1. Introduction

The tourism industry across the world came to a halt with the emergence of a COVID-19 pandemic. With consecutive lockdowns coupled with the fear of travelling, the once crowded airports ended up being nearly vacant. Similarly, to other countries, the Australian tourism industry performance faced a sheer drop and is gradually recovering since the removal of travel restrictions in 2022. The highly competitive business environment and the volatility in the tourism industry has highlighted the need for local businesses in Australia to shift towards data driven decision making. To keep the momentum of tourism recovery and to further improve the performance of the tourism industry, it is of utmost importance that businesses are equipped with the necessary tools and insights to adjust with the dynamic nature of the industry whilst incorporating insights that could make strategies and products more attractive to tourists. Data on inbound Australian tourist arrivals are readily available, but they all seem to be scattered and hence lacking a single 360-degree view for a business user. The lack of such a user-friendly insights portal is what sparked the development of this project. The purpose of this project is to develop an interactive insights portal that can be utilized by local businesses and others interested in making more informed decisions. It will enable the end user to keep a pulse on the market, identify the best areas for investments or opportunities within an area. Furthermore, behavioral insights that could be incorporated into business strategies to make products and services more attractive. The purpose of this report is to detail out the development of the Australian Inbound Tourism Insights Portal.

2. Five Design Sheet methodology

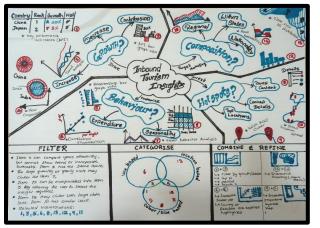
The purpose of the five-design sheet methodology is to formalize a process of designing visualizations without being held back by the technical details. It is centered around the end users' requirements and enables the user to design effectively and efficiently. This methodology was utilized in designing the Inbound tourism Insights Portal.

When considering the requirements of the end users, who are businesses in the tourism industry, there are four key pieces of information they would require for day-to-day operations as well as for investment decisions. First is keeping a pulse on the tourism market **growth**, the area and granularity of this may vary from business to business. Understanding which tourist inflows are stable, growing and are on the declining can help to develop more effective business strategies. Secondly, understanding the similarity and differences in tourist **behavior** will enable market segmentation and customization of products and offers that may be much more attractive to the tourists. This is followed by need to understand the **composition** of tourists by location, this could allow the business users to understand which areas to invest in if they are interested in specific target markets or which nationalities should be the target market if a specific location is in mind. Finally, they need an in-depth understanding of tourism **hotspots**, this will allow them to identify areas with most tourist content and hence ideal for investments.

Figure 2.1 is an image of sheet 01 which is the result of the initial brainstorming sessions. The ideas root from the four areas highlighted above. Growth requires visualizations that can enable the end user to carryout growth analysis and have the flexibility to deep dive into granular details. Visualizations such as line graphs, highlight table, year-to-year comparison bubble charts for each nationality, year-on-year change in visitors bar chart were considered. The behavior of tourists can be analyzed via visualizations

such as key variable bar charts sorted, a scatterplot for relationship analysis, a heatmap to analyze seasonality of visitors or a line chart across the 12 months color coded by the year. The composition of the tourists in a state can be obtain analyzed with a choropleth graph coupled with a bar graph of visitors by country of residence. Furthermore, a ribbon chart can be utilized to understand the change in composition of tourists across time. Lastly, the hotspots can be analyzed with a proportional symbol map or a choropleth map along with two summary tables for the end user to have access to granular information. From this stage a mixture of visualizations were filtered which consists of line graphs, maps, size based and color-based graphs. Each of the four areas with two visualizations were filtered and refined to create sheet02, sheet03 and sheet04.

Figure 2.2 is of design sheet 02. A line graph of the number of short-term tourist visits across time was selected coupled with a highlight table. Line graphs are mostly used in trend analysis due to their simplicity for the end user to understand and the fact that it's ideal for trend analysis. This brought upon the idea of utilizing something simple but equipping it with everything required to make it flexible to use. The highlight table will summarize the average visitors per year from other countries along with the growth rate. The growth rate will be the key performance indicators for which will be color coded and marked to emphasize the changes. The combination of these two visualizations along with multiple filters will provide the end user with a lot of flexibility to obtain the insights that are relevant to their needs while maintaining simplicity.



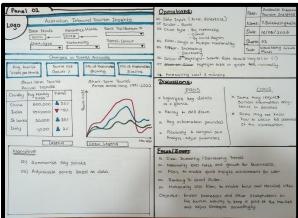


Figure 2.1: Design Sheet01

Figure 2.2: Design Sheet02

When it comes to design sheet03 (Refer figure 2.3), a scatterplot of average expenditure per night vs average nights stayed was selected, along with a heat map of tourist visits across the 12 months by each nationality. The scatterplot was mainly used to highlight the inverse relationship between these variables while providing the user to highlight the average expenditure behavior of tourists based on their nationality. This will enable businesses users to keep tabs on their spending habits and incorporate these insights into business plans and pricing. The heatmap was used to highlight the seasonality in tourist visits based on the country of residence to incorporate these insights into business plans.

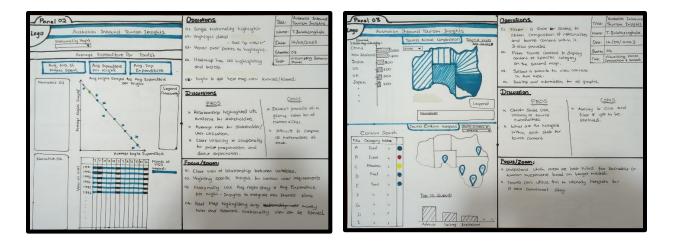


Figure 2.3: Design Sheet03

Figure 2.4: Design Sheet04

Design sheet04 revolves around insights on opportunity identification. This includes understanding tourist composition across states, for which a choropleth map coupled with a bar graph was selected (Refer figure 2.4). On the other hand, it also needs to be able to provide the end user with insights on tourism content available in specific are in order to material an opportunity. For this a proportional symbol map with two highlight tables was selected. One summarized the category of tourist content found and the quantity, while the other summarized the specific tourist content details such as name, contact details, location etc.

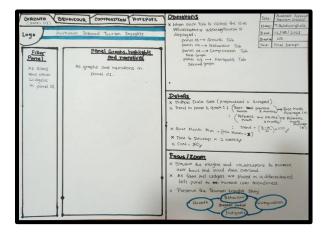


Figure 2.5: Design Sheet05

Once each of the detailed designed sheets was completed with the objective of catering insights to the business community in the tourism industry, the final step was to compile all visualizations in a creative, user-friendly and a clear flow. Each of the 3 design sheets ensured to meet the requirements of the end user, therefore with the aim of portraying a clear story, each of the four insights were allocated a separate tab in sheet05 with the filters placed in a specific location (Refer figure 2.5). This was done to avoid overloading the end user with information, while keeping the visualization simple and clear with a smooth flow from one topic to another.

3. Visualizations with R Shiny

3.1 Australian Inbound Tourism Roadmap

The overall layout of the Australian Inbound Tourism Dashboard consists of a yellow header with the title of the project (R shiny headed colors are limited to 6). The colour is closer to "yellow ochre" which is easy on the eyes but ensures to capture attention without fail. The Tourism Australia logo is placed on the left-hand side to signify the project owner, the black side bar with the tabs allows easy access to the visualizations whilst maintaining a clear roadmap illustrating the story behind the topic of interest and highlighting the information available on the portal. The background is maintained as light grey or white to ensure more comfort to the end user when working with a screen for long hours. The landing page was designed to convey the main message which is the flow of the four main areas of interest in the subject of Australian Inbound Tourism (Refer Figure 3.1 below). The portal provides insights of the four areas which includes Tourism growth, tourist behavior, composition and tourism hotspots. Furthermore, it provides a summary of what is provided in each of the tabs. A user can click on the three lined menu icon to open the navigation panel and select the tab of interest. Furthermore, the lower section provides the hyperlinks to the original datasets that were utilized in the development of the portal.

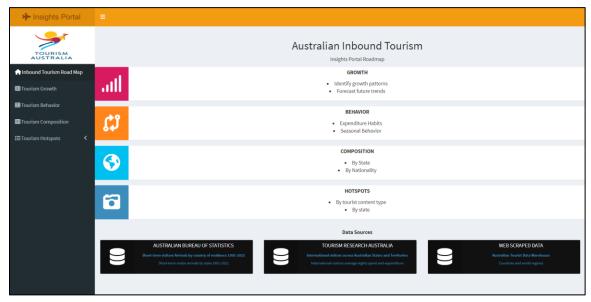


Figure 3.1.1: Insights Portal Roadmap

3.2 Analyzing Inbound Tourism Growth

When working with dashboards or any sort of visualizations, majority of the time people tend to want the ability to drill down to obtain more information. They require the flexibility to utilize it as a tool to answer the questions they have without it being a more general product. The trend analysis tab has been equipped with a "date range" filter so people can customize the visualization according to their needs (Refer Figure 3.2.1). For example, some may want to understand tourist nationality wise growth prior covid or in another scenario they might want to understand the recovery rates based on the country of residence. Furthermore, a "chart type" filter was incorporated incase the business users have the need to understand tourism growth

in terms of world regions (Such as Asians, Europeans, Africans...etc) or even the overall tourism arrival trends (Refer figure 3.2.3). In addition, a "**Top countries**" filter have been provided ranked by the average number of visits per year to allow the users to avoid congesting the visualizations (Refer figure 3.2.4).

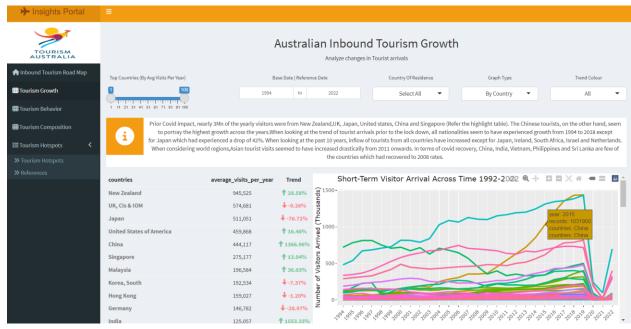


Figure 3.2.1: Tourism Growth Insights

The interactive line graph consists of a tooltip to highlight the required information from the graph. The colors utilized are from the "ggplot default" color palette. The main purpose of color in this visualization was to signify its line graphs of tourist arrivals by multiple nationalities. Regardless of the quantity of the lines mapped on the graph, the colors provided were significantly distinct. This is most apparent when analyzing 10 or less nationalities at a time (Refer Figure 3.2.3). On the left-hand side, a summary of the tourist arrivals by country of residence is given, along with the average visits per year and the growth rate based on the date range selected. The highlight table is colour coded with green to signify growth or red to signify a decrease. Furthermore, these key performance indicators are marked with an upward arrow to signify growth and a downward arrow for a decline. Both colour and mark visual variables are utilized to increase the distinctiveness in the visualization.

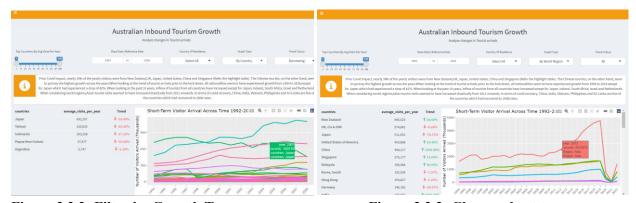


Figure 3.2.2: Filter by Growth Type

Figure 3.2.3: Change chart type

Figure 3.2.2 shows the use of "Trend colour" filter. In the image it's been used to filter out all the tourists whose visits have been declining from the year 2008 to 2018. In order to provide the end user the ability to customize the visualizations, a "Country of Residence" filter is provided.

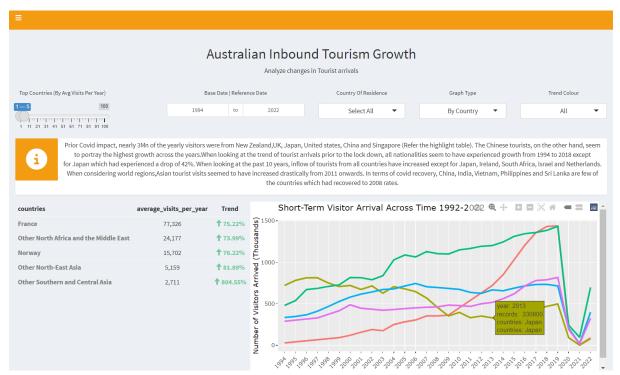


Figure 3.2.4: Top N Countries



Figure 3.2.5: Filter by Country

3.3 Analyzing Inbound Tourist Behavior

Humans are creatures of habits. Their way of life and preferences vary which defines them. But in today's competitive environment, businesses deep dive into market segmentation and customer analytics to better cater to their target market. Similarly, the aim of the insights portal is to allow the business users to keep tabs on the average expenditure habits of tourists from different nationalities (Refer Figure 3.3.1). Understanding their expenditure preferences better will pave the way to more attractive pricing, products, and services. For this requirement we have used a scatter plot, a traditional way of establishing a relationship between two variables. This specific form of visualization was selected with the end user in mind. In today's age nearly everyone is familiar with scatterplot. Its simplicity and wide awareness will be ideal when presenting to a busy business community. But most of all, it clearly establishes the presence or a lack of a relationship which is the type of evidence needed to convince a group of individuals. This interactive scatterplot consists of markers along with the line of best fit to emphasize the relationship present (Figure 3.3.1). It has a tooltip to easily provide the user with further information and it is linked to the a "country of residence" filter. When a country is selected the marker changes color from blue to orange and converts to a square shape which gets enlarged. Size, color, location and shape are utilized to differentiate it from the rest.

After which, a heatmap is utilized to summarize and provide insights on the average monthly visits by different nationalities along with their seasonal behavior. The interactive heatmap is colour coded using a sequential colour scheme referred to as "Heat" (Refer figure 3.3.2). "Veridis" color palette was also considered due to its robustness even with those who are colorblind, but the "heat" palette was selected for two reasons. The requirement was to have a single colour palette which gradually darkens but does not portray multiple colours when doing so. The clear convergence from cream colour to dark brown highlights seasonality quite clearly.

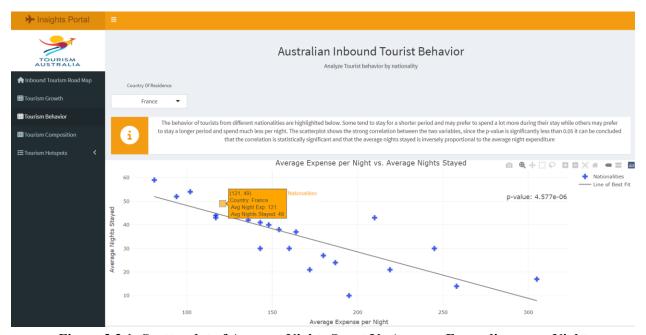


Figure 3.3.1: Scatterplot of Average Nights Spent Vs Average Expenditure per Night

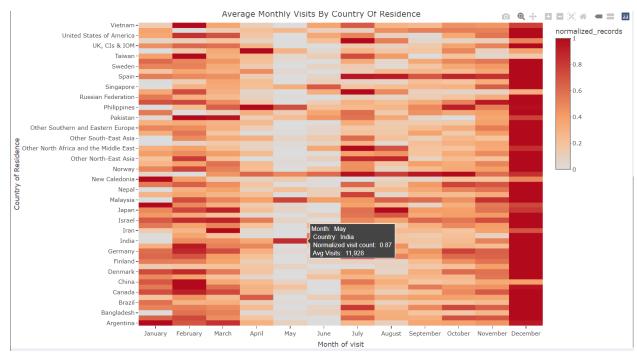


Figure 3.3.2: Heatmap of Average monthly visits by Country of Residence

3.4 Analyzing Inbound Tourist Composition by State

Deciding on a location can be one of the most important decisions one can make whether it's for a personal tour or for a business investment. In the tourism industry this decision can be either what are the tourist compositions within an area of interest or, if a specific target market has already decided, which areas have the highest compositions of them in order to invest in. This thought process led to realizing the need for a user to drill down both ways, one being based on nationality what are their preferred locations or based on a location to identify which nationalities are most available to cater to. Therefore, a choropleth map was used to visualize the tourist arrival density across the states (Refer figure 3.4.1) which highlights their preferred destinations. A sequential color pallete from RColourBrewer known as "YlOrBr" was utilized to highlight the density in the map. It was preferred due to its clear sequential shades and the sensitivity on the eyes. Moreover, this visualization is coupled with a bar graph that summarizes the total visits in 2019 by the country of residence. These interactive visualizations are linked to two filters, one being by "Country of Residence" and the other being "By State" to enable the user to drill down if needed (Refer figure 3.4.1 & 3.4.2).

A choropleth map was preferred due to its ability to summarize large amounts of information and in turn highlight patterns or areas of interest. It provides an effective and an efficient way of visually comparing values. The bar graph was coupled together due to its simplicity and its ability to summarize the leading countries, the ones at the bottom and the ones which stand out quickly and effectively. This is referred to as composition information. Its value addition is seen when it is utilized in unison with other filters and charts.



Figure 3.4.1: Tourist Arrival Year 2019

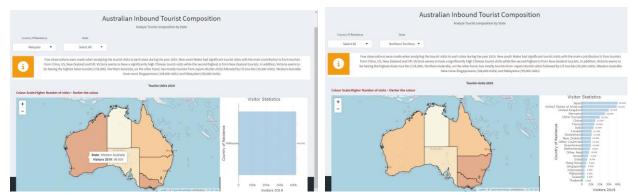


Figure 3.4.2: Malaysian Tourist Visits

Figure 3.4.1: Northern Territory Tourist visits 2019

3.5 Identifying Tourism Hotspots

When investing in a location within this industry is of paramount importance to consider the tourist content availability within the area. Sometimes the most cost-effective place might not be the best decision due to the lack of tourist content. This will tremendously lower the interest of their target market. Tourists in general prefer areas with proper food and drink, accommodation, events, and attractions. Investing in an area where these lack might not be fruit full. This section of the insights portal is dedicated to making use of and communicating insights generated from the data maintained by the Australian Data Warehouse.

The visualizations selected to the portray this information was a proportional symbol map to high light opportunities using shape, size, location and colour to effectively differentiate between the tourist content while highlighting hotspots of opportunities (Refer figure 3.5.1). The choropleth map is linked with both a highlight table summarizing the quantity and type of tourist content available at the location and a column graph summarizing the suburbs with the most tourists (Refer Figure 3.5.5). Further below, a highlight table

with a google panel like view of the granular details of the tourist content listing such as the title, address, phone number etc. are displayed for end user convenience (Refer figure 3.5.4).

In order to add in user convenience three dashboard filters have been incorporated. A "Tourist Content Category" filter which can concentrate all insights within the dashboard tab to focus on the specific content type. This can be useful for both tourists and as well as for business users within the same industry. Secondly, a "State filter was incorporated to identify the tourist content listings within the state and within each suburb. This was followed by the inclusion of a "Tourist Content Quantity" filter which enables the user to filter out locations by the quantity of listings available in that area (Refer figure 3.5.3).

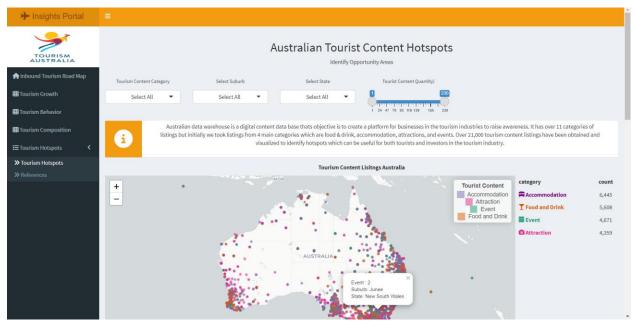


Figure 3.5.1: Tourism Content Listing Australia



Figure 3.5.2: Filter Adelaide

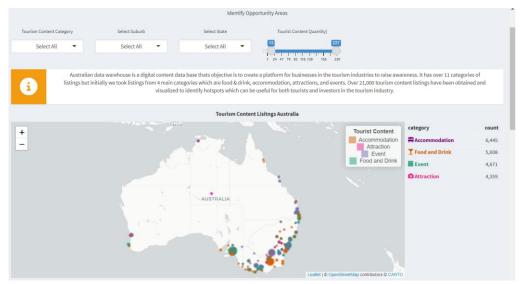


Figure 3.5.3: Filter by Tourism Content Quantity



Figure 3.5.4: Information Panel

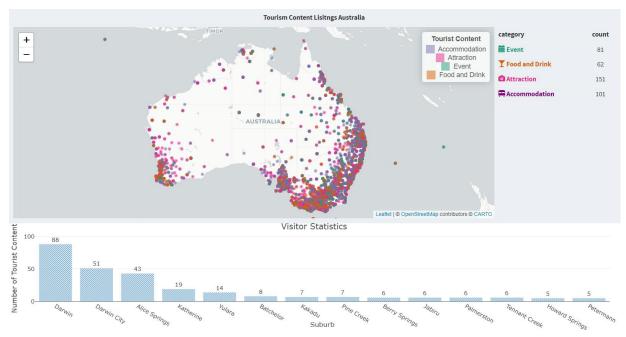


Figure 3.5.5: Filter Northern Territory

3. Insights Generated

3.1 Growth

Tourism in Australia has been growing significantly from 1994 (Approximately 2Mn), the nation was at its peak in 2019 reaching a rate of 9Mn tourist arrivals per year. When drilling down the next level of detail, tourists from all nationalities were increasing but Asian visitors have increased dramatically since 2009. When looking at the top 14 contributors to the Australian Tourism Industry in terms of annual visits, New Zealand and Chinese tourists have been increasing significantly year on year since 2004. Prior Covid impact, nearly 3Mn of the yearly visitors were from New Zealand, UK, Japan, United states, China and Singapore. When looking at the trend of tourist arrivals prior to the lock down, all nationalities seem to have experienced growth from 1994 to 2018 except for Japan which had experienced a drop of 42%. When looking at the past 10 years, inflow of tourists from all countries have increased except for Japan, Ireland, South Africa, Israel and Netherlands. In terms of covid recovery, China, India, Vietnam, Philippines and Sri Lanka are few of the countries which had recovered to 2008 annual visitor rates.

3.2 Behavior

As highlighted prior, humans are creatures of habit. Learning to understand their preferences better means the higher the chance of being able to deliver better products and services to their end. When considering the average night stay and the average night expenditure, a clear inverse relationship was identified. The p-value generated was significantly low amounting to 4.5 x 10^{\(^{\chi}\)} (-6). Hence, since the p-value is lower than 0.05, the hypothesis test can be concluded stating that there is sufficient evidence to reject the null hypothesis and conclude that the those who spend less during their visits tend to look for longer stay.

H₀: Those who spend less, do not prefer longer periods of stay Hₐ: Those who spend less, prefer longer periods of stay

Understanding, the seasonality of tourists from various nationalities will open up opportunities for product, service or pricing customization. It will enable businesses to develop products, offers and finalize prices that the end user will find more attractive. For instance by observing the scatterplot, it becomes clear that most Asians prefer longer stay for lower prices when compared with tourists from European countries. Observing the seasonality heatmap showed that majority of tourist arrivals are during the moths of February, March and December. Furthermore, highlights the average number of visitors for each of the months.

3.3 Composition

When considering the preferred destination of the tourists, without a doubt New South Wales and Victoria were identified as the most visited states. New South Wales had the highest tourist visits with mainly those from China, New Zealand, and UK. Victoria seems to have a significantly high Chinese tourist visits amounting to 667,000 while the second highest is from New Zealand amounting to 346,000 visits. Other Asian tourists seems to be highest in Victoria with 178,000 visits in 2019. The composition in the other states seem to be rather similar except for South Australia had the highest number Korean tourists amounting to 55,000 visits and UK tourists amounting to 65,000 visits. Northern Territory Australia contrastingly has mostly tourists from Japan (40,000 visits) followed by US tourists (39,000 visits). Western Australia shows a higher preference by Singaporeans (104,000 visits) and Malaysians (99,000 visits).

3.4 Hotspots

Majority of the hotspots are concentrated in the Eastern region of Australia, mostly in New South Wales and Victoria. The top suburbs include Adelaide, Sydney, Melbourne, Perth, Wagga Wagga, Hobart, Bright and Ballarat Central which includes the most popular city as assumed. When it comes to less popular states such as Northern Territory the top suburbs highlighted are Darwin (88 listings), Darwin city (51 listings) and Alice springs (43 listings) are potentially suburbs. The highest content listings for attractions are found in the suburbs Adelaide (Could be due to the size of the suburb), Darwin, Melbourne and Perth. Furthermore, the quantity of listings can be obtained for even the lesser-known suburbs such as Geraldton in Western Australia. It is said to have 18 listings. Like wise these insights in the hands of an investor with the flexibility to drill through for more information will promote data driven decision making.

4. Reflection

As all project developers experience, the most time consumption for the development of this project was at the stage of data gathering and data preparation. At this stage itself there were a few limitations faced such as the state wise tourist visitation by the country of residence was only available for two years. A much more accurate picture would have been derived if a larger history was available. In addition, the average nights spent, night expenditure and trip expenditure were for tourists based on country of residence. If this information had been obtained within each state the derived expenditure habits would be much more accurate. When considering the data structure, the small tourist inflow Asians countries were grouped

together as Other Asian countries which limits the granularity of the insights that be given to specific end user incase a request arises. Getting access to raw data will allow anyone interested to drill down as the require without any limitations. When it comes to the visualization aspect of the project, within a short period of time the development of a comprehensive dashboard using Rshiny was possible. There were few limitations in colour choices when it came to the design of the dashboard as the colours are limited by default in Rshiny. But the best colours taking into account the end user sensitivity on the eyes, the possibility of long work hours and colour blindness were taken into consideration were ever possible. When it comes to the end project I would further included interactivity in the Roadmap page to directly switch tabs based on the roadmap area of interest the user had clicked on.

5. Project Continuity

Six data sets were utilized in developing this project. Two tabular data sets from Australian Bureau of Statistics. One consisted of short-term movement visitors' arrivals by their country of residence from the year 1991-2022 tabular data (380 rows x 70 columns) while the other consisted of visitor arrivals by state from the year 1991-2022 tabular data (391 rows x 10 columns) which had to preprocessed utilizing R Loops to provide a simplified structure that allows for flexibility when visualizing. Further two additional data sets were utilized from Tourism Research Australia, one consisted of international visitors by nationality across Australian States and Territories (5 separate sheets of data) along with country of residence wise expenditure data. The last 2 data set were restructured to a single dataset to enable the development of a more comprehensive visualization. Finally, over 20,000 tourism content was web scraped via a web scraped tool known as Octoparse which resulted in spatial data set consisting of (21000 rows x 19 columns). The data set contained addresses from which postcodes, suburbs and state was extract prior geocoding. A further world region data set was scraped in order to provide a deeper analysis into world region wise tourism growth which led to the finding of the Asian tourist inflow is resulting in the dramatic increase in tourist arrival. The tedious data pre-processing stage results in three clean and structured data sets which the entire portal is built on. The relevant codes and programs are currently available in order to automate the process eventually to enable it to become the Australian Inbound Tourism Insights Portal.

5. Conclusion

The visualization approaches taken in this project have taken into consideration the requirements of the end user. A business user prioritizes time over everything else, therefore the final dashboard should be able to provide easy and quick navigation, clear and simple visualizations with the freedom to drill down as desired. The objective has been to maintain simplicity in the visualizations but deliver a complex dashboard from within that that allows flexibility to the end user. From the five design sheets the visualizations cater to this requirement were selected and arranged in a manner each complement one another. Hence, providing an insightful visualization. The development has ensured to address the four key areas identified as important the end user which includes understanding Australian inbound tourism in terms of growth, tourist behavior, composition by location and tourism hotspots to invest in. The coding and additional work put in are readily available for the next stage of development which is to provide the delivery for end user test and afterwards proceed with automation. Tourism industry has been growing in Australia, this project will support the business community in the tourism industry to move toward data driven decision making which will lead to the improvement in business value.

6. Bibliography

Australian Bureau of Statistics. (2022). Short-term Movement visitors Arrivals by their country of residence from the year 1991-2022 [Data set]. Retrieved from

 $\underline{https://www.abs.gov.au/statistics/industry/tourism-and-transport/overseas-arrivals-and-departures-australia/dec-2022/340105.xlsx$

Australian Bureau of Statistics. (2022). Short-term visitor arrivals by state from the year 1991-2022 [Data set]. Retrieved from https://www.abs.gov.au/statistics/industry/tourism-and-transport/overseas-arrivals-and-departures-australia/dec-2022/3401011.xlsx

Tourism Research Australia. (2020). International visitors by nationality across Australian States and Territories [Data set]. Retrieved from

https://www.tra.gov.au/ArticleDocuments/185/InternationalVisitorSurvey(IVS)EarlyReleaseSummaryJune2020.xlsx.aspx

Tourism Research Australia. (2020). International visitor average nights spent, night expenditure and trip expenditure by nationality [Data set]. Retrieved from https://www.tra.gov.au/ArticleDocuments/185/IVS_TOURISM_RESULTS_YE_MAR_2020.xlsx.aspx

Australian Tourist Data Warehouse. (n.d.). Tourist content listings [Data set]. Retrieved from https://atdw.com.au/our-listings/?pge=1

U.S. Department of State. (n.d.). Countries and world regions [Data set]. Retrieved from https://www.state.gov/countries-and-areas-list/