Aviation Islands

**Project Plan**

Around 7 weeks to complete the project work:

**Week Mon, 4th Oct**:

* Project Plan.
* Initial analysis of dataset.
* Setup a private git repository to store the project files.

**Week Mon, 11th Oct**:

* Deciding the tools to clean the data (Python or R).
* Finding the problem with the data formats and choosing appropriate pre-processing techniques to clean the tables in each sheet.

**Week Mon, 18th Oct**:

* Exploratory analysis of the pre-processed data. And decision on what story to which we want to convey with our findings and through various visualisation techniques.

**Week Mon, 25th Oct**:

* Applying machine learning techniques to find the correlations between the island data and find the patterns which emerge during exploratory analysis.
* Draw possible hypotheses from the clean data.

**Week Mon, 1st Nov**:

* Can we get a better understanding of the data by using tools such as Power BI? Will it be a more convincing way to narrate the understandings?

**Initial rough understanding of the dataset:**

**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.