Aviation Islands

Tarun Malarmaran

Submitted for the Degree of Master of Science in

Data Science and Analytics



Department of Computer Science

Royal Holloway University of London

Egham, Surrey TW20 0EX, UK

November 26, 2021

**Declaration**

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 Where other published and unpublished source materials have been used, these have been acknowledged.

**Word Count**:

**Student Name**: Tarun Malarmaran

**Date of Submission**: November 26, 2021

**Signature**:

**Abstract**

The single airport island nations’ data were gathered from various source including World Bank’s Dataset. This project aims at cleaning this ambiguous data with a lot of missing values using appropriate imputation methods wherever necessary and form a subset for various aviation island nations the data for a period of thirteen years.

To find correlation of the datapoints generated by these island’s data for each year to discover various attribute’s dependency and find corelation between islands based on a specific attribute to draw hypothesis based on certain pattern of growth. Applying various data visualising techniques to answer specific queries on finding similarities and inter-relationships such as area, population etc of different islands.

**Contents**

1 Introduction 3

1.1 Dataset Initial Analysis 3

1.1.1 Sub-subsection 3

1.1.2 Sub-subsection 3

1.2 Subsection 3

2 Section two 4

3 Section three 5

# Introduction

The dataset is about the island nations across the world. The data recoded in this dataset are about the economic statistics of the all the islands with their population data, geographic area and aviation statistics such as number of flights bound and via various modes of transportation.

The problem arises when we try to incorporate some meaningful insights from this dataset. The various statistical data are vaguely recorded in various sheets in the Microsoft Excel file which makes nearly impossible to make sense. To resolve this situation, various data cleaning and imputation methods have been incorporated in this project to perform the analysis.

This is a sample section. The heading should be in style ‘Heading 1’.

Please do not change the styles used for headings, paragraphs, bullets, and so on.

Do not change the layout of the document either. For lists, use the ‘bullet’ style as follows:

* Bullet one
* Bullet two

Resume rewriting in the ‘Normal’ style.

## Dataset Initial Analysis

**Sheet**: The table in this sheet of the dataset seems to have the data of 40 island countries, with the attributes being about United nations and non-United Nations developing states. About population and year, it was recorded, area, number of arrivals from in different years.

**Sheet 1**: This sheet contains the data of traffic between Mauritius and Rodrigues which appears to be in a wide format but the column attributes are unclear, and around 70% of the cells have missing values.

**Sheet 2**: This sheet contains the individual data for Mauritius island, the attributes are almost same as the ones recorded in sheet 5, but the data here is in wide format.

**Sheet 3**: This sheet contains the individual data for Seychelles island, the attributes are almost same as the ones recorded in sheet 5, but the data here is in wide format.

**Sheet 4**: This sheet appears to have 2 separate tables, the wide table on the top appears to have some data about the rest of the world, Mauritius and Seychelles. The tall format table below seems to have the expenditure comparison among the three islands: Antigua and Barbuda, Mauritius and Seychelles.

**Sheet 5**: This sheet contains the data of 27 different islands recorded over 13 years of time span, where for each year the population, area, GDP, number of incoming flights, hotel rooms, visitor’s average expenditure, number of day visits, number of arrivals from different modes of transportation, number of arrivals from different counties, and certain inbound/outbound tourism expenditure attributes.

**Sheet 6**: This contains the data from dutch territory, the attributes are almost same as the ones recorded in sheet 5, but the data here is in wide format.

### Sub-subsection

This is an example of a sub-subsection. The heading should be in style ‘Heading 3’.

### Sub-subsection

This is another example of a sub-subsection

## Subsection

This is an example of a subsection.

# Section two

Sections start automatically in a new page.

# Section three

**References**

The list of sources that you have used, which should have been referenced in the body of the report.

1. Plotly: Scientific and Interactive open-source data visualisation application integrated with python 3. <https://plotly.com/>