**Welcome to the Little Red Dot**

**Data analysis of International Arrivals to Singapore (2000-2018)**

This project examines data on international arrivals to Singapore from 2000-2018 to uncover interesting trends and provide actionable insights. We begin by performing data visualisation of and investigating general trends, before conducting regression analysis to identify significant factors influencing total arrival numbers. These findings are then interpreted and translated into actionable insights for the relevant stakeholders in Singapore’s tourism industry.

**Project Objectives**

Welcome to the Little Red Dot was initiated to study trends in international arrivals to Singapore and extended later to determine key factors affecting these numbers. With tourism being a key pillar of Singapore’s economy (contributing about 4% of its GDP), it is crucial to draw insights from relevant data to inform the decisions of various stakeholders.

The project is also a platform for me to apply the skills I have picked up over summer towards a self-initiated data analytics project.

**Thought Process and Methods**

The project initially began with only a single data set of international arrivals to Singapore, categorised by year, month and country (1978-2019). That was enough to conduct some data visualisation and trend analysis across time to answer the following questions:

1. What is the general trend of international arrivals to Singapore? Are there reasons for a fall in certain years?
2. Where are the bulk of international arrivals from?
3. Are there ‘cyclical effects’ on international arrivals each year? Why?

However, we wanted to go further and ascertain statistically significant factors influencing international arrival numbers to Singapore. Thus, more data sets were added to the project: GDP per capita (World Bank), Total Population (World Bank) and distances/common language (CEPII). Both cross-sectional and fixed effects regression analysis were performed to answer the following questions:

1. What is the effect of GDP per capita on international arrivals?
2. What is the effect of total population on international arrivals?
3. Does sharing a common language increase international arrivals?
4. How does geographic distance affect international arrivals?

**Key Findings**

General trends

* International arrival numbers generally increased between 2000-2018, although it fell in some years due to adverse shocks such as SARS (2003), the Global Financial Crisis (2009) and MH370 (2014)
* China, Indonesia and India are the top countries contributing to international arrivals in Singapore – Visitor numbers from China and India have been growing steadily and consistently
* There is a cyclical pattern in arrivals to Singapore every year: arrival numbers peak in July, slump in September and surge again in December

Influential factors

* As total population and GDP per capita of a country grows, we can expect to see an increase in international arrivals from that country
* As distance between a country and Singapore grows, we can expect to see a fall in total international arrivals from that country
* Sharing a common language with Singapore is not statistically significant towards international arrival numbers to Singapore

**How to read/run the Project**

Installation and Packages required

* Import any required Python packages, type “pip install (package name)” into the Anaconda Prompt
* To run the R codes on Jupyter notebook, type “conda install -c r r-essentials” on your terminal – it will install the R kernel and some important R packages (e.g. dplyr, ggplot2, etc.)

**Description of Repository Files**

All files related to the project can be found in the folder “Welcome to the Little Red Dot”

1. Welcome to the Little Red Dot\_Project Slide Deck.pdf – Slide deck reporting the project process, analysis and insights
2. Original Datasets folder – Complete and original data sets used in the project
3. [1] Data Cleaning and Merging.ipynb – Codes used to create the final merged data set
4. [2] Welcome to the Little Red Dot.ipynb – Data analysis and visualisations of general international arrivals trends
5. [3] Regression Analysis & Conclusion.ipynb – Cross-sectional and fixed effects regression analysis and Conclusions
6. country\_population.csv – Tidy data of total population in each country (2000-2018)
7. df\_2018\_model.csv – Sliced data set to be used for cross-sectional regression analysis
8. df\_merged.csv – Single merged data set incorporating total international arrivals, GDP per capita, total population, distance from Singapore and common language of each country
9. df\_model\_main.csv – Cleaned and transformed data set for regression analysis
10. gdp\_per\_capita.csv – Tidy data of GDP per capita in each country (2000-2018)
11. sg\_intl\_arrivals.csv – Tidy data of international arrivals from in each country (2000-2018)
12. sgp\_dist\_lang.csv – Tidy data of a country’s distance from Singapore and whether it shares a common language

**Acknowledgements**

Thank you for your interest in my project. I hope you enjoy looking through the codes and slide deck in this repository.

I would like to acknowledge the SMU Business Intelligence and Analytics Club for the opportunity to be part of the Data Associate Fast Track Programme. The programme allowed me to pick up Python 3 and apply it to perform data analytics.

Finally, this project would not have been possible without the help and support of my partner, Jocelyn.