

**Technical Report: Social Distancing Problem**

**Authors: Dharmarlou Bowen (BCU)**

**Date: 31/05/2021**

**Wordcount: 1500 MAX**

**Pagecount: TBD?**

**Confidential: YES – CDT DEPARTMENT ONLY**

**Turn the Marking Criteria into an image after you have completed the report!!!**

Contents

[Introduction 4](#_Toc73440425)

[Theoretical Basis 5](#_Toc73440426)

[Empirical Basis 6](#_Toc73440427)

[Conclusions 8](#_Toc73440428)

[References 8](#_Toc73440429)

[Appendix A 9](#_Toc73440430)

[Appendix B 10](#_Toc73440431)

[Appendix C 11](#_Toc73440432)

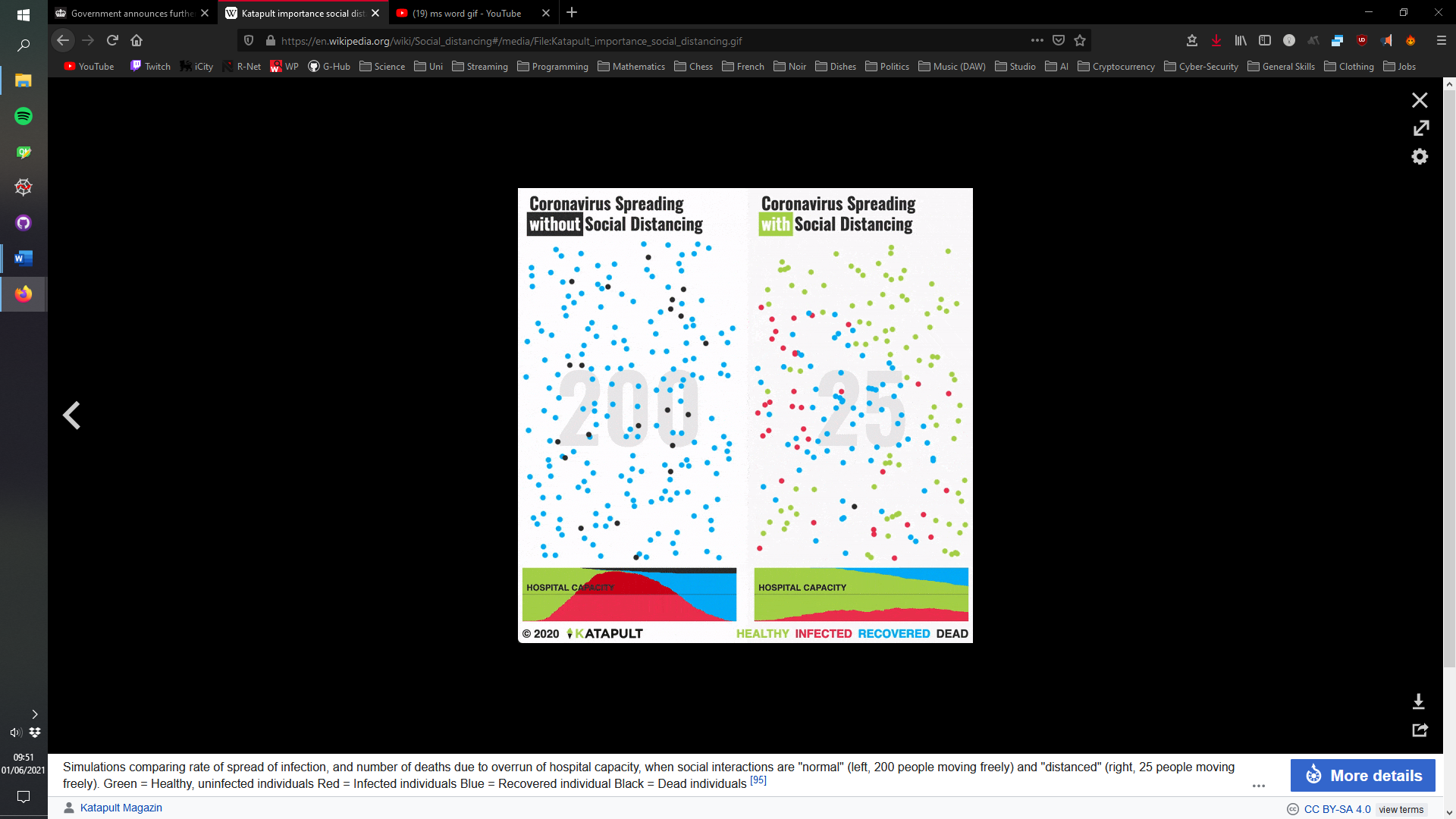
Executive Summary

The ambition of this report is to evaluate the theoretical and empirical development of an algorithmic solution to assist in the *Social Distance Problem* plagues our nation. The solution will distinguish the occurrences of an individual passing within 1 metre of another individual whilst propagating through a supermarket, which is in correlation to a risk factor. This factor appertaining to if an entity’s risk of infection is high or low due to their amount of interactions performed with others whilst in the supermarket premises. The solution will be used to assist current track-and-trace methods used to prevent the spread of the COVID-19 virus. Once the geographical data from the customers have been extracted, the algorithm will have sort individuals into those that express the greatest infection risk to the least. The development of such a solution is highly practical given its indirect ability to not just identify and individual’s risk factor, but if they show signs of infection then their geographical data would present who else they may have infected.

# Introduction

On the 12th of march the UK governing bodies announced that the country was leaving its “contain Phase” and entering its “delay phase”, which gave rise to stricter *Social Distancing* methods as it was noted that over the coming weeks from the announcement that “we will be introducing further social distancing measures for older and vulnerable people” where they would have to self-isolate regardless of showing any symptoms. [**1,** ln: 1-23] This further demonstrated through the developing severity of the rate of infection that led to the addition of more restrictions from the 20th of march onward in the uk. Non-essential services of hospitality such as pubs, bars, and restaurants. As the virus was measured through the months it was noted that such guideline and “measures will be reviewed on a monthly basis” [**2,**  ln: 28]

The image (**Figure 1.0**) demonstrates why avoiding contact is paramount. It’s an image compares two hypothetical simulations of how the virus would propagate through a crowd. The key distinguishing factor being if social distancing guidelines were adhered to.

It expresses that there is a direct correlation between and number of deaths due to a hospital’s capacity being maxed out, when social interactions were orthodox. There is a key at the bottom of the image to present further insight into the colour coding of each entity.

*(Fig 1.0)*

On the left, each person continues common direct interaction within the close proximity of bystanders.

On the right, only 25 out of 200 people move freely. All others self-isolate at home. It takes much longer for everyone to get infected. Hospitals can treat all patients because they don’t arrive all at once. [**3,** ln: 8-14]

This can be used to set the theoretical underpinning for plausible and effective solutions to counter this problem as social distancing is down to the public’s appropriate behaviour.

# Theoretical Basis

With the *social distancing* case study, it provides core sub-problem to be solved is the ability to identify if a person has passed within 1 meter of someone whilst propagating through a supermarket. Once a solution has been developed to perform this, the data can be sorted into those that have the highest infection rate due to interactions in comparison to the lowest. Additionally, it is also important to provide insight into what positions had an occurrence of interactions in correlation to recipients.

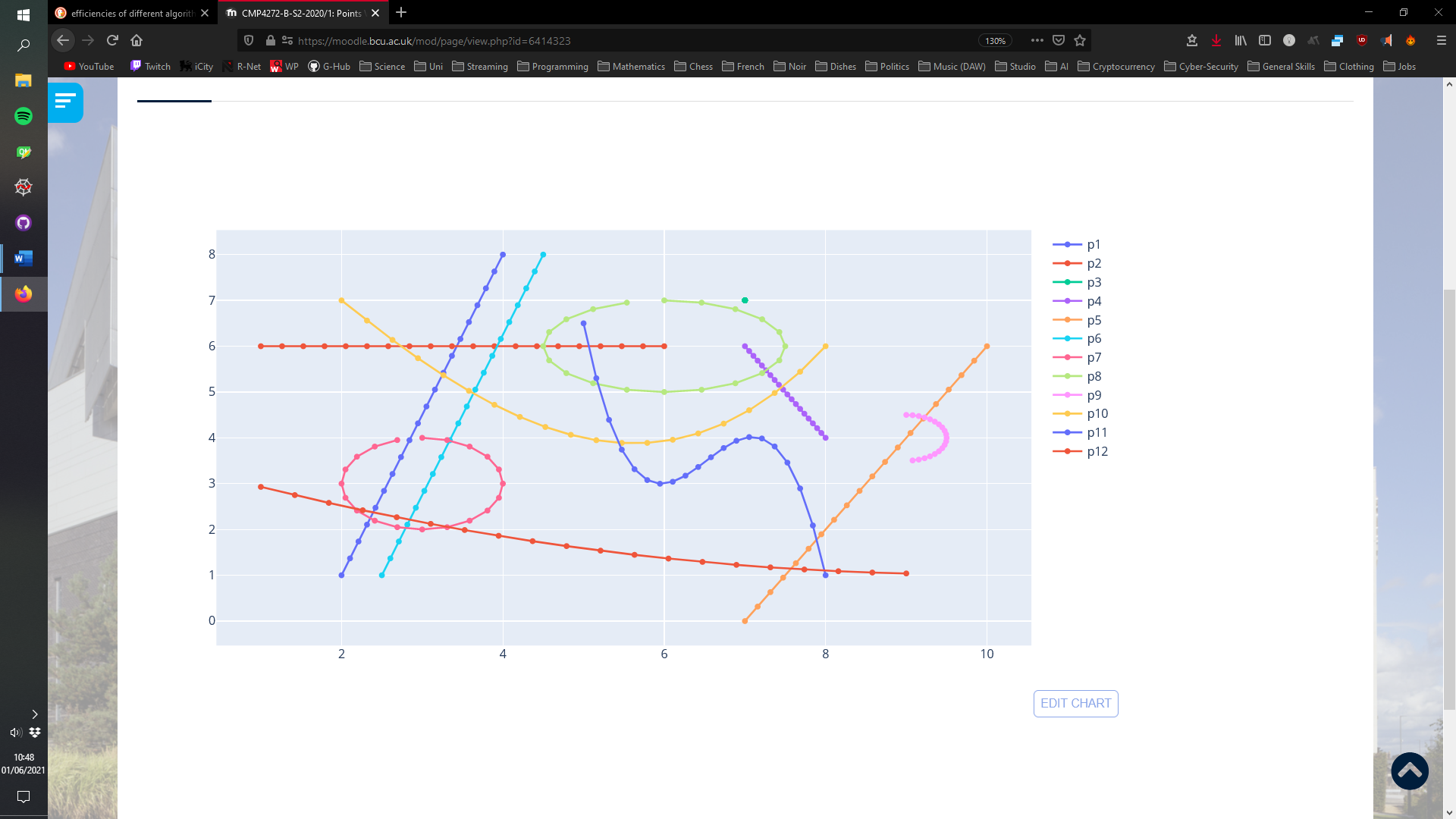
Problem decomposition:

1. Identify interactions
2. Present position in correlation to recipient of interactions
3. Sort people into the highest and lowest probable of spreading infection
4. Planning, Static Efficiency & Correctness
5. What data structures & Algorithms are suited to the problem and why?

Having 3 core sub-problems provide appropriate clarity to decide what data structures and algorithms would be the most suitable to attack such problems. The data that is being used can be graphed as it’s concerned with the geometric data (2D-array) of a person, displaying the following attributes:

* 20 different positions (using x and y coordinates)
* 1 of 12 people

The image (**Figure 1.1**) below is a data visualisation of each person’s propagation through a supermarket at different times. Attempting to analyse interactions through simple observation is possible to an extent but highly inefficient and imprecise if done manually. This becomes especially apparent if larger data sets are being analysed.



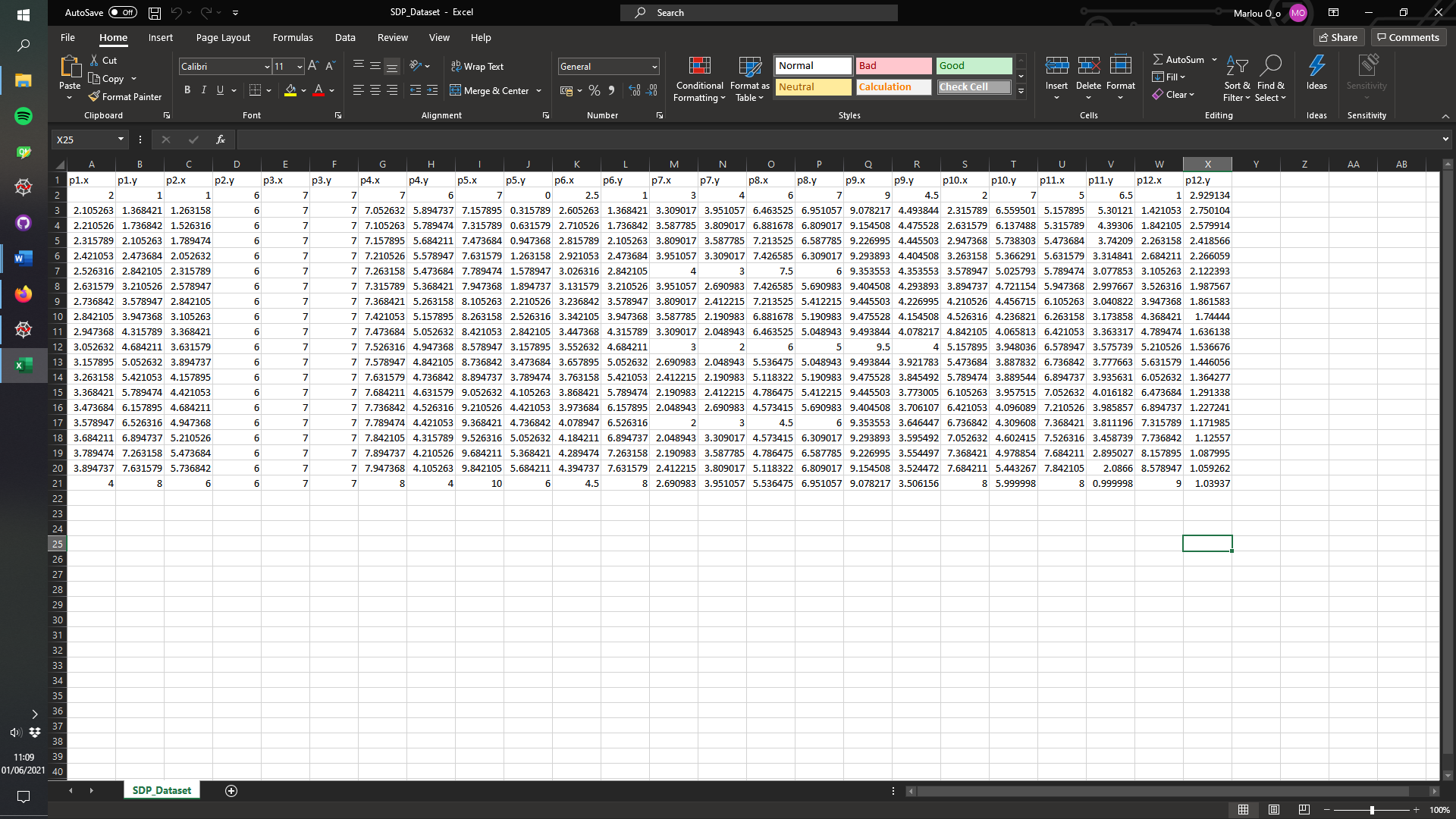
*(Fig 1.1)*

The image was created using visualisation software that uses the coordinates to plot the data on a graph. Acquiring this data is the key foundational task as well as being able to access it without having to rerun any functions to calculate a given coordinate. Storing this data in a CSV file is the most appropriate in this case as accessing it for analysis with the use of libraries such as pandas will streamline workflow.

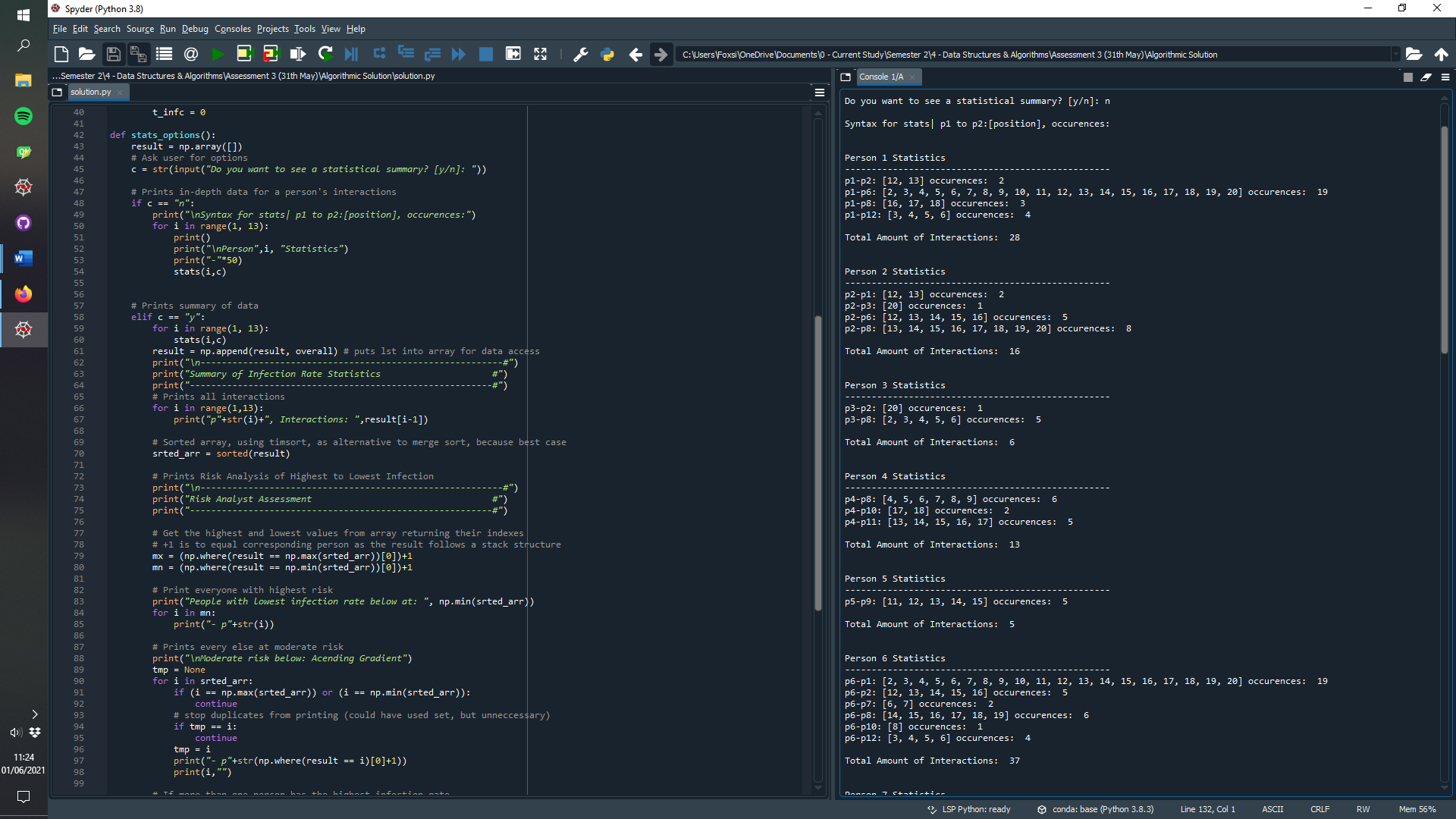
The sorting of data is the last problem to be encountered as multiple functions may be required to perform the correct comparisons between people. As a person has 20 different positions and each position has to be checked against 11 others the amount of check begins to accumulate. Creating a method that can utilise past checks from previous people can help save time on having to perform repeated checked.

If p1 is checked against p2, that data can be used in p2’s collection instead of having to check p2 against p1. From a time, complexity perspective this would put the operations below the input size but depending on the structures used space complexity may be effected.

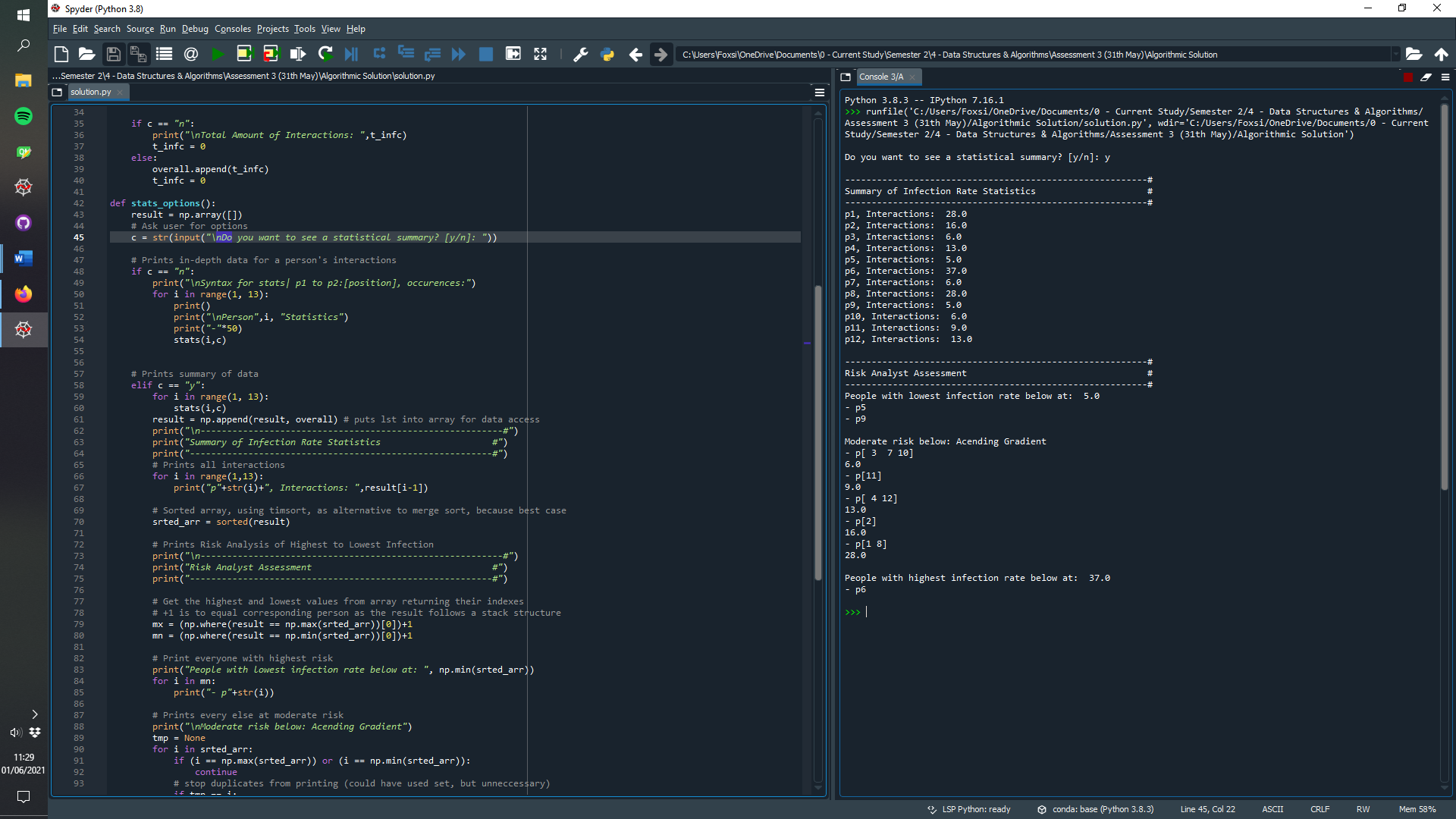
# Empirical Basis

Assertions can be made that a person should have 20 possible position and each position should have 2 coordinates (x and y). This data was exported to csv file (**figure 2.0**) using the function in that can be found in appendix A.

*(Fig 2.0)*



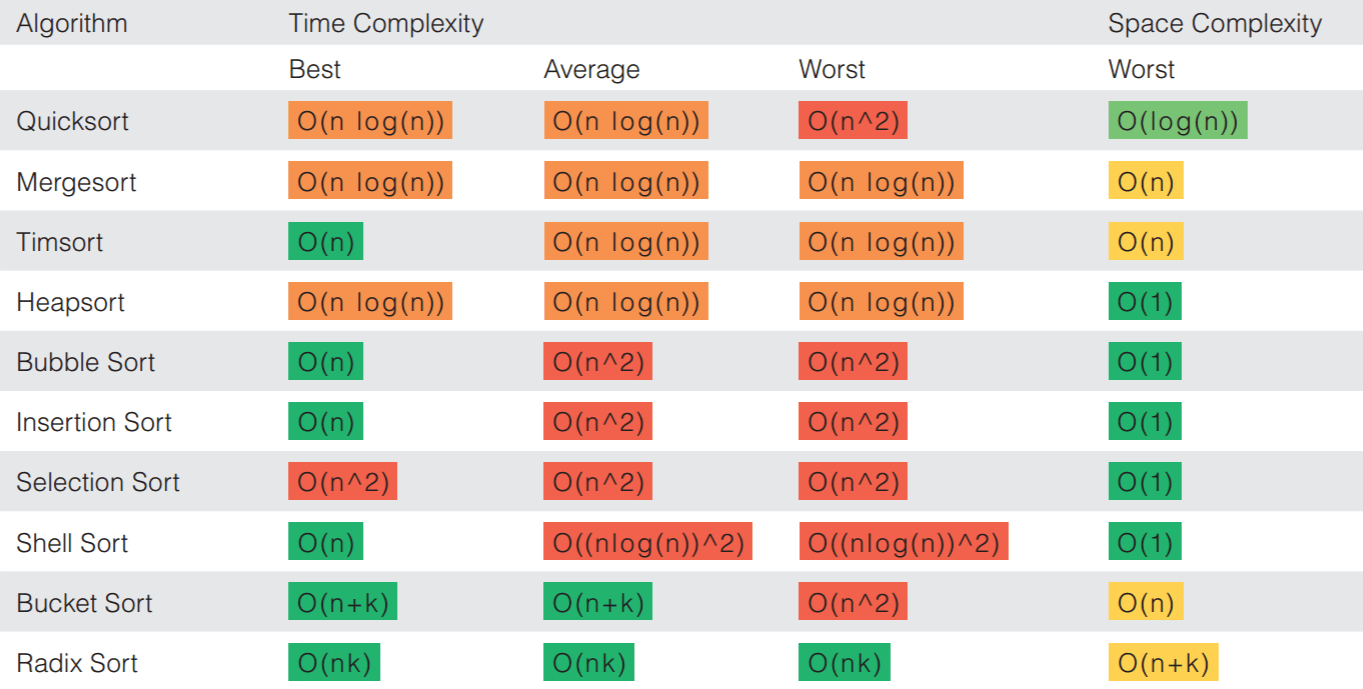
Having the necessary assertions in place would ensure the data was accessible. Figure 2.1 displays some of the output if the user chooses to not see a statistically summary of the data from the code in appendix C. Two decisions are offered to the user. To either display each person’s statistics of who they interacted with (p1 to p2) and in what positions [2,3,4 positions].

In the instance where the user chose to display the summary of infection rate statistics, they would be presented with:

* Total interactions
* Risk analyse of high to low infection rate

For the infection rate the person with the highest total interactions (p6) at 37 would be deemed the most at risk while p5 and p9 only has 5. Everyone in between is considered moderate but ascending in order from highest at the bottom to lowest at the top. p[n1,n2,n3] = multiple had the same amount of interactions.

All of this was achieved through the use of the code in

A dictionary was used for data for position association. Dict = {p1-p2: positions} was the structure followed for storing interactions. Which made keeping track of the positions and recipients simpler as with the appropriate structure.

All this data was then passed to a function that could analyse in appendix C. Timsort was used for sorting the data as its best case of is O(n) due if the input is below a certain fresh hold, the data sorted was only 12 inputs as each person has a total interaction value. Merge sort’s best case does not make a great alternative in comparison.

# Conclusions

I was able to analyse and produce a solution to output the person at highest risk of infection and the lowest. This was the initial goal of the problem to be solved, but additional functionally as implemented to accommodate additional functionality to display interactions at different positions.

# References

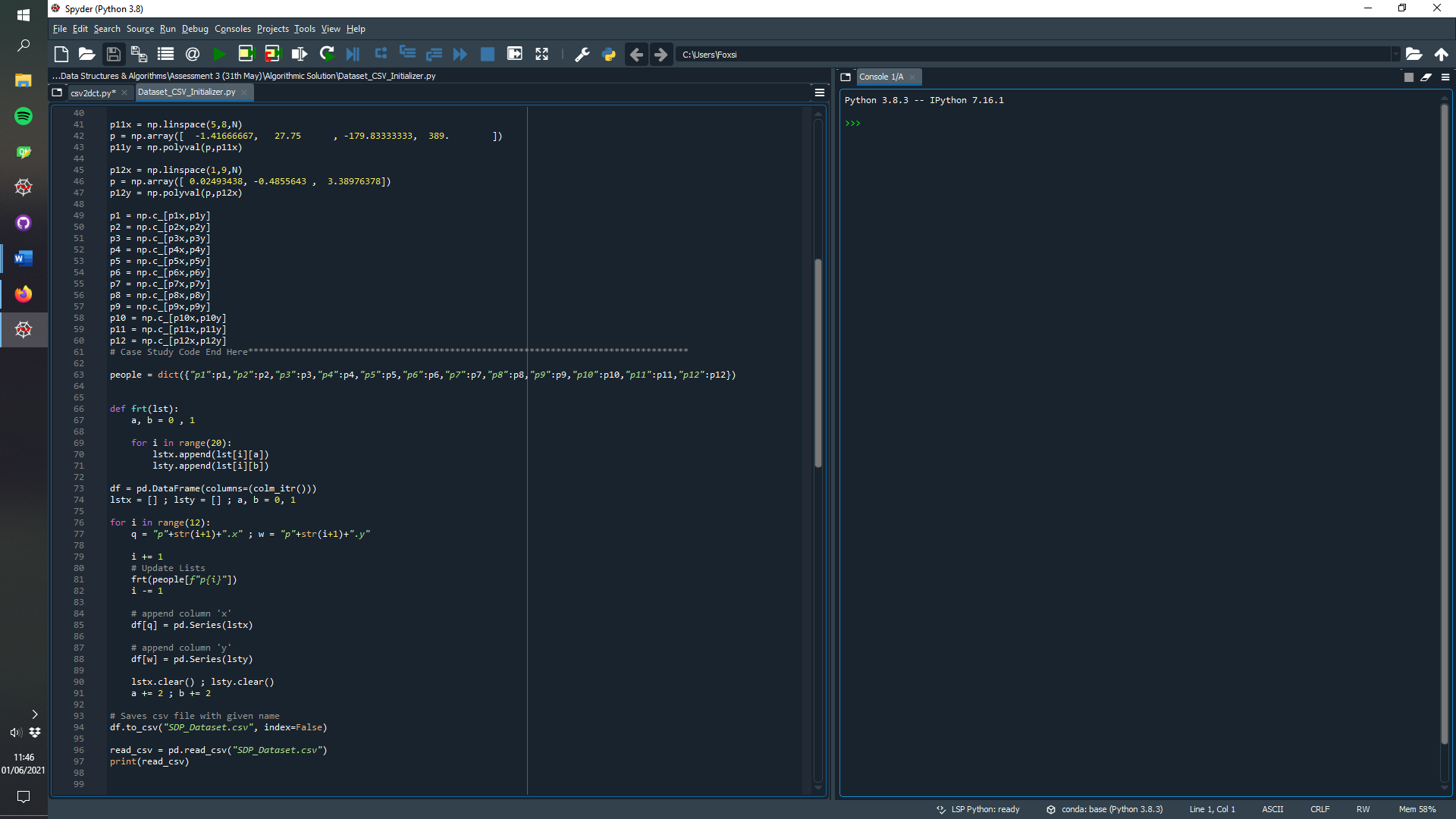
Refencing of any other sources used to support report, Examples below:

[1] Department of Health and Social Care, Press Release, “*COVID-19: government announces moving out of contain phase and into delay*” (12 March 2020) :<https://www.gov.uk/government/news/covid-19-government-announces-moving-out-of-contain-phase-and-into-delay>

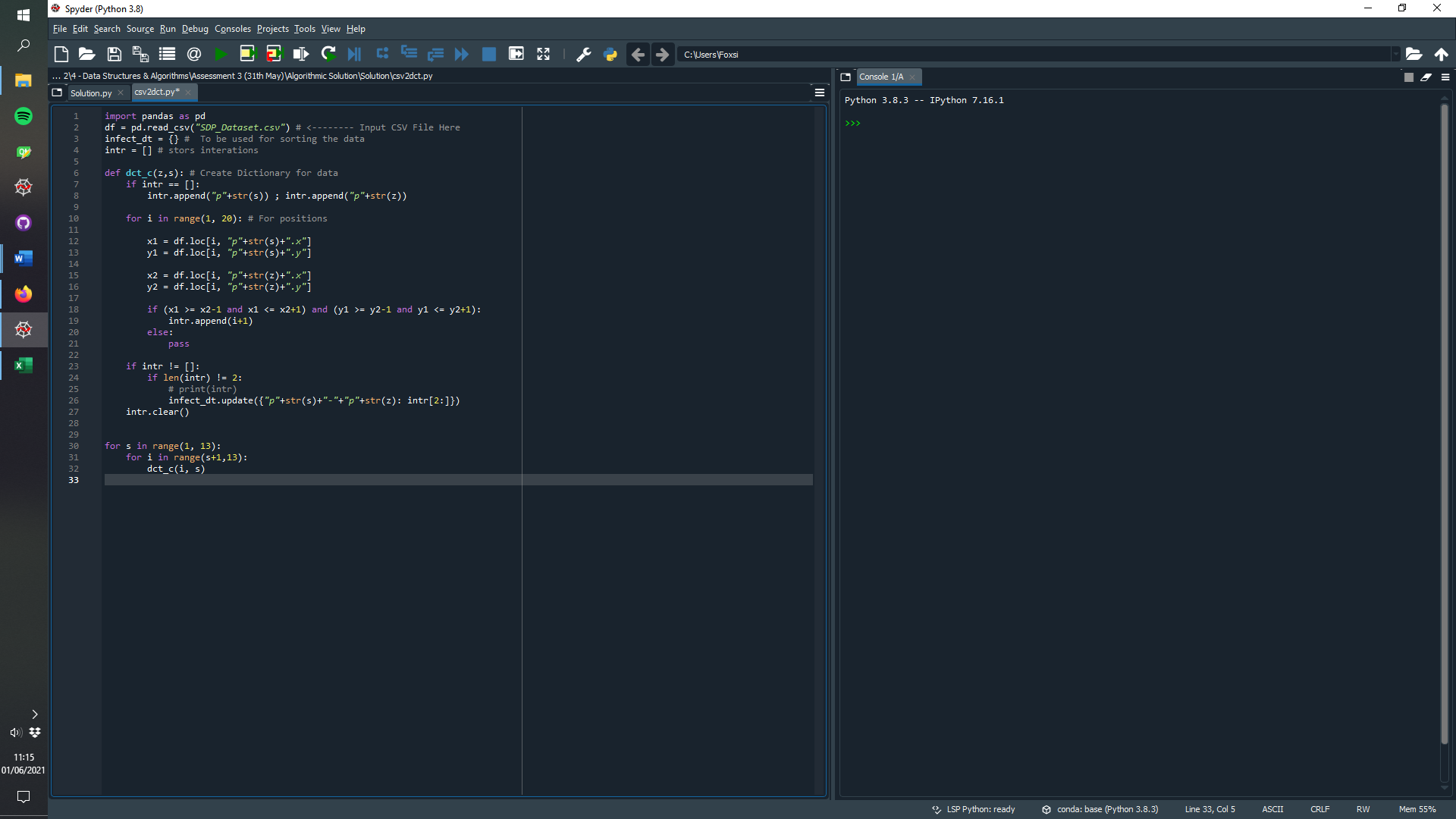
[2] Government announces further measures on social distancing, Press release, (20 March 2020): <https://www.gov.uk/government/news/government-announces-further-measures-on-social-distancing#history>

[3] Stevens, Harry (2020-03-14). "These simulations show how to flatten the coronavirus growth curve". Washington Post. Archived from the original on 2020-03-30. Retrieved 2020-03-29.: <https://www.washingtonpost.com/gdpr-consent/?next_url=https%3a%2f%2fwww.washingtonpost.com%2fgraphics%2f2020%2fworld%2fcorona-simulator%2f>

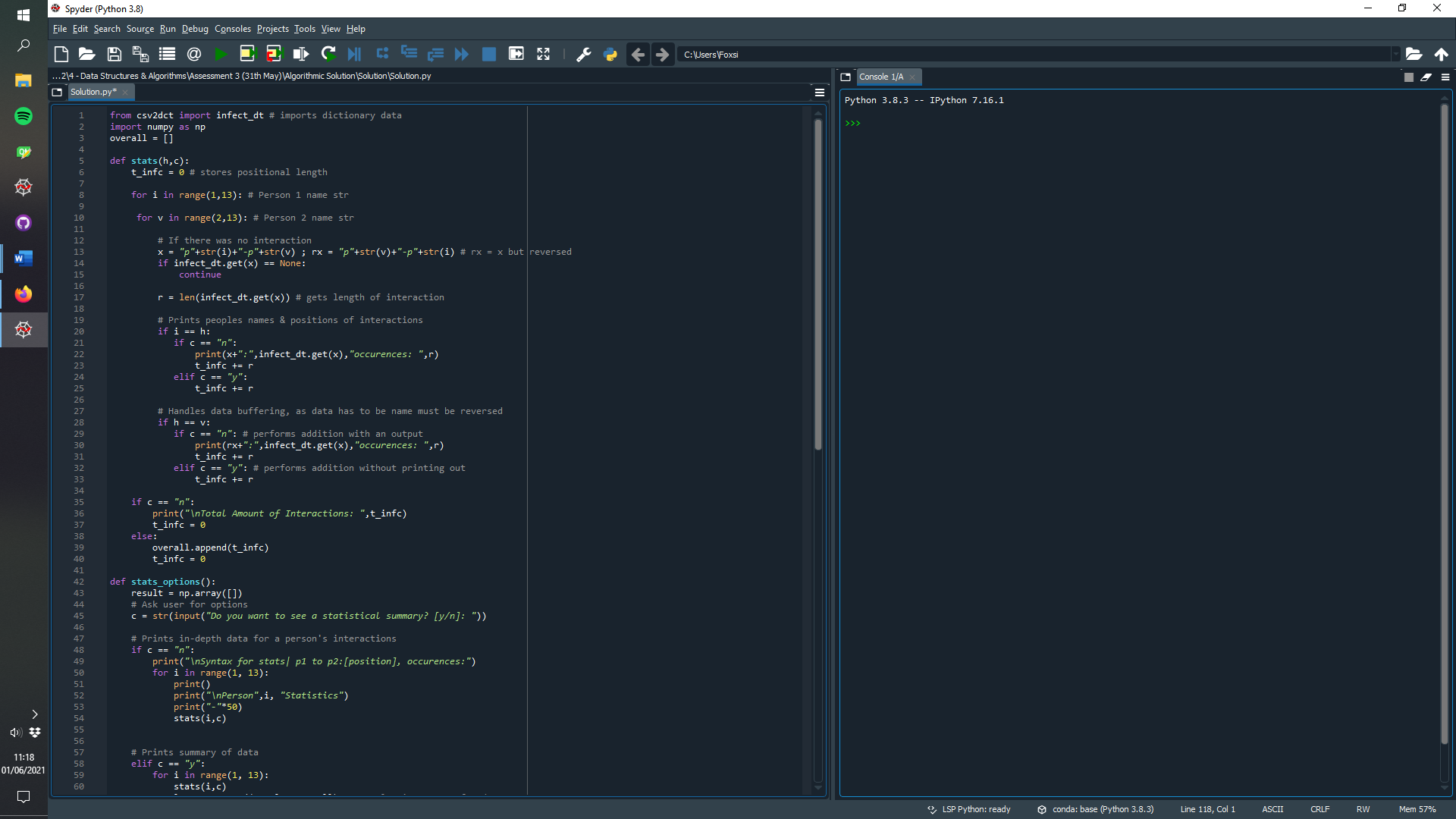
# Appendix A

Python code used in print data to csv file:

# Appendix B

Python code to convert csv to dict

# Appendix C

used to print statical data (bulky more presents statistics in a comprehensive manner)

