Southampton Solent University

MSc Applied Artificial Intelligence and Data Science

Academic Year 2021-2022

R.P.Machaka

Analysing sales data for business growth

Supervisor: Dr Drishty Sobnath

September 2022

This report is submitted in partial fulfilment of the requirements of Solent University for MSc Artificial Intelligence and Data Science

Contents Page

Acknowledgements

Firstly, I would like to extend my deepest gratitude to my ever so hardworking and accommodating lecturers Drishty.S, Femi.I, Prins.B and Dr Shakeel. I want to thank you for the knowledge you have imparted during this Masters

My completion of this project would not have been successful without the support of my colleagues Georgette, Adrian, Margaret, and Isaac. I appreciate the time and effort they have dedicated as a group to share knowledge and skills within the Information Technology Industry.

Lastly, to my caring family I would like to thank you from the bottom of my heart for believing in me, you recognized a lot of potential in me that I did not realize I possess. My heartfelt thanks.

Abstract

This masters project focuses on analyzing sales data to ensure a distinct understanding within National Tickets company as well as derive effective and constructive business decisions. According to National Tickets executives, the company has not been able to track trends in their sales data for the past years and they consider lack of analyzing sales data is the chief contributor to the near demise of their business. The primary aim of this project is to ensure detailed and transparent insights are drawn from the sales data provided by National tickets, to instigate valuable and well thought out business decisions. The system of gathering, cleaning, manipulating and visualising data aids organisation's executives to make informed decisions (Kościelniak et al,2015).

Data analysis is one of the essential processes that ought to be performed for business success this is merely because data driven companies are able to experiment numerous value propositions on different segments of clients to ascertain which are most effective. In addition, analysing sales data exposes many valuable insights that the company can utilise to cut cost and improve productivity.

To satisfy the request from National Tickets executives I utilised data cleaning and analytics tools such as Microsoft Excel and Python, additionally I utilised Streamlit a tool that supports website application building to ensure results are presented distinctively to the end user. It is of vital importance to note that results attained indicated that most of National Tickets clients are individual clients and therefore there is an urgent need to create marketing strategies tailor-made for corporate clients to maximize profits. Additionally, the difference noted between the highest salesperson and the lowest salesperson in terms of sales achieved during the stipulated period, is extremely enormous and leaves a lot to be desired. National Tickets executives ought to strategize on how they can offer thorough training to their staff as well as re-strategize their recruitment process to ensure highly skilled individuals are employed for the benefits of the company. Moreover, performance audits in order to identify and execute tailored training in order to engage employees accordingly and work effectively and efficiently.

Chapter 1: Introduction and Background

1.1 Introduction

Analysing trends in sales data has gradually become a pertinent service to earmark business growth. Taking a closer look at the elapsed years coupled with the dynamism of technology, flight operators have had to content with erratic business operations. Henceforth, this project aims to draw insights from National Tickets sales data to determine which flight operators are the most valuable in terms of business growth, with the aim of maximizing profits as well as developing strategies to rectify the least contributing components to the business. Analysing data exposes crucial trends that are likely be lost in bulk information, these trends can enhance the effectiveness of a business. Madamanchi and Sreekrishna (2020).

National Tickets Zimbabwe is a travel agency established in 2015 that operates in the city of Harare in Zimbabwe, their core business is selling flight tickets to individual and corporate clients. The main thrust of the company is flight travelling; however, it seeks to manoeuvre into the accommodation arena as a diversification bid. That in mind, this calls for the need to analyse the operations of various stakeholders that National Tickets has been having a symbiotic relationship with as a roadmap to the future endeavours and many prospects. Data analysis is the art of transforming raw data into meaningful insights for the benefits of the organisation Madamanchi and Sreekrishna(2020) and this is what the project stems from in its endeavour to analyse the data that is available from the previous performances of the company.

1.2 **Statement of Problem**

National Tickets has recently encountered several upheavals that have threatened the demise of the business. Executives of this company strongly believe the problems are emanating from lack of analysing sales data to determine trends and business performance. The company is in urgent need of information such as type of clients mostly contributing to their income, staff performance, customer behaviour and time periods that clients travel the most. In addition, to undertake the accommodation arena the company would like to identify the most travelled routes and build their new business in the destination that most clients travel to.

1.3 Aims and Objectives

The time and effort involved in securing contracts with operators can be lengthy and tedious therefore business decisions to cancel contracts with the least performing operators will be made and all efforts will be diverted to the high contributing operator. The project aims to find developing opportunities, uncover possible threats, disclose more industry insights, and strengthen decision-making for the purposes of business success.

The rationality behind examining sales data is to instil growth within the organisation that will help in attaining resilience in the market, elevated market dominance, commercial risk mitigation through diversification and to minimize the threat of competition. Data that is processed in a meaningful manner supports the organisations decision-making procedure.

Chapter 2: Literature Review

Overview

This chapter seeks to explore various literature in relation to the impact of big data analytics on effective decision-making for the purposes of business success. The study strives to provide in-depth scrutiny performed by the researcher on Big data analytics to ensure enhanced comprehension of the research.

2.1 Big Data Analytics

According to Davenport and Jill (2013), Big data is classified as higher and richer data that depicts more details about behaviours, activities and events that happened all around. In agreement, Laney et al. (2013),alludes that, big data is the "*high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making*". According to Favaretto et al. (2020) big data is depicted as a game changer in techniques organisations function across various industries. Insights extracted from big data facilitate organisations to yield new products, services and business models that satiate client needs. Apart from the competitive advantage derived from analysing big data there are also numerous benefits that stand to be gained, this was expressed by, Das and Kumar (2013) in agreement to, Favaretto et al. (2020) view.

The usage of big data analytics in existing business settings produce a competitive edge and uplifts the efficiency of data exploration. Kościelniak and Puto (2015) declared through a research study that big data analytics has the ability to elevate decision-making and support the digging out of meaningful insights that have not been recognized before. This Literature of big data analytics displays a considerable link amid the utilisation of big data analytics and organisations success. This literature has further indicated how big data analytics allows organisations to scrutinize, monitor, evaluate and achieve business goals across lens of data.

The field of data analytics has been uttered in primary sections which include data creation, data procurement, data storage, data analysis, data visualisation and value creation initiated through effective decision making Saggi and Jai(2018). The digital transformation era has led organisations recognize the magnitude of initiating right decisions at the right interval, which is induced from relying on appropriate data (Joseph and Gaba,2020). In this aspect it has been

noted that well-versed decision-making methods that incorporate big data as inputs can be utilized to tackle contemporary obstacles in businesses. Big data allows originated time-sensitive decisions and aids organisations to invigilate modern trends swiftly and professionally within the industry. The latter permit firms to influence organisational aptitude that instils new business opportunities and generate developed products and services for clients Halaweh and Massry(2015).

To add on, big data is also perceived from the process of gathering, cleaning, manipulating, mining and evaluating based on the '5 Vs' which are velocity, value, veracity, volume and variety to extract momentous value that establish performance and originate viable returns Wamba et al., (2015). It is of vital importance to put into consideration the constantly changing business environment has compelled businesses around the globe to endeavour competitive lead through the use of modern technology to analyse data and enrich well thought out decisions. In this regard, Mazzei and Noble (2017) contend that diverse and complex datasets have impacted the formulation of analytical techniques and methods for re-defining productivity in all sectors.

Decision-making driven by data analysis has the capability to be utilized to amend methods in which resources are handled and lead to the success of businesses. To fulfil the aforesaid research gaps, the purpose of this study is to investigate the impact of analysing big data to extract meaningful insights for the purposes of business success and growth. To add on, the study seeks to understand the extraction of insights from National Tickets sales data and its impact on decision-making and business growth, this approach supplies a distinct understanding on the role of big data analytics on organisational success. More specifically the study findings validate the latter as the insights drawn from analysing National Tickets sales data indicated various areas of concern that may have contributed to the near demise of the business and urgent need for decision-making. Furthermore, the study justifies the impact of data analysis on the relationship between decision-making and business growth.

It is essential to note that a study similar to this research was once conducted by Ahmed et.al(2022), who researched on the impact of data analysis on project success. The aforementioned study critically analysed the effects of big data analytics on project success and used a survey instrument where data from 135 respondents was collected. The study later utilised the sample size to draw conclusions on the relationship between big data analytics,

decision-making and project success. The researchers further checked for skewness and kurtosis to ensure there is normal distribution in the data collected during surveys before they performed regression analysis. To add on, the research was geographically limited to Information Technology and telecommunication in Pakistan therefore drawing conclusions based on that leaves a lot to be desired. This study seeks to conceal the gap as it is utilising a tourism industry which is different as compared to the previously mentioned study.

In my opinion, the study was not maximised to its greater strengths due to the sample size. In their limitations section, the study did not mention the disadvantages of using a small sample size to draw inferences. A small sample size increases the margin of error and conceives voluntary response bias which depicts the study irrelevant. The outcome of such research is likely to be skewed with no normal distribution. A study with sufficient sample size is likely to have added predictive power due to enough data points gathered to make satisfactory inference analysis. Unlike the aforementioned study, this study is using more data points to draw conclusions on the impact of data analytics on effective decision-making for business success.

This study explores National Tickets sales data to ensure meaningful information is attained for the benefits of good decision making and business growth. In contrast with the aforementioned research, this study is utilising sales data that has been collected from January 2019 to May 2022 and contains 1,629 observations. The use of a big dataset is attributed to more accurate findings hence the decision to utilise big data to exonerate bias from the study.

2.2 Big Data Analytics and Business Growth

Big data has the ability to shape businesses and be central to the growth of such businesses as can be buttressed with how companies like Amazon, eBay and Alibaba as online selling hubs have had to also rely on big data to scrutinise the factors surrounding performance of their platforms and user interactions. Big Data strives to collect and analyse mass volumes of data and the footrails that clients leave unravel significant customer behaviour Madamanchi and Sreekrishna (2020). The insights drawn from analysing customer data, is what has made it possible for companies to provide products or services that suit the customers and is able to provide gratification to customers thereby in turn giving them room for business expansion.

Business growth is heavily reliant on understanding the market you operate in. This has been the case in point where due to the chaotic 2 years the world has endured, most companies have

had to adapt or rather rely on Big data analytics as most business shifted to online due to the Covid-19 pandemic. United Nations Conference of Trade and Development (2021) bemoans that, "as lockdowns became the new normal, businesses and consumers increasingly went digital, providing and purchasing more goods and services online, raising e-commerce's share of global retail trade from 14% in 2019 to about 17% in 2020". In this view, the need for more online presence in the business circles due to the pandemic, the salience of big data could not have been more precious. Wang and Zhao (2016) expressed their view on how big data utilisation is widely accepted by many industries and companies, which leads to improved business processes for many huge demands and applications in multiple industries. These assertions all add weight to how big data is of paramount importance to business growth especially with the world going digital due to the pandemic thus even companies had to adapt and move with the flow. This need feeds on how the online service provision is heavily reliant on understanding the clientele's needs which can be channelled towards new business ventures for the company and Alsghaier et. Al (2017) alludes that big data analytics aids organizations to utilize their data and employ it to find new opportunities.

2.3 Sales Data Analysis

Sales data refers to information regarding company performance and important insights around customer behaviour. Analysing sales data helps identify under-performing sections within the organisation and allows the company to further investigate the reason behind under-performing and also use the findings in refining products, services to increase productivity. Additionally, there is a possibility of identifying products that no longer bring value to the company and dropping these products will aid in minimizing costs. This allows the business to focus more attention on products that generate more revenue and profits.

In agreement, Zikopoulos et al. (2013) mentioned, It is essential to note that analysing sales data delivers numerous crucial insights that the organisation can utilise to reduce costs, improve products, and increase productivity. Reliable sales data, permits managers to effectively assess staff performance and rectify any issues related to staff performance within the organisation.

Research performed by Federick Reichheld of Bain and Company indicates that 5% increase in repeat clients, increases the organisations profit by 25% to 95%. The latter can be attained by using information generated from sales data analysis that reveals the clients who buy the most

and ensuring the sales team maximize efforts on these clients. Sales data analysis aids in identifying the uppermost reason that influence clients to churn and allows the company to reach out and thoroughly address the customers apprehensions.

Utilizing sales data to improve advertising techniques allows organisations to identify the most profitable clients and divert efforts to targeting those clients as they are good fit for the firm Manyika et al. (2011). In a certain period of time the company is able to collect data that contain information on how to tailor marketing techniques and sales efforts to suit the needs of the most contributing clients and significantly expand revenue and profits.

Chapter 3: Research Methodology

Chapter Overview

The research methodology for this study is quantitative as it focuses on measuring the quantity of values in sales data variables to assess business performance. The research is utilising historical records collected from January 2019 to May 2022.

A stepwise technique was used as one of the pre-requisites for this project was to communicate with the managing executives of National Tickets to ensure the project meets their expectations and the methods are tailor-made to suit their objectives. To guarantee that the clients' goals were met, there was essential need to have a strategic plan on which tools were suitable to be used for the project and how they were going to be utilized. The continual communication with the client facilitated in the guiding the project.

Performing Explanatory Data Analysis helps in exposing business performance. National Tickets seeks to make continuous improvements to the services they are offering clients, to identify which of its branches within Zimbabwe is performing well, as well as identify which routes are being visited more to allow business growth by factoring in accommodation services in the most visited destinations. After making the objective crystal clear, tools suitable and in accordance with what had been discussed were considered and put in place. Data analysis and graphic designing tools such as Microsoft Excel, Python, Canva and Streamlit were used to analyse and visualize data to ensure distinct and well-define understanding among all stakeholders.

3.1 Data Collection

The sales data being used for this project is primary data provided by from National Tickets Zimbabwe. It is vital to note that this historical data was gathered over three years and 5 months, it dates back from January 2019 to May 2022. The sales data being utilised was supplied in excel format containing 14 columns and 1,629 rows.

To achieve the goal of this project it is vital to understand how and why National Tickets collect their sales data. The company has a call centre that receives all incoming calls from clients who want to book tickets or who may have enquiries in relation to services being

offered by the company. The call centre records details of the caller and ask the for more details from the caller, this is merely because when travelling from one country to another border security checks are completed, and this requires that all details on your flight ticket are correct including your name.

After a sale has been finalised the call centre personnel then enters all details into the system including amount paid, this is how the sales data is gathered. The overall goal is identifying trends in sales data to conceive constructive business decisions based on the insights derived from data. The table below patently explains the variables in the sales data that was used for this project.

Fig1.Data Description Table

Field Variables	Field Description	Type of Variable
Date	The date the flight ticket was sold	Independent
Customer Name	The name of the customer who purchased a ticket	Independent
Route	The route and destination the client is travelling to	Independent
Flight	The flight embarked	Independent

ZWL Cost	The total cost of the ticket in Zimbabwean Dollar (ZWL)	Independent
Total Cost	The total cost in (USD) before National tickets Mark up.	Independent
Sale	The total sale after National Tickets mark up	Dependent
Payment Method	The payment method used by client to purchase ticket	Independent
Paid or Not Paid	This variable shows if client paid or did not pay	Independent
Receipt Number	The number on receipt	Independent

	issued after payment was received	
Invoice Number	The number on invoice issued, a bill with a list of services provided	Independent
Salesperson	The National Tickets staff who closed the sale	Independent
Client Type	This variable shows if the client who purchased flight ticket is corporate or individual	Independent
Account Type	The type of account payment was received from	Independent

3.2 Environment

The project utilized data analytics tools such as Microsoft Excel, Python version 3.10 and Streamlit version 1.2.0. The tools mentioned in the previous statement were deployed to extract meaningful insights and trends in the data. It is of vital importance to note that the data

provided is extremely untidy as the company was not using Excel formulars while entering data therefore there was immense need to clean the dataset before extracting any insights. The dataset size is compatible with Microsoft Excel hence why part of data cleaning was executed in Microsoft Excel. Microsoft Excel contains functions that aid in cleaning data to ensure data used for Explanatory Data Analysis is good quality which is essential for accurate results.

3.3 Microsoft Excel

3.3.1 Data Cleaning using Microsoft Excel

Data cleaning is an essential stage that ought to be performed before analyzing the data. The importance of data cleaning is driven from the logic that, it eliminates noisy data, that is likely to produce biased results if left ignored. To start with, there was need to remove characters such as **=, -, $, /,** which were within the dataset. As mentioned in the previous paragraph no formulars where being applied when data was entered in the system therefore data contained different characters which were mostly resulting from typing errors. The solution used to solve this problem was using the **find and replace** method in Microsoft Excel. The method helps in finding any characters within the data that may have been a result of typing errors and give the option to replace them with the correct characters. This method was adopted to also ensure python will be able to read in the file without any hiccups. The first image below displays the raw dataset with typing errors and the second image displays the dataset after characters were removed.

Name	Column3	Airline	ZWL Cost	Total Cost
T.Dzimbanhete	Hre-Jnb	SAA		$ 490.00
L.Dakota (family)	Hre,Lun	KQ		$ 667.00
Makomborero	Lhr-Hre	E.K/Satguru		$ 1,187.00
Mr Goto (team)	Hre-EBB	E.T		$ 1,046.00
U.Dube	Ticket Change	SAA		$ 40.00
E.Chanakira(TTCS)	Hre-Jnb	SAA-Satguru		$ 942.00
D.Chanakira (TTCS)	Hre-Jnb	SAA/Satguru		$ 942.00
E.Chanakira (TTCS)	Jnb-Hre (change)	SAA/Satguru		$ 10.00
P.Chiobvu(TTCS)	Hre-Mpm	SAA		$ 430.00
T.Chisambiro	Hre=Los	KQ		$ 471.00
E,Chanakira (TTCS)	Hre-Jnb	SAA/Satguru		$ 942.00
Mrs Dube	Hre=Moscow	E.K/Satguru		$ 8,438.00
I.Chilapula	Hre-Jnb	B.A		$ 173.00
I.Chilapula	Jnb-Hre	S/AA		$ 196.00
Mutize(family)	Hre-Jnb	SAA		$ 854.00
K.Chitembo	Buq=Hre	Fastjet		$ 283.00
Goto (ticket Change)	EBB-Hre	KQ		$ 150.00

harare to johanessburg	saasatguru	1050
harare to johanessburg	saasatguru	995
johanessburg to harare (change)	saasatguru	0
harare to maputo	saa	450
harare to lagos	kq	500
harare to johanessburg	saasatguru	1050
harare to moscow	e.ksatguru	8845
harare to johanessburg	b.a	200
johanessburg to harare	saa	235
harare to johanessburg	saa	965
bulawayo to harare	fastjet	293

It is essential to note that, the excel file provided had columns that contained inconsistent values this is mainly because the company was not following a particular formular to enter their data. In this case Python would read the same person as two different people because names were entered in different formats which is in caps lock and small caps. Analysing this type of dataset will produce highly biased results as the system is assuming a single individual to be

two individuals. To solve this problem there was need to apply the formular (=LOWER ()) and pass it to all the values in the specified column to make all the values uniform. The first image below shows the dataset before making it uniform and the second image shows how the dataset appeared after applying and passing the aforementioned formular.

Paid	Receipt	Inv	Column1	Client Type	Account Name
Y			Noreen	Individual	
Y			Noreen	Individual	6EV9LG
Y			nothando D	Individual	5PGPCK
Y			Noreen	Individual	
Y			fari	Individual	
Y			Fari	Individual	
Y			fari	Corporate	SG8MFU
Y			Nothando s	Individual	620LP9
Y			Fari	Individual	S5HA3E
Y			Fari	Corporate	TG3S03
Y			Fari	Individual	
y			Fari	Individual	
Y			Fari	Individual	
Y			Fari	Corporate	RTIWKE
y			praiSe	Individual	7KEUV8
Y			Praise	Individual	7KEUV8
Y			Fari	Individual	SW89IP

360	cash	Yes	noreen	individual
360	cash	Yes	noreen	individual
144	cash	Yes	vonai	individual
230	cash	Yes	vonai	individual
380	cash	Yes	fari	individual
408	cash	Yes	fari	corporate
408	cash	Yes	fari	corporate
408	cash	Yes	fari	corporate
408	cash	Yes	fari	corporate
408	cash	Yes	fari	corporate
1233	cash	Yes	noreen	individual
235	cash	Yes	macmilan	individual
1010	cash	Yes	praise	individual
201	cabstransfer	Yes	nothando	individual
27	cash	Yes	noreen	individual
120	cash	Yes	fari	individual
240	cash	Yes	fari	individual

Also, the company executive managers communicated that they were not able to read some of the shorthand that was within their dataset, therefore they were finding it difficult to understand data gathered. The ‘Route’ column consisted of shorthand of destinations visited by clients therefore, one of their requests was to solve this problem to allow the company to have a clear picture of the routes travelled during January 2019 and May 2022. To rectify this issue there was need to consult **skysacnner.com** and check all the codes to reveal the names of all the shorthand in the ‘Route’ variable, after finding the names I used the find and replace function in Excel to replace these codes with full names of route. The first image below displays the column with shorthand and the second image show the column with the full route names.

Route
Hre-Jnb
Hre - DLM
Hre - DLM
JNB- HRE
Hre-Nce
Hre-Nce
Hre-jnb
Hre-jnb
Hre-Vfa
Hre-jnb
Hre-Jnb
Jnb-Cpt
Jnb-Cpt
Hre-Dxb
Hre-Nce
Hre-Nce
Hre-Dar
Dar-Hre

harare to johanessburg
harare to dalaman
harare to dalaman
johanessburg to harare
harare to france
harare to france
harare to johanessburg
harare to johanessburg
harare to victoria falls
harare to johanessburg
harare to johanessburg
johanessburg to capetown
johanessburg to capetown
harare to dublin
harare to france
harare to france
harare to dar salam
dar salam to harare

3.3.2. Dimensionality Reduction using Microsoft Excel

Firstly, there was need to erase some variables in the dataset that were not contributing to accurate results. Columns such as ‘Account Name’, ‘Invoice Number’, ‘Receipt Number’, ‘ZWL Cost’. These columns were removed using Excel for a couple of reasons which include, not related to the domain problem being solved as communicated by the Executive Managers and also the variables had noticeably missing data. The image below justifies the latter as it indicates that more than 3 quarters of data is missing from the ‘Receipt Number’ variable.

Count of Receipt
235

To add on, further dimensionality reduction was performed in Microsoft Excel for columns such as ‘ZWL Cost’. The reason behind deleting this column was mainly because the dataset provided contained two columns that showed amount received in two currencies which are **USD** and **ZWL**. Zimbabwe currently has recorded an inflation rate of 96.4% according to Inflationdata.com therefore, analyzing sale performance using ZWL is likely to mislead stakeholders as the figures are prone to change any time without warning. This also worked to the advantage of this project as reducing data dimensions is one of the essential techniques that need to be performed to ensure all redundant data that is likely to mislead the outcome is removed.

Furthermore, is vital to note that when travelling, a client is expected to provide their clear and accurate identification documents which carry their personal information such as date of birth and full names. As a travel agency National Tickets Zimbabwe is expected to record client’s personal information to ensure all details entered on the ticket are correct to avoid any inconvenience while travelling. However, due to the General Data Protection Act there was need to delete the ‘Customer Name’ column from the dataset as it is against the law to use such data without consent from the individuals involved.

3.4 Python

The coding language which was used for this project is Python Language version 3.10 using Pycharm as the Integrated Development Environment. Python has inbuilt libraries such as Pandas and NumPy that assist in understanding, pre-processing, and transforming data. The

logic behind using python for this project is it consist of numerous open-source Python libraries that allow data manipulation, visualisation, mathematics, machine learning to mention just a few. The table below show how different Python libraries were applied to the development of this project

Fig 2. Python libraries utilised

Library	**Application**
Panda	To understand the dataset and to check for the count of any missing values within the dataset. The library was also used to clean, analyse, and manipulate data
Streamlit	To create interactive, website application with visuals
Seaborn	To perform statistical plotting that can visualize trends in the data
Matplotlib	To visualize trends in the sales data

3.4.1 Data Understanding and Manipulation using Pandas

Pandas is a built-in library in Python that contains prolific built-in functions to allow sound data understanding, representation, and manipulation. The lucidity to use pandas for this dataset emanated from the fact that, Pandas aids in sorting data according to the conditions the programmer sets and it also groups and divides data according to set preferences. Pandas provides a group of instructions and key functions that help in reviewing data in a more understandable way. Lastly, pandas supports in-depth analysis, and decent ways of handling of statistics.

To understand the dataset, I imported data into Python and read the csv using Pandas library. Firstly, checking the data head and tail were essential processes to ensure the dataset imported is the correct. To further understand data I used (d.shape) to check data dimensions,(d.types) to check data type and the describe function to get a description of the data. The description

function provides descriptive statistics of the dataset that is very essential when understanding and manipulating data.

It is of vital importance to note that when seeking to understand data there is need to check the unique values in all the columns to help Data Analysts understand the column structure and values. Checking the unique values in this dataset was very essential to confirm whether there are any duplicate in columns such as ‘Client Type’ and ‘Paid or Not’. The previously mentioned columns are binary variables which means they only have two values therefore using the unique function quickly exposes if there are any duplicate or typing errors in these columns. The image below clearly displays that the ‘Paid/Not’ variable has an error which needs to be fixed as it shows three unique values instead of two values. It is also fundamental to note that checking unique values also helps in having a basic analysis of the dataset. This can be witnessed by the image below; the dataset has 1,629 rows in total and ‘Date’ column is displaying 590 unique values. A swift conclusion that, during January 2019 and May 2022 an average of 2 tickets were sold per day however, to validate this, there is need for further deeper analysis.

```
These are the unique values in your data Date              590
Customer Name     1115
Route              229
Flight             149
Total Cost         632
Sale               405
Payment Method      37
Paid/Not             3
Receipt Number     177
```

Also, there was need to manipulate the ‘Date’ column from object to datetime because the analysis to be performed includes time series therefore datetime help in dealing with dates effectively. The datetime module provides classes that help in manipulating date and time to help in effective data analysis. The use of this module is crucial to ensure the analysis carried is distinct and in line with the objectives of this project.

```
#Changing Date to datetime
my_data['Date'] = pd.to_datetime(my_data['Date'])
```

Lastly, as part of data understanding there was need to check for the sum of null values in all the columns to ensure the dataset to be analysed contains adequate data fit for analysis. This function helps with information of how many values are missing in each column and allows us to make decisions on how best to deal with the missing data without disrupting the objective of the project and ensuring the method selected to deal with missing data does not deviate from the domain problem. The below image displays the number of missing values in every column of the dataset. It is crystal clear that all the columns contain null values that are less than 1% of the dataset. The column that has the most missing values is 'Paid/Not', however the missing value of this column is still less than 1% this was calculated by 0.01 multiplied by the total rows 1,629 which gives 16.29. In this scenario a decision to perform dimensionality reduction by deleting null values was made this is merely because, the missing values were less than 1% and erasing them would not affect the accuracy of the data analysis to be performed.

```
Date              3
Route            12
Flight            3
Sale             14
Payment Method   14
Paid/Not         15
Sales Person     13
Client Type      14
```

3.4.2 Univariate Analysis using Python

It is of vital importance to note that this project aims and objectives were driven from what National Tickets executives desired to understand from the data they provided. Throughout

communication with client, they communicated that, they require answers to the following questions:

1. **The percentage of corporate clients who purchased tickets during January 2019 and May 2022?**
2. **The range of sales during January 2019 and May 2022 to measure business performance?**
3. **The percentage of clients who are currently owing the company for services rendered during January 2019 and May 2022?**

To answer the question one, there was need to perform univariate analysis on the 'Client Type' variable. The suitable solution for this question was plotting a pie chart of the 'Client Type' column using matplotlib an inbuilt visualisation function in Python. Plotting a pie chart was deemed appropriate as it would help to vividly show the type of clients who purchased more tickets during the given period as well as display the percentage of corporate clients who purchased tickets during January 2019 and May 2022.

Furthermore, to give a solution to question two, there was need to analyse the 'Sales' variable. This was univariate analysis as it was analysing just one variable. This analysis was performed by plotting a violin diagram using seaborn a built-in visualisation library in Python. The violin plot was utilised in a way that it would clearly show the range of sales during January 2019 and May 2022. The use of violin plot was necessary as it aids in observing the distribution of sales.

Lastly to answer the question number three, there was need analyse the 'Paid/Not variable'. This variable is binary as it only contains two outcomes which are clients who paid and clients who have not paid. To get the answer for this question there as need to plot a pie chart that patently displays the percentage of flight tickets not paid. The reason behind performing this analysis, is for the company to understand the fraction of clients who still owe the company to date.

3.4.3Bivariate Analysis using Python

Bivariate is essential when performing exploratory data analysis. Bivariate analysis involves the process of analysing only two variables to draw insights and relevant answers that suit the domain questions. In this project there was fundamental need to perform bivariate to ensure the

clients questions are answered and to give the client a clear picture of the hidden insights in the dataset. The following questions were posed by the client, and they steered the use of bivariate analysis in this project.

1. **How much was made from corporate clients in comparison to individual clients?**
2. **How much was made by the highest and lowest Salesperson?**
3. **The number of times the same flight ticket was purchased by clients**

To answer the first question there was need to perform bivariate analysis on two variables which are the 'Sale' and 'Client Type' variable. The analysis strived to see the number of sales made from the two types of clients. To perform this analysis, I used pandas groupby function and plotted a relplot using seaborn. Visualising the answer was performed to give distinct response to the client to ensure they understand how much of their income came from corporate clients during the start of 2019 to mid-2022.

To add on, there was need to analyse the 'Sale' and 'Salesperson' variables to draw insights on how much was made by the highest salesperson. According to the client, they desire to know the number of sales made by the highest and lowest salesperson and this is merely because it will help them establish the difference between their current employees. If the difference is significant, they will decide to relieve the lowest salesperson off their duties, which helps the company in minimizing cost as no income is being generated by the salesperson. Also, another way would be to offer thorough marketing training to the lowest salesperson to help them achieve more sales.

To analyse the sales data furthermore, there was need to perform bivariate analysis which included analysing the 'Client' and 'Flight' variables to identify which client type of clients purchased certain tickets and the number of tickets purchased by the client. To achieve the former, I used the groupby function to group clients by the flight tickets purchased and further the value counts function to determine how many of the tickets were bought from the same flight operator. It is of vital importance to note that this allows National Tickets Zimbabwe to understand customer behaviour and preference. The analysis is crucial in exposing the type of flights preferred by a single group of clients. In this case, National Tickets will use this information when implementing their marketing strategies, it allows them to determine which flights they can advertise to individual or corporate clients based on information generated

from this analysis. This gives National Tickets the opportunity to identify their targeted audience for certain flights.

In addition, bivariate analysis was essential in determining the total amount of sales each flight operator made. To achieve the latter, I analysed the ‘Sale’ and ‘Flight’ variable using the groupby function and using the function .sum() to get the total amount made. Performing this type of analysis helps National Tickets identify the least contributing flight operator and the highest contributing. In this scenario the company can make decisions such as cancelling contracts with the least performing operator, or they can re-strategize their marketing techniques for that flight and invest on marketing ideas to ensure flight tickets are also purchased from the least performing operator.

3.5 Streamlit

Streamlit is a platform that is used for visual analytics, and it was utilized in this project to draw and display clear insights for National Tickets Zimbabwe. The use of Streamlit was very essential in this study because it is compatible with big datasets and allows the creation of interactive dashboards much faster as compared to Microsoft Excel and Google Sheets. In most cases, to join several data sources one needs to thoroughly know programming language, however Streamlit syntax language is not too complicated to understand therefore, it caters for both intermediate level and professional programmers. The syntax in the image below justifies the former.

The rationale behind using Streamlit is because it allows the creation of interactive dashboards that are very essential for this study. Dashboards are essential as they tend to massively expose trends and enhance productivity within the company. This is merely because they induce convenience by visually presenting large amounts of business data on a single dashboard simultaneously which is imperative for valuable business decisions and growth.

3.5.1 Website Application Development using Streamlit

To ensure the results generated from this project are well understood by the client and distinct there was need to create a website application that would allow the users to generate desired information. It is fundamental to note that visualization was included, and interactive dashboards created to ensure easy understanding of the analysis. Data visualisation is crucial in aiding National Tickets to recognize trends in their sales data. It is vital to note that, the survival of any business is depended on data understanding, this is merely because understanding the trends in datasets allows companies to plan accordingly and place protective measures if need be.

To build the website application I used python libraries which are:

Fig 3. Python libraries utilised for Streamlit

Library	**Application**
Pandas	To analyse and manipulate data using pandas built in functions
Streamlit	Use of functions such as • **st.write** to print codes • **st.session _state** for storing data without vulnerability to interruptions • **st.radio** to create the structure of element in the application • **st.button** to create buttons for users in the application • **st.markdown** for visual separation to section the application

Additionally, I utilised PIP in install-to-install libraries such as Requests and Streamlit – Lottie. To incorporate animation in the website application to make it more interactive there was need to visit the Lottie files website and search for suitable animations that fit with the theme of the application. After identifying the desired animations there was need to copy the link and paste

them in the code. The Requests library was used to fetch the requested address of the website page that was being used in the code and Lottie allows to display the link pasted in our code as animation. To execute this task, I created a function that would fetch the JSON file and if the request encounters an error the function returns nothing.

```
def load_lottieurl(url):
    r= requests.get(url)
    if r.status_code != 200:
        return None
    return r.json()
```

To create sections in the website application I created containers for each section and called each container as I created a section using the **with** function. Session State is a Streamlit inbuilt function that was utilised to share variables amid reruns per session and is capable of storing data without interruption. A good example is when you implement session state in your application it means various people from different parts of the world can use the application at the same time without any disruptions. It is vital to note that for session state to function the keys need to be initialised by checking if the variable is not in session state and assigning a value to the variable.

```
if 'number_of_rows' not in st.session_state or 'type' not in st.session_state:
    st.session_state['number_of_rows'] = 5
    st.session_state['type'] = 'Categorical'
```

I used the if loop for conditional decision making in the code to enable the show more and show fewer buttons execute their functions. The if loop is a conditional argument that allows the show more button to add a row on the table when clicked and the show fewer to subtract one row from the table. To add on, I created a sales dashboard that displays the total sales, topflight and top route. The former was executed by assigning variables to the codes **value_counts** and .**sum()** .

```
#assigning variables
    total_sales = int(sale_data['Sale'].sum())
    flight = str(sale_data['Flight'].value_counts().head(1))
    Route = str(sale_data['Route'].value_counts().head(1))

#assigning columns and calling variables created earlier
    left_column,middle_column,right_column = st.columns(3)
    with left_column:
        st.subheader('Total Sales')
        st.subheader(f'US $ {total_sales:,}')
    with middle_column:
        st.subheader('Top Flight')
        st.subheader(f' {flight}')
    with right_column:
        st.subheader('Top Route')
        st.subheader(f' {Route}')

    st.markdown('---')
```

Also, there was need to create a dashboard that incorporates all the variables and plot performance of at least 30 values of all the variables on a bar graph. To ensure the smooth execution of this I created a radio button widget which gives the user, the option to select which variable they want to view and to change from categorical to Numerical. The former was done by assigning values to the types variable and creating a handle click function that allows the change button to function when clicked. To finish of this task I created an if loop that I passed the argument if the type selected by the user is categorical the application should fetch and display value count head (30) in each value in the types variable as a bar chart. The same was done for the Numerical, else the user selects Numerical the application fetches and displays the description of the values in Numerical.

```
#Creating the selectbox to allow the user to select a column
    column = st.selectbox('Select a column', types[st.session_state['type']])

#creating a function to use for the change button
    def handle_click(new_type):
        st.session_state.type = new_type


#creating the two options (categorical and numerical) and assigning the change function
    type_of_column = st.radio('What kind of analysis',['Categorical','Numerical'])
    change = st.button('Change',on_click=handle_click, args=[type_of_column])

#creating a loop to iterate through the columns
    if st.session_state['type'] == 'Categorical':
        dist = pd.DataFrame(sale_data[column].value_counts()).head(30)
        st.bar_chart(dist)

    else:
        st.table(sale_data[column].describe())
```

Finally, there was need to create a side bar that the user can opt to dismiss or keep when the application is running. The side bar, was created using streamlit inbuilt function **with.sidebar**. I made use of the lottie files animation to display an email animation and created an email form below, that allows the user to enter their email, message and click send. The email form was created using the submit form website where a link to submit a form is copied and I pasted it in the code. To ensure I received the email from user I had to edit the link and enter my email address. To add on, there was need to customise the email form, which I did by creating a

style.css file that contains all the code for configuration including colours for the send button and the size of the text area, to further customize the latter I hid Streamlit branding from displaying on the application.

To add on the side bar I created a rate bar that allows the user to rate the application, the rate bar displays rating submitted by the user. To execute this I utilised the **st.select_slider** function and used the star emoji code I pulled from **webfx website,** this was to display a star emoji along with the message 'Your rating is'. To finish of

f the side bar, I displayed my name as the developer of the application and pulled an animation from **lottie files website** to display below. To configure the application and put the desired colours to make it more appealing I created a toml file and configured the application theme inside the file.

```
with st.sidebar:
    st.title('Rate this Application')
    options = ['Poor','Not Bad','Good','Very Good','Exceptional']
    ratings = st.select_slider('Choose your Rating',options=options)
    st.write("Your rating :star: is:", ratings)
```

3.6 Canva

To circumvent crowding my website with visualisations I decided to design an infographic video with Canva, a graphic design platform. The video contains all the essential information extracted from the exploratory data analysis. To incorporate this video on my website application,I used Streamlit inbuilt function called **st.video**. This helped improve the graphic user interface to be well-organised and distinct.

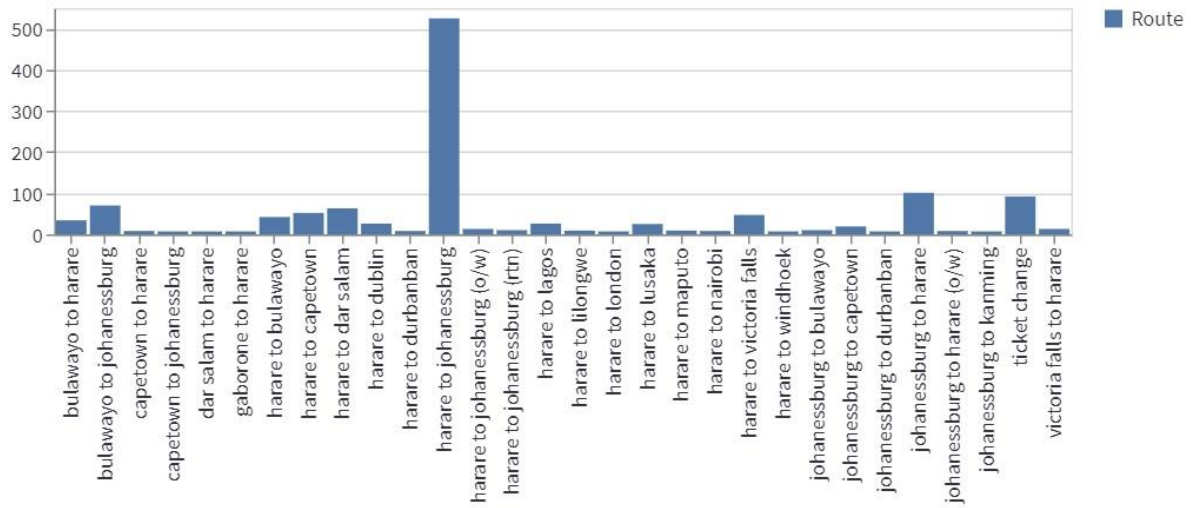
Chapter 4: Results and Findings

Overview

Exploratory Analysis is very essential for business growth as it unravels threats, opportunities, customer, and product behaviour. This type of analysis also helps in digging out insights that are crucial for future planning to ensure business sustainability. Early warning systems are one the most important features derived from Exploratory Data Analysis to help businesses identify future threats and opportunities and prepare accordingly. It is of vital importance to note that both Exploratory Data Analysis and Prescriptive Analysis were performed on National Tickets data in accordance to clients request and the following results were observed.

4.1 Route Findings

As part of Exploratory Data Analysis there was need to find out the most travelled routes during the period 2019 and 2022. Finding out the most travelled routes was very essential as it helps National Tickets decide on which country to expand their business. National Tickets executives are striving to grow their business and it is vital to point out areas that have business potential. The image below displays Harare to Johannesburg (South Africa) as the most travelled routes with a count slightly above 500 tickets purchased. In this case, it is evident that there is business potential in Johannesburg, South Africa and National Tickets can stimulate their idea to manoeuvre into accommodation business using Johannesburg, South Africa as their starting point.To add on, it is of vital importance to note that inside the ‘Route’ variable, National Tickets also included tickets changed. The below image shows that a count of about 90 ticket changes were recorded. Ticket changes are triggered by a bunch of reason, one of them which is National tickets staff mistakenly booking the wrong route for a client. In this case it fundamental for National Tickets to ensure thorough training is provided for staff. Ticket changes caused by staff negligence are very critical to business performance as they can frustrate and drive out customers hence why this should be rectified with urgency.



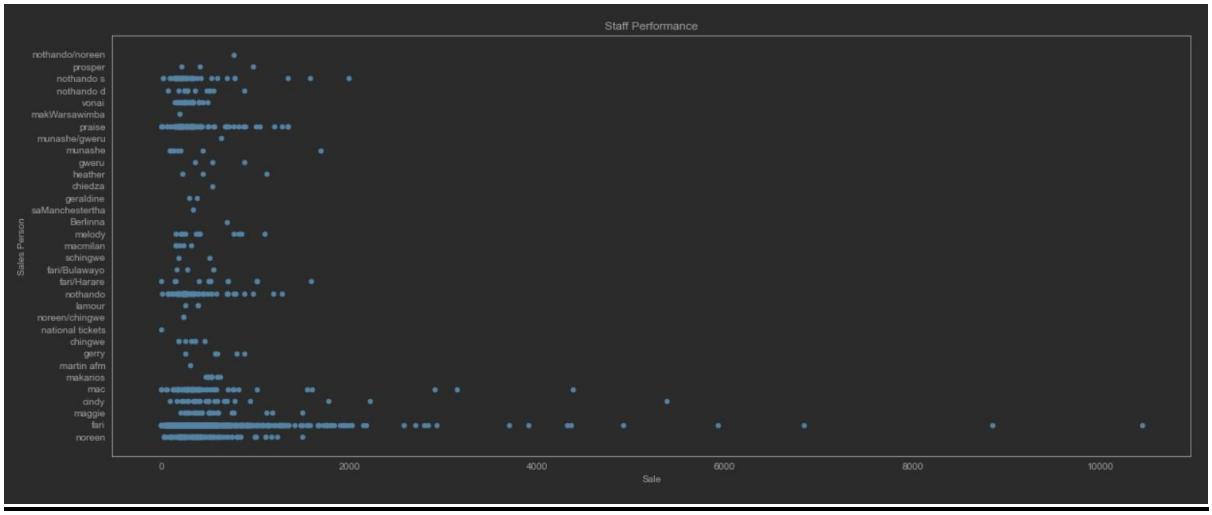
To further validate the latter, there was need to find out the exact number of clients who travelled from Harare to Johannesburg. It is quite evident that the most destination that clients travel to is Harare to Johanessburg. This is further backed up by the image below which displays 525 tickets were purchased by clients travelling to Johanessburg followed by 101 tickets purchased by clients travelling from Joahnessburg to Harare. In this regard, there is clear evidence that there are more business oppotunities in Johanessburg and therefore the company should consider diverting marketing efforts on this destination to ensure marketing strategies taior-made for this route are originated for the purposes of maximizing revenue and profits.

```
harare to johanessburg      525
johanessburg to harare      101
ticket change                92
bulawayo to johanessburg     70
harare to dar salam          63
Name: Route, dtype: int64
visa itenerary                    1
johanessburg to dublin            1
penalty                           1
```

4.2 Staff Performance Findings

To satisfy the request of finding out the highest and the least salesperson, there was need to plot a scatterplot for the ‘Salesperson’ and ‘Sale’ variables. The reason behind using a scatter plot is because it vividly displays how National Tickets staff have been performing from January 2019 to May 2022. Taking a closer look on the image below, Fari is the highest salesperson with some of his sales slightly above $10,000 whereas National Tickets only had one sale which is $0. The scatterplot further displays that some of the sales team had only one sale during 2019

and 2022 these people include Martin atm, Makwarimba and Nothando/Noreen. The reason behind digging these insights is to expose and assess staff performance and to induce decision-making such as minimizing company cost, by dismissing staff members that not performing well as no income is being generated from their work. The company can also decide on offering intensive training to the least performing salesperson to stimulate productivity with the company.



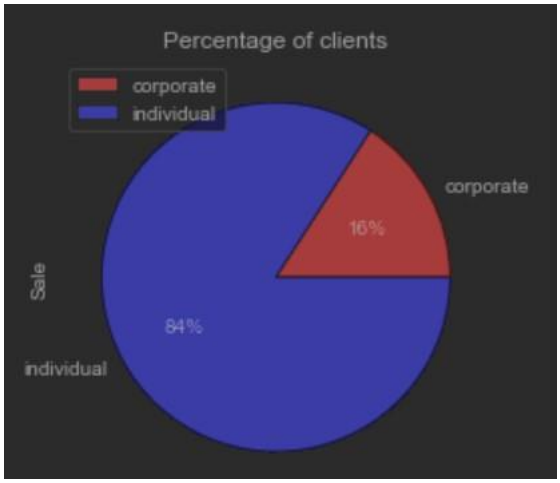
To further validate the information displayed by the scatter plot above I used the 'Salesperson' and 'Sale' variable to analyse how much was generated by each salesperson. The image below distinctly shows that Fari is the highest salesperson with $489,296 followed by Cindy who sold tickets worth $18,616. The image below validates the information presented by the scatterplot which displayed Fari as the highest followed by Cindy.

Sales Person	
Berlinna	695.00
chiedza	545.00
chingwe	2126.00
cindy	18616.00
fari	489296.24

4.3 Customer Findings

During the period January 2019 and May 2022, National Tickets had two types of clients which are Individual Clients and Corporate Clients. Individual clients refer to customer who purchase flight tickets for their personal travelling whereas corporate clients refer to customers who

purchase tickets for business travel. To find out the percentage of both clients I created a pie chart. The image below shows that 84% of clients who purchased tickets during January 2019 to May 2021 where individual whereas 16% where corporate clients. A pie chart was reckoned as the superlative way to visualize this data as it makes use of colour codes to ensure the information is presented distinctively.



To cross check the information presented by the pie chart, I plotted a boxplot using the ‘Sale’ and ‘Client Type’ variables. The image below shows the results derived from the box plot, it is quite distinct that both the pie chart and box plot display the same information which is, most income made during 2019 and 2022 was generated from individual clients.

To further understand clients, it was essential scrutinize their preferred flights. This was done to help National Tickets understand customer behaviour and originate their marketing strategies based on the information revealed from this analysis. I used the ‘Client Type’ and ‘Flight’ variables to count how many flight tickets were bought by a certain group of clients.

The first image below shows that during January 2019 and May 2022, corporate clients bought 27 Fastjet tickets, 25 SA Airways tickets and 21 SA Satguru. The second image shows that individual clients purchased 448 Fastjet tickets which validates the previous analysis presented

by the pie chart, that showed individual clients purchased more tickets compared to corporate clients. The second image below goes further to show that individual clients also purchased 87 AirlinkBerlinry tickets and 62 Berlinry tickets.

Client Type	Flight	Flight
corporate	fastjet	27
	saa	25
	saasatguru	21
	f.nsatguru	13
	emiratessatguru	10

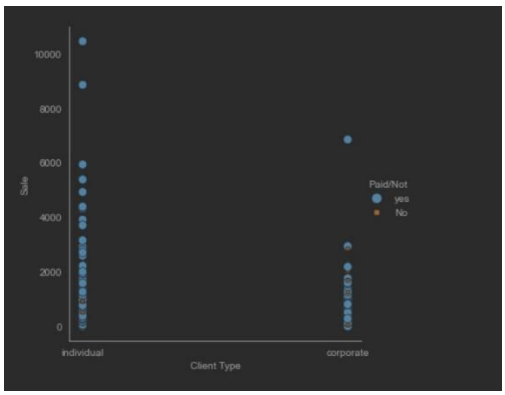
Client Type	Flight	Flight
individual	fastjet	448
	airlinkBerlinry	87
	Berlinry	62
	saasatguru	59
	b.a	57
	saa	53

To further analyse the above, there is need for National Tickets to invest more on marketing material tailored-made for corporate clients. It is evident that there is potential for more sales from corporate clients but this can be achieved if the company maximize their marketing efforts and create marketing strategies that target corporate clients. However, on the other hand this analysis should factor in the fact that the sales data was gathered during the peak of covid-19 when most business where not operating well. The former can be a contributory factor to the poor performance of corporate clients.

It is patent that the top three flights prefered by corporate clients is Fastject, SA Airways and SA Satguru therefore the company should strive to centre most of their marketing material around these flights and the same should also be done for flights preferred by individual clients.

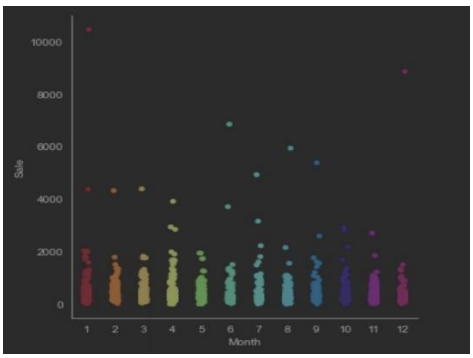
The company can use the information displayed above for individual clients to restrategize their marketing strategies for the least performing operators. It shows individual clients purchased 448 Fastject tickets, the company should apply the same marketing strategy being used for Fastjet to market other flights, this will help the company score more sales.

It was also essential to plot the most clients who owe the company. This would give the company a clear pitcure of the type of clients who prefer credit. The image below clearly shows that the most clients who have not paid are corporate clients. This information gives the company the platform to instigate decisions on how to deal with clients who prefer credit over paying cash. It is essential for the company to note that offering services and products on credit is a fragile decision that requires well-thought out strategies on how to tackle the entire process, else there is risk of minimizing revenue and profits and in worst case scenarios business demise.



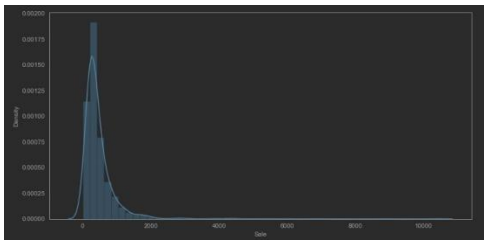
4.4 Sales Findings

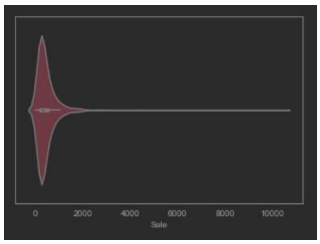
Understanding the periods that generated more income was one of the pre-requisites of this project to ensure set objectives are satisfied. The logic behind analysing the best performing months is to allow managers to make decisions on which months require extreme marketing efforts to maximise sales. The first image below shows July is the month that generates the most income followed by January and April. This can be justified by the fact that Easter holidays are in April and therefore more people tend to travel. In this regard National Tickets can look into designing marketing material related to Easter holidays and consider Easter promotional strategies to secure customers. The second image below is a catplot that was plotted using seaborn to validate the information displayed in the first image.



To add on, it was essential to find the range of sales during January 2019 and May 2022. The former allows National Tickets to identify the value of the products they are offering. To find out the latter, there was need to plot a density plot and violin plot for further validation. The first image below is a density plot, and the second image is a violin plot. It is evident that most

sales range between $0 and $2000 which means out of 1,629 tickets sold about three quarters of these tickets had a value of $2000 or less paid.

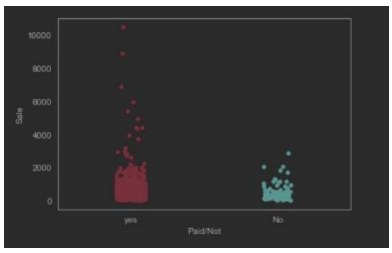




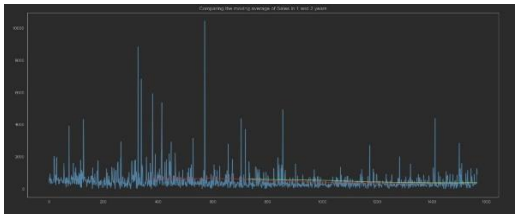
The project further analysed the top 5 amount paid by clients. The logic behind analysing the former is to have an idea of how much clients are able to pay in a single purchase. The information permits the company to create tailor-made products factoring in the maximum amount clients are capable of paying. The image below shows the maximum value that was paid for a tickets during January 2019 and mid-2022 is $10,440 followed by $8,845. This further validates the aforementioned analysis that identified the most paying customers as individual clients with one of the them paying $10,440.

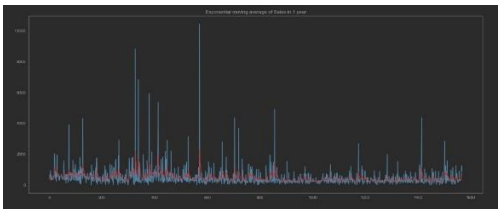
	Sale
570	10440.0
326	8845.0
338	6840.0
379	5926.0
413	5375.0

The image below shows the distribution of how much has been paid and how much is being owed. The plot seeks to display information on how much income has been received by the company versus how much the company is owed. The logic behind this plot is to give a distinct picture to the company of the status of their sales. It is evident that most of the money owed are tickets that range from about $100 and $1,000, a conclusion can be made that most clients are comfortable receiving products that are not above $2,000 on credit.



It is essential to determine the moving average of sales to assist National Tickets executives understand the trend direction of sales. The Simple Moving Average of sales is calculated by summing all the sales made during the period January 2019 and May 2022 and dividing the total by the number of time periods. The advantage of performing the latter is it smooths out short term fluctuations and helps company executives understand forecast trend and monitor fluctuation. The first image was comparing the moving average in one year and the moving average in 2 years using the Simple Moving Average (SMA). To provide wholesome information I also plotted Exponential Moving Average on the second image, which is different from Simple Moving Average in the sense that it is more sensitive to prices and places weight to recent prices. Exponential and Simple moving average simply show the trends in prices to help understand the movement of prices. The overall go for potting both moving averages is to establish the direction of sales based on historical data provided.

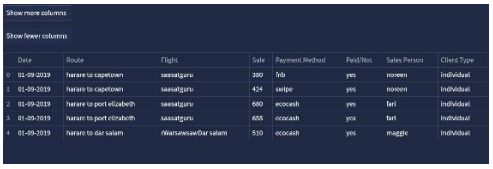
. 



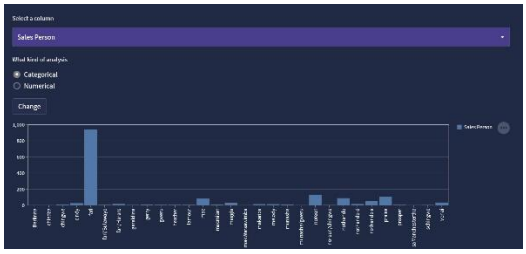
4.5 Use Interface Results

The image below is the welcome note of the application that is displayed together with animation from lottie files. The animation was added to give a rough idea of what the application is about, to intensify the data analysis part of the project.

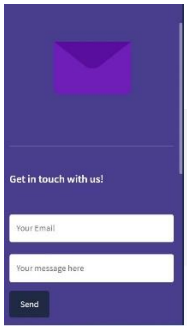
The first image below, shows how the dataset is presented in a table with options to show more or fewer columns. The second image shows brief insights of the data on total Sales, topflight and top Route.

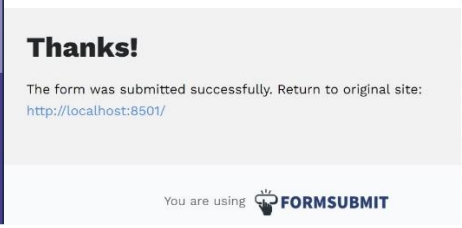


The image below shows how the user is given options to choose which variable they want displayed and the insights are presented in a bar graph.

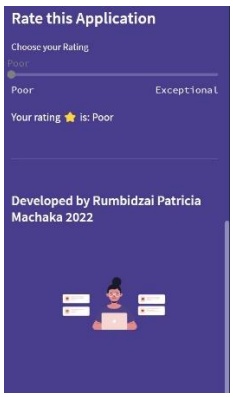


The first image below shows the email form created on the sidebar, after clicking the send message button the user receives the message in the second image below. The message sent by the user is then sent to the email that I incorporated in the code as shown by the third image.





The last part of the side bar displays a rating bar that allows the user to use based on how they find the application. Lastly the developers name is displayed, and animation is added below to give the statement weight.



The image below shows the video created using Canva. The video was incorporated into the website application and when user presses the play icon the video starts displaying insights.

To run the application type **streamlit run app.py** in the terminal.

Chapter 5: Discussion and Analysis

Chapter Overview

This chapter gives an overall assessment of the project, it incorporates the critical analysis of how the project was executed and the outcome of the project. The chapter further points out the limitations of the project carried out and how it can be rectified in the future. Lastly the chapter stretches a sum up of the project, distinctly incorporating the aim and objectives of the project and the extent they were achieved.

5.1 Discussion

This project has been successful in unravelling meaningful insights from the data provide by National Tickets. Various trends have been identified from the sales data and brought to the managers attention for urgent decision making to ensure high productivity within the company. It is of vital importance to note that the findings of this study indicate the positive effects of analysing big data for effective decision-making, business growth and maximising revenue and profits which is the response variable of this project.

The latter findings are similar to the findings revealed by Ahmed et.al(2022) on the study conducted on impact of data analysis on project success. The aforementioned scholar discovered the positive impact of big data analytics on organisations performance and the findings further supports the idea of utilizing big data analytics for business decision by revealing the indirect influence of decision making on organisational business success.

5.2 Limitations

The sales data collected by the company does not include the age of the clients that purchased tickets this is a drawback as it may hinder in analysing the age group of National Tickets clients. It is important to identify the age group of your clients for example, knowing if your clients are students and extending discounts that are tailor-made for students to attain the company's primary objective which is business growth. Therefore, without the age of the clients it is impossible to perform the latter.

It is vital to note that the project should have considered forecasting sales to allow National Tickets executives to understand what their sales are likely to be based on the historical data provided. However, this task was overlooked, this is because the sales data that was been utilised ranged from January 2019 to May 2022 which was the peak of the covid-19 pandemic. In this regard, sales data was gathered during a period when the tourism industry was not functioning properly, given this scenario forecasting sales using this data would have generated biased results as the company was not operating at its max. Sales forecasting is essential as it allows companies to plan for the future however in this case it would have misled the company.

5.3 Further Research Direction

It is crucial to incorporate more industries in the big data analytics research to give a wholesome and distinct understanding of the effect of big data analytics on business success in all industries. There is a chance some companies do not drive their success from big data and therefore research for such organisations ought to be performed to find out why and how. To add on, this research was centred on the impact of big data analytics on tourism business success similar to Ahmed et.al(2022) who researched on the impact of big data analytics on project success in telecommunications. Both studies are limited to two industries, there is essential need to research on the impact of big data analytics on other industries such as Healthcare, Agriculture, Law, Transportation to mention a few.

5.4 Conclusion

The project strived to draw meaningful insights from National Tickets Zimbabwe sales data to evoke constructive business decisions for immense business growth and maximising productivity, revenue, and profits. The main objective of this study was fulfilled as various insights that displayed trends within the data were unravelled and this stimulated urgent decision-making to ensure the smooth running of the company and to shun business demise. The utilization of various data cleaning, understanding, manipulation and visualisation were implemented to achieve the aims and objectives of the study as rationalised by the managers. This analysis will aid in uncovering innovative business possibilities that would have been overlooked otherwise, by so doing, growth and profits are attained, and business is strengthened. Various studies have investigated on the impact of data analysis on business

growth; however, most scholars did not go further into building website application with interactive dashboards after analysing the data, therefore this project strives to conceal that gap.

Thank you

References

Ahmed, E and Yüzbas, D., 2016.*Big data analytics: integrating penalty strategies*. Int. J. Manag. Sci. Eng. Manag. 11 (2), 105–115.

Ahmed, R., Hussain, A and Philbin, P., 2021. *Moderating effect of senior management support on the relationship between schedule delay factors and project performance*. Eng. Manag. J. 33 (3), 1–20

Akter, S. et al, 2019. *Analytics-based decision-making for service systems: A qualitative study and agenda for future research*. Int. J. Inf. Manag. 48, 85–95.

Al Nuaimi,K., Khan, M and Ajmal, M.,2021. *The role of big data analytics capabilities in greening e-procurement: A higher order PLS-SEM analysis*. Technol. Forecast. Soc. Change 16.

Ahmed,R. , Shaheen ,S. and Philbin,S.,2022. *The role of big data analytics and decision-making in achieving project success*. Journal of Engineering and Technology Management .School of Engineering, London South Bank University, UK

Alsghaier, et al.,2017. *The Impact of Big Data Analytics on Business Competitiveness. Proceedings of the New Trends in Information Technology* (NTIT-2017) The University of Jordan, Amman, Jordan.

Bag, S. et al, 2020. *Big data analytics as an operational excellence approach to enhance sustainable supply chain performance*. Resour. Conserv. Recycl. 153, 104559. Barney, J.B., 2001.

Bauer, D. and P.Curran., 2005. *Probing interactions in fixed and multilevel regression: Inferential and graphical techniques*. Multivar. Behav. Res. 40, 373–400.

Botterhuis, L.et al., 2010. *Monitoring the future. Building an early warning system for the Dutch Ministry of Justice*. Futures 42, 454–465.

Brands, K., 2014. *Big data and business intelligence for management accountants*. Strateg. Finance. 95, 64–65.

Carter, P., 2011. *The effects of Project Management Information Systems on decision making in a multi project environment*. Int. J. Project Management.30 (2), 162–175

Chae, B., D.Olson and C.Sheu., 2014. *The impact of supply chain analytics on operational performance: a resource-based view*. Int. J. Prod. Res. 52 (16), 4695–4710.

Davenport, H. and Jill, D., 2013. *Big data in big companies*. International Institute for Analytics.

Favaretto, M.et al, 2020. *What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade.* PLoS One 15 (2), pp 228–247.

Germann, F. et al, 2014. *Do retailers benefit from deploying customer analytics?* J. Retail. 90 (Issue), pp 587–593.

Halaweh, M., and Massry, A., 2015. *Conceptual model for successful implementation of big data in organizations.* J. Int. Technol. Inf. Manag. 24 (2), 2.

Hugh J.,2019. *Update Tutorial: Big Data Analytics: Concepts, Technology, and Applications*. Communications of the Association for Information Systems, pp.364–379.

Hajli, N.et al.,2019. *Understanding market agility for new product success with big data analytics*. Industrial Marketing Management

Haid, M. and Albrecht, J.,2021. *Sustainable Tourism Product Development: An Application of Product Design Concepts*. Sustainability, 13(14), p.7957.

Ko´scielniak, H., Puto, A., 2015. *BIG DATA in decision making processes of enterprises.* Procedia Comput. Sci. 65, pp 1052–1058.

Kowalczyk, M. and Buxmann, P.,2015. *An ambidextrous perspective on business intelligence and analytics support in decision processes: Insights from a multiple case study*. Decision Support Systems, 80, pp.1–13.

Laney, D, Taylor, P and Gartner, A.,2013. *Big data means big business*. Gart. INC.

Lee, G. and Xia, W.,2010. Toward Agile: *An Integrated Analysis of Quantitative and Qualitative Field Data on Software Development Agility*. MIS Quarterly, [online] 34(1), p.87.

Melville, N., Kraemer, K. and Gurbaxani, V.,2004. *Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value*. MIS Quarterly, [online] 28(2), pp.283–322.

Madamanchi, B and Sreekrishna, T., 2020. *Survey on Growth of Business using Data Analytics for Business Intelligence in Real-Time world*. Gorteria Journal, Vol 33 (12), pp 407-415.

Mazzei, M. and Noble, D., 2017. Big data dreams: a framework for corporate strategy. Bus. Horiz. 60 (3), pp 405–414.

Manyika, J.et al., 2011. *The Next Frontier for Innovation, Competition, and Productivity.* McKinsey Global Institute

Marr, B., 2015.*Big data : using smart big data, analytics and metrics to make better decisions and improve performance*. Chichester, West Sussex, United Kingdom ; Hoboken, New Jersey: John Wiley And Sons, Inc.

Melewar, T., 2014 *Advancing Branding Research: A Qualitative Agenda.* Emerald Group Publishing Limited.

Power, D., 2014. *Using 'Big Data' for analytics and decision support*. J. Decis. Syst. 23 (2), 222–228.

Pramanik, D., Mondal, S. and Haldar, A., 2020. *A framework for managing uncertainty in information system project selection: an intelligent fuzzy approach*. Int. J. Manag. Sci. Eng. Manag. 15 (1), 70–78.

Queiroz, M. and Telles, R., 2018. *Big data analytics in supply chain and logistics: an empirical approach.* Int. J. Logist.

Rahi, S., 2017. *Research design and methods: a systematic review of research paradigms, sampling issues and instruments development*. Int. J. Econ. Manag. Sci. 6 (2), 1–5.

Russom, P., 2011. *Big data analytics*. TDWI Best. Practices. Rep., Fourth Quart. 19 (4), 1–34.

Schroeck,M.et al .,2012. *Analytics: the real-world use of big data*. IBM Glob. Bus. Serv. 12, 1–20.

Serrador, P. and Turner, R., 2015., *The relationship between project success and project efficiency*. Project. Management. Journal. 46 (1), 30–39.

Shah, T., 2022. *Can big data analytics help organisations achieve sustainable competitive advantage? A developmental enquiry*. Technol. Soc. 68.

Saggi, M.K., Jain, S., 2018. *A survey towards an integration of big data analytics to big insights for value-creation*. Inf. Process. Manag. 54 (5), 758–790.

Schrage, M., 2016. *How the big data explosion has changed decision making*. Harv. Bus. Rev. 25, 1–6.

Wang, Z. and Zhao, H., 2016. *Empirical Study of Using Big Data for Business Process Improvement at Private Manufacturing Firm in Cloud Computing. In Cyber Security and Cloud Computing (CSCloud).*3rd International Conference, pp 129-135 .

Watson, H.,2018. *Revisiting Ralph Sprague's Framework for Developing Decision Support Systems*. Communications of the Association for Information Systems, 42, pp.363–385.

Yorulmaz, Ö.,2017. *The Relationship between Socio-Economic Development, Corruption and Health Indicators*: Application of Partial Least Squares Structural Equation Modeling. Alphanumeric Journal, 5(2), pp.191–206.

Yoo, K.-H. and Gretzel, U.,2011. *Influence of personality on travel-related consumer-generated media creation*. Computers in Human Behavior, 27(2), pp.609–621.

Zikopoulos, P.et al.,2013. *Harness the Power of Big Data*: The IBM Big Data Platform. McGraw-Hill, New York, NY.

Zhao, Y.et al.,2021. *Research on operational research-based financial model based on e-commerce platform*. Information Systems and e-Business Management.

InflationData.com., (2022). *Worldwide Inflation by Country*. [online] Available at: https://inflationdata.com/articles/2022/05/21/worldwide-inflation-by-country-2022/.

UNCTAD (2021)., *How COVID-19 triggered the digital and e-commerce turning point | UNCTAD*. [online] unctad.org. Available at: https://unctad.org/news/how-covid-19-triggered-digital-and-e-commerce-turning-point.