AVIATION ISLANDS

Jathin Varma Mandapati

Submitted for the Degree of Master of Science in

Artificial Intelligence

Department of Computer Science
Royal Holloway University of London
Egham, Surrey TW20 0EX, UK

August 31, 2022

Declaration

This report has been prepared based on my own work. Where other published and unpublished source materials have been used, these have been acknowledged.

Word Count: 18450

Student Name: Jathin Varma Mandapati

Date of Submission: 31st August 2022

Signature:

Abstract

Nowadays, big data is mostly utilized for the analysis of vast amounts of data. It is utilised with many various systems, including customer databases, transaction processing systems, and more. Big data will be used in this context to assess all the elements relating to economic indicators of island countries. This can be useful for assessing all the components of the offered island countries' economic indicators. The dataset is obtained from world bank website. The publicly accessible data is in an unprocessed state. In the research report, it has discussed that Economic indicators are statistically based economic activities that enable study of past economic performance and future performance forecasting. Various aspects, including income, GDP, employment rate, production, commercial activity, security markets, international data, and federal finance, are connected to various economic indicators. Economic indicators are based on how well a country is doing, as the research has shown. The country's GDP and growth rate, which are based on various factors including personal consumption, business investment, government spending, and net exports, are the most significant economic factors. On the other hand, debt, debt ratio, and debt cycle are also crucial economic elements that contribute to the country's expansion. The publicly accessible data is in an unprocessed state. This raw data needs to be transformed into meaningful data that, when properly evaluated, may assist businesses in offering crucial insights. The utilisation of Big Data is seen as a vital aspect in today's environment since it has the capacity to derive insights from a given dataset, which is necessary given that accurate data interpretation and thorough analysis are required.

Link to the dataset website –

https://databank.worldbank.org/source/world-development-indicators

Table of Contents

1 Chapter 1: Introduction

1.1 1.1 Chapter overview

Economic indicators are the statistical economic activities that allow the analysis of economic performance and prediction of the future performance. Economic indicators are associated with different factors such as income, GDP, employment rate, production, business activity, security markets, international statistics, and federal finance (Arel-Bundock *et al.* 2019). The research has demonstrated that economic indicators are based upon the enhancement of the countries. GDP and its growth rate are the most important economic factor of the country that is based upon different factors such as personal consumption, business investment, government spending, and net exports. On the other hand, debt, debt ratio and debt cycle are also important factors of the economy that have a positive significant impact on the growth of the country.

This is illustrated as lower the number of debts in a country, higher will be the growth. Debt ratio depends on the debt equity ratio to debt GDP ratios. Inflation is also having a positive significant impact in a country as if inflation expectations determine the way inflation would be involved in future along with including the interest rate of the country (Adebayo *et al.* 2021). Big data is majorly used in recent days for the purpose of analyzing large amounts of data. It is used with different systems such as transaction processing systems, customer databases and many more. In this context all the factors related to economic indicators of island countries are going to be analyzed with the help of big data. This can be helpful for the purpose of analyzing all the aspects related to economic indicators of the provided island countries.

1.2 1.2 Research Background

In this assignment our main objective is to understand the different economic indicators of the provided island countries using big data. Currently, data is considered as a very important source for companies to succeed. The data

which is available in the public domain is in a raw form. This raw data has to be converted into meaningful data which can help the companies to provide critical insights if it is analyzed in a proper way (Ahmad *et al*. 2020). As there is a need for correct data interpretation and proper analysis, the use of Big Data is considered as a critical factor in today's scenario because it has a potential of deriving insights from a given dataset.

In this project we have selected a dataset that has details on various islands across the globe. The data has been extracted from the World Data Bank website. The details extracted are the economic, educational and tourism indicators for a particular island. We would be using visualization techniques to identify the relation between the variables present in the data. We would be using line charts to identify the trends of the indicators (Arel-Bundock *et al*. 2019). These would help us to understand about the condition of tourism in these countries, growth of the island nations over the last decade, growth in education, and the other economic indicators in these countries.

It has been known that in the 21st century data is considered as the new oil. It is because data is a very valuable resource when the extraction is done properly like oil. The meaningful data can help us to provide critical insights if it is analyzed in a proper way. Due to this emerging need for proper analysis and correct data interpretation, Big Data Analytics is considered as an influential domain in the present scenario (Assaf *et al.* 2019). This analytics of Big Data has a huge potential for extracting the insights of any particular dataset for deriving important insights from it. When this refined data is assessed and used correctly, it will lead to multiple positive outcomes. The importance of analytics of Big Data in predicting the overall position of an economy cannot be ignored.

The analysis from the refined data of the dataset will eventually lead to various positive outcomes and can be used by the countries to enhance their GDP. In the present scenario, most of the countries use different models to predict the demand of their economy using Big Data which is not possible just by looking at

the data (Benkraiem *et al*. 2019). A proper analysis of the consumer data will help the countries to make informed decisions which would ultimately increase the GDP of the country and reduce the unemployment of the country.

1.3 1.3 Aim and objectives

The aim of this research is based on understanding different economic indicators of the provided island countries using big data.

Research Objectives

Few research objectives are discussed below:

- To understand the concept of economic indicators
- To analyze the significance of economic indicators for countries
- To understand the significance of big data
- To implement big data to analyze different economic indicators of provided island countries

1.4 1.4 Research Question

Few research questions are discussed below:

1. What is the concept of economic indicators?
2. What is the significance of economic indicators for countries?
3. What is the significance of big data?
4. How to implement big data to analyze different economic indicators of provided island countries?

1.5 1.5 Research significance

This research can be highly useful for the purpose of enhancing different aspects of economic indicators and provided island countries. This has been illustrated as the provided island countries are lagging in the GDP rate due to several factors that can be improved with the help of analyzing the areas which is leading towards a degradation in the performance. However, there are also important economic factors of the island countries covered with the help of which the indicators can be enhanced in such a way that it can lead towards an advanced

growth of the country. As per the view of Foroni *et al.* (2021), The research can also be useful for the managers of companies within the concerned island countries in terms of using information related to the factors that should be included for the enhancement of customer satisfaction so that the company growth can be enhanced, and it can lead towards contributing to the economic growth of the country.

1.6 1.6 Research hypothesis

The hypothesis that the report is going to conduct is that the variables present in the dataset are statistically significant to predict if the value of the GDP growth of a country would be positive or negative. For the analysis of the economic indicators, the report is going to use the Logistic Regression algorithm. The logistic regression technique would predict the value of target variables (0 or 1) using the logit function. In the report, it would be using a Logistic regression model to test the hypothesis and then would understand the significance of each variable in the contribution of a positive or negative GDP growth (Guo *et al*. 2021).

Initially we would be doing an analysis to understand the trend of the indicators provided in the data for each country. This would help us to identify the trend of these indicators in the last decade. After that, we would be doing an intra-country analysis which would help us to identify the relationship of the provided indicators for a particular country. We would be conducting a correlation analysis and visualize that to understand which indicator has a direct or indirect relationship with each other (Hasnisah *et al.* 2019).

1.7 1.7 Structure of the dissertation

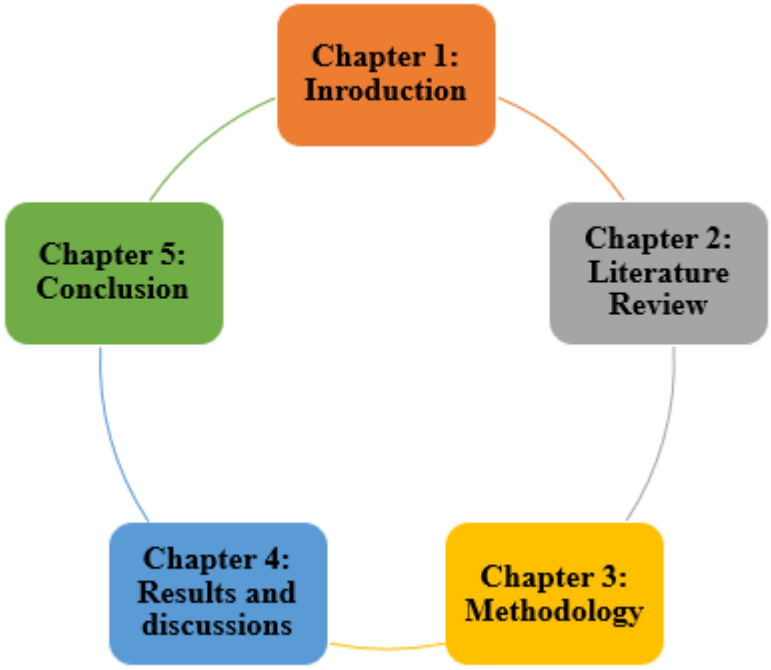


Figure 1.1: Structure of the dissertation

(Source: Self-developed)

2 Chapter 2: Literature review

2.1 2.1 Chapter overview

The economic indicators are associated with different factors of the country such as GDP, inflation, debt, debt ratio, budget, crude oil prices, interest rates and many more. Economic indicators are the most important factors for the country that depends on growth. Big data is a set of large datasets with the help of which different analysis is done. Big data contains a large variety of data with increased volume and more velocity with the help of which the analysis of insights becomes easier. In this context big data is used for the purpose of analysing economic indicators of provided island countries so that the insights of the economic indicators of provided island countries can be done (Hussain *et al*. 2020). As per the view of Kush lev *et al.* (2022), this chapter is based upon reviewing the literature based on different factors of economic indicators of island countries that have been provided. The concepts of economic, educational and tourism indicators are going to be analysed along with a proper analysis of significance of each of the factors on the country's growth. The demonstration of big data and its usefulness in analysing the accurate insights of the data are also going to be done so that it can be effectively used in the further research at the time of generating different visualizations associated with economic indicators of the provided island countries.

2.2 2.2 Analysing the concept of economic, educational and tourism indicators

Through several studies it has been found that the positive Association between economic and tourism growth is developing. As per the islands which are characterized by limiting the factors for human capital, tourism plays a crucial role for the economic growth. As there are some countries and Islands which are still developing economically, thus are found to be dependent on the tourism factors for the economic growth for their country (Lim *et al.* 2022).

It has been shown in a research study that the effect of tourism on the island is developing significantly. As per the author it shows that tourism has mostly a positive impact on the economic growth of the island (Nawaz *et al*. 2021). As the author wants to address, international tourism has more profit than domestic tourism for the economic growth of the island. Similarly, there is another study which shows that the researchers suggested that the growth could come up with a new economic growth plan on island as if it is below the GDP of their threshold for the low economical state of the island, the adequate levels of understanding they on the island, for education and the development of the tourism (Magazine *et al*. 2022). Through this case the studies show the small Islands have a low per capita income.

Different study shows that the investigation of direction in between causality tourism and economic growth. The utilization panel of islands in the period between 1995 to 2007 it reviews the causality between tourism and the economic growth are short but has a long run in them. The relationship between tourism and economic growth supports their backs and has a huge economic Evaluation they have a nice exchange of income between the islands of the country. The positive relation between tourism and growth that the data represents in the period of examining the growth impact by multiplying the measures of tourism by cleaning the parks and hotels and giving them hygienic food as they liked. As per the view of Nawaz *et al*. (2021), they only have a limitation for international tourism as they do not know what they are, where they have come from and what they need. That was quite a change for them. The arrivals of them may not reflect a true impact of growth in change and are not accounted for instance assuming no change in the total tourist arrivals.

A study by Thacker et al. in 2012 conventional Douglas production was used while investigating the compositions of the tourism which plays a crucial role in economic growth and educational forms in the islands shows the capital accumulation and total factor productivity. The result of tsp in particular shows the transition of intensive agriculture sector as the tourist wants to see the scenery

of the island and the Beauty behind that in many countries tourism has been enhancing their growth by keeping their environment clean and tourist friendly environment around them so they can enjoy themselves while they are visiting that Island and by that tourism helps the economic growth of the island (Nguyen *et al.* 2021). Due to this tourism many of the islands have been recovered significantly in an economical way.

2.3 2.3 Evaluating the significance of economic indicators

The significance of economic indicators is basically that it indicates the state of an economy and understanding. They are for traders, investors, and analysts. They provide information which helps them Discover new opportunities by adjusting their portfolios and by their understanding and the need for what the economic indicators have suggested to them. There are five leading indicators which are most useful to follow are durable goods order, manufacturing orders, building permits, yield curve and the stock market. As per the view of Nti *et al.* (2019), The economic indicators are most important and have the most comprehensive measure for overall economic performance on the GDP of the island. They can be divided into two categories or groups. Mostly the economic indicators have specifically schedule for their release allowing the investors to prepare for certain information in the time of month or the year (Adebayo *et al.* 2021).

The data on the financial position moves their changes before the economy. That's why the name of their category changes for everyone in a while. Considering the information that they may gain economic growth, but they can be incorrect too (Philip *et al*. 2022).

Coincident indicators include GDP, retail sales and employment levels have been seen with specific activities done economically (Nguyen *et al.* 2021). As per the view of Plagborg-Møller *et al.* (2022), the metrics show the activity to the region, the policymakers who follow the real-time deals through the economists. Lagging indicators including CPI, interest rates, GNP and unemployment rates are

seen only in specific economic activity occurring. That implies, the set data shows the information that event has happened. The trailing indicators are also known as technical indicators which usually follow large economic shifts.

Interpreting economic indicators

The indicators which are useful to interpret it correctly. History has shown the strong correlation between economic growth which is measured by GDP and corporate with profit growth, the specific growth of the company is through earning on the indicators by GDP it is nearly impossible. No denying of gross domestic product, existing home sales and interest rate or other indexes, the activity level of a major portion of the overall economy is why investment and the cost of money through spending has been shown in this figure (Nguyen *et al.* 2021).

The stock market as an indicator

The leading indicators for cars that the economy is headed on the top of the stock market itself. The most crucial leading indicator is the people who everyone looks up to. As the stock prices in the stock market rises and falls the forward-looking performance makes and it indicates the economy direction through its earning and estimates and accuracy. As per the view of Rehman et al. (2029), The market may suggest the estimate of earnings through the stocks are overly high in economic rates conversely, they have a down market which may indicate with her company earnings it also expects to suffer. The limitations to the usefulness in the stock market can be indicated because the performance is not guaranteed through its estimates as they have risk in them. It is a subject which manipulates the street traders and the corporations by the manipulation of the inflating stock prices, high volume trades and the complex financial derivative strategies through creative Accounting Principles as they contain that they are both legal and illegal (Saint Kediri *et al.* 2019).

2.4 2.4 Critically analysing the educational indicators and its significance in country growth

The educational achievement for the country can be measured through the terms indicators for adult literacy rate, years of schooling, enrolment ratios in schools and teacher pupil ratio. The indicators give an idea for the schools to spread the education that may or may not be the reflection of the quality of education in a country (Nguyen *et al.* 2021). As per the view of Rehman *et al.* (2029), The education is the human right which develops the powerful drive for one to be the vital player for their life and for the country by reducing the poverty each and every family will be happy for they will be main economic powerhouse of the family so that they may eat every day and be happy, peace, gender equality and stability in their home. Delivers a large consistent Returns for the income on that important factor this ensures the Equalities for their opportunities on their needs of doing Things (Hassan and Romilly, 2018). As per the view of Plagborg-Møller *et al.* (2022), The adult literacy rate must be over 15 population and the literacy rate must be from 15 to 24 years old population with a gender equality they must do without any misconception and do their job and by expanding on the adults and continuing further education in total expanding on education, enrolment ratios of age group of 3-5 years is the age where the students Must start their studies (Sen *et al*. 2019). The primary grade is pupils having attended some from the organization by ECCE programmes, their habitations is Having a primary school facility in 1 km diameter so the students of each place can attend that school easily without facing any problems the primary schools that have toilet facilities, so the faculties and the students doesn't have to go through with any such problem under school's jurisdiction. The five keys for educational development are by maintaining students' achievement, attendance, rate of graduation rates, and discipline referrals (Sorci *et al*. 2020). The school quality for the success Which indicates that it had been referred to as it had the 5th indicators which a required conjunction with the other four indicators students' growth, progress in achieving proficiency in English language, academic achievement, graduation rate.

The major indicators for the development in education is that the educational indicators can be classified into the size or quantity, efficiency, equality and quality. The access to schooling and facilities is said to be one of the important components of UEE. For knowing these facilities are equal in distribution or not for their teaching and understanding the students whether they are good in studies or not. As per the view of Plagborg-Møller *et al.* (2022), The meaning of indicators in the education statistics is to be described as it plays the key role in schooling to permit the evaluation by monitoring the students and the Teachers of those schools and by performing exams and tests on the students whether that they are being taught properly or not. It shows they can be getting scholarships for the university or the colleges (Udemba *et al.* 2021). The main attributes for the education indicators are perseverance, prior knowledge, demographic variables, readiness of schools, barriers of learning and the Lerner aptitude. Some examples for the education indicators may be given are time spent on homework, home and school language and gender differences in Earning through which the students must be treated equally (Hussain *et al.* 2020).

2.5 2.5 Observing the tourism indicators and its importance in growth of the Island

Indicators are considered practical instruments that help tourism managers diagnose the state of the location and recognize and evaluate problems that need to be resolved to raise the level of sustainability of the tourism destinations. When indicators are used effectively, they can develop into essential management tools, serving as performance indicators that give managers and all other tourism participants' vital information. Effective indicators can deliver timely information to address urgent problems and support the sustainable growth of a location (Assaf *et al.* 2019). There is a demand for indications that aid in evaluating and coordinating environmental sustainability in order to track the progress of ecological sustainability and to enhance the planning and implementation. Environmental, economic, and social metrics all have been found for the three

facets of sustainable tourism activities. Islands and other peripheral, distant, and insular places have benefited greatly from tourism. As per the view of Benkraiem *et al*. (2019), these locations are often distinguished by their isolation, fragility, inadequate resources, unemployment problem, and high transportation costs, all of which are comparative disadvantages. Due to the increasing popularity of island travel, development is a crucial and inevitable issue, resulting in consequence; most islands are relying heavily on tourism as a key driver of economic growth. The relevance assessment, impact assessment, and economic evaluation are really the three main analyses used to determine the economic consequences of tourism. One industry that has a lot of momentum is tourism, which is generally growing faster than the world economy.

Significance analysis

This analysis of the importance is significantly concerned with the extraction of the structure and size of the tourism sector from the information of the national account and abroad. This system usually focuses on calculating the proportion of expenditures and GDP that the tourist industry accounts for. As per the view of Foroni *et al.* (2021), The holistic approach furthermore encompasses a variety of requirements for defining the terms "tourist" and "visitor," along with "surroundings as usual," "tourism destinations consumption characterization," and "gross fixed capital structure in the tourist industry," "tourist and tourist industry product lines and actions," and methods for determining the charitable donation of the overarching tourism industry in the domestic product" (GDP) (Guo *et al*. 2021). Additionally, it provides information on how to calculate the travel industry's commitment to various kinds of consumption and activity and presents that information in numbers and graphs.

Impact analysis

This type of analysis concentrates on how money spent by outside visitors affects the economic growth of a region, whether it has significant affect (on hospitality

industry, transportation, exhibitions, and retail outlets) or even an incidental one (on revenue and employment), while also keeping in mind the way that the tourist industry interacts with other economic sectors. This methodology is distinguished by adaptability regarding the degree of circularity, which might be used to examine the significant effects of tourism on the city's population in its entirety or examine the effects of a particular paying tourist on a certain category of tourism activity (Hasnisah *et al.* 2019). As per the view of Khurshid and Khan (2021), the framework employs a variety of approaches, such as information tables for analysing the impacts of rising tourism expenditure on the diverse economic sectors and by examining the interdependencies among these industries.

Cost-benefit analysis

The Cost-Benefit Analysis is a method of analysis that focuses on the viability and effectiveness of developmental activities from the perspective of society. It is thus interested in determining the private and societal advantages (like the intellectual benefits of tourist activities) as well as the private and societal expenses (such as implications of the environmental impacts of tourism.

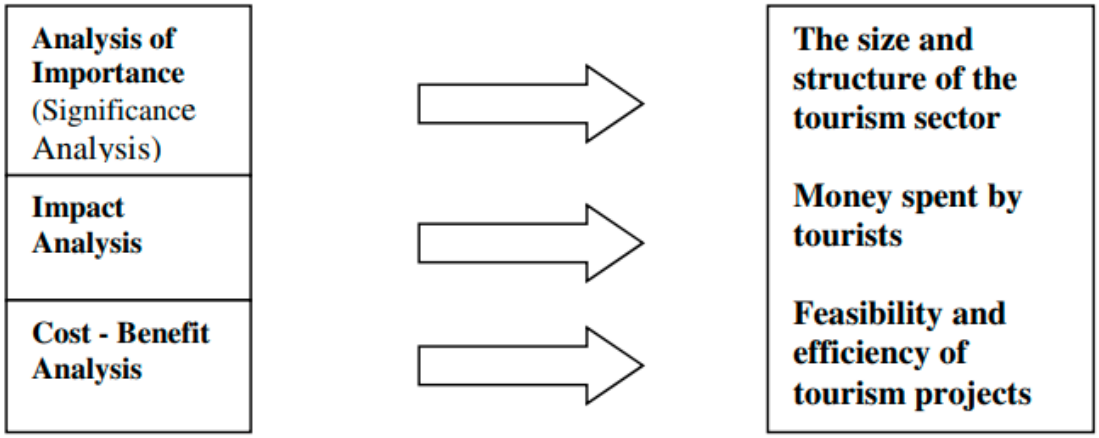


Figure 2.1: Main three indicators for measuring the economic impact of the tourism

(Source: Karintseva *et al.* 2019)

It has identified several important concerns affecting the growth of tourism on islands, keeping in mind the significance of tourism. Whether or not it is necessary, the growth of tourism has both beneficial and detrimental effects on the economy, society, and the environment that must be thoroughly examined. Statistics clearly demonstrate the significance of the tourist industry; without which, even the largest immigrant economy may experience instability (Karintseva *et al.* 2019). Regardless of economy and financial slump of the previous five years, the United States tourism industry generates around $110 billion in monetary revenues yearly, which is ten times more than the European Union (Khan *et al.* 2022). This makes tourism a significant reason solution to overcome economic collapse. Throughout this essay, the repot will talk about how to gauge the growth of the tourist industry as well as the relationship between travel and the reduction of company complexity and environmental sustainability (Jena *et al*. 2021).

2.6 2.6 Analyzing the factors associated to economic indicators

Economic and financial statistics are often divided into three categories by economists such as leading, leading, or coincidental. Remarkable coincidence and trailing indications are an excellent starting point since they help predict which industry might well be headed while offering companies some conformation regarding where the marketplace is and how it has previously. As per the view of Kushner *et al.* (2022), the below section has represented different factors which has influenced their economic impact of the tourism industry and discussed different input variables for the analysis. The basic requirement it has maintained for the improvement of the economic contribution which has impacted the tourism development. Few factors such as the number of the visitor days, spending the amount par visitors, types of the visitors and the trip purposes and also the development of the economic model for calculating the multiplier effects (Khurshid and Khan, 2021). Other than that, there is also moderating effect of this special limitation of this survey area. Considering the visitation is properly

maintained to improve the economic impact of the tourism sector according to the national skill and the domestic tourism has represented the distributive effects (Hussain *et al*. 2020). Such visitors have spent there on vacation in the home country and would have effectively visited at another destination. In this case, only incoming tourists have provided that additional input for this national economy, however, they might be argued that domestic trip can also be about it its trip abroad which would significantly lead to the leakage from this national economy.

Positive impacts

Numerous islands have viewed the growth of tourism as a natural outcome; "it appears like a common consensus between islands, particularly the tropical islands, that they are no effective decision but to embrace conventional sustainable tourism and international visitors as either a certainty," says one islander (Nawaz *et al*. 2021). According to several tourism academics, islands' reliance on tourism to assist them to get beyond their size restrictions has grown significantly. It is debated whether economic repercussions of tourism must be utilized on a national level when taking visitation into account because most those of tourism infrastructure only indicate distributed effects. These tourists would have either spent their holiday there and still elsewhere then, depending on where they were traveling from. Only foreign visitors increase the country's economic output (Hussain et al. 2020). One could counter that traveling within the country could prevent travel overseas, which would also cause economic leakage.

Income balance of the payment and foreign exchange

Numerous studies have revealed how important tourism has become to the economies of very many locations. For instance, tourism revenue contributes 58% of the Gross Domestic Product (GNP) on the island of Bermuda, while it contributes almost 50% of the Economy in the Caribbean Islands. Additionally, tourism is crucial to many pacific islands as a provider of additional foreign

currency. The country's infrastructure is developed, its revenue is increased, and a sense of cultural interaction between locals and visitors has sown thanks to tourism. In numerous locations, tourism generates a sizable number of jobs (Nguyen *et al.* 2021). The demand and availability for a state's currency in the foreign industry at such a specific rate of exchange are referred to as the marketplace payment balance (Lim *et al.* 2022). The market payment equilibrium would be in balanced by accident if the price level was fixed. Purchasing or purchasing countries besides the one's own is referred to as currency currencies. Its cost of a unit in relation to a different currency is the value at which one nation's currency could be converted into that of the another.

Employment

Tourism has become one of the most prominent and expanding industries in the world economy. This industry contributes significantly to the growth of a nation's economy. The economies of the countries can benefit from a rise in tourism, particularly in terms of GDP and employment outcomes. Diverse employment prospects are generated by tourism and tourism-related industries in a variety of industries, including lodging, restaurants, transport, tour operators, tour operators, and tourist and cultural destinations (Nawaz *et al*. 2021). Additionally, the employment that is brought about by tourism seems perhaps just as significant as income generation. In contrast, 10% of occupations in American Samoa are tied to tourists, and part-time work linked to tourism improved average earnings in the Faroe Islands. Most islands' jobs, though, are typically low-paying and socially unequal.

Diversification of the economic environment: development

Despite what the opponents claim, several tiny islands are now utilizing tourist experience as a strategic approach to boost their economic growth, develop more rapidly, and expand their economy. As a means of economic diversifying, it has sometimes been proposed that tourism should really be promoted or perhaps even

launched, as in the instance of Seychelles, where the World Economic forum pushed this market throughout the forecast period to lessen the nation's excessive reliance on production of sugarcane (Lim *et al.* 2022). As per the view of Magazine *et al*. (2022), the infrastructures of the island are what promote tourism, which has in turn strengthens the social circle of both the islands. Tourist industry also raises the living standards on island nations. Both have been improved in this regard by tourism. The development of tourism, the improvement of the infrastructure, and the socioeconomic development has properly maintained to improve the tourism factors. Some islanders may choose to depart because of the poor condition of living and chances there. Potentially, the tourist industry, which offers transportable and perpetually renewed resources, might be a means of resolving some of these issues.

2.7 2.7 Evaluating the factors influencing educational indicators

One type of tourism is academic tourism, which helps students learn more effectively and expands their knowledge. The attraction of learning new skills and attaining enlightenment from the best places on earth has led to a growth in commitment to promotion. Statistics that reflect important facets of education are known as educational indices. This information allows for the evaluation and oversight of schools, educators, programs, and individuals. It is believed that the degree to which each indicator's data may show an impact on student results should be the focus of evaluate the ion of instructional indicators (Lim *et al.* 2022). This article also offers ideas for developing a framework to evaluate the reliability of indicators and to pinpoint the immediate and long-term consequences of the factors discussed on student results.

As per the view of Nawaz *et al*. (2021), there are criteria for reliable indicators. Finally, conclusions about the value of developing consistent predictors are formed regarding research in education policy. The conservation of the natural environment is now seen as a commitment as the productive capacity of tourism has now become universally acknowledged. Therefore, on the plus side, tourism

may be recognized for increasing environmental awareness and offering motivation for conservation practices. Most island administrations are now putting a greater focus on "clean up" initiatives, as in the cases of Tonga and the Cook Islands. Other than that, the degree of adult literacy, the enrolment rate in schools, the number of years spent in school and the instructor proportion are some examples of metrics that can be used to assess a nation's educational success. As per the view of Nti *et al.* (2019), these metrics provide an overview of how widely knowledge is dispersed, but they might not accurately represent how well educated a nation is. Price competitiveness, infrastructural investment, environmental stewardship, technological development, human resource management, the degree of openness, socioeconomic evolution, and human tourists are the seven major characteristics.

This system of indicators addresses important aspects of tourism sustainable supply chain management, including financial gains, unemployment, seasonally, availability, demand, contentment of visitors and locals, sustainable development, and the efficient use of water, energy, and waste. The amount of education and understanding in a community is one of the fundamental measures of the progress of its industries (Lim *et al.* 2022). As a result, academic achievement university education seen has a major priority that supports both the growth of the economy and the advancement of society as a whole.

2.8 2.8 Analysing the factors of tourism indicators

Decision-makers can develop plans that will help instead of the preservation by evaluating the impact of on the environment from tourism. The International Sustainable Tourism Criteria (GSTC) can be employed by decision-makers to assess how tourism impacts the environment, surrounding area, and traditional culture. The key consideration in planning, especially preparing for tourist, is time. The possibility of accomplishing tourism-related societal and economic development depends on it. National governments and global non-profit groups are paying an increasing amount of attention to the tourism industry (Nguyen *et al.*

2021). Although the phrases are rarely employed in this manner, modifications in these principles' nonlinear characteristics themselves in election outcomes, and sustainable and resiliency strategies serve as the foundation for several political disputes. As per the view of Nti *et al.* (2019), Different indicators are outlined for evaluating the relationship between the resilience and sustainability in the project development and the sustainability may significantly referred different approaches and challenges for the improvement of the tourism department. The below table has shown that challenges, link to the implementation, the interpretation and Measurement, prevention of the challenges and also data method and resources for the improvement of the tourism industry.

Core indicators	**Description**
Tourism impact and performance	• ***Tourism direct Gross Domestic Product (GDP)*** (It has lead the international measurement of the tourism contribution to the GDP) • ***Exports of the tourism services*** (It measures The contribution of the tourism sector for the exports) • ***Overnights in different types of the accommodation*** (It measures of the tourism flows in the accommodation) • ***Inbound tourism revenue according to the visitor by the source market*** (It measures of the economic activity of the visitors in the tourism sector)
Ability of the tourist destination for delivering the quality and the competitive tourism service	• ***Labor productivity in the tourism services*** (It measures the entire evidence of the entire product potential of this tourism economy) • ***Visa requirement in the country entry*** (It measures the visa requirement features included the method of the visa insurance.) • ***"Purchasing power parity (PPPs)"*** and the tourism prices (It measures the tourism competitiveness using the purchasing power.)

Attractiveness of the tourism destination	• ***Visitor satisfaction*** (It measures the demand and also the attractiveness value according to the future and present competitiveness) • ***Creative and cultural resources*** (It measures of the Creative and cultural attractions, event and activity of the country) • ***Biodiversity and natural resources*** (It measure of the country's stock of the natural assets)
Economic opportunities and policy responses	• ***Tourism Action Plan*** (it represents a measurement which has indicated the effectiveness in assisting of the improvement of the competitiveness of the tourism industry)
Supplementary indicators	
Tourism Impact and performance	***Growth market and market diversification*** (Eat the text the measurement for capturing the broad basis of the performance in different source market and focused on the growth market which has received the highest score in the tourism sector.)
Ability of the destination for delivering quality and the competitive services in the tourism sector	• ***Employment in the tourism by education level, the type of contract and age*** (It Detect the measurement of the accessibility regarding the development of the competitiveness in the tourism sector.) • ***Consumer price index for the tourism*** (It Detect the measurement of the consumer price.) • ***Inter modality and their air connectivity*** (It Detect the measurement of the competitiveness which has revealed in the flight time and fare air routes from the main market and the passenger numbers)
Attractiveness of the	***Better life index*** (It measures the focus version of the tourism as per our

tourist destination	index)
Future development indicators	
Ability of the destination for delivering the competitive tourism and quality	• ***Maintenance of the organisational mortality rate (***It measures the enterprise activity and the business churn.) • ***The government budget appropriate options for the tourism*** (It measures the measurement of the national government tourism expenditure as per capital.)
Economical opportunity and policy responses	• ***Structure of the tourism supply chain*** (It measures the index measurement of the industrial classes, competitiveness, thickness, potential and existing.) • ***Use of the e - tourism and the other innovative services*** (It measures Measurement of the innovation and also the use of the social media in this tourism industry.)

Table 2.1: Supplementary, list of core and development indicators

(Source: Nguyen *et al*. 2021)

2.9 2.9 Understanding the concept of big data

Even predicting which new products might succeed in a market is possible with the use of big data. Effective usage of big data technology can result in personalized offers for tourists that are catered to their wants and interests. They can get better experiences that are centered on the demands of the clients. These kinds of data can be useful to tourism boards and businesses in many different ways (Nguyen *et al.* 2021). This entails focusing marketing efforts, creating

packages based on the preferences of consumers, and selecting which nations to target for customer acquisition. Moreover, these findings can greatly aid in decision-making and enhance how the tourism sector functions (Sen *et al*. 2019).

Participants in the tourist industry can now base their judgments on statistics and information driven by numbers. Every step of the way through the travel planning process, they may pinpoint certain types of potential consumers. As per the view of Nti *et al.* (2019), they could also improve service quality and efficiency. Even predicting which product innovations might succeed in a market is possible with the use of big data. Effective usage of big data technology can result in personalized services for travelers that are catered to their wants and preferences. They can get the positive quality which is focused on the demands of the clients. Big data can be utilized in instantaneously to anticipate and reply to visitor demands almost instantly, as well as a decision support tool to foresee future trends. According to the study, there is a growing tendency toward the usage of big data in the tourism industry. Big data raises expectations for a better comprehension of tourism demand and the modification of supply by travel agencies to satisfy traveler needs and maintain profitable operations (Philip *et al*. 2022).

Big data is a crucial idea to understand in the hospitality sector since it can assist hotel owners and other business executives in spotting significant patterns and trends. As a result, it may aid in enhancing customer service, optimizing marketing initiatives, and improving revenue management. The use of booking information to develop products, competitive tools, and targeted advertising is a specialty of the travel industry. These sectors are able to produce highly individualized emails for their consumers thanks to the enormous amount of data that is collected during the bookings process (Udemba *et al*. 2021). As per the view of Rehman et *al.* (2029), big data gives hospitality businesses the ability to act based on live metrics and real-time data and helps them reframe important questions about how they conduct their research, gather their data, and use it. The tourism

industry, taken in its broadest sense, includes all companies that directly supply products or services to support business, pleasure, and leisure travel.

2.10 2.10 Analyzing the Big data incorporation in analyzing indicators of Island

The economists currently use Big Data analytics to understand the overall economic health of a country in a better way. An appropriate analysis of these indicators can help the analysts to frame informed decisions which would help to increase the overall GDP of the country (Vogt Isaksen, 2019). This research paper will focus on giving details of how, what and why of the technology of Big Data in a detailed manner which would help us to identify the effect of Big Data technology on the economic indicators (Verma *et al.* 2021). The countries need to analyze the trend of the economic indicators which is changing at a rapid pace to sustain itself in the long run in such a competitive scenario. This can be effectively done using the technology of Big Data which would help us to gather insights from the pattern of the indicators. This detailed analysis will keep the countries in an advantageous position compared to other countries which is not using the Big Data technology to garner insights.

The Big Data analytics has different stages, which has discussed in below section:

Storage of data: There is a huge set of data that is generated in every fraction of a second. This data has to be stored in a secure place to use it for further analysis (Wang *et al.* 2020).

Mining of data: This step consists of identifying the trends of the complex and large datasets which is generally available in raw and unstructured form. Before proceeding with the mathematical and statistical data analysis, it is important to identify the pattern and the trends of the data in an appropriate manner (Verma *et al.* 2021).

Analytics: It involves a process of analyzing the data which has been converted to a structured form. This step will help to get helpful insights from the data (Wang and Wang, 2020). The algorithms used in this stage used advanced forms of mathematical, IT and statistical techniques to get key information from a dataset available in a structured form.

Visualization of data: This step involves the graphical illustration of the data. This helps the analysts to get an idea about the data efficiently by recognizing the trends of the data and the outliers present in the data (Verma *et al.* 2021).

In the next part it would be discussing about the logistic algorithm which we would be using in the analysis:

Logistic Regression: This algorithm is a kind of supervised classification algorithm where the dependent variable or the target variable can have only discrete values for a given data. It usually builds a regression model to forecast the probability that whether the set of independent variables belongs to category 1 or 0. As per the view of Verma *et al.* (2021), This model is usually used for predictive analytics. It tries to estimate the event probability based on the different set of independent variables. As the output is a probability, the value of the dependent variable is in the range of 0 and 1. In this technique, a logit transformation is used on the odds which is the event success probability divided by the event failure probability (THANDA, 2022).

2.11 2.11 Conceptual framework

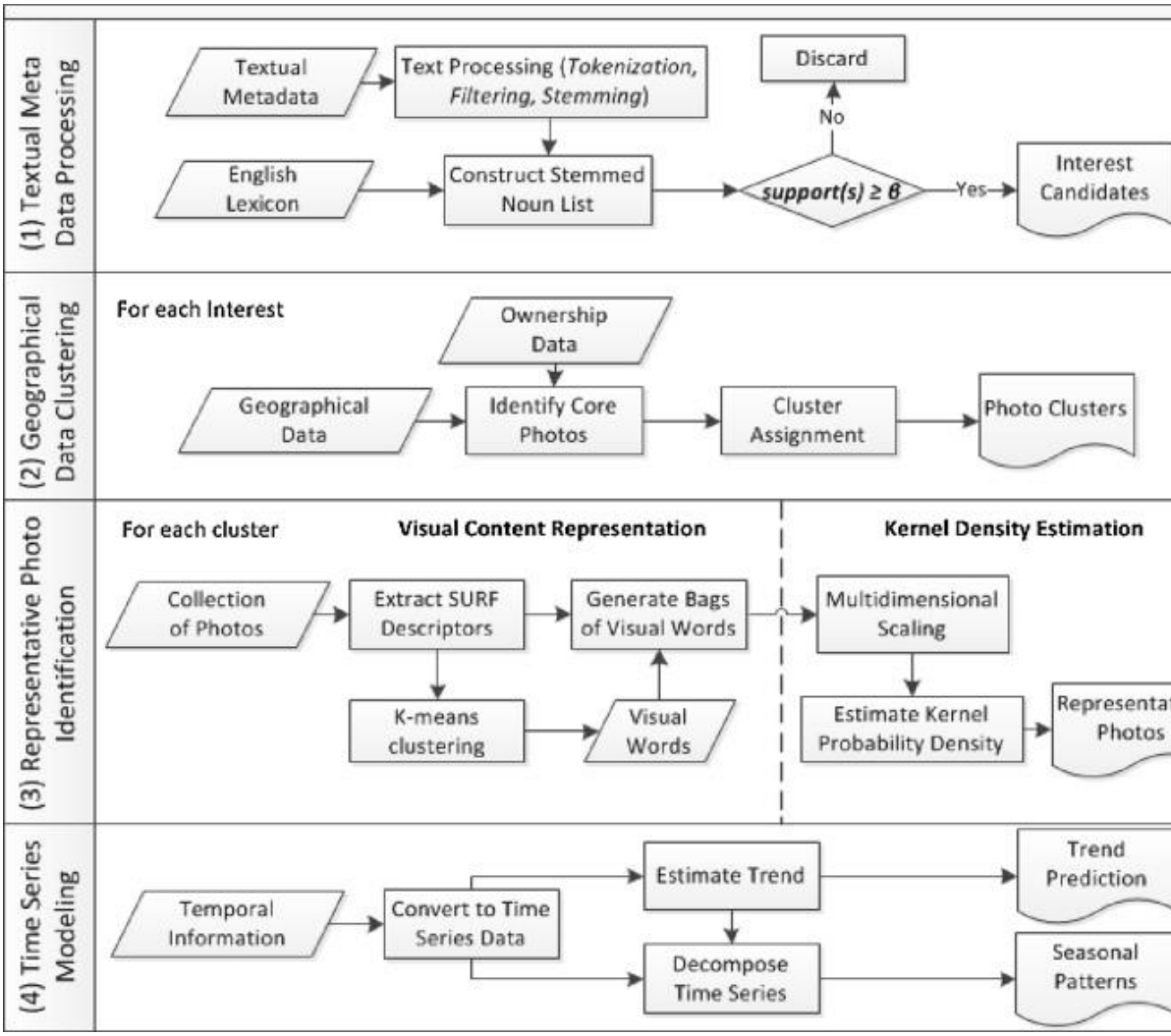


Figure: Conceptual framework

(Source: Self-created)

2.12 2.12 Literature gap

This essay provides a sympathetic analysis of the context and studies around big data in tourism, including its general characteristics, categorization, methods of analysis, and current state of research. The instance of the tourism model is used to highlight the difficulties in carrying out big data tourist research. In doing so, it makes the case that the key to exploiting big data in the tourism sector is to

efficiently implement heterogeneous data structures from many sources. As per the view of Plagborg-Møller *et al.* (2022), This calls for advanced techniques and a cutting-edge conceptual framework that enable us to investigate putative mechanisms incorporated into our reality but are challenging to identify and validate using conventional approaches. Moreover, there is an abundance of data about tourism, but it has long been difficult to turn it into knowledge that is valuable and utilize it to guide good decisions. Customs and border control forms must frequently be filled out when attempting to cross country boundaries, disclosing personal information is typically required when booking a hotel room and bank card and Etihad airways companies also frequently accumulate a significant amount of individual data (Nguyen *et al.* 2021). However, much of this data is proprietary and not accessible for educational studies, even though the organizations that gather it may carefully analyze it. The regular publication of some data, such as national tourist statistics, offers helpful context for both research and practice.

Taxation, travel marketing, infrastructural problems, security concerns, and cross-border laws are the top issues facing the tourism industry. Too many tourist attractions lack visitor-friendly amenities. Sometimes tourists or travelers may think that travel advertising is overblown. Companies need qualified data specialists to manage these contemporary technology and huge Data tools. To work with the technologies and make sense of enormous data volumes, these experts will include data scientists, data analysts, and data engineers (Sen *et al*. 2019). Lack of big data experts is one of the problems that each company has to deal with. This frequently occurs because, while data handling tools have advanced quickly, most professionals haven't. There must be concrete actions done to close this gap.

2.13 2.13 Summary

The significance assessment, performance measurement, and economic evaluation seem to be the three primary analyses used to determine the economic

consequences of tourism. One industry that has a lot of momentum is tourism, which is generally growing faster than the world economy (Verma *et al*. 2021). Due to its considerable contribution to the overall gross domestic product, tourism is a key industry for the advanced markets that were studied in this regard. It can also say that, for most of the period under consideration, several different countries under investigation saw an increase in their national GDP due to the tourism income they engendered (Sen *et al*. 2019). However, after the worldwide financial crisis took hold throughout 2008, every one of the four nations had a comparable decline in Gross domestic product. Our primary goal in this assignment is to use big data to comprehend the various economic indicators of the offered island countries. Data is currently seen as a very crucial source for businesses to flourish. The publicly accessible data is in an unprocessed state (Sorci *et al*. 2020). This raw data needs to be transformed into usable data, which, if properly analyzed, may assist businesses in offering crucial insights. The utilisation of Big Data is regarded as a significant aspect of today's environment since it has the capacity to derive insights from a given dataset, which is necessary for accurate data interpretation and effective analysis.

3 Chapter 3: Methodology

3.1 3.1 Chapter overview

Secondary data will be gathered using a qualitative research methodology from a variety of pertinent books, periodicals, publications, and other sources. The success of the qualitative approach will be evaluated using the exploratory technique. Depending on the objectives to be achieved, a suitable technique must be carefully chosen when conducting primary market research (Edema *et al.* 2021). In this section, the process for data collecting, analysis, and sampling will be thoroughly explained.

3.2 3.2 Research philosophy

The research philosophy serves as a roadmap to help the researcher analyse the research challenge, choose a method and strategy, and acquire the data needed to explain the problem. The two major pillars of conducting research appear to be positive thinking and interpretative thinking (Sen *et al*. 2019). These two ideologies look into what knowledge, humanity, and knowledge are all about. Self-improvement is frequently linked to natural science research, especially when observations and testing are conducted using organized scientific methods. The study is methodically and logically carried out in order to comprehend a certain event and predict its consequences, either directly or indirectly as a result of causal relationships. Other than that, the interpretation maintains that each person's perspective on reality shapes how society is structured (Udemba *et al.* 2021). As a result, a different, more quantitative method is used to investigate and comprehend societal challenges. The interpretive idea, which contends that society's structural constitution is excessively subjective since it is influenced by views, is used to approach this topic. This seems to be quite consistent with the results, given that judgments and, by extension, buying behaviour are essentially

subjective and considerably vary from person to person. Such social sciences cannot be grasped without personal interpretation.

3.3 3.3 Research approach

The research Approch explains the researcher's approach to the theories and facts. The alternatives include using a deductive method, deduction, and marketing and advertisement research techniques. Deductive method study begins with data; using the gathered information, the researchers develop a general understanding before developing a framework and a set of hypotheses (Udemba *et al.* 2021). Deductive research approaches the study challenges by first applying the recognized theories to them. Data are gathered to create a new theory, and then data are gathered again while a deductive approach is used to support the theory. The conceptual work has undergone various revisions as a result of the deductive and inductive approaches, which have also outlined how the research aims would be put into practice (Wang and Wang, 2021).

3.4 3.4 Research design

In the research report, the researchers gather data using a deductive method in an effort to come up with a novel idea. Researchers gather information to form a specific hypothesis, and then they gather information once more using a deductive method to support their idea. The straightforward, rational approach was less effective due to the lack of recent data on the tourism development. According to this deductive research approach, it has been anlysed that Businesses are spending more on the hiring of qualified professionals. To get the most out of the current employees, training programmes must also be offered to them (Sen *et al*. 2019). The purchase of information analytics products powered by artificial intelligence/machine intelligence is another crucial action performed by enterprises. Professionals with a basic understanding of data science but no data science expertise frequently recommend these Big Data Tools. By taking this action, businesses can save a tone of money on hiring.

3.5 3.5 Data collection method

Both secondary and primary data are incorporated into this study. Secondary research refers to information that has already been acquired. To acquire primary data, such as that from focus groups and depth interviews, a variety of techniques can be employed, including as observation, interviews, experimentation, and motivating research. The main technique used to collect the crucial data for the study was semi-structured conversations. For this investigation's secondary research, both printed and online materials were used. Digital resources were provided from the university bookshop, and internet media was sourced from the Primo and Google Scholar websites (Vogt Isaksen, 2019). An analysis of the literature has been conducted to look at past research, scholarly papers, and books, as well as inquiries within those fields of second-hand stores, notably high-end stores, with regard to the incorporation of independent sources according to the conceptual basis. In the research project, it has been stated that Companies' efforts to use Big Data fail due to a lack of understanding. Employees might not be familiar with the definition, sources, processing, and storage of data (Verma *et al.* 2021). Data experts may be aware of what's going on, but others may not have a clear picture. Employees might not preserve the backup of sensitive data, for instance, if they don't comprehend the value of knowledge storage. They were unable to effectively store data in databases. Because of this, it is difficult to get this crucial data when needed.

3.6 3.6 Data Sampling

With non-probability sampling, it is challenging to foresee who will be selected from the community. Throughout the market analysis, using this logical sample data is typical practise. Non-probability sampling techniques should never be used for statistical significance or to draw conclusions about the broader population from which the sample was taken. Because it was impossible to conduct a randomly selected sample without having access to the entire population of Local Residents, non-probability sampling has been employed as a fallback strategy

(Wang *et al.* 2020). Due to the fact that probability sampling requires numerical computations to calculate the sampling, it has been established that it is now too time-consuming and challenging for this kind of research. Since they can be utilized to approach evaluation and emotions, interviews are appropriate for research enquiries. The convenience sampling method was therefore used in this investigation. This kind of sampling comprises selecting components from the statistical sample that are useful or simple for the researcher to obtain (Verma *et al.* 2021). However, using this convenient sampling strategy could result in bias. Due to insufficient representation, the findings cannot be extended to the entire community. There may be both planned and impromptu conversations. Individual inquiries are not preplanned; rather, as the dialogue continues, new ideas naturally arise. On the other hand, semi-structured dialogues include a plan that tries out specific predefined questions.

3.7 3.7 Ethical consideration

The proper storage of these enormous volumes of data is one of the most urgent problems with big data. Companies' data centers and databases are storing an ever-growing amount of knowledge. These data sets become more difficult to manage when they expand rapidly over time. The majority of the information is unstructured and is derived from text files, movies, audio, and other sources. This implies that the database does not contain them. When working with Big Data, there can be serious implications and repercussions that have an impact on a company's reputation, customer relationships, and eventually revenue (Wang *et al.* 2020). Compressing, power leveling, and duplication are three contemporary approaches that businesses use to manage these massive data collections. By using compression, data can be made smaller overall by reducing the amount of bits included inside. The elimination of redundant and undesired material from an information set is known as duplication. Organizations can store data in various storage tiers thanks to data prioritization. It guarantees that the data is kept in the ideal location for storage (Verma *et al.* 2021). Depending on the volume and

relevance of the data, the data tiers are frequently flash memory, cloud platform, and cloud service provider. Businesses are indeed selecting Big Data technologies including Hardtop, NoSQL, and others.

Both internal and external trust might be undermined even by the perception of unethical data processing. In the tourist and hospitality sectors, ethics is a fundamental idea that is essential to the operation of a firm. It is an important component of guiding strategic goals, short-term goals, and long-term goals and objectives for the sustainability of an organisation. As per the view of Plagborg-Møller *et al.* (2022), Poor ethics will make it difficult for a hotel business to retain clients, and employee morale will suffer. However, there will be moral conundrums, in which a human must decide between acting immorally or morally.

3.8 3.8 Pre-processing and Imputation

a. Selection of a proper dependent variable: The dataset related to the economic indicators of a country has multiple variables inside it. From the dataset it is important to understand which is the correct variable that has to be selected as the output variable which has to be predicted during the development of the model. From this dataset, we have created a condition which assigns the dependent variable to be 1 if the growth of the GDP of a country for a particular year is greater than equal to zero. The value of the dependent variable is 0 when the growth of the GDP of a country for a particular year is negative.
b. Converting variables into proper form: After reading the dataset, we observed that the variable Time was loaded as a float type variable. We have converted the variable to an object type for further analysis.

c. Scaling the independent variables: The process of scaling the independent variables is also known as the process of normalization. This step reduces the problem of certain data being outliers or pretty small or huge compared to other values in the same column. The data is converted into values that belong to a specific range. The maximum value of this range and the minimum value of this range is decided automatically if it is not pre-defined. This process is important since it reduces the chance of

introduction of anomalies in the data due to the presence of outliers which significantly reduces the performance of the model.

d. Segregating the dataset into Training set and Test set: The output of the scaled or the normalized process is put through the input of the segregation process where the data is split into training dataset and test dataset. The normal method considers 80% of the data to be a part of the training set. This data is used for building a model using a defined algorithm. The rest of the 20% data is a part of the test set. This data is used for comparing the results of prediction that is generated from the trained model and the actual model result.

4 Chapter 4: Result and findings

4.1 4.1 Chapter overview

Tool and technology

In the research analysis it would be using a logistic regression model and then understand the significance of each of them in establishing the null hypothesis which considers that the independent variables in the dataset are significant in the prediction of the positive or negative growth of GDP of a country. In this machine learning technique, the forecasting of the values of the output variable is done based on logic transformation calculations. The model will have different independent variables which would represent the categories of the different economic indicators whose pattern has to be predicted. The dependent variable will yield a result of 1 or 0. A value of 0 would indicate that the GDP growth for a particular country in a year is negative and a value of 1 indicates that the GDP growth for a particular country in a year is positive.

For conducting this analysis, we would be using Python. In python we would be using different libraries such as Matplotlib, seaborn, pandas, and NumPy to read and visualise the values in the dataset. This would help us to identify the trend of a particular parameter of a country as compared to other countries. We would be comparing a particular indicator for different countries. The dataset has to be scaled such that there is no presence of bias or anomaly in the dataset. The dataset would be split into training set and test set. The training set will have 80% of the data and the rest 20% would be assigned to the test set. The comparison of the test values and the predicted values of the trained model will help us to identify the performance of the model.

It would be visualizing the trends of the economic indicators for all countries in the last decade using bar plot, scatter plot and the line plot. Also, we would be performing an intra-country analysis to understand the relationship of

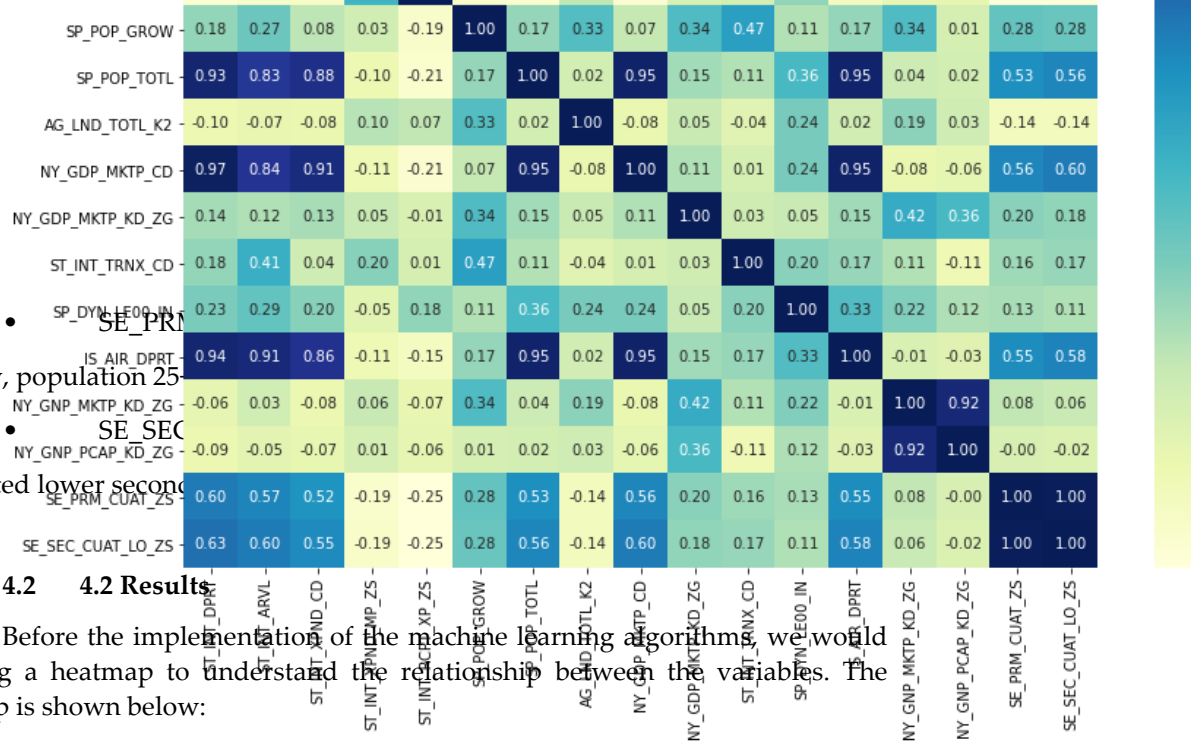
certain indicators with the GDP of the country. This will enable us to understand whether they have a direct or indirect relationship with the country's GDP.

Dataset Explanation

The dataset that is going to be used for the analysis has various fields. The name and the description of the fields are as follows:

- Country Name: Name of the country
- Country Code: Code of the name of the country
- Time: Year of analysis
- Timecode: Code for the year of analysis
- ST_INT_DPRT: International tourism, number of departures
- ST_INT_ARVL: International tourism, number of arrivals
- ST_INT_XPND_CD: International tourism, expenditures (current US$)
- ST_INT_XPND_MP_ZS: International tourism, expenditures (% of total imports)
- ST_INT_RCPT_XP_ZS: International tourism, receipts (% of total exports)
- SP_POP_GROW: Population growth (annual %)
- SP_POP_TOTL: Total Population of the country
- AG_LND_TOTL_K2: Land area (sq_ km)
- NY_GDP_MKTP_CD: GDP (current US$)
- NY_GDP_MKTP_KD_ZG: GDP growth (annual %)
- ST_INT_TRNX_CD: International tourism, expenditures for passenger transport items (current US$)
- SP_DYN_LE00_IN: Life expectancy at birth, total (years)
- IS_AIR_DPRT: Air transport, registered carrier departures worldwide
- NY_GNP_MKTP_KD_ZG: GNI growth (annual %)
- NY_GNP_PCAP_KD_ZG: GNI per capita growth (annual %)

- SE_PRM
primary, population 25-
- SE_SEC
completed lower secon



4.2 4.2 Results

Before the implementation of the machine learning algorithms, we would be using a heatmap to understand the relationship between the variables. The heatmap is shown below:

The boxes which have a deeper shade signifies a strong positive correlation and the lighter shade boxes signifies a weak positive or weak negative correlation between the variables.

The algorithm which we would be training would use the Logistic Regression method. The output for accuracy, confusion matrix, r-squared value, mean square error, and the root mean square error is given below:

Accuracy: 0.8571428571428571
Confusion matrix:
[[1 4]
[2 35]]
MSE equals-> 0.14285714285714285
RMSE equals 0.3779644730092272
R-squared equals-> -0.362162162162162

Other Models include

- Random Forest

Accuracy: 0.8095238095238095
Confusion matrix:
[[3 2]
[6 31]]
MSE equals-> 0.19047619047619047
RMSE equals 0.4364357804719847
R-squared equals-> -0.8162162162162161

- K Neighbours Classifier

Accuracy: 0.7619047619047619
Confusion matrix:
[[0 5]
 [5 32]]
MSE equals-> 0.23809523809523808
RMSE equals 0.4879500364742666
R-squared equals-> -1.2702702702702702

- Gaussian NB Classifier

accuracy is 52.88 %
std is 16.33 %

Since, Logistic Regression gives us the best result s, it is what we conside r going forward.

The accuracy output indicates that the independent variables of the model have a capacity of predicting the output of the dependent variable with an efficiency of 85.71%. The values of the root mean square error and the mean absolute error which gives the estimated difference between the actual values and the values of the predicted model should be as low as possible. The degree of fit or the explanation of the difference in proportion of the dependent variable by the independent variable which is signified by the r-squared value should be as high as possible for a model to be more efficient.

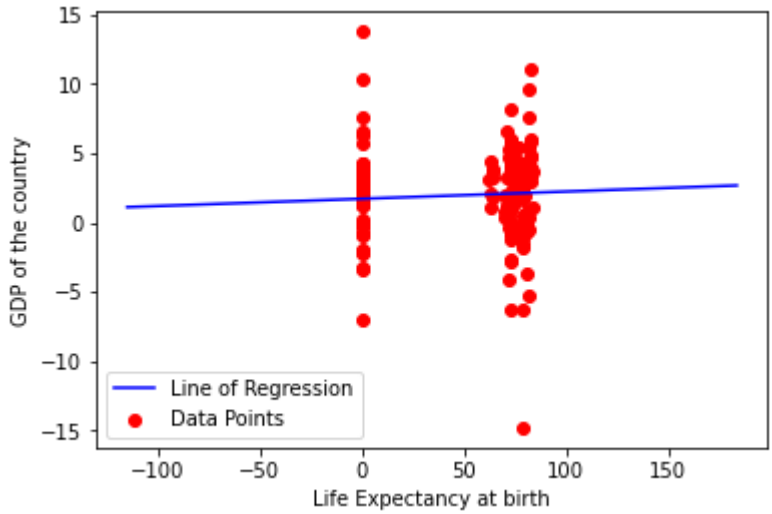


Figure 4.1: Result 1

(Source: Jupyter notebook)

From the above figure, it can be concluded that there exists a direct relationship between the life expectancy at birth and gap of the country. The line regression has calculated the estimation of the regression coefficient value based on the predicted wrights. This has identified the data points of the gap of the country by implementing the line of regression method.

The confusion matrix is used for the calculation of Accuracy, Precision, Recall (TPR), F1-Score, False positive rate, False negative rate, Sensitivity, and Specificity of a model. From the above confusion matrix, we can conclude that the True Negative is 2, and the True Positive is 33. The Type 1 error or the False positive is equal to 6 and the Type 2 error or the False Negative is equal to 3. The significance of the independent variables in the prediction of the dependent variable is done using the OLS Regression whose output is given below:

OLS Regression Results

==========

Dep. Variable: op R-squared: 0.254

Model:	OLS	Adj. R-squared:	0.145
Method:	Least Squares	F-statistic:	2.326
Date:	Wed, 31 Aug 2022	Prob (F-statistic):	0.00545
Time:	11:49:18	Log-Likelihood:	-54.147
No. Observations:	126	AIC:	142.3
Df Residuals:	109	BIC:	190.5
Df Model:	16		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.2460	0.288	4.328	0.000	0.675	1.817
x1	0.1111	0.057	1.939	0.055	-0.002	0.225
x2	0.1897	0.908	0.209	0.835	-1.609	1.988
x3	-0.1911	0.402	-0.476	0.635	-0.987	0.605
x4	0.1740	1.017	0.171	0.865	-1.842	2.190
x5	-0.0202	0.090	-0.224	0.823	-0.198	0.158
x6	0.0979	0.080	1.229	0.222	-0.060	0.256
x7	0.4053	0.241	1.684	0.095	-0.072	0.882
x8	-0.1686	0.415	-0.406	0.686	-0.992	0.655
x9	0.0442	0.128	0.344	0.732	-0.210	0.299
x10	-0.0696	0.302	-0.230	0.818	-0.668	0.529
x11	-0.1095	0.060	-1.818	0.072	-0.229	0.010
x12	0.2667	0.525	0.508	0.612	-0.773	1.306
x13	0.2984	1.113	0.268	0.789	-1.908	2.505
x14	0.1723	1.090	0.158	0.875	-1.988	2.333
x15	1.0103	0.788	1.282	0.202	-0.551	2.572
x16	-1.1086	0.874	-1.268	0.207	-2.841	0.624

Omnibus:	11.266	Durbin-Watson:	2.126
Prob(Omnibus):	0.004	Jarque-Bera (JB):	12.181
Skew:	-0.734	Prob(JB):	0.00226
Kurtosis:	2.590	Cond. No.	142.

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

To understand the significance of every independent variable, we need to analyse the value of p for each of them. The below table shows the meaning of each corresponding variable (Sharma, 2020):

Variable in regression output	**Name of the independent variable**
x1	Time
x2	ST_INT_DPRT
x3	ST_INT_ARVL
x4	ST_INT_XPND_CD
x5	ST_INT_XPND_MP_ZS
x6	ST_INT_RCPT_XP_ZS
x7	SP_POP_GROW
x8	SP_POP_TOTL
x9	AG_LND_TOTL_K2
x10	ST_INT_TRNX_CD
x11	SP_DYN_LE00_IN
x12	IS_AIR_DPRT
x13	NY_GNP_MKTP_KD_ZG
x14	NY_GNP_PCAP_KD_ZG
x15	SE_PRM_CUAT_ZS
x16	SE_SEC_CUAT_LO_ZS

Figure 4.2: Result 2

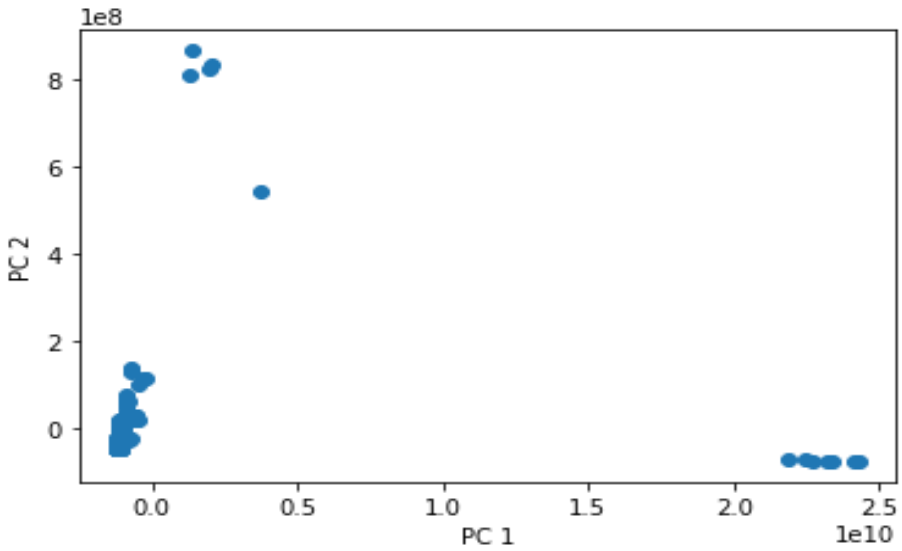
(Source: Jupyter notebook)

The threshold value of p for an independent variable to be statistically significant is equal to 0.05. The variables x11 (SP_DYN_LE00_IN) and x16 (SE_SEC_CUAT_LO_ZS

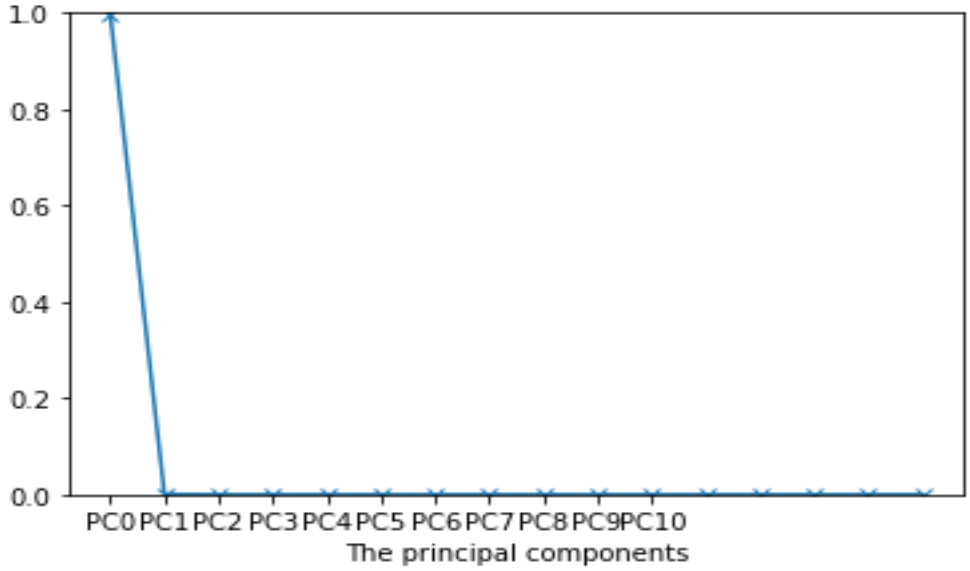
) Have a p-value which is less than 0.05. We can conclude from the above output that the Total Life expectancy at birth in years and educational attainment, at least completed lower secondary, population 25+, total (%) (cumulative) are the variables which are statistically significant in forecasting the value of a positive or negative GDP growth of a country. As most of the independent variables have a value of p which is greater than 0.05, we cannot reject the null hypothesis.

PCA Implementation

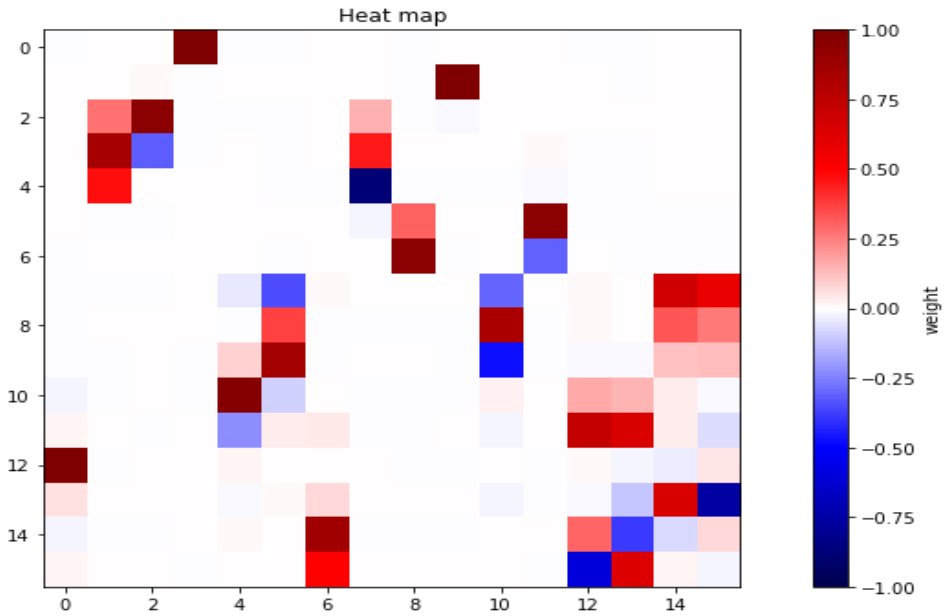
In this analysis, we are going to implement the Principal Component Analysis clustering method. The clustering figure is given below:



It is evident that there are 3 number of clusters. The principal component visualisation is shown below:



In the next visualisation, we have checked the distribution of the data points using a heat map:



From the above figure, we can conclude that the values are not equally distributed. The Principal component explanation variance ratio is given below:

[9.99128535e-01 8.71236790e-04 2.01056401e-07 2.17342423e-08

5.69192305e-09 3.24259570e-12 1.21836153e-12 4.25197737e-17
3.90178486e-17 2.10176562e-17 1.05887723e-18 4.85202849e-19
2.09390187e-19 1.23758922e-19 3.05435717e-20 2.55602507e-21]

It can be concluded from the above output that the 1st component of PCA explains 99.12% of the total variation in the total dataset.

Now, we would be applying visualizations and observe the descriptive statistics which would help us to understand about the intra-country indicator relationship and the trend of the indicators for each country

BAHRAIN

The descriptive statistics of Bahrain indicates that the average Number of departures in International tourism is lower than that of Number of arrivals in International tourism. The line plot visualisation for checking International tourism, receipts (% of total exports) in Bahrain is given below:

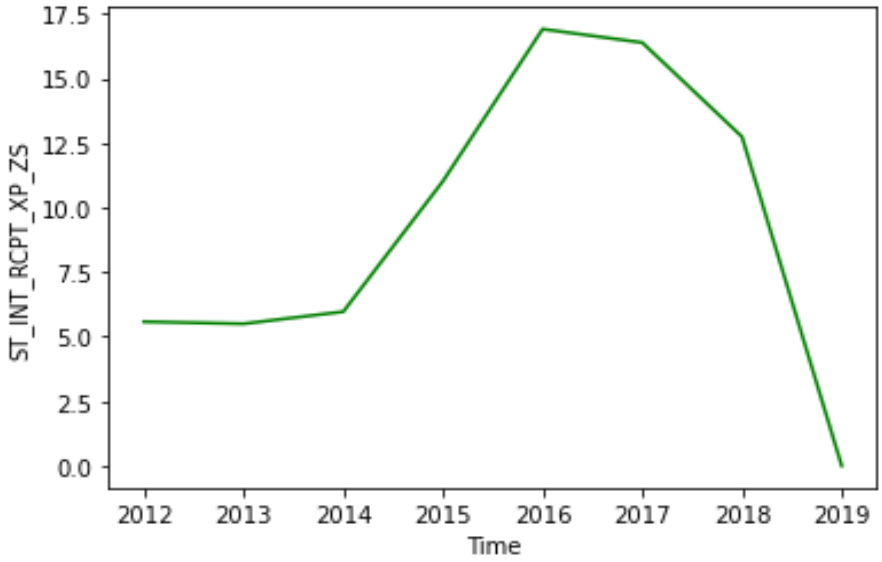


Figure 4.3: Result 3

(Source: Jupyter notebook)

From the above output, we can understand that international tourism receipts had reached a peak between 2016 and 2017 and after that it has gradually decreased in Bahrain. The values of 2019 are not given.

The next plot below shows the relationship between International tourism, number of arrivals and GDP (current US$) in Bahrain:

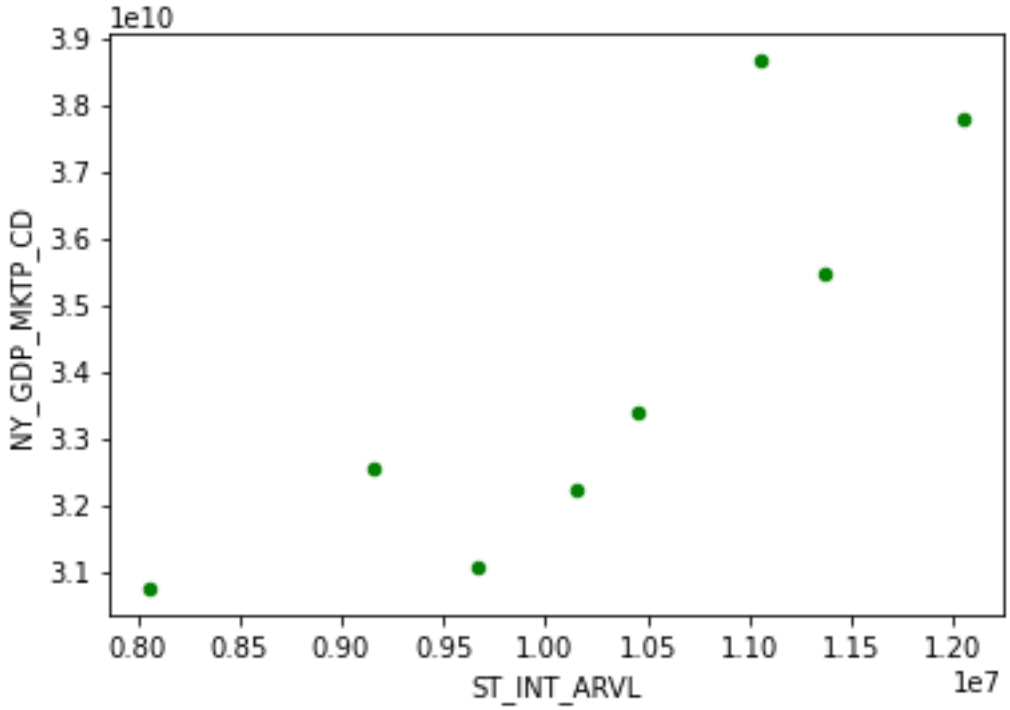


Figure 4.4: Result 4

(Source: Jupyter notebook)

From the above scatter plot, we can conclude that the relationship between the GDP of Bahrain and the number of arrivals in international tourism have a direct relationship which means that with the increase in arrivals in international tourism, there is an increase in the GDP value of Bahrain.

BARBADOS

The descriptive statistics of Barbados indicates that the average value of Expenditures in US$ for International tourism for Barbados is lower than that of Bahrain. The line plot visualisation for checking International tourism, expenditures (current US$) in Barbados is given below:

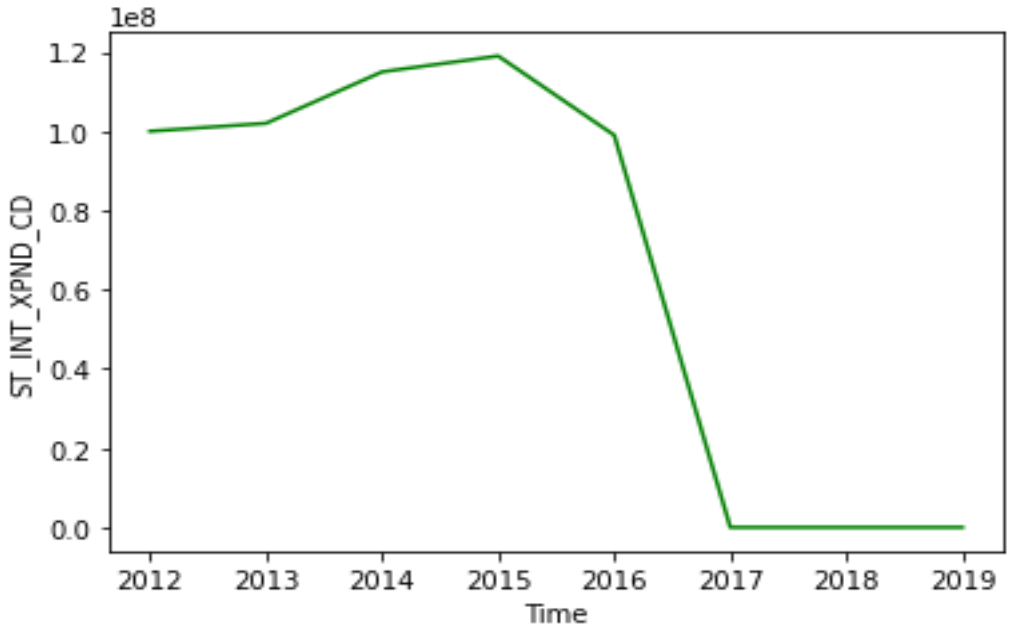


Figure 4.5: Result 5

(Source: Jupyter notebook)

The above figure indicates that the peak value of Expenditures in US$ for International tourism for Barbados happened in the year 2015 which decreased in the year 2016. The next below plot shows the relationship between Life expectancy and GDP (current US$) in Barbados:

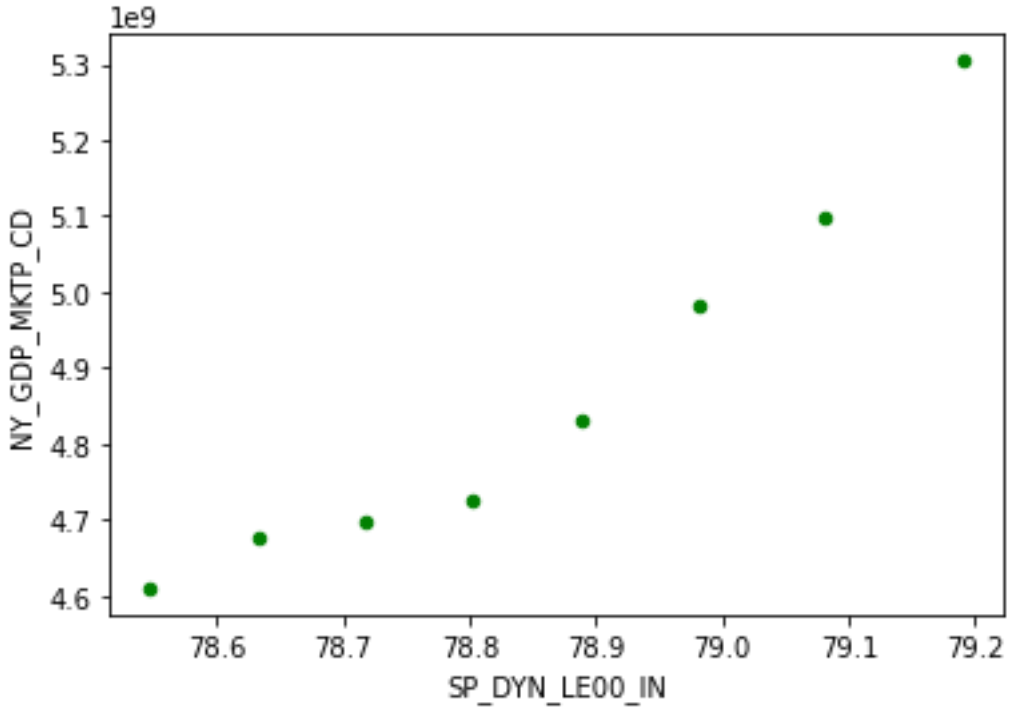


Figure 4.6: Result 6

(Source: Jupyter notebook)

The above visualization indicates that the Total Life expectancy at birth given in years and the GDP of Barbados in US$ has a direct relationship.

BERMUDA

The descriptive statistics of Bermuda indicates that the average Expenditures as a % of total imports for International tourism for Bermuda is higher than Bahrain and Barbados. The bar plot visualisation below is for checking International tourism, expenditures (% of total imports) in Bermuda:

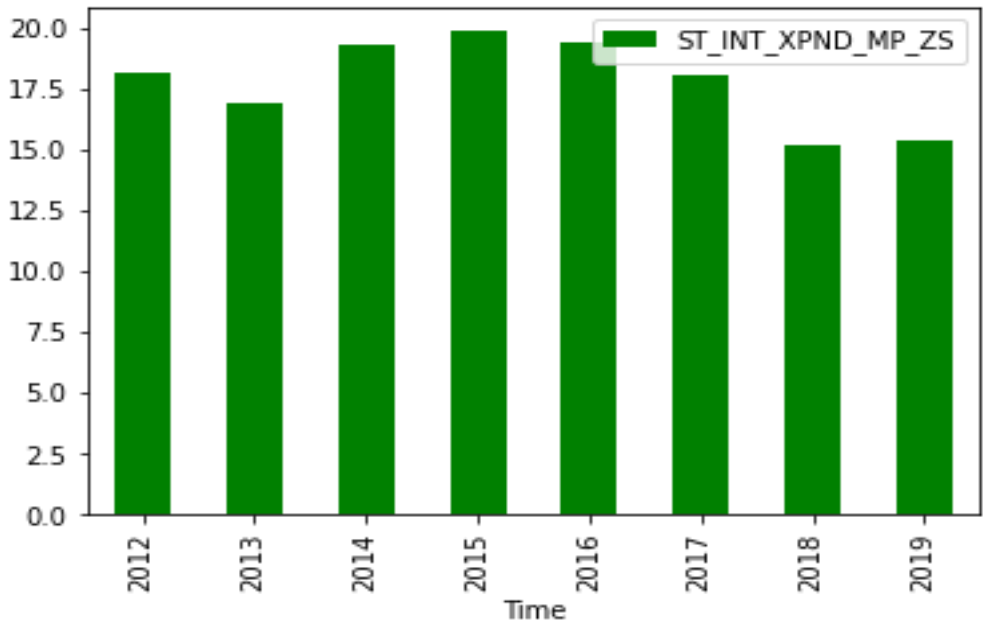


Figure 4.7: Result 7

(Source: Jupyter notebook)

The above visualisation indicates that the value of the Expenditures as a % of total imports for International tourism for Bermuda had reached a peak in the year 2015 and it started to decrease from that time. The lowest value of the Expenditures as a % of total imports for International tourism in Bermuda was recorded in 2019 for the last decade.

The next below plot shows the relationship between International tourism, receipts (% of total exports) and GDP (current US$) in Bermuda:

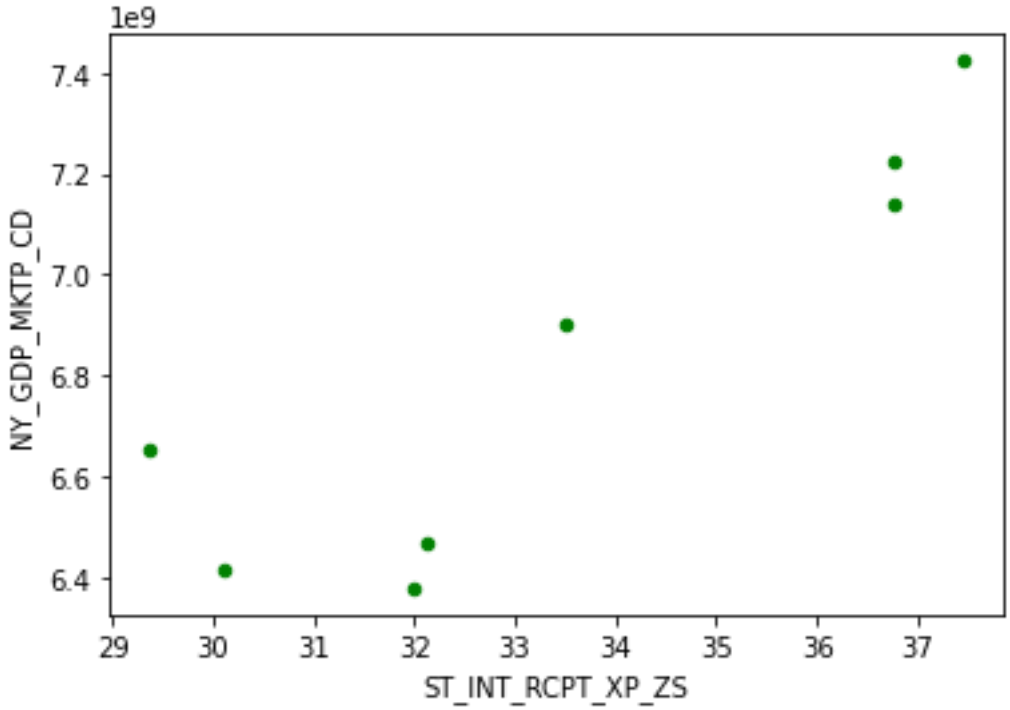


Figure 4.8: Result 8

(Source: Jupyter notebook)

The above picture indicates that there is a direct relationship between the GDP of Bermuda in US$ and the Receipts as 8% of total exports for International tourism.

COMOROS

The descriptive statistics of Comoros indicates that the average Expenditures for passenger transport items in US$ for International tourism for Comoros is lower than Bahrain and Bermuda. The below bar plot is for checking International tourism, expenditures for passenger transport items (current US$) in Comoros:

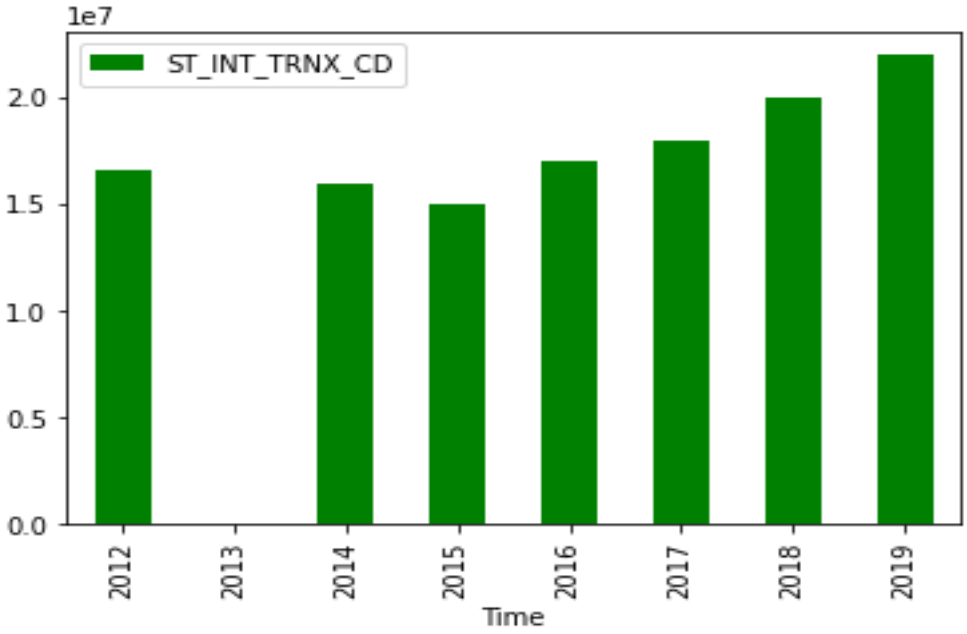


Figure 4.9: Result 9

(Source: Jupyter notebook)

From the above figure, it can be observed that the value of Expenditures for passenger transport items in US$ for International tourism for Comoros has increased from the year 2015 to 2019 after seeing a decrease in the year 2015 as compared to 2014. The value for the year 2013 is not given.

The next below plot shows the relationship between life expectancy and GDP (current US$) in Comoros:

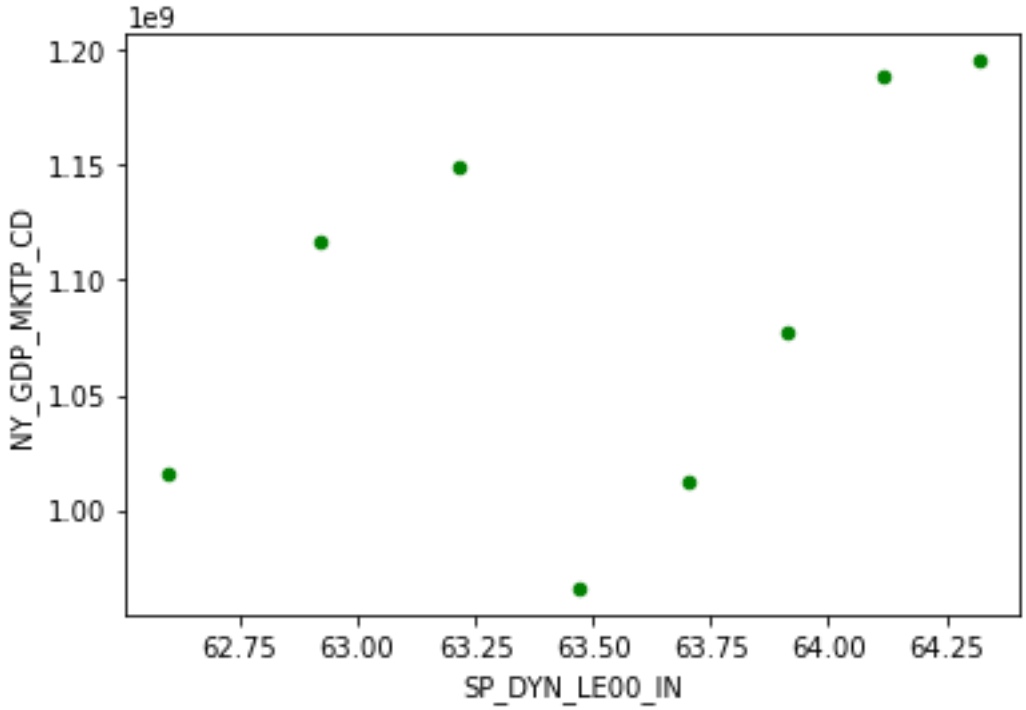


Figure 4.10: Result 10

(Source: Jupyter notebook)

The above scatter plot which gives a relationship between the GDP of Comoros in US$ and the total population suggests that for a certain range of population, the country has a direct relationship with the GDP and for another specific population range, the relationship with country's GDP is indirect.

MALTA

The descriptive statistics of Malta indicates that the average Population % completed at least completed secondary education aged 25+ for Malta is higher than Bahrain. The below bar plot visualisation is for checking educational attainment, at least completed lower secondary, population 25+, total (%) (cumulative) in Malta:

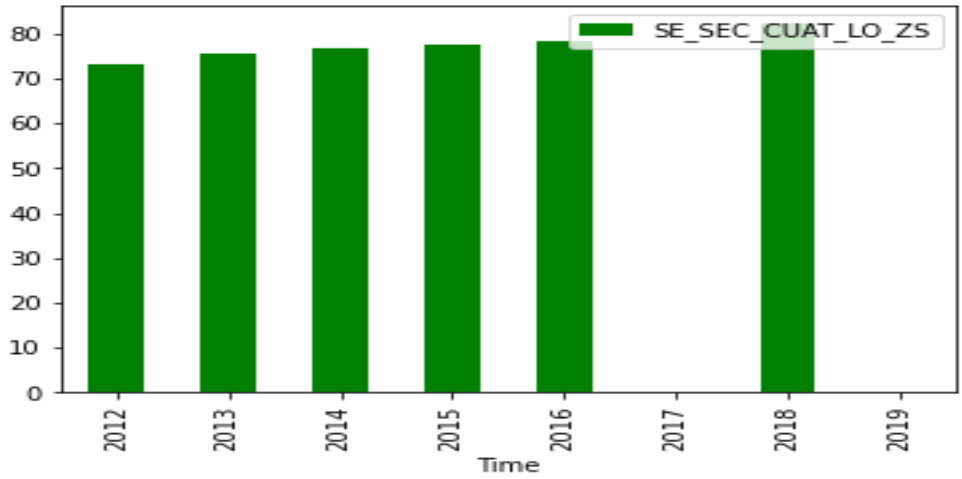


Figure 4.11: Result 11

(Source: Jupyter notebook)

From the above figure, it can understand that in the last decade, the Population % completed at least completed secondary education aged 25+ has increased in Malta. The peak has been reached in the year 2018. The values for the year 2017 and 2019 are not given.

The next below plot shows the relationship between International tourism, number of departures and GDP (current US$) in Malta:

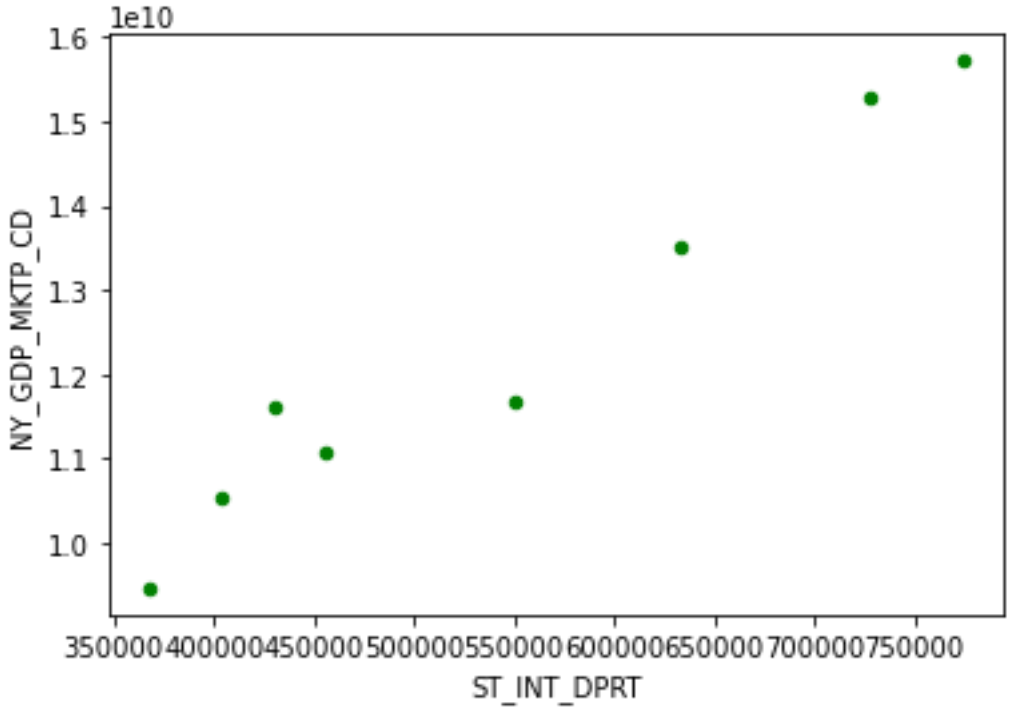


Figure 4.12: Result 12

(Source: Jupyter notebook)

From the above line plot we can conclude that there exists a direct relationship between the GDP of Malta in US$ and the Number of departures in International tourism.

MARSHALL ISLANDS

The descriptive statistics of Marshall Islands indicates that the average GDP of the country in US$ is lower than Comoros and Malta. The below scatter plot visualisation is for checking GDP (current US$) in Marshall Islands:

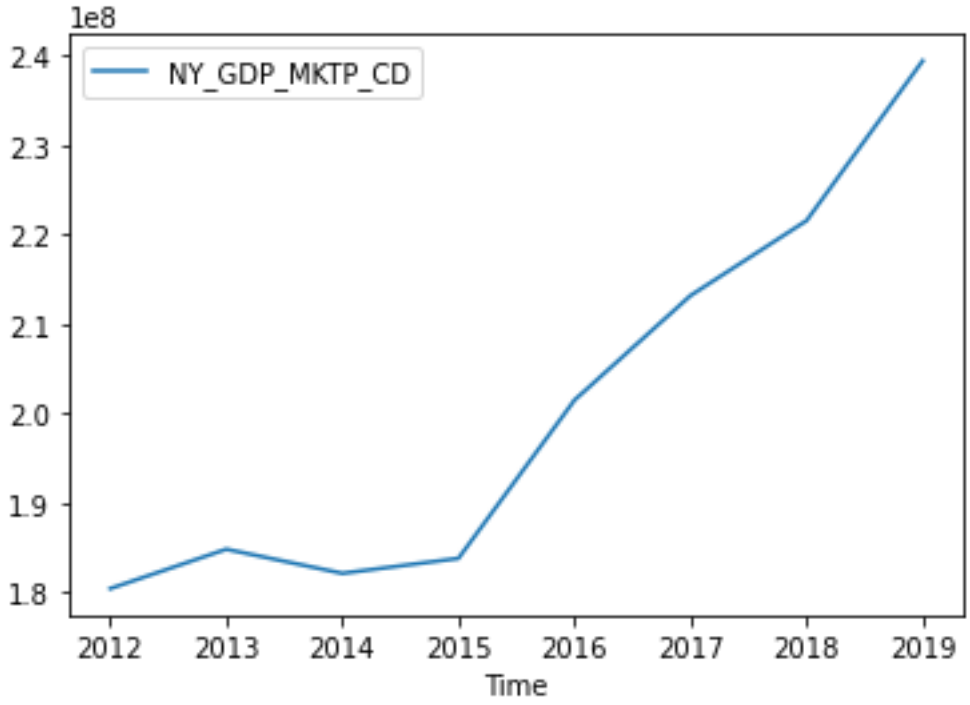


Figure 4.13: Result 13

(Source: Jupyter notebook)

It can observe from the above picture, that the GDP of Marshall Islands in US$ had increased very less from 2012 to 2015 but after the year 2015, the GDP has increased significantly and the peak was observed in the year 2019.

The next below plot shows the relationship between International tourism, expenditures for passenger transport items (current US$) and GDP (current US$) in Marshall Islands:

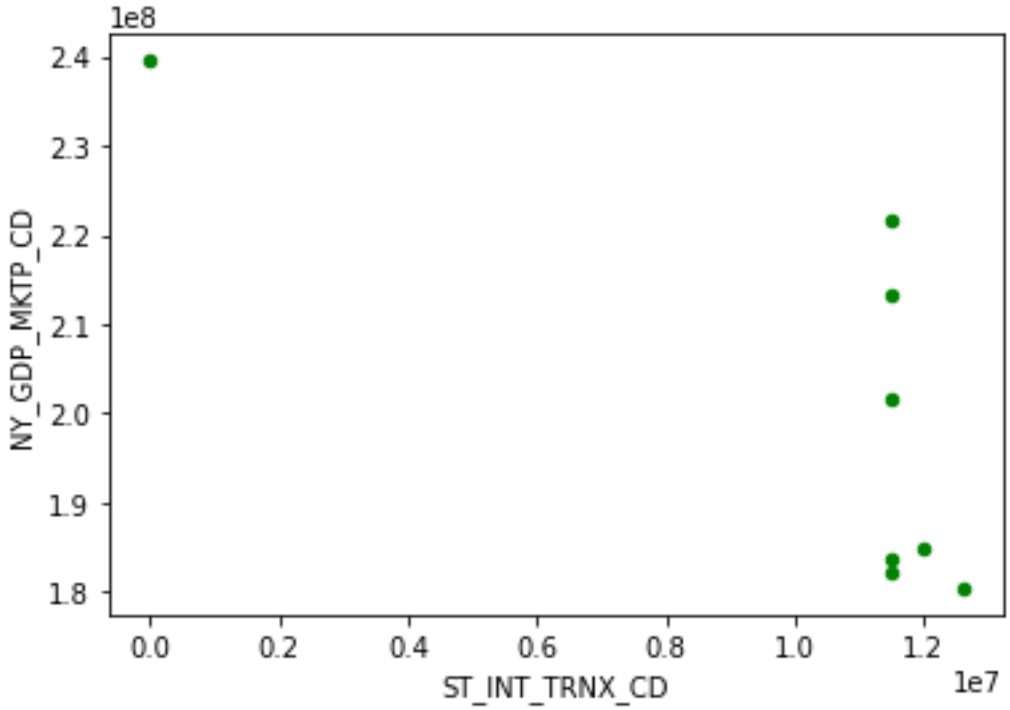


Figure 4.14: Result 14

(Source: Jupyter notebook)

From the above picture, we can understand that the relationship between the GDP of Marshall Islands in US$ is indirectly related with the Expenditures for passenger transport items in US$ for International tourism.

MAURITIUS

The descriptive statistics of Mauritius indicates that the average Number of departures in International tourism is higher than Marshall Islands but lower than Malta. The below line plot visualisation is for checking International tourism, number of departures in Mauritius:

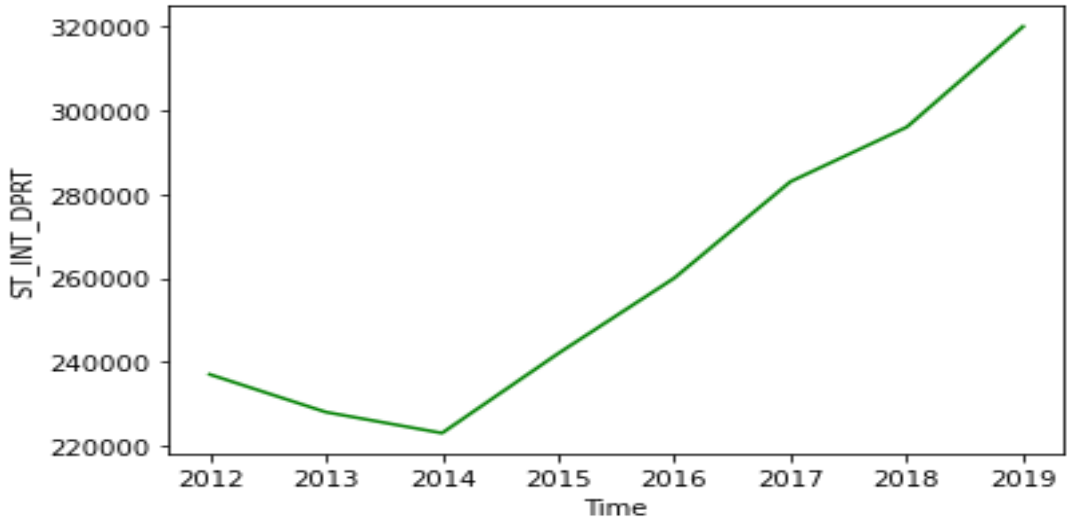


Figure 4.15: Result 15

(Source: Jupyter notebook)

The above visualisation makes it evident that the value of Number of departures in International tourism for Mauritius had decreased from 2012 to 2014 but after the year 2014, there has been an exponential increase in its value.

The next below plot shows the relationship between Air transport, registered carrier departures worldwide and GDP (current US$) in Mauritius:

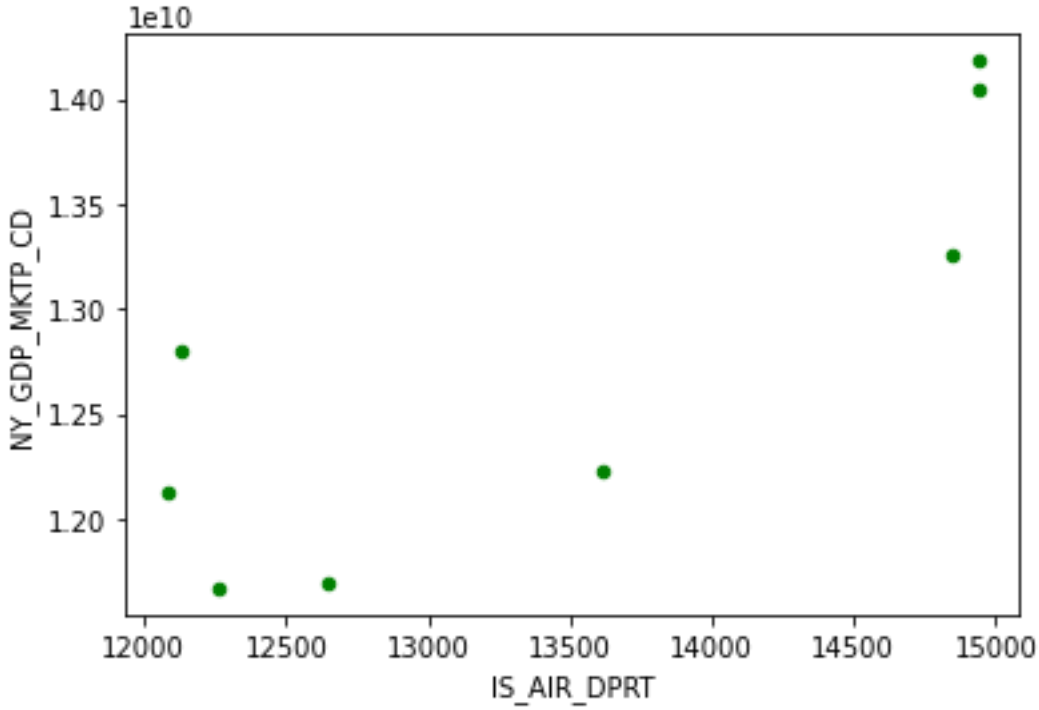


Figure 4.16: Result 16

(Source: Jupyter notebook)

The above figure indicates that the relationship between the GDP of Mauritius in US$ has a direct relationship with Expenditures for passenger transport items in US$ for International tourism after a value of 12600 for the later one.

SAMOA

The descriptive statistics of Samoa indicates that the average Number of arrivals in International tourism is lower than Mauritius and Malta but higher than Marshall Island. The bar plot below is for checking International tourism, number of arrivals in Samoa:

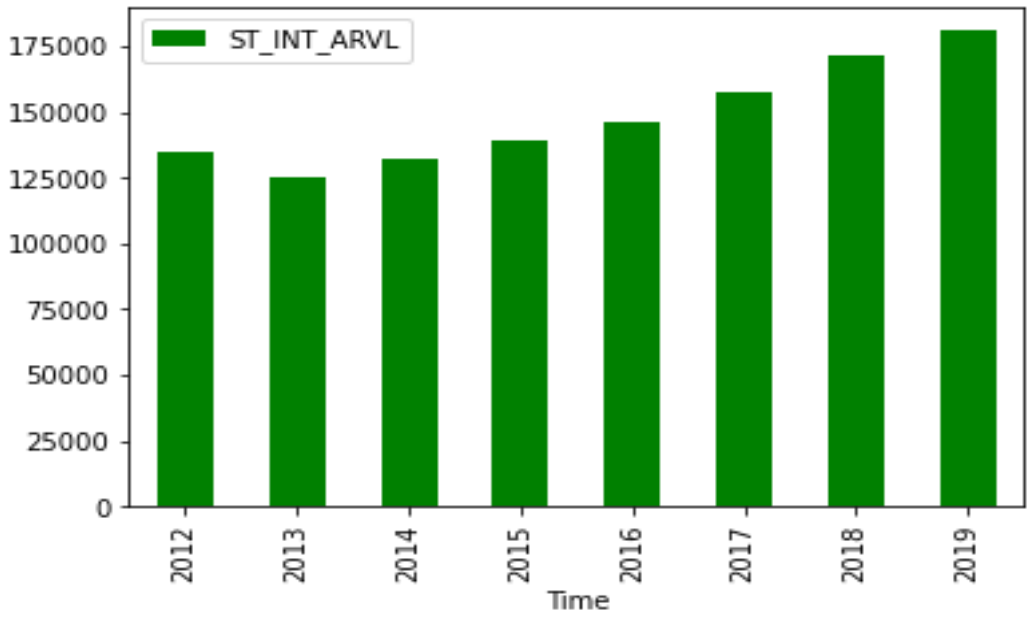


Figure 4.17: Result 17

(Source: Jupyter notebook)

The figure makes it evident that the Number of arrivals in International tourism had decreased from 2012 to 2013 but after that, it has increased significantly and the peak is in the year 2019.

The next below plot shows the relationship between population growth and GDP (current US$) in Samoa

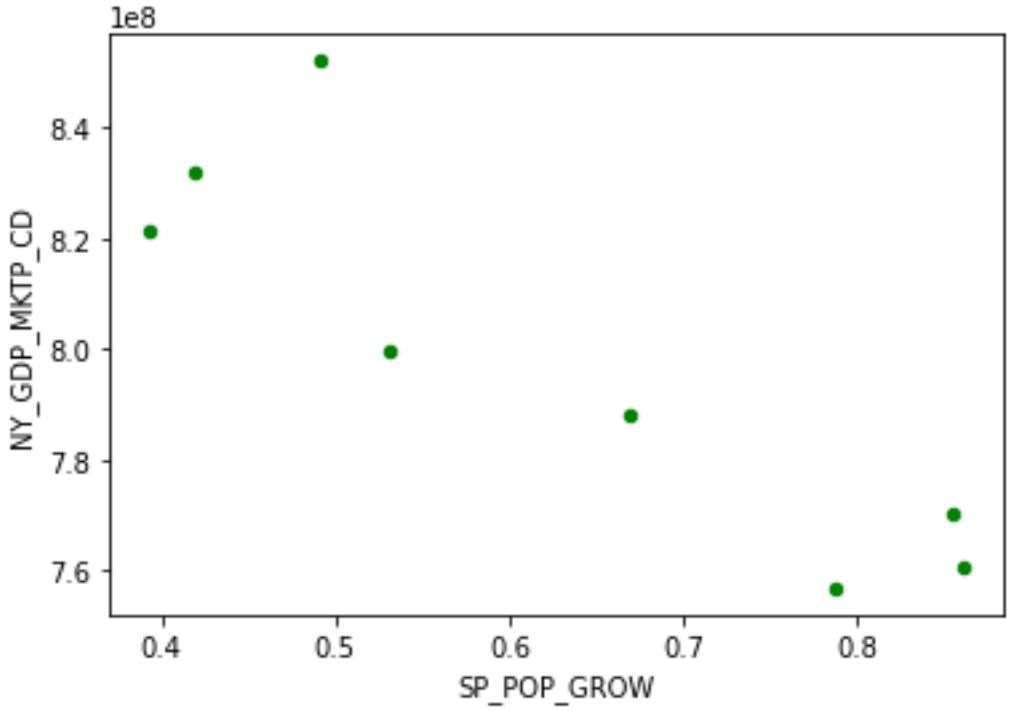


Figure 4.18: Result 18

(Source: Jupyter notebook)

The above picture indicates that there is an indirect relationship between the Registered carrier departures of air transport worldwide and the GDP of Samoa in US$.

CAYMAN ISLANDS

The descriptive statistics of Cayman Islands indicates that the average GDP growth is higher than Samoa but lower than Mauritius. The below bar plot is for checking the GDP growth of Cayman Islands:

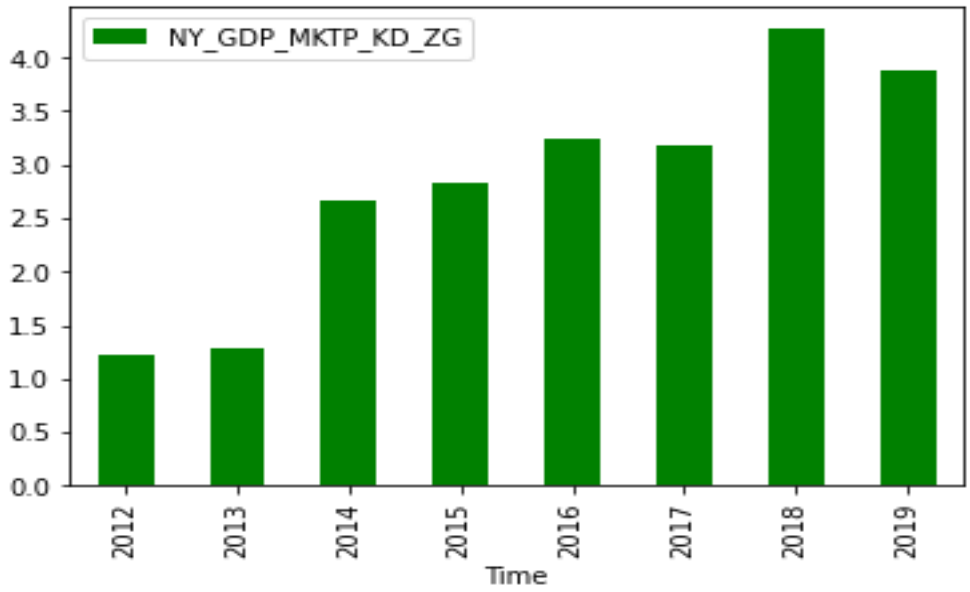


Figure 4.19: Result 19

(Source: Jupyter notebook)

The above visualization indicates that the GDP growth of Cayman Islands in the last decade has been positive. The peak value of GDP growth was observed in the year 2018.

The next below plot shows the relationship between International tourism, number of arrivals and GDP (current US$) in Cayman Islands:

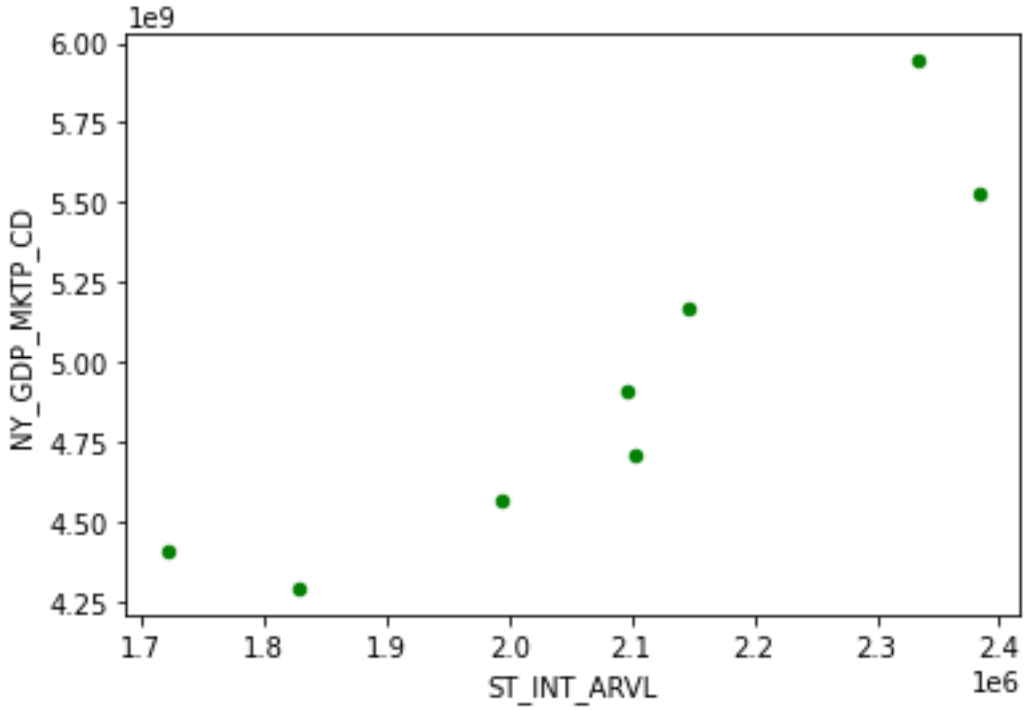


Figure 4.20: Result 20

(Source: Jupyter notebook)

It can be concluded from the above visualisation that the relationship between the Number of arrivals in International tourism and the GDP of Cayman Islands is direct in nature.

TONGA

The descriptive statistics of Tonga indicates that the average Expenditures in US$ for International tourism is lower than Cayman Islands but higher than Samoa. The below bar plot is for checking International tourism, expenditures (current US$) of Tonga:

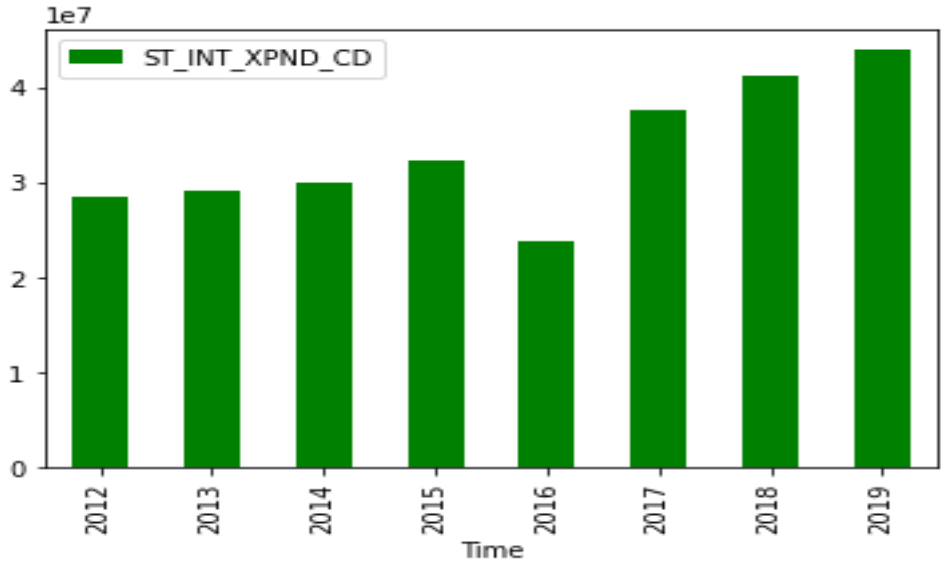


Figure 4.21: Result 21

(Source: Jupyter notebook)

From the above picture, we can understand that the value of Expenditures of Tonga in US$ for International tourism has increased significantly from the year 2016 to 2019. The maximum value was observed in the year 2019.

The next below plot shows the relationship between International tourism, expenditures (current US$) and GDP (current US$) in Tonga:

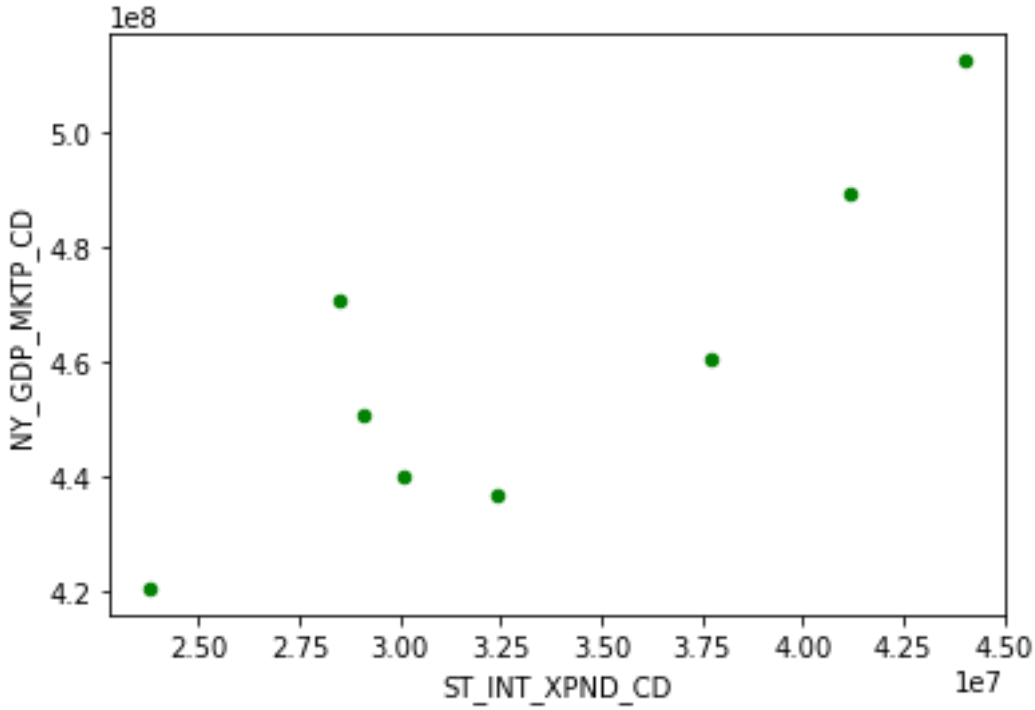


Figure 4.22: Result 22

(Source: Jupyter notebook)

The above figure indicates that the relationship between the GDP of Tonga in US$ and the Number of arrivals in International tourism is direct in nature.

TUVALU

The descriptive statistics of Tuvalu indicates that the average Number of arrivals in International tourism is lower than Tonga and Cayman Islands. The scatter plot visualisation below is for checking International tourism, number of arrivals in Tuvalu:

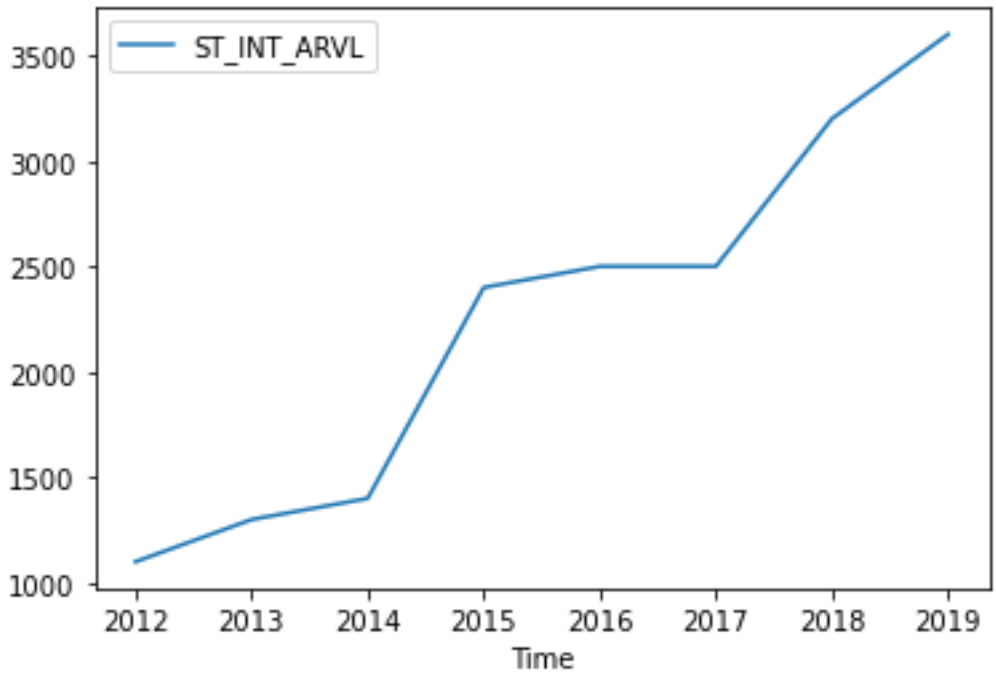


Figure 4.23: Result 23

(Source: Jupyter notebook)

The figure above shows that the Number of arrivals in international tourism for Tuvalu has increased significantly in the last decade. The maximum value was observed in the year 2019.

The next below plot shows the relationship between total population and GDP (current US$) in Tuvalu:

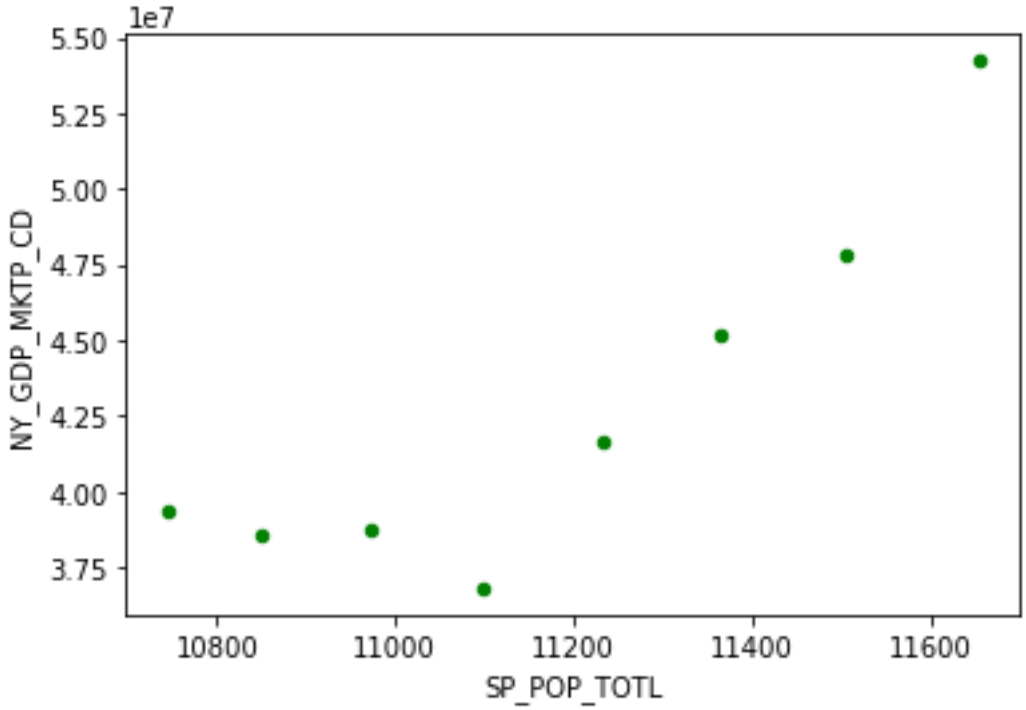


Figure 4.24: Result 24

(Source: Jupyter notebook)

The visualisation above concludes that the total population initially has an indirect relationship with the GDP of Tuvalu in US$ but after a certain value of the population, the relationship is direct in nature.

PALAU

The descriptive statistics of Palau indicates that the average Annual % growth in population is lower than Tuvalu but higher than Tonga. The below scatter plot visualisation is for checking Population growth (annual %) in Palau:

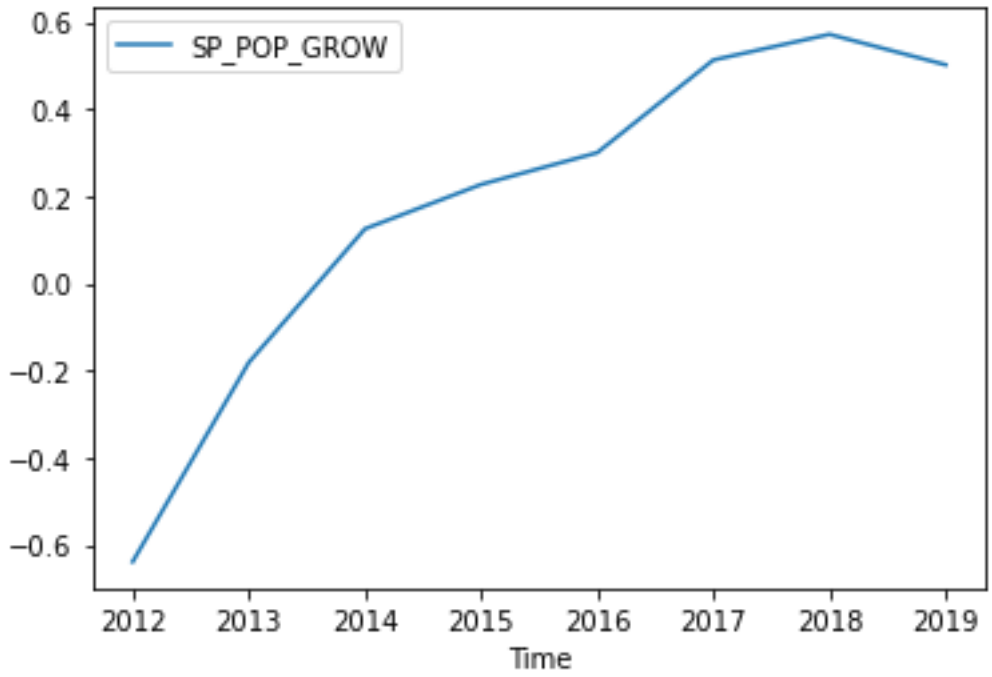


Figure 4.25: Result 25

(Source: Jupyter notebook)

From the above figure, it can observe that the population growth in Palau has increased from 2013 to 2019. The peak growth was observed in the year 2018.

The next below plot shows the relationship between total population growth and GDP (current US$) in Palau:

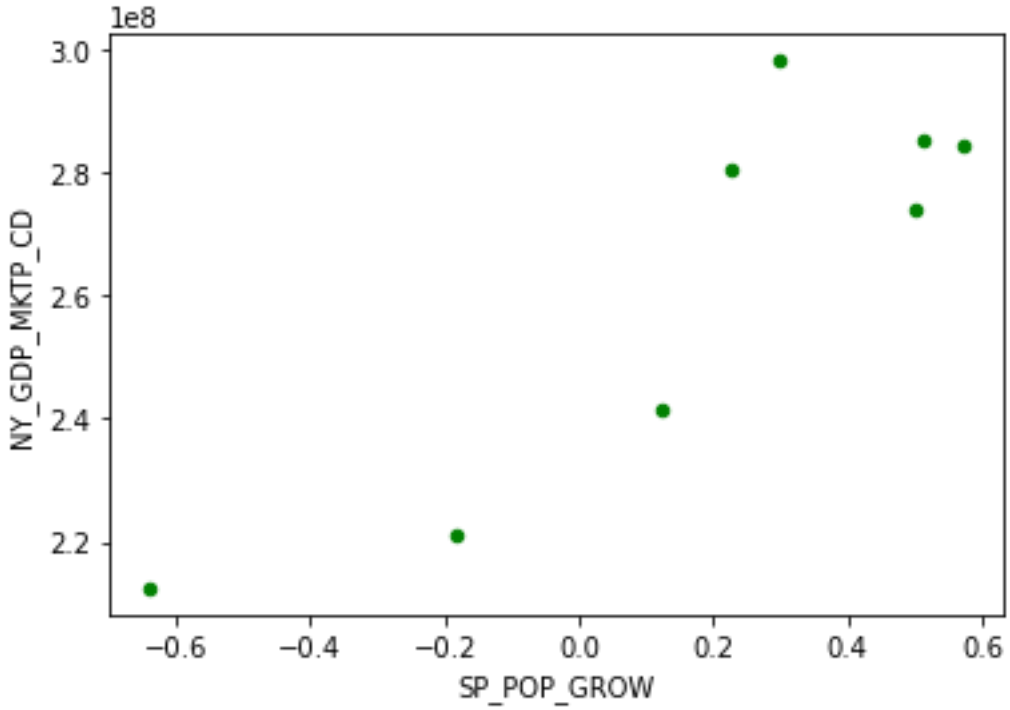


Figure 4.26: Result 26

(Source: Jupyter notebook)

The total population of Palau and the total GDP is US$ has a direct relationship till a certain range and after that the relationship between the both the mentioned variables is indirect in nature.

SINGAPORE

The descriptive statistics of Singapore indicates that the average Total Life expectancy at birth given in years is higher than Samoa and Tonga. The below scatter plot visualisation is for checking International tourism, expenditures for passenger transport items (current US$) in Singapore:

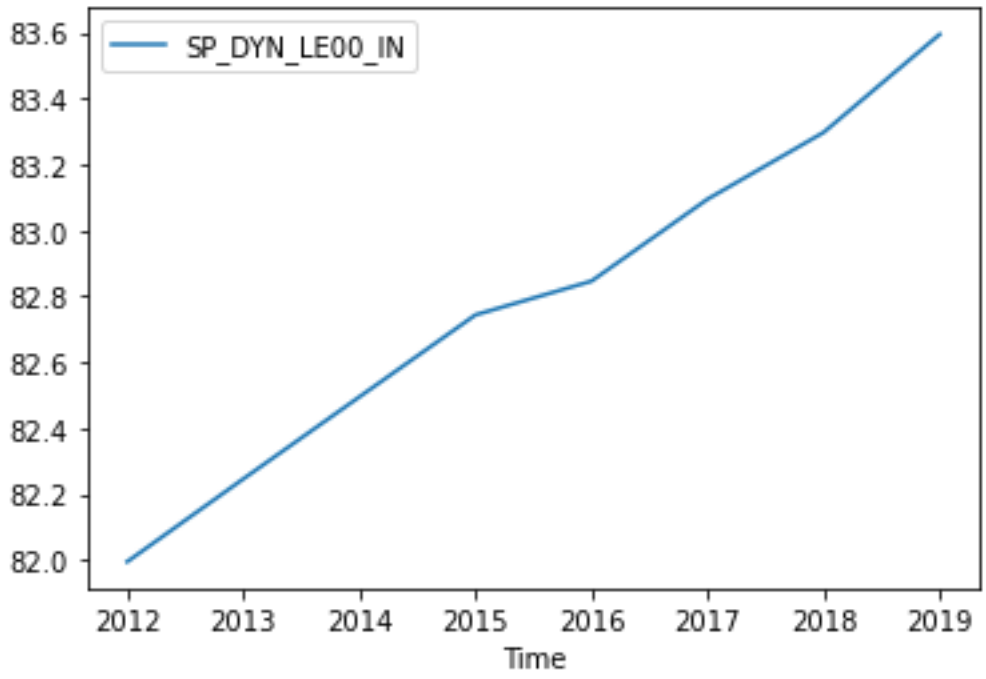


Figure 4.27: Result 27

(Source: Jupyter notebook)

It can conclude from the above visualisation that the Total Life expectancy at birth given in years for Singapore has increased significantly in the last decade. The maximum value was observed in the year 2019.

The next below plot shows the relationship between Land area (skim) and GDP (current US$) in Singapore:

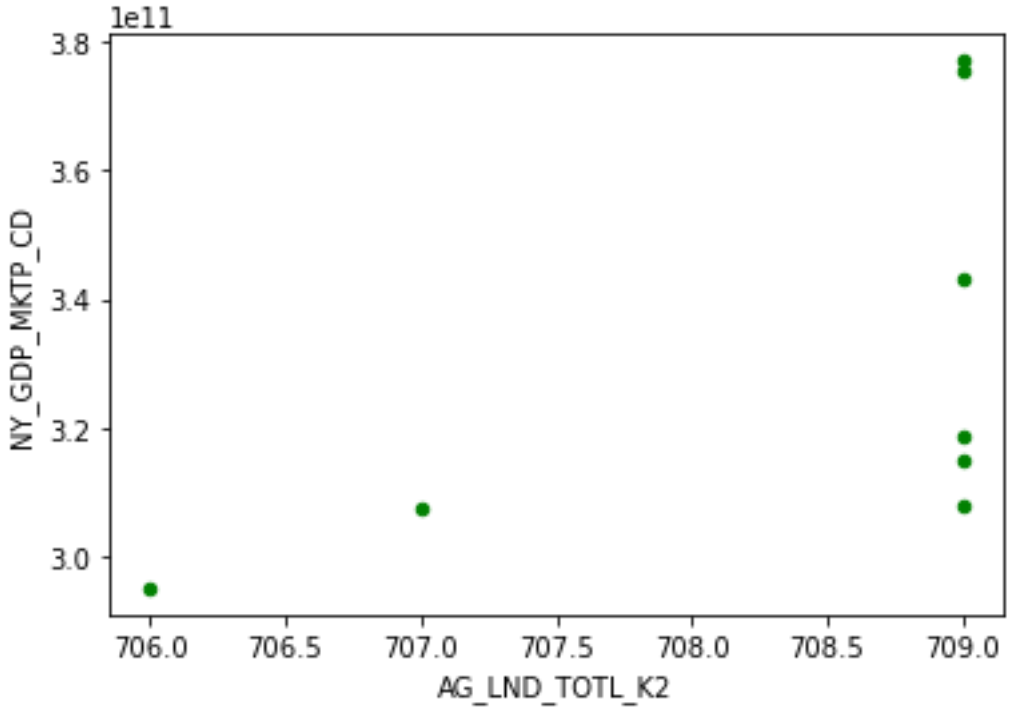


Figure 4.28: Result 28

(Source: Jupyter notebook)

From the above figure, we can conclude that there exists a direct relationship between the GDP of Singapore in US$ and the Area of Land area in sq_ km.

SOLOMON ISLANDS

The descriptive statistics of Solomon Islands indicates that the average Number of arrivals in International tourism is lower than Singapore and Palau. The below bar plot visualisation is for checking the International tourism, number of arrivals of Solomon Islands:

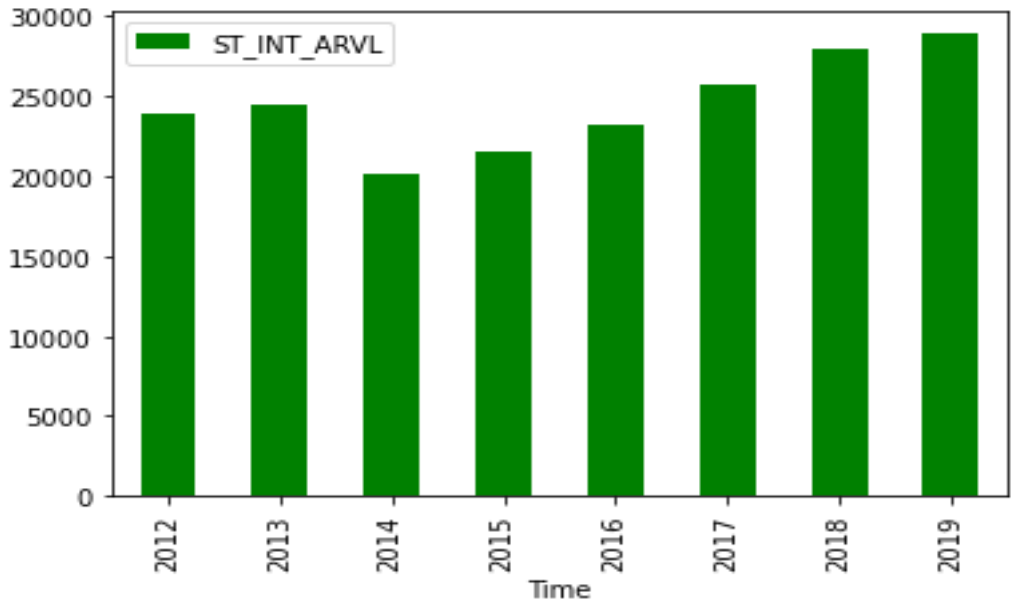


Figure 4.29: Result 29

(Source: Jupyter notebook)

From the above picture we can conclude that the Number of arrivals in international tourism has increased from the year 2014 to 2019. In the last decade, the value had seen a reduction for the first time in the year 2014.

The next below plot shows the relationship between Life expectancy at birth, total (years) and GDP (current US$) in Solomon Islands:

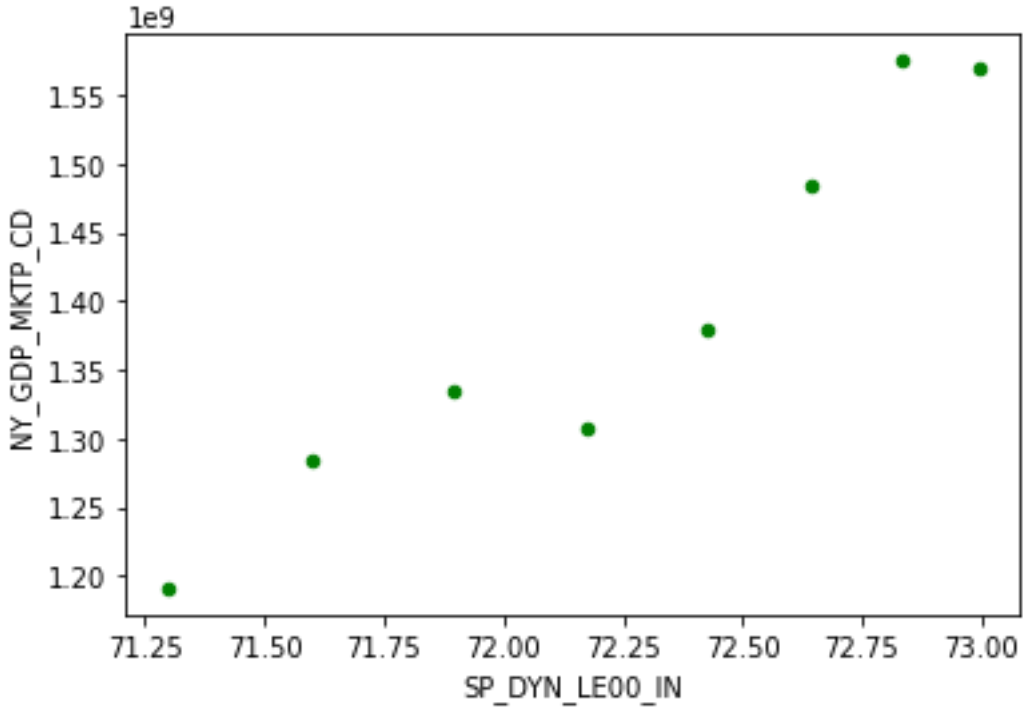


Figure 4.30: Result 30

(Source: Jupyter notebook)

It can observe from the above picture that there exists a direct relationship between the GDP of Solomon Islands in US$ and the Total Life expectancy at birth given in years.

TRINIDAD AND TOBAGO

The descriptive statistics of Trinidad and Tobago indicates that the average Annual % growth in population is lower than Singapore and Solomon Islands. The below bar plot visualisation is for checking the population growth of Trinidad and Tobago:

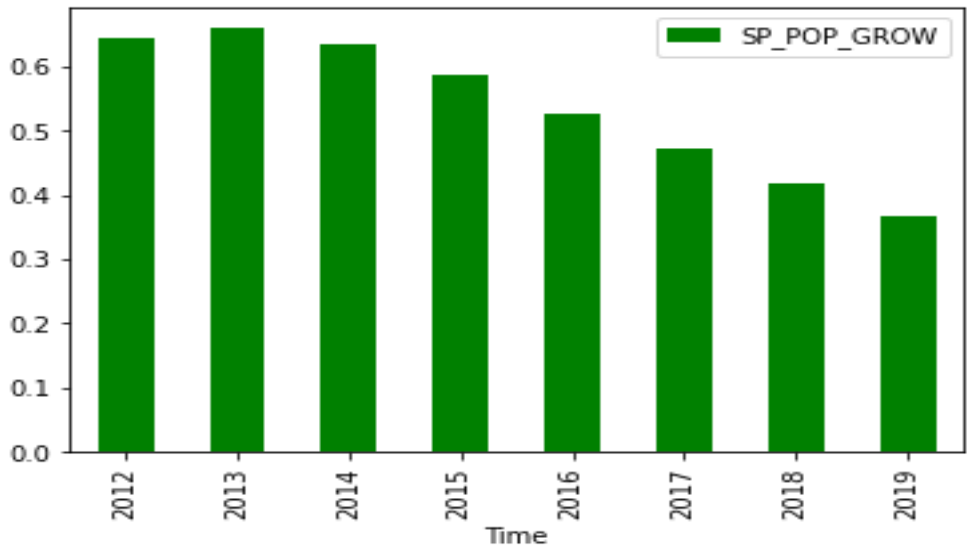


Figure 4.31: Result 31

(Source: Jupyter notebook)

The above picture indicates that the value of population growth in Trinidad and Tobago has reduced during the last decade. The peak value was observed in the year 2013.

The next below plot shows the relationship between Life expectancy at birth, total (years) and GDP (current US$) in Trinidad and Tobago.

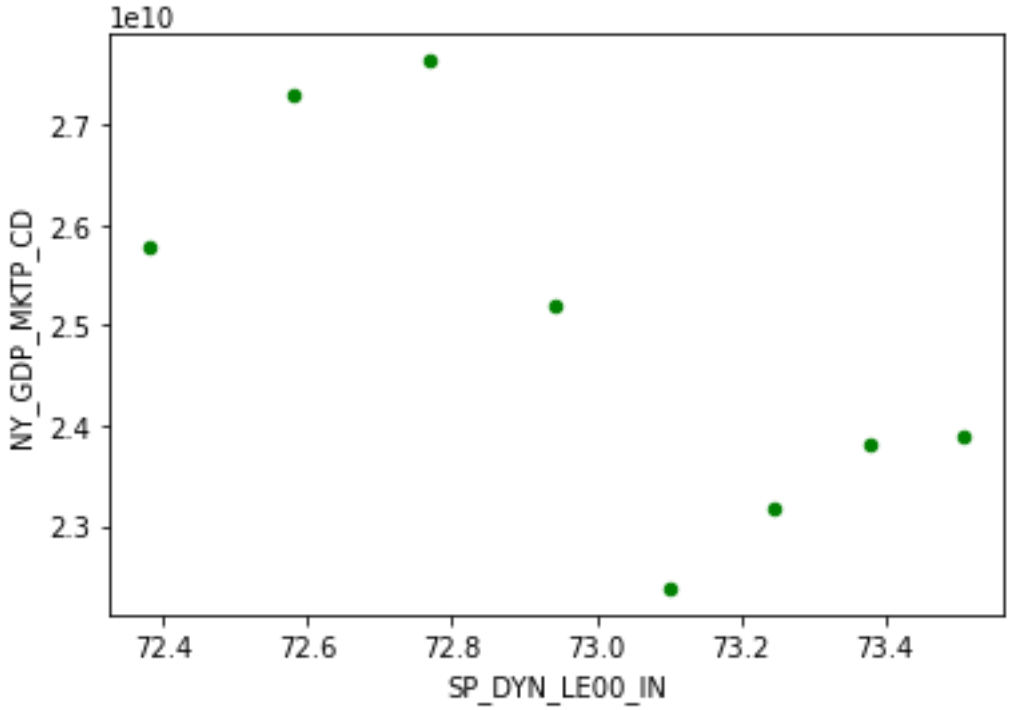


Figure 4.32: Result 32

(Source: Jupyter notebook)

From the visualisation, we can understand that the Total Life expectancy at birth given in years and the GDP of Trinidad and Tobago had a direct relationship for a certain range of life expectancy and for another particular range of life expectancy, the relationship is indirect.

SAN MARINO

The descriptive statistics of San Marino indicates that the average GDP of the country in US$ is lower than Trinidad and Tobago but higher than Solomon Islands. The below bar plot is for checking the GDP (current US$) of San Marino:

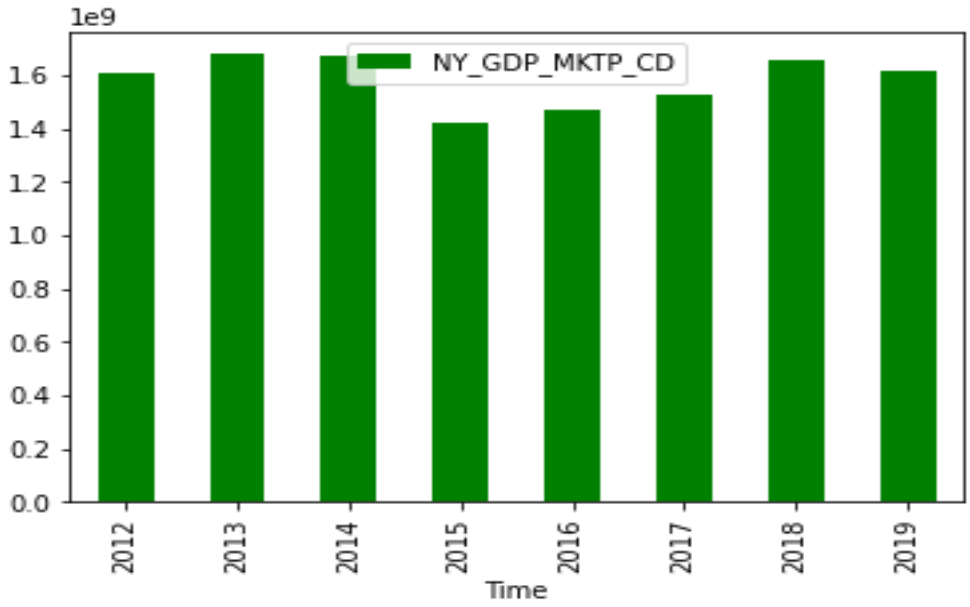


Figure 4.33: Result 33

(Source: Jupyter notebook)

It can observe from the above figure that the value of GDP of San Marino in US$ has been almost similar during the last decade. The maximum value was observed in the year 2013.

The next below plot shows the relationship between the population growth and GDP (current US$) in San Marino:

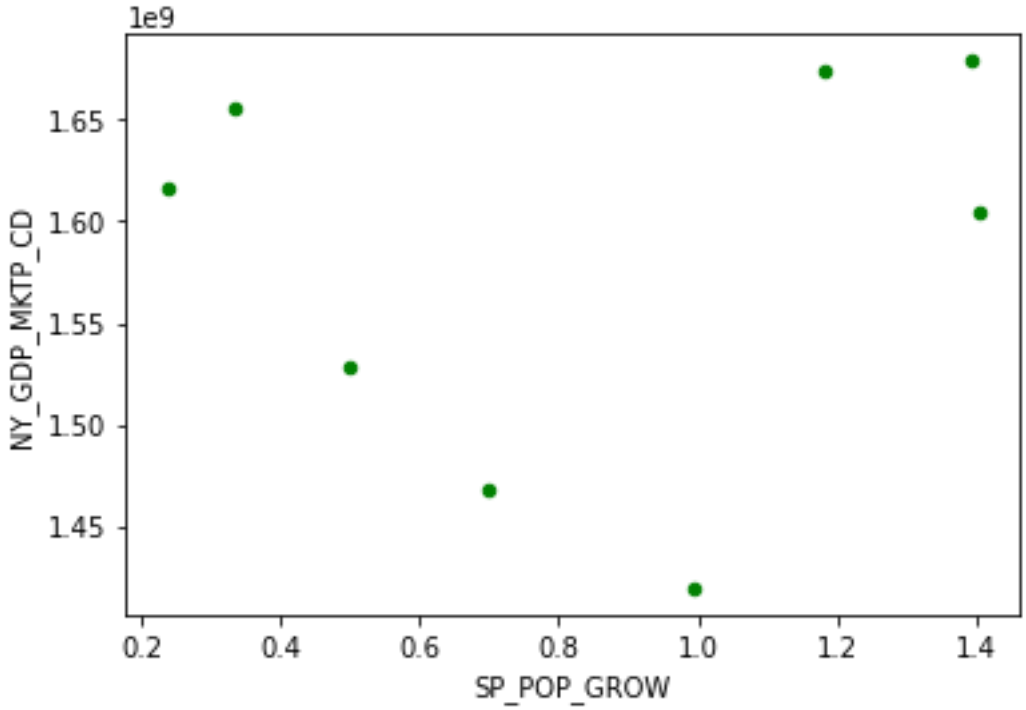


Figure 4.34: Result 34

(Source: Jupyter notebook)

As evident from the above figure, the relationship between the Annual % growth in population and the GDP of San Marino in US$ is V-shaped.

CABO VERDE

The descriptive statistics of Cabo Verde indicates that the average Annual GDP growth % is higher than San Marino and Trinidad and Tobago. The below bar plot is for checking the GNI growth (annual %) of Cabo Verde:

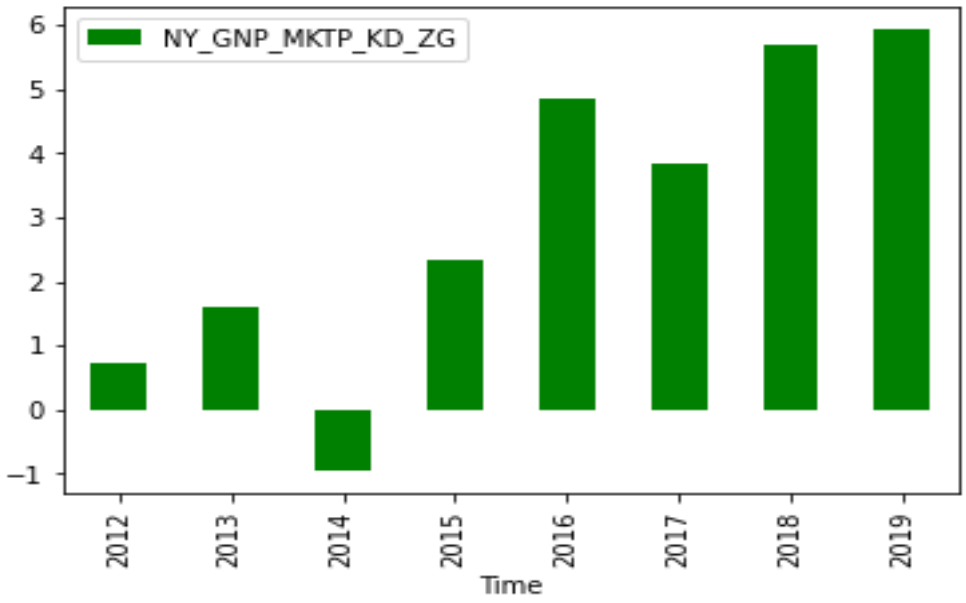


Figure 4.35: Result 35

(Source: Jupyter notebook)

It can observe from the above visualisation that the Annual % growth in GNI for Cabo Verde has been fluctuating. There was a negative GNI growth in the year 2014.The annual GNI growth has increased from 2017 to 2019 and the peak was observed in the year 2019.

The next below plot shows the relationship between International tourism, expenditures for passenger transport items (current US$) and GDP (current US$) in Cabo Verde:

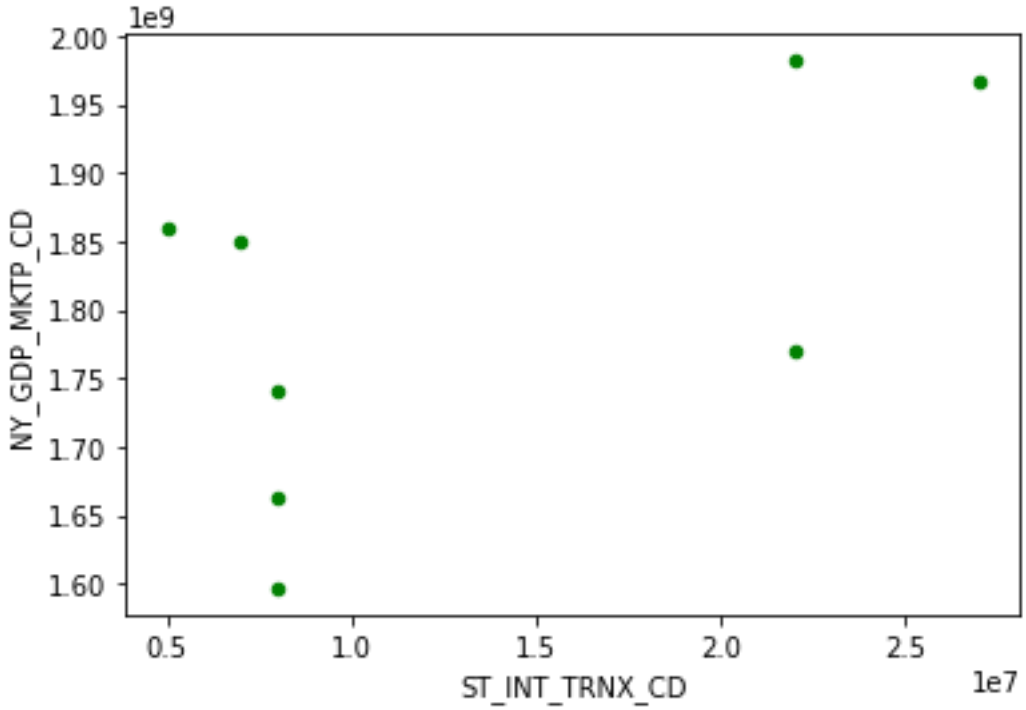


Figure 4.36: Result 36

(Source: Jupyter notebook)

It can see the above picture and say that the relationship between the Annual % growth in population and the GDP of Cabo Verde in US$ is direct in nature.

SEYCHELLES

The descriptive statistics of Seychelles indicates that the average Number of departures in International tourism is lower than Singapore but higher than Samoa. The below scatter plot visualisation is for checking the International tourism, number of departures in Seychelles:

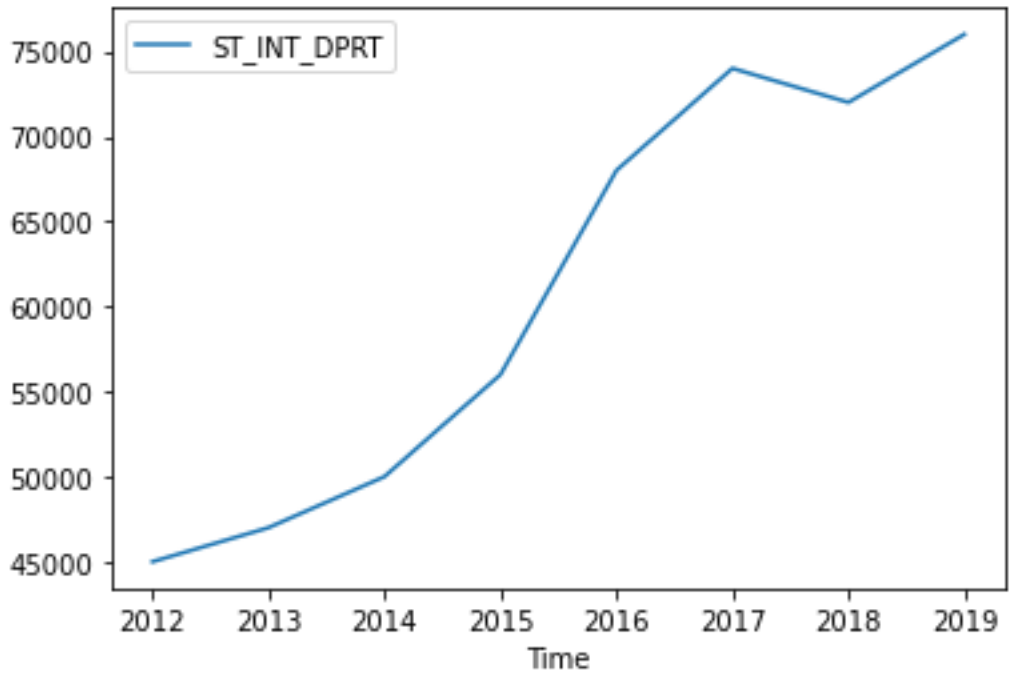


Figure 4.37: Result 37

(Source: Jupyter notebook)

The above visualisation indicates that the Number of departures in International tourism has increased significantly in the last decade for Seychelles. The maximum value was observed in the year 2019.

The next below plot shows the relationship between Air transport, registered carrier departures worldwide) and GDP (current US$) in Seychelles:

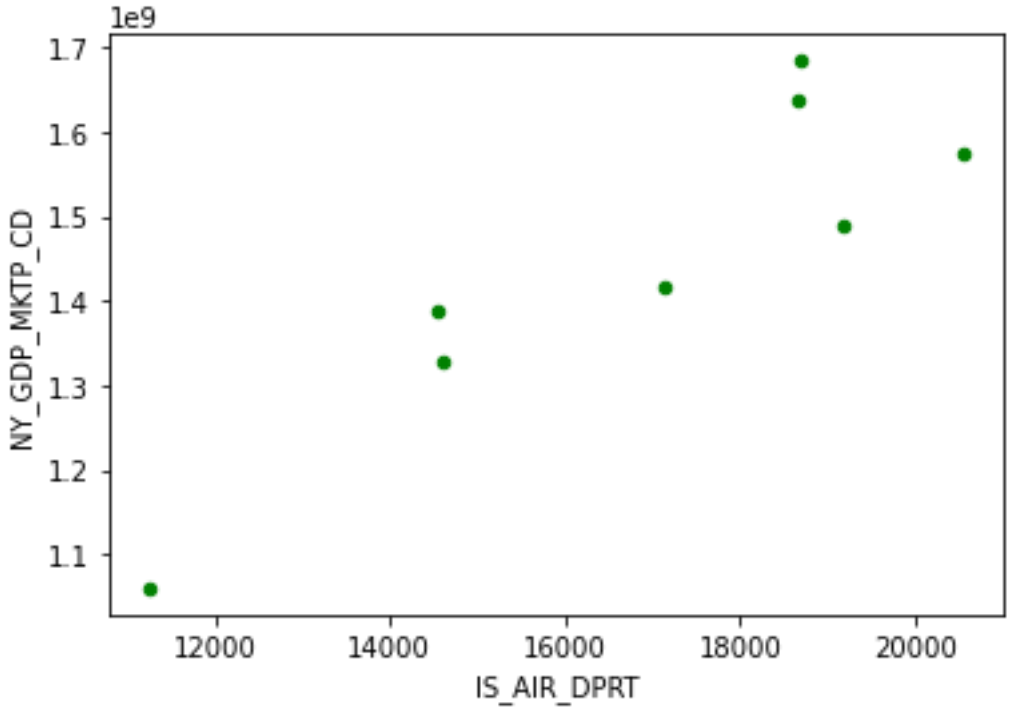


Figure 4.38: Result 38

(Source: Jupyter notebook)

From the above figure, it can be concluded that there exists a direct relationship between the GDP of Seychelles in US$ and the Registered carrier departures of air transport worldwide.

BAHAMAS

The descriptive statistics of Bahamas indicates that the average Annual % growth in GNI per capita is lower than Cabo Verde and Solomon Islands. The below bar plot visualisation is for checking GNI per capita growth (annual %) in Bahamas:

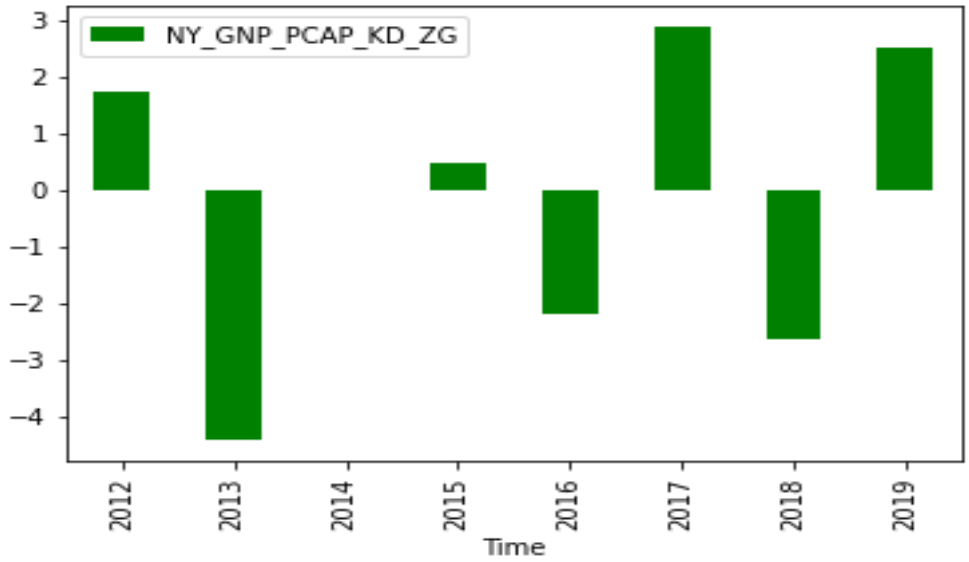


Figure 4.39: Result 39

(Source: Jupyter notebook)

The above visualisation tells us that in many years, the Annual % growth in GNI per capita was negative in the last decade. The lowest GNI per capita growth was observed in the year 2013 and the highest GNI per capita growth was seen in the year 2017.

The next below plot shows the relationship between Life expectancy at birth, total (years) and GDP (current US$) in Bahamas:

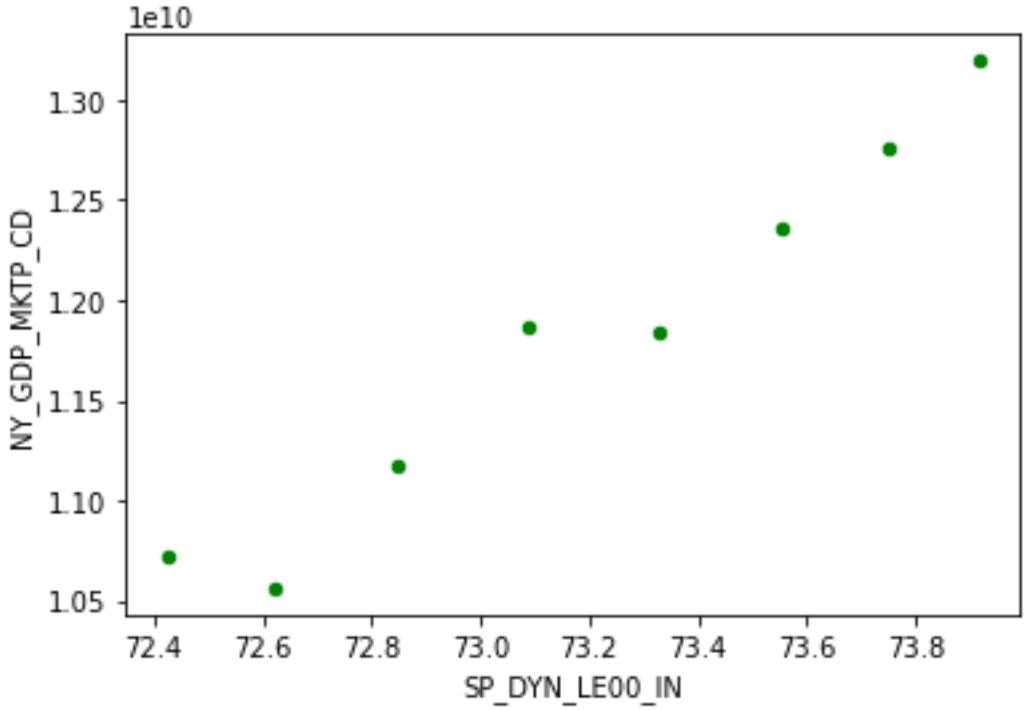


Figure 4.40: Result 40

(Source: Jupyter notebook)

The line plot above indicates that there exists a direct relationship between the GDP of Bahamas in US$ and the Total Life expectancy at birth given in years.

BRITISH VIRGIN ISLANDS

The descriptive statistics of British Virgin Islands indicates that the average Total Population is lower than Seychelles and Bahamas. The below scatter plot visualisation is for checking the total population in British Virgin Islands:

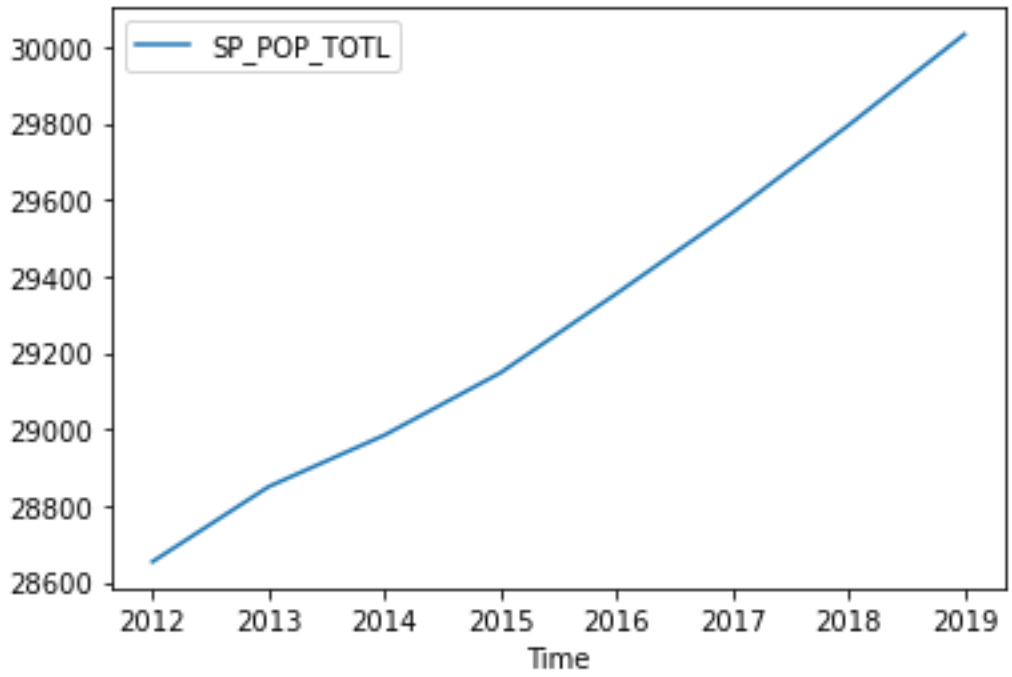


Figure 4.41: Result 41

(Source: Jupyter notebook)

It can observe from the above scatter plot that the Total Population of British Virgin Islands has increased significantly in the last decade. The peak is observed in the year 2019.

The next below plot shows the relationship between International tourism, number of departures and Total Population in British Virgin Islands:

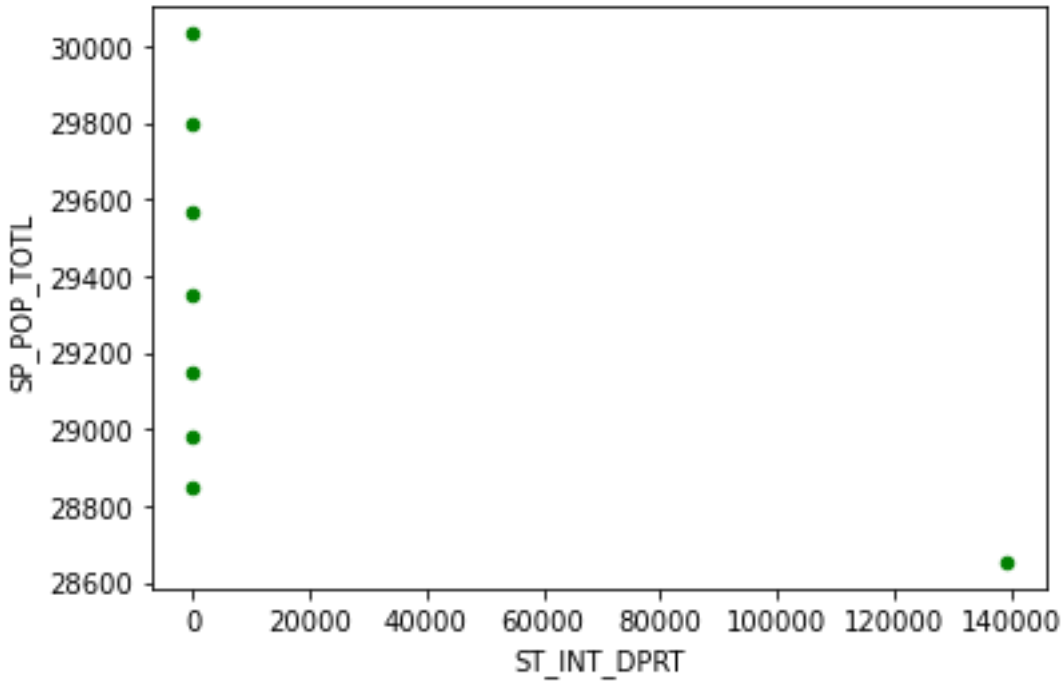


Figure 4.42: Result 42

(Source: Jupyter notebook)

The line plot shown above indicates that there exists an indirect relationship between the Total Population of British Virgin Islands and the Number of departures in international tourism.

VIRGIN ISLANDS (US)

The descriptive statistics of Virgin Islands (US) indicates that the average Annual % growth in population is lower than British Virgin Islands and Bahamas. The below scatter plot visualisation is for checking the population growth in Virgin Islands (US):

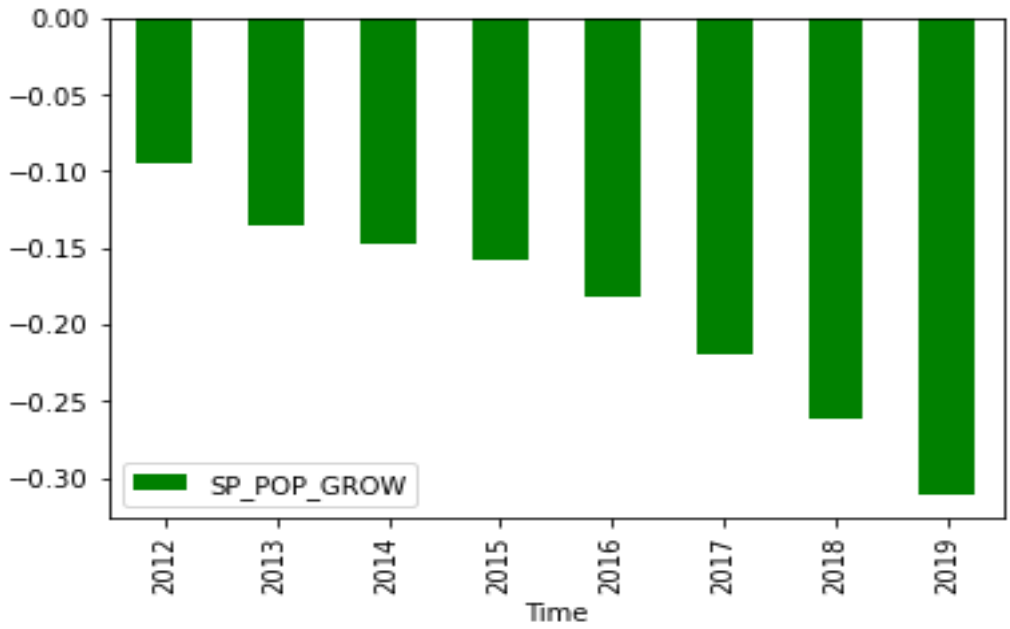


Figure 4.42: Result 42

(Source: Jupyter notebook)

The above bar plot indicates that there has been a negative growth in population for Virgin Islands (US) in the last decade. The minimum growth was observed in the year 2019.

The next below plot shows the relationship between Population growth (annual %) and GDP (current US$) in Virgin Islands (US):

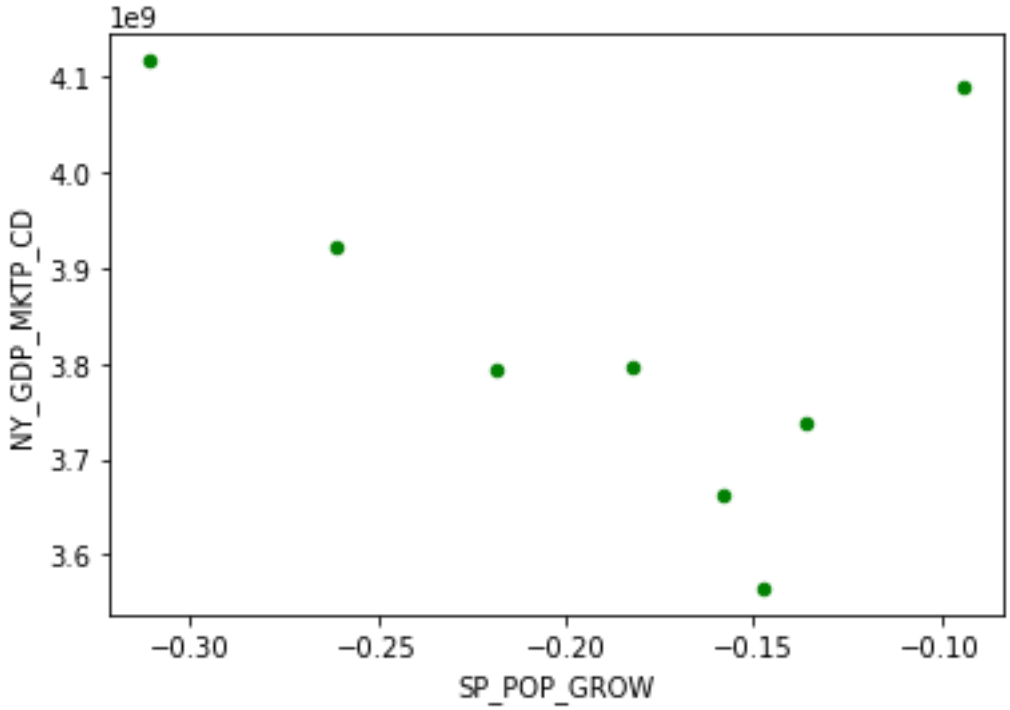


Figure 4.43: Result 43

(Source: Jupyter notebook)

The above scatter plot makes it evident that the relationship between the Annual % growth in population and the GDP of Virgin Islands (US) in US$ is V-shaped.

4.3 **4.3 Discussions**

The benefits of adopting big data in tourism are evident. Using big data would prevent sampling bias because it would not have to take this into account. The entire set of data is available, regardless of the data you want to gather. The only issue is if you have the necessary tools to acquire them. These data are likewise entirely user-generated, whether intentionally or unintentionally, and readily available, without any prerequisites or presumptions for their generation (Saint Awadari *et al.* 2019). Second, big data can be a reliable source of information to contrast with the conventional methods of gathering tourism-related data from other sources. It might be considered a fresh information flow. The third attribute of big data is its ability to dynamically capture information in real-time. Other than that, the third attribute of big data is its ability to dynamically capture information that is offered at any time and even live data that are being generated. Big data in the travel and tourism sector surely has a favourable impact on research thanks to its exceptional features.

The interaction between big data and tourism stimulates a variety of more practical and effective techniques as well as more creative study topics. Numerous internet travel platforms offer a limitless supply of data sources, and continuously updated analytical techniques make complex analyses visible and approachable. Big data in tourism is still in its infancy, and despite the fact that current researchers have carried out numerous studies to assess and analyse travelers, tourism businesses, and destinations, it displays with sparkling and broad prospects in the future. This study would like to make the case that the key to exploiting tourism big data is to make good use of data from numerous sources with heterogeneous structure by evaluating the general features, types, analytical techniques, and research progress. As per the view of Plagborg-Møller *et al.* (2022), this calls for advanced techniques and cutting-edge analytical frameworks that enable us to investigate putative mechanisms incorporated into our reality but challenging to identify and validate using conventional approaches.

4.4 **4.4 Summary**

From the above analysis, it has been stated that only the number of guests who have stayed in registered accommodations is provided by counting overnights. Surveys are used to gather information on overnights in a few categories of registered lodging. Private and unregistered commercial lodging are typically excluded from coverage. These data are only consistently produced by a small number of nations. In many nations, the informal housing market is significant, and with that importance come numerous positive but underappreciated contributions as well as difficulties, such as establishing uniform standards. The accuracy result shows that the model's independent variables can accurately predict the dependent variable's output 79.55% of the time (Sorci *et al*. 2020). The estimated difference between the actual values and the anticipated model's values, expressed as the root mean square error and mean absolute error, should be as little as feasible. For a model to be more effective, the r-squared value, which represents the degree of fit or the explanation of the difference in proportion of the dependent variable by the independent variable, should be as high as possible.

5 Chapter 5: Summary

5.1 5.1 Chapter Overview

We have analysed the dataset related to the economic indicators of the island countries. During our analysis, we have considered the positive or negative GDP growth of the island countries as the dependent variable. We have conducted a linear regression analysis and the OLS regression results showed that the two variables which are statistically significant are Total Life expectancy at birth in years and educational attainment, at least completed lower secondary, population 25+, total (%) (cumulative) as they have a p-value of less than 0.05 which is the threshold value. As most of the independent variables have a p-value greater than 0.05 we can reject the null hypothesis.

5.2 5.2 Linking with Objectives

For semi-structured interactions, having particular questions prepared in advance gives scientists more assurance that they'll cover all the bases with regard to the subject at hand. Strong national management promotes beneficial relationships with these other countries through qualitative sampling in order to share resources and coordinate responses as well as a common goal and strategy. Following the suggested adjustment in accordance with the model, plans and a vision that are in line with the organization's goal and that incorporate technology are developed. As per the view of Plagborg-Møller *et al.* (2022), the initiatives would promote staff training in order to improve the network infrastructure and the adoption of appropriate evaluation methods. The third component of this change model, which is related to information exchange regarding the introduction of remote location surveillance systems for treating and diagnosing patients, and the team leader plays a crucial role in promoting awareness of such functionalities and attributes of the system. However, the component doesn't seem to be sufficiently satisfied, increasing the risk factors for the proposed changes. In a business,

information is gathered from a variety of sources, including social networking pages, ERP software, customer logs, financial reports, emails, PowerPoint presentations, and employee-written reports. It could be difficult to organize reports after combining all of this data (Udemba *et al.* 2021).

To better evaluate a nation's overall economic health, economists are currently using Big Data analytics. The analysts can make well-informed decisions that will contribute to raising the nation's overall GDP with the aid of an adequate examination of these data. This study paper will concentrate on providing specific information concerning the how, the what, and why of the Big Data technology in order to understand how it affects economic indicators. To survive over time in such a competitive environment, the nations must examine the trend of the economic indicators, which are changing quickly (Nguyen *et al.* 2021). Utilizing big data technology, which enables us to draw conclusions from the pattern of the indicators, is an efficient way to accomplish this.

5.3 5.3 Recommendation

Even predicting which new products might succeed in a market is possible with the use of big data. Effective usage of big data technology can result in personalized offers for tourists that are catered to their wants and interests. They can get better experiences that are centered on the demands of the clients. Big data is responsible for precise decision-making in the tourism industry, as it is in many other corporate sectors. They frequently participate in price strategy optimization, service personalization, travel marketing, and customer demand forecasting. Although domestic travel has picked up again in many nations, it can only partially offset the decline in foreign travel (Nguyen *et al.* 2021). As per the view of Plagborg-Møller *et al.* (2022), The OECD currently predicts that international travel would decline by almost 80% in 2020. International tourism flows are not expected to significantly improve until well into 2021, and even then it will probably take some time. Adopting proper regulations for tourism investment, building adequate infrastructure, working with other important stakeholders, offering high-

quality goods and services, and installing facilities for health and safety are only a few examples of such efforts. Individuals from different walks of life can benefit from the opportunities provided by tourism, which also supports vulnerable populations including minorities, young people, and women. As people learn about and get to know one another better, it fosters compassion. The study of the dataset's enhanced data will eventually produce a number of beneficial results that the countries can employ to raise their GDP (Nguyen *et al.* 2021). Currently, the majority of nations employ various models to estimate their economy's demand using Big Data, as this is not achievable by simply looking at the data. An accurate study of the consumer data will aid nations in making decisions that will ultimately boost their GDP and lower their unemployment rates.

5.4 5.4 Research future scope

According to future study on tourism big data, social media, and internet travel alter traditional travel patterns by showing up before, during, and after the journey for both travelers and tourism businesses. In exchange, travelers' active participation gives tourism businesses the chance to pinpoint visitor travel traits and further delve into their preferences in order to deliver a satisfying service and an engaging experience. We-media software and travel apps are becoming more and more common on mobile devices, which are fostering the development of new business models in this sector. Big data in tourism can be a significant marketing and communication avenue to help a place build its brand. The effectiveness and quality of destination management will also constantly improve thanks to smart tourism and new big data applications in picturesque areas. On the one hand, it can give visitors a pleasant and informed experience; on the other side, it can also help scenic regions become better at controlling crises in real time and forecasting visitor numbers. Overall, the future application of big data in tourism is quite promising.

6 Reference List

Website

Sharma, N., 2020. *Ordinary Least Squared (OLS) Regression.* [Online]

Available at: https://medium.com/analytics-vidhya/ordinary-least-squared-ols-regression-90942a2fdad5

[Accessed 10 08 2022].

THANDA, A., 2022. *What is Logistic Regression? A Beginner's Guide.* [Online]

Available at: https://careerfoundry.com/en/blog/data-analytics/what-is-logistic-regression/

[Accessed 10 08 2022].

Journal

Adebayo, T.S., Akinola, G.D., Karkalla, D., Beken, F.V., Marbella, S. and Semeion, O.S., 2021. Economic performance of Indonesia amidst CO2 emissions and agriculture: a time series analysis. *Environmental Science and Pollution Research*, *28*(35), pp.47942-47956.

Adebayo, T.S., Akinsola, G.D., Kirikkaleli, D., Bekun, F.V., Umarbeyli, S. and Osemeahon, O.S., 2021. Economic performance of Indonesia amidst CO2 emissions and agriculture: a time series analysis. *Environmental Science and Pollution Research*, *28*(35), pp.47942-47956.

Adebayo, T.S., Awassa, A.A., Karkalla, D., Akinola, G.D. and Mwamba, M.N., 2021. Can CO2 emissions and energy consumption determine the economic performance of South Korea? A time series analysis. *Environmental Science and Pollution Research*, *28*(29), pp.38969-38984.

Adebayo, T.S., Udemba, E.N., Ahmed, Z. and Karkalla, D., 2021. Determinants of consumption-based carbon emissions in Chile: an application of non-linear ARDL. *Environmental Science and Pollution Research*, *28*(32), pp.43908-43922.

Ahmad, M., Khattak, S.I., Khan, A. and Rahman, Z.U., 2020. Innovation, foreign direct investment (FDI), and the energy–pollution–growth nexus in OECD region: a simultaneous equation modeling approach. *Environmental and Ecological Statistics*, *27*(2), pp.203-232.

Ahmadov, L.R. and Hajiyev, M.R. 2021, The Role of Air Transport in International Trade and Its Importance in The Development of The Agricultural Sector, Vara din Development and Entrepreneurship Agency (VADEA), Vara din.

Alrewas, K.N., Ruwais, A.M., Alaafin, S.B. and Allawah, N. 2022, "ECONOMIC ANALYSIS OF INDICATORS OF THE COMPETITIVENESS OF SAUDI DATE EXPORTS", Independent Journal of Management and Production, vol. 13, no. 2, pp. 829-840.

Andreaka, N.O., Andreaka, V.M. and Falchuk, H.B. 2022, "Assessment of ecological and economic competitiveness of regions using factor analysis", IOP Conference Series. Earth and Environmental Science, vol. 1049, no. 1, pp. 012084.

Anielak-Sobczyk, K. 2022, "A Bank's Intellectual Capital and Its Importance in Building Competitiveness on The Example of Polish Listed Banks", Economic i Prado, vol. 21, no. 1, pp. 5-24.

Arel-Bundock, V., Blais, A. and Dassonneville, R., 2019. Do voters benchmark economic performance? *British Journal of Political Science*, *51*, pp.1-13.

Assaf, A.G., Li, G., Song, H. and Tsongas, M.G., 2019. Modeling and forecasting regional tourism demand using the Bayesian global vector autoregressive (BGVAR) model. *Journal of Travel Research*, *58*(3), pp.383-397.

Beaudry, S., Ghosh, S. and Kumar, P. 2021, "Constructing a Coincident Economic Indicator for India: How Well Does It Track Gross Domestic Product?", Asian Development Review, vol. 38, no. 2, pp. 237-277.

Benkraiem, R., Lahaina, A., Malouda, A. and Shahbaz, M., 2019. The asymmetric role of shadow economy in the energy-growth nexus in Bolivia. *Energy policy*, *125*, pp.405-417.

Coze, A. and Mania, G. 2022, "SUSTAINABILITY AND ITS IMPORTANCE IN ENVIRONMENT PROTECTION", Romanian Economic and Business Review, vol. 17, no. 1, pp. 23-31.

Dumba, E.N., Ginger, H., Beken, F.V. and Karkalla, D., 2021. Economic performance of India amidst high CO2 emissions. *Sustainable Production and Consumption*, *27*, pp.52-60.

Duminda, S. and van Beuke ring, P., 2021. Understanding public concern about climate change in Europe, 2008–2017: the influence of economic factors and right-wing populism. *Climate Policy*, *21*(3), pp.353-367.

Ever, F. and Sarfaraz, H.Z. 2022, "Evaluating economic freedom via a multi-criteria MEREC-DNMA model-based composite system: case of OPEC countries", Technological and Economic Development of Economy, vol. 28, no. 4, pp. 1158-1181.

Foroni, C., Marcelino, M. and Stefanovic, D., 2020. Forecasting the Covid-19 recession and recovery: Lessons from the financial crisis. *International Journal of Forecasting*.

Guo, L.N., She, C., Kong, D.B., Yan, S.L., Xu, Y.P., Khayat Nezhad, M. and Gloxinia, F., 2021. Prediction of the effects of climate change on hydroelectric generation, electricity demand, and emissions of greenhouse gases under climatic scenarios and optimized ANN model. *Energy Reports*, *7*, pp.5431-5445.

Hasnisah, A., Alina, A.A. and Taiba, C.M.I.C., 2019. The impact of renewable energy consumption on carbon dioxide emissions: Empirical evidence from developing countries in Asia. *International Journal of Energy Economics and Policy*, *9*(3), p.135.

Hassan, O.A. and Romilly, P., 2018. Relations between corporate economic performance, environmental disclosure and greenhouse gas emissions: New insights. *Business strategy and the environment*, *27*(7), pp.893-909.

Horneck, T., Richter, A., William, S.H., Baste, M., Bollmann, S., Bonn, A., Bamberger, J., Dietrich, P., Seminole, B., Grote, R., Hein old, B., Keller, A., Luteus, M.L., Vara, E.M., Pisonia, S., Punyas Ena, S.W., Rakes, D., Richter, R., Sickle, W., Ingold Steffan-Dewenter, Theodoros, P., Treadle, R., Wenchang, B., Wenchang, M., Wolke, R. and Dunker, S. 2022, "An integrative environmental pollen diversity assessment and its importance for the Sustainable Development Goals", Plants, People, Planet, vol. 4, no. 2, pp. 110-121.

Hussain, H.I., Slusarczyk, B., Kamarudin, F., Thaker, H.M.T. and Szczepańska-Woszczyna, K., 2020. An investigation of an adaptive neuro-fuzzy inference system to predict the relationship among energy intensity, globalization, and financial development in major ASEAN economies. *Energies*, *13*(4), p.850.

Hussain, H.I., Surzhyk, B., Kamrudin, F., Thaker, H.M.T. and Szczepańska-Woszczyna, K., 2020. An investigation of an adaptive neuro-fuzzy inference system to predict the relationship among energy intensity, globalization, and financial development in major ASEAN economies. *Energies*, *13*(4), p.850.

Janko, F., Bottle, Z. and Goyri, R. 2022, "Vienna's South-Eastern Hinterlands: Regional Development in the Austrian-Hungarian Border Area, 1910–2011", European Countryside, vol. 14, no. 2, pp. 232-257.

Jena, P.R., Majhi, R., Kalli, R., Managi, S. and Majhi, B., 2021. Impact of COVID-19 on GDP of major economies: Application of the artificial neural network forecaster. *Economic Analysis and Policy*, *69*, pp.324-339.

Khan, Z., Bade, R.A. and Nawaz, K., 2022. Natural resources and economic performance: Evaluating the role of political risk and renewable energy consumption. *Resources Policy*, *78*, p.102890.

Khurshid, A. and Khan, K., 2021. How COVID-19 shock will drive the economy and climate? A data-driven approach to model and forecast. *Environmental Science and Pollution Research*, *28*(3), pp.2948-2958.

Khurshid, A. and Khan, K., 2021. How COVID-19 shock will drive the economy and climate? A data-driven approach to model and forecast. *Environmental Science and Pollution Research*, *28*(3), pp.2948-2958.

Korsakov, T., Bacevich, L., Drozdov, O., Mikhailova, A. and Kurchatovia, E. 2021, "Digital Economy and its Importance in the Development of Small and Medium Innovative Enterprises", International Journal of Advanced Computer Science and Applications, vol. 12, no. 11.

Kristeva, O.I., Melnyk, L.H., Kuratko, O.V., Dentaria, I.B. and Ericolin, A.O., 2019. Disruptive technologies for the transition of digital economies towards sustainability.

Kush lev, K., Radisic, N. and Diener, E., 2022. Subjective well-being and PR sociality around the globe: Happy people give more of their time and money to others. *Social Psychological and Personality Science*, *13*(4), pp.849-861.

Li, J., Liu, D. and Cai, M. 2022, "Water pollution and administrative division adjustments: A quasi-natural experiment in Chaozhou Lake, China", Plops One, vol. 17, no. 3.

Lim, W.M. and To, W.M., 2022. The economic impact of a global pandemic on the tourism economy: The case of COVID-19 and Macao's destination-and gambling-dependent economy. *Current Issues in Tourism*, *25*(8), pp.1258-1269.

Magazine, C., Mele, M., Schneider, N. and Valet, G., 2020. The relationship between nuclear energy consumption and economic growth: evidence from Switzerland. *Environmental Research Letters*, *15*(9), p.0940a5.

Milicevic, R., Mitrovic, D. and Simonović, Z. 2022, "SUSTAINABLE TOURISM DEVELOPMENT ACCORDING TO EU INDICATORS, IN THE FUNCTION OF ECONOMIC GROWTH OF NIŠKA BANJA", Ekonomika, vol. 68, no. 2, pp. 15-29.

Moistens, T.L., Moistens, T.G., Yuri, E., Lanker, Z. and Vasa, L. 2022, "Economic affordability of food as a component of the economic security of Ukraine", PLoS One, vol. 17, no. 3.

Nawaz, M.A., Hussain, M.S., Kamran, H.W., Ehsanullah, S., Maneen, R. and Shari, F., 2021. Trilemma association of energy consumption, carbon emission, and economic growth of BRICS and OECD regions: quantile regression estimation. *Environmental Science and Pollution Research*, *28*(13), pp.16014-16028.

Nguyen, D.K., Huynh, T.L.D. and Nasir, M.A., 2021. Carbon emissions determinants and forecasting: Evidence from G6 countries. *Journal of Environmental Management*, *285*, p.111988.

Nguyen, D.K., Huynh, T.L.D. and Nasir, M.A., 2021. Carbon emissions determinants and forecasting: Evidence from G6 countries. *Journal of Environmental Management*, *285*, p.111988.

Nit, K.O., Adekoya, A. and Woori, B., 2019. Random forest based feature selection of macroeconomic variables for stock market prediction. *American Journal of Applied Sciences*, *16*(7), pp.200-212.

Pavlik, M. and Pisa, P. 2020, "Strategic management controlling system and its importance for SMEs in the EU", Problems and Perspectives in Management, vol. 18, no. 3, pp. 362-372.

Pereira, R.C. and da Gama Bahia, R. 2021, "Rhinoliths: Can Its Importance on a Large Scale Be Promoted by a Microscale and Invisible Phenomenon?", Frontiers in Marine Science,

Philip, L.D., Emir, F. and Alzola, A.A., 2022. The asymmetric nexus of entrepreneurship and environmental quality in a developing economy. *International Journal of Environmental Science and Technology*, *19*(8), pp.7625-7636.

Plagborg-Møller, M., Reichlin, L., Ricco, G. and Haenszel, T., 2020. When is growth at risk? *Brookings Papers on Economic Activity*, *2020*(1), pp.167-229.

Rehman, A., Rauf, A., Ahmad, M., Chandio, A.A. and Deyan, Z., 2019. The effect of carbon dioxide emission and the consumption of electrical energy, fossil fuel energy, and renewable energy, on economic performance: evidence from Pakistan. *Environmental Science and Pollution Research*, *26*(21), pp.21760-21773.

Rohrer, K., Murtaza, A., Mahdi, K. and Ali, T. 2021, "Assessing the effects of desertification control projects using socio-economic indicators in the arid regions of eastern Iran", Environment, Development and Sustainability, vol. 23, no. 7, pp. 10455-10469.

Safarov, Q.A., gaze, G.V. and Galkin, S.V. 2022, "Application of statistical dependencies to evaluate economic indicators of production efficiency by the example of oil fields in Azerbaijan", IOP Conference Series. Earth and Environmental Science, vol. 1021, no. 1, pp. 012003.

Saint Kediri, S., Beken, F.V. and Sarkodie, S.A., 2019. Contemporaneous interaction between energy consumption, economic growth and environmental sustainability in South Africa: what drives what? *Science of the total environment*, *686*, pp.468-475.

Sen, D., Guna, M.E. and Tunk, K.M., 2019. Forecasting annual natural gas consumption using socio-economic indicators for making future policies. *Energy*, *173*, pp.1106-1118.

Sen, D., Günay, M.E. and Tunç, K.M., 2019. Forecasting annual natural gas consumption using socio-economic indicators for making future policies. *Energy*, *173*, pp.1106-1118.

Shinkaruk, O., Boysen, V., Bedsheets, N., Goncharovian, N., Stepanek, O., Radioneon, M. and Pliskova, O. 2022, "Risk assessment of sports activity and the impact of economic and demographic indicators during the COVID-19 pandemic", Journal of Physical Education and Sport, vol. 22, no. 3, pp. 672-679.

Sorci, G., Favre, B. and Morando, S., 2020. Explaining among-country variation in COVID-19 case fatality rate. *Scientific reports*, *10*(1), pp.1-11.

Stapinski, N., Zimin, G. and Brownie, D. 2021, "The Complementary Currency Sialon in Poland and Its Importance for the Development of Local Economy Entities during the COVID-19 Pandemic Lockdown", Sustainability, vol. 13, no. 16, pp. 9184.

Udemba, E.N., Güngör, H., Bekun, F.V. and Kirikkaleli, D., 2021. Economic performance of India amidst high CO2 emissions. *Sustainable Production and Consumption*, *27*, pp.52-60.

Verma, P., Dumka, A., Bhardwaj, A., Ashok, A., Kestwal, M.C. and Kumar, P., 2021. A statistical analysis of impact of COVID19 on the global economy and stock index returns. *SN Computer Science*, *2*(1), pp.1-13.

Verma, P., Dumka, A., Bhardwaj, A., Ashok, A., Kotwal, M.C. and Kumar, P., 2021. A statistical analysis of impact of COVID19 on the global economy and stock index returns. *SN Computer Science*, *2*(1), pp.1-13.

Vogt Isaksen, J., 2019. The impact of the financial crisis on European attitudes toward immigration. *Comparative Migration Studies*, *7*(1), pp.1-20.

Waheeda, R. 2022, "The Significance of Energy Factors, Green Economic Indicators, Blue Economic Aspects towards Carbon Intensity: A Study of Saudi Vision 2030", Sustainability, vol. 14, no. 11, pp. 6893.

Wang, R., Mirza, N., Abueva, D.G., Abbas, Q. and Xiong, D., 2020. The nexus of carbon emissions, financial development, renewable energy consumption, and technological innovation: what should be the priorities in light of COP 21 Agreements? *Journal of Environmental Management*, *271*, p.111027.

Wang, Y. and Wang, J., 2020. Modelling and prediction of global non-communicable diseases. *BMC Public Health*, *20*(1), pp.1-13.

Wang, Y. and Wang, J., 2020. Modelling and prediction of global non-communicable diseases. *BMC Public Health*, *20*(1), pp.1-13.

7 A **Packages** Installation**.**

Project was done with the help of python, Anaconda Jupyter Notebook. The code, presented in .ipynb file holds the code to operate the project.

Python packages such as NumPy, Pandas, Sklearn, Matplotlib, Seaborn, Statsmodels are used.

- import numpy
- import pandas
- import sklearn
- import matplotlib
- import seaborn
- import statsmodels

Above commands are to check if the python packages are installed in the OS(Operating System), if else it shows an error.

What does each python package do and how to run the notebook.

NumPy

NumPy helps in supporting mathematical computing in Python, Arrays and matrices comes under the category of NumPy.

Pandas

It is used to perform data analysis and manipulation of data. A data structure named Data-frame makes it easier to achieve these tasks

Sklearn

A tool in Python, to perform data prediction analysis and to import machine learning algorithm required for the project. It is an open-source tool

Matplotlib

Creating visualization plots of different manner is its background.

Seaborn

A tool based on Matplotlib, which provides interface for plots in Python

Statsmodels

Is a python module that provides functions for the estimation of different statistical models as well as conducting tests and exploration.

Pip

To install python Packages, we need Pip and to install Pip, https://pip.pypa.io/en/stable/installation/

We can install packages as shown below.

- pip install numpy
- pip install pandas
- pip install ~U scikit-learn
- pip install matplotlib
- pip install seaborn
- pip install statsmodels

8 B Professional **Issues**

The project's first and foremost issues rests in segregating the data and preparing it. It took some time to identify the goal of the project and to aim for it. The goal is to identify correlation between economy growth of an island with its several other attributes and draw conclusions based on patterns of growth. The modules that assisted me in moving towards the completion of the project are Data Analysis, Programming for Data Analysis and Machine Learning, therefore it all started with getting to know the concepts taught, more specific and detailed.

As for the outcome, I once again went back to study the aim of the project given in its description, which says to understand the relationship between economic (GNP) and various other data related to the islands. I went forward and made analysis(graphs) on every island, which gives us correlation between their different attributes, drawing hypothesis from them and mention if they have a common pattern of growth.