# Group 16
# Web Application Proposal

## Visualising London's international visitors


**Members:**

Aidan Coyne
Monkgogi Galeitsiwe
Oreeditse Mogobye
Raluca Paiajen
Tiffany Yu

# Table of Contents

# Introduction

## Problem statement

In 2016, the UK recorded 37.6 million overseas visitors (Office for National Statistics, 2017) that contributed £22.5 billion to the economy, around 53% of which was spent in London (Condor Ferries, 2020). While international visitors are a key part of London's economy, the tourism sector struggles to use data analytics to share information effectively (London First, EY, 2019). The International Passenger Survey provides extensive data on international visitors, but the information is not widely available in the form of data visualisations. Additionally, the visualisations that do exist are limited. They do not segment international visitors (e.g. by travel purpose or country of origin) or only show information relating to the top ten countries where people are coming to London (Budget Your Trip, 2020) (Champion Traveler, 2020).

## Opportunity

There is an opportunity to improve data analytics and provide more channels to convey information about overseas visitors in London. A solution would be designing data visualisations that represent trends in overall overseas visitors to London and allow for data segmentation using appropriate filters. Additionally, all countries where visitors come from can be encapsulated using dynamic interactive filtering in visualisations to capture more information than currently available. The information would be available on a website, thus making information easily accessible to a range of different audiences.

## Target audience

- Anyone interested in the demographics and specificities of international visitors in London
- Businesses whose target audiences include international travellers in London
- The government who must assess how effective London is in attracting visits, for different purposes, and among different demographic groups to effectively develop policy
- International individuals who are planning on traveling to London
- Consultants doing projects that require a general overview about visitors in London
- Students/academics doing academic research that require literature on visitors in London

## Goal

To create a website with interactive and engaging data visualisations to help relevant audiences (e.g. tourism businesses and travellers) better understand the patterns of overseas visitors to London. Also, to create an environment where users can share thoughts and ideas on visualisations and interact with other users.

## Aims

- Users of the website will better understand where visitors in London are travelling from, why people visit, how long visitors are staying, and how much they spend per stay duration.
- UK businesses will better understand the average amount of money a person from each country spends when visiting London.
- Transport agencies will better understand how people are traveling to London and be able to adapt for future scenarios.
- International travellers to London will be able to better predict their trip budget.
- Those working in the airline industry, hotel industry, tourism industry, and other relevant industries can easily inform and update themselves on visitor trends.
- Users of the website can share ideas about each visualisation and share information about themselves, fostering an interactive and personal user experience.

## Smart objectives

- By November 30th, our dataset will be clean of NaN values and only columns related to the variables: year, quarter, country of origin, duration of stay, mode, purpose, number of visits, amount spent, and nights spent will remain.
- By February 15th (reading week) the test coverage for the code relating to our application will be 90% and by March 26th it will be 95%.
- By February 15th, our relational database will be populated by 100% of the datapoints from our cleaned dataset.
- By March 26th (or before the due date), the mobile-friendly web application will be fully functional and all of our 'must have' requirements will be implemented.
- By March 26th (or before the due date), we will have written at least 75% of our 'Should have' requirements.
- By March 26th (or before the due date), our application will have been tested by at least 10 students and be ready to be deployed.

# Requirements elicitation

A software requirement is a condition or capability needed by the user to solve a problem or archive an objective (Visual Paradigm, 2020). Requirement analysis involves constant communication with the users and client to define the specific features they require. However, we do not have a real client or users. Thus, we used observation and brainstorming elicitation techniques to identify user requirements for our project. As a first step, we designed a context diagram to help define the scope of our system.

## Context diagram

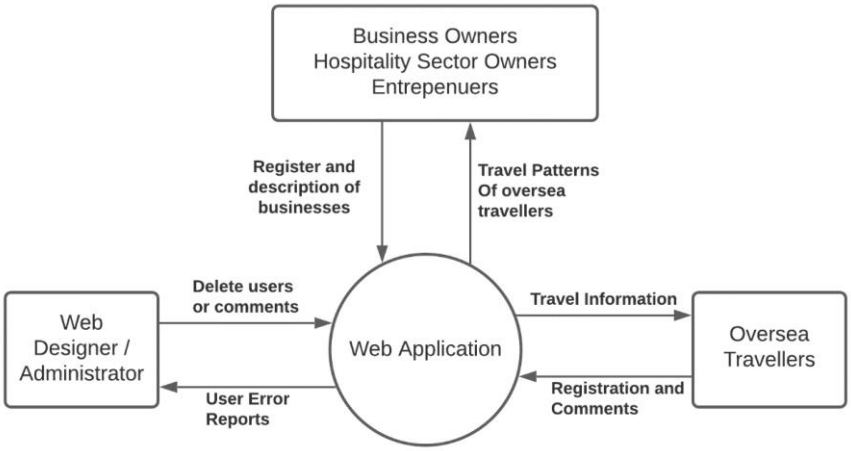



*Figure 1: Context Diagram*

After agreeing on the boundary for our web application, we gathered requirements through brainstorming and observation.

## Brainstorming

Brainstorming can be to generate ideas and create a consensus on an idea that will work to solve a problem (Sonego, Brainstorming, 2017). We brainstormed user requirements as a group by putting virtual sticky notes on a virtual white board using Padlet (Appendix A). Then we went through each sticky note individually to reach a consensus on whether to add it to our user stories. Brainstorming as an initial elicitation technique was quite effective as it allowed for virtual collaboration and helped us generate numerous ideas at once. Through brainstorming, we came up with a variety of features to make our data visualisations more interactive and user-friendly (e.g. drop-down menus for data filtering instead of tabs).

### Observation

Observation is an elicitation technique that requires a member of the team to navigate an online system with similar features, while the rest of the team observes. Our team had one individual navigate and understand the data visualisation for Top 500 Passwords on the website Information is Beautiful (McCandless & Quick, 2020). Information is Beautiful was the ideal platform to observe as we wanted the primary focus of our web application to be around data visualisations. We used the graphic for Top 500 Passwords because it had similar interactive filters that we wanted to implement in our application. We adopted a passive approach where we only logged the individual's actions and difficulties while they spoke aloud. From this process, we observed that:

1. The infographic was too complicated with three axes. It should be easier to understand
2. The user wanted to comment directly on the page with the graphic instead of sharing to social media
3. If we listed all filters at the top it would be too crowded. We should have a drop-down menu instead

## Requirements specification

### User stories

Consistent with the XP method that we propose to follow, the requirements have been captured as user stories. A user story describes a specific use case (what a user wants and why) in the perspective of an end-user, to ensure the users' interests are at the core of the design. The syntax and format of the categorised cells help developers digest necessary information (what, why, for who) about various use cases (Bondarenko, 2017). Initially, we categorised user stories by user (e.g. general user, tourist, web designer, business owner), which encouraged us to think in the perspective of the user and specify more requirements. However, we found that categorising by functionality was more comprehensible and intuitive.

### Prioritising the user stories

When running a project with a fixed time limit, it is important to prioritise the user stories and develop the most important ones that reflect the primary objectives of the web application.

For this project we have decided to use the MoSCoW technique to help us categorize each user story's importance. This technique is relatively simple but effective in projects with straightforward aims and requirements, making is suitable for our project. We chose MoSCoW over planning poker, because our inexperience made it difficult to assign numeric values to each requirement.

MoSCoW divides requirements into the following categories (Agile Business Consortium, 2019):

- **M**ust Have (MH - green) – vital requirements that are necessary to meet project aims, objectives and business needs.
- **S**hould Have (SH - blue) – requirements which are important but not vital. The software solution is still viable without them.
- **C**ould Have (CH - yellow) – desirable features that will have minimal impact if they are not included.
- **W**on't Have this time (WH - red) – lowest impact requirements that the team have decided to completely omit due to limitations in timeframe.

| RL | | **Register and login** |
|---|---|---|
| 1 | MH | As a user I want to be able to create an account quickly and easily to start using certain functions of the website, e.g. commenting and saving data visualisations. |
| 2 | MH | As a user, I want to be able to reset my password if I forgot it or want to change it. |
| 3 | MH | As a user, I want form validation when I register my account (error messages when username, password, or email is invalid) to speed up the process of registering. |
| 4 | MH | As a web application designer, I want the user to be able to create a strong password and can verify it, to meet the requirements of the data protection. |
| 5 | CH | As a user, I want the website to remember my username to make it easier to login when I revisit the website. |

| UP | | **User profile** |
|---|---|---|
| 1 | MH | As a business owner, I want to be able to create a profile and upload a description to promote my business when I am active on the platform. |
| 2 | MH | As a user, I want to be able to revisit saved data visualisations through my profile so that I won't have to manually filter graphs every time I revisit the page. |
| 3 | MH | As a consultant, I want to be able to revisit saved graphs so I can load them quickly for presentations to clients. |
| 4 | MH | As a user, I want to be able to visit other users' profiles and read their descriptions to know more about the users who posted certain comments. |
| 5 | MH | As a user, I want to be able to change my profile at any time to update relevant information. |
| 6 | WH | As a business owner, I want to be able to embed social media (Twitter feed, Instagram posts etc.) on my profile to promote my brand. |

| CO | | **Comments** |
|---|---|---|
| 1 | MH | As an administrator, I want to delete comments that go against community guidelines. |
| 2 | MH | As a student, I want to be able to comment on graphs to ask for opinions that could be useful to a report I am writing. |
| 3 | MH | As a tourist interested in travelling to London, I want to be able to read and share interesting discussion about visitor spending, tourism industries etc. that may be relevant to my trip. |
| 4 | MH | As a business owner, I want to be able to view discussion on each data visualisation to see how other business owners are interpreting the data and applying it into their business plans. |

| | | |
|---|---|---|
| 5 | SH | As a user, I want to be able to upvote interesting comments that I like to engage with the community on the platform. |
| 6 | SH | As a user, I want to be able to see the comments with the most upvotes first so I can see the popular comments first. |
| 7 | SH | As a young user, the app should be able to able to automatically filter vulgar language to protect me and other users. |
| 8 | SH | As a user, I want to be able to reply to comments to create a thread about a particular discussion. |

| DV | | **Data Visualisation** |
|---|---|---|
| 1 | MH | As a user, I want the data visualisations to be easy to navigate and interactive. |
| 2 | MH | As a restaurant owner in a tourist heavy area, I want to know where tourists are coming from so that I can change my menu according to their tastes. |
| 3 | MH | As a transportation investor, I want to be able to see what transport people are using to come to London and whether it differs by their purpose of stay. |
| 4 | MH | As a consultant, I want to view different charts and graphs to communicate visitor trends in a consultancy project. |
| 5 | MH | As a student/academic, I want to save and access different charts and graphs to use for an academic report. |
| 6 | CH | As a consultant, I want to be able to view a side-by-side comparison of different charts to communicate key differences between markets in a presentation for clients. |

| GE | | **General Functionality** |
|---|---|---|
| 1 | MH | As a user, I want to be able to access the service using my mobile phone, tablet or PC so that I can use whichever device I have at the time or the one is most confident in using. |
| 2 | MH | As a web designer I want the system and service to be intuitive and contain sufficient guidance such that clients can use it without additional help from the web designer. |
| 3 | MH | As a user, I want to be able to easily contact the developers to report errors or inappropriate activity on the website. |
| 4 | MH | As an administrator, I want error messages to be vague to improve our security. |
| 5 | WH | As a user, I want to be able to change the language of the website to a preferred language. |

| MI | | **Miscellaneous Functionality** |
|---|---|---|
| 1 | CH | As a business owner, I want to be able to have a section on my profile specifically for hyperlinking my website to direct other users to it. |
| 2 | WH | As a business owner, I want to be able to privately message other business owners I found in the comments section to gain insight on their strategy. |
| 3 | WH | As a business owner, I want to be able to run ads next to relevant visualisations, e.g. tour bus ads next to graphs about transportation methods. |

## Use case diagram

A Unified Modelling Language (UML) use case diagram is the primary form to represent requirements in software development. It specifies how the application is expected to behave and helps developers design from the user's perspective. Our use case diagram helped provide a visual summary of the web application design from a high-level of abstraction. Furthermore, it helped us contextualise our system and drive the design for our UML class diagram.

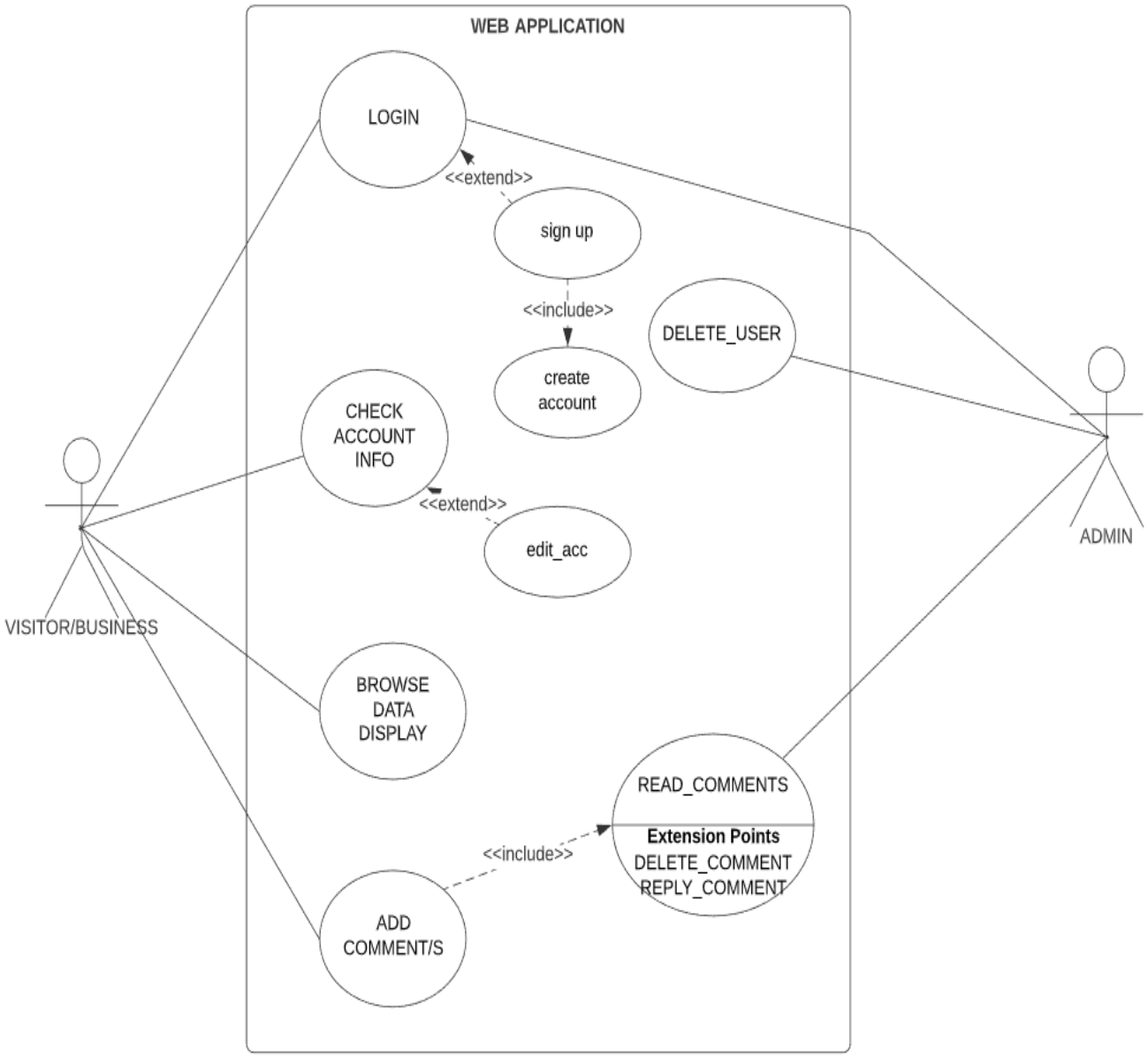


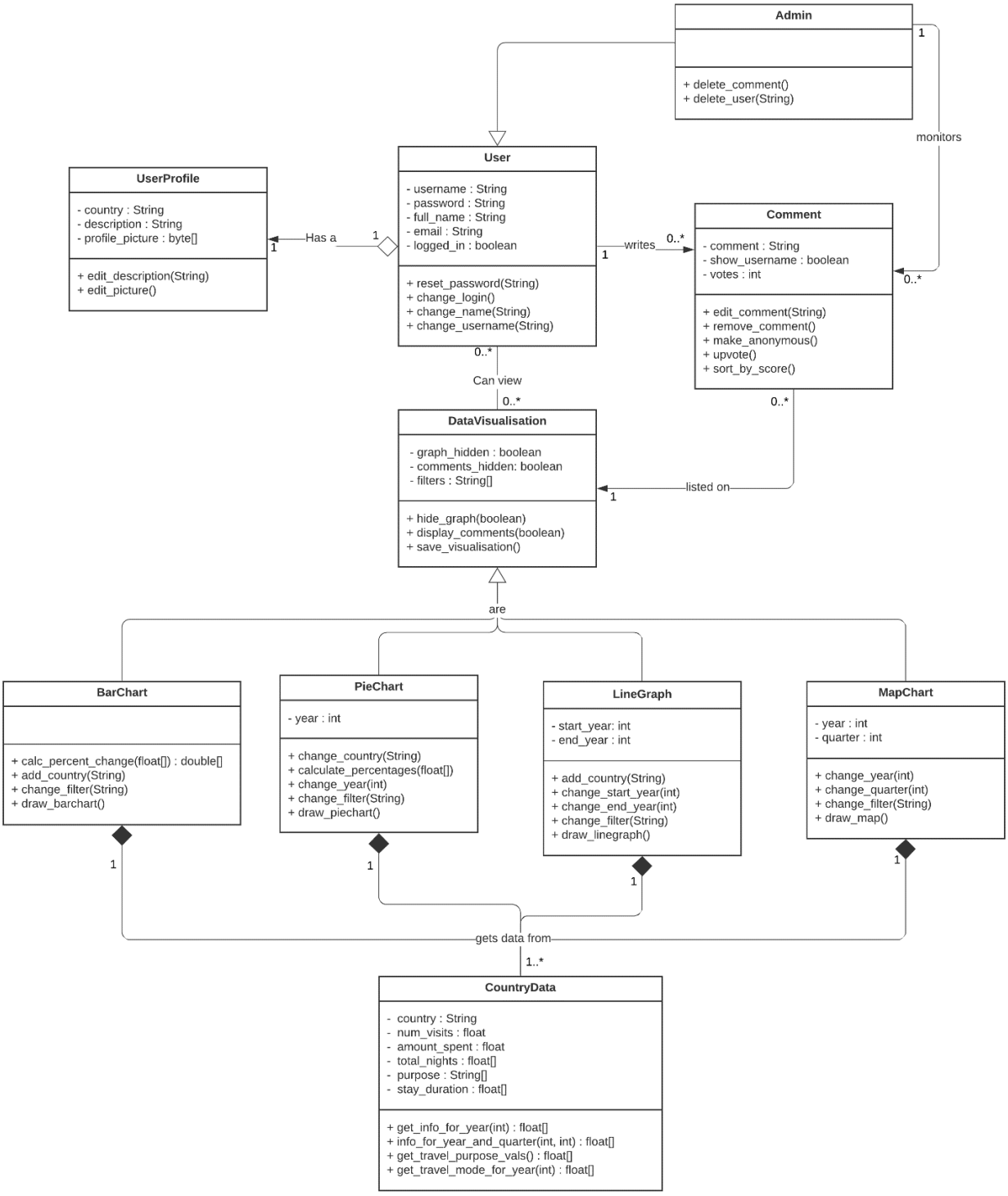

*Figure 2: Use Case Diagram*

## Class diagram

Our class diagram (Figure 3) below captures the objects and relationships within our web application. Our application will be structured into 10 different classes and adheres to object-oriented design principles such as completeness, sufficiency, primitiveness, high cohesion, and low coupling. For example, to maintain high cohesion we separated DataVisualisation into several helper classes – MapChart, LineGraph, PieChart, and BarChart. They each have their own specific methods and their own draw functions to make the classes specific and focused.

We additionally considered low cohesion through the use of a composition and aggregation relationships between DataVisualisation and the four chart type classes and User and UserProfile respectively. However, we chose to use inheritance instead of aggregation where it was appropriate. We have Admin inherit User because an Admin 'is a' user. Similarly, each individual chart (pie chart, line graph, etc.) 'is a' data visualisation. Each of the inherited classes will be using all the functionality of their parent class and nothing will need to be changed. While using inheritance goes against the principle of low cohesion, it was the appropriate design choice for our application.

*Figure 3: UML Class Diagram*



## Activity diagram

The following activity diagram (Figure 4) details one of the main features of the website: viewing and saving a data visualisation. More activity diagrams detailing core functionality can be viewed in Appendix C.

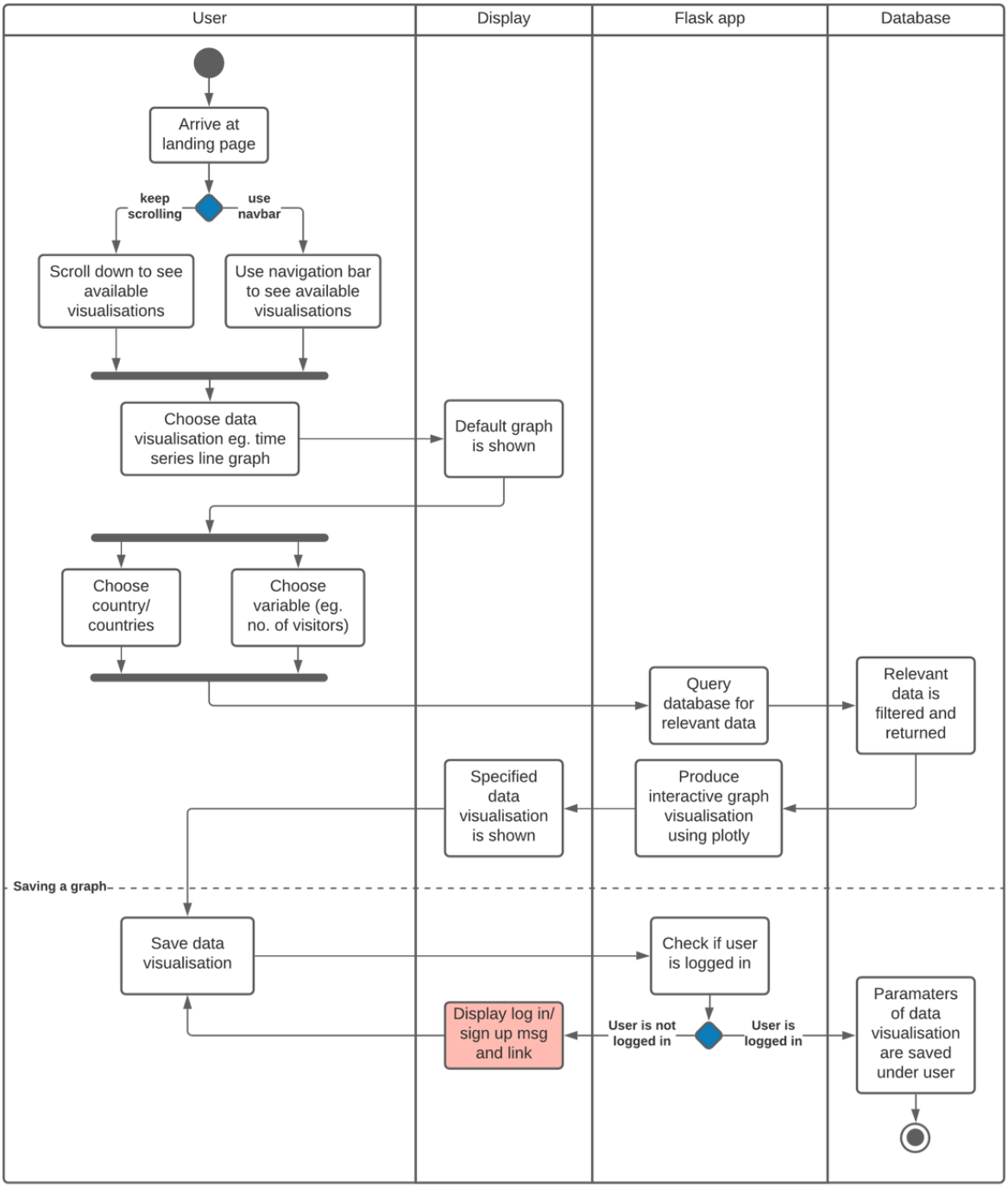



*Figure 4: Activity diagram for viewing and saving a data visualisation*

# Interface design

## Wireframes

The Wireframes are designed for 1920 x 1080 pixels, 16:9 aspect ratio desktop screens. We chose to represent this aspect ratio as it is the most common resolution in the world (StatCounter, 2020).

More wireframes detailing "About", "Contact", "Create account", "Sign into your account" and "Forgot your password" pages can be found in Appendix D.

### Home and landing page

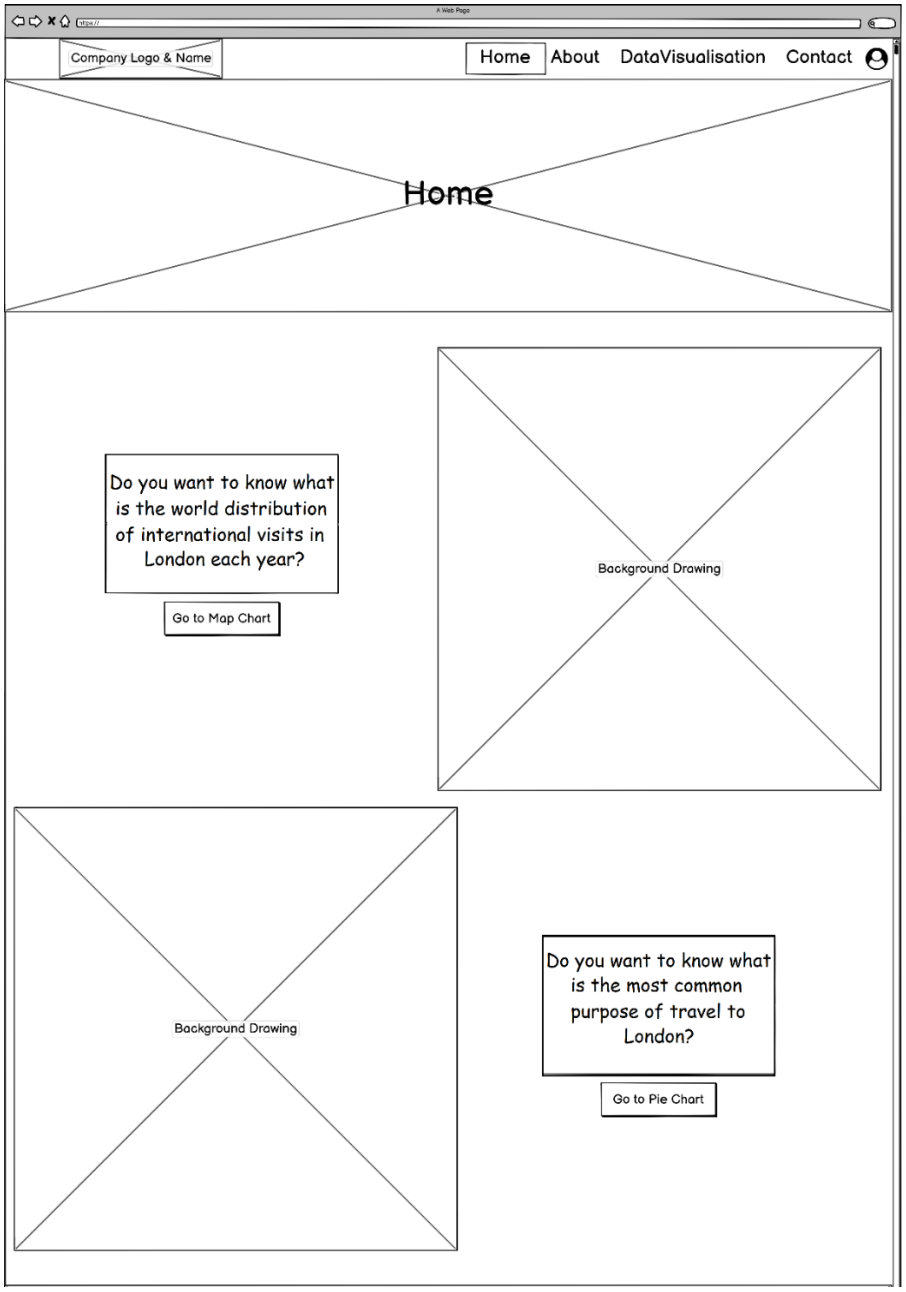



*Figure 5: Home and Landing Page - Part 1*

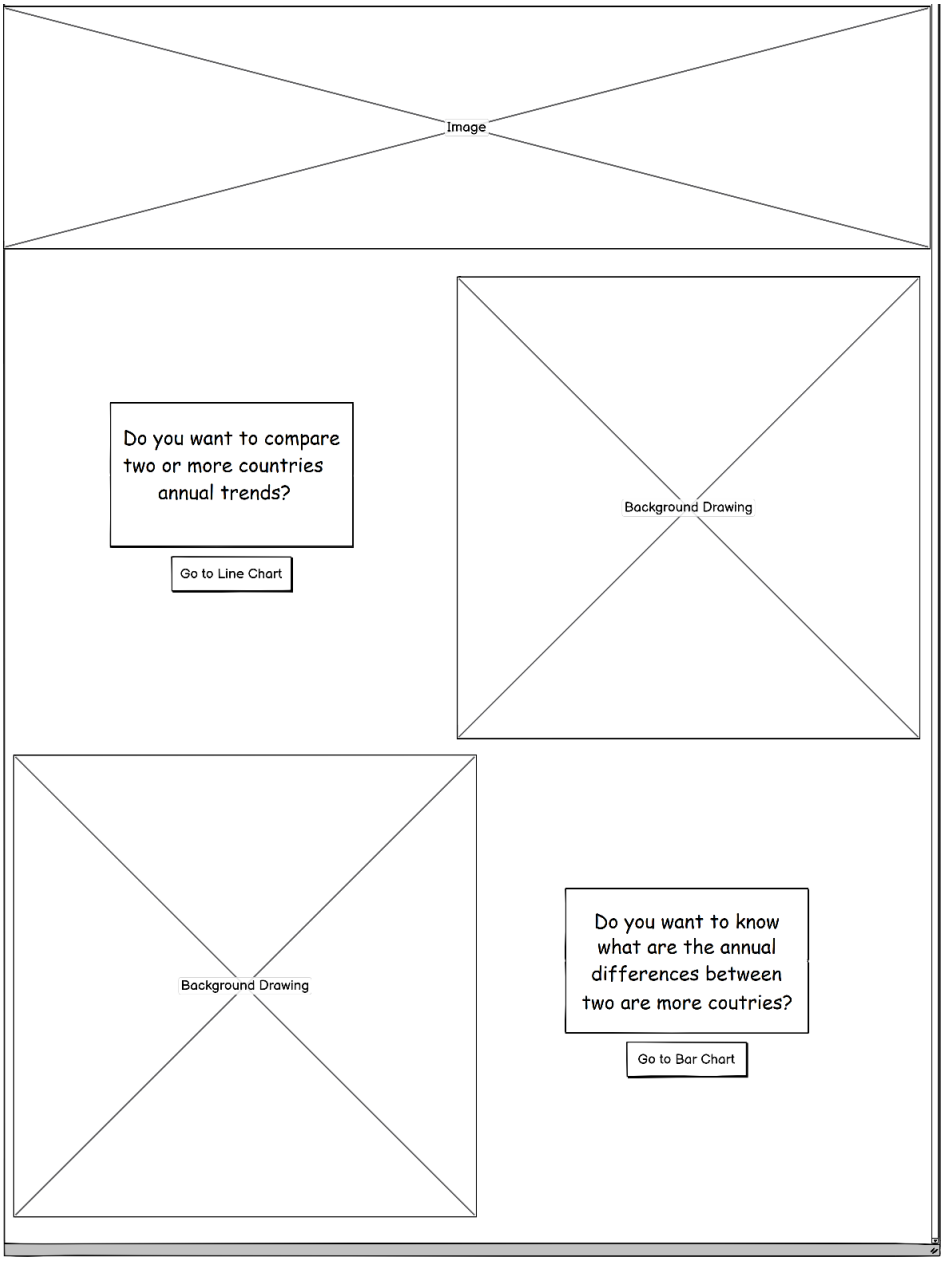



*Figure 6: Home and Landing Page - Part 2*

**Data visualisation pages**

The user can access the main "Data Visualisation" pages ("Map Chart", "Pie Chart", "Line Chart", "Bar Chart") by either: clicking the corresponding graph icon displayed at the bottom of the "Data Visualisation" main page; using the navigation tabs that appear when hovering over the "Data Visualisation" tab; or by clicking on the corresponding buttons located on the "Home" page.

By providing three ways of reaching the page of interest, the user saves time and naturally navigates through the website.

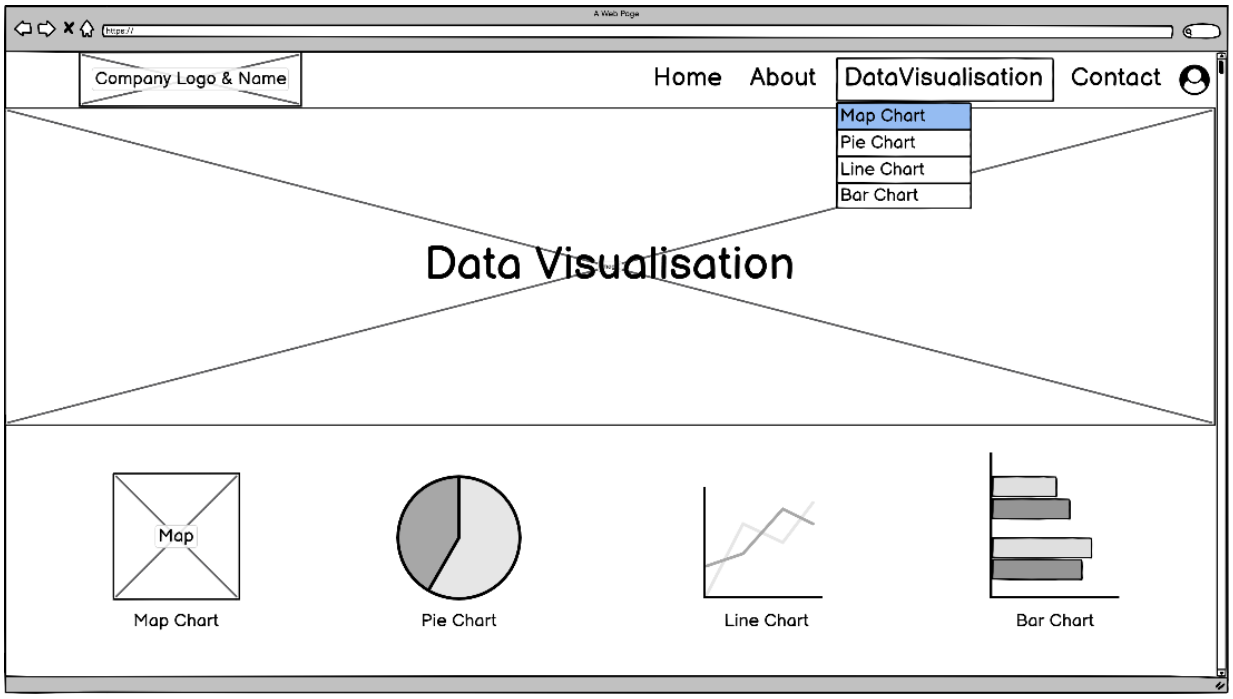



*Figure 7: Data Visualisation Page with Navigation Tabs*

**Design rationale**

Each chart has a dropdown button which gives the user the ability to easily change the variable that the graph displays. There are also drop-down buttons for years, quarters or countries depending on the type of the chart.

The sliders at the bottom of the charts give the user the ability to easily navigate throughout the years. The play, pause and stop buttons control the animation of the data change throughout the years.

Each chart has its own comment section where the users can comment regarding the information provided on the page.

The "Save" icons can be used to store the parameters in the database, which can be reviewed by accessing the "User Profile" page, "Saved Data Visualisations" section.

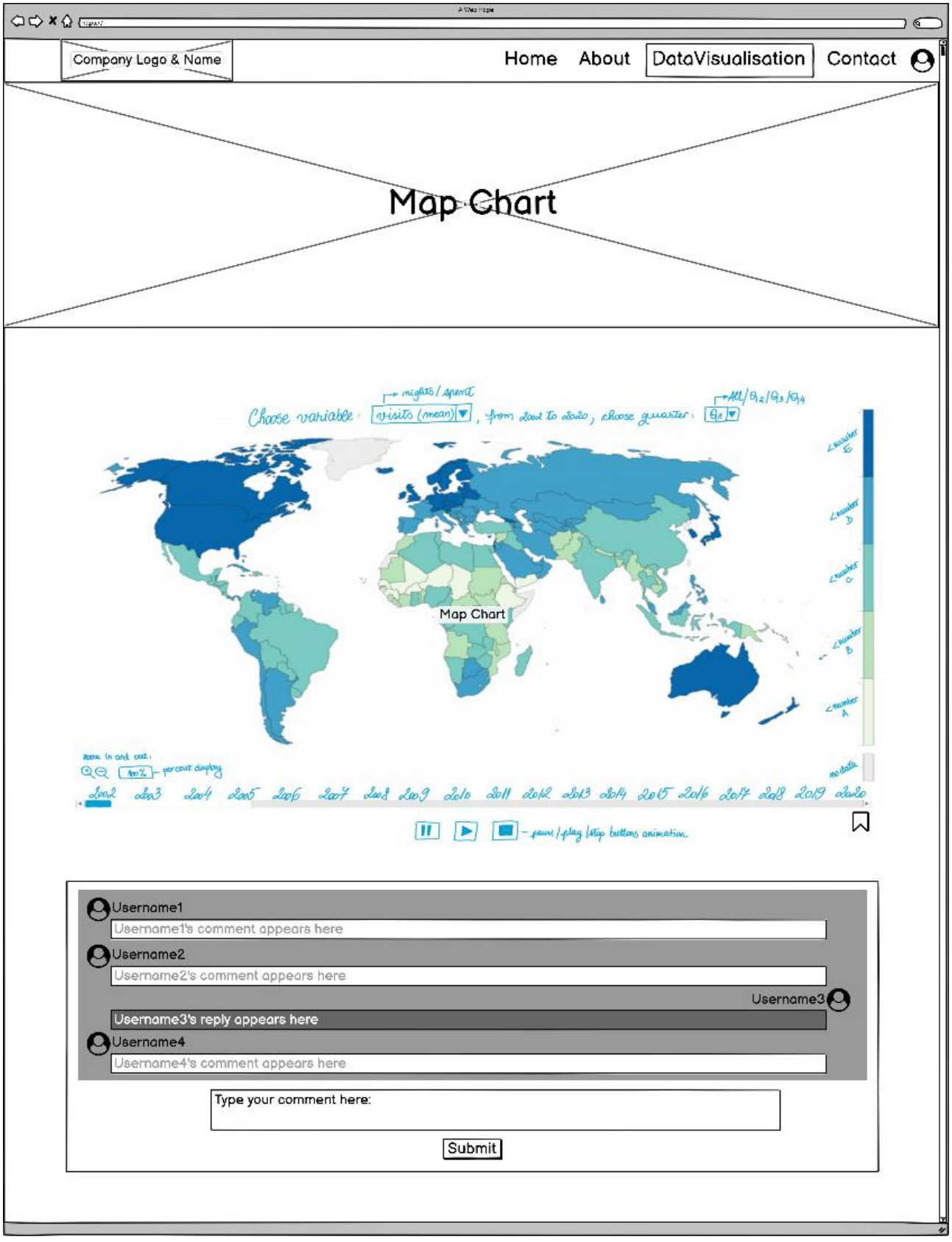



*Figure 8: Data Visualisation, Map Chart Page*

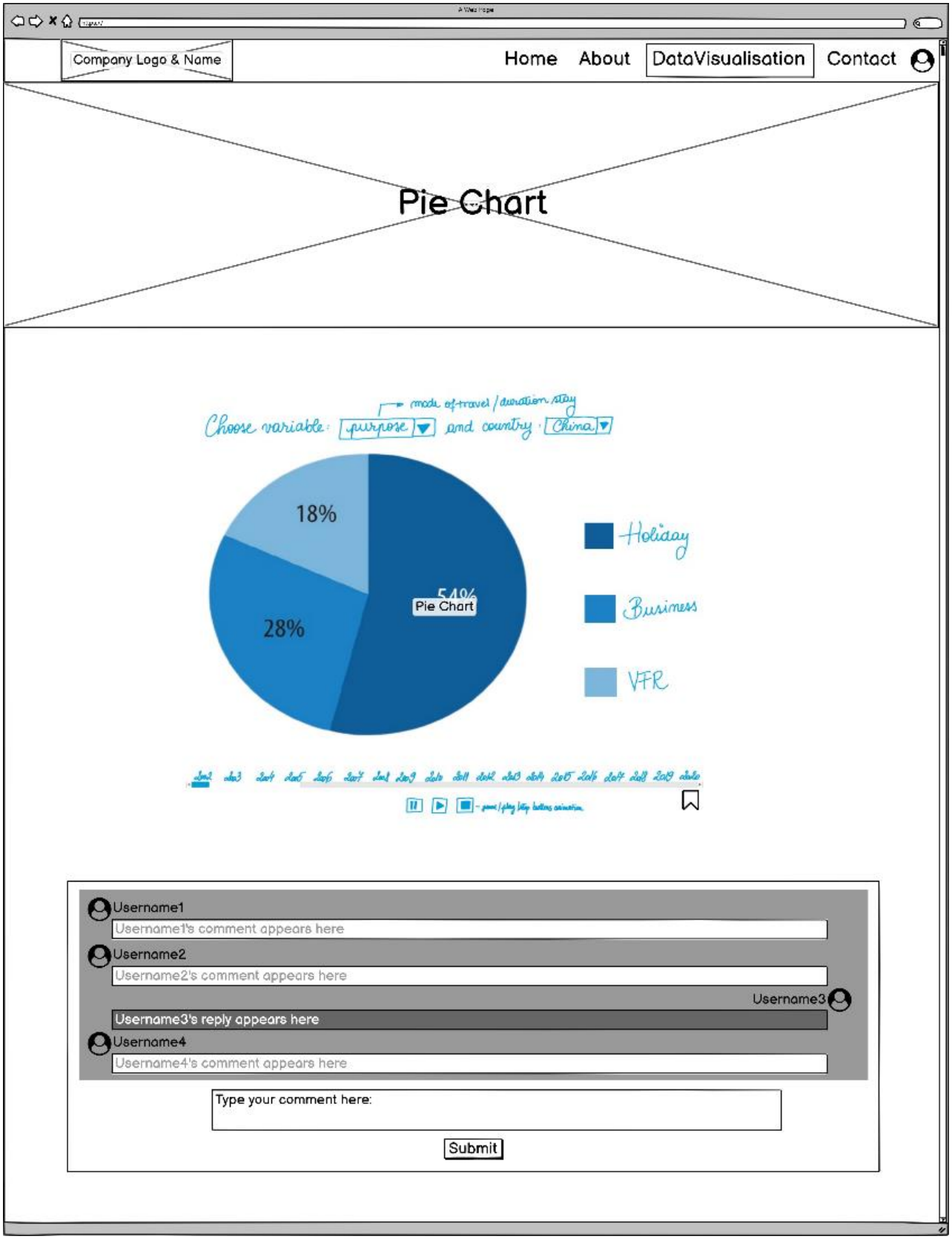



*Figure 9: Data Visualisation, Pie Chart Page*

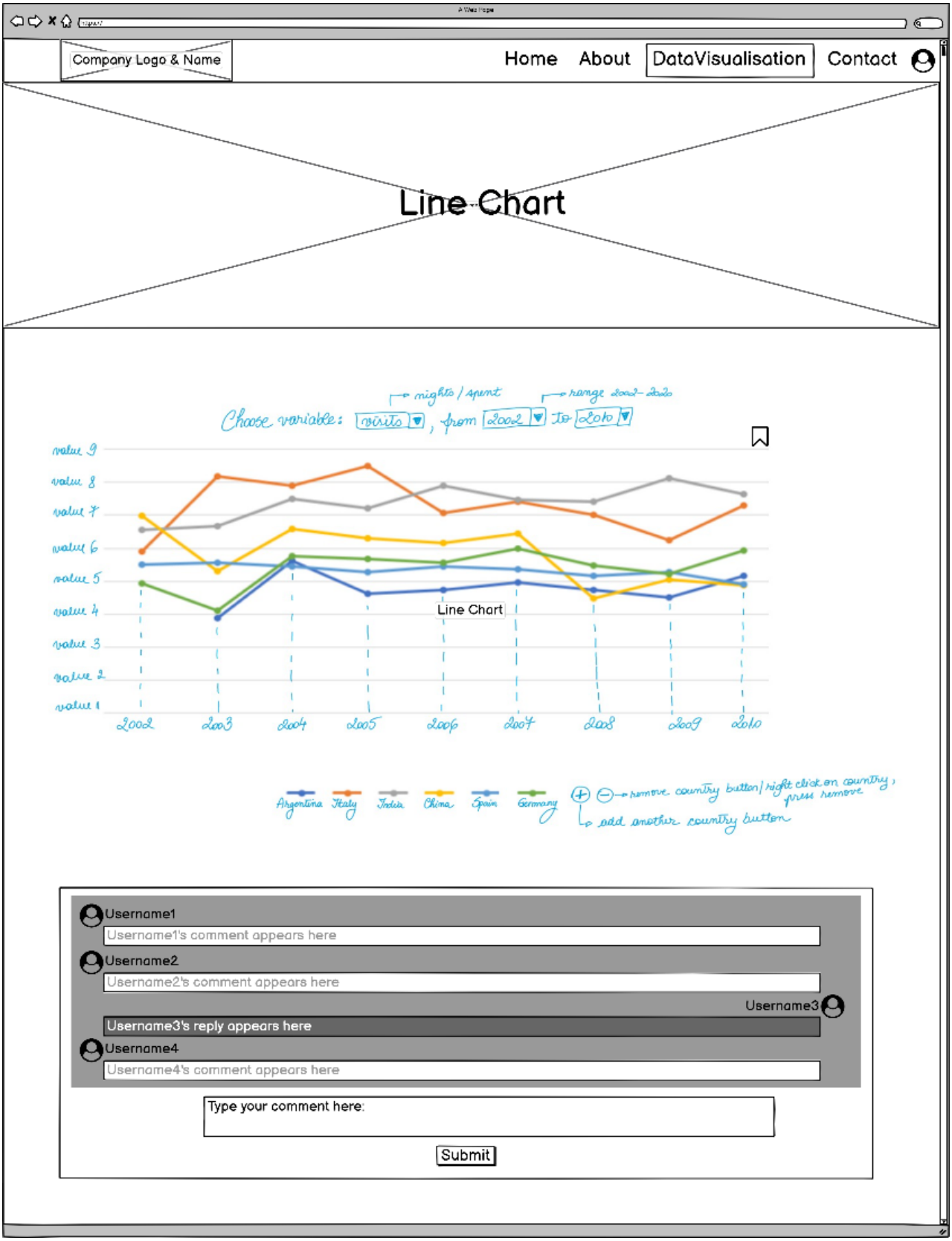



*Figure 10: Data Visualisation, Line Chart Page*

In order to be compared, countries can be added in the charts by pressing the '+' button and removed by selecting them and pressing the '–' button.

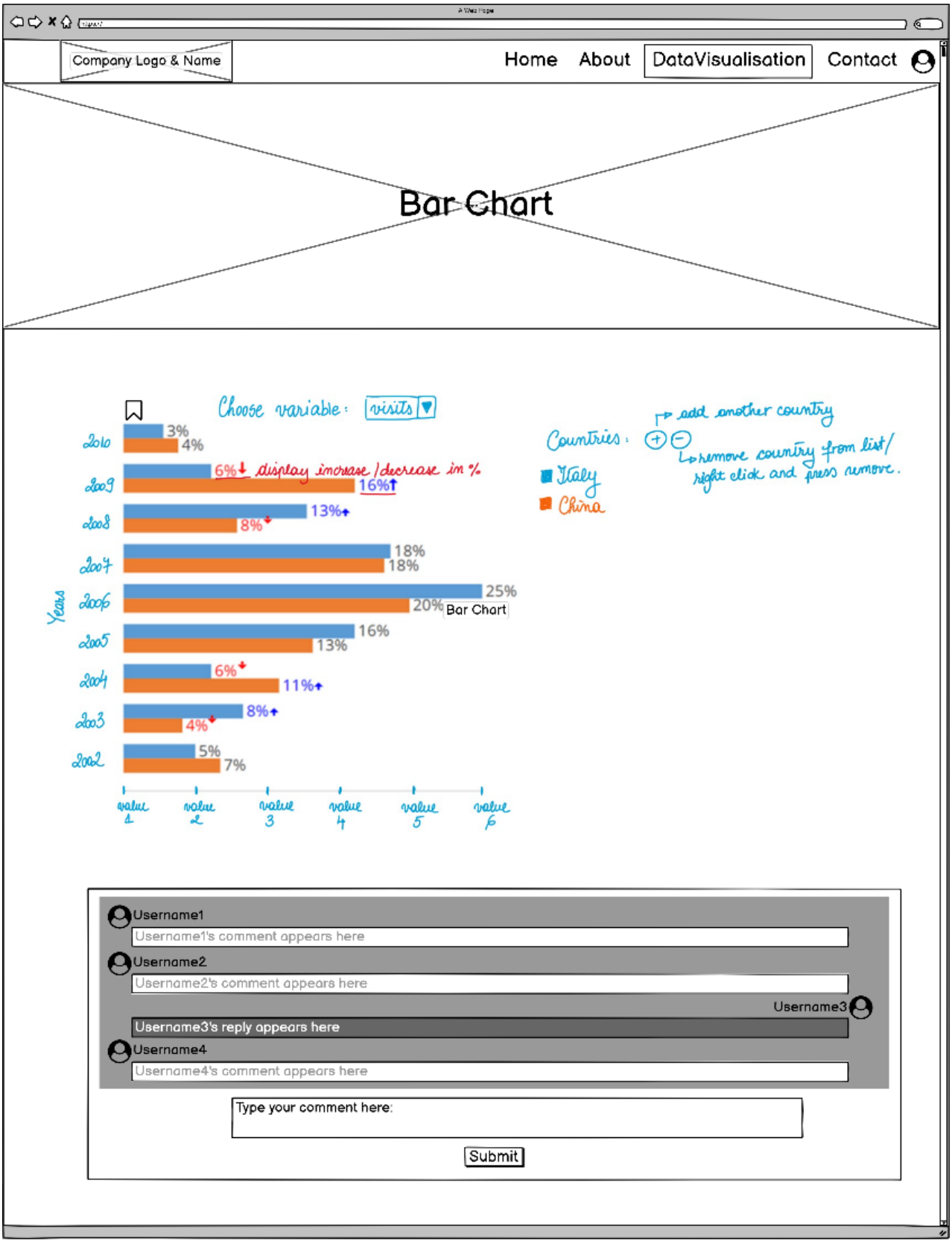



*Figure 11: Data Visualisation, Bar Chart Page*

## User profile page

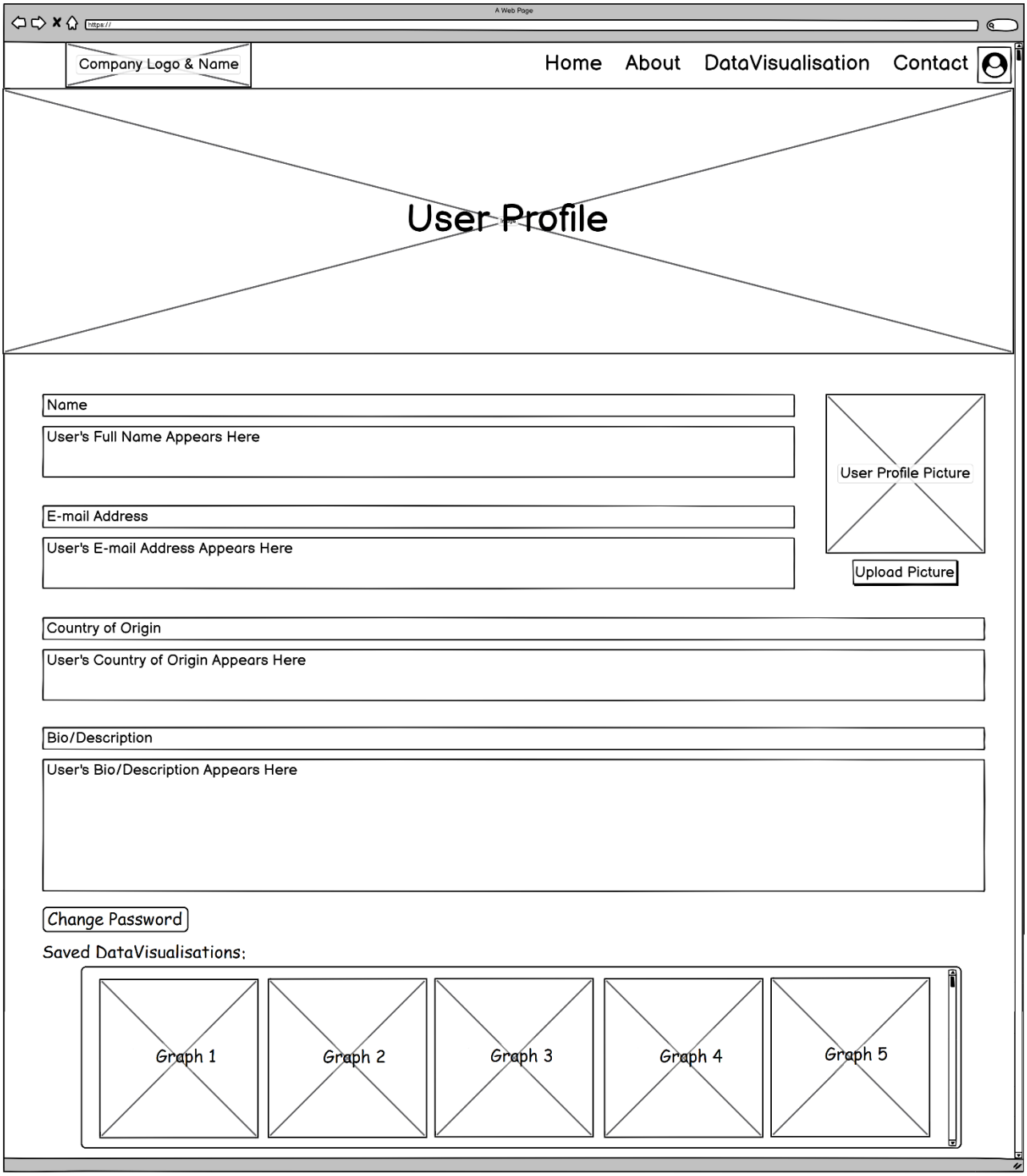



*Figure 12: User Profile Page*

### Further features to be implemented within the wireframes

- A search tool will be added in the main tab bar to give the users the ability to search for other users' profiles by typing in their usernames/names/e-mails.
- Like and dislike buttons will be added in the comment boxes, right below the comments, in order to give the users the ability to up-vote or down-vote comments. The number of likes and dislikes will be displayed next to the buttons.
- A report errors section/tab will be added so that the users can reach to the developers to report errors or inappropriate activity on the website.

## Database

### Conceptual design

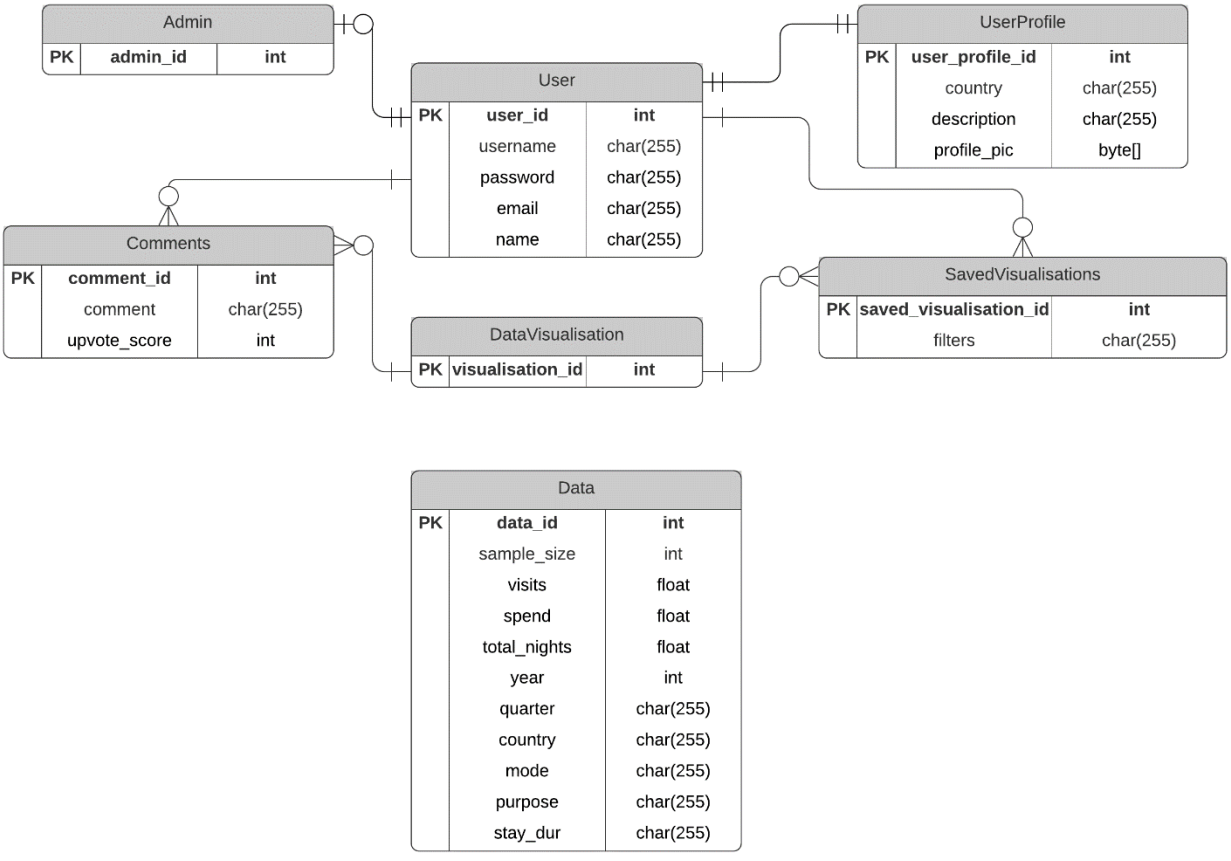



*Figure 13: Conceptual Database Design*

## Logical design

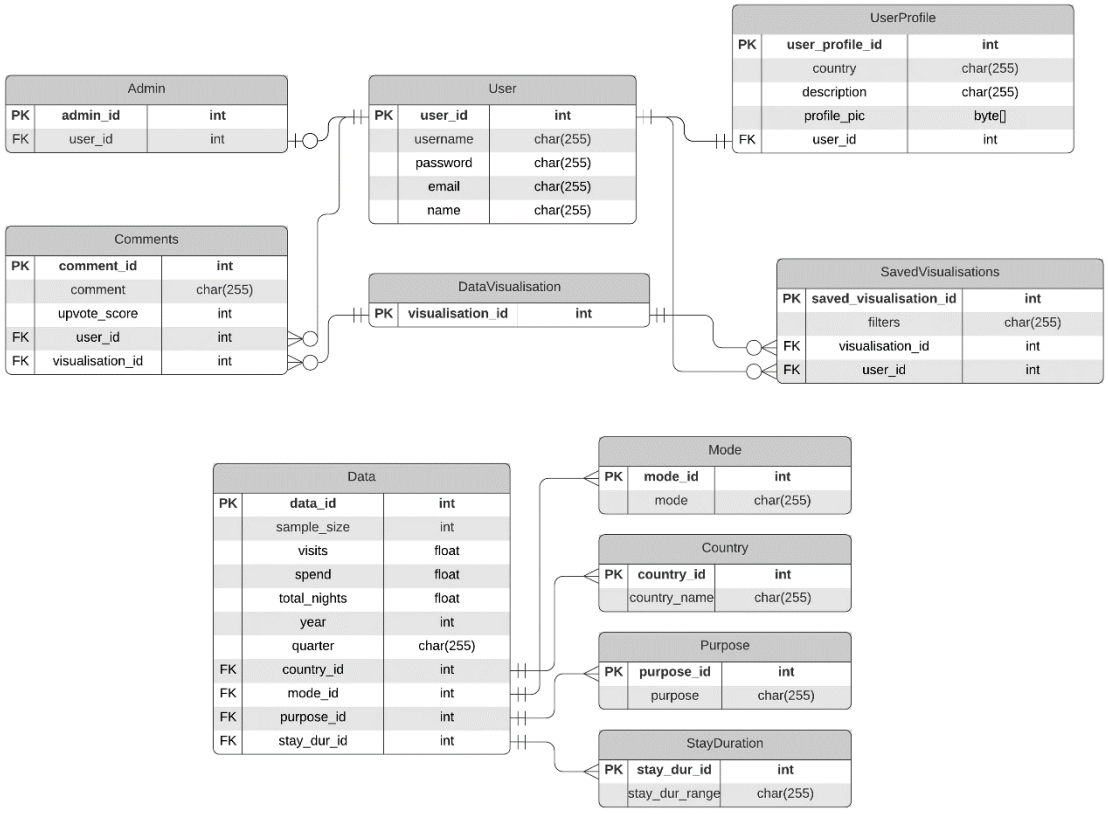



*Figure 14: Logical Database Design*

The logical design (Figure 14) has been normalised to 3rd form. The data for our visualisations is included as a table in our database design but is still subject to change if it is later found that using an application programming interface (API) is more convenient. We associated all the sections in the data that would be strings (e.g. country name) with numbers that are used as foreign keys in the Data table.

On the following page, the entities and their attributes are listed out with constraints.

## Tables with constraints

**User**

| *Key* | *Attribute* | *Data Type* | *Constraint* |
|---|---|---|---|
| PK | user_id | int | NOT_NULL, AUTO_INCREMENT, UNIQUE |
| | username | char(255) | NOT_NULL |
| | password | char(255) | NOT_NULL |
| | email | char(255) | NOT_NULL |
| | name | char(255) | NOT_NULL |

**Comments**

| *Key* | *Attribute* | *Data Type* | *Constraint* |
|---|---|---|---|
| PK | comment_id | int | NOT_NULL, AUTO_INCREMENT, UNIQUE |
| | comment | char(255) | |
| | upvote_score | int | DEFAULT 0 |
| FK | visualisation_id | int | NOT_NULL |
| FK | user_id | int | NOT_NULL, CASCADE_DELETE |

**Data**

| *Key* | *Attribute* | *Data Type* | *Constraint* |
|---|---|---|---|
| PK | data_id | int | NOT_NULL, AUTO_INCREMENT, UNIQUE |
| | sample_size | int | |
| | visits | float | |
| | spend | float | |
| | total_nights | float | |
| | year | int | |
| | quarter | char(255) | |
| FK | country_id | int | |
| FK | mode_id | int | |
| FK | purpose_id | int | |
| FK | stay_dur_id | int | |

**UserProfile**

| *Key* | *Attribute* | *Data Type* | *Constraint* |
|---|---|---|---|
| PK | user_profile_id | int | NOT_NULL, AUTO_INCREMENT, UNIQUE |
| | country | char(255) | |
| | description | char(255) | |
| | profile_pic | byte[] | |
| FK | user_id | int | NOT_NULL, CASCADE_DELETE |

**SavedVisualisations**

| *Key* | *Attribute* | *Data Type* | *Constraint* |
|---|---|---|---|
| PK | saved_visualisation_id | int | NOT_NULL, AUTO_INCREMENT, UNIQUE |
| | filters | char(255) | |
| FK | visualisation_id | int | NOT_NULL |
| FK | user_id | int | NOT_NULL, CASCADE_DELETE |

**DataVisualisation**

| *Key* | *Attribute* | *Data Type* | *Constraint* |
|---|---|---|---|
| PK | visualisation_id | int | NOT_NULL, AUTO_INCREMENT, UNIQUE |

**Mode**

| *Key* | *Attribute* | *Data Type* | *Constraint* |
|---|---|---|---|
| PK | mode_id | int | NOT_NULL, AUTO_INCREMENT, UNIQUE |
| | mode | char(255) | |

**Country**

| *Key* | *Attribute* | *Data Type* | *Constraint* |
|---|---|---|---|
| PK | country_id | int | NOT_NULL, AUTO_INCREMENT, UNIQUE |
| | country | char(255) | |

**Purpose**

| *Key* | *Attribute* | *Data Type* | *Constraint* |
|---|---|---|---|
| PK | purpose_id | int | NOT_NULL, AUTO_INCREMENT, UNIQUE |
| | purpose | char(255) | |

**StayDuration**

| *Key* | *Attribute* | *Data Type* | *Constraint* |
|---|---|---|---|
| PK | stay_dur_id | int | NOT_NULL, AUTO_INCREMENT, UNIQUE |
| | stay_dur_range | char(255) | |

*Figure 15: Tables with Constraints*

All attributes in the User table are not null. This is because all this information will be required when the user registers and thus should have a value. For the Comments table, all upvote scores start at 0, which is why 0 is the default value. If the user who posts the comments is deleted, all their comments will subsequently be deleted. Therefore, the Comments table will cascade delete when the corresponding User is deleted. Similarly, if a user is deleted, their UserProfile will also be deleted and their SavedVisualisations. Therefore, cascade delete is used on UserProfile and SavedVisualisations tables.

# Dataset

## Dataset description, columns and format

Our dataset describes visitors to London from overseas (Office for National Statistics, 2020). A table of raw data is included, as well tables that conveniently aggregate the raw data by year, quarter and country of origin. The data spans from January 2002 to March 2020.

For every quarter and every country, the following columns are outlined: number of visits, total £ spent, total number of nights. This information is made more granular with trip specific data, which include the columns: duration of stay, mode of transport, and purpose of travel.

The data is either in the form of float numbers or text. The text values are limited within categories: country, mode of transport (sea, air, tunnel), quarter (January – March, April – June, July – September, October – December) and purpose of travel (holiday, business, study, VFR, miscellaneous). Textual data does not pose too much of a problem due to its limited variability. A detailed table outlining the variables in each column can be found in Appendix E.

## Sample methodology and data reliability

The data was collected through the International Passenger Survey (IPS) and conducted by Office for National Statistics, which follows the EU General Data Protection Regulation (GDPR), confirming that the data was collected in accordance with it (Office for National Statistics, 2020) (Office for National Statistics, n.d.).

The International Passenger Survey (IPS) produces estimations that are based on interviews with a random sample of passengers entering and leaving the UK on the most important air, tunnel and sea routes. This means that 90% of international passengers have the chance of being sampled, making the information within the dataset reliable and accurate (Horsfield, 2017).

Travelers passing through passport control are randomly selected for interview. Only interviews carried out at the end of a visit are used to generate estimates of expenditure and stay.

## Data quality and ethics

There is a high ethical consideration for data collected on human participants. The response rates are high, usually over 80%. However, there are some refusals (less than 2%), indicating that interviews were conducted on a purely voluntary and anonymous basis.

The data is available for us to use as it is provided under the UK Open Government License.

## Data cleaning and processing

The data we are using does not need any additional cleaning. Some countries do not have any data available during earlier years. For example, no data was collected from Indonesia from 2002 to 2011 (Appendix F). These missing values are represented as blanks in the overall yearly summary by country. If any data was collected in a year it is always complete. This is because the International Passenger Survey (IPS) calculates an estimated value for individuals who do not fill out a particular section with values (Office for National Statistics, 2020). Therefore, there is no incomplete datapoints that need to be identified and cleaned.

The missing yearly data points will not affect our visualisations as the number of countries that have data gaps is very small and so they will adjust to the information that is available. If information for a particular country only exists from 2011 onwards, then a user will only be able to filter by those years.

We initially considered omitting data points which say that individuals spend £0 while visiting London. It initially seemed like a data entry error. However, we assume that some individuals might have had their journey paid or sponsored by a third party and therefore they did not spend any money during their visit. Moreover, there are few rows where spending is reported to be £0, which will be insignificant when calculating average spend per country or per quarter. Therefore, we decided not to omit the £0 datapoints.

Our dataset does not provide the total number of visits from every country, for every purpose. It also does not segment visits to London by country and by purpose. We would have to aggregate this data ourselves. Therefore, a way to approach this problem is to sum the visits column by purpose for every country.

A feature of our bar charts will be representing percentage change of certain values of countries over the years. We will process this information within our application by having a function take in values of a certain filter (e.g. total number of visits), calculate the percentage change from year to year, and return the processed information.

## Data for project

All the data provided within the dataset will be used to answer the questions of the project. However, we will also have to calculate the percentage change between certain values of countries over the years and the total number of visits for a specific purpose. The questions we aim to answer by using this dataset and the further processing of it are listed in the following section of the report.

## Questions we aim to answer

- What is the total number of international visits, nights spent, and amount of money spent in London (each quarter of) each year? How does this differ by country?
- What is the mean number of visits, nights spent, or amount of money spent in London per year, over the past 18 years? How does this differ by country?
- What is the most common purpose of travel, mode of travel and duration of stay in London in the past 18 years? How does this differ by country?
- What are the top 10 countries that have the most people travelling to London?
- What are the top 10 countries whose visitors spend the most money when traveling to London?
- Does the purpose of the visit affect the overall spending?
- Which quarter is the most visited by international tourists? How does this differ by country?
- How much does the total amount of money spent from international visitors in London vary by each quarter of each year?
- What is the purpose of the travel that brings the most income to London?

## Implementation Approach

Our team has decided to use a combination of agile methods Extreme Programming (XP) and Scrum for software engineering and project management. Cross Industry Standard Process for Data Mining (CRISP-DM) will be used for data science requirements.

XP's focus on simplicity, user stories, paired programming, and automated testing are the primary reasons why it is ideal for our team. We have a small, remote development team, dynamically changing software requirements, risks caused by fixed time projects, and technology that allows for automated unit tests. Simplicity and having a functional project are necessary before implementing higher levels of functionality and design, especially due to the short time limit of our project. The paired programming is ideal as we are a relatively inexperienced team of software developers. While it might be faster to have one individual program one section, our quality of software will be higher with paired programming. As automated tests and continuous integration are part of the project challenge, having them included in our software development framework will ensure that they are part of our work.

A component of XP that we have decided not to use is test driven development (TDD). TDD requires a large time commitment for writing and designing extensive tests. Since we are inexperienced in writing tests and designing full stacks of software, eventual changes in design will cause us to rewrite tests. To prevent this, tests will be written concurrently with software.

A simplified version of Scrum sprints will be used to apply XP's iterative and frequent small releases, allowing us to examine and review the project's progress throughout the entire software development life cycle. A Scrum Master is a useful leadership role that would help inspect our teamwork and facilitate meetings, ensuring that we are on track. Weekly sprints will include a Scrum meeting at the start of the week (for sprint planning) and Scrum deadlines at the end of the week. Since the product specification is relatively straightforward, instead of having a single Product Owner that manages and sets the Product Backlog, we could collaborate and agree on individual items. Furthermore, we will be using Jira's Scrum Board. As we are all working remotely, Jira's Scrum Board will be useful to increase communication, transparency, and establish clear tasks within our team.

To tackle the data science section of the project, we will use the CRISP-DM framework. The 6 major phases, especially business understanding, modelling and evaluation, will guide our sprint plans for relevant data science requirements. The cyclical process of evaluation and improvement will give us the opportunity to ensure requirements are being met at the highest standard. Evaluation of data analysis results or visualisations will be a group activity and can take place in sprint reviews.

Some limitations would be there is a large time investment with pair programming. It is not as efficient but is important as we are relatively new at software development. Secondly, there is typically a daily stand-up which might not be possible for students. However, meetings twice a week will suffice for the limited complexity of the project.

# Testing

### Development testing – unit testing and component testing

Development testing will focus on making sure that each class and function does what it is intended to and is not defected. This stage will comprise of mainly unit test and component test suites. Scenario testing and acceptance testing in the next stages will ensure that the whole software system is tested, fully integrated and complete.

Unit tests will be made as code is written. A separate file within the project directory will be made to hold the test scripts. Each class or function will each have a python file with their respective test classes. The python module pytest will be used for our test suite.

Component tests will be written once the main classes (user, userProfile, comment, data Visualisations etc.) have been developed to a workable degree. Test requirements will be developed by categorising each route or page as a component. For example, the user profile page will have component tests that evaluate the functionality of the user and userProfile classes in combination with functions for querying the database for user information, retrieving saved graphs and more.

### Release testing – scenario testing, performance testing, stress testing

Scenario testing will evaluate the validity of our system against the requirements we outlined. The user specification will be used to create scenarios and then test cases.

The diagram below (Figure 16) shows how specifications from user stories (in this case, the "data visualisation" user stories table from page 7) provide scenarios, then test cases. For these scenarios, the test cases overlap and every scenario (detailing the user choosing various graphs) must check for the same test cases. More scenario test case diagrams for core functionality can be found in Appendix C.

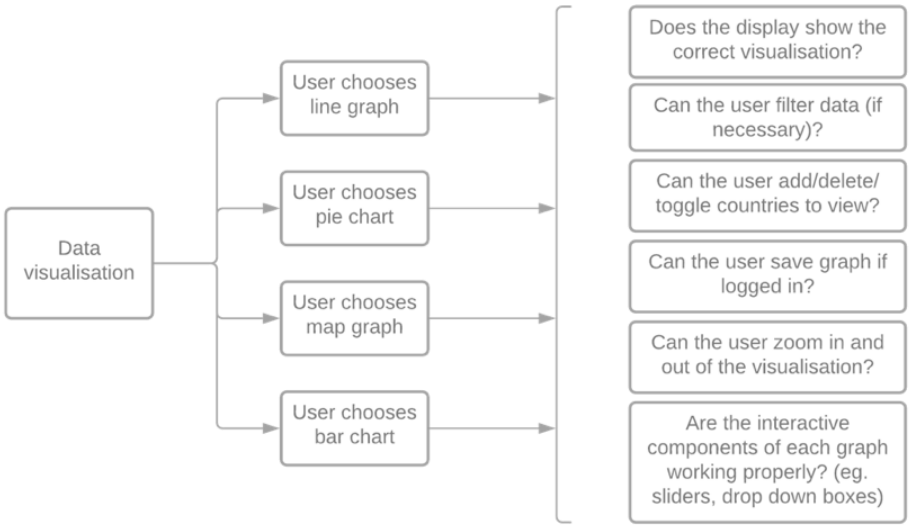



*Figure 16: Creating test cases from scenarios on data visualisation requirements*

Basic performance testing and stress testing will be done to ensure users can enjoy functionality without the system overloading. The most important parts of the website are the data visualisation pages, where data must be filtered and presented, and the comments, which must be displayed and sorted instantaneously. Hence, the system will be tested for the following measurable targets:

- Be able to handle loading 200 random, concurrent data visualisations (hypothetically for 200 different users)
- Be able to sort and display 100 concurrent comments
- Be able to sort and display 100 concurrent upvotes/downvotes

Achieving these targets does not just mean the system successfully produces the relevant visualisations and comments, but also that they are produced within a reasonable time frame. Ideally, each target should take less than 3 seconds. However, due to the limitation of the cloud deployment service we use, we can accept an elapsed time of around 5 seconds.

Moreover, stress tests stretching each of the targets will be done to grasp the capacity of the system. Line graphs showing the average time elapsed for increasing loads will be used to evaluate how efficiently our system runs. This could potentially help us identify areas in our programming where we could decrease time and space complexity.

**User testing – alpha testing and acceptance testing**

The final stage of testing – user testing – will consist of alpha testing and acceptance testing. For alpha testing, a group of users (<10 people) will be invited to test our web application and fill in a survey where they can rate functionality, answer questions about whether user requirements have been fulfilled, and provide constructive criticisms. The users will likely be students because, as university students ourselves, we have access to students that are willing to test early versions of a product and can work closely with us to provide detailed feedback. With alpha testing, we aim to find any remaining bugs and defects, as well as extend requirements testing previously done using scenarios in preparation for deployment.

After evaluating the criticisms and implementing the improvements, acceptance testing will be done with a slightly larger group of users (10-20) as our final test before deployment. The select group of users will consist of audiences given in the user stories: a restaurant owner in a tourist area, an investor in the tourism/transportation industry, a consultant, and a student or academic. Users will be asked about to give feedback on our requirements and then to evaluate the product and judge whether it is fit to be released.

Beta testing will not be done as we have a restrained time limit for the project to be completed; we may not have enough time to allow users to freely experiment and raise issues as they use the product. Furthermore, finding a larger group of users in our target audience will be difficult and require more time.

## Member Contribution

| | |
|---|---|
| **Aidan Coyne** | Database Design \| Introduction \| UML Class Diagram \| Requirement Elicitation - Observation \| Data – Data Cleaning and Processing \| Bibliography |
| **Monkgogi Galeitsiwe** | Table of contents \| UML Class Diagram (contribution) \| Use Case Diagram \| Requirements specification (partial contribution) \| Appendix A |
| **Oreeditse Mogobye** | Analysis - Requirement elicitation \| Requirement specification - User stories, categorisation, prioritisation \| Context diagram<br><br>Dataset – Questions we aim to answer (partial contribution)<br><br>Appendix A \| Bibliography |
| **Raluca Paiajen** | Interface design – Wireframes \| Design rationale \| Further features for wireframes<br><br>Dataset - Sample methodology and data reliability \| data quality and ethics \| Data cleaning and processing (partial contribution) \| Questions we aim to answer<br><br>Appendix D, E, F |
| **Tiffany Yu** | Problem statement \| User story tables \| Activity diagrams \| Data description, columns and format \| Implementation \| Appendix B, C |

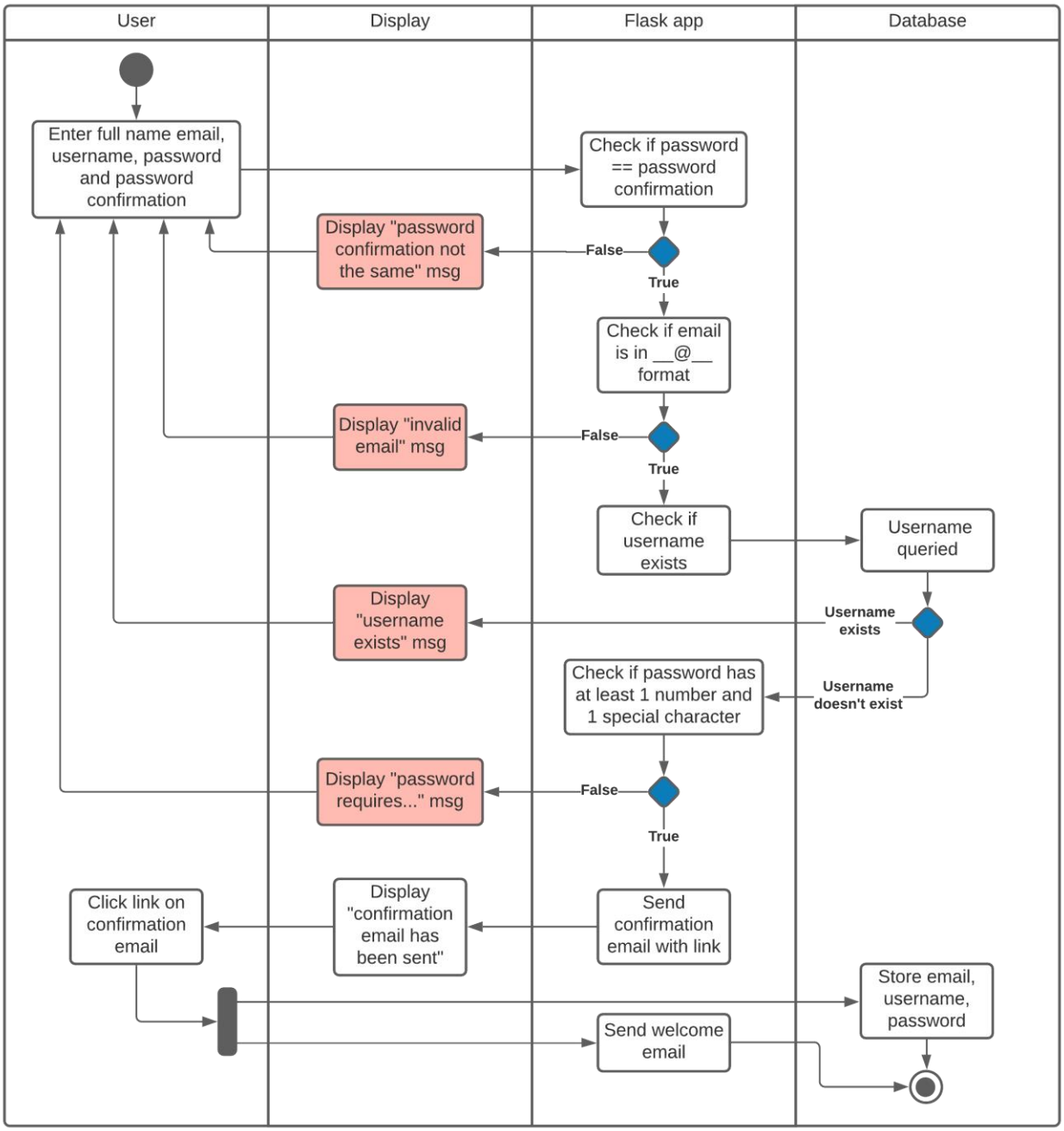
# Appendices

## Appendix A

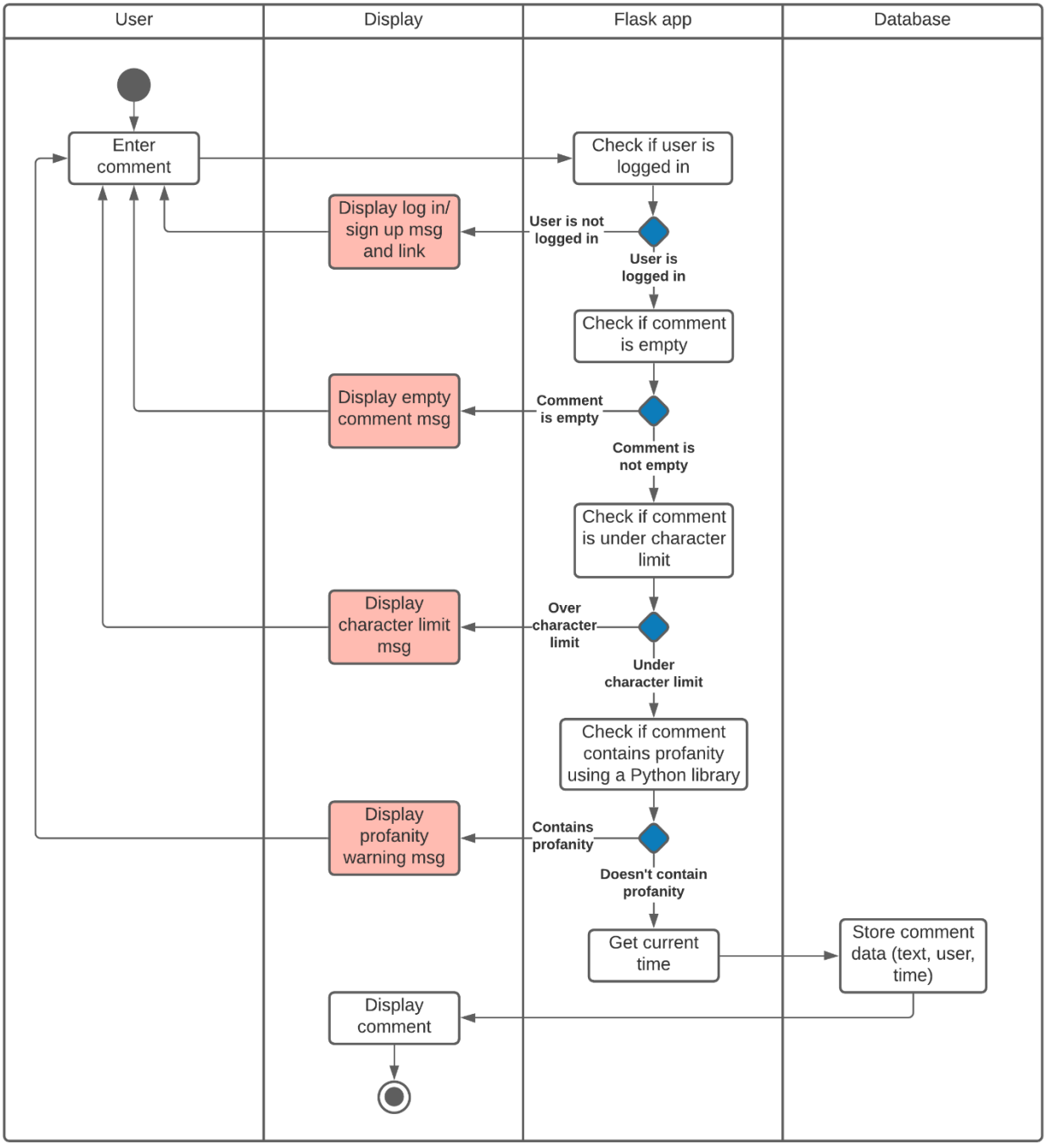
Brainstorming session using the Padlet.

**Appendix B**

The following activity diagram shows the process of verification and data storage when a user creates an account.

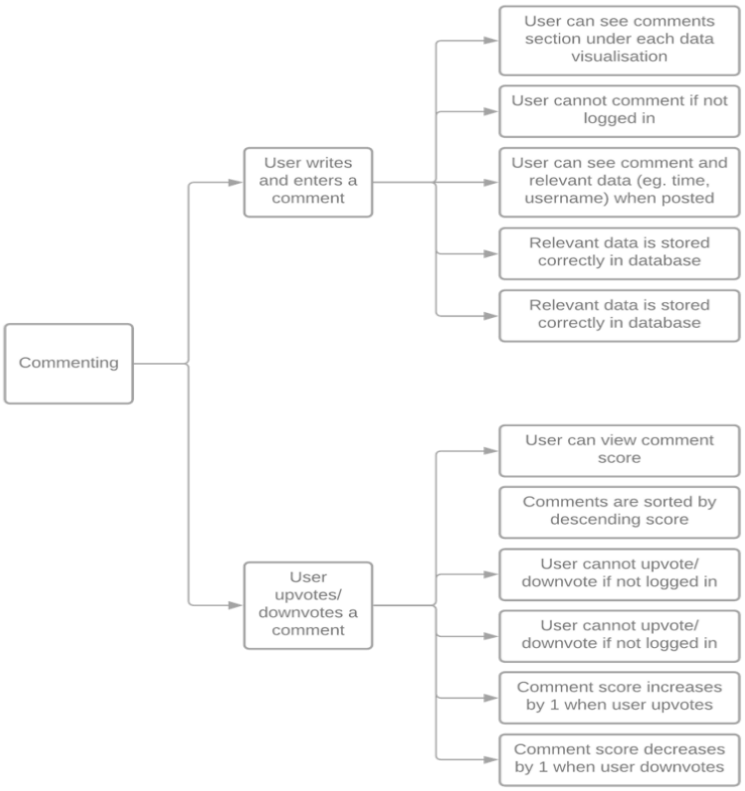


A similar activity diagram that explains how a comment is displayed and stored into the database.

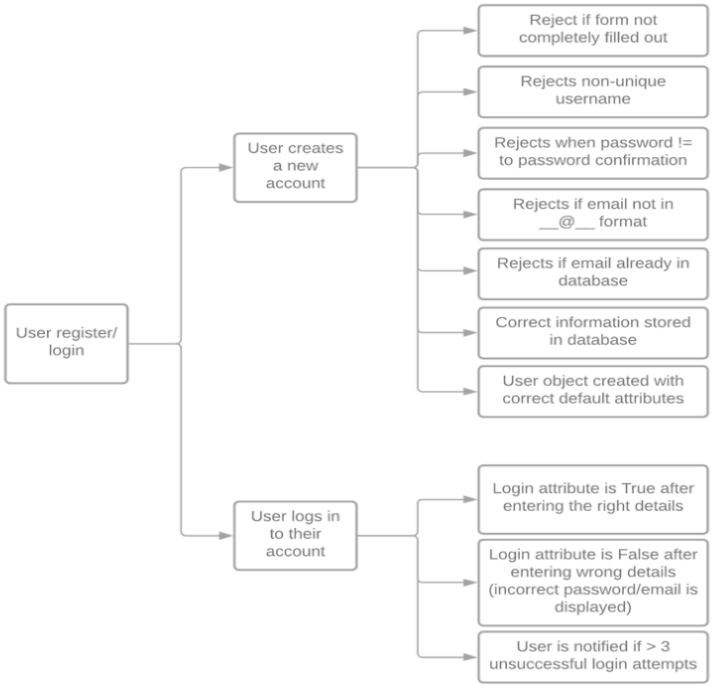


## Appendix C

Scenario testing diagram showing the production of test cases from scenarios about commenting.

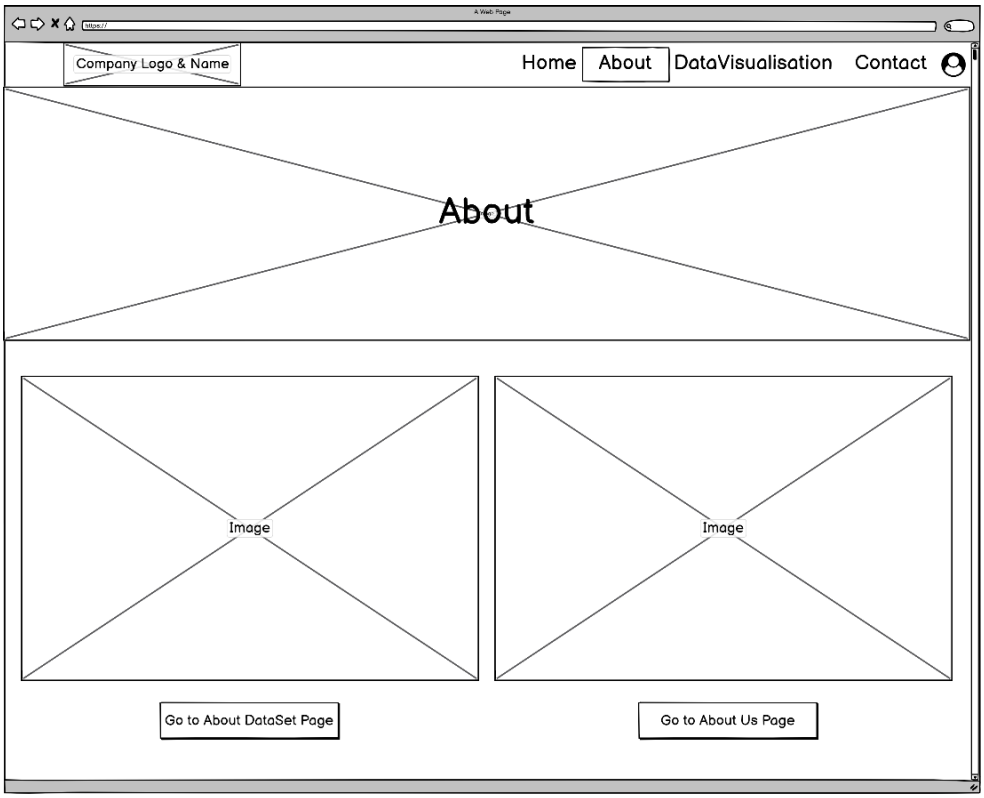


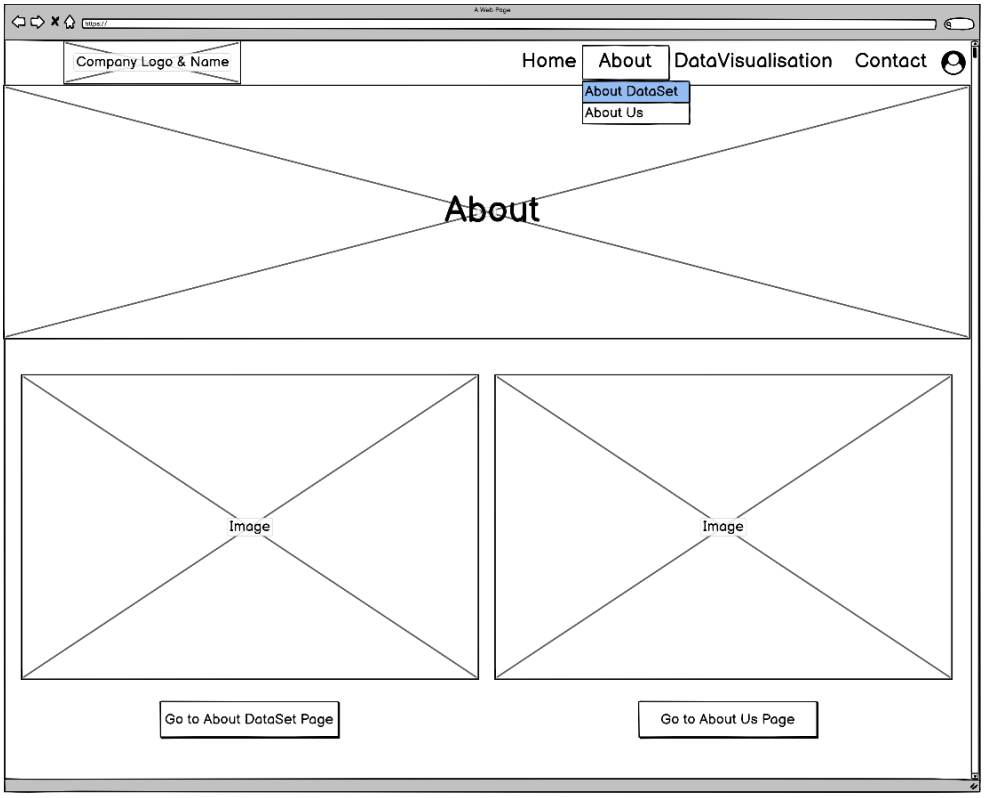

A similar scenario testing diagram showing test cases from scenarios relating to user login/signup.

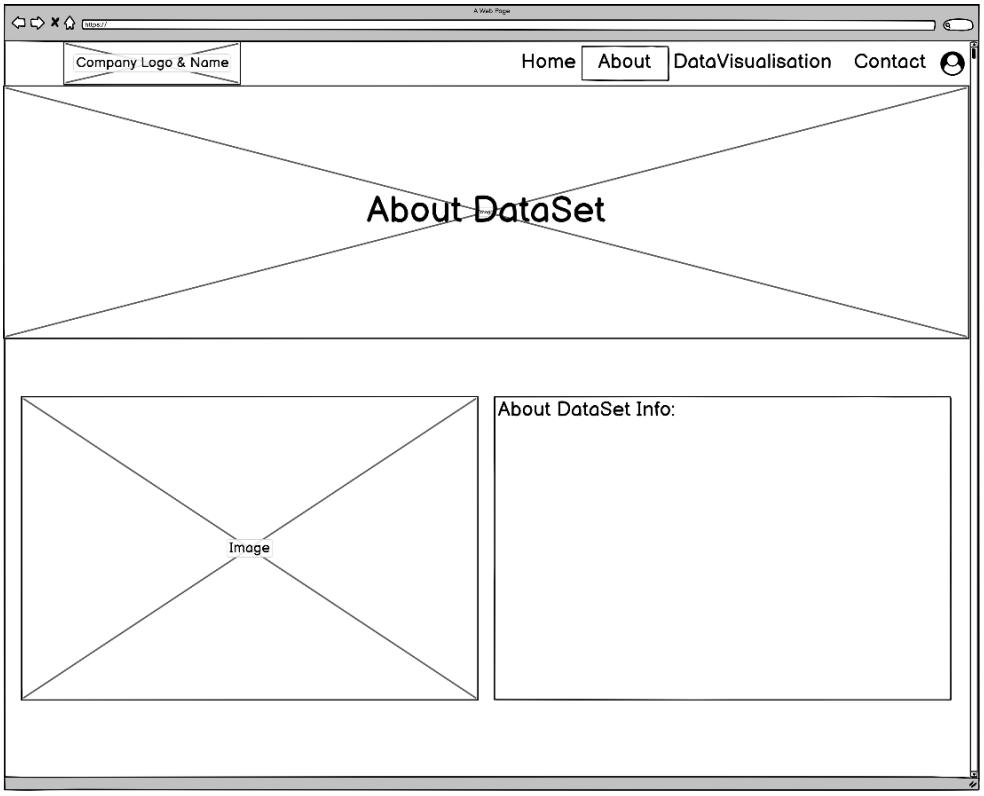


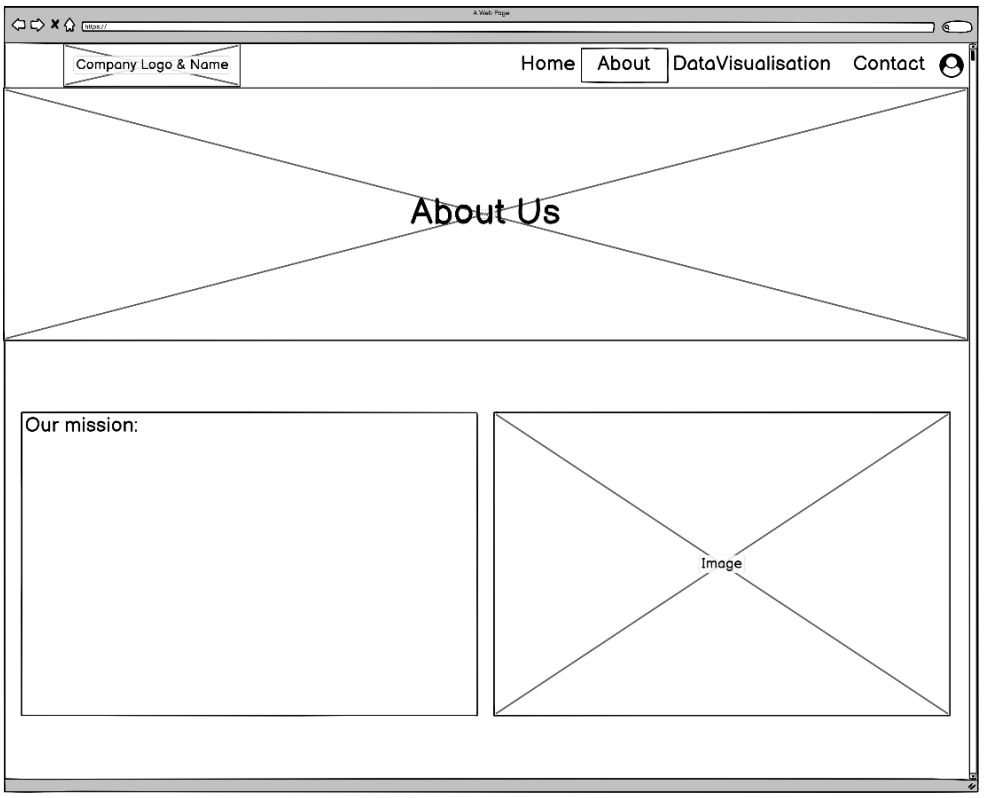
## Appendix D

## About Pages

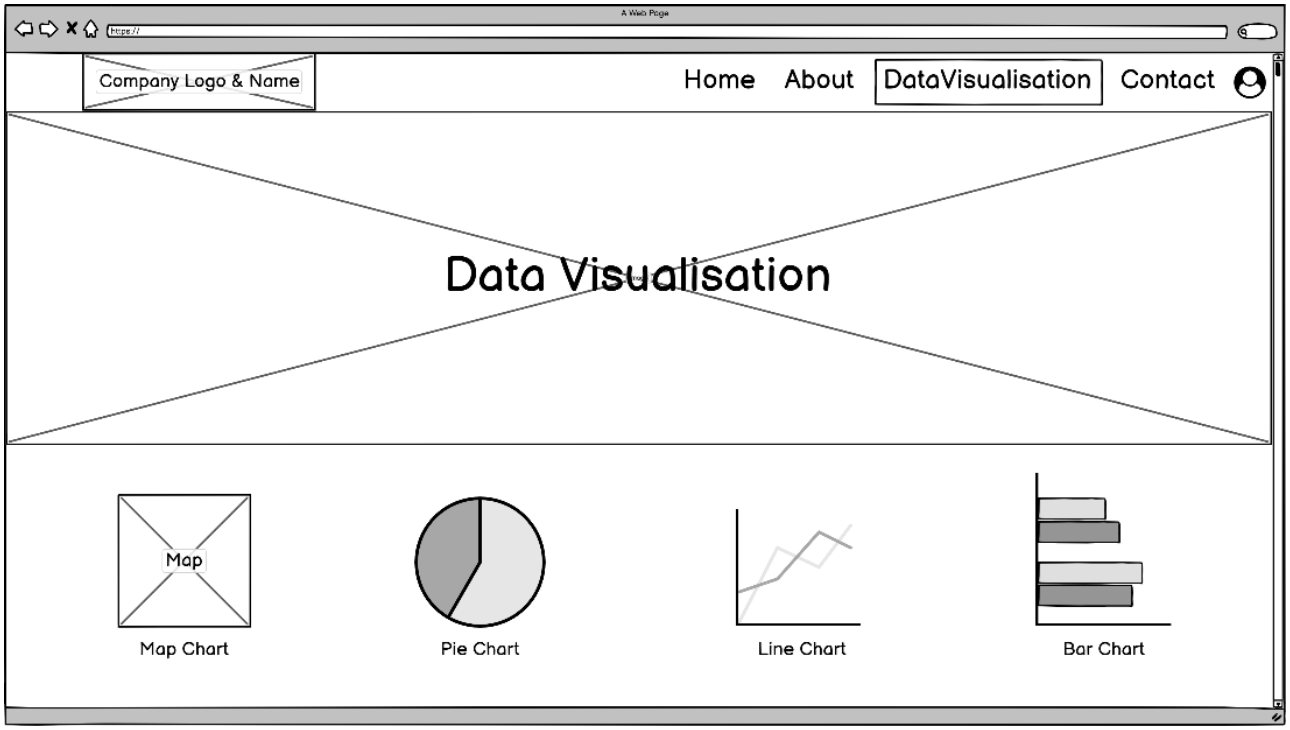


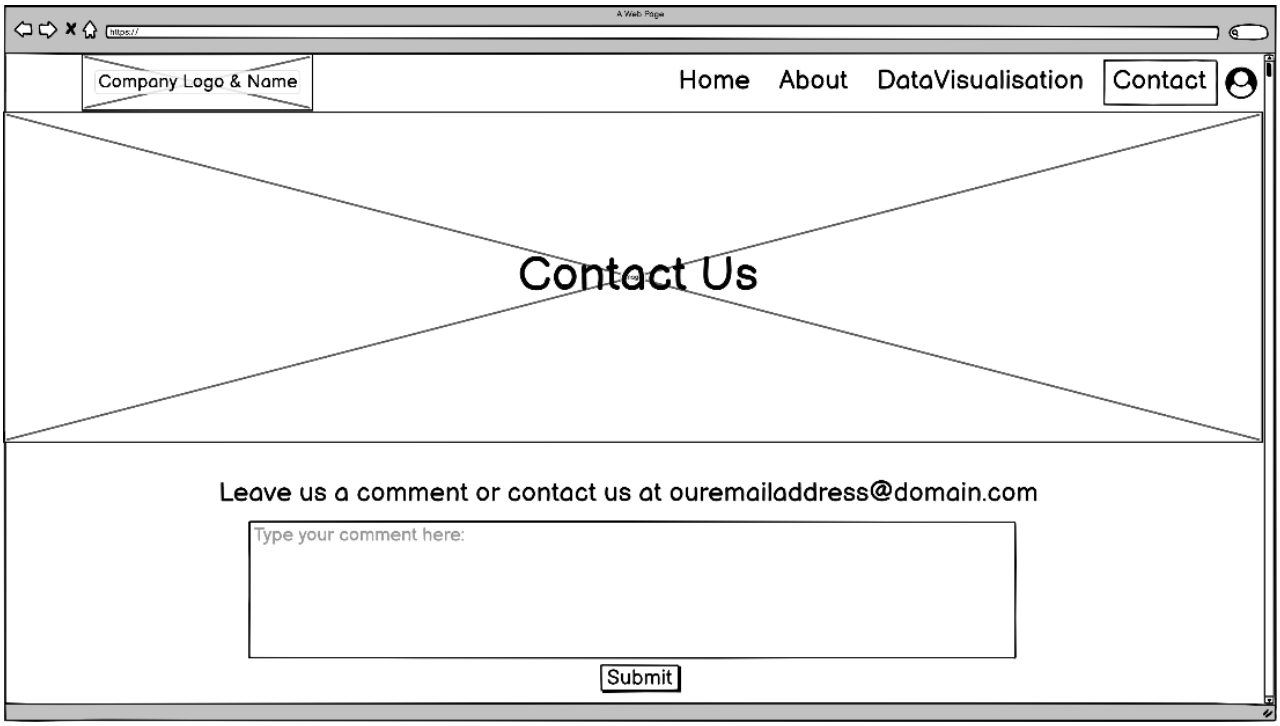

A Web Page
https://
Company Logo & Name
Home
About
DataVisualisation
Contact
About DataSet
Image
About DataSet Info:


A Web Page
https://
Company Logo & Name
Home
About
DataVisualisation
Contact
About Us
Our mission:
Image

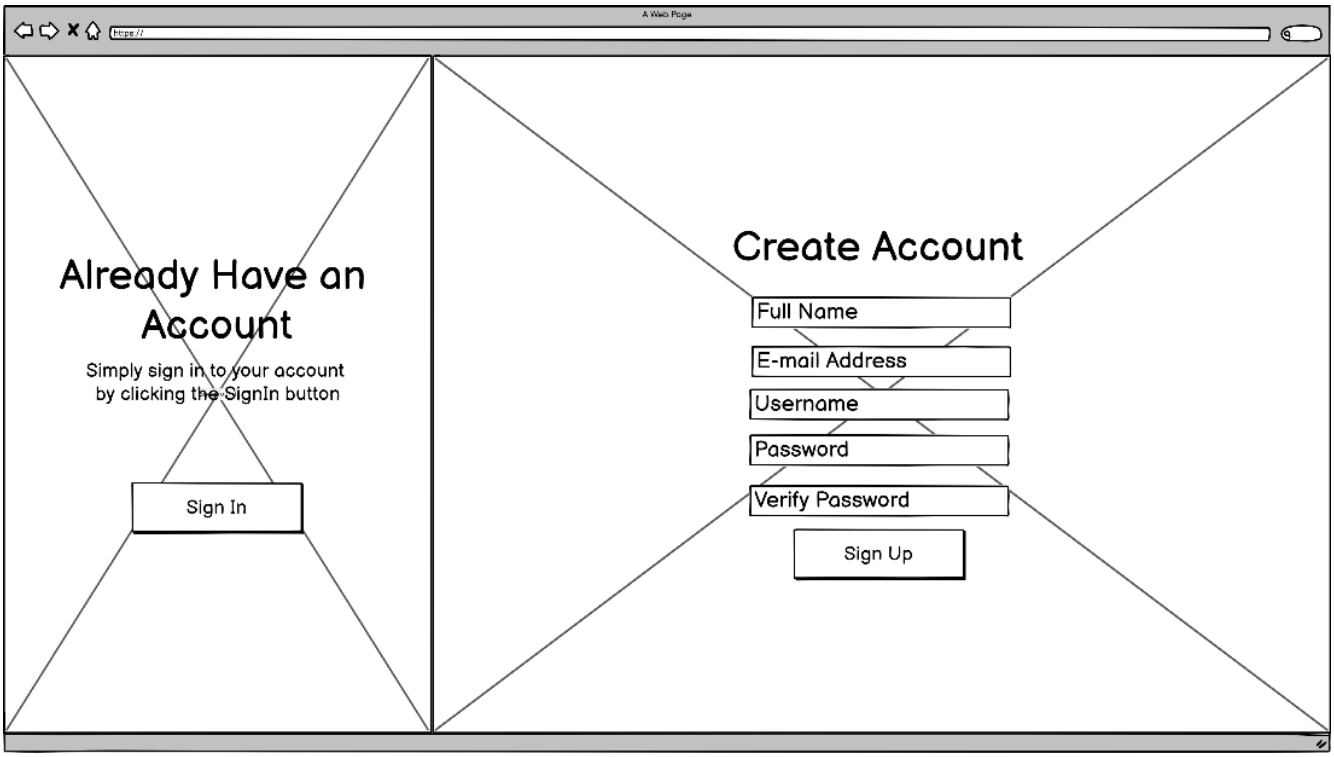
**Data Visualisation Page**



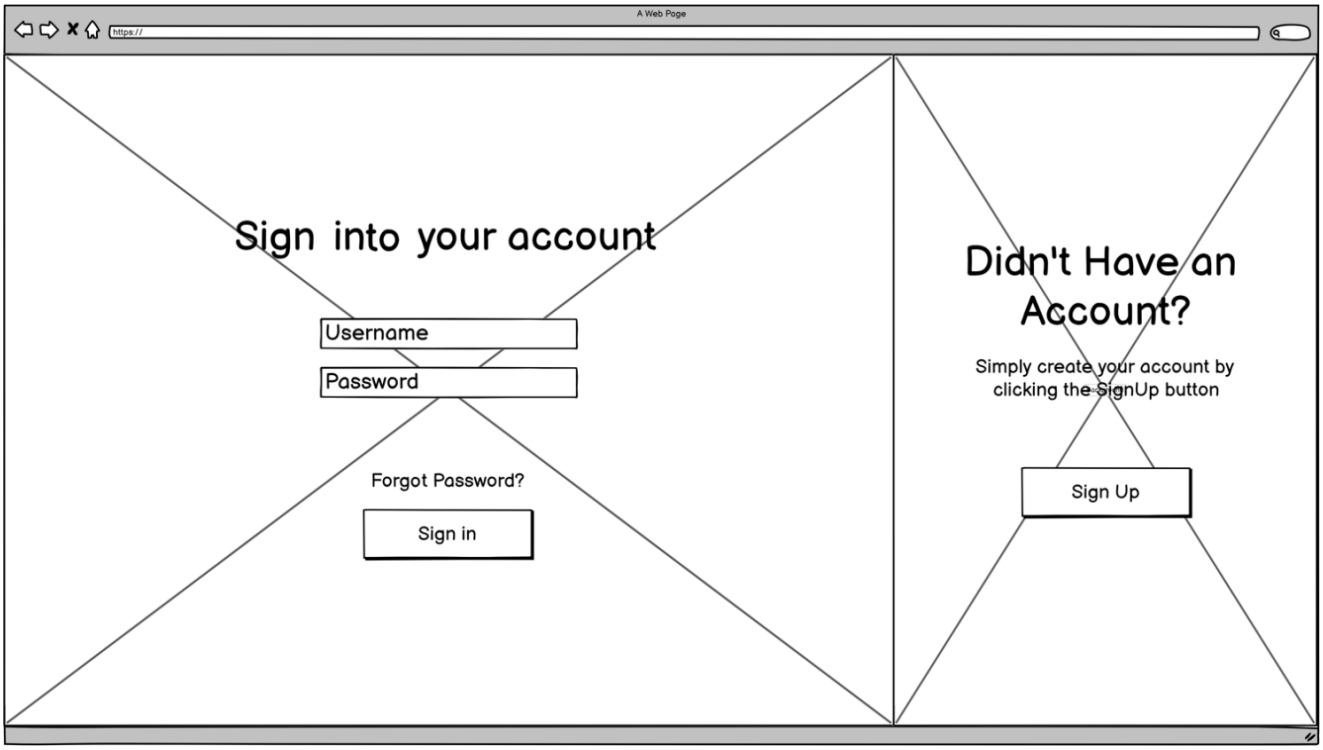

**Contact Page**



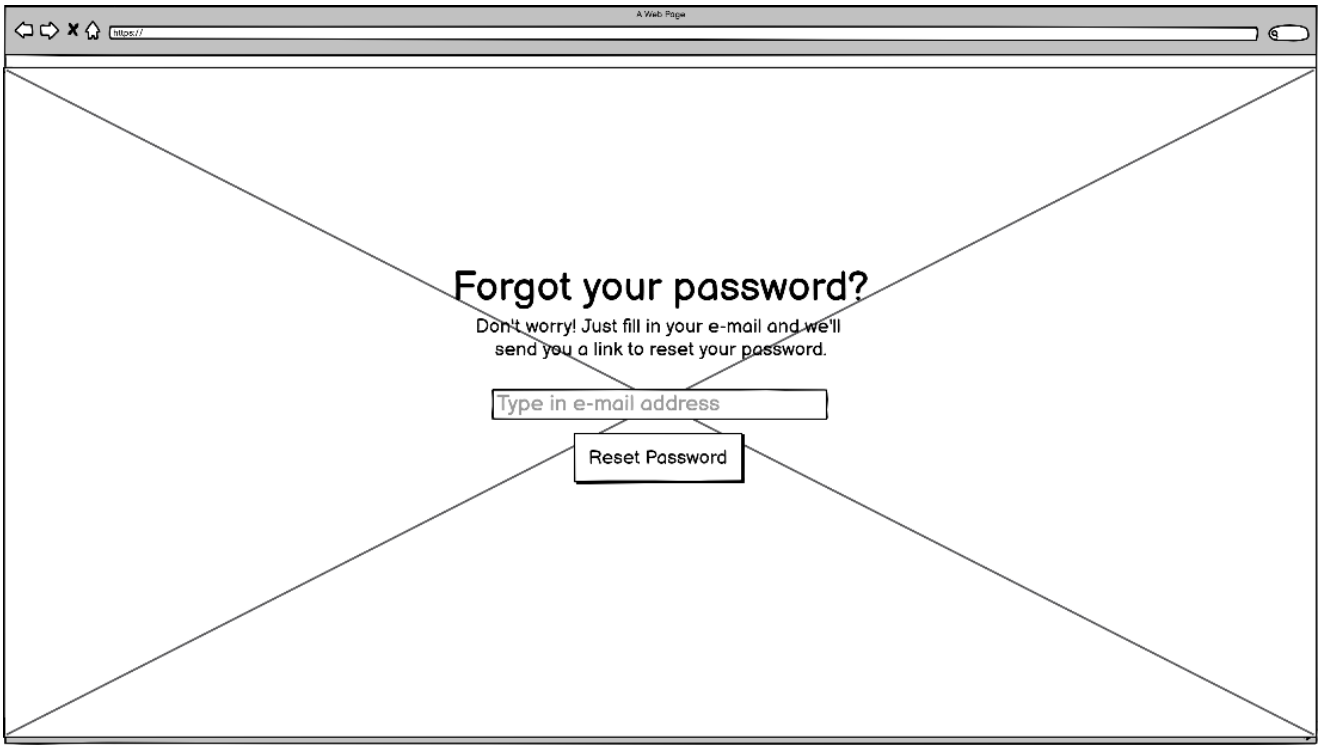
**Create Account Page**



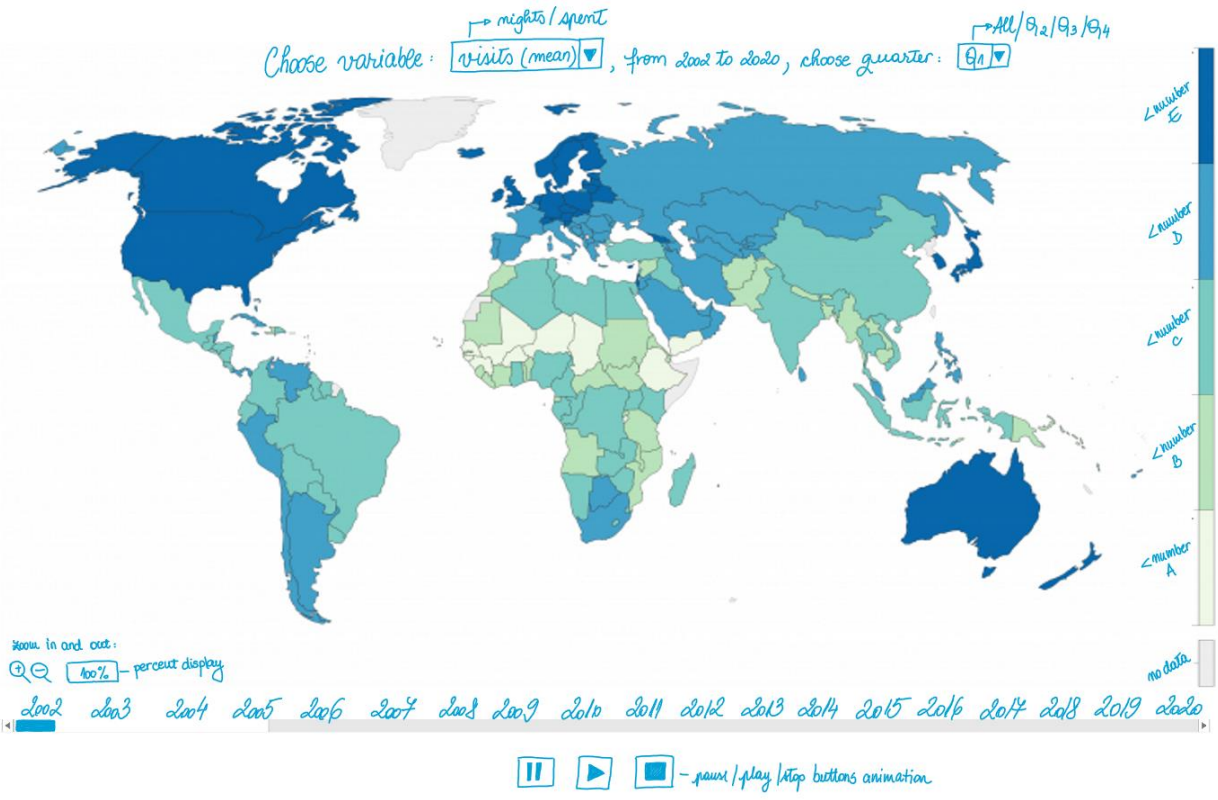

**Sign into your account Page**

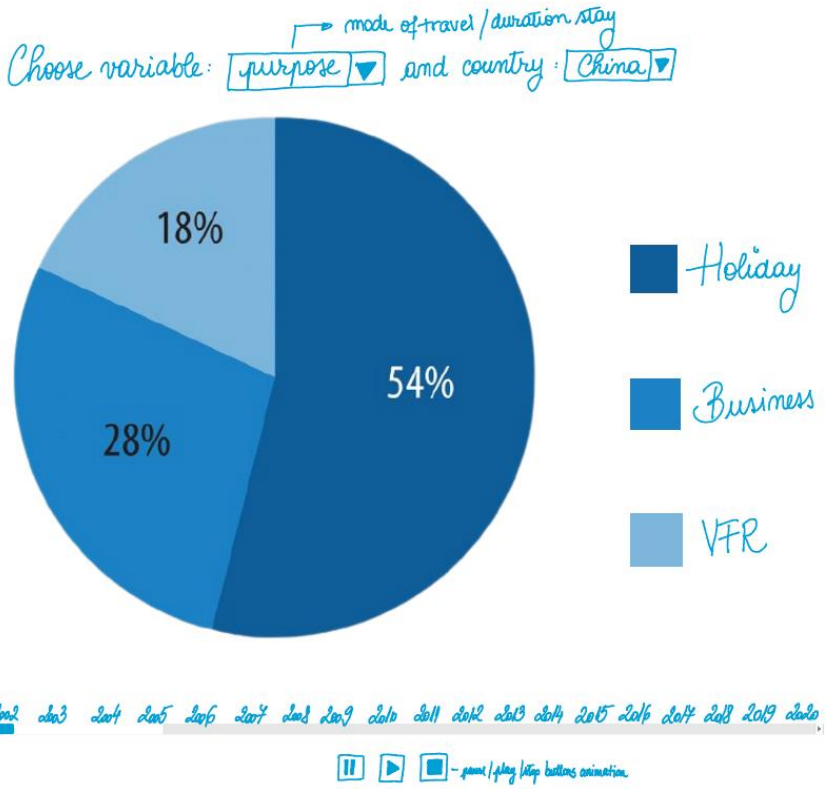


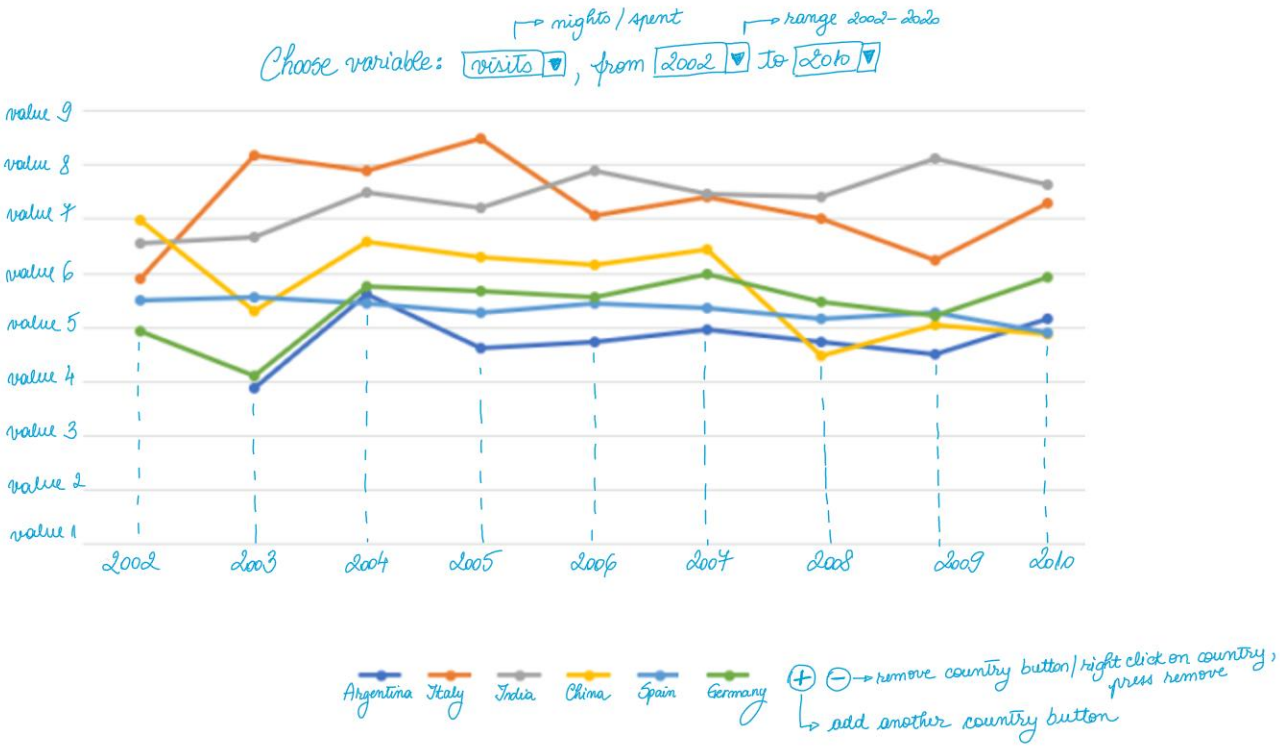
## Forgot your password page

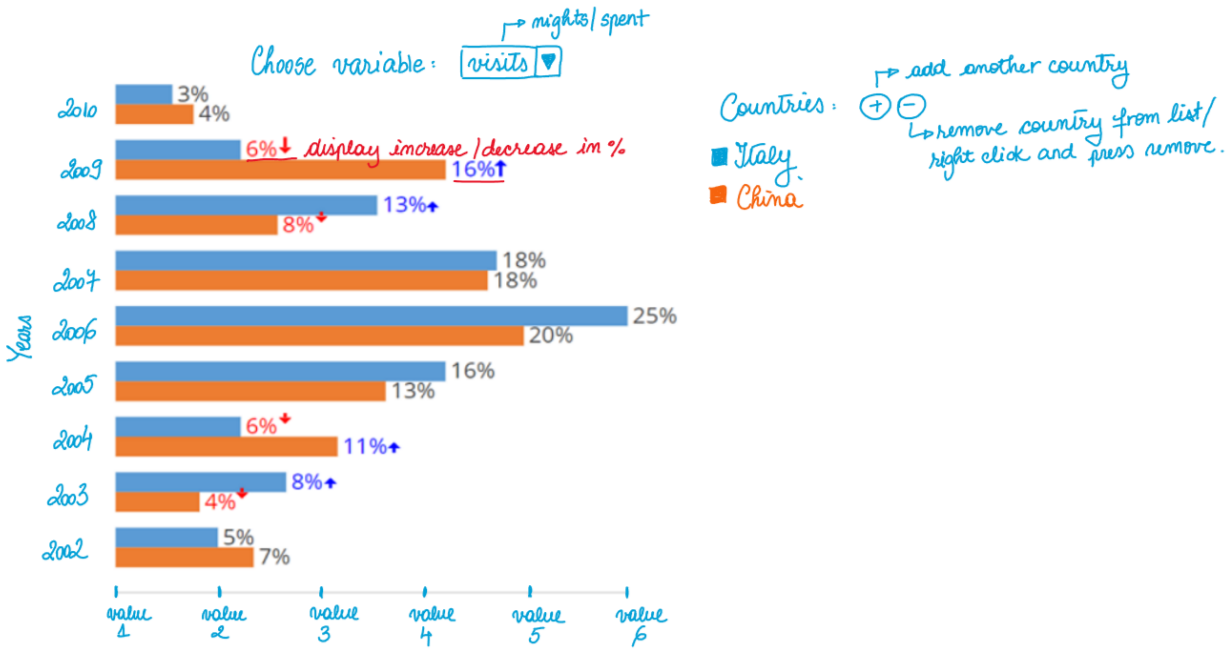



## Charts




mode of travel / duration stay
Choose variable: purpose and country: China
18%
54%
28%
Holiday
Business
VFR
2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020
pause / play / stop buttons animation


nights / spent
range 2002 - 2020
Choose variable: visits, from 2002 to 2010
value 9
value 8
value 7
value 6
value 5
value 4
value 3
value 2
value 1
2002
2003
2004
2005
2006
2007
2008
2009
2010
Argentina
Italy
India
China
Spain
Germany
remove country button / right click on country, press remove
add another country button


Choose variable: visits
nights/spent
Countries:
add another country
remove country from list/ right click and press remove.
Italy
China
display increase/decrease in %
Years
2010
2009
2008
2007
2006
2005
2004
2003
2002
3%
4%
6%
16%
13%
8%
18%
18%
25%
20%
16%
13%
6%
11%
8%
4%
5%
7%
value 1
value 2
value 3
value 4
value 5
value 6

**Appendix E**

A detailed table outlining the variables in each column of the dataset.

| **Columns** | **Variables** |
|---|---|
| Year: | Integer (4) from 2002 to 2020 |
| Quarter: | Q1: January - March |
| | Q2: April - June |
| | Q3: July - September |
| | Q4: October - December |
| Market (List of countries): | Argentina, Australia, Austria, Bahrain, Belgium, Brazil, Bulgaria, Canada, Chile, China, Czech Republic, Denmark, Egypt, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, India, Indonesia, Irish Republic, Israel, Italy, Japan, Kenya, Kuwait, Luxembourg, Malaysia, Mexico, Netherlands, New Zealand, Nigeria, Norway, Oman, Other Africa, Other Asia, Other Central & South America, Other Eastern Europe, Other Middle East, Other Southern Africa, Other Western Europe, Pakistan, Poland, Portugal, Qatar, Romania, Russia, Saudi Arabia, Serbia, Singapore, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Turkey, United Arab Emirates, USA |
| Duration of Stay: | 1-3 nights |
| | 4-7 nights |
| | 8-14 nights |
| | 15+ nights |
| Mode of Travel: | Air |
| | Sea |
| | Tunnel |
| Purpose of Travel: | Holiday |
| | Business |
| | Study |
| | VFR |
| | Miscellaneous |
| Visits: | Float numbers (000s) |
| Spend: | Float numbers (£m) |
| Nights: | Float numbers (000s) |
| Sample: | Float numbers |

## Appendix F

An example of one part of the dataset that has missing data points.

| LONDON | Visits (000s) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Country of origin | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
| Argentina | 14.6 | 21.5 | 23.2 | 20.9 | 28.4 | 30.5 | 20.1 | 55.4 | 52.4 | 61.1 | 93.7 | 93.5 |
| Australia | 436.9 | 450.7 | 477.7 | 552.6 | 578.8 | 595.3 | 583.7 | 571.2 | 624.3 | 681.3 | 596.9 | 687.3 |
| Austria | 109.5 | 129.0 | 128.7 | 122.3 | 140.8 | 163.7 | 141.7 | 168.8 | 165.9 | 164.9 | 153.0 | 142.8 |
| Bahrain | | | | | | | | | | 12.1 | 21.5 | 23.9 |
| Belgium | 288.6 | 306.2 | 359.2 | 300.5 | 342.8 | 351.1 | 307.7 | 349.9 | 488.8 | 406.4 | 470.3 | 530.7 |
| Brazil | 64.7 | 46.1 | 58.6 | 65.4 | 91.5 | 108.9 | 150.9 | 108.4 | 149.2 | 211.3 | 215.2 | 220.0 |
| Bulgaria | | | | | | | 44.3 | 55.7 | 49.3 | 52.6 | 53.9 | 55.2 |
| Canada | 358.3 | 317.2 | 368.9 | 399.2 | 439.5 | 478.1 | 499.7 | 371.2 | 389.5 | 428.8 | 402.9 | 434.5 |
| Chile | | | | | | | | | | 18.7 | 16.7 | 29.0 |
| China | 39.5 | 42.6 | 61.0 | 58.1 | 64.8 | 79.2 | 60.1 | 46.7 | 55.4 | 80.6 | 104.1 | 96.4 |
| Czech Republic | 63.6 | 79.7 | 126.1 | 106.9 | 112.0 | 134.1 | 169.7 | 144.8 | 109.7 | 108.7 | 144.6 | 136.5 |
| Denmark | 191.4 | 209.8 | 238.1 | 286.7 | 291.3 | 312.3 | 298.4 | 332.2 | 273.8 | 340.2 | 325.7 | 378.5 |
| Egypt | 19.7 | 25.0 | 21.6 | 19.5 | 33.6 | 21.8 | 31.6 | 30.5 | 31.0 | 29.9 | 35.0 | 39.1 |
| Finland | 66.3 | 59.3 | 78.1 | 97.5 | 102.6 | 99.4 | 94.9 | 94.4 | 103.6 | 124.5 | 130.7 | 134.2 |
| France | 1,102.9 | 1,190.9 | 1,258.1 | 1,367.3 | 1,447.5 | 1,342.4 | 1,409.9 | 1,528.2 | 1,624.3 | 1,600.5 | 1,681.3 | 1,904.0 |
| Germany | 889.4 | 891.7 | 1,172.9 | 1,280.9 | 1,274.7 | 1,196.6 | 1,052.4 | 1,068.9 | 1,257.4 | 1,214.5 | 1,199.1 | 1,295.3 |
| Greece | 97.8 | 96.6 | 109.8 | 111.8 | 102.5 | 99.6 | 124.4 | 116.9 | 104.5 | 134.4 | 92.3 | 110.8 |
| Hong Kong | 96.1 | 78.8 | 82.9 | 88.9 | 96.6 | 85.1 | 84.5 | 86.7 | 84.4 | 97.9 | 92.7 | 99.5 |
| Hungary | 43.7 | 45.1 | 83.5 | 120.9 | 119.5 | 149.1 | 108.2 | 102.4 | 75.6 | 80.7 | 104.9 | 88.2 |
| Iceland | 20.1 | 23.7 | 49.8 | 43.0 | 46.7 | 46.1 | 32.3 | 23.7 | 19.9 | 24.5 | 25.4 | 29.0 |
| India | 135.2 | 130.4 | 160.3 | 166.0 | 229.6 | 219.7 | 241.6 | 177.9 | 235.9 | 235.3 | 231.5 | 243.6 |
| Indonesia | | | | | | | | | | 19.2 | 18.5 | 17.9 |
| Irish Republic | 628.9 | 623.6 | 682.8 | 726.5 | 718.4 | 732.3 | 736.1 | 690.4 | 656.6 | 620.3 | 595.9 | 610.6 |
| Israel | 152.7 | 122.7 | 127.4 | 106.9 | 138.4 | 102.7 | 125.7 | 124.9 | 122.8 | 120.5 | 108.2 | 141.0 |
| Italy | 541.6 | 601.4 | 765.4 | 643.7 | 851.2 | 837.7 | 981.0 | 747.6 | 932.2 | 1,009.3 | 962.8 | 1,072.2 |
| Japan | 294.5 | 232.5 | 265.0 | 242.9 | 245.2 | 220.1 | 167.8 | 180.5 | 168.2 | 166.4 | 177.9 | 155.3 |
| Kenya | 31.6 | 19.8 | 22.1 | 16.5 | 15.9 | 12.7 | 24.6 | 14.9 | 18.5 | 17.8 | 20.8 | 22.3 |
| Kuwait | 31.1 | 36.2 | 32.5 | 33.8 | 43.6 | 46.7 | 35.4 | 49.0 | 41.9 | 52.8 | 50.5 | 105.7 |
| Luxembourg | 23.9 | 16.9 | 38.4 | 25.9 | 28.2 | 31.7 | 35.0 | 27.9 | 26.8 | 40.0 | 40.1 | 51.2 |
| Malaysia | 60.5 | 46.7 | 65.0 | 52.8 | 52.7 | 47.0 | 45.6 | 77.9 | 89.8 | 82.5 | 72.0 | 113.4 |
| Mexico | 53.0 | 57.9 | 52.2 | 55.0 | 69.5 | 38.1 | 50.4 | 60.7 | 51.9 | 62.2 | 69.8 | 91.4 |
| Netherlands | 492.4 | 619.2 | 578.3 | 606.7 | 659.8 | 665.3 | 654.1 | 683.9 | 621.2 | 631.2 | 637.9 | 686.7 |
| New Zealand | 82.7 | 100.5 | 101.9 | 121.7 | 124.5 | 119.7 | 103.1 | 108.8 | 111.3 | 95.5 | 104.5 | 93.7 |
| Nigeria | 93.0 | 85.6 | 120.6 | 119.8 | 118.4 | 125.0 | 113.6 | 116.5 | 124.6 | 106.1 | 109.9 | 120.7 |
| Norway | 178.4 | 182.0 | 218.0 | 265.9 | 302.2 | 306.6 | 337.6 | 324.1 | 372.1 | 382.1 | 396.8 | 480.1 |
| Oman | | | | | | | | | | 8.1 | 8.5 | 12.1 |
| Other Africa | 102.6 | 100.9 | 113.5 | 93.5 | 98.6 | 110.8 | 83.3 | 93.9 | 87.8 | 77.1 | 88.5 | 113.4 |
| Other Asia | 68.6 | 56.0 | 65.5 | 76.0 | 60.8 | 73.4 | 71.8 | 51.9 | 59.7 | 59.6 | 59.9 | 75.7 |
| Other Central & South A | 106.0 | 98.0 | 103.2 | 108.2 | 108.5 | 119.2 | 76.1 | 90.5 | 84.6 | 85.0 | 84.8 | 128.3 |
| Other Eastern Europe | 79.3 | 74.4 | 86.2 | 285.3 | 404.8 | 485.0 | 352.5 | 266.9 | 261.2 | 291.3 | 291.1 | 307.6 |
| Other Middle East | 80.6 | 66.9 | 83.2 | 77.8 | 97.0 | 93.2 | 91.6 | 115.1 | 97.4 | 55.8 | 45.5 | 40.4 |
| Other Southern Africa | 19.0 | 11.9 | 13.6 | 12.2 | 11.4 | 12.3 | 11.7 | 5.6 | 1.8 | 21.3 | 19.4 | 17.0 |
| Other Western Europe | 80.3 | 75.1 | 101.9 | 81.8 | 106.9 | 122.8 | 118.2 | 122.6 | 96.6 | 94.1 | 94.8 | 103.5 |
| Pakistan | 43.0 | 35.6 | 49.2 | 58.3 | 66.1 | 55.5 | 47.7 | 38.4 | 32.5 | 31.4 | 40.6 | 39.7 |
| Poland | 115.3 | 167.2 | 277.8 | 425.2 | 528.2 | 432.7 | 451.3 | 359.6 | 355.7 | 353.7 | 396.2 | 419.2 |
| Portugal | 97.8 | 97.5 | 97.0 | 113.6 | 120.8 | 144.1 | 133.8 | 136.7 | 172.4 | 165.6 | 153.0 | 138.0 |
| Qatar | | | | | | | | | | 24.6 | 31.7 | 43.6 |
| Romania | | | | | | | 101.1 | 107.8 | 121.9 | 127.9 | 109.8 | 158.6 |
| Russia | 143.2 | 160.1 | 193.4 | 133.2 | 159.6 | 165.1 | 125.0 | 85.7 | 119.1 | 154.4 | 155.3 | 147.2 |
| Saudi Arabia | 42.8 | 34.1 | 37.7 | 37.0 | 39.5 | 56.5 | 63.8 | 58.6 | 54.2 | 66.8 | 74.4 | 90.9 |
| Serbia | | | | | | | | | | 18.8 | 17.8 | 16.3 |
| Singapore | 57.5 | 52.6 | 57.2 | 49.9 | 70.9 | 63.4 | 77.0 | 64.7 | 89.2 | 88.9 | 106.7 | 91.5 |

## Bibliography


Agile Business Consortium. (2019). *MoSCoW Prioritisation*. Retrieved December 2020, from https://www.agilebusiness.org/page/ProjectFramework_10_MoSCoWPrioritisation

Bondarenko, A. (2017). *How to Write a Good User Story: with Examples & Templates*. Retrieved December 2020, from https://stormotion.io/blog/how-to-write-a-good-user-story-with-examples-templates/

Budget Your Trip. (2020). *Travel Budget for London*. Retrieved December 2020, from https://www.budgetyourtrip.com/united-kingdom/london

Champion Traveler. (2020). *Champion Traveler*. Retrieved December 2020, from https://championtraveler.com/

Condor Ferries. (2020). *UK Tourism Statistics 2020*. Retrieved December 2020, from https://www.condorferries.co.uk/uk-tourism-statistics

Horsfield, G. (2017). *International Passenger Survey*. Retrieved December 2020, from https://www.tmi.org.uk/files/documents/resources/GilesHorsfield-ONS-InternationalPassengerSurvey.pdf

London First, EY. (2019). *Tourist Information: Mapping the local value of international visitors.* London First.

McCandless, D., & Quick, M. (2020). *Information is Beautiful*. Retrieved December 2020, from https://informationisbeautiful.net/visualizations/top-500-passwords-visualized/

Office for National Statistics. (2017). *Travel Trends: 2016*. Retrieved December 2020, from https://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/articles/traveltrends/2016

Office for National Statistics. (2020). *International Passenger Survey (IPS) Methodology*. Retrieved December 2020, from https://webarchive.nationalarchives.gov.uk/20160115212418/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/travel-and-transport-methodology/international-passenger-survey/index.html

Office for National Statistics. (2020, September). *Number of International Visitors to London.* Retrieved December 2020, from London Data Store: https://data.london.gov.uk/dataset/number-international-visitors-london

Office for National Statistics. (n.d.). *Data Principles*. Retrieved December 2020, from https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/dataprinciples

Sonego, D. (2015). *Document Analysis*. Retrieved December 2020, from https://babokpage.wordpress.com/techniques/document-analysis/

Sonego, D. (2017). *Brainstorming*. Retrieved December 2020, from https://babokpage.wordpress.com/techniques/brainstorming/

StatCounter. (2020). *Screen Resolution Stats Worldwide*. Retrieved December 2020, from https://gs.statcounter.com/screen-resolution-stats

Visual Paradigm. (2020). *Requirement Analysis Techniques*. Retrieved December 2020, from https://www.visual-paradigm.com/guide/requirements-gathering/requirement-analysis-techniques/