

PROFESSIONAL TRAINING REPORT
SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering degree Information Technology

By

DINESH RAJA RADHA
(40120028)



DEPARTMENT OF INFORMATION TECHNOLOGY
SCHOOL OF COMPUTING
SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade “A” by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI, TAMIL NADU- 600 119

APRIL – 2023



SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

(www.sathyabama.com)

DEPARTMENT OF INFORMATION TECHNOLOGY

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **DINESH RAJA RADHA** who carried out the project entitled "**FOREIGN VISITORS PREDICTION**" under my supervision from January 2023 to April 2023.

Internal Guide

Dr.J. Jabez M.E., Ph.D.,

HEAD OF THE DEPARTMENT

Dr. R. Subhashini M.E., Ph.D.,

Submitted for Viva voce Examination held on 28.04.2023

Internal Examiner

External Examiner

DECLARATION

I, **DINESH RAJA RADHA** hereby declare that the Project Report entitled "**FOREIGN VISITORS PREDICTION**" done by me under the guidance of **DR.J. JABEZ M.E., Ph.D.** at Sathyabama Institute of Science and Technology (Deemed to be University), Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600119 is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology degree in Information Technology.

DATE: 28.04.2023

PLACE: Chennai



SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge our sincere thanks to Board of management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey our thanks to **Dr. T. SASIKALA M.E., Ph.D., Dean, School of Computing** and **Dr. R. SUBHASHINI M.E., Ph.D., Head of the Department of Information Technology** for providing us the necessary support and details at the right time during the progressive reviews.

I would like to express our sincere and deep sense of gratitude to our Project **Dr.J. JABEZ M.E., Ph. D.**, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

TRAINING CERTIFICATE

No. : 34338


एन एस आई सी
N S I C

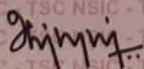
राष्ट्रीय लघु उद्योग निगम लिमिटेड
NATIONAL SMALL INDUSTRIES CORPORATION LIMITED
(A Govt. of India Enterprise)

NSIC - TECHNICAL SERVICES CENTRE
Sector B-24, Guindy Industrial Estate, Ekkaduthangal, Chennai - 600 032.

CERTIFICATE

This is to certify that Mr. **DINESH RAJA RADHA**
S/o. Shri. **RAJA APPADURAI**, student of III year
B.Tech. (IT), Sathyabama Institute of Science and
Technology has undergone "**SKILL DEVELOPMENT
PROGRAM ON DATA ANALYTICS**" conducted by
us for a period of one week (30 Hours) from
09.03.2023 to 14.03.2023.




HEAD OF TRAINING


HEAD OF CENTRE

ABSTRACT

The increasing availability of trajectory recordings has led to the mining of a massive amount of historical track data, allowing for a better understanding of travel behaviors by revealing meaningful motion patterns. In the context of human mobility analysis, the problem of motion prediction assumes a central role and is beneficial for a wide range of applications, including for touristic purposes, such as personalized services or targeted recommendations, and sustainability studies related to crowd management and resource redistribution. This paper tackles a particular case of the trajectory prediction problem, focusing on large-scale mobility traces of short-term foreign tourists. These sparse trajectories, short and non-repetitive, lack spatial and temporal regularity, making prediction analysis based on individual historical motion data unreliable. To face this issue, we hereby propose a deep learning-based approach, taking into account the collective mobility of tourists over the territory. The underlying semantics of motion patterns are captured by means of a long short-term memory (LSTM) neural network model trained on pre-processed location sequences, aiming to predict the next visited place in the trajectory. We tested the methodology on a real-world big dataset, demonstrating its higher feasibility with respect to traditional approaches. Our focus and intended application are related to tourists' mobility within the growing field of smart tourism. Smart tourism integrates tourism resources with information technologies to design intelligent services to provide valuable outcomes to tourists and tourism-related industries. The development of smart tourism is particularly embodied in four main aspects, namely tourism experience, tourism management, tourism service. sustainability studies related to crowd management and resource redistribution. In general, this study contributes to the expansion of tourists' mobility analysis in the direction of actively integrating artificial intelligence into the tourism sector.

TABLE OF CONTENTS

Chapter No	TITLE	Page No.
	ABSTRACT	v
	LIST OF FIGURES	vii
1	INTRODUCTION	1
	1.1 FORECASTING	2
	1.2 FOREIGN VISITORS FORECASTING	4
	1.3 FOREIGN VISITORS' PREDICTION	5
	1.4 SPECIFICATION	6
2	LITERATURE SURVEY	8
3	METHODS AND ALGORITHM USED	9
	3.1 Project Development	9
	3.2 Data Analytics	10
	3.3 Description Analysis	11
	3.4 System Requirements	15
	3.5 Power Bi	16
	3.6 Methodology	17
4	RESULTS AND DISCUSSION	23
5	CONCLUSION AND FUTURE WORK	26
	REFERENCES	30
	APPENDIX	27
	A. SCREENSHOTS	27

LIST OF FIGURES

Figure No	Figure Name	Page No
3.1	PROJECT DEVELOPMENT	9
3.2	EXCEL DATA	18
3.3	SUM OF VISITORS BY YEAR	19
3.4	SUM COUNT BY DURATION	20
3.5	SUM OF COUNT BY COUNTRY	20
3.6	SUM OF COUNT BY EHTHINICITY	21
3.7	SUM OF COUNT BY GENDER	22
3.8	FOREIGN VISITORS' PREDICTION	23
3.9	FOREIGN VISITORS' PREDICTION	24
3.10	FOREIGN VISITORS' PREDICTION	25

CHAPTER – 1

INTRODUCTION

As a highly applicable tool, prediction has been widely applied to various fields of society and has achieved good application results and as an important driving engine of the national economy, tourism plays a vital role in the development of all countries. As a tourism great nation, inbound tourism has important position in the development of China tourism market, and is an important source of China foreign exchange earnings. With the increasing number of foreign tourists, the effective forecasting of China inbound foreign visitors has become an important reference for the pricing and resource allocation of the tourism market. However, traditional studies based on historical statistics have a certain hysteresis quality, which limits the short-term effectiveness of prediction. In addition, forecasting based on traditional historical data are always affected by uncontrollable factors such as natural disasters, thereby, it also reduces the accuracy of prediction. To overcome this obstacle, tourism demand forecast based on the internet. But in terms of methods, most existing researches still focus on measurement models such as time series analysis and regression analysis. Although it is convenient and easy to use, in the case of too many variables, the result of multicollinearity is often difficult to avoid. In order to solve this problem, many articles generally perform principal component analysis on multiple independent variables but this operation will greatly weaken the amount of information carried by the independent variables. To solve this problem, this paper uses the Random Forest (RF) method which is suitable for multiple variable scenarios to predict the number of China inbound foreign visitors based on Google Trends data. The validity of the data and the rationality of the method are both verified by higher prediction accuracy.

1.1 FORECASTING

Forecasting is the process of projecting past sales demand into the future. Implementing a forecasting system enables you to assess current market trends and sales quickly so that you can make informed decisions about the operations. It is a planning tool that enables businesses to chart their next moves and create budgets that will hopefully cover whatever uncertainties may occur.

1.1.1 Types of Forecasting

- Quantitative
- Qualitative

1.2 FOREIGN VISITORS FORECASTING

Foreign visitors forecasting is the process of predicting the number of tourists or visitors that will travel to a particular destination from other countries. It is an important task for tourism industry stakeholders such as hotels, airlines, and governments, as it helps them plan for and manage tourism activities. Foreign visitors forecasting is a critical task for the tourism industry, and accurate predictions can help stakeholders make informed decisions about tourism activities, investments, and marketing strategies. They can also help enterprises to plan their marketing and advertising efforts. Foreign visitors forecasting can help businesses make informed decisions about allocating resources. Foreign visitors forecasting is not an exact science, and uncertainty is always involved. However, sales forecasts can give businesses a good idea of what to expect in the future and help them plan accordingly.

1.2.1 Pros and Cons

Pros:

1. Predicting the number of foreign visitors to a particular destination can help tourism industry stakeholders allocate resources more efficiently. For example, hotels can better manage their room inventory, and airlines can adjust their flight schedules to meet demand.
2. By accurately predicting visitor numbers, tourism industry stakeholders can develop more effective marketing strategies. This can help attract more visitors to a destination and increase tourism revenue.
3. Accurate predictions of foreign visitor numbers can help governments and other stakeholders measure the economic impact of tourism on a destination. This information can be used to inform policy decisions and allocate resources to support the tourism industry.
4. In the event of a natural disaster or other disruptive event, tourism industry stakeholders can use forecasting data to adjust their operations and minimize the impact on visitors.
5. Predicting the number of foreign visitors can help tourism industry stakeholders plan their budgets and investments. This includes marketing budgets, capital expenditures, and revenue projections.

Cons:

1. Predicting tourism demand is inherently difficult, and there are many factors that can influence visitor numbers that are beyond the control of stakeholders. This includes unexpected events such as natural disasters, political instability, or global pandemics.
2. Potential for overestimation or underestimation: If forecasting methods are inaccurate or if unforeseen events occur, tourism stakeholders may overestimate or underestimate visitor numbers. This can lead to underutilized resources or missed opportunities.
3. Inaccuracy of forecasting methods: No forecasting method is perfect, and some methods may be more accurate than others depending on the specific context. It is important for stakeholders to carefully consider the strengths and limitations of different methods when making decisions based on forecasted data.
4. Difficulty in predicting long-term trends: While forecasting can be useful for short-term planning, it is often more difficult to predict long-term trends in tourism demand. This is because many factors that influence tourism are subject to change over time, such as economic conditions, travel restrictions, and consumer preferences.

1.3 FOREIGN VISITOR PREDICTION:

Foreign visitor prediction is the process of forecasting the number of tourists or visitors that will travel to a particular destination from other countries. This is an important task for stakeholders in the tourism industry, including hotels, airlines, and local governments, as it helps them plan for and manage tourism activities. Accurate predictions of visitor numbers can help stakeholders allocate resources more effectively, plan their budgets and investments, and target specific markets more effectively.

1.4 SPECIFICATIONS

1.4.1 HARDWARE CONFIGURATION

- Processor - I5
- Speed - 3 GHz
- RAM - 8GB (min)
- Hard Disk - 500 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - LCD
- Ethernet connection :(LAN)

1.4.2 SOFTWARE CONFIGURATION

- Microsoft Excel sheet
- Power BI Application
- Web Browser

CHAPTER - 2

LITERATURE SURVEY

2.1 Big Data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model

The prediction of tourist numbers is important for Destination Management and Marketing. While most existing methods rely on well-structured statistical data, using web search queries of the destination to forecast its tourist arrivals is a new way to apply Big Data analytics. However, there are no studies exploring correlation of weather, temperatures, weekends and public holidays with tourism destination arrivals and web search queries of the destination, respectively. This study uses the Vector Autoregressive modeling to examine the Granger causality between actual arrivals of the studied cultural tourism destination and its web search queries, and to explore the correlation mentioned above. The striking result is that weather has no correlation either with actual arrivals of the studied cultural tourism destination, or with its web search queries. Meanwhile, unlike previous researchers who discuss the predictive power of web queries on actual tourism flows, this study emphasizes their reciprocal predictive powers upon each other. The originality of this study is exemplifying the utilization of Big Data analytics in the tourism domain with Big Data datasets, data capture techniques, analytical tools, and analysis results. This study further digs possible reasons for an identified short time lag length ($p = 2$), to provide insights for Destination Management and Marketing.

2.2 Big data from dynamic pricing: A smart approach to tourism demand forecasting

The task of demand forecasting is often based on dynamic statistical models – mainly involving time series analysis – or model-free machine learning techniques. Such models present one main drawback that limit their application in long-term forecasting: the need to identify informative variables at very large lags or to accept increasingly wider confidence intervals in the case of dynamic forecasting. Predicting the sectoral demand is not simple even in the short term, when small area or high-frequency details are requested. In fact, official statistics are not usually published

with high time-space detail or – when available – they are neither complete nor timely. In this paper, we explore the possibility to exploit public big data to partially overcome these problems in all those industries where online dynamic pricing is spreading as a marketing-management practice. In particular, we focus on the accommodation sector, as it is important for the economies of many locations worldwide and above all, characterized by a huge number of decision makers interested in making daily forecasts about arrivals for circumscribed periods like summer, weekends, or days when fairs or other relevant events are held. Buono et al. (2017) highlight the potential of online price data in nowcasting and forecasting inflation, though this kind of big data has not been extensively studied in macroeconomics. In the field of tourism, according to Li et al. (2018), big data are drawn from three primary sources: social media, devices, and operations. They originate mainly from the behaviour of customers (even potential), while online sales prices are among the few operations generated by the supply side. Surprisingly, big data analytics have not yet considered hotels' pricing strategies – i.e., the whole set of prices that a hotel publishes on an Online Travel Agency (OTA) at different advance bookings. As we further argue, these data represent one of the most effective examples of shared knowledge regarding businesses' expectations about future demand, allowing for high-frequency prediction, even in a small geographic area.

2.3 A systematic literature review on the use of big data for sustainable tourism

Big data has attracted substantial interest from researchers as a means to obtain insights to find solutions, new strategies, and to uncover hidden potentials for an array of purposes (Mayer-Schonberger & Cukier). Big data is a term that primarily describes datasets that are so large, unstructured and complex that they require advanced and unique technologies to store, manage, analyse and visualize data (Gunther et al.). Datasets with such features are emerging from multiple economic sectors. Among them, the tourism sector has distinctive characteristics, like the intangibility and perishability of products and the vulnerability to natural, economic and political phenomena. The tourism industry thrives on information. As such, research on tourism requires large volumes of vigorous, updated, timely and relevant information to support

and help decision-making processes. Big data can deliver up-to-date and immensely informed inferences regarding behaviour and human activity that enhance the tourism industry (Xu & Whitmarsh). Through the tourist, a plethora of data sources can inform decisions made at different phases, such as before, during and after a trip. One example of a popular data source is social media, which is used by tourists and also tourism businesses to communicate, find or deliver relevant information or any recommendation, and to obtain information about important news and crises. Among the branches of tourism, sustainable tourism has understandably gained traction in recent years (Demunter,). Sustainable tourism is not only about achieving economic purposes, but also assessing environmental, and social dimensions with the focus on mitigating or remediating environmental, social and economic impacts (Loaiza et al). This type of tourism also has an important role in the achievement of the 2030 Agenda of Sustainable Development Goals (SDGs) (UNWTO & UNDP). In particular, it can accelerate the accomplishment of all goals in the agenda because it has a multiplier effect on other sectors.

2.4 Sustainable tourism research: an analysis of papers published in the Journal of Sustainable Tourism

Decision-making in tourism development and planning is becoming increasingly complex as organizations and communities have to come to terms with the competing economic, social and environmental demands of sustainable development. Geographical Information Systems (GIS) can be regarded as providing a tool box of techniques and technologies of wide applicability to the achievement of sustainable tourism development. Spatial (environmental) data can be used to explore conflicts, examine impacts and assist decision-making. Impact assessment and simulation are increasingly important in tourism development, and GIS can play a role in auditing environmental conditions, examining the suitability of locations for proposed developments, identifying conflicting interests and modelling relationships. Systematic evaluation of environmental impact is often hindered by information deficiencies but also tools for data integration, manipulation, visualization and analysis. GIS seems particularly suited to this task. This paper examines the progress tourism related organizations are making towards the utilization of GIS and its integration with the principles of sustainable development in the UK.

CHAPTER – 3

METHODS &ALGORITHM USED

3.1 PROJECT DEVELOPMENT:

In this section, there is an overview of the process of data mining and data modelling, from collecting the data, through the data preparation and finally the data modeling. Data cleaning and formatting can be considered as the most critical part of the whole project according to several data scientists. Figure 3.1 shows how the process of data mining works to extract knowledge using algorithms.

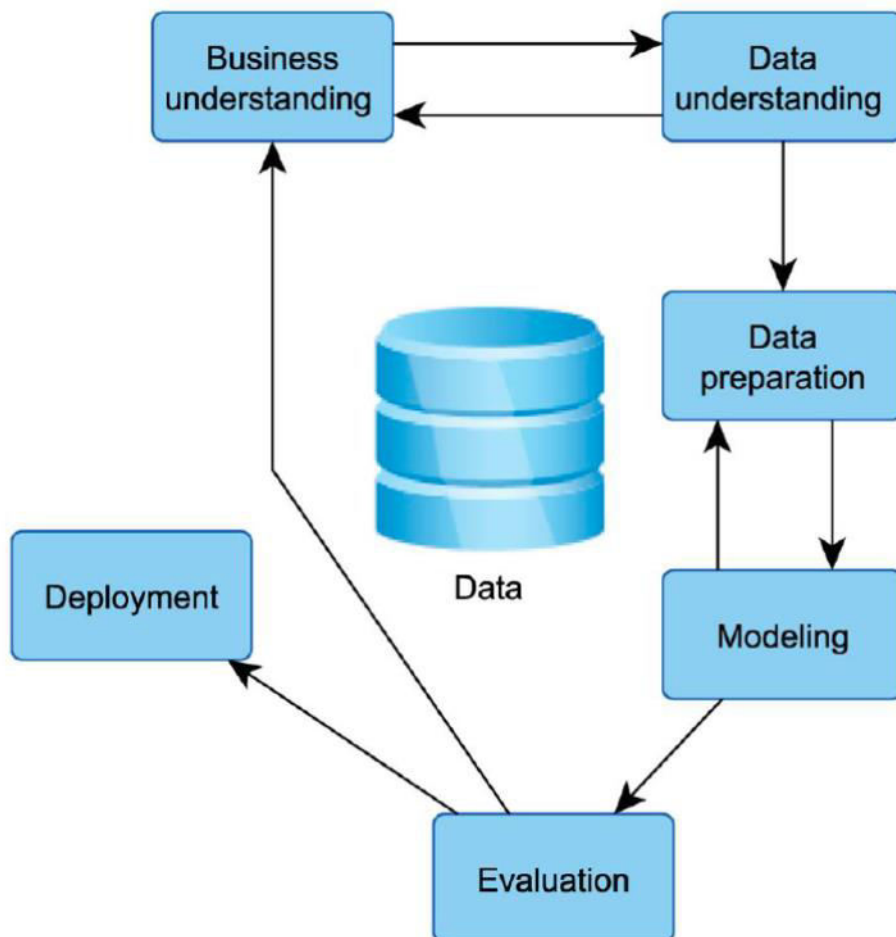


Fig:3.1 Project Development

3.2 DATA ANALYTICS:

The process of data analysis, or alternately, data analysis steps, involves gathering all the information, processing it, exploring the data, and using it to find patterns and other insights. The process of data analysis consists of: Data Requirement Gathering: Ask yourself why you're doing this analysis, what type of data you want to use, and what data you plan to analyses.

3.2.1 Data Collection:

Guided by your identified requirements, it's time to collect the data from your sources. Sources include case studies, surveys, interviews, questionnaires, direct observation, and focus groups. Make sure to organize the collected data for analysis.

3.2.2 Data Cleaning:

Not all the data you collect will be useful, so it's time to clean it up. This process is where you remove white spaces, duplicate records, and basic errors. Data cleaning is mandatory before sending the information on for analysis.

3.2.3. Data Analysis:

Here is where you use data analysis software and other tools to help you interpret and understand the data and arrive at conclusions. Data analysis tools include Excel, Python, R, Looker, Rapid Miner, Charito, Meta base, Rehash, and Microsoft Power BI.

3.2.4. Data Interpretation:

Now that you have your results, you need to interpret them and come up with the best courses of action based on your findings.

3.2.5. Data Visualization:

Data visualization is a fancy way of saying, "graphically show your information in a way that people can read and understand it." You can use charts, graphs, maps, bullet points, or a host of other methods. Visualization helps you derive valuable insights by helping you compare datasets and observe relationships.

3.2.6 Types of Data Analysis:

The four types of data analysis are:

1. Descriptive Analysis
2. Diagnostic Analysis
3. Predictive Analysis
4. Prescriptive Analysis

3.3 DESCRIPTIVE ANALYSIS:

Descriptive analytics answer the question, “What happened?”. This type of analytics is by far the most commonly used by customers, providing reporting and analysis centered on past events. It helps companies understand things such as:

- How much did we sell as a company?
- What was our overall productivity?
- How many customers churned in the last quarter?

Descriptive analytics is used to understand the overall performance at an aggregate level and is by far the easiest place for a company to start as data tends to be readily available to build reports and applications. It’s extremely important to build core competencies first in descriptive analytics before attempting to advance upward in the data analytics maturity model. 23 Core competencies include things such as data modeling fundamentals, communicating data with right visualization and basic dashboard skills. It’s likely you’ve adopted some form of descriptive analytics internally, whether that be static P&L statements, PDF reports, or reporting within an analytics tool. For a true descriptive analytics program to be implemented, the concepts of repeatability and automation of tasks must be top of mind. Repeatability in that a data process is standardized and can be regularly applied with minimal effort (think a weekly sales report), and automation in that complex task (VLOOKUPS, merging of excel spreadsheets, etc.) are automated—requiring little to no manual intervention.

3.3.1. DIAGNOSTIC ANALYSIS:

Diagnostic analytics, just like descriptive analytics, uses historical data to answer a question. But instead of focusing on “what”, diagnostic analytics addresses the critical question of why an occurrence or anomaly occurred within your data. Diagnostic analytics also happens to be the most overlooked and skipped step within the analytics maturity model. Anecdotally, I see most customers attempting to go from “what happened” to “what will happen” without ever taking the time to address the “why did it happen” step. This type of analytics helps companies answer questions such as:

- Why did our company sales decrease in the previous quarter?
 - Why are we seeing an increase in customer churn?
 - Why is a specific basket of products vastly outperforming their prior year sales figures?
- Diagnostic analytics tends to be more accessible and fit a wider range of use cases than machine learning/predictive analytics. You might even find that it solves some business problems you earmarked for predictive analytics use cases. Being at the diagnostic analytics phase likely means you’ve adopted a modern analytics tool. Most modern analytics tools contain a variety of search-based, or lightweight artificial intelligence capabilities. ²⁴ These features allow for detailed insights a layer deeper (for example: the Key Drivers visualization in Power BI, or Qlik’s search-based insight functionality). To be clear, these are an effective lightweight means to address diagnostic analytics use cases but are not a means to a full-scale implementation. Software vendors like Sisuu have built their core business around addressing diagnostic analytics use cases (what they call “augmented analytics”) and are a great bet. Diagnostic analytics is an important step in the maturity model that unfortunately tends to get skipped or obscured. If you cannot infer why your sales decreased 20% in 2020, then jumping to predictive analytics and trying to answer “what will happen to sales in 2021” is a stretch in advancing upward in the analytics maturity model.

3.3.2. PREDICTIVE ANALYSIS:

Predictive analytics is a form of advanced analytics that determines what is likely to happen based on historical data using machine learning. Historical data that comprises the bulk of descriptive and diagnostic analytics is used as the basis of building predictive analytics models. Predictive analytics helps companies address use cases such as:

- Predicting maintenance issues and part breakdown in machines.
- Determining credit risk and identifying potential fraud.
- Predict and avoid customer churn by identifying signs of customer dissatisfaction.

At the outset of any predictive analytics build, three core elements need to be established:

- Identify a problem to solve,
- Define what is you want to predict, and
- State what you will achieve by doing so. To start you should collect existing data, organize data in a useful way to allow for data modeling, cleanse your data and review overall quality, and finally determine your modeling objective. 25 While modeling takes up the spotlight in predictive analytics, data prep is a crucial step that needs to happen first. This is why organizations with a rock-solid foundation in descriptive and diagnostic analytics are better equipped to handle predictive analytics. Simply put, the time and effort to prep, transform, and ensure data quality for retrospective reporting has already taken place. The groundwork should be relatively well laid to quickly identify and leverage data for the modeling phase. I always encourage customers with well-defined KPIs and business logic in a specific business reporting area (think sales reporting for example) to use that as the first predictive analytics use case. The goal is to derive value quickly, and there is no better place to start than an area where you know data is well defined and of high quality.

3.3.3. PRESCRIPTIVE ANALYSIS:

Prescriptive analytics is the fourth and final pillar of modern analytics. Prescriptive analytics pertains to true guided analytics where your analytics is prescribing or guiding you toward a specific action to take. It is effectively the merging of descriptive, diagnostic, and predictive analytics to drive decision making. Existing scenarios or conditions (think your current fleet of freight trains) and the ramifications of a decision or occurrence (parts breakdown on the freight trains) are applied to create a guided decision or action for the user to take (proactively buy more parts for preventative maintenance). Prescriptive analytics requires strong competencies in descriptive, diagnostic, and predictive analytics which is why it tends to be found in highly specialized industries (oil and gas, clinical healthcare, finance, and insurance to name a few) where use cases are well defined. Prescriptive analytics help to address use cases such as:

- Automatic adjustment of product pricing based on anticipated customer demand and external factors.
 - Flagging selects employees for additional training based on incident reports in the field.
- 26 Prescriptive analytics' primary aim is to take the educated guess or assessment out of data analytics and streamline the decision-making process. Prescriptive analytics is commonly considered the merging of descriptive, diagnostic, and predictive analytics. Getting started isn't so much a step-by-step list but rather the time and effort up front to build your competencies within the analytics maturity curve. Simply put, there is no starting point in prescriptive analytics without the requisite first three pillars of modern analytics being established first. If you're ready for prescriptive analytics, then quantifying your call to action and the underlying criteria will be the first requirement. For example: if the use case is to call corrective action for an employee (i.e. – additional training based on poor performance) then the factors that necessitate this action must be firmly established and the action itself must be clearly defined. Knowing how each kind of analytics helps you better understand your data and how to use it.

3.4 System Requirements:

3.4.1 Windows:

- Microsoft® Windows® 7/8/10 64-bit
- 4 GB RAM minimum, 8 GB RAM recommended
- 2 GB of available disk space minimum,
 - 4 GB Recommended 500 MB for IDE + 1.5 GB for Android SDK and emulator system image
- 1280 x 800 minimum screen resolution

3.4.2 Mac:

- Mac® OS X® 10.10 Yosemite or higher, up to 10.14 macOS Mojave
- 4 GB RAM minimum, 8 GB RAM recommended
- 2 GB of available disk space minimum,
 - 4 GB Recommended 500 MB for IDE + 1.5 GB for Android SDK and emulator system image

3.4.3 Linux:

- GNOME or KDE desktop - Tested on Linux based on Debian.
- 64-bit distribution capable of running 32-bit applications
- GNU C Library glibc 2.19 or later 17
- 4 GB RAM minimum, 8 GB RAM recommended
- 2 GB of available disk space minimum.

3.5 POWER BI:

Power BI is an interactive data visualization software product developed by Microsoft with a primary focus on business intelligence. It is part of the Microsoft Power Platform. Power BI is a collection of software services, apps, and connectors that work together to turn unrelated sources of data into coherent, visually immersive, and interactive insights. Data may be input by reading directly from a database, webpage, or structured files such as spreadsheets, CSV, XML, and JSON. Microsoft Power BI is a business intelligence (BI) platform that provides nontechnical business users with tools for aggregating, analyzing, visualizing and sharing data. 27 Power BI's user interface is fairly intuitive for users familiar with Excel, and its deep integration with other Microsoft products makes it a versatile self-service tool that requires little upfront training. Users can download an application for Windows 10, called Power BI Desktop, and native mobile apps for Windows, Android and iOS devices. There is also Power BI Report Server for companies that must maintain their data and reports on premises. That version of Power BI requires a special version of the desktop app -- aptly called Power BI Desktop for Power BI Report Server.

3.5.1 USES OF POWER BI:

Microsoft Power BI is used to find insights within an organization's data. Power BI can help connect disparate data sets, transform and clean the data into a data model and create charts or graphs to provide visuals of the data. All of this can be shared with other Power BI users within the organization. The data models created from Power BI can be used in several ways for organizations, including the following:

- telling stories through charts and data visualizations;
- examining "what if" scenarios within the data;
- creating reports that can answer questions in real time and help with forecasting to make sure departments meet business metrics. Power BI can also provide executive dashboards for administrators or managers, giving management more insight into how departments are doing

3.6 METHODOLOGY:

The project is implemented in Microsoft excel and Microsoft power Bi.

- We need to collect the data set for our product, over a year or 365 days.
- So, we are creating a data set for our product
- That includes:
 1. Count
 2. Month
 3. Student
 4. Strength
 5. Duration
 6. Mode of Travel
 7. Location
 8. Ethnicity
 9. Home Country
 10. Gender
 11. Age

Now enter the dataset in Microsoft excel sheet as per the date and month as show in the fig 3.2

Project-14 dataset - Excel (Product Activation Failed)

Search

Dhana lakshmi

File Home Insert Page Layout Formulas Data Review View Help

Share

Undo Cut Copy Paste Format Painter

Font: Calibri, 11, A⁺, A⁻, B, I, U, Color, Background Color, Merge & Center

Alignment: Left, Center, Right, Indent, Decrease Indent, Increase Indent, Wrap Text

Number: General, Text, Percentage, Fraction, Decimals, Thousands Separator, Comma Separator, Negative numbers in parentheses, Show zero for empty cells

Styles: Conditional Formatting, Format as Table, Cell Styles

Cells: Insert, Delete, Format

Editing: Fill, Sort & Filter, Find & Select, Clear

O21

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Count	Month	Student	Strength	Duration	Mode Of Travel	Location	Ethnicity	Home Country	Gender	Age							
2	528555	January	No	Group	5	Air	Delhi	Asian	UK	Male	5							
3	117961	February	Yes	Solo	8	Land	Kolkata	American	USA	Female	7							
4	157991	March	No	Solo	3	Air	Bangalore	Black	Cannada	Non-Binary	8							
5	103472	April	Yes	Solo	10	Water	Chennai	White	Russia	Female	10							
6	537490	May	No	Group	2	Air	Mumbai	Latino	USA	Male	12							
7	128575	June	No	Group	5	Land	Chennai	White	France	Female	15							
8	157748	July	No	Solo	4	Land	Kochi	American	Australia	Female	17							
9	141308	August	Yes	Group	8	Air	Delhi	Asian	Cannada	Female	20							
10	488012	September	No	Solo	2	Air	Bangalore	American	USA	Male	24							
11	115838	October	Yes	Solo	5	Land	Chennai	Latino	Australia	Male	26							
12	164735	November	No	Group	5	Water	Mumbai	Black	France	Male	28							
13	167902	December	No	Group	7	Air	Kolkata	Asian	Russia	Non-Binary	30							
14	236798	January	No	Group	4	Air	Mumbai	White	Russia	Male	32							
15	297789	February	No	Solo	3	Water	Chennai	American	Cannada	Male	35							
16	189573	March	No	Solo	4	Air	Mumbai	Asian	USA	Female	37							
17	256879	April	No	Group	5	Land	Chennai	Black	UK	Non-Binary	40							
18	127753	May	Yes	Group	10	Air	Delhi	White	UK	Male	43							
19	278564	June	No	Solo	2	Water	Bangalore	Latino	Russia	Male	47							
20	297501	July	No	Group	5	Air	Chennai	American	France	Non-Binary	50							
21	197564	August	Yes	Solo	6	Air	Kolkata	Latino	Australia	Non-Binary	54							
22	197533	September	Yes	Solo	6	Land	Kochi	Asian	Cannada	Female	56							
23	267454	October	No	Group	2	Air	Delhi	Black	UK	Female	58							
24	355679	November	No	Solo	1	Air	Kochi	American	USA	Male	63							
25	357809	December	No	Group	3	Air	Chennai	Black	France	Male	70							

Sheet1 Sheet2

Ready

34°C Partly sunny

Search

ENG IN

18:10 19-04-2023

FIG: 3.2 EXCEL DATA

Sum of Visitors: by Year and Month

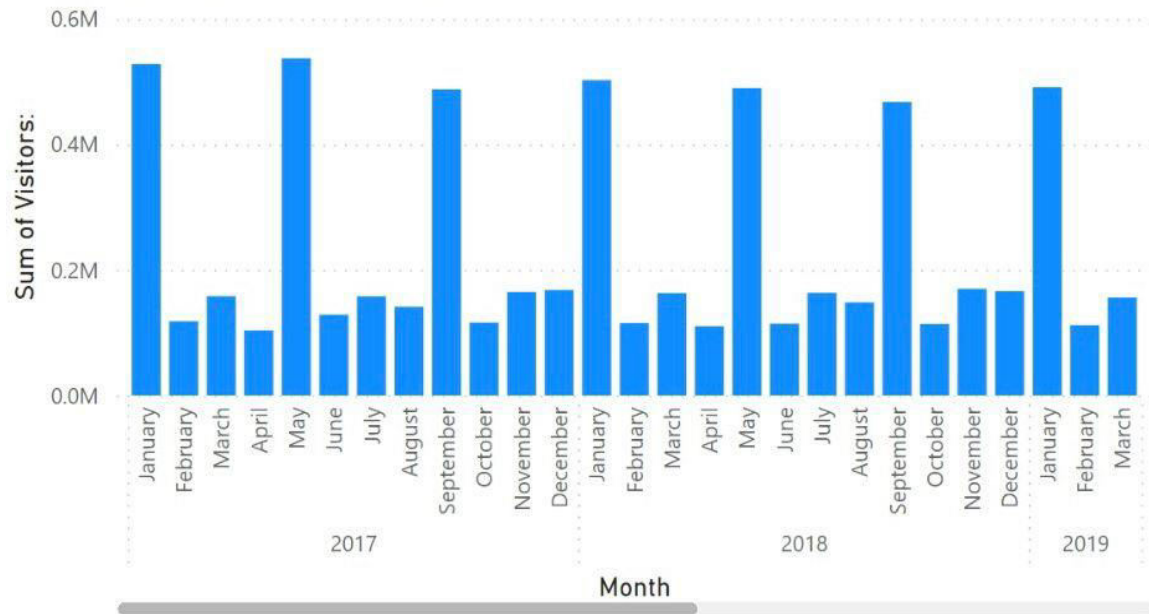


FIG: 3.3 SUM OF VISITORS BY YEAR

- In FIG:3.3 graph is between sum of Visitors and Month.
- In 2017 May has highest number of visitors.
- From 2017 to 2019 all April month has least number of visitors.

Sum of Count by Duration

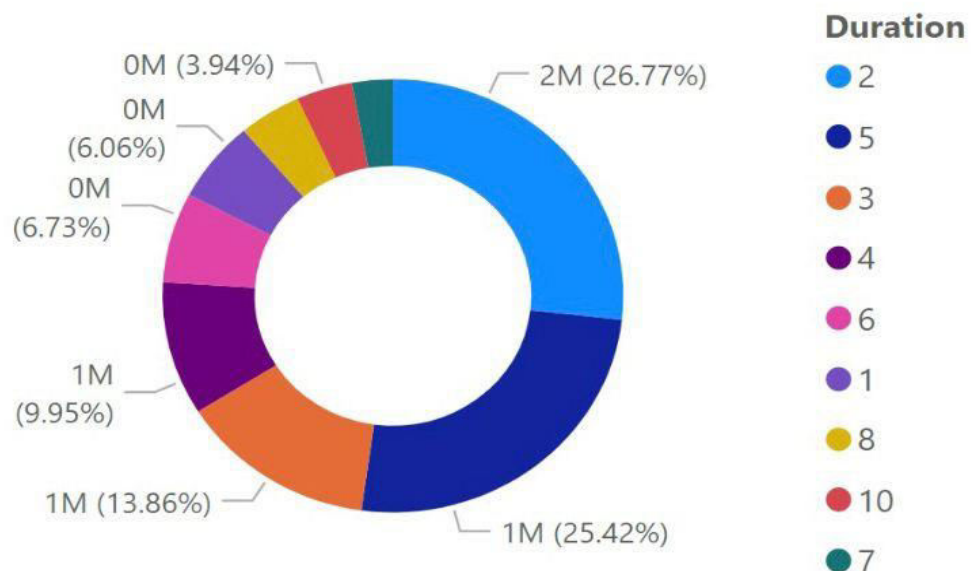


FIG: 3.4 SUM COUNT BY DURATION

- This Chart(fig:3.3) is between Sum of Count by Duration.
- People spend more duration at 5.
- 3 and 6 has same almost same duration.
- 7 has the least duration.

Sum of Count by Home Country

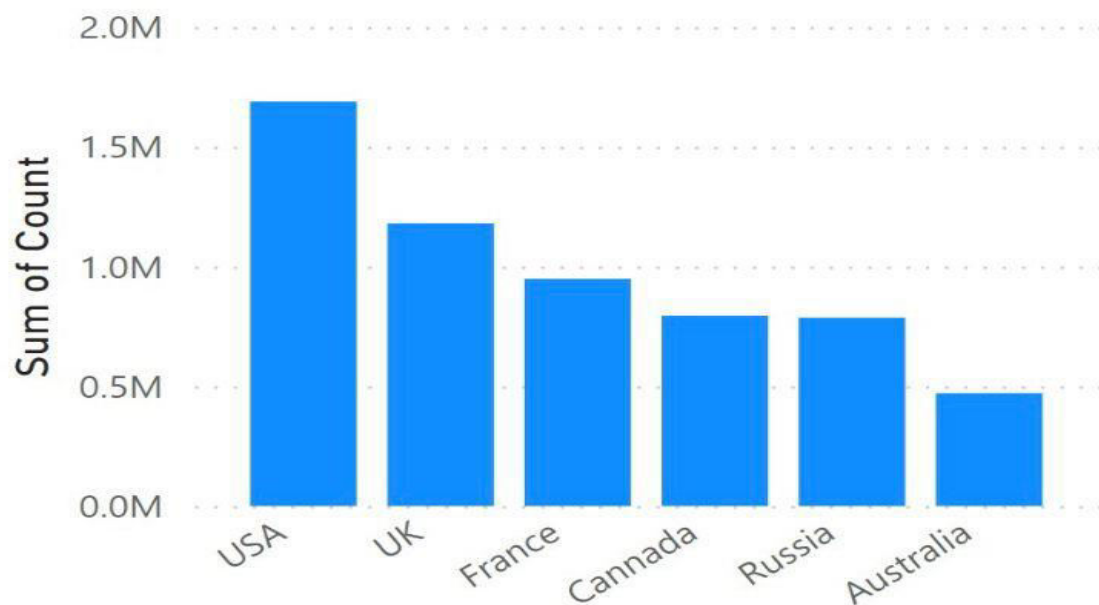


FIG: 3.5 SUM OF COUNT BY HOME COUNTRY

- USA visitors is more among all others.
- Canada and Russia as equal visitors.
- Australia as least Visitors among all countries.

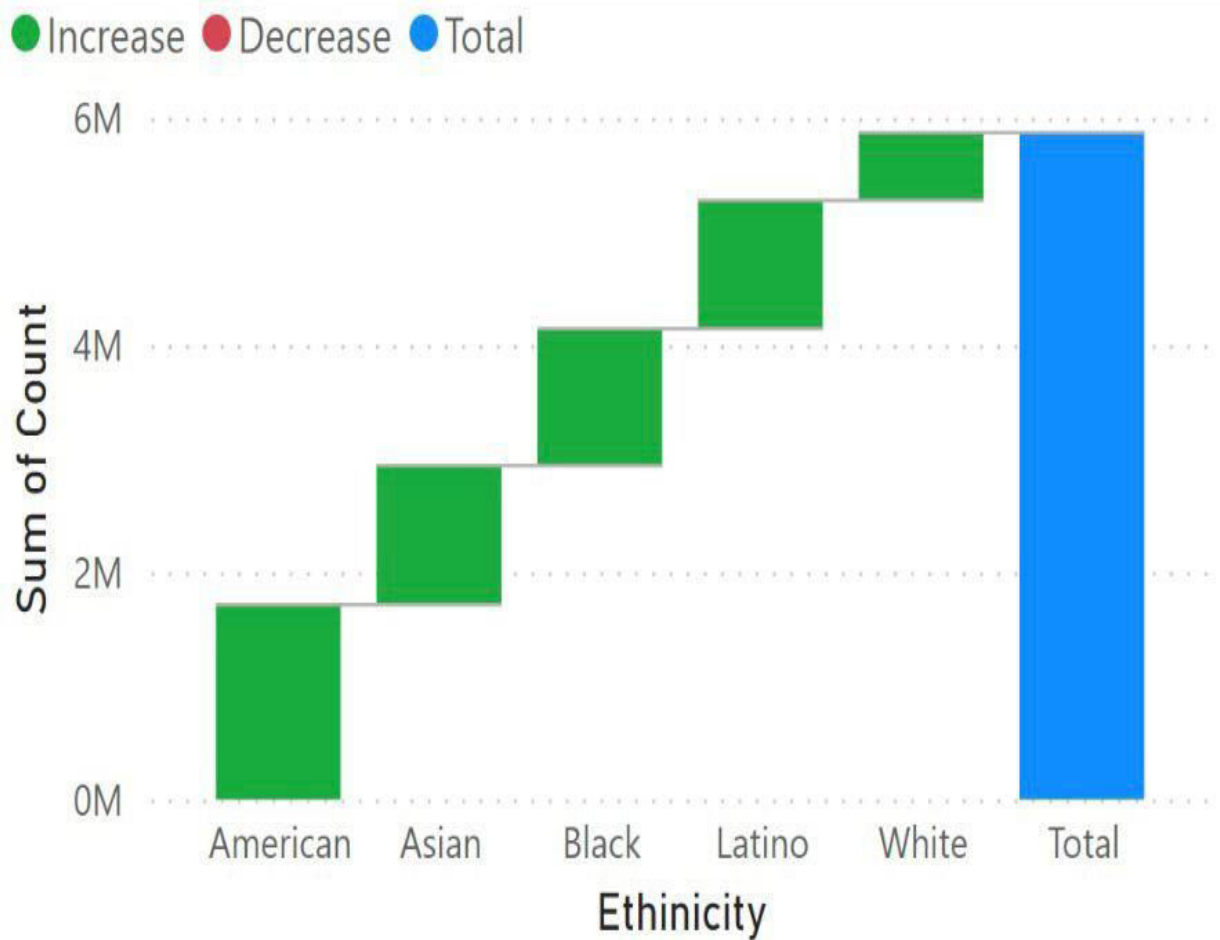


FIG: 3.6 SUM OF COUNT BY ETHNICITY

- American's have highest count than others.
- Black and Asian people as equal number of visitors.
- White people as least Visitors.

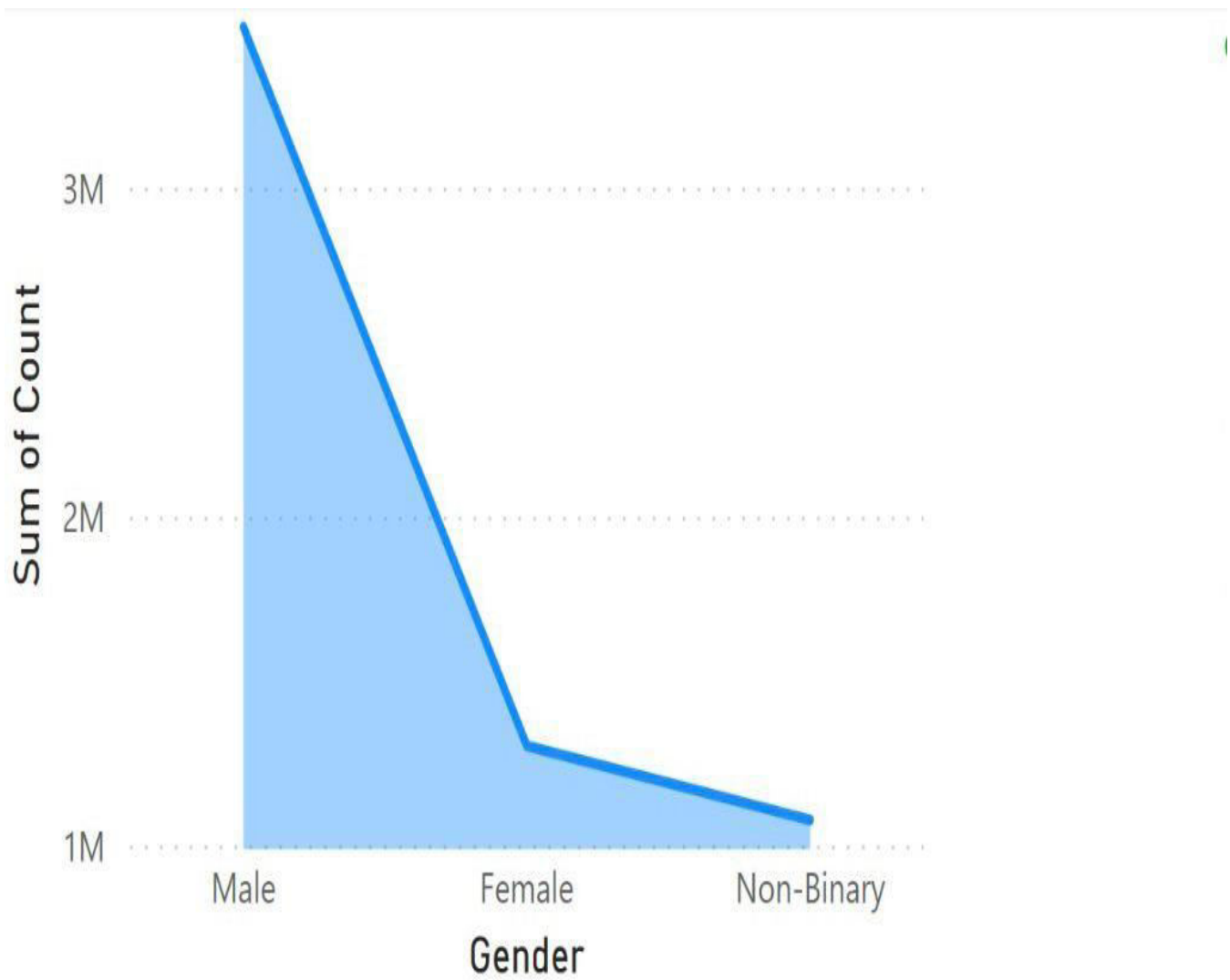


FIG: 3.7 SUM OF COUNT BY GENDER

- This graph is between Sum of count and Gender.
- There are more Male visitors than female visitors.
- Here we can also see non-binary visitors.

CHAPTER – 4

RESULT & DISCUSSION

The performance of the Prediction is mostly focused on Prediction accuracy, Accuracy in each table and confusion matrix which shows the number of predictions of each table which can be compared to the instances of each table. From the data visualization it was observed that the May month has the Visitors since it's a summer season we will have more offers which will attract customers to buy the product and also Velachery has highest sales among all other areas.

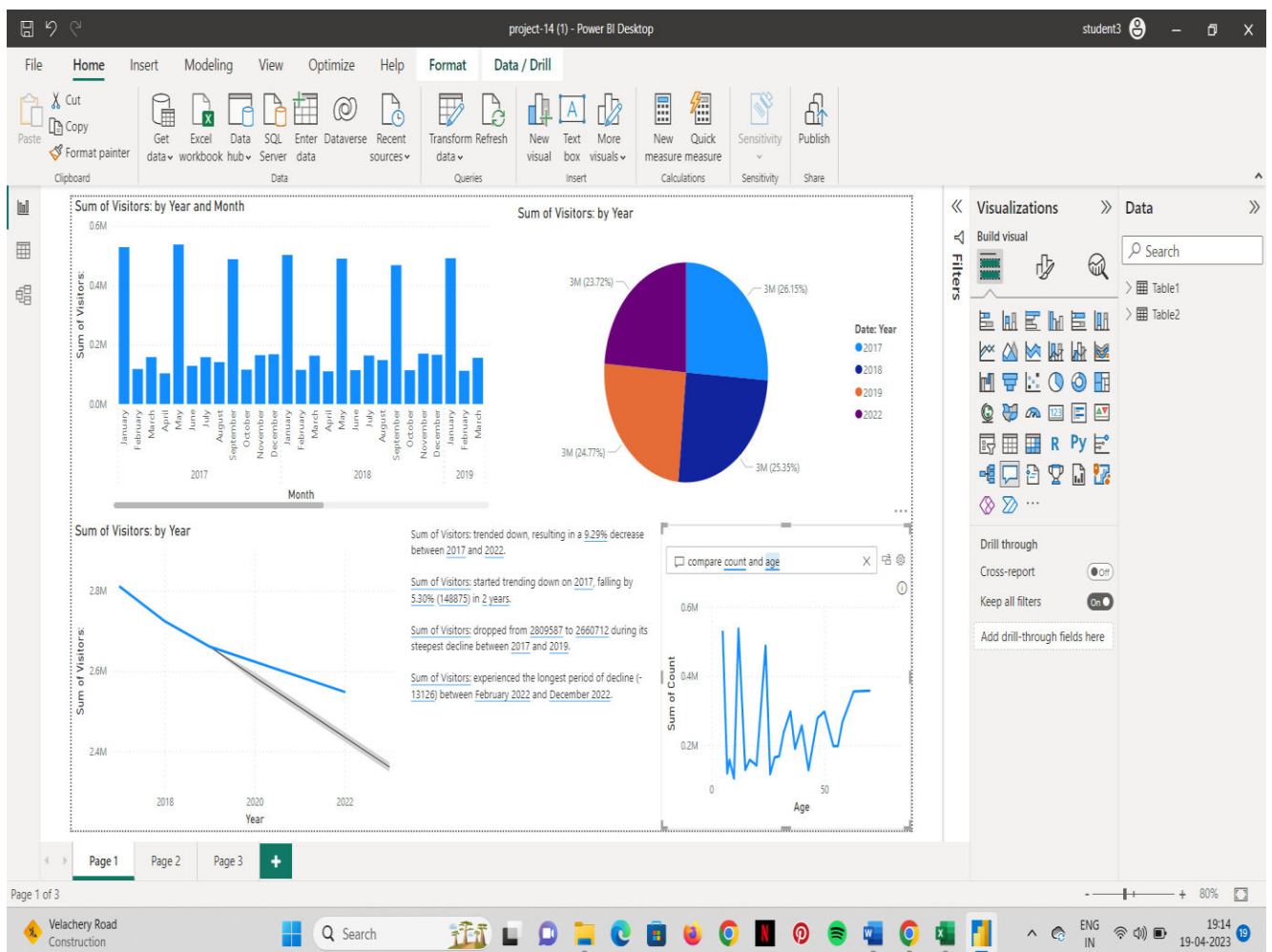


FIG:3.8 FOREIGN VISITORS PREDICTION

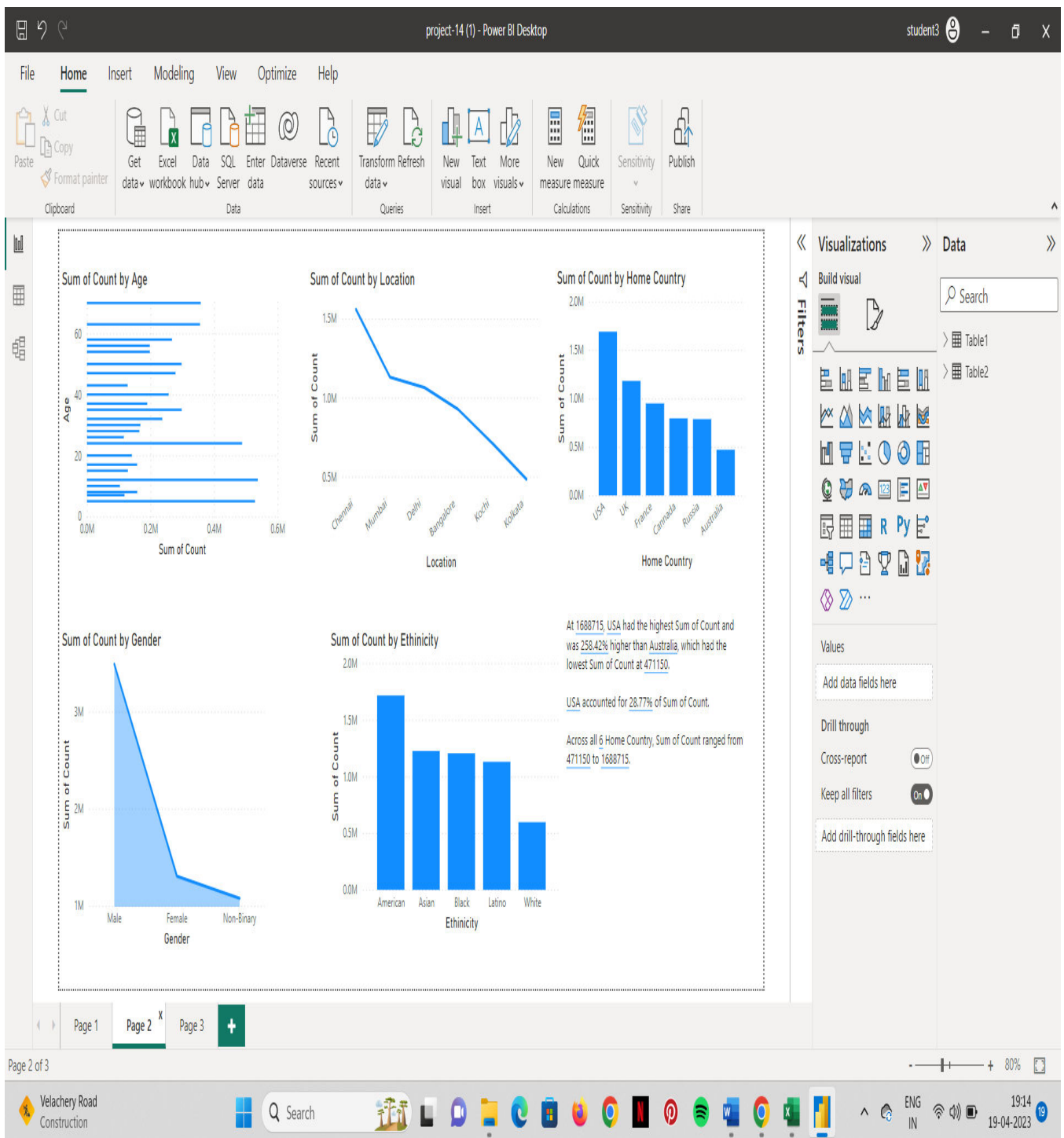


FIG:3.9 FOREIGN VISITORS PREDICTION

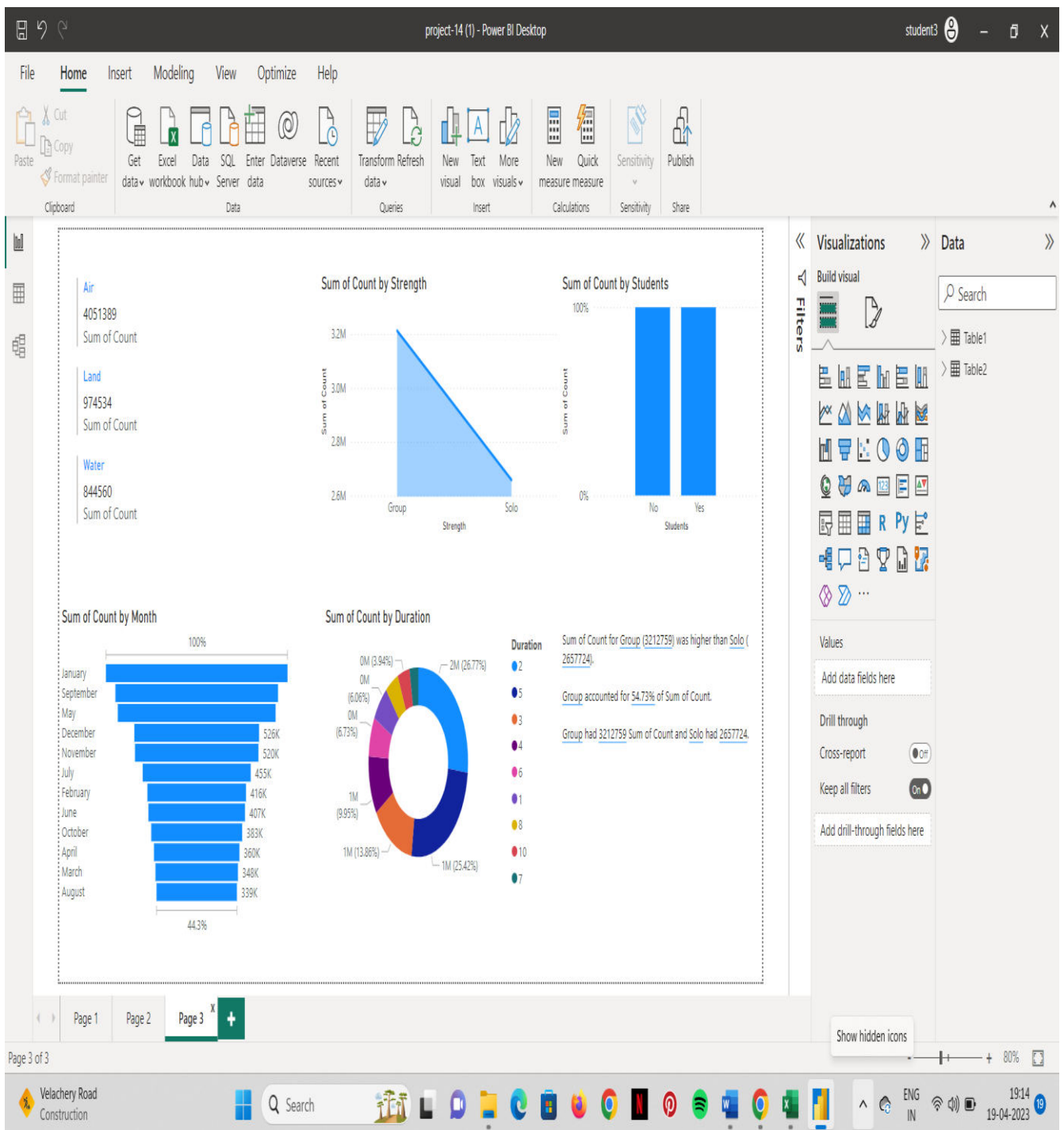


FIG:3.10 FOREIGN VISITORS PREDICTION

CHAPTER – 5

CONCULSION & FUTURE WORK

CONCLUSION:

Predicting future foreign visitor arrivals is critical for tourism planning and marketing purposes since it is currently one of India's key areas for economic growth. The Box-Jenkins method was used in this study to analyze data on monthly foreign visitor arrivals in India. Forecasters and government decision-makers can both profit from this work in different ways. It serves as a manual for improving sample predictions in the prediction context. These findings assist government decision-makers in developing appropriate policies, planning and providing resources to society, and making investment decisions. his study also found evidence of seasonal changes, indicating the need for appropriate action to address the problem of declining numbers of foreign visitors and tourism revenue during low season. For instance, the government should run tourism campaigns, ease international traveler visa requirements, create welcoming travel policies, introduce direct flights to cut down on the number of tourists arriving by road, and appoint tourism ambassadors to market the nation's tourist destinations abroad as well as promote domestic tourism.

FUTURE WORK:

The predictions presented in this study provide strong evidence about India's future foreign tourist profile and the reasons for consumers' travel abroad behavior. The developed machine-learning model is effectively trained even though the training data is limited and uncertain. The training accuracy of the model confirms that the dominant character of the departing foreign visitors by reasons is correctly captured. In addition, the learned model is robust and stable since it can provide smooth and bounded for the next ten years of multi-dimensional predictions. In particular, increases in arrivals due to health, transit, and travel offer valuable opportunities for Indian tourism to gain a competitive advantage in these areas.

APPENDIX SCREENSHOTS

Project-14 dataset - Excel (Product Activation Failed)

Search

Dhana lakshmi

File Home Insert Page Layout Formulas Data Review View Help

Undo Paste Copy Format Painter Clipboard Font Alignment Number Styles Cells Editing

O21

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	Count	Month	Student	Strength	Duratic	Mode Of Travel	Location	Ethnicity	Home Country	Gender	Age							
1	528555	January	No	Group	5	Air	Delhi	Asian	UK	Male	5							
2	117961	February	Yes	Solo	8	Land	Kolkata	American	USA	Female	7							
3	157991	March	No	Solo	3	Air	Bangalore	Black	Cannada	Non-Binary	8							
4	103472	April	Yes	Solo	10	Water	Chennai	White	Russia	Female	10							
5	537490	May	No	Group	2	Air	Mumbai	Latino	USA	Male	12							
6	128575	June	No	Group	5	Land	Chennai	White	France	Female	15							
7	157748	July	No	Solo	4	Land	Kochi	American	Australia	Female	17							
8	141308	August	Yes	Group	8	Air	Delhi	Asian	Cannada	Female	20							
9	488012	September	No	Solo	2	Air	Bangalore	American	USA	Male	24							
10	115838	October	Yes	Solo	5	Land	Chennai	Latino	Australia	Male	26							
11	164735	November	No	Group	5	Water	Mumbai	Black	France	Male	28							
12	167902	December	No	Group	7	Air	Kolkata	Asian	Russia	Non-Binary	30							
13	236798	January	No	Group	4	Air	Mumbai	White	Russia	Male	32							
14	297789	February	No	Solo	3	Water	Chennai	American	Cannada	Male	35							
15	189573	March	No	Solo	4	Air	Mumbai	Asian	USA	Female	37							
16	256879	April	No	Group	5	Land	Chennai	Black	UK	Non-Binary	40							
17	127753	May	Yes	Group	10	Air	Delhi	White	UK	Male	43							
18	278564	June	No	Solo	2	Water	Bangalore	Latino	Russia	Male	47							
19	297501	July	No	Group	5	Air	Chennai	American	France	Non-Binary	50							
20	197564	August	Yes	Solo	6	Air	Kolkata	Latino	Australia	Non-Binary	54							
21	197533	September	Yes	Solo	6	Land	Kochi	Asian	Cannada	Female	56							
22	267454	October	No	Group	2	Air	Delhi	Black	UK	Female	58							
23	355679	November	No	Solo	1	Air	Kochi	American	USA	Male	63							
24	357809	December	No	Group	3	Air	Chennai	Black	France	Male	70							

Sheet1 Sheet2

Ready 34°C Partly sunny

Project-14 dataset - Excel (Product Activation Failed)

Search

Dhana lakshmi

File Home Insert Page Layout Formulas Data Review View Help

Undo Paste Copy Format Painter Clipboard Font Alignment Number Styles Cells Editing

D349

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
	Date	Visitors																					
1	01-01-2017	9206																					
2	02-01-2017	10732																					
3	03-01-2017	12527																					
4	04-01-2017	13961																					
5	05-01-2017	16292																					
6	06-01-2017	19607																					
7	07-01-2017	22145																					
8	08-01-2017	25144																					
9	09-01-2017	28622																					
10	10-01-2017	31531																					
11	11-01-2017	34794																					
12	12-01-2017	38526																					
13	13-01-2017	8840																					
14	14-01-2017	10305																					
15	15-01-2017	11881																					
16	16-01-2017	13688																					
17	17-01-2017	15825																					
18	18-01-2017	18685																					
19	19-01-2017	20912																					
20	20-01-2017	23569																					
21	21-01-2017	27061																					
22	22-01-2017	29404																					
23	23-01-2017	32030																					
24	24-01-2017	35179																					
25	25-01-2017	1654																					
26	26-01-2017	1850																					
27	27-01-2017	2468																					

Sheet1 Sheet2

Ready 31°C Partly cloudy

Project-14 dataset - Excel (Product Activation Failed)

Search

Dhana lakshmi

File Home Insert Page Layout Formulas Data Review View Help

Undo Paste Copy Format Painter Clipboard Font Alignment Number Styles Cells Editing

Calibri 11 A+ A-

B I U Merge & Center Conditional Formatting Format as Table Cell Styles Insert Delete Format AutoSum Fill Sort & Filter Find & Select

D349

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
64	04-03-2017	1114																					
65	05-03-2017	1250																					
66	06-03-2017	1424																					
67	07-03-2017	1624																					
68	08-03-2017	1990																					
69	09-03-2017	2137																					
70	10-03-2017	2334																					
71	11-03-2017	2435																					
72	12-03-2017	2484																					
73	13-03-2017	2594																					
74	14-03-2017	3930																					
75	15-03-2017	4600																					
76	16-03-2017	5258																					
77	17-03-2017	6386																					
78	18-03-2017	7421																					
79	19-03-2017	8710																					
80	20-03-2017	9620																					
81	21-03-2017	11106																					
82	22-03-2017	12702																					
83	23-03-2017	14383																					
84	24-03-2017	15811																					
85	25-03-2017	17332																					
86	26-03-2017	1533																					
87	27-03-2017	1831																					
88	28-03-2017	2095																					
89	29-03-2017	2378																					
90	30-03-2017	2723																					

Sheet1 Sheet2

Ready 31°C Partly cloudy Search 19:40 19-04-2023

Project-14 dataset - Excel (Product Activation Failed)

Search

Dhana lakshmi

File Home Insert Page Layout Formulas Data Review View Help

Undo Paste Copy Format Painter Clipboard Font Alignment Number Styles Cells Editing

Calibri 11 A+ A-

B I U Merge & Center Conditional Formatting Format as Table Cell Styles Insert Delete Format AutoSum Fill Sort & Filter Find & Select

D349

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
151	30-05-2017	4303																					
152	31-05-2017	5055																					
153	01-06-2017	5576																					
154	02-06-2017	6701																					
155	03-06-2017	6889																					
156	04-06-2017	7206																					
157	05-06-2017	7992																					
158	06-06-2017	2104																					
159	07-06-2017	2524																					
160	08-06-2017	2871																					
161	09-06-2017	3179																					
162	10-06-2017	3658																					
163	11-06-2017	4272																					
164	12-06-2017	4485																					
165	13-06-2017	5027																					
166	14-06-2017	5557																					
167	15-06-2017	5786																					
168	16-06-2017	6519																					
169	17-06-2017	7225																					
170	18-06-2017	1568																					
171	19-06-2017	1915																					
172	20-06-2017	2189																					
173	21-06-2017	2417																					
174	22-06-2017	2765																					
175	23-06-2017	3207																					
176	24-06-2017	3347																					
177	25-06-2017	3778																					

Sheet1 Sheet2

Ready 31°C Partly cloudy Search 19:40 19-04-2023

Project-14 dataset - Excel (Product Activation Failed)

Search

Dhana lakshmi

File Home Insert Page Layout Formulas Data Review View Help

Undo Paste Copy Format Painter Clipboard Font Alignment Number Styles Cells Editing

Calibri 11 A⁺ A⁻ B I U Merge & Center Wrap Text Conditional Formatting Format as Table Cell Styles Insert Delete Format AutoSum Fill Sort & Filter Find & Select

D349

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
267	23-09-2017	3075																					
268	24-09-2017	3340																					
269	25-09-2017	3252																					
270	26-09-2017	3264																					
271	27-09-2017	3587																					
272	28-09-2017	3831																					
273	29-09-2017	3810																					
274	30-09-2017	4080																					
275	01-10-2017	4011																					
276	02-10-2017	4012																					
277	03-10-2017	4333																					
278	04-10-2017	2970																					
279	05-10-2017	3368																					
280	06-10-2017	3685																					
281	07-10-2017	3817																					
282	08-10-2017	4227																					
283	09-10-2017	4605																					
284	10-10-2017	4577																					
285	11-10-2017	4865																					
286	12-10-2017	5150																					
287	13-10-2017	5434																					
288	14-10-2017	6207																					
289	15-10-2017	6739																					
290	16-10-2017	2177																					
291	17-10-2017	2494																					
292	18-10-2017	2756																					
293	19-10-2017	2847																					
294	20-10-2017	3306																					

Sheet1 Sheet2

Ready 31°C Partly cloudy 19:40 19-04-2023

REFERENCES

- [1] Elgin, A.; Gokcu, M.; Gül, S.; Kazdal, A. *Turizmin Büyüme ve İstihdam Üzerindeki Etkileri*; The Central Bank of the Republic of Türkiye: Ankara, Türkiye, 2021; Volume 2, pp. 1–9.
- [2] Gilal, F.G.; Paul, J.; Gilal, N.G.; GiR, G. The role of organismic integration theory in marketing science: A systematic review and research agenda. *Eur. Manag. J.* **2022**, *40*, 208–223.
- [3] Kozak, M. Comparative analysis of tourist motivations by nationality and destinations. *Tour. Manag.* **2002**, *23*, 221–232.
- [4] Birdir, K.; Akgöl, Y. Gastronomy Tourism and Analysis of Gastronomy Experiences of the Foreign Tourists Visiting Türkiye.
- [5] Andreu, L.; Kozak, M.; Avci, N.; Cifter, N. Market segmentation by motivations to travel: British tourists visiting Türkiye. *J. Travel Tour. Mark.*