistic regression

ACKNOWLEDGEMENT

The preparation of this proposal required a lot of guidance and assistance from many people and I am extremely fortunate to have got this.

I am grateful to our Program Co-ordinator – Asst. Professor Mr. Anil Marsani, Asst. Professor Dr. Pradeep Kumar Shrestha and Asst. Professor Dr. Rojee Pradhananga, for their support, guidance and motivation.I would like to thank Mr. Ashok Kumar Shrestha, Deputy team leader ATC enhancement project for his guidance and support.

Name: Nishub Luitel

Roll no: 075/MSTrE/010

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

PAX	Number of passengers
CAAN	Civil Aviation Authority of Nepal
IATA	International Air Transport Association
SP	Stated Preference
GDP	Gross Domestic Product
CPI	Consumer Price Index
TIA	Tribhuvan International Airport
SE	Standard Error
RMSE	Root Mean Square Error
PPP	Purchasing power parity
VFR	Visiting friends and relatives
EM	Employment migration
MLR	Multiple linear regression
SPSS	Statistical package for social sciences
IVs	Independent Variables
ROC	Receiver operating characteristic curve
MOCTCA	Ministry of culture, tourism and civil aviation
MOLESS	Ministry of labour, employment and social security
LCU	Local currency unit
USD	United States dollar

CHAPTER 1: INTRODUCTION

1.1 Background

Transportation plays a pivotal role in the progress of any nation, as evidenced by historical records indicating that countries with well-established transportation systems have experienced accelerated development. The globalized nature of today's world has led to rapid growth in the aviation industry, surpassing other modes of transportation. In this dynamic market, meticulous infrastructure planning is crucial, necessitating the forecasting of passenger demand to facilitate the development of essential facilities such as airport terminal buildings, runways, and aprons.

Forecasts are essentially educated predictions of future aviation activities, supported by thorough assessment and analysis. They serve as the foundation for making informed decisions about terminal plans. Given the impact of air travel on transportation networks and the environment, accurate forecasts of future air travel demand and a comprehensive understanding of the factors influencing it are vital components in shaping effective transportation policies.

The surging demand for air travel has brought Tribhuwan International Airport (TIA) to a point where its capacity to manage air traffic is nearly saturated. While efforts have been made to expand TIA's runway and other facilities, technical constraints limit further expansion. Decisions regarding capacity expansion are challenging, as expansion may not always be the optimal solution. The aviation sector is intricately linked to spatial development, and airports within the TIA network are interconnected in terms of socioeconomic factors and passenger movements. Consequently, the case for capacity expansion relies not only on the overall growth in air traffic but also on how it is distributed among alternative airports within the region.

In multi-airport areas, analyzing the potential impacts of capacity or policy changes on air travelers' airport choices has become a crucial aspect of long-term transport strategies. The shift in demand between airports not only affects the commercial viability of individual airports but also has significant repercussions on support structures such as auxiliary businesses and seemingly unrelated sectors like local hotels. The surge in demand necessitates improvements in surface access facilities, potentially exacerbating ground transport congestion in metropolitan areas, such as Kathmandu metropolitan city in our case.

Conversely, a decrease in demand can have equally substantial effects, impacting the survival of local businesses and related jobs, such as taxi companies, which often owe their existence to the success of the airport. For instance, a decline in demand for local transport services due to reduced air traffic may lead to significant job losses or even render purpose-built business structures, like a new bus terminal, partially obsolete. Furthermore, the rise in traffic has resulted in high levels of congestion in some flight corridors, particularly in the airspace surrounding large multi-airport regions, underscoring the need for careful consideration of the effects of capacity changes on airspace congestion.

Observing the trend of increasing passenger demand at TIA, the establishment of an alternative airport appears imperative to accommodate this demand. Therefore, our study aims to predict the demand for international air passengers at TIA in the coming years. It focuses on developing an air passenger demand model using socioeconomic variables that influence demand. Additionally, the study aims to analyze the impact of a new airport to be constructed as an alternative to TIA, specifically the airport at Nijgadh. The approach involves studying the choice behavior of travelers regarding airports, utilizing a stated preference survey to collect data and develop an airport choice model.

1.2 Problem Statement

Nepal has seen continuous growth in both domestic and international air passengers. TIA, being the only international airport till the date have almost reached its saturation and as a counter measure a new airport at Nijgadh have been proposed.

Development of new airport or capacity enhancement of the existing airport requires huge amount of investment and is a matter of concern in the national stage. So, deciding whether or not to carry out project, proper forecasting of passenger demand is necessary. Also, in case of TIA and Nijgadh airport, both airport location not being too far, choice of the passengers to choose either airport comes into play which affects the demand and capacity of both airports. So, to understand the demand distribution between airports, a proper airport choice model is necessary.

1.3 Objective

This thesis study aims to develop an international airlines passenger prediction model for TIA using multiple regression model and Airport choice model using binary logistic regression. The specific objectives are:

- To develop demand forecasting model for international air traffic for TIA.
- To develop airport choice model between TIA and Nijgadh international Airport.

1.4 Limitations

New international airports in Pokhara, Bhairahawa are recently under operation and airport at Nijgadh is proposed. There is no pre-existing data considering international passenger flow. So, the consideration of impact of these airport on TIA could not be incorporated in both demand forecasting modelling and airport choice modeling.

For demand forecasting model, due to the availability of passenger movement data on CAAN's database from 1998 onwards only, large sample could not be achieved in the modelling.

1.5 Organization of Report

This report is organized in five chapters as described below;

Chapter 1: Introduction

It generalizes background, problem statement, objective and limitations of the study. The need of study and the contribution of the findings from the study is briefly described in this section.

Chapter 2: Literature Review

The review of articles and journals relevant to the study were described briefly in this section. International and national level papers related to demand forecasting, stated preference survey and choice modelling were reviewed throughout the study and important findings needed for the study was used.

Chapter 3: Methodology

The methods adopted for analysis, data collection, model development was discussed in this section. Typically, two models i.e. demand forecasting model and airport choice model was developed. Detailed procedure for model development has been described in this section.

Chapter 4: Data analysis

The data and results from the developed model have been discussed in this section.

Chapter 5: Conclusion and Recommendations

The conclusions from the study and recommendation according to the result were presented in this section.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Many researches have been carried out on the topic demand forecasting, airport choice modelling, and stated preference survey design. The literatures required for the thesis work were reviewed thoroughly and are described below.

2.2 Demand forecasting

(Wang & Song, 2010) conducted a concise overview of global air travel demand from 1950 to 2008, offering a comprehensive review of existing studies in this field. The review focused on the progress of research, publication sources, geographical concentration, catalysts for air travel demand, methodologies for demand modeling and prediction, analysis of demand flexibility, and the connection between air travel demand and leisure and tourism. The primary objective was to critically evaluate current research on air travel demand, providing guidance for scholars interested in exploring this specific subject area.

(Rengaraju & Arasan, 1992) conducted extensive studies involving the calibration of a city pair model for domestic air travel demand based on 40 city pairs in India. The model, calibrated using combined cross-sectional data through multiple linear regression analysis, underwent validation through cross-validation and a backward prediction method. Results indicated a significant increase in total air travel demand across all 40 city pairs in the study in 2001, assuming a 50% increase in service frequency. The study also emphasized the need to segment the demand model based on travel distance and city size within the study region.

(Karlaftis et al., 1996) established a time series model with econometric independent variables to forecast international passengers at Frankfurt airport. The paper aimed to assess the predictive capability and forecasting precision of air-travel demand models, using statistical data from Frankfurt and Miami International Airports. Findings suggested that simpler models with fewer independent variables performed as effectively as more complex ones, highlighting the significant influence of external factors on air-travel demand. Multicollinearity challenges were noted in models with multiple explanatory variables.

(Bastola, 2017) developed an econometric model for estimating passenger air transport demand in Nepal, linking it to GDP and the number of tourist arrivals. (Erraitab, 2016) formulated a regression model for air travel demand forecast in Morocco, highlighting the impact of socioeconomic characteristics of countries of origin on demand. (Duwal, 2012) developed an air travel demand model for TIA using regression analysis, trend analysis, and exponential smoothing, with exponential smoothing proving to be the most effective.

(Seraj et al., 2001) created various models for air travel demand using different combinations of explanatory variables through stepwise regression. The focus was on forecasting international air travel demand in Saudi Arabia, with population size and total expenditure identified as primary determinants.

(Dargay & Hanly, n.d.) developed a model to analyze the demand for air travel to and from the UK, considering factors such as income, airfares, foreign trade, exchange rates, and domestic price levels. Results indicated price sensitivity in leisure travel among UK residents and highlighted income as a key driver across all cases. The review of existing literature revealed diverse models linking travel demand to independent variables, typically socioeconomic measures, with R2 commonly used to assess forecasting quality.The results presented in this paper indicated that the demand for international air travel among UK residents for leisure purposes displayed a relatively high sensitivity to price changes, suggesting that a portion of the observed increase in air travel over the past decade could be attributed to declining airfares. On the other hand, business travel seemed to be less responsive to fare fluctuations and was primarily influenced by other factors, particularly foreign trade. For non-UK residents, the price sensitivity appeared to be similar for both leisure and business travel. In all cases, income (or trade) emerged as a key driver of air travel demand, although relative price levels and exchange rates also played significant roles. Additionally, the study revealed substantial lagged effects, with short-term elasticities being considerably smaller than long-term elasticities.

The review of the existing literature discovered that a range of models have been developed for the forecasting of air-travel demand at airports. All the models cited established an logical relationship between travel demand and a set of independent variables. These variables are usually socioeconomic measures, expressing the attraction for air-travel. A common measure of effectiveness used to evaluate the forecasting quality of the various models is the R^2

2.3 SP survey and choice modeling

(Cherchi & Hensher, n.d.) conducted a workshop addressing challenges in the design and implementation of Stated Preference (SP) surveys, with a focus on reviewing current practices and suggesting corrections. The research explored recent advancements in SP models, highlighting their growing significance in demand modeling and prediction. Key topics included a comparative evaluation of choice and preference models, the importance of considering scaling when integrating various data types, the attractiveness of SP data in enriching revealed preference models, the use of hierarchical designs for complex experiments, and techniques for accommodating dynamics within SP modeling.

(Jiang et al., 2014) conducted a survey on self-driving and bus commuting, developing a Binary Logit (BL) model for commute mode choice. The study identified 11 influential factors, including gender, age, and personality traits, and calibrated the BL model using data from 280 commuters, validating it with another dataset of 106 commuters. Findings indicated that factors such as age, education, and monthly income influenced self-driving choice, while bus choice was sensitive to travel time, trip frequency, and peak-hour congestion. The model demonstrated high accuracy in estimating values.

(Johnston et al., 2017) presented contemporary best-practice suggestions for stated preference (SP) studies used in decision-making, drawing on collective knowledge from peer-reviewed literature. The suggestions covered various SP techniques for estimating values related to both usage and non-usage scenarios, including contingent valuation and discrete choice experiments.

(Gautam, n.d.) developed a model to calculate the Value of Travel Time (VTT) in the Kathmandu Valley. Using the Revealed Preference/Stated Preference (RP/SP) methodology and employing a multinomial logit model for RP data and an uncorrelated mixed logit model for SP data, the study determined a VTT of Rs. 114.65 per hour.

(Abuhamoud, 2011) aimed to understand the mode choice behavior of car users in Tripoli, employing a binary logistic model to analyze factors influencing the choice between private cars and government buses. The findings suggested measures to encourage car users to shift towards public transportation.

(Alkabaa & Taylan, 2021) introduced an efficient approach for crafting a stated-choice survey, utilizing an orthogonal array design to collect data aimed at enhancing inventory accuracy within electricity company warehouses. The study selected eight factors with two levels each and employed a stated-choice factorial survey with four alternatives, addressing inventory accuracy issues.

2.4 Orthogonal design

(Zurovac & Brown, n.d.) provided insights into the utility of orthogonal designs, emphasizing key considerations in their design and implementation. They highlighted the significance of employing orthogonal designs in quick-cycle comparative effectiveness research, particularly for rigorously and simultaneously testing various aspects during the delivery of complex interventions. This approach is crucial as current policy research often focuses on the effectiveness of broad concepts without delving into the detailed implementation of interventions. The flexibility of orthogonal designs was underscored, allowing researchers to decide whether to test multiple interventions concurrently or a smaller number with greater precision.

An orthogonal design is a meticulously crafted survey design that ensures the independence of predictor variables, avoiding multicollinearity issues in regression analysis. Essentially, orthogonal designs aim to minimize correlations between predictor variables, facilitating the interpretation of each predictor's effects on binary outcomes.

The advantages of employing an orthogonal design in binary logistic regression for stated preference surveys are manifold:

Interpretability: By guaranteeing independence among predictor variables, logistic regression coefficients become more interpretable. Researchers can articulate clear statements about how each predictor influences the likelihood of choosing one alternative over another.

Efficiency: Orthogonal designs optimize the allocation of survey responses, allowing researchers to extract maximum information from a minimal number of survey questions. This enhances the efficiency of both data collection and analysis.

Reduced Bias: Orthogonal designs contribute to the reduction of bias in estimating logistic regression coefficients, resulting in more accurate and reliable results.

Generalizability: Results obtained from orthogonal designs are more likely to be applicable to the broader population of interest. The design's emphasis on minimizing biases and maximizing efficiency enhances the generalizability of findings.

2.5 Airport choice

(Harvey, 1987) formulated a multinomial logit model to examine air travel behavior in selecting departure airports within a region with multiple airports. Using data from a 1980 survey of air passengers in the San Francisco Bay Area, separate multinomial logit models were developed for business and non-business travelers. The study emphasized that ground access time and the frequency of direct air service to the chosen destination played a pivotal role in shaping airport usage patterns. The key conclusion drawn was that a straightforward logit model, relying on two variables - airport access time and flight frequency to the chosen destination - provided a reasonably accurate representation of airport selection behavior, particularly among business travelers. Notable differences were observed in the decision-making behavior of non-business travelers, with flight frequency and access time holding less significance for this group.

(Windle & Dresner, 1995) constructed a logistic model for predicting airport choice among passengers, using passenger data from the Washington, D.C./Baltimore area. The study identified that airport access time and flight frequencies were crucial factors in forecasting airport choice. It observed that reduced access time and increased flight frequencies were more critical for business travelers compared to non-business travelers. Further analysis focused on passengers within reasonable proximity of multiple airports, revealing that the significance of access time diminished while flight frequencies gained importance in this subset. The inclusion of variables related to a passenger's prior experience with a specific airport proved to be statistically significant, indicating a tendency for passengers to stick with familiar airports.

(Hess et al., 2007) conducted a Stated Preference (SP) survey to model airport and airline choice behavior in the U.S., incorporating airfare, access time, flight time, and airline and airport allegiance as survey attributes. The findings highlighted a greater willingness among business travelers to accept higher fares for shorter access times compared to holiday or VFR travelers. Business travelers also showed a higher inclination to invest in reducing schedule delays and improving on-time performance. Distinct population groups exhibited differences in willingness to pay for features such as early arrivals and elite frequent-flier accounts, with business travelers being more responsive in this regard.

(Marcucci & Gatta, n.d.) developed a choice model using Stated Preference (SP) data focused on a multi-airport region in central Italy with four competing airports. The research aimed to assess the robustness of previous studies in regions with differing characteristics and to estimate significant attributes influencing the choice of origin airport. The study emphasized the importance of diversifying and validating results across various institutional, regulatory, and demand contexts. The approach utilized SP data, implemented a traditional segmentation approach, and introduced the concept of cut-offs to differentiate between compensatory and non-compensatory decisions.

(Saffarzadeh et al., n.d.) explored the airport selection behavior of air travelers in the Tehran multi-airport region using a Binary Logit model and Stated Preference (SP) data. The study, conducted through direct personal interviews in May 2011, revealed the significant impact of factors such as flight frequency, public access options to airports, and airport taxes (airfares) on airport choice.

CHAPTER 3: METHODOLOGY

3.1 Framework of the study

Initially literatures related to the topics like demand forecasting, choice modelling, stated preference survey techniques etc were studied and relevant information from the previous literature were noted. After having general idea about previous literature data collection was done which was then analysed to develop the required model.

The methodological framework of the study is illustrated in the Figure 3.1.

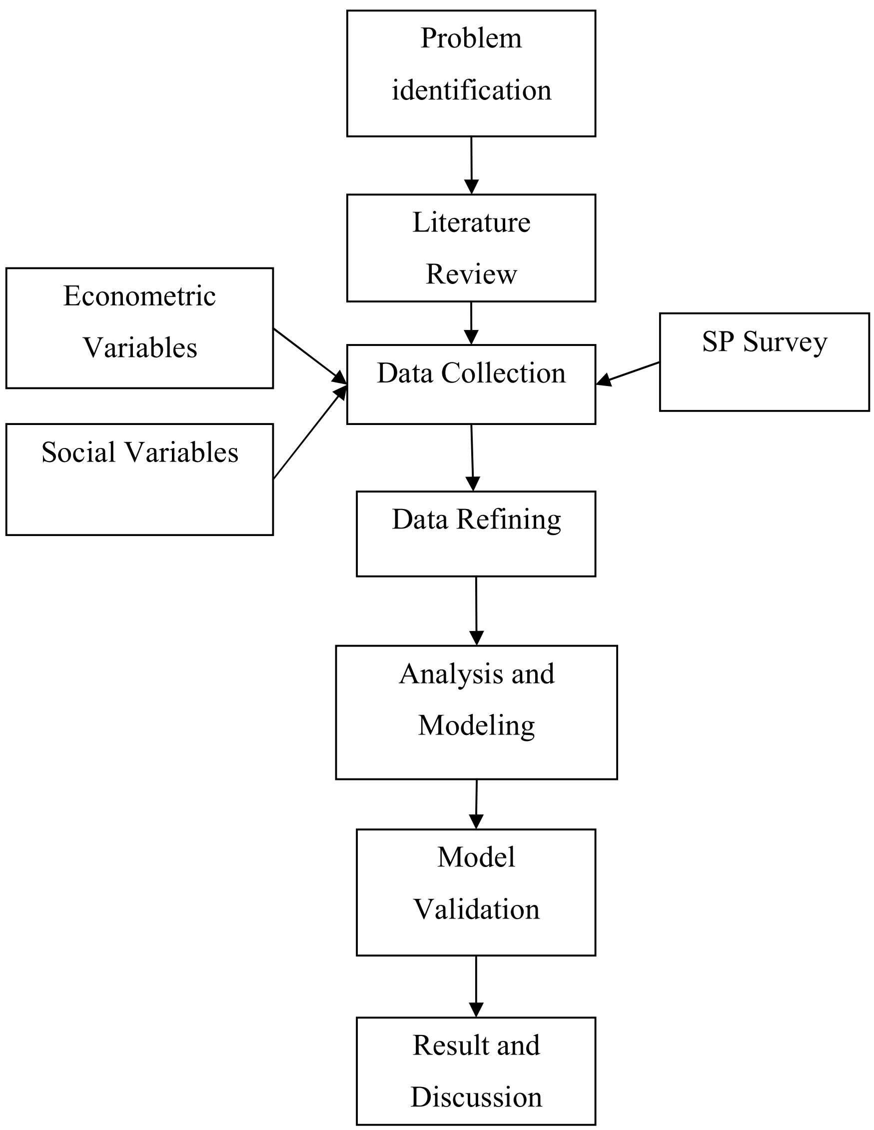

Figure 3.1 : Framework of the study

3.2 Study Area

For airport choice modelling study population was departing passengers in TIA check in area. SP survey was carried out to assess if people are willing to choose TIA or Nijgadh airport for departure given choice sets based on attributes.

3.3 Sample Size

For airport choice modelling sample size was determined using following formula (Dios Ort´uzar & Willumsen, 2011)

Where,

$n = \bar{n}/(1+ \bar{n}/N)$

N is total population

n is sample size from finite population

$\bar{n}$ is sample size from infinite population Sample size for infinite population is calculated using:

$\bar{n} = S^2/se(\bar{x})^2$ ……………………………………………………….3.1

Where,

S^2 = variance of population

$se(\bar{x})$ is standard error of sampling population

For 95% confidence level, sample size is calculated as:

$se(\bar{x})^2 = 0.1\mu /1.96 = 0.051\mu$

$\bar{n} = S^2/ (0.051\mu) 2 = 384cv^2$ ……………………………………………………….3.2

Taking coefficient of variation as 1 and $\bar{n}/N$ being very small, minimum sample size is taken as 384 which was collected for perception survey. A total of 403 observations from Passengers at TIA check in terminal were collected in perception survey.

3.4 Data Collection

The data required for demand modelling was extracted from the published report of several institutions. Historical PAX movement data in TIA have been collected from the reports published by CAAN. Similarly, Econometric variables like GDP, CPI, exchange rates, PPP, required for the analysis have been extracted from the reports published by the world bank. Data from the year 1998 to 2019 have been used in the analysis.

Similarly data required for airport choice modelling was collected by conducting SP survey.. An official recommendation from the campus was submitted to TIA office and entry was allowed up to check in area of TIA where printed forms were distributed to commuting passengers and one to one survey was conducted. Following information were gathered from the survey.

- Origin of the respondent
- Destination of the respondent
- Purpose of travel
- Access cost
- Access time

3.5 Air Passenger Demand Modeling

3.5.1 Identification of Independent Variables

In forecasting air-travel demand, it is of critical importance both to ensure the statistical validity of the models, and to select the models with the higher potential forecasting accuracy. In this study several variables that might affect the air travel demand has been considered and analysis was done to select the best model which is statistically valid and have a good forecasting power. Following variables were taken as explanatory variable for demand forecasting model and are described briefly below.

a. **GDP (gross domestic product)**:It is the total monetary or market value of all the finished goods and services produced within a country's borders in a specific time period. The GDP is the total of all value added created in an economy. The value added means the value of goods and services that have been produced minus the value of the goods and services needed to produce them. Historic data of GDP for Nepal, World and GDP per capita was extracted from the database of the world bank and measurement unit was in USD. Three types of GDP was taken in consideration assuming the economic indication of Nepal by the variables GDP(Nepal) and GDP per capita(Nepal) and of other countries by the variable GDP(World).

b. **CPI(consumer price index)**: The Consumer Price Index (CPI) is a measure that examines the weighted average of prices of a basket of consumer goods and services, such as transportation, food, and medical care. The CPI is calculated by taking the cost of the basket in a given year and dividing it by the cost of the same basket in the base year, then multiplying by 100. The changes in the CPI were used to assess price changes associated with the cost of living. 2010 was considered base year and CPI being ratio was unit less. The formula is:

$$CPI=\frac{cost\ of\ basket\ in\ current\ year}{cost\ of\ basket\ in\ base\ year} * 100 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.3$$

c. **Labour permit**: It gives the number of Nepali Citizen who has been given authority to work in foreign countries. Required data was collected from the report published by MOLESS. The unit for measurement was number of person.

d. **Remittance**: Remittances are funds transferred from migrants to their home country. They are the private savings of workers that are spent by their families in the home country for food, clothing and other expenditures, and which drive the home economy. Data for remittance was extracted from the database of the World Bank. The unit for measurement was USD.

e. **Number of tourist arrivals**: It is the total number of foreign citizens who has arrived to Nepal as a tourist by both land air. Required data was collected from the report published by MOCTCA. The unit for measurement was number of person.

f. **Purchasing power parity (PPP) conversion factor**: It is a spatial price deflator and currency converter that controls for price level differences between countries, thereby allowing volume comparisons of gross domestic product (GDP) and its expenditure components. This conversion factor is for household final consumption expenditure. The unit for measurement was unit less.

g. **Official exchange rate**: It refers to the actual, principal exchange rate and is an annual average based on monthly averages (local currency units relative to U.S. dollars). Exchange rate brings the economic factor of other countries along with Nepal. Required data was collected from the database published by the World Bank. The unit for measurement was USD.

Demand = f [GDP,CPI, Remittance, number of labour permit, number of tourist arrivals, exchange rate, PPP]

GDP = Gross domestic product

CPI= consumer price index

PPP=purchasing power parity

In this study, a hypothesis was made assuming air passenger demand is the function of above mentioned variables i.e. GDP, CPI, remittance, number of labour permit, number of tourist arrivals, exchange rate, PPP.

3.5.2 Regression model development

Regression analysis is employed to establish the connection between a dependent variable and one or more independent variables. In the context of this study, the analysis involves multiple linear regression, which is applied when there are multiple independent or explanatory variables. The mathematical representation of multiple regression takes the form:

$$y = a + b * x_1 + c * x_2 + \cdots + \mathcal{E} \ldots\ldots\ldots\ldots\ldots\ldots\ldots .3.4$$

y = Dependent variable

x1, x2= Independent variables

a = Intercept

b, c = slope

Ɛ = Residual error

3.5.3 Criteria for Model Selection

The R2 value, known as the coefficient of determination, indicates the proportion of the total variance in the dependent variable that is elucidated by the independent variable. This metric serves as an assessment of the goodness of fit of the function to the data. It is computed by comparing the sum of the squares of the residuals (the differences between predicted and actual values) and the sum of the squares of regression. The sum of the squares of the residuals reflects the discrepancy between the estimated and actual data, akin to the sum of the squares within ANOVA.

For a given value of X1, the regression line predicts the dependent variable as (Y_1)̂. While the regression line accounts for a portion of the deviation of this observation from the mean, specifically explaining the proportion (Y_1)̂-Y ̅, it is not flawless. There exists a discrepancy, represented by Y_1-(Y_1)̂, which remains unexplained by the regression equation. Consequently, the deviation of Y1 from the mean can be categorized into following components.

$Y_1 - \bar{Y}$ = Total deviation of Y_1 from the mean

$\widehat{Y_1} - \bar{Y}$= Explained deviation from mean

$Y_1 - \widehat{Y_1}$= Unexplained deviation of Y_1

If we square and sum up for each observation, we obtain the complete components of variation for the response variable.

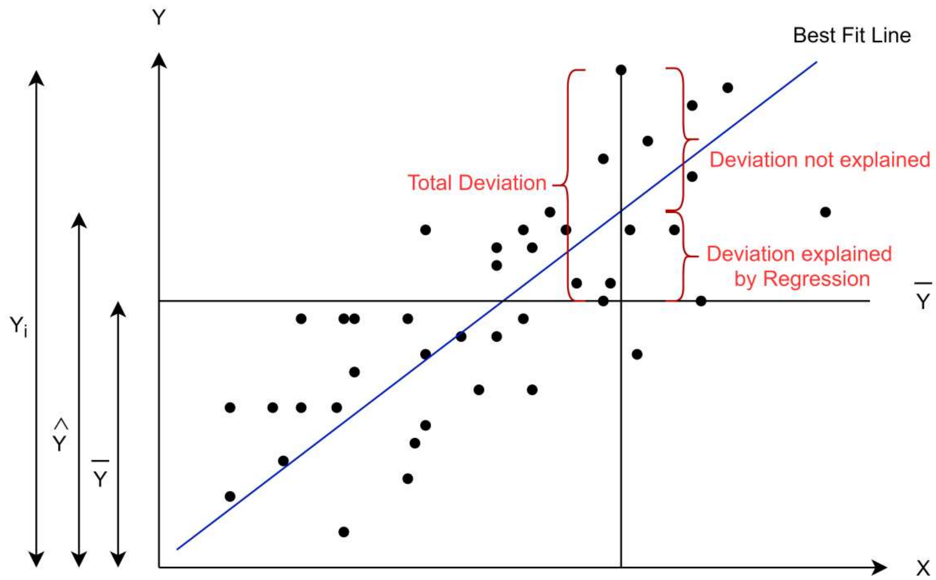

Figure 3.2:Scattered diagram for unexplained variation

TSS – Total Sum of Squares = $\sum_i (Y_i - \bar{Y})^2$

MSS – Model Sum of Squares = $\sum_i \left(\hat{Y} - \bar{Y}\right)^2$

RSS – Residual Sum of Squares =$\sum_i (Y_i - \bar{Y})^2$

The R^2 value is calculated as:

$$R^2 = 1 - \frac{Unexplained\ variation}{Total\ variations} = 1 - \frac{RSS}{TS} \ldots\ldots\ldots\ldots.3.5$$

The *"t"* statistic corresponding to a particular coefficient estimate is a statistical measure of the confidence that can be placed in the estimate. The *"t"* statistic is obtained by dividing the value of the coefficient by its standard error. The higher the value of *"t"*, the greater is the statistical significance of the relationship between the explanatory variable and the dependent variable, and the greater is the confidence that can be placed in the estimated value of the corresponding coefficient.

" t" value =x/SE……………………………………..……………..3.6

Where, *x* is the estimated value of the coefficient of the explanatory variable, and *S.E.* is the Standard error of the estimated value of the coefficient.

3.5.4 Model validation

Multiple linear regression is built on several key assumptions, and validating these assumptions is crucial to ensure the reliability and accuracy of the model. Some important assumptions of multiple linear regression and methods for their validation is given below.

- **Linearity Assumption:**

Assumption: The relationship between the dependent variable and the independent variables is linear. This means that the change in the dependent variable is proportional to the change in each independent variable.

Validation:

Scatter plots: Creating scatter plots of the dependent variable against each independent variable helps to visually check for linearity.

- **Independence Assumption:**

Assumption: The residuals are independent of each other. This means that the error in predicting one observation does not depend on the errors in predicting other observations.

Validation:

Durbin-Watson test: A statistical test that checks for autocorrelation in the residuals. Durbin Watson significance table can be used to validate the assumption.

- **Homoscedasticity Assumption:**

Assumption: The variance of the residuals is constant across all levels of the independent variables. In other words, the spread of residuals is the same for all values of the independent variables.

Validation:

Residual plots: Plotting residuals against the predicted values or independent variables can conclude homoscedasticity assumption is met if there is consistent spread of residuals.

- **Normality of Residuals:**

Assumption: The residuals follow a normal distribution.

Validation:

Normal probability plot (Q-Q plot): Examining whether the residuals follow a straight line on a Q-Q plot.

Shapiro-Wilk or Anderson-Darling tests: These are formal statistical tests to check for normality. If the Shapiro- wilk statistic is negative then normality assumption is met.

- **No Perfect Multicollinearity Assumption:**

Assumption: There is no perfect linear relationship among the independent variables. In other words, no independent variable can be exactly predicted by a linear combination of the others.

Validation:

Calculating the Variance Inflation Factor (VIF) for each independent variable. A high VIF (typically > 10) suggests multicollinearity.

3.6 Choice modelling

Choice modelling refers to the distribution of choices of human according to the utilities of available alternatives. In Nepal, the idea behind Nijgadh International Airport was to provide an alternative and alleviate the pressure on TIA, which has been facing challenges related to congestion and capacity limitations. The new airport is expected to handle a significant portion of the international air traffic and contribute to the overall development of aviation infrastructure in Nepal while Pokhara International Airport and Gautam Buddha International Airport were constructed with the purpose to serve tourism. So, in this study alternative to TIA was taken as Nijgadh Airport. To determine the utility of TIA against Nijgadh Airport data was collected using SP survey at check in area of TIA. The data collected from SP survey was processed to develop airport choice model.

3.6.1 SP survey design

Stated preference (SP) survey helps to understand how people behave or interact with a new situation which may be hypothetical scenarios. This survey is especially useful in cases where no real world data exists to make any conclusion. In our study some possible scenario of the proposed international airport with the existing scenario of TIA was given to the respondents asked them to choose best according to their preference.

SP questionnaires were developed for the purpose of which several elements that are related are described below.

a. Alternatives

Respondents were asked to choose between two or more alternatives. In case of Nepal, Gautham Buddha international airport and Pokhara international airport were already constructed but Nijgadh international airport was proposed specifically to serve as alternative to TIA. So, in this study TIA was taken as one alternative and Nijgadh airport was taken as second alternative.

b. Attributes

These are the characteristics possessed by each alternatives i.e. TIA and Nijgadh airport. In this study airport access time, airport access cost and announced delay were taken as attribute. Attributes were further divided into different levels to create scenarios or experiment.

c. Levels

Access time and access cost are easily quantifiable quantities and hence these attributes were taken as numeric variable while announced delay being difficult to quantify, was taken as categorical variable. Since the study was based on hypothetical scenario at Nijgadh International Airport, attributes access cost and access time was compared with that of TIA and three levels were assigned based on equal deviation (greater or lesser or equal) as compared to TIA while announced delay was assigned either less or greater than that of TIA accordingly Following levels were adopted for the stated preference survey.

- Access time
- o 1.5 times of TIA
- o Equal to TIA
- o 0.5 times of TIA
- Access cost
- o 1.5 times of TIA
- o Equal to TIA
- o 0.5 times TIA
- Announced delay
- o Greater than TIA
- o Less than TIA

Questionnaire design

An orthogonal design was adopted for creation of choice sets. Since different attributes with different variables creates large number of choice sets, multicollinearity arises severely among the predictior variables. So, to avoid multicollinearity and to ensure the predictor variables are independent of each other orthogonal design is chosen in this study. In this study orthogonal design was created using SPSS where 9 different scenarios were developed out of which 4 scenarios were taken for survey. Four choice set used for survey is given in Table 3.1.

Table 3.1: Choice sets for SP survey

Choice sets for survey			
Card ID	Access time	Access cost	Announced delay
1	1.5 times of tia	0.5 times of tia	less than tia
2	0.5 times of tia	1.5 times of tia	greater than tia
3	equal to tia	1.5 times of tia	less than tia
4	1.5 times of tia	1.5 times of tia	less than tia

3.6.2 Binary logistic regression model development

Binary logistic regression serves as an extension of simple linear regression, specifically tailored for situations where the dependent variable takes on a dichotomous or binary form. When dealing with a binary dependent variable with only two possible outcomes, such as gender or response status, simple linear regression is unsuitable. Logistic regression, on the other hand, is the statistical methodology employed to forecast the connection between predictors (independent variables) and a predicted variable (the dependent variable) in binary scenarios. Logistic regression requires two or more independent variables or predictors, which can encompass continuous (interval/ratio) or categorical (ordinal/nominal) variables.

1. Log Transformation

Among various types of transformations employed to make skewed data more closely approximate normality, the log transformation stands out as perhaps the most widely utilized. Both log transformations and square root transformations have proven effective in bringing skewed distributions closer to a normal distribution. By applying a log transformation to p values, we create a link with the normal regression equation. The log distribution, also known as the logit of p or logit(p), results from this transformation of p values.

2. Hypothesis Test

In logistic regression, hypotheses are of interest:

the null hypothesis, which is when all the coefficients in the regression equation take the value zero, and

the alternate hypothesis that the model currently under consideration is accurate and differs significantly from the null of zero, i.e. gives significantly better than the chance or random prediction level of the null hypothesis.

3. Likelihood Ratio Test for Nested Models

The likelihood ratio test relies on the -2LL ratio and serves as an examination of the significance of the disparity between the likelihood ratios (-2LL) of the researcher's model with predictors (referred to as model chi-square) and the likelihood ratio of the baseline model containing only a constant. Significance at the 0.05 level or below indicates that the researcher's model, including predictors, differs significantly from the one containing only a constant (where all 'b' coefficients are zero). This test quantifies the enhancement in fit achieved by the explanatory variables compared to the null model. The chi-square statistic is employed to evaluate the significance of this ratio. In large samples, the difference between the two -2LogL values follows a chi-square distribution.

3.6.3 Model Validation

Model training was done using data collected at airport and the model coefficients from the trained model was used to validation data set which was collected about after 7 months in same location. The number of validation data set was about 12.5% of original data set.

Different metrics used to assess the model performance are:

a. Accuracy:

Accuracy is a common evaluation metric that measures the proportion of correctly predicted instances (both true positives and true negatives) out of the total number of instances. It provides an overall view of how well the model performs across all classes.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots 3.7$$

However, accuracy can be misleading, especially when dealing with imbalanced datasets, where one class greatly outweighs the other. In such cases, achieving high accuracy might come from the model simply predicting the majority class. Therefore, accuracy should be considered alongside other metrics.

b. Precision:

Precision, also known as positive predictive value, measures the proportion of true positive predictions among all instances that the model predicted as positive. In other words, it quantifies how many of the predicted positive instances were actually correct.

$$Precision= \frac{True\ positives}{True\ positives+Fa\quad positives} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.8$$

Precision is important when the cost of false positives is high.

c. Recall (Sensitivity or True Positive Rate):

Recall, also known as sensitivity or true positive rate, measures the proportion of true positive predictions among all actual positive instances. It quantifies how well the model captures all positive instances.

$$Recall= \frac{True\ positives}{True\ positives+Fa\quad negatives} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. 3.9$$

Recall is crucial when the cost of false negatives is high. For instance, in medical screenings, high recall indicates that the model is effectively identifying most of the true positive cases.

d. F1 Score:

The F1 score is the harmonic mean of precision and recall. It provides a balanced measure that considers both false positives and false negatives. The F1 score is useful when precision and recall have different importance in the context of your problem.

$$F1= \frac{2*precision*recall}{precision+recall} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. 3.10$$

The F1 score rewards models that have a good balance between precision and recall. It is especially valuable in imbalanced datasets where one class is significantly more prevalent than the other.In summary, accuracy, precision, recall, and F1 score are fundamental metrics for evaluating classification models. Each metric has its own focus and use case, and understanding these metrics helps in assessing the model's strengths and weaknesses in different scenarios.

CHAPTER 4: RESULT AND DISCUSSIONS

4.1 Multiple linear Regression model

4.1.1 Model 1

A model incorporating the variables linked with economic activities was attempted to develop for this study. Going through literatures in the related topic, it was found that several important factors like GDP, income, car ownership, oil export and import, CPI, employment migration etc showed remarkable impact on the air travel demand.

A regression model was developed considering all of the above mentioned variables i.e.(PPP conversion factor, number of labour permit, Tourist arrival, Exchange rate, world GDP, GDP per capita, Remittance, CPI, GDP constant LCU) as an explanatory variable. A regression equation was found with R^2 of 0.998 as given in Table 4.1. It was found that all variables except tourist arrival, number of labour permit and exchange rate was found insignificant (i.e. $p>0.05$) .Besides coefficients of the variables being insignificant, multicollinearity as indicated by VIF value (>5) was found among the predictor variables as given in Table 4.3. Hence, the model was found to be statistically invalid. As a finding from result of first regression model as shown in Table 4.4 variables highly correlated ($r>0.8$) and variables having similar meaning were excluded from the analysis.

Table 4.1 Model Summary-Model 1

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.999[a]	.998	.996	.07090	2.486

Table 4.2 ANOVA-Model 1

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	29.081	9	3.231	642.883	.000[b]
	Residual	.060	12	.005		
	Total	29.142	21			

Table 4.3 Model coefficients-Model 1

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.063	.750		.084	.934		
	GDP_consant_LCU	1.009E-12	.000	.351	1.515	.156	.003	310.523
	GDP_per_capita	-.001	.001	-.372	-2.104	.057	.006	181.651
	Remittance	1.236E-10	.000	.300	1.458	.170	.004	244.813
	Tourist_arrival	1.864E-6	.000	.410	6.595	.000	.045	22.376
	CPI	.011	.006	.444	1.987	.070	.003	288.996
	Exchange_rate	-.022	.008	-.288	-2.625	.022	.014	69.982
	number.of.labour.permit	2.223E-6	.000	.300	3.404	.005	.022	45.030
	world.GDP	-2.279E-14	.000	-.379	-2.033	.065	.005	202.068
	PPP.conversion.facor	.060	.030	.283	1.952	.075	.008	122.129

Table 4.4: Collinearity statistics for all variables

	pax_movement	GDP_consant_LCU	GDP_per_capita	Remittance	Tourist_arrival	CPI	Exchange_rate	number.of.labour.permit	world.GDP	PPP.conversion.facor
pax_movement	1									
GDP_consant_LCU	.978**	1								
GDP_per_capita	.992**	.985**	1							
Remittance	.990**	.988**	.991**	1						
Tourist_arrival	.926**	.902**	.924**	.889**	1					
CPI	.985**	.988**	.985**	.992**	.893**					
Exchange_rate	.893**	.913**	.896**	.916**	.795**	.945**	1			
number.of.labour.permit	.841**	.791**	.826**	.853**	.622**	.810**	.711**	1		
world.GDP	.956**	.948**	.954**	.957**	.851**	.928**	.780**	.893**	1	
PPP.conversion.facor	.987**	.974**	.984**	.991**	.870**	.988**	.913**	.864**	.947**	1

4.1.2 Model 2

A sub model with CPI, number of labour permit, Tourist arrival, and Exchange rate as a predictor variable was fed into MLR analysis and following output was obtained. The model developed had R^2 value of 0.996 as presented in

Table 4.5 which is almost perfect fit for the model.

Table 4.7 showed that the predictor variables CPI, tourist arrival and number of labour permit had a significant p value i.e. 0.000, 0.000, 0.000 (< 0.05) while exchange rate was found insignificant with p value 0.064. Also, issue of multicollinearity was found in the model where an independent variable (CPI) had VIF value of 41.51 and exchange rate had VIF value of 13.756 which is greater than 10 and hence statistical fit of model was not justified.

Table 4.5 Model summary-Model 2

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.998[a]	.996	.995	.08099	1.689

Table 4.6 ANOVA-Model 2

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	29.030	4	7.258	1106.493	.000[b]
	Residual	.112	17	.007		
	Total	29.142	21			

Table 4.7 Model coefficients-Model 2

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.016	.221		.072	.944		
	Tourist_arrival	1.366E-6	.000	.300	7.168	.000	.128	7.790
	CPI	.017	.002	.662	6.849	.000	.024	41.517
	Exchange_rate	-.008	.004	-.110	-	.064	.073	13.756

					1.979			
	number.of.labour.permit	1.452E-6	.000	.196	6.207	.000	.226	4.428

4.1.3 Model 3

This model included number of tourist arrival, number of labour permit and CPI as independent variable results are presented in Table 4.8. Although all variables entered into the model was found to be significant ($p<0.05$) , multicollinearity was found in the variable CPI with VIF value exceeding 10 and hence the model was rejected.

Table 4.8 Model coefficients-Model 3

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-.412	.049		-8.420	.000		
	Tourist_arrival	1.557E-6	.000	.342	8.786	.000	.173	5.795
	number.of.labour.permit	1.674E-6	.000	.226	7.566	.000	.293	3.409
	CPI	.013	.001	.496	9.548	.000	.097	10.326

4.1.4 Model 4

This model was developed with number of labour permit, Tourist arrival, and Exchange rate as a predictor variable seemed to have significant result. Table 4.9 shows the predicted model had R^2 value of 0.986 which implies that 98.6% of variance in air passenger demand can be explained by the independent variables of the model.

Overall model was found to be good fit with p value less than 0.05 as shown in Table 4.10.

Table 4.11 showed that among the dependent variables number of tourist arrival was found to be the strongest predictor with factor of 0.520 when all other variables are controlled. Also the significance of all variables was recorded less than 0.05 which indicates that all variables i.e. number of tourist arrival, exchange rate and number of labour permit had significant contribution in the model.VIF value less than 5 for all variables indicated there was no problem of multicollinearity among the predicted variables. The equation for the model can be written as:

Air passenger demand(in millions)= -1.25+ 2.414*10^{-6} *(number of tourist arrival)+ 0.017*exchange rate +2.624*10^{-6}*(number of labour permit)……………………… 4.1

Table 4.9 Model summary-Model 4

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.993[a]	.986	.983	.15260	1.023

Table 4.10 ANOVA-Model 4

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	28.723	3	9.574	411.147	.000[b]
	Residual	.419	18	.023		
	Total	29.142	21			

Table 4.11 Model coefficients-Model 4

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std e	Beta			Tolerance	VIF
1	(Constant)	-1.248	.229		-5.449	.000		
	Tourist_arrival	2.414E-6	.000	.530	11.287	.000	.362	2.764
	Exchange_rate	.017	.004	.220	4.211	.001	.292	3.422
	number.of.labour.permit	2.62E-6	.000	.354	8.735	.000	.486	2.057

Figure 4.1 gives the idea about the normality of residuals where almost every residual closely follow a straight line from lower left to upper right corner which verifies the normality assumption of the model. Furthermore, Shapiro-wilk statistic in Table 4.12 being insignificant ensures the normality of residuals in the developed model.

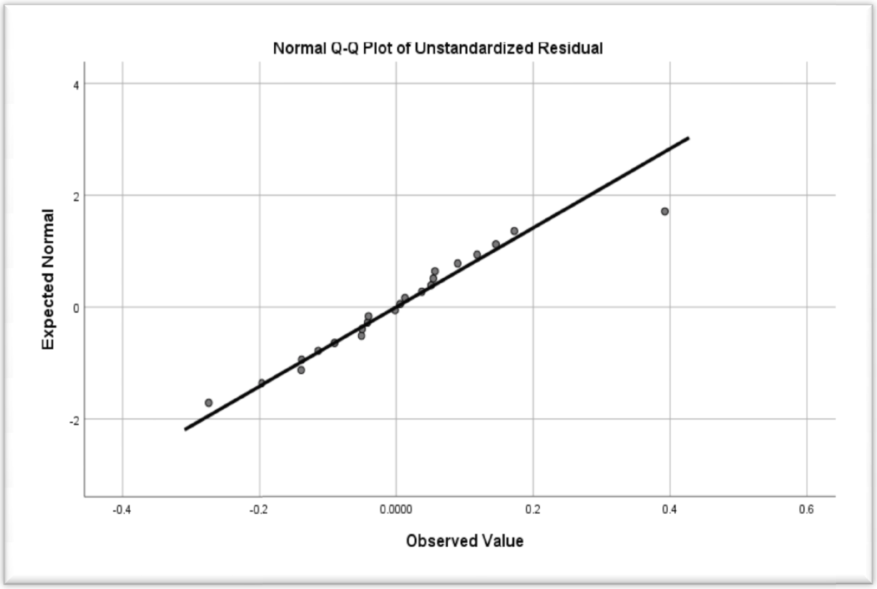

Figure 4.1: Normal Q-Q plot-Model 4

Table 4.12 Normality test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.117	22	.200*	.963	22	.558

Figure 4.2 given below shows the residual plot where points on the graph doesn't form any pattern and are randomly distributed which implies the homoscedasticity assumption of the model is preserved.

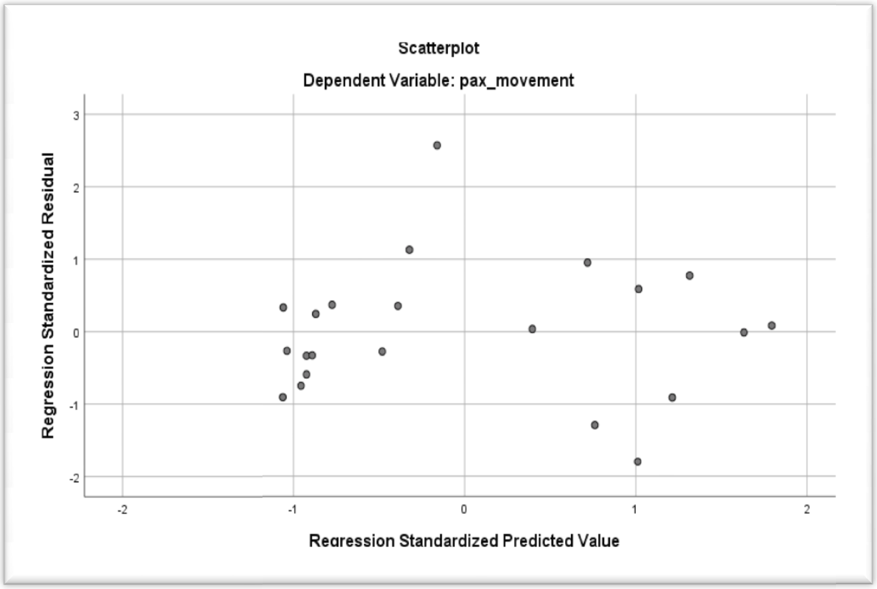

Figure 4.2: Scatter plot-Model 4

Figure 4.3 given below shows the points on the scatter plot between the dependent variable and independent variables forms an approximately straight line which verifies the linearity assumption for the model.

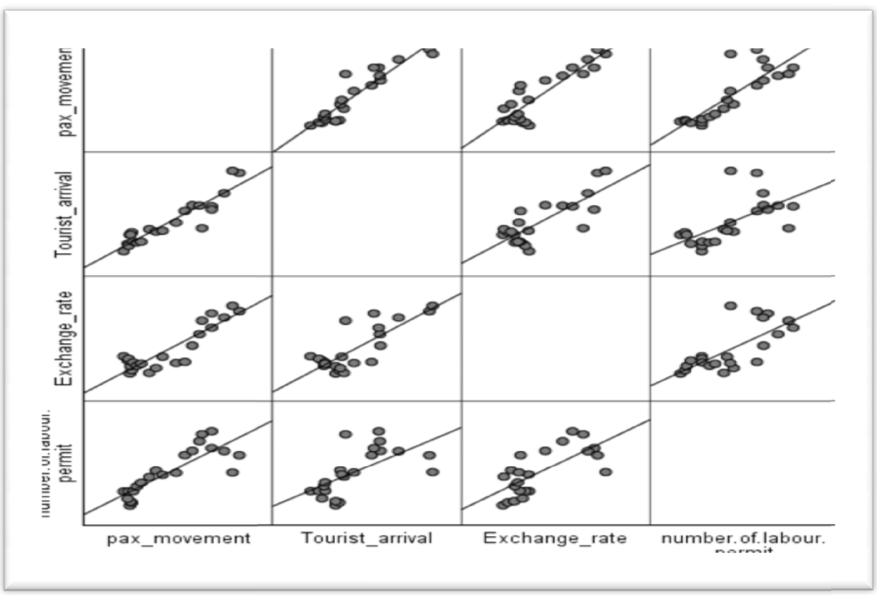

Figure 4.3: Linearity plot-Model 4

4.2 Airport Choice Model

4.2.1 Summary of data from SP survey

The distribution of collected data from the survey is presented in bars shown below. Figure 4.4 shows the distribution of respondents according to their district of origin. Respondents from various district participated in survey with dhanusa being highest in the sample.

Figure 4.5 shows the distribution of respondents according to their gender. About 80% of respondents were male while 20% were female.

Figure 4.6 shows the distribution of respondents according to their age group. Around 48% of respondents were into the age group 21-30 followed by 31-40, 41-60 and so on in decreasing order.

Figure 4.7 shows the distribution of respondents according to their destination country. Australia, Canada, UK, USA, Thailand were the most preferred destination.

Figure 4.8 shows the distribution of respondents according to their purpose of travel. Around 50% of the respondents travelled abroad for employment and around 35% of the respondent travelled for study. Less than 10% of respondents travelled abroad for business, VFR and holiday.

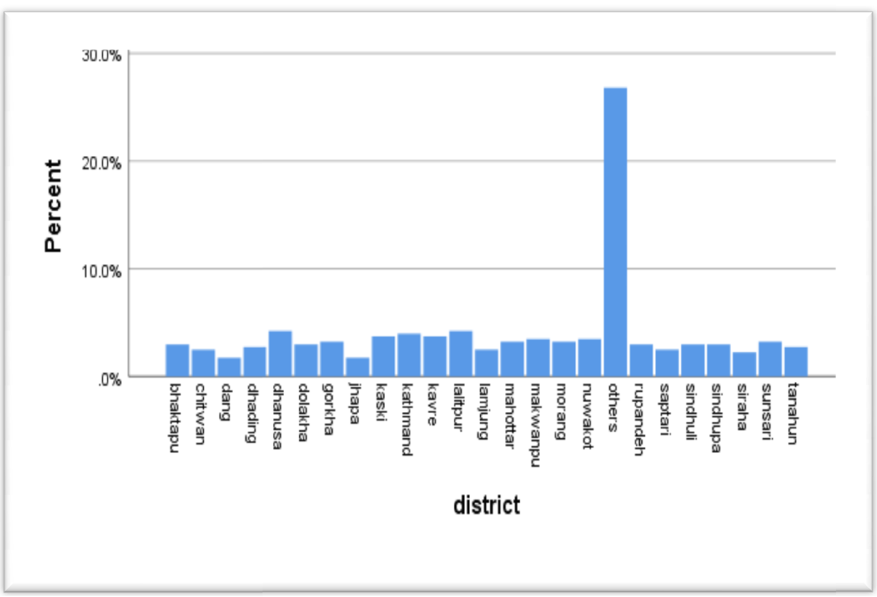

Figure 4.4: Distribution of data wrt district

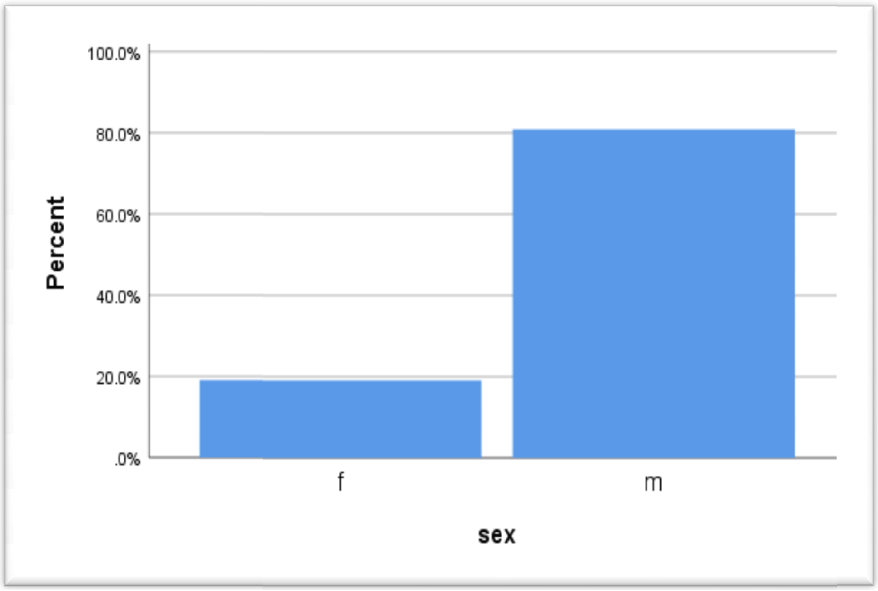

Figure 4.5:Distribution of data wrt sex

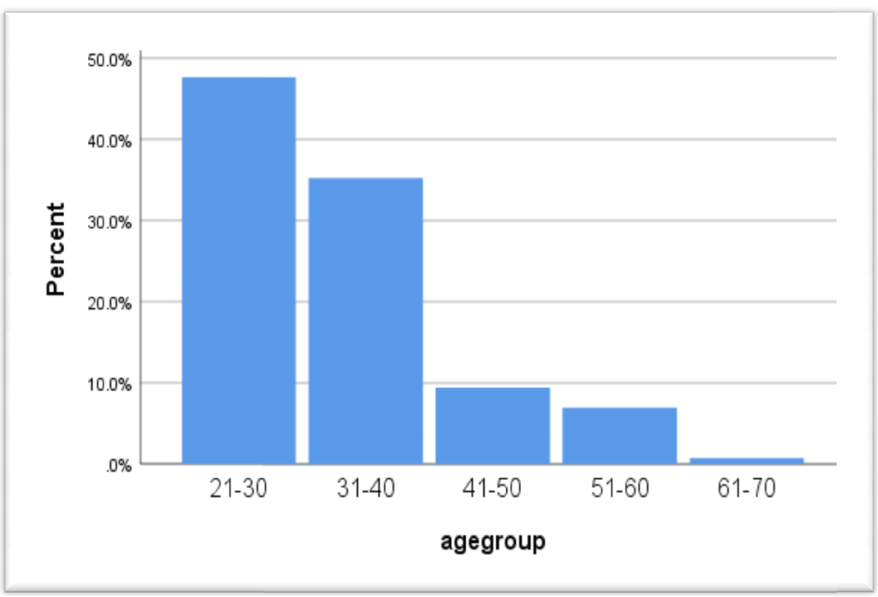

Figure 4.6: Distribution of data wrt age group

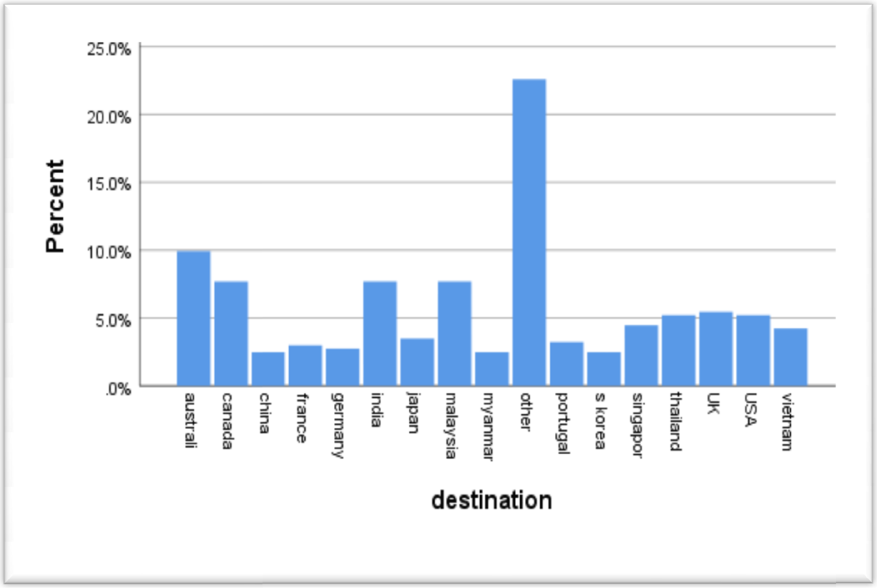

Figure 4.7: Distribution of data wrt destination

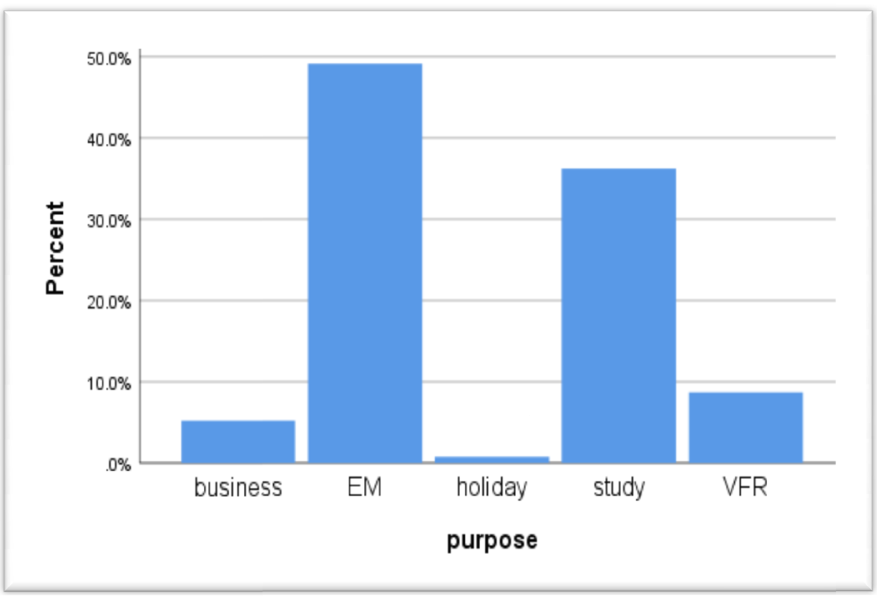

Figure 4.8: Distribution of data wrt purpose

4.2.2 Binary logistic model

Binary logistic regression was performed on the data collected from SP survey. The coding for airport choice is given in Table 4.13 and coding for categorical independent variable is given in Table 4.14.

Table 4.13: Coding for Airport Choice

Original Value	Internal Value
no	0
yes	1

Table 4.14: Coding for categorical variable

		Frequency	Parameter coding
			(1)
announced_Delay	less than TIA	1209	0
	greater than TIA	403	1

Logistic regression was performed in R. The goodness of fit test for the model as shown in Table 4.16 is represented by omnibus test of model coefficients (p value < 0.05) which indicates good fit of model. As shown in Table 4.15, the model as a whole explained between 16.4. % (Cox and Snell R square) and 21.9% (Nagelkerke R squared) of the variance in choosing the airport. As shown in Table 4.17, all of the three independent variables (access time, access cost and announced delay) made a unique statistically significant contribution (p<0.05) to the model. Furthermore from Figure 4.9, area under ROC was noted 0.715 which indicates the model has predicting ability greater than prediction by chance (=0.5).

Analysis of deviance is shown in Table 4.18 .Three models with different variables were developed and compared with full model (model with all variables).

First comparison was done between the full model and null model (i.e. model without any variables). It showed the deviance of full model (1944.8) was less than that of null model (2233.6) which verifies the significance of independent variables in the model.

Next comparison between full model with model having only access time showed that deviance of full model (1944.8) was less than other model (2223.2) which verifies the full model was better than model with access time only as independent variable.

Similarly, last comparison between full model and model with access time and access cost as independent variable showed results where deviance of full model (1944.8) was lesser than deviance of other model (2179) which signifies the performance of model with all variables was better than other three models described above.

The strongest predictor for choosing Nijgadh international airport was announced delay, recording an odds ratio of 0.108. This indicated that when announced delay at Nijgadh airport is greater than that of TIA, odds of choosing Nijgadh airport reduces by 90%, controlling for all other factors in the model. The odds ratio of 0.97, 0.94 was recorded for access time and access cost respectively which indicated that increase in 1 hour in access time to Nijgadh airport decreases the odds of choosing Nijgadh airport by 6% while increase in Rs 1000 in access cost for Nijgadh airport reduces the odds of choosing Nijgadh airport by 2%, controlling for all other factors in the model. The reason for access cost and access time being far less significant than announced delay might be the geography of Nepal. It's a small country and no vast difference in cost and time occurs between two places that are not too far away and that mentality of respondents might be evident in being access cost and access time less significant. Furthermore, the fact that TIA has reached almost saturation and air traffic congestion at TIA is evident to whole country and announced delay increases with the rise in air traffic. This might be the reason for announced delay having very high impact in airport choice.

Table 4.15: Model summary

Step	Mc fadden	Cox & Snell R Square	Nagelkerke R Square
1	0.13	.164	.219

Table 4.16: Omnibus test

		Chi-square	df	Sig.
Step 1	Step	288.691	3	.000
	Block	288.691	3	.000
	Model	288.691	3	.000

Table 4.17: Model coefficients

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lowe	Upper
Step 1[a]	Access_time	-.022	.010	5.032	1	.025	.978	.959	.997
	access_cost	-.062	.017	13.937	1	.000	.940	.910	.971
	announced_Delay(1)	-2.226	.162	187.872	1	.000	0.108	0.079	0.148
	Constant	0.771	.10	96.799	1	.000	.233		

Table 4.18: Analysis of deviance

Model 1: airport.choice ~ access.time + access.cost + announced.delay					
Model 2: airport.choice ~ 1					
model	**Resid. Df**	**Resid. Dev**	**Df**	**Deviance**	**Pr(>Chi)**
1	1608	1944.8			
2	1611	2233.6	-3	-288.85	< 2.2e-16
Model 1: airport.choice ~ access.time + access.cost + announced.delay					
Model 2: airport.choice ~ access.time					
model	**Resid. Df**	**Resid. Dev**	**Df**	**Deviance**	**Pr(>Chi)**
1	1608	1944.8			
2	1610	2223.2	-2	-278.41	< 2.2e-16
Model 1: airport.choice ~ access.time + access.cost + announced.delay					
Model 2: airport.choice ~ access.time + access.cost					
model	**Resid. Df**	**Resid. Dev**	**Df**	**Deviance**	**Pr(>Chi)**
1	1608	1944.8			
2	1609	2179	-1	-234.25	< 2.2e-16

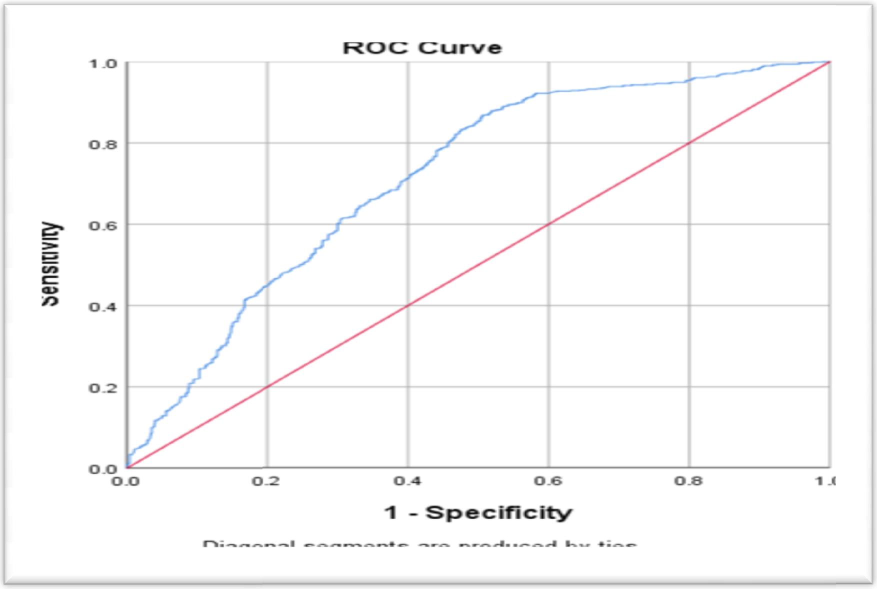

Figure 4.9: ROC curve

4.2.3 Model validation

Data collected for validation was processed to calculate precision, recall, accuracy, specificity and F1 score for the model. This study hypothesized, both model (trained data and test data) would achieve greater than chance performance (=50%) for all performance metrices.

A precision value of 0.617 for trained data indicated that out of all instances predicted positive by the model , 61.7% are truly positive which is greater than 50% (random chance) while precision value of 0.647 for test data indicated that out of all instances predicted positive by the model , 64.7% are truly positive. Only 3% variance in precision was found between test and train data.

Accuracy value of 67.4% for trained data and accuracy value of 72.1 % for test data showed approximately 5% variance between two models. It implied that developed model correctly classified 67.4% of cases out of total cases where model predicted Nijgadh as chosen airport which is greater than 50%(random chance).

Recall value of 86.9% for trained data and 95.1% for test data showed approximately 9% variance between two models. It implied that developed model correctly predicted 86.9% of the cases out of total cases where Nijgadh airport was chosen.

An F1 score of 0.722 was indicative of a well-performing binary classification model that had struck a good balance between precision and recall. It suggested that accurate positive predictions (choosing Nijgadh airport) were made effectively while both false positives (predicting positive but recorded negative) and false negatives (predicting negative but recorded positive) were minimized. The summary of metrices used to access the overall prediction capacity of models is shown in Table 4.19

Table 4.19: Confusion matrices of train and test data

Test data	***predicted positive***	***predicted negative***	**Trained data**	***predicted positive***	***predicted negative***
actual positive	**97**	**5**	***actual positive***	**682**	**103**
actual negative	**53**	**53**	***actual negative***	**423**	**404**
precision	**0.647**		***precision***	**0.617**	
recall	**0.951**		***recall***	**0.869**	
F1	**0.770**		***F1***	**0.722**	
accuracy	**0.721**		***accuracy***	**0.674**	
specificity	**0.500**		***specificity***	**0.489**	

4.2.4 Model as per travel purpose

A model for the passengers based on the purpose of their travel was made for the passengers with employment migration (EM), study and business as their purpose of travel.

- **Model for passengers travelling for EM:**

A summary for the model for the respondents having EM as the purpose of travel is shown in Table 4.21. Access cost and announced delay were significant with ($p<0.05$) while access time did not have much impact on the model. It implies that the group of people who travel abroad for employment to earn were sensitive about the cost rather than access time. Access cost being important than access time may be due to the obvious reason that respondents travelling abroad for employment are travelling for cost related/earning issues.

Table 4.20: Model summary- EM

	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	958.490	.161	.215

Table 4.21: Model coefficients-EM

Variables in the Equation								
	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Access_time	-.024	.014	3.086	1	.079	.976	.950	1.003
access_cost	-.056	.023	5.959	1	.015	.945	.903	.989
announced_Delay(1)	-2.198	.228	92.858	1	.000	0.11	5.759	14.079
Constant	.813	.206	45.245	1	.000	.251		

- **Model for passengers travelling for business:**

A summary for the model for the respondents having business as the purpose of travel is shown in Table 4.23. Access time and access cost were found to be insignificant for business travelers while announced delay had the impact on choice of airport for those travelers with business as purpose of travel.

Table 4.22: Model summary-Business

	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
	98.504	.185	.247

Table 4.23: Model coefficients-Business

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
	Access_time	-.006	.048	.018	1	.892	.994	.905	1.091
	access_cost	-.045	.070	.410	1	.522	.956	.833	1.097
	announced_Delay(1)	- 2.547	.819	9.673	1	.002	12.772	2.565	63.592
	Constant	.473	.381	7.288	1	.007	.126		

- **Model for passengers travelling for study:**

A summary for the model for the respondents having study as the purpose of travel is shown in Table 4.25. Access cost and announced delay were significant with ($p<0.05$) while access time did not have much impact on the model. It implies that the group of people who travel abroad for study were found to be sensitive about the cost rather than access time. (shrestha R.,2021) suggested that majority of Nepali students travelled abroad where the concept of earning with learning was the main factor of travel. So, students being motivated to travel by the factor of earning which is related to reducing cost, this might be the reason for students not considering access time as an important factor rather they are more interested in minimizing access cost and delay at the airport.

Table 4.24: Model summary-Study

	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	705.757	.161	.214

Table 4.25: Model coefficients-Study

		B	S.E.	Wald	df	Sig.	Exp(B)
	Access_time	-.024	.018	1.830	1	.176	.976
	access_cost	-.069	.030	5.296	1	.021	.933
	announced_Delay(1)	-2.227	.274	65.984	1	.000	.107

	Constant	.716	.182	35.984	1	.000	.221

4.2.5 Model as per travel origin:

Three different models was extracted from model according to the origin of the passengers region wise. Due to low number of data in sub model, passengers originating from province 1 and 2 was grouped in a single model which represents eastern region of the country, passengers originating from province 3 and 4 was grouped in a single model which represents central region of the country and passengers originating from province 5, 6 and 7 was grouped in a single model which represents western region of the country.

- **Model for passengers from eastern region**

Table 4.27 below shows that, model for passengers originating from the eastern part of Nepal are influenced by all the variables where increase in Rs 1000 in access cost decrases the odds of choosing Nijgadh airport by 9% while increase in 1 hour in access time decreases the odds of choosing Nijgadh airport by 10% and if the announced delay is greater than that of TIA, the odds of choosing Nijgadh airport decreases by 94%. The reason for announced delay being the most sensitive factor while access time and access cost having lesser impact might be due to geography of Nepal i.e. distance and cost to reach TIA is almost equivalent from mid point of eastern region. Respondents from the places at immediate proximity to Nijgadh airport could not be analyzed separately due to less number of samples available.

Table 4.26: Model summary-Eastern region

	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	568.308	.191	.254

Table 4.27: Model coefficients-Eastern region

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Access_time	-.093	.027	11.464	1	.001	.911
	access_cost	-.109	.036	9.359	1	.002	.897
	announced_Delay(1)	-2.768	.345	64.420	1	.000	.063
	Constant	1.826	.357	26.138	1	.000	6.207
.							

- **Model for passengers from central region**

Table 4.29 below shows that, model for passengers originating from the central part of Nepal are influenced by all the variables where increase in Rs 1000 in access cost decrases the odds of choosing Nijgadh airport by 6% while increase in 1 hour in access time decreases the odds of choosing Nijgadh airport by 7% and if the announced delay is greater than that of TIA, the odds of choosing Nijgadh airport decreases by 92%. The reason for announced delay being the most sensitive factor while access time and access cost having lesser impact might same as in case of eastern region. Respondents from the places at immediate proximity to TIA could not be analysed separately due to less number of samples available.

Table 4.28: Model summary-Central region

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
	975.205	.188	.250

Table 4.29: Model coefficients-Central region

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Access_time	-.069	.031	5.032	1	.025	.933
	access_cost	-.064	.026	6.011	1	.014	.938
	announced_Delay(1)	-2.573	.246	109.654	1	.000	.076
	Constant	.885	.171	26.711	1	.000	2.423

- **Model for passengers from western region**

Table 4.31 below shows that, model for passengers originating from the western part of Nepal are influenced by access cost and announced delay only. The p value for the coefficient assigned with access time was noted to be greater than 0.05 which implies the non significant contribution of access time variable. Whether the airport is located at Nijgadh or Kathmandu, its still far and equivalent in distance for the people of western region. This might be the reason for people from western side of country not considering access time significant while making the choice between airports.

Table 4.30: Model summary-Western region

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	373.073	.118	.157

Table 4.31: Model coefficients-Western region

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Access_time	-.028	.019	2.116	1	.146	.972
	access_cost	-.069	.029	5.548	1	.019	.933
	announced_Delay(1)	-1.743	.368	22.394	1	.000	.175
	Constant	1.062	.347	9.388	1	.002	2.893

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The study analyzed the demand for international air passenger in TIA using the econometric variables. Previous study had been done to model demand using several variables like GDP, number of tourist arrivals, CPI etc. Developed models with CPI, GDP, remittance, number of labour permit, number of tourist arrivals etc as explanatory variables yielded a good statistical result but since nearly perfect correlation between the variables was found as indicated by large VIF value, the model was rejected and a new model with CPI, number of labour permit, and exchange rate was developed and good statistical fit was found and the model was taken for the forecasting. The forecasted model could be used by planners to design terminal facilities and could be used for other strategic decisions.

Developed Airport choice model could be used to design new airport considering the impact of attributes like access time, access cost and announced delay. As announced delay is the only attribute that is totally related to a typical airport while access time and access cost can be influenced by several factors like comfort, safety, reliability etc and it was found that announced delay had far more significant effect in airport choice as compared to other variables. So, planners may be interested in focusing more on development of airport related attributes like terminal space facilities, flight frequencies, parking etc. Similarly, model based on purpose of travel showed different result where access time was insignificant for passengers travelling for EM and for study while both access cost and access time was found insignificant for passengers travelling for business. Also, model based on origin showed that passengers from western region were non sensitive about access time while passengers from other region were sensitive about all the attributes.

5.2 Recommendations

Following factors could be considered for further studies.

- Market share analysis could give better prospect for air travellers demand analysis.
- Other attributes like flight frequency, airport charges could be included to construct airport choice model
- Impact of other airports like Gautam Buddha international airport, pokhara international airport could be considered in airport choice study.

REFERENCES

Abuhamoud, M. A. A. (2011). *Modeling of Transport Mode in Libya: A Binary Logit Model for Government Transportation Encouragement*.

Alkabaa, A. S., & Taylan, O. (2021). *Designing an Effective Questionnaire using Orthogonal Array for Improving Inventory Accuracy at Electricity Warehouse*.

Bastola, D. P. (2017). *Air Passenger Demand Model (Apdm): Econometric Model For Forecasting Demand In Passenger Air Transports In Nepal*. *1*(4), 5.

Cherchi, E., & Hensher, D. A. (n.d.). *Stated preference surveys and experimental design: An audit of the journey so far and future research perspectives*. 14.

Dargay, J., & Hanly, M. (n.d.). *The Determinants of the Demand for International Air Travel to and from the UK*.

Erraitab, E. (2016). An Econometric Analysis Of Air Travel Demand: The Moroccan Case. *European Scientific Journal, ESJ*, *12*(7), 367. https://doi.org/10.19044/esj.2016.v12n7p367

Gautam, A. (n.d.). *Estimation of Value of Travel Time Saving for Commuter Trips:A case study of Kathmandu*.

Harvey, G. (1987). Airport choice in a multiple airport region. *Transportation Research Part A: General*, *21*(6), 439–449. https://doi.org/10.1016/0191-2607(87)90033-1

Hess, S., Adler, T., & Polak, J. W. (2007). Modelling airport and airline choice behaviour with the use of stated preference survey data. *Transportation Research Part E: Logistics and Transportation Review*, *43*(3), 221–233. https://doi.org/10.1016/j.tre.2006.10.002

Jiang, K., Zhang, N., & Feng, Z. (2014). *A Binary Logit Model for Commute Mode Choice: A Case Study of Hefei*. 3029–3040. https://doi.org/10.1061/9780784413623.290

Johnston, R. J., Boyle, K. J., Adamowicz, W. (Vic), Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., & Vossler, C. A. (2017). Contemporary Guidance for Stated Preference Studies. *Journal of the Association of Environmental and Resource Economists*, *4*(2), 319–405. https://doi.org/10.1086/691697

Karlaftis, M. G., Zografos, K. G., Papastavrou, J. D., & Charnes, J. M. (1996). Methodological Framework for Air-Travel Demand Forecasting. *Journal of Transportation Engineering*, *122*(2), 96–104. https://doi.org/10.1061/(ASCE)0733-947X(1996)122:2(96)

Marcucci, E., & Gatta, V. (n.d.). *A Stated Preference survey for airport choice behaviour*.

Rengaraju, V. R., & Arasan, V. T. (1992). Modeling for Air Travel Demand. *Journal of Transportation Engineering*, *118*(3), 371–380. https://doi.org/10.1061/(ASCE)0733-947X(1992)118:3(371)

Saffarzadeh, M., Yazdan-Panah, M., & Taherpour, A. (n.d.). *Airport Choice Model for Tehran Metropolitan Area*.

Wang, M., & Song, H. (2010). Air Travel Demand Studies: A Review: 航空旅行需求研究综述. *Journal of China Tourism Research*, *6*(1), 29–49. https://doi.org/10.1080/19388160903586562

Windle, R., & Dresner, M. (1995). Airport Choice in Multiple-Airport Regions. *Journal of Transportation Engineering*, *121*(4), 332–337. https://doi.org/10.1061/(ASCE)0733-947X(1995)121:4(332)

Zurovac, J., & Brown, R. (n.d.). *Orthogonal Design: A Powerful Method for Comparative Effectiveness Research with Multiple Interventions*.

APPENDIX A

Choice set :Orthogonal design

Card List				
	Card ID	time	cost	delay
1	1	equal to tia	0.5 times of tia	greater than tia
2	2	1.5 times of tia	0.5 times of tia	less than tia
3	3	0.5 times of tia	1.5 times of tia	greater than tia
4	4	equal to tia	1.5 times of tia	less than tia
5	5	0.5 times of tia	0.5 times of tia	less than tia
6	6	equal to tia	equal to tia	less than tia
7	7	0.5 times of tia	equal to tia	less than tia
8	8	1.5 times of tia	1.5 times of tia	less than tia
9	9	1.5 times of tia	equal to tia	greater than tia

SP survey sample questionnaire:

AIRPORT CHOICE SURVEY

SECTION 1: DEMOGRAPHIC INFORMATION (जनसांख्यिकीय विवरण)

Place of origin(स्थायि ठेगाना)

Province: (प्रदेश) [ ]

District: (जिल्ला) [ ]

Age: [ ]

Sex (लिङ्ग) Female (महिला) [ ] Male(पुरुष) [ ]

SECTION 2: AIR TRAVEL INFORMATION

Number of international travels by flight(अन्तराष्ट्रिय उडान को संख्या)

First(पहिलो) [ ] More than 1(एक भन्दा [ ]

बढि)

Destination country (तपाइ जान लाग्नु भएको देश)

Purpose of travel (वैदेशिक भ्रमण को उदेश्य)

Business (व्यापार)

Holiday (घुमघाम)

Employment's migration (वैदिशिक रोजगार)

Visiting family and relatives (आफन्त तथा साथिभाइ संग भेटघाट)

study(पठनपाठन)

SECTION 3: AIRPORT ACCESS INFORMATION

Did you come to airport directly from home? (के तपाइ आफ्नो वासस्थान बाट सिधै विमानस्थल आउनु भएको हो?)

yes **No**

If,yes how much time did it take you to get to Airport? (यदि हो भने तपाइलाइ एयरपोर्ट आउन कति समय लाग्यो?)	**If,no where did you stay before you get to Airport? (यदि होइन भने तपाइ एयरपोर्ट आउन काहाँ भयो?)**
	At hotel At relatives/ friends other

Which mode of transport did you use to get to airport? (तपाइ ले एयरपोर्ट आउन प्रयोग गर्नु भएको साधन?) एक भन्दा बढि साधन प्रयोग गर्नु भएको भए खुलाइदिनु होला

public bus(सार्वजनिक बस)

Airplane(हवाइजहाज)

Car (कार)

Taxi(ट्याकसि)

Others(अन्य)	

How much did it cost you to get to airport from your place of origin?(तपाइ लाइ आफ्नो स्थाइ वासस्थान बाट एयरपोर्ट आउन लागेको खर्च ?)

SECTION 3: AIRPORT CHOICE

Four different airport scenario for the proposed Nijgadh International Airport is presented in the table below. Please select the choice of your airport i.e. between Tribhuwan International Airport and Nijgadh International Airport

- **please tick (✓) the Airport option you choose** -

OPTION 1

	ATTRIBUTES				Remarks
AIRPORT OPTIONS	Access time(एयरपोर्ट पुग्न लाग्ने समय) (hours)	access cost (एयरपोर्ट पुग्न लाग्ने खर्च)(in rupees)	Announced delay (उडाएन मा ढिलाइ) (in hours)	**Choice of Airport**	
TIA					*please fill the row according to your experience*
NIJGADH AIRPORT	Equal to TIA	1.5 times of TIA	less than TIA		*please compare the conditions of Nijgadh airport and make your choice*

OPTION 2

	ATTRIBUTES				Remarks
AIRPORT OPTIONS	Access time(एयरपोर्ट पुग्न लाग्ने समय) (hours)	access cost (एयरपोर्ट पुग्न लाग्ने खर्च)(in rupees)	Announced delay (उडाएन मा ढिलाइ) (in hours)	**Choice of Airport**	
TIA					*please fill the row according to your experience*
NIJGADH AIRPORT	0.5 times of TIA	1.5 times of TIA	greater than TIA		*please compare the conditions of Nijgadh airport and make your choice*

OPTION 3

	ATTRIBUTES	Remarks

AIRPORT OPTIONS	Access time(एयरपोर्ट पुग्न लाग्ने समय) (hours)	access cost (एयरपोर्ट पुग्न लाग्ने खर्च)(in rupees)	Announced delay (उडाएन मा ढिलाइ) (in hours)	**Choice of Airport**	
TIA					*please fill the row according to your experience*
NIJGADH AIRPORT	1.5 times of TIA	1.5 times of TIA	less than TIA		*please compare the conditions of Nijgadh airport and make your choice*

OPTION 4					
	ATTRIBUTES				Remarks
AIRPORT OPTIONS	Access time(एयरपोर्ट पुग्न लाग्ने समय) (hours)	access cost (एयरपोर्ट पुग्न लाग्ने खर्च)(in rupees)	Announced delay (उडाएन मा ढिलाइ) (in hours)	**Choice of Airport**	
TIA					*please fill the row according to your experience*
NIJGADH AIRPORT	1.5 times of TIA	0.5 times of TIA	less than TIA		*please compare the conditions of Nijgadh airport and make your choice*

APPENDIX B

Source code for choice modeling in R

```
#for full model

data<-read.csv("data for R1.csv",header=TRUE,sep=",")

str(data)

data$announced.delay<-as.factor(data$announced.delay)

str(data)

model<-glm(airport.choice~1,data=data,family="binomial")

summary(model)

finalmodel<-
glm(airport.choice~access.time+access.cost+announced.delay,data=data,family="binomial")

summary(finalmodel)

exp(coef(finalmodel))

anova(model,finalmodel,test="LRT")

library(DescTools)

PseudoR2(finalmodel,which=c("McFadden","McFadden Adj","Nagelkerke","CoxSnell"))

finalmodel1<-glm(airport.choice~access.time+access.cost,data=data,family="binomial")

anova(finalmodel,finalmodel1,test="LRT")

finalmodel2<-glm(airport.choice~access.time,data=data,family="binomial")

anova(finalmodel,finalmodel2,test="LRT")

#for model with business as purpose of travel

data<-read.csv("data for R1 business.csv",header=TRUE,sep=",")

str(data)

data$announced.delay<-as.factor(data$announced.delay)
```

```
str(data)

model<-glm(airport.choice~1,data=data,family="binomial")

summary(model)

finalmodel<-
glm(airport.choice~access.time+access.cost+announced.delay,data=data,family="binomial")

summary(finalmodel)

exp(coef(finalmodel))

#for model with EM as purpose of travel

data<-read.csv("data for R1 EM.csv",header=TRUE,sep=",")

str(data)

data$announced.delay<-as.factor(data$announced.delay)

str(data)

model<-glm(airport.choice~1,data=data,family="binomial")

summary(model)

finalmodel<-
glm(airport.choice~access.time+access.cost+announced.delay,data=data,family="binomial")

summary(finalmodel)

exp(coef(finalmodel))

#for model with study as purpose of travel

data<-read.csv("data for R1 study.csv",header=TRUE,sep=",")

str(data)

data$announced.delay<-as.factor(data$announced.delay)

str(data)

model<-glm(airport.choice~1,data=data,family="binomial")
```

```
summary(model)

finalmodel<-
glm(airport.choice~access.time+access.cost+announced.delay,data=data,family="binomial")

summary(finalmodel)

exp(coef(finalmodel))
```

APPENDIX C

Sample of collected data

sn	province	district	sex	age	destination	purpose	*Total time (hour)*	*Total cost*	access time	access cost	delay	Aiport choice(Nijgadh=1)
1	7	kailali	m	32	UAE	EM	13.5	2700	13.5	4.05	0	1
								0	6.75	4.05	1	1
								0	20.25	4.05	0	0
								0	20.25	1.35	0	1
2	3	dolakha	m	23	canada	study	2.5	500	2.5	0.75	0	1
								0	1.25	0.75	1	0
								0	3.75	0.75	0	0
								0	3.75	0.25	0	1
3	1	Jhapa	m	37	Qatar	EM	7.15	1500	7.15	2.25	0	0
								0	3.575	2.25	1	0
								0	10.725	2.25	0	1
								0	10.725	0.75	0	1
4	1	Sunsari	f	32	Qatar	EM	8	1500	8	2.25	0	1
								0	4	2.25	1	0
								0	12	2.25	0	0
								0	12	0.75	0	1
5	3	Kavre	m	27	USA	study	2	400	2	0.6	0	0

								0	1	0.6	1	1
								0	3	0.6	0	1
								0	3	0.2	0	0
6	2	Saptari	m	25	Qatar	EM	7	1350	7	2.025	0	1
								0	3.5	2.025	1	1
								0	10.5	2.025	0	0
								0	10.5	0.675	0	0
7	2	dhanusa	m	23	USA	study	9	1600	9	2.4	0	1
								0	4.5	2.4	1	1
								0	13.5	2.4	0	1
								0	13.5	0.8	0	0
8	5	dang	m	55	UK	VFR	15	3000	15	4.5	0	1
								0	7.5	4.5	1	0
								0	22.5	4.5	0	0
								0	22.5	1.5	0	1
9	1	Ilam	f	27	Bahrain	EM	18	3960	18	5.94	0	0
								0	9	5.94	1	0
								0	27	5.94	0	0
								0	27	1.98	0	1
10	2	Mahottari	m	27	malaysia	EM	7	1350	7	2.025	0	0
								0	3.5	2.025	1	1
								0	10.5	2.025	0	0
								0	10.5	0.675	0	1
1	3	sindhuli	m	33	UAE	EM	5	105	5	1.57	0	0

1								0		5		
								0	2.5	1.57 5	1	0
								0	7.5	1.57 5	0	0
								0	7.5	0.52 5	0	1
1 2	3	bhaktap ur	m	46	Japan	Busin ess	1	200	1	0.3	0	0
								0	0.5	0.3	1	0
								0	1.5	0.3	0	0
								0	1.5	0.1	0	1
1 3	3	Nuwak ot	f	23	USA	study	2	450	2	0.67 5	0	0
								0	1	0.67 5	1	0
								0	3	0.67 5	0	0
								0	3	0.22 5	0	1
1 4	4	gorkha	m	34	s korea	EM	7	150 0	7	2.25	0	1
								0	3.5	2.25	1	1
								0	10.5	2.25	0	0
								0	10.5	0.75	0	0
1 5	1	morang	f	27	USA	study	11	210 0	11	3.15	0	1
								0	5.5	3.15	1	0
								0	16.5	3.15	0	1
								0	16.5	1.05	0	1
1 6	2	saptari	m	26	Qatar	EM	9	180 0	9	2.7	0	0
								0	4.5	2.7	1	1
								0	13.5	2.7	0	1
								0	13.5	0.9	0	0

17		siraha	m	33	Qatar	EM	7	1500	7	2.25	0	0
	2							0	3.5	2.25	1	1
								0	10.5	2.25	0	1
								0	10.5	0.75	0	1
18	5	Argakhachi	m	35	Qatar	EM	15.25	3000	15.25	4.5	0	0
								0	7.625	4.5	1	1
								0	22.875	4.5	0	0
								0	22.875	1.5	0	1
19	3	Dhading	m	33	UAE	EM	3	500	3	0.75	0	1
								0	1.5	0.75	1	0
								0	4.5	0.75	0	0
								0	4.5	0.25	0	1
20	3	dhading	m	64	USA	VFR	6	1100	6	1.65	0	1
								0	3	1.65	1	0
								0	9	1.65	0	0
								0	9	0.55	0	1