Visualisation Design Coursework 1

COMP0034 – Application Programming for Data Science

Table of Contents

Target audience

Main target audience:

- Hotel chain business owners

Other target audiences:

- Restaurant chain business owners
- Other hospitality business owners such as café owners

The target audience for the choropleth map is mainly hotel and hospitality business owners who are looking for finding a country to open a new hotel. This target audience wants to learn and compare different countries to see which countries have the highest number of international tourist arrivals as it is advantageous for them to open a hotel in those countries to increase revenue and profits (1) . They also would want to see the top countries in a specific region, for opening hotels in that specific region. This target audience is professional and may have a familiarity with the data since tourism is a crucial factor affecting the revenue of the hospitality industry (2).

Another goal of this dashboard in the future, after visualizing the necessary information for the best country and or/region to open their business location, is for hospitality business owners to have the chance to advertise and collaborate by posting their locations of, for example, any existing or planned hotel. This allows other hospitality businesses such as restaurants to contact the existing or planned business, who may have also used the dashboard to find the best country to open their business. This can allow the opening of a restaurant within or next to a hotel, thereby increasing profits and revenue for both (3).

All the visualisations use the same target audiences, with the same main goals outlined before and in the persona (in the Persona section below).

Target audience questions answered by entire app:

1. Which countries have the highest average number of arrivals over at least the last 10 recorded years?
2. For each year, which countries have the highest number of arrivals?
3. What are the rankings of countries in terms of the highest average number of arrivals over at least the last 10 recorded years?
4. How many arrivals are there in the countries with the highest number of arrivals each year?
5. What are the highest countries within each region (official groups of countries recognized by the world data bank (4)) for the highest number of tourist arrivals for any given year?
6. Are there any gaps or inconsistencies (possible anomalies) in number of arrivals over the years for different countries?
7. What are the trends (e.g. positive, negative or no correlation) over at least 10 years, for a particular country?
8. How do the trends (e.g. positive, negative or no correlation) compare between two countries?

Persona for main target audience: Hotel business owners

Figure 1 - Persona for Hotel Business Owners

Visualisation 1: Choropleth Map

1.1 Questions the visualisation is intended to address

1. For each year, which countries have the highest number of arrivals?
2. What are the highest countries within each region (official groups of countries recognized by the world data bank (4)) for the highest number of tourist arrivals for any given year?
3. How many arrivals are there in the countries with the highest number of arrivals each year?

1.2 Implemented Design Screenshot

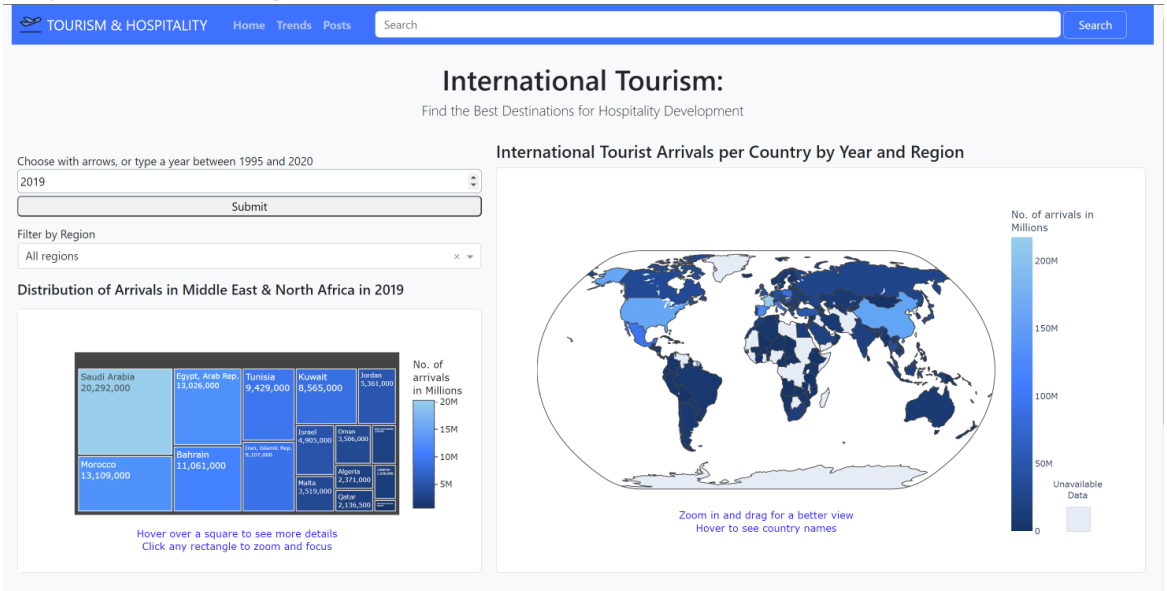


Figure 2 - Choropleth figure, also showing navigation bar

1.3 Explaining the design

A choropleth map was used as it is very powerful in utilizing a colour scale to represent the range of one type of data variable in a geographical context (5). That is true for this case, where there is one variable being represented: the number of arrivals in a specific year. In the original state of the figure without call-backs run, it shows the data for all regions unfiltered and shows how the number of arrivals is distributed among every country. The navigation bar chosen was a primary bootstrap blue. The main colour chosen for the colour scale was a custom range of blue shades, with the middle colour of the colour scale the exact same blue as the navigation bar. This was done first of all for visual consistency (6) as consistent colours are easier on the eye and therefore more effective at communication the desired information rather than distracting with colours. The second reason is that it has been found that blue conjures emotions of professionalism (7) and trust so is a good choice for a more professional target audience, like the ones I am targeting which are hotel business owners and any other hospitality business owners.

Two types of interactive functions were added, which change both the choropleth and the tree map (which will be discussed later). The first type was an input field for choosing a year between 1995 and 2020, which were the years of data available. The year can be either typed or the arrow keys can be clicked to choose a year. Then the submit button must be clicked for the output figure to change. This type of input was chosen over a year slider as the user can quickly type or choose a year of their choice which saves some time over sliding a slider over the large range of 25 years. Additionally, there was a

lot of data columns of 25 different years, which means that there would be visibly too many options across the span of the page. This does not look visually appealing. Also, that many options do not fit the screen in a responsive manner when the user uses a smaller screen device such as a phone as the text would be too close together (8). This was one of the user's requirements in the persona (Figure 1). Hence, due to these reasons, the input field was added, with a caption to inform the user the range of years, saving a lot of space as well as exhibiting very good responsive behaviour on smaller screens. The submit button outline was also made slightly darker to alert the user to click it for the charts to update the year.

The second type of interactivity added was the dropdown selector to choose a region from the 7 world data bank defined regions (4) . This was intended to answer the second target audience question, which allows the user to select a particular region. This type of selector was chosen as there are only 7 regions which is a good low amount of items in a dropdown to prevent the user being overwhelmed with options (9) . It was also used as it helps to preserve screen space and allow space for the other figure in the same column, the tree map.

It was decided to place the chart in one column that is slightly wider than the second column which contains the dropdown, which leaves enough space that for the choropleth to be viewed clearly. It also leaves some room for compromise to ensure the tree map in the column beside it can also be clear. The "natural earth" projection type was also used as it had an aesthetically pleasing round shape and was visually clearer than the other possible 2D projections, such as "mercator", for example shown in Figure 3 below. That type made the important continents of focus with most of the available data like Europe and Asia small but made Antarctica needlessly large, which is unnecessary since there was no data for Antarctica in the dataset.

Figure 3 - "mercator": Alternate projection example for choropleth (not used)

The columns were also set to update their layout on medium sized and smaller screens like tablets and phones, such that the entire second column with the interactive elements and tree map and would move position to be below the choropleth. 87% of employees in businesses are expected to use their mobile phones (10) so responsiveness is essential to our target audience who are business owners.

The map has a hover label that includes both the exact country name and informs the user of the number of arrivals in each country, therefore it intends to addresses the first target audience question.

An additional grey coloured shape (matching the missing countries on the choropleth) and an annotation was added to create a custom legend next to the normal legend of the colour scale. This was to inform the user that the grey sections were countries with missing data, which was a problem of the original dataset provided to me, but this was done to prevent misleading the user into thinking that colour represents the top countries for arrivals. This is also why I ensured the highest part of the colour

scale to represent those top countries was a bright blue, almost turquoise colour to allow it to stand out.

1.4 Evaluation of design

For the choropleth map, the dataset was for number of arrivals from the international tourism dataset which was pre-prepared.

This map effectively communicates the locations of top countries for number of tourist arrivals and it answers successfully the target audience question of which countries have the highest number of arrivals per year with the carefully varying colour scale. This scale is not only visually appealing with its consistency with the overall blue theme, but also conveys a sense of assurance and professionalism(7). Therefore. this colour scale is perfect for the very professional target audience. Additionally, the interactivity succeeds in allowing the user to change the year easily with buttons or the freedom to type and also allows freedom to choose a region which only shows the colour for countries in the chosen region. Consequently, this interactivity answers the first two target audience questions well.

The ability to hover is an extremely powerful feature as it allows the target audience user to gather an insight on exactly what country and its number of arrivals in that year is and thereby answers the third target audience question well.

A weakness of the map is that it doesn't have the ability to show labels of each country without hovering, which makes it hard to simply glance at the map and know which country is which for users without that knowledge. This weakness is more prevalent for a user using a phone where they don't have the option to simply hover over the country and see the name and details. However, this can be overcome by the user simply clicking each country and dragging and zooming in on the map to view a specific country. Also, it can be argued in the context of the target audience, an international hospitality business owner may already be familiar with the countries since international tourism is a crucial form of income for this sector (2).

Another weakness is the responsiveness of the choropleth figure itself when a large size screen is resized to be slightly less wide, it is the only chart that does not resize to fit the screen, despite having a responsive bootstrap container and relevant settings. I tried to fix this issue in many ways, but I could not find a resolution and found it to be a bug on the plotly side. Although it may look slightly unattractive on the small screens, which is what the user wanted in the persona, it wasn't a need. Furthermore, the user can still see the map on a small screen by simply dragging the map or the screen horizontally.

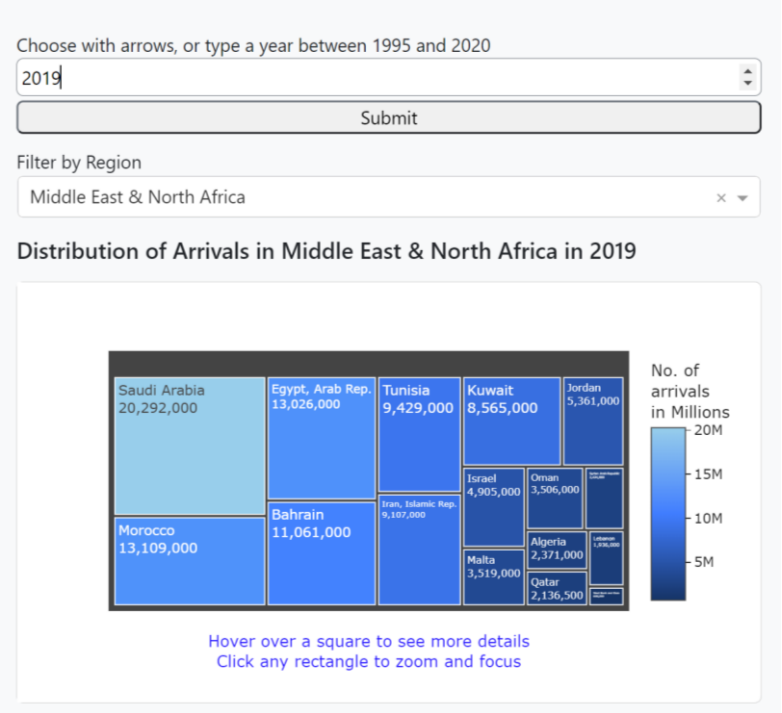
Additionally, in the dataset there were so many different countries (around 195). Similarly, the range of number of arrivals ranged from relatively smaller magnitudes of 100 thousands to relatively very high magnitudes of 100 millions Therefore, it was hard to easily distinguish the colours between countries that are in the lower quartile of this range. However, the aim of the countries is to convey the top countries in arrivals which are easily distinguishable by the light colours, so this is a good tradeoff. The weakness is only significant when "All regions" are selected, which shows every country on the same map but when each individual region is selected, the colour scale range adjusts better to rank between just the countries within a particular region, which is therefore still very useful. By highlighting these countries, this figure answers very well the first 2 target audience questions.

Visualisation 2: Tree map chart

2.1 Questions the visualisation is intended to address

1. For each year, which countries have the highest number of arrivals?
2. What are the highest countries within each region (official groups of countries recognized by the world data bank (4)) for the highest number of tourist arrivals for any given year?
3. How many arrivals are there in the countries with the highest number of arrivals each year?

2.2 Implemented Design Screenshot



2.3 Explaining the design

The call backs and interactivity options are connected to the choropleth figure in the first visualisation and so will not be repeated but they answer the same target audience questions as before and follow the same reasoning (see Visualisation 1 section).

The tree map was chosen to allow the user to better visualise the same information that the choropleth map uses, but in a simple and clearer way. It was chosen as this is one of the best ways to visually present many different tiles of data from tens to hundreds of data (11) . It therefore, exceeds the pie chart in its ability to present proportional data when there is a lot of data. This was the case with my dataset. The tree map filters and displays data for a selected region, but still has a lot of data in each region, so the pie chart was not used as the slices would be so small and converge to lines.

Another reason why the tree map was chosen is because it allows the powerful feature of being able to click on each square, consequently zooming into only that country and its data to effectively see the number of arrivals, even for a small square. It therefore intends to help the target audience hospitality business owners to choose a specific country in a more interactive and fluid manner.

The choice of colours of blue was again deliberate and compliments the choropleth with the exact same colour scale for the magnitude of number of arrivals. the

2.4 Evaluation of design

The ability to click individual square in this way, is extremely powerful as it acts as a visual aid to the choropleth map which could not show labels

A weakness of the tree map is when there are large amounts of data showing, the proportionally smaller squares do

Visualisation 3: Bar chart

3.1 Questions the visualisation is intended to address

1. Which countries have the highest average number of arrivals over at least the last 10 recorded years?

3.2 Implemented Design Screenshot

3.3 Explaining the design

The bar chart was chosen to represent the rankings as the bars make it easy to quicky and simply understand the hierarchy of a dataset, simply with the height of bars (12). The bar colour was chosen as the exact same blue colour as the navigation bar and the other charts on the page, in order to, again convey consistency, which is valuable in a dashboard to make it less visually distracting (13) . Additionally, the use of the blue colour was used as it generally evokes emotions of professionalism (7) and trust so is a good choice for a more professional target audience, like the ones I am targeting which are hotel business owners and any other hospitality business owners. Therefore, it is a suitable colour choice. Another smaller consideration for the reason why colour blue chosen for every chart rather than red, for example is to account for individuals with colour-blindness since red-green colour blindness is the most common form (14). A colour scale was not used as the heights of the bars are sufficient to show scale and ranking.

The bars were chosen to be laid out across the entire width of the page. This is to ensure when a high number of rankings is chosen, for example, 13 with the slider, the bars would appear larger and wider. For the smaller screens such as phones that the target audience would like for the app to be used on, this will look more visually appealing. Hence, this layout across the screen was used with a responsive container.

Interactivity was added in the form of a slider as this is a good way to represent a low number of choices, or rankings in this case in the full view of the user. The user can easily select which ranking they desire by sliding to it rather than having to select the number from a dropdown. Tooltips were added to make the chosen state of the graph output more obvious and clearer. The tick labels were removed off the x axis as they can be distracting (15) when there are words rather than numbers directly below the bars. The y axis tick labels were kept since they represent numerical data and make the positioning of the bars relative to the data it represents (number of arrivals) clearer to the hospitality business owner. The x axis title was also removed as it is obvious from the graph title and the bar labels themselves

what the bars represent, in this case the country names. There is also hover data which clearly shows the exact 10-year average of arrivals for each country bar.

3.4 Evaluation of design

The bar chart does an excellent job at answering the target audience question as it displays in terms of the height of the bars, the top countries for tourist arrivals. Therefore, it allows the hotel or hospitality business owner to efficiently choose a location for their hotel with a visual representation of the best locations. The slider feature is also powerful at formatting the rankings to show as many ranks as the user requires and fluidly updates the chart accordingly.

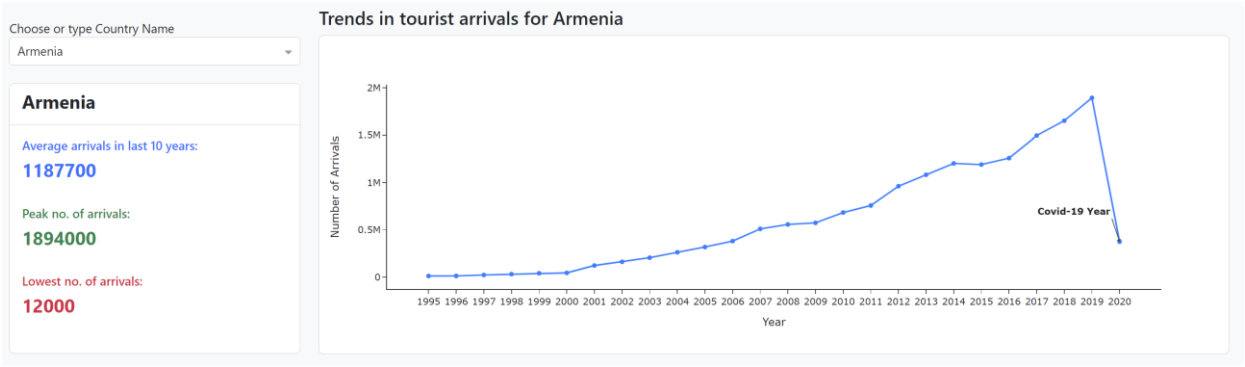
A weakness of the design is the fact that the user may have to hover over the bar to see the exact numbers which may be difficult for the target audience if they are using a phone or smaller device without a mouse.

Visualisation 4: Line chart with markers for one country and stats card

4.1 Questions the visualisation is intended to address

1. Are there any gaps or inconsistencies (possible anomalies) in number of arrivals over the years for different countries?
2. What are the trends (e.g. positive, negative or no correlation) over at least 10 years, for a particular country?

4.2 Implemented Design Screenshot



4.3 Explaining the design

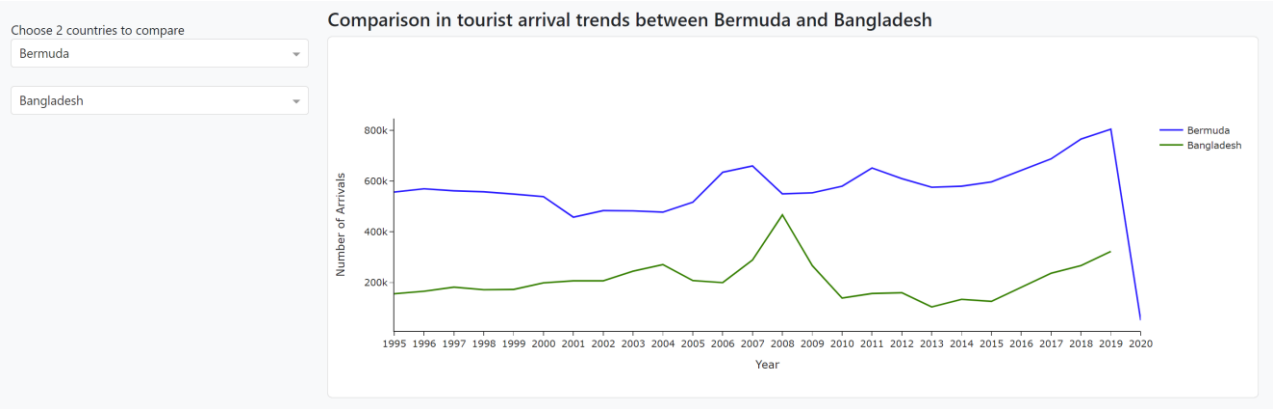
4.4 Evaluation of design

Visualisation 5: Line chart with for comparing 2 countries

4.1 Questions the visualisation is intended to address

1. How do the trends (e.g. positive, negative or no correlation) compare between two countries?
2. Are there any gaps or inconsistencies (possible anomalies) in number of arrivals over the years for different countries?

4.2 Implemented Design Screenshot



4.3 Explaining the design

4.4 Evaluation of design

The line chart for comparing 2 countries does a very good job at distinguis

A weakness of this chart design is the lack of comparative statistical data. Although it does visually show the differences in the trends like of two countries. For example, Libya has an overall negatively decreasing line slope which suggests negative correlation between the time period over the years compared to number of arrivals. However, in comparison to another country line like Curacao, we can see the line and points tend to increase in slope such that the two lines cross. This shows how the Curcao country is much better to put a hotel in terms of the target audience user as hotel owner since it has been proven to be an increasingly popular country so better to choose to build a hotel here for revenue and longevity. As well as the difference between two countries for number of arrivals each year, it also shows the difference in actual levels as tourism arrivals between two countries which further helps the target audience to decide, if the two countries have positive correlation and increase, the final choice of where to build a hotel is left down to which line is higher. This is because the higher line indicates more number of arrivals and hence answers the second target audience question extremely well.

A weakness is that a stats card could have been included to compare the percentage differences over the given year period in order to compare the countries better. This may have answered the target audience question of comparing trends between two countries better. However, this was decided not to be included to avoid repetition since the previous chart had a stats card.

Another weakness is that there could have been an annotation on the year 2020 to indicate a drop due to covid-19, which was indicated in the previous chart but in this chart the positioning of the marker would not position accurately, possibly due to there being 2 lines. However, this is the reasoning behind

there being 2 line charts as the first line chart in the previous visualisation addresses this issue. This lack of annotation may mislead the user if the have not used the previous one as they might wonder why there is a sudden drop in a positively increasing trend line.

A prevalent weakness with most visualisations that have multiple countries, is that the lines for countries with very low relative number of arrivals compared to a country with very high causes the lower line to appear quite squashed. This may misinform the hotel business owner when comparing trends as they may not be able to clearly see the increase or decreasing trend nature of a country and therefore slightly hinders the ability of the design to answer the first target audience question.

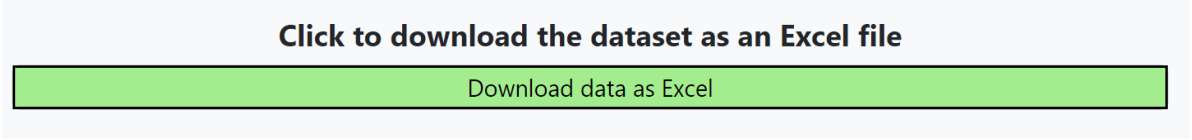
Additional features that answer target questions and user requirements

Additional feature 1 Download Button:

1.1. Questions the feature is intended to address

1. Can I download the data so I can analyze it myself or use it for methods such as machine learning algorithms to make predictions?

1.2 Implemented Design Screenshot



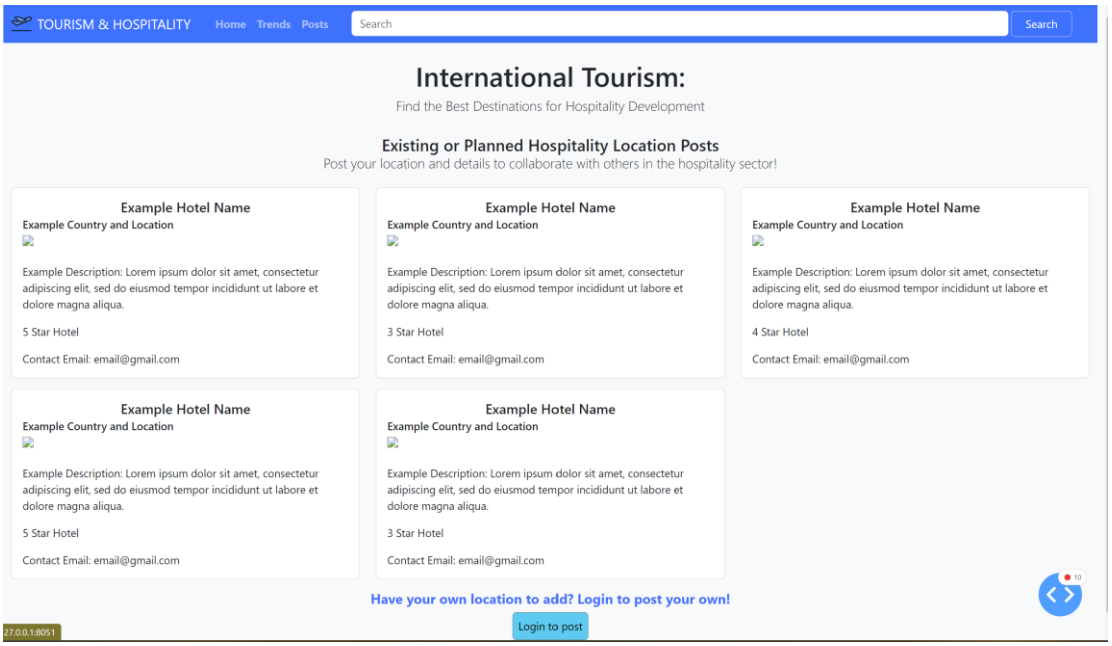
1.3 Explaining the design

1.4 Evaluation of design

Additional feature 2 Location posts page with working login functionality to access ability for user to post their own location

2.1 Questions the feature is intended to address

1.2 Implemented Design Screenshot:



1.3 Explaining the design

1.4 Evaluation of design

References:

1. Radu D, Huidumac C, Rossela N, Costel N. The Correlation between the Number of Tourists and the Number of Nights Spent in the Hotel: Analysis Indicator of Hotel Business Efficiency. Communications of the IBIMA [Internet]. IBIMA Publishing; 2010 [cited 2023 Feb 14]; 1–6. Available from: https://www.researchgate.net/publication/42831847_The_Correlation_between_the_Number_of_Tourists_and_the_Number_of_Nights_Spent_in_the_Hotel_-_Analysis_Indicator_of_Hotel_Business_Efficiency.

2. Travel, Tourism & Hospitality | Statista [Internet]. [cited 2023 Feb 15]. Available from: https://www.statista.com/markets/420/travel-tourism-hospitality/.

3. How a hotel restaurant can become profitable - The Business Journals [Internet]. [cited 2023 Feb 15]. Available from: https://www.bizjournals.com/bizjournals/how-to/growth-strategies/2013/11/heres-how-hotel-restaurants-can-go.html.

4. World regions according to the World Bank [Internet]. [cited 2023 Feb 14]. Available from: https://ourworldindata.org/grapher/world-regions-according-to-the-world-bank.

5. What Is a Choropleth Map and Why Are They Useful? - Population Education [Internet]. [cited 2023 Feb 14]. Available from: https://populationeducation.org/what-is-a-choropleth-map-and-why-are-they-useful/.

6. Qu Z, Hullman J. Keeping Multiple Views Consistent: Constraints, Validations, and Exceptions in Visualization Authoring. IEEE Trans Vis Comput Graph. IEEE Computer Society; 2018; 24(1):468–77.

7. The Psychology of Design: The Color Blue in Marketing & Branding [Internet]. [cited 2023 Feb 14]. Available from: https://www.impactplus.com/blog/the-psychology-of-design-the-color-blue-in-marketing-branding.

8. Should I Use a Slider on My Webpage? Slider Pros and Cons [Internet]. [cited 2023 Feb 14]. Available from: https://www.rcreative.marketing/wordpress/should-use-slider-webpage-pros-and-cons/.

9. When and How to Use Drop-down Menus | University IT [Internet]. [cited 2023 Feb 14]. Available from: https://uit.stanford.edu/blog/when-and-how-use-drop-down-menus.

10. Personal smartphones at work | CIO [Internet]. [cited 2023 Feb 15]. Available from: https://www.cio.com/article/236574/employees-feel-pressure-to-use-personal-smartphones-at-work.html.

11. Shneiderman B. Tree Visualization with Tree-Maps: 2-d Space-Filling Approach. ACM Transactions on Graphics (TOG) [Internet]. 1992 [cited 2023 Feb 15]; 11(1):92–9. Available from: https://towardsdatascience.com/treemaps-why-and-how-cfb1e1c863e8#:~:text=The%20big%20difference%20with%20pie,or%20may%20not%20be%20hierarchical.

12. Why Do Bar Charts Work? [Internet]. [cited 2023 Feb 15]. Available from: https://www.forbes.com/sites/evamurray/2019/04/04/why-do-bar-charts-work/.

13. Bach B, Freeman E, Abdul-Rahman A, Turkay C, Khan S, Fan Y, et al. Dashboard Design Patterns. IEEE Trans Vis Comput Graph. IEEE Computer Society; 2023; 29(1):342–52.

14. Types of Color Blindness | National Eye Institute [Internet]. [cited 2023 Feb 15]. Available from: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/color-blindness/types-color-blindness.

15. Talbot J, Lin S, Hanrahan P. An extension of Wilkinson's algorithm for positioning tick labels on axes. IEEE Trans Vis Comput Graph. 2010; 16(6):1036–43.

For why hotel business owners want countries with high level of tourism, the link between number of arrivals and hotels:

https://ibimapublishing.com/articles/CIBIMA/2010/813597/

For removing ticklines on categorical bar charts, also for omitting axis titles:

https://doi.org/10.1145/3056540.3064955