**4216: Specification Document**Team 6  
  
**Contents:**

|  |  |  |
| --- | --- | --- |
| **Section** | **Detail** | Page No. |
| **1** | Project Synopsis |  |
|  | Design Stages |  |
| **2** | Team Building |  |
| **3** | Project Management Approach |  |
| **4** | Data Chosen and final selection |  |
| **5** | Assessment, Insights & Extraction |  |
| **6** | Algorithm exploration and optimisation |  |
| **7** | Concept/Logical/Implementation Stage |  |
| **8** | Design that identifies interaction with user |  |
| **9** | Description of the implementations |  |
| **10** | Testing |  |
| **11** | Evaluation |  |
| **12** | Project Closure |  |
| **13** | Leaning |  |
| **14** | Conclusion |  |

**Cohort Members:**   
Cameron March  
Ethan Ibrahim  
Laura J. Phillips  
Luke Citirne  
Luke Curran, Niamh Walsh  
Shane Mendes  
  
  
  
  
  
  
  
  
**Project Synopsis**The main project required us to build a successful and supportive team, with a collaborative design, in which we would design and implement a data analysis tool that could output the meaningful information and have it displayed in graphs, it would also allow the user to have the ability to sort through the data in the python shell in terms of narrowing down the larger set of data by splitting it into years, months and seasons.  
  
We decided as a team to use a standard project management techniques in order to manage the project, and it was early agreed that each team member would be responsible for leading the team for two weeks to perform a programming and administrative function and also provide the use of ad-hoc support to each other. This was going to ensure that the project as a whole would run smoothly as well as allowing all team members to obtain a fulfilling an successful learning experience and would output a results in a successful project delivery with all the team members working together.  
  
**Design stages**When it came to the acquisition of the data that we would be using for the project, as a team we considered several such data sets, such as climate change as well as criminal activity data however both of these data sets where rejected for reasons explored such as incomplete records or other things such as low volume of the dataset (not giving us enough data to create a reasonable application, or effective visualisations. After which, we evaluated whether temperature records would allow us to successfully allow us to give the required output as stated in the specification.   
  
We concluded as a team that the new data set, we acquired would suffice, and we decided that it was suitable for us to use with over 13,000 quality records that covered a wide range of time period including 1900 to 2013. Project readiness explores the process that is used by management companies to assess the completeness of a project and preparation to date, which identifies key information gaps and records that team members must take, and make to ensure that the project we create will suffice for the specific criteria that we must fulfil. This allows us to create a system in which the actions that are needed to be taken are enforced and completed, allowing for a robust working system for the entire team, and allows time for mistakes in terms of project preparation.  
  
**Team Building**Early on in development the team decided that it was important to plan for specific concepts that where asked of us when it came to the project, as such was the team building that would be acquired and assessed throughout the project – therefore allowing the team to work closely and well with each other, but also ensure that each member of the team Table

Description automatically generated(and the rest of team) where aware of each other’s strengths and weaknesses before, during and after (improvement) of the project itself. Allowing skillsets and assets of the team to be used in the best places possible to ensure the best possible outcome in any specific tasks that would be set out to be achieved by members.  
  
As a team we needed to appreciate our individual skill strengths and weaknesses for each member and the table provided illustrates our standing position when it came starting the progress on the project and easily highlights each team members strengths and weaknesses, allowing for room for growth, learning, and education on such areas as stated. Although the scores recorded where not captured scientifically, the exercise did help us understand the potential roles each of us could naturally play in the birth and conception of the project and beyond this on that scope. It also easily highlighted what was expected of each team member. It also allows us to view our entire team strengths and weaknesses to assess how we would work as a team.  
  
Graphical user interface, application

Description automatically generated**Project Management Approach**Design and concept of the document, as well as assessing the team to create a solid foundation for the project is one thing that works well together, but this doesn’t cover how the project was managed throughout, and fails to cover the systems that where put into place to ensure that the tasks where completed on time.   
  
Within GitHub, we had created a repository which was used mainly to share the files that would be used for the project so all team members could easily share, discuss, access and change files for the project. We created this in a way in that its organised into folders and quickly and easilly accessable by all members of the repository. People would know what they where doing, where it was, and that it could be easily accessed by any of the team members in a quick fashion without there being any hassle. **These assets where instrumental in helping us easily manage the project effectily and easily.**Graphical user interface, application, email

Description automatically generatedHere is a look into the planning documents that we had created in order to easily manage the project to the most efficient degree, which could be easily accessed and edited by the team, allowing us as well to easily show growth and changes in the team, from the start of the project, right through to the end. We believed this was the most effective way in not only showing how we work as a team, but how we evolved and learnt (from each other) during the project.  
  
Graphical user interface, text, application, Word

Description automatically generated  
  
Above, you can see that we used specific documents at the end of each meeting to make notes of everything that was discussed by each team members, the actions that where taken during the meeting and what was discussed; done in a format on a week to week basis that allowed team members who hadn’t otherwise attended to quickly ascertain what was achieved from the previous meeting. This is **IN ADDITION** to the regular team meetings we would have to heavily enforce the ease of the structure, but to ensure everything is organised and everyone is busy, working and progressing for the greater good.  
  
Additionally, this allowed everyone to be aware of how the project was growing, split into sections for easily clarity, to make this as easy and as simple as possible. As well as this we had other documents to show growth of the project itself, moving deeper into the production of the project itself and the impact it has on us as a team. Think of the management of the project like a layered cake; A GitHub with documents concluding the meeting, then documents above this one, covering the nitty-gritty of the project itself, and a fine-tooth deep dive into every task that needed doing.  
  
  
Table

Description automatically generated  
  
Above is an example of our “Milestone Plan” as explained above, this allowed us to easily refer to points or plots in the project, with relative ease; this split up into sections, set out in such a way that it can be easily edited and changed/added too. Stating the scope of the task at hand, weather it was completed, what section it belonged too (GitHub, Pre-start, Project Start, etc) As well as the person doing that project the start and required date for the task to be finished, structured in a chronological order, as well as highlighting in which sections of the skills it would require and use (using this to assign the best person to a job)  
  
This document and structure works in the same way as the meeting minutes, which as stated previously can be easily referred to, created in such a way it’s as easy to use and edit, but allows team members to, in a matter of seconds understand: Project progress, current tasks, time frames, etc.  
Text

Description automatically generated In projects like this communication is the most important thing! Following the heavy pre-planning of the structure of how the project would run, and how the project would be managed, we all agreed that communication was the best way to ensure that a project flows is communication. The first thing we did as a team was set up a WhatsApp group so that (as well as the aforementioned documents) that everyone is aware of the progress, issues or support request, first hand, as soon as required, requested or posted.  
  
We then agreed to set up regular face-to-face meetings at 12:00PM, in room 173 before our tutorial had begun, which would start an hour later – We would then continue that meeting until 13:30PM, this provided invaluable as we where easily able to manage the workload, discuss any issues effectively, support each other and set up action plans for this, adding this to our meeting minutes. As mentioned previously; each week we recorded the minutes of the meeting of the minutes and uploaded them for GitHub for the reasons explained above.  
**Data Chosen and Final Selection**Diagram

Description automatically generated *Fig. 1: Original data overview.*

|  |  |
| --- | --- |
| Reduce the number of records available for analysis. | 13,000 global weather records were too many for us to process, so we chose to trim down the volume. |
| Simplify data content | By filtering UK records out from the global data we obtained a more manageable data set of around 1300 records. (see fig. 1 above). |
| Data split | This data was then allocated evenly amongst all seven team members. |

The data we chose was UK weather temperature data from the years 1900-2013, made up of average temperature measurements from the first of the month throughout each of the 113 years. It covers both land and ocean temperatures, with twelve records for each year in this period. Originally, the data was in three tables: Global Temperature, Country Land Temperature, and City Temperature (see fig. 2 below). When it came to the data considered by the group and the rational for the final selection of the data, there were several different data sources that where consider by the team to begin with: To start, the climate change data that we had looked at and secondly the criminal data that we had explored.  
  
Both of these where unsuitable we found for a number of reasons, these include records of the data being incomplete as well as some of the data having small quantity of records overall, which wouldn’t allow for a interesting or useful data visualisation when it came to creating the larger project and eventually presenting. This has since been discounted with the third piece of data that we reviewed, we accepted as a team because of the large number of records that we have (13,000+) and all data entities where complete and useable, which allowed us a large base of data to choose from. We feel that the data set we chose has been flexible in the 3 larger areas that did data allowed us to choose from and this also provides meaningful comparison between those three key areas, these three key areas being, but not limited to the Land Minimum and Maximum temperature as well as the Land and Ocean temperature average over time, as well as the land average and land maximum temperatures as well as their uncertainties.  
  
As seen in figure one above, the data we focused on was on a global appeal and was considered because of the data that was available, which covered global temperatures, however due to the size of the source it would have been too cumbersome for us to use as this project, however we were able to remove and narrow down the data for specifics which would include for London UK, and this data set has been underpinned for the data that we would use for the tool.  
  
Diagram

Description automatically generated  
  
*Fig. 2:* an *entity relationship diagram for the original three tables, showing the fields, relationships, and linking keys for each, and is used to demonstrate the complexity of the tables and led to the reason why we reduced the functionality to London alone.*

**Assessment, Insight and Extraction**We decided to combine all of these into one table for ease of access and data cleanliness. When first analysing the data for eligibility, we split the data into equal chunks of eighteen years, which made it easier to analyse. As such, we decided to take a similar approach when drawing up our list of requirements for the software design.

It was our intention that the user should be able to utilise the historical data throughout the period to be able to identify changes between and patterns in the temperatures, by allowing the user to search and filter different seasons and months for years. We needed the data to be easy to manipulate, reflecting this in our chosen design: the data will be explored by the user mixing and matching various filters, including month, season, and year. The landing page will have options for all three, though within each the user will be able to add the other filters should they wish to do so. The result will be a dynamic graph, either a pie chart or a bar chart depending on which is most appropriate, that changes with the data the user chooses.

Assessment of the potential insights that might be extracted from the data set and the interfaces that are needed to interact with application from the user’s side including the requirements. The insight of the chosen data is the visualise the average temperature in London within a 130-year period, this analysis can be further broken down to have the ability to show patterns over years, seasons or months. The system that we’re developing is designed to provide the end user with the functionality capability for searching (or breaking down) the temperature data between 1900 and 2013. This is underpinned by the 13,000+ records captured during this period. We got the data from:

[**https://www.kaggle.com/yamqwe/global-climate-change-datae?select=GlobalLandTemperaturesByMajorCity.csv**](https://www.kaggle.com/yamqwe/global-climate-change-datae?select=GlobalLandTemperaturesByMajorCity.csv)**.**

Kaggle provided the data that was used to develop the tool that we would be developing, and as stated previously, we had taken the large amount of data and broke it down to just show London, UK, which gave us more than enough data to use to create interesting visualisations and a reasonable working application. The user page for viewing all the data you can either view this from the command shell in which the user is able type the kind of data that they want to view, this allows users (in a way) to narrow down the type of data that they’re wanting to see and use and have insight on. This allow the data to be more flexible as well as making sure that it’s dynamic and useful, as this user drills down on the data with the data changing, being dynamically changing graphs shown here.

The interactive interfaces will provide insights by the user on the following things: Seasons and pie charts, range of temperatures across the years that are shown (Bar charts, which will be broken down) as well as the Years/Bar chart/Horizontal charts. The 113 years will be split into the 16 data points amongst the access. With the user giving the input that they can use using the shell in Python.

The potential analysis from the tool that we could provide often correlates to the assessment between the land and the temperature and potentially backed with Ocean temperatures and vice versa, however, due to the time constraints, this is complexity may not be guilt into the tool at this time. The software that was used to develop the tool would include software packages such as Visio Studio, Python, Matplotlib, whilst they themselves may seem complex, together it allows us to complement the objective of the assignment and the tool that we’d be creating by allowing us to develop the code in python but provide the user with visualisation of the data which is split dynamically and generated by the user input – the data itself is changing, providing the dynamic visualisation of the data itself.

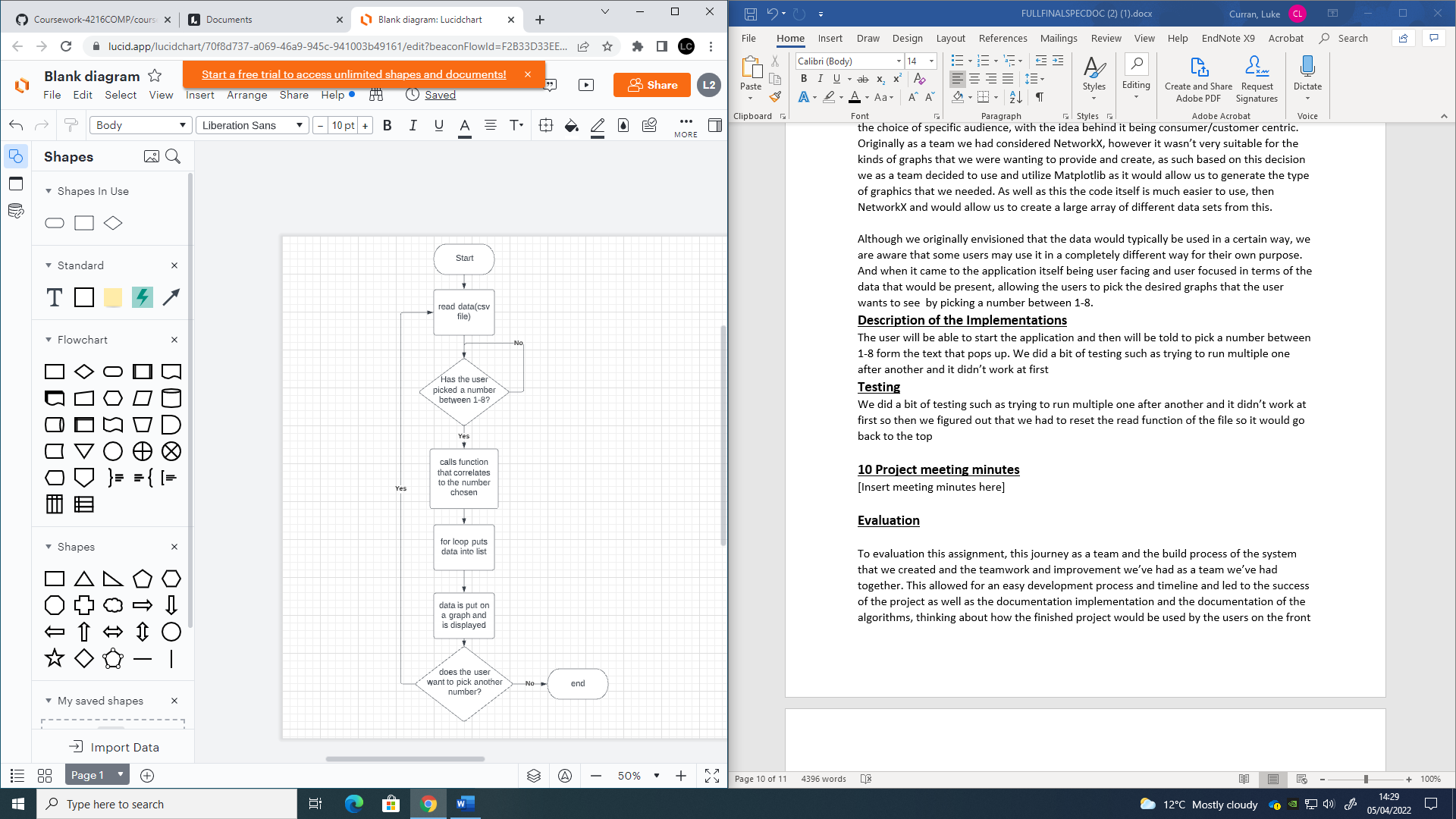
**Our requirements for the software:**

* Allow the user to filter data by:
  + User choice on what graph they want
* Produce graphs that changes with the data the user filters
  + line charts
  + scatter charts
* Be intuitive and easy to use
* Read from the .CSV file to obtain the data needed
* Allow the user to draw conclusions as to temperature patterns in the data
* EXT: write to a .txt file to produce a downloadable file of filtered data.

**Algorithm exploration, optimisation and justification**We have decided to utilise the Matplotlib library to create our graphs, as it would allow us to do so dynamically by simply changing the values of the variables assigned, rather than needing to recode the graph each time a new filter is added. It is our hope that by making use of dynamic graphs, the user can easier draw conclusions from our data, thereby achieving point 5 in our software requirements while achieving point 2.

We have also decided to utilise semi-identical functions for each filter; as we intend to import the data into a list, functions would streamline the code, allowing for a simpler and more efficient program. Each search function – one for seasons, one for months, and one for year – would be almost identical, simply swapping out variables as needed.

When it comes to the algorithmic exploration of what we’d be using with a description of any optimisations that would be made, for this we took the process as such that we were going to read the data from a large data file (data set) and then allow the user to input the type of data that they wanted to view (from the command shell in python), allowing them to narrow down or (search) the data, including year, season and month, an example of this could be when they’re wanting to look at the seasons for all 113 years, or the winter seasons for those years for example. The code starts off with imports ,we need ed to import CSV and matplotlib to get the graphs working, after that we had to read in the file after that we made it so that the file would be read by each row, and every time it hits a column that needs to be indicated by an index, it would take that piece of data and store it into an empty list that has already been made. After which , we used the matplot lib to start making the graph by setting the axis by importing the lists on the axis, next it was to make general aesthetic to setting the style of the ticks and the style of them, and the range and also setting the overall style of the graph itself.



**Concept, Logical and Physical model**The optimisation that we’d mentioned earlier about already and that we looked at the design needs and identified the areas in which we could re-use the code with adaptions simply with the elements such as variables and use their names, when it came to the individual visualisations that were being created, and then fed into the larger project as a whole, this was an important step because it ensures that all the visualisations can work together with each other without any breaking, as well as allowing them all to work as intended, this also included the variable names. As a team, we followed the data model of the process that includes **concept, logical and physical model**.

As part of the **concept** process we captured what we understood about the data, which was the requirement and when we understood that as well as the rest of the team, as well as on a higher concept level, we were able to challenge our thinking in how we would process and understand the assignment and how we could have a plausible output, when it came to the **logical** side of things and how it’d work.

When we looked at what areas we could simplify when it came to the larger data set such as the volume of data that was present, we had – as mentioned earlier, reduced this to London, UK, this allowed us to simplify the capability of the application as well as the design, but still allowed for the same output as present for a working model. This also gives is the capacity to design a robust and working system rather than as a project failure due to time. The functionality that we had developed, we are confident that it could be replicated with comparative ease to provide a further insight covering the land and ocean temperatures, but as a full team we felt that we could possibly lose sight of the objective, due to overcomplicating the task at hand.

Started off with making a skeleton by making a basic graph that everyone could work off of or to talk about how we decided to have each person’s graph inside their own function so it could be called, at the very bottom there is a necessary input from the user which means to multiple nested if statement that will allow the user to select a chosen function. **Design that identifies interaction with the user**As mentioned above when it came to the design of the system - the system that we’re developing is designed to provide the end user with a functional capability for narrowing down, or (searching) the temperature data between the data events of 1900 and 2013. This is the underpinning of the previously mentioned 13,000+ data set files that we’d narrowed down to ensure that it gave us enough data to work with, but enough data that it would still remain interesting when it comes to the implementation of the applications and visualisations themselves.  
  
Aiming at the general public, this shows that the data visualization more clearly including the choice of specific audience, with the idea behind it being consumer/customer centric. Originally as a team we had considered NetworkX, however it wasn’t very suitable for the kinds of graphs that we were wanting to provide and create, as such based on this decision we as a team decided to use and utilize Matplotlib as it would allow us to generate the type of graphics that we needed. As well as this the code itself is much easier to use, then NetworkX and would allow us to create a large array of different data sets from this.  
  
Although we originally envisioned that the data would typically be used in a certain way, we are aware that some users may use it in a completely different way for their own purpose. And when it came to the application itself being user facing and user focused in terms of the data that would be present, allowing the users to pick the desired graphs that the user wants to see by picking a number between 1-8.  
**Description of the Implementations**The user will be able to start the application and then will be told to pick a number between 1-8 form the text that pops up. We did a bit of testing such as trying to run multiple one after another and it didn’t work at first **Testing**We did a bit of testing such as trying to run multiple one after another and it didn’t work at first so then we figured out that we had to reset the read function of the file so it would go back to the top  
**10 Project meeting minutes**<https://canvas.ljmu.ac.uk/groups/50120/files>   
**Evaluation**  
To evaluation this assignment, this journey as a team and the build process of the system that we created and the teamwork and improvement we’ve had as a team we’ve had together. This allowed for an easy development process and timeline and led to the success of the project as well as the documentation implementation and the documentation of the algorithms, thinking about how the finished project would be used by the users on the front end. We believe that the final program is simple to use and does it’s job, allowing the user to execute the system easily and use simply.  
  
**Learning**  
During the course of the project as a whole, the team’s dynamics changed based on the skills that we learnt from each other and the skills we assessed and attained because of this, many of us where generally surprised with the skills that we thought we had and we’re able to develop as a result of the assignment. This takes us back to the team building that we did earlier on in the project, which allowed us to outline our strengths.   
  
**Conclusion**In conclusion to the project, as a team, we felt as though we performed with maturity as well as respecting one another and we were generally pleased with the outcome of the project and how we evolved and changed around it and adapted to the new tasks that it would bring us on a weekly basis. We had systems in place right from the start to ensure that the procedure of the project would run smoothly, efficiently, and quickly and proved that with the correct planning and management that we were able to complete the project to a passing degree.   
  
Thinking about the overall teamwork of the project, having communication as one of the majority factors in the project, as well as assessing and adapting to the strengths of each team member and seeing how we were at the start and where we are now. The system design that we built allowed us to complete it that satisfied the outset of tasks that we were given from the start and we worked well as a team together, with the success of the system coming from the communication and development of the management systems and documents that we had set out in the start of the project.