"Evaluating Ideal Study Destinations: an insightful data story"

Name: Taha Rizvi
Student ID: 35461411

Introduction

Education has the power to transform individuals, societies and the world at large. The significance of education is such that people do not mind travelling thousands of miles to learn and grow. In 2020 there were approximately 6.36 million international students; this number keeps on increasing with time. The aim of this data story is to be of use to if not all at least some of these students. There are several reasons to travel to a different country; it might be because the quality of education that is available in a country is unmatched, or sometimes students want to explore the world and not just be "a frog in the well", there are some privileged ones who just go abroad for the sake of it, to have a fun. No matter the reason behind the decision, studying in a different country is both emotionally and financially taxing.

If the decision is taken without proper research it comes back to haunt the student, because if they are not aware about the exact amount of money required to sustain themselves in the new country it overwhelms them when they face the reality. This leads to stress, anxiety and depression thereby affecting their academics. I intend to achieve the below mentioned goals through my data-story.

Goals:

Students and their parents get a proper approximation about the expenditure involved while pursuing education from a foreign country. Any user going through my data story should be able to choose between the countries to study in based on the rankings of the University. They should be able to calculate prices for grocery items, electricity, and transportation for several cities in different countries. I have also added filters throughout the story to help the user customize the visualization and get a clearer picture of what they are looking for. They should be able to know about the return on investment (R.O.I) that is amount of money invested in the studies and the earning potential after completing the degree.

Datasets:

Data Set 1: The Education Price Index
Source of Data Set: https://n26.com/en-eu/the-education-price-index
Rows*Columns=19*49
There are 15 columns with currency Euro, I have created new columns by converting the currency to US dollars so that this data can be used with reference to first data set. This dataset provides us the Tuition Fees for courses in various countries. It also provides us the salaries of the most common occupations when compared with cost of living dataset it can help us get some great insights.

The columns related to salary	Salary of a Doctor ($), Salary of a Nurse ($), Salary of a Lawyer ($), Salary of a Civil Engineer ($), Salary of a Software Developer ($), Salary of a Teacher ($)
The columns related to cost	Cost of Becoming a Doctor ($), Cost of Becoming a Nurse ($), Cost of Becoming a Lawyer $), Cost of Becoming a Civil Engineer ($), Cost of Becoming a Software Developer ($), Cost of Becoming a Teacher ($)

Data Set 2: Global Cost of Living

Source of Data Set: https://www.kaggle.com/

Rows*Columns=4957*58

Column/Features	Category based on the description
x1 to x3	Different types of Meals
x10 to x27	Grocery Items
x28 to x33	Travel Expenditure
x34 to x35	Cost of Cars
x36	Electricity
x37 to x41	Schooling for Kids
x42 to x47	Clothing
x48 to x55	Accommodation

Important points regarding the data sets:

1. For the purpose of this report we are considering the data from Education price index, so we have filtered out the same countries (i.e. 46 countries) from the cost of living dataset as well, selecting 3373 rows out of 4957 rows from the cost of living data set.
2. In case of Education price index data set there were few missing values I have imputed those values by searching the values from the internet.
3. Our datasets connect as through Education price index we can browse through salaries and tuition fees country wise, and once we have settled on a country we can use the Global cost of living dataset to do a detail analysis of the country through city wise data. The datasets join through the country column.
4. There were tilde operators (`) in-front of some city names they have been fixed. Other spellings were also rectified.

5. Remainder of the missing values in the data set have been imputed with median, as when skewness of the feature was calculated we could see the columns were highly skewed had they not been skewed we could have used mean values for imputation.

Personas

The personas developed for this report and data story are chosen such that each persona represents one class of audience out of primary (Ava), secondary (Liam) and tertiary (Maya). The audiences are explained below.

Primary Audience: Students who are deciding which country they should choose to study abroad.

Secondary Audience: The consultation Agencies who guide the students in choosing the correct location, they can provide them specific solutions based on the problems they might have identified for a location of their choice.

Tertiary Audience: Parents/ Sponsors for the students who will be aiding the students financially

****

Name: Ava

Age: 21

Education and employment: Ava is in the final year of her undergraduate degree. She is pursuing a 6-month internship with an IT firm.

Location: Bengaluru

Family Status: Ava lives with both her parents and a younger sibling.

Online behavior: Ava is online usually for educational purposes either watching videos to clear some concepts or solving online practice tests, she takes occasional breaks to browse through her social media accounts. She is not online for long stretches of time.

Goals: She wants to study abroad for her master's degree in Computer Science. Needs to make an informed decision based on factors like cost of living, tuition fees and potential job prospects in different countries.

Challenges: Due to financial constraints seeking a balance between affordable living cost and quality education.

****

SCENARIO I:

Ava is in the last year of completing her bachelor's degree; and wants to pursue her masters from a foreign country. She has been occupied with her final year project and an ongoing internship. Ava uses the available online resources to research about different countries. The large amount of data that she has to browse for selecting an appropriate location to study has overwhelmed her. As a student who is also an intern it is not possible for her to spare the time to do this research. To help Ava out an interactive data story can be developed that allows Ava to input her preferences or constraints and get the result based on that. This data story helps Ava to narrow down her options.

BASIC FLOW OF EVENTS /USE CASE I

1. Ava is given access to **Evaluating Ideal Study Destinations: an insightful data story**
2. Ava clicks on the on the first tab (University score tab).
3. She can see a world map with University scores; these scores are directly proportional to the quality of Universities present in the country.
4. Uses the Slider on the right to change the range of University scores, since good University is a priority she drags the lower range to 85 instead of the default 71.
5. The countries outside the range are not highlighted anymore.
6. Clicks on the second tab (tuition fee tab) to view the cost of one year at the university in different countries.
7. Clicks on the filter on the right and enters the lower range and upper range as per her budget.
8. Has a quick look on the tuition fee page, goes back and forth to the Score tab and tuition fee tab to find the perfect balance.
9. Clicks on the accommodation tab to check the variation in rents in different countries.
10. Clicks on the cost of grocery items tab, since she has already narrowed down the countries, she wants to know the cost of living in various cities of the country.
11. Ava is satisfied with the data she has obtained, she observed some additional metrics which she can also analyze if required.

SCENARIO II:

In this scenario Ava who is in the final year of her graduation has been granted 30% coverage on her living expense abroad for her master's degree by an organization. So now she wants to focus on the quality of education and the amount required for completing the degree. Cost of living is still a constraint but not the one with highest priority. She needs to find a country to study where total cost of a degree becomes the top criterion along with salary earned after completing the degree. Ava wants to strike a balance between total expenditure on University fees and the salary earned after the degree. **Evaluating Ideal Study Destinations** is the data story that she will use to sort out the countries as per her requirements.

BASIC FLOW OF EVENTS /USE CASE II

1. Ava accesses the interactive data story.
2. Ava clicks on the second tab to compare cost of a degree and salary earned.
3. Clicks on the filter on the right and selects Software Developer.
4. The visualization undergoes a transition.

5. Country wise data for both the metrics is available side by side, making it easier for Ava to sort out the countries of her choice.
6. Clicks on University scores tab.
7. She can see a world map with University scores; these scores are directly proportional to the quality of Universities present in the country. Uses the Slider on the right to change the range of University scores. In this scenario she has relatively better financial situation so she uses it to her advantage and filters the countries with range 90 to 100.
8. Clicks on the accommodation tab to see the city wise costs of rented apartments.
9. Clicks on cost of groceries to compare city wise difference in cost.
10. Makes the final decision as she has enough information for the same.
11. Closes the data story as she is satisfied with her shortlist.

****

Name: Liam

Age: 35

Education and employment: Liam holds a degree in International relations (BSc.) and currently works for Pathfinder Consultancy.

Location: Birmingham

Family Status: Married, doesn't have any children yet.

Online behavior: Liam is very active online, a large part of his job is to connect with clients over zoom calls and helping them with the options for specific countries to study in based on their requirements.

Goals: Provide personalized guidance to students by understanding their aspirations and financial situation.

Challenges: Staying up to date about accurate data on cost of living, tuition fees of Universities, job prospects in various countries.

****

SCENARIO III:

Liam works at Pathfinder Consultancy; he is one of the senior employees of the company. His colleague was supposed to give a seminar tomorrow on the variation in cost of food items across different countries/cities, the seminar aims at helping the parents of the students who are going for a degree abroad, away from any unexpected monetary burden as there have been cases where students have gone unprepared and then ended up stressed and depressed. His colleague informs him late at night about his unavailability for tomorrow's session and requests him to fill in for him. Liam has to make a presentation but there is not enough time for him to search the internet. An interactive data story can be used to help him out.

BASIC FLOW OF EVENTS /USE CASE III

1. Liam decides to use **Evaluating Ideal Study Destinations.**
2. Moves on to the fourth tab which has visualizations on cost of grocery items.
3. Two visualizations/charts appear on the screen.
4. One describes the total cost of grocery items city wise and the other shows the cost of individual item country wise(i.e. average of all prices of the cities in the specific country)
5. Clicks on the desired country from the filter on the right.
6. Visualization makes adjustment based on the selected country.
7. Liam takes several screenshots of the data for his presentation.
8. Clicks on one the specific city bar in the first bar chart.
9. Visualizations change in case of the second bar chart as the first one acts as a filter.
10. Liam is pretty happy with the detailed information he is able to find and continues to take screenshots and notes wherever necessary.
11. Liam browses through other tabs as well just in case there is a question from the audience.
12. Exits the data story as his presentation his ready.

SCENARIO IV:

In this scenario, Liam has an appointment with one of his clients who is planning to study abroad for his master's. His basic requirements are similar to what generally students ask for like ranking of the University should be good, the tuition fees should be within his budget, living expenses should be within a certain range. However this client made a special request from him, he explained that since he will be very busy with his studies, projects and course works more often than not, he won't be able to cook meals for himself and he is a big foodie, he wants to consider the cost of different meals at restaurants in various countries. Liam would have had to do a lot of research based on this request, but since Liam has access to the tool AcademiaPick Pro his task can be done quite easily.

BASIC FLOW OF EVENTS /USE CASE IV

1. Opens the data story Evaluating Ideal Study Destinations.
2. Click on the tab Restaurant Prices Country/City wise.
3. There are two charts visible, one representing the cost of a meal at the restaurant and the other representing individual choices.
4. Clicks on the drop down on the right to choose the Country.
5. Costs of meal at restaurant in different cities of the country appear.
6. Clicks on the meal type filter where he can choose between 3 options (Inexpensive meal, three course meal for 2 and a Mcmeal)

7. Clicks on the choice of drink to change the drink in the meal and thus sees a change in the prices of the meal.
8. Clicks on the Water filter to add water to the meal.
9. He takes down notes and screenshots wherever necessary and the moves on to tabs like Top University Score, Tuition fees to input the criteria provided by the client.
10. Liam has got his shortlist ready and he exits the data story.

****

Name: Maya

Age: 41

Education and employment: Maya has a master's degree in Psychology she currently works as student councelor for an International School.

Location: Pokhara, Nepal

Family Status: Separated, with one child

Online behavior: Maya is a modern age woman who is very tech-savvy, she is required to create presentations, excel sheets for her work. She is usually online for work.

Goals: To find a suitable country for her son to study in.

Challenges: Balancing her son's career aspirations and his hobbies and making an informed decision

****

SCENARIO V:

Maya works for an international school, she has had the opportunity to travel to different countries for her work, and she wants her son to have that exposure as well. She wants her son to pursue masters from a foreign country. She has an idea about some countries but wants to explore all the options before taking the decision. Due to her busy schedule she has hired a consulting agency for help. She has made it clear that she is hiring them to get the full service she wants all the details from cost of education, how much would grocery cost, return of investment i.e. cost of education vs. salary comparison, cost of accommodation. Money is no object for Maya she just wants the best education for her son. Also she wants to know the cost of fitness centers in various locations as her son is an athlete and she wants to make sure that there is no compromise in that area. One of the employees of the consultancy agency (say Liam) can use the data story to provide all details to Maya.

BASIC FLOW OF EVENTS /USE CASE V

1. Liam begins browsing through the data story.
2. The first tab is University scores, Liam begins the shortlisting process.
3. Liam then views the tuition fee tab, since Maya has no criteria regarding the fees, he doesn't need to use any filters.
4. Liam moves on to the next tab which compares the cost of completing a degree and the salary earned post degree completion.
5. He moves to the cost of grocery items and Restaurant prices tab to take note of various details he makes use of the filters for various customizations.
6. Since the fitness center costs were something Maya wanted to know , Liam finds these details in the last tab.
7. Liam has the required data to provide Maya.

SCENARIO VI:

In this scenario Maya since she has had the exposure of visiting different countries, she wants her son to pursue a degree from a foreign country and she wants to be there for him; she plans on moving along with her son. She feels she is capable enough to find a job in any country, so she wants the best country for her son. Since she will be moving along with her son accommodation prices, electric bills and the daily food items cost will have to be taken under consideration. Since she is pretty close to the field being a counselor, she has the data sets Education Price Index and Global cost of living available. But since the research part is difficult using an excel sheet; one of her colleagues suggested her to use the data story for a clear picture of the pricing in different countries.

BASIC FLOW OF EVENTS /USE CASE VI

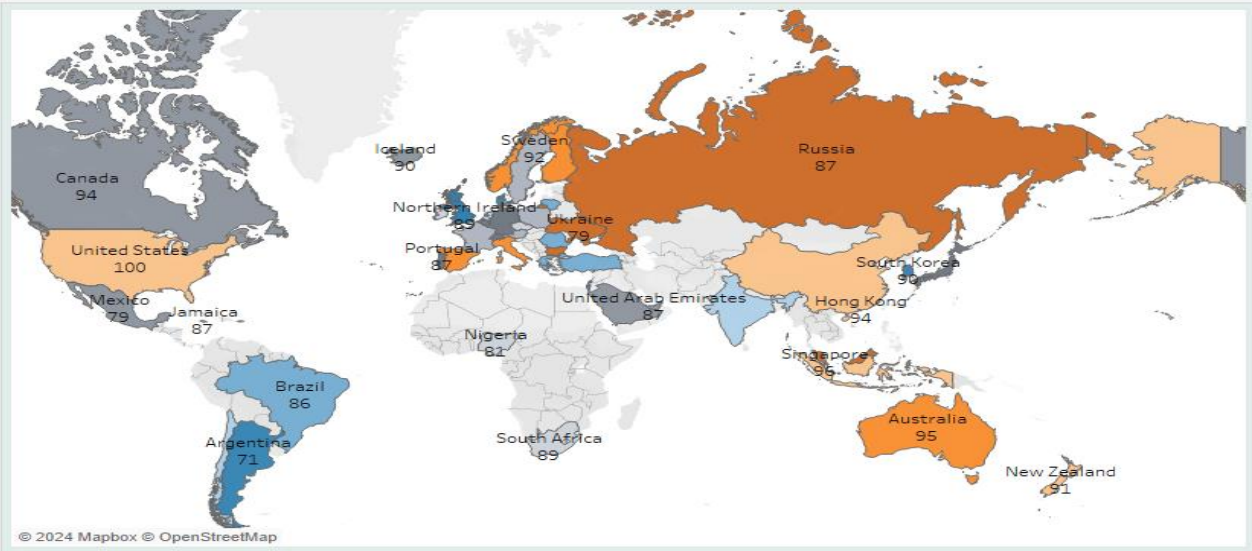
1. Maya browses through the Data Story.
2. Maya browses through the data story tabs; she begins with cost of grocery prices.
3. She checks the cost of grocery item prices she uses the country filter to see the prices in cities, on the bar chart below are depicted prices of individual items, she clicks on one of the bars of cities the visualization below changes to the cost of individual items for that city.
4. She goes ahead with exploring the transport expenses tab, where she can compare the travel costs through taxi, local transport etc. She can get the cost of cars in the area as well. She shortlists some countries based on the data.
5. After comparing the cost of living Maya moves to the University Scores tab to find the best rated countries based on her shortlist.
6. Maya clicks on the return on investment tab to see the comparison between cost of a degree and salary earned after the degree.

7. Finally after analyzing the accommodation tab and the electricity bills visualization in various cities she is able to make the decision.

Visualizations

The color palette used in the visualization is chosen in consideration with people who are color blind. Our story flows through several representations of the data set, for each visualization the graph selected has been done considering better expression of the data, ease of understanding and also the size of data has at times affected the type of chart chosen for representation.

Visualization 1:



The first tab of the data story represents Top Universities Score of the country. This Score is calculated on the basis of three factors:

- The university score which is highest in the country
- The sum of scores for top scoring universities in the country
- Average score of top scoring Universities, adjusted for the number of students enrolled [1].

This basically helps us to rank the countries according to the quality of Universities they have. The visualization has a filter which can help the user customize the highlighted countries on the basis of score on the slider.

I decided to go with map representation because I think it enhances the user experience, this could be done simply through a table or bar chart, but using the map representation provides ease of readability and

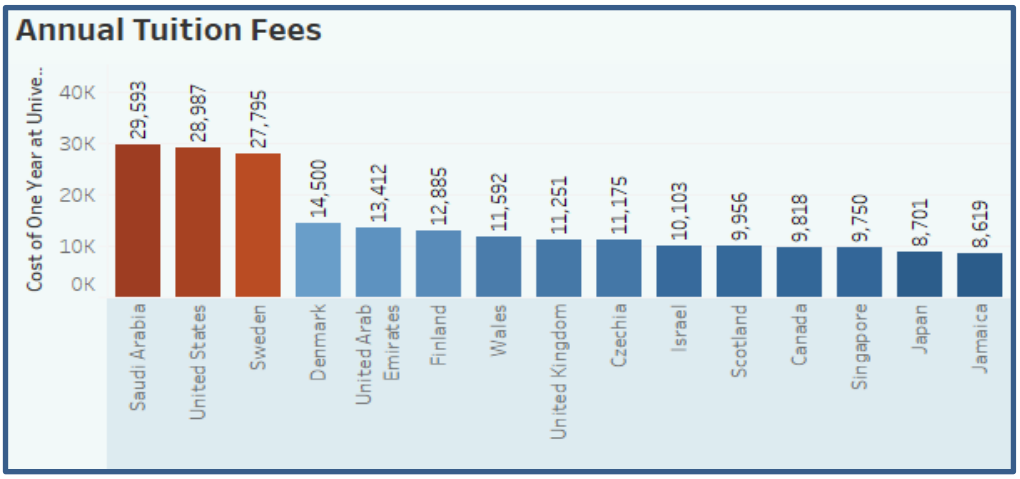
also it caters to the audience which for my data story is mainly the students. When researching about their dream country a student is very excited and good visuals make the experience more fun for them.

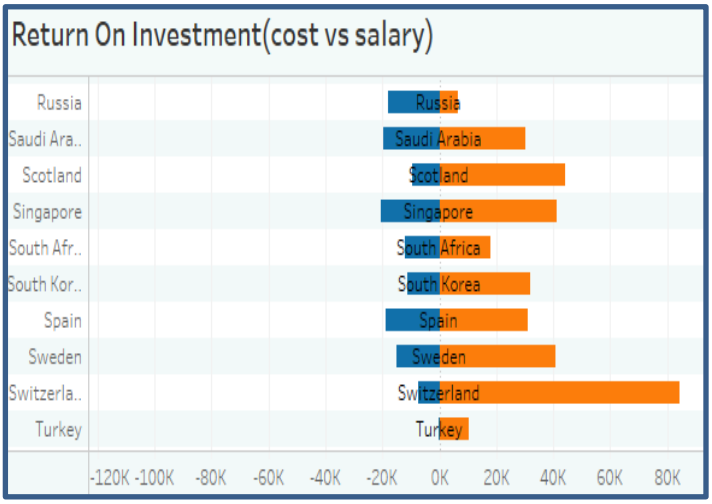
I have chosen this metrics to start my data story as this where generally students begin their search the university rankings.

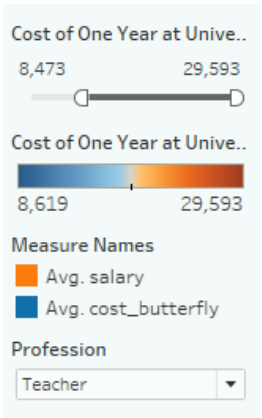
This visualization caters to the Scenario 1 and Scenario 2 as Ava is now focused on good quality education.

United States, Singapore, Australia, Switzerland and Germany are the top five countries based on the University score, whereas Bulgaria and Argentina have the lowest score. However the score cannot be the only criteria for selecting a country, we need to consider factors like cost of education, probable salaries once the student has completed their degree and the living expenses.

Visualization 2:







After checking the rankings, before users do any further investigation, they will have to check if they can afford studying in the Universities present in the countries of their choice. Hence the story moves on to

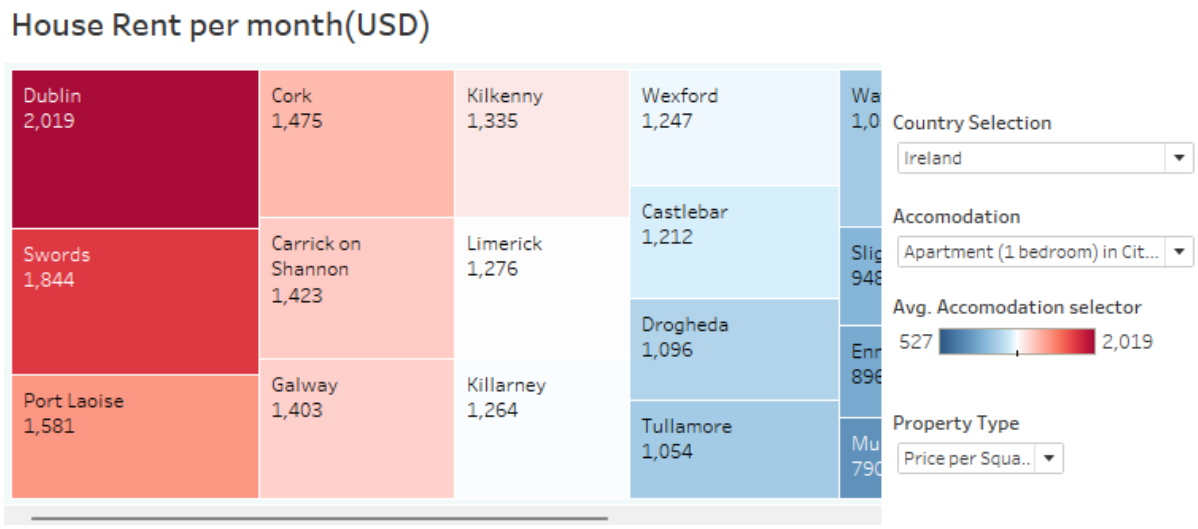
tuition fees visualization which is a bar chart representation which has been sorted in descending order. The color palette is chosen in such a way that brown represents the higher end of the tuition fees and blue the lower end. There is also a slider which can be used to filter countries based on the tuition fees. Columns used here are Cost of one year at the University ($), all columns for cost and salaries of different professions.

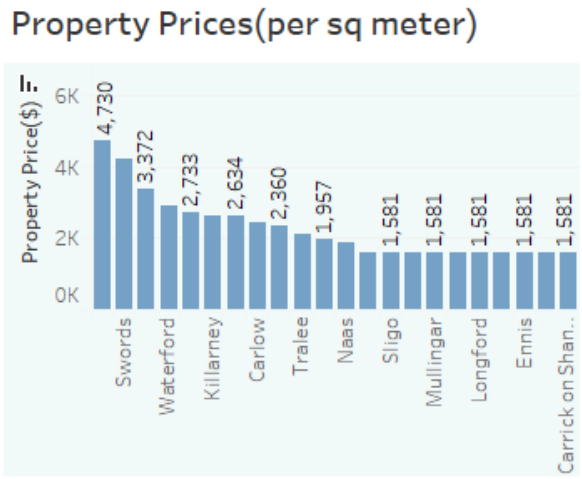
I felt that the tuition fee representation should be simple to understand hence bar chart was used. It is useful in scenarios I and IV. The second visualization used here is a butterfly chart.

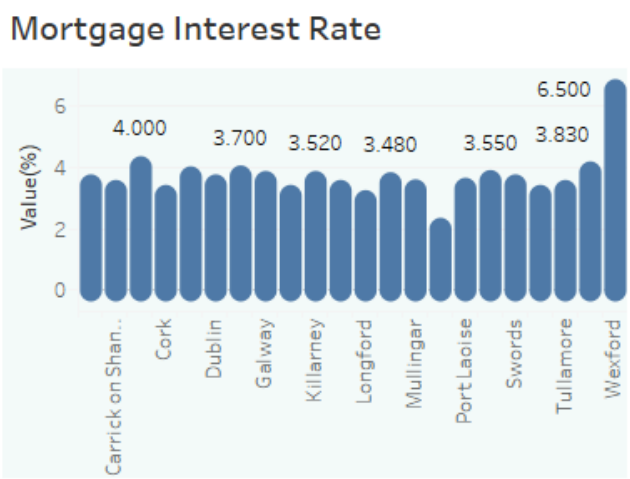
I have placed these visualizations together because after finding the tuition fees the obvious question a user will have is will their earnings after the course justify the amount they have spent.

The choice of butterfly chart is very intentional as I wanted the user to be able to compare both these parameters (cost and salary) instantly. Orange color is used to depict salary earned and blue is used to show the amount of money it takes to complete a degree. This is very useful for scenario II.

Visualization 3:



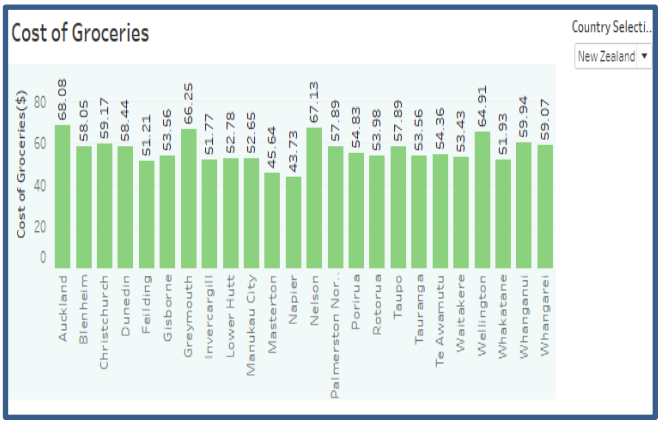


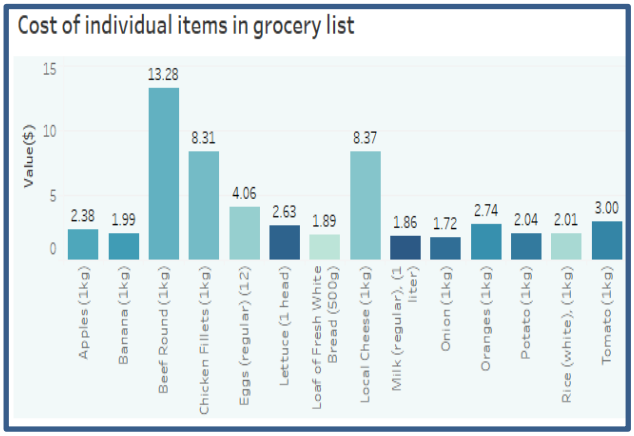


The flow of the story moves towards the accommodation visualizations, the thing that people will be most worried about is where they will live.

To represent the rent per month of apartments I have used tree maps,
The reason for selecting tree maps was that it just made it easier to understand the difference in rents of different cities. The size of the rectangles is proportional to the rent values which are pretty helpful in differentiating between rent of two cities. Bar chart representation was the option I chose for property prices and the mortgage rates, for mortgage edges I used curved edges in the bar chart to avoid monotony. It is used in scenario VI.

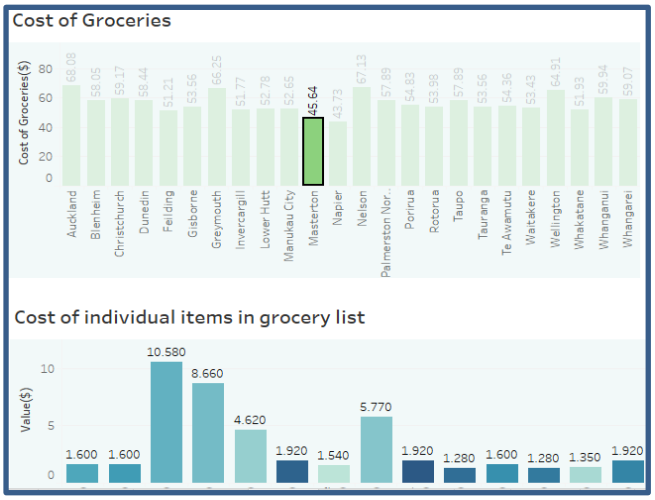
Visualization 4:





Our story so far has covered the quality of education, tuition fees, and return on investment. The basic human needs are food and shelter, out of these, shelter has been discussed. Now the next two tabs will discuss the food part. First we present the total cost of grocery items in different cities of a country. Country can be altered through the drop down parameter. Then the bar graph below shows us the average cost of individual items in the selected country.

The reason for selecting bar graphs was because I wanted to use the bars of the first graph as filters so once you select a bar i.e. select a city the second graph now gives you cost of items in that city. It is useful for scenarios IV and VI.

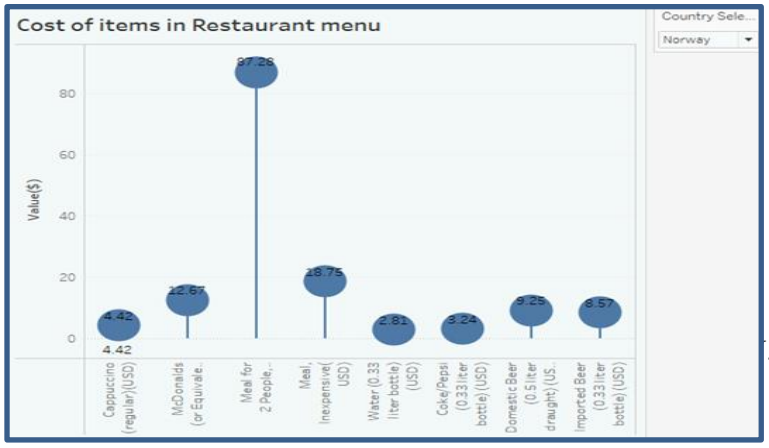


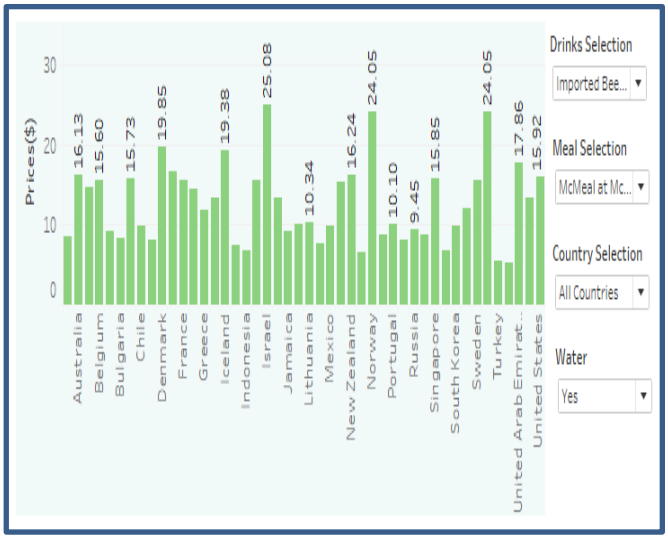
Visualization 5:

This visualization represents cost of eating in a restaurant in various cities. There is 1 bar chart and one lollipop chart, the second graph represents the total amount of money required to eat a meal in restaurant the first one represents the cost of individual items to give a more detailed view.

The filters present allow users to vary the type of meal and beverage they choose, and the filter for countries helps the user to look into various countries.

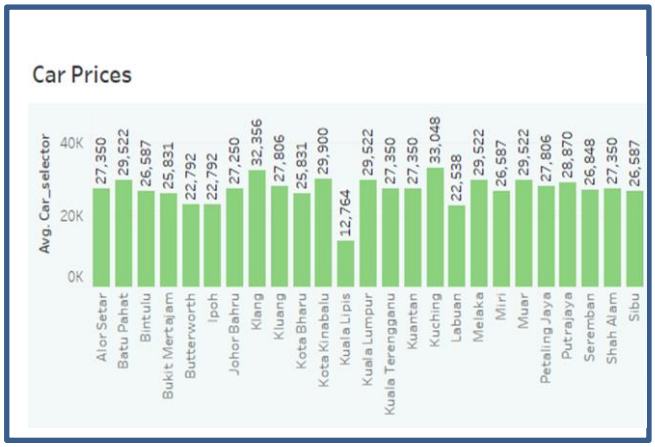
The data is similar to that of the grocery prices so I am using Bar chart and Lollipop charts here for a clear representation of prices.

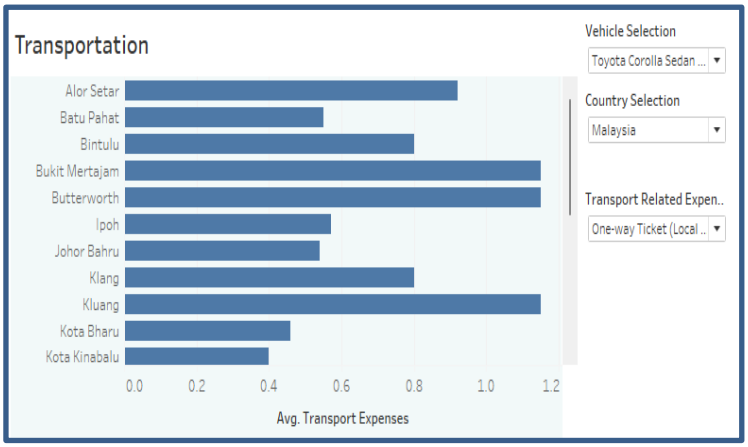




Visualization 6:

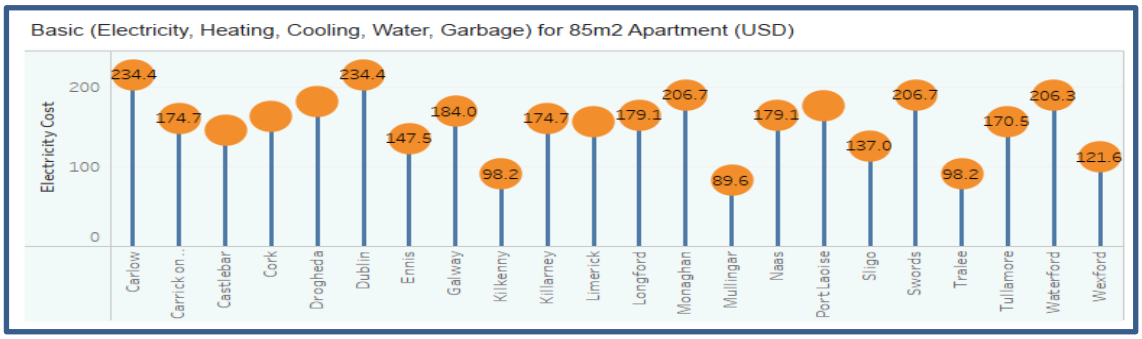
Our narrative has covered several important aspects required to sustain one-self in a foreign country. However we haven't yet seen the cost of transportation once you start living in the city. So this visualization Gives us an idea about the cost of local transports, travel pass and cost of cars in different cities. Through filters type of vehicle, mode of local transport and the country can be changed.





Bar charts have been used here because if we try to use a bubble chart or tree map in case when number of cities are high the labels don't appear which just makes it difficult to understand thus bar chart is preferred. These charts came in use in scenario VI.

Visualization 7:



Our story has been successful in covering majority of the criteria that are necessary but I have kept these two charts for the last because they are not so commonly considered. There are few people who would like to keep a track of the electricity charges for ex. Scenario VI where Maya is moving with her son or in scenario V where fitness club tracking was a request from the client.

Lollipop charts for electricity and bar chart for fitness clubs are used because due to the presence of large number of cities it is not visually appealing to use charts like tree map and bubble charts.

Conclusion:

Hopefully this data story was successful in taking its users through a structured journey where the end destination is a user who is well aware and prepared for the challenges he/she is about to face and thereby making them at ease with at least the financial aspect of the process. Everything from Country rankings (University Score), Tuition fees, Cost of a degree, potential salaries have been considered in the story. Through filters the aim was to cater to a wide audience because no two students are the same they have different personalities and they make different choices.

References:

1. n26.com. (n.d.). The Education Price Index 2022 — a Study by N26. [online] Available at: https://n26.com/en-eu/the-education-price-index.
2. Numbeo.com. (2009). Cost of Living. [online] Available at: https://www.numbeo.com.
3. whed.net. (n.d.). World Higher Education Database (WHED) Portal. [online] Available at: https://whed.net/home.php.