# COMP1002 | Data structures and Algorithms

Final Assignment: Optimising Urban Parcel Logistics Using Data Structures and Algorithms



#### Description

CityDrop Logistics is expanding its operations in a metropolitan area and required me to develop a suite of inter-connected software modules to optimise there delivery network.

I ended up developing a system that is divided into four modules that are interconnected:

- 1. Graph Based Route Planning:- This allows CityDrop Logistics to determine the most effective and efficient paths from a hub to a destination
- 2. Hash Based Customer Lookup :- This allows CityDrop Logistics to retrieve customer information quickly
- 3. Heap Based Parcel Scheduling :- This allows CityDrop Logistics to prioritise deliveries based on urgency
- Sorting of Delivery Records: This allows CityDrop Logistics to organise records at the end of the day usign Merge and Quick sorts

#### Getting Started

#### Dependencies

- Python 3.13
  - o 

    ✓ Numpy

#### Recommended

- Microsoft VSCode
  - o 🗹 Pylance
  - o 🗹 Python
  - o 🗹 Python Debugger
  - ✓ Rainbow CSV

Name: Muhammad Annas Atif

Student ID: 22224125

Email: 22224125@student.curtin.edu.au

# Curtin University: School of Electrical Engineering, Computing and Mathematical Sciences Curtin

# Contents

Overvi	ew1						
1.	Graph Based Route Planning						
2	. Hash Based Customer Look Up						
3.	Heap Based Parcel Scheduling						
4.	Sorting Delivery Records						
Data S	Structures Used						
1.	Why Graphs?3						
2.	Why Hash Tables?						
3.	. Why Heaps?4						
4.	Why use Sorting Algorithms?						
Modul	Module Integration						
Algori	thm Complexity Analysis						
Sample	e Output7						
	Fig 1 – Loading of Test Graph with Output						
	Fig 2 – Loading CSV File with Test Customers into Hash Table with Output9						
	Fig 3 – Raw Heap Output (Not a Sorted List Based on Priorities)10						
	Fig 4 – Removing Items from Heap Until Empty Message Shows Up Followed by Exporting Removed Heap Items as CSV then Reimporting Them for Sorting11						
	Fig 5 – Returning to Main Menu and Performing Merge Sort with Time Output12						
	Fig 6 – Performing Quick Sort with Time Output						
	Fig 7 – 100 Element Merge Sort Random Data with Time14						
	Fig 8 – 100 Element Quick Sort Random Data with Time15						
	Fig 9 – 100 Element Merge Sort Nearly Sorted Data with Time						
	Fig 10 – 100 Element Quick Sort Nearly Sorted Data with Time17						
	Fig 11 – 100 Element Merge Sort Reversed Data with Time						
	Fig 12 – 100 Element Quick Sort Reversed Data with Time						
Reflect	tion						
Limita	tions and Assumptions						

# Overview

CityDrop Logistics is expanding its operations in a metropolitan area and required me to develop for them, a suite of inter-connected software modules to optimise their delivery network.

In the end, I ended up developing four inter-connected modules of software that communicate between one another that allows CityDrop logistics to effectively and efficiently add, schedule and sort customer packages and parcels.

Each system that I implemented utilised the various self implemented data structures and algorithms such as graphs, hashes, heaps, linked lists, stack and queues, quick sort and merge sort.

The systems that I developed include:

## 1. Graph Based Route Planning

This module allows CityDrop logistics to manipulate their delivery network by allowing them to add and remove delivery hubs/intersections as well as the routes between those hubs and intersections. It also allows them to display their network and perform automated tasks such as determining the shortest path from a hub/intersection to a destination (Dijkstra's Shortest Path), performing Breadth-First searches and Depth-First searches to determine accessible delivery zones from a hub/intersection as well as any cycles in their network that would be inefficient.

This software module utilises the *graph* data structure. Additionally, this module easily and simply integrates into the subsequent modules.

#### 2. Hash Based Customer Look Up

This module allows CityDrop logistics to manage customers in their company. This module allows them to add/remove customers as well as check if a customer exists. Furthermore, it allows CityDrop logistics to load/save customer data from/to a CSV file for easy data management. The system is able to detect whether a customer being added to the system has a valid address from a hub/intersection which is loaded from module one and similarly the system is able to handle collisions in the hash table which can be verified by the use of the collision example. When a customer is added the system is able to update delivery status and the hash table can be displayed. The module takes graph data from module one to determine estimated delivery time (Dijkstra's Algorithm) and the aforementioned valid destinations and hubs.

This module utilises the *hash table* data structure and this module communicates with subsequent modules.

#### 3. Heap Based Parcel Scheduling

This module allows CityDrop logistics to schedule parcels from customer data from module two which can easily be imported directly or from a CSV file from the prior module. This module also calculates the priority of the delivery based on the following formula

$$P = (6 - p) + \frac{1000}{T}$$

Where P is the calculated priority, p is the customer priority from 1 to 5, and T is the estimated delivery time. This information is retrieved from the previous module. The heap itself extracts the scheduled deliveries based on highest calculated priority and this data can then be loaded into subsequent or prior modules or exported as a CSV file. When customer data is loaded into this module, the system updates the delivery status to "In-Transit" and when a delivery is completed, it can be removed from the scheduler (heap) where its status is set to "Delivered", the completed deliveries can then be exported to a CSV file.

This module utilises *heaps* as its data structure and this module communicates with subsequent modules.

## 4. Sorting Delivery Records

This module allows CityDrop Logistics to sort the deliveries using *quick* and *merge* sorts at the end of the day for reporting purposes. Completed deliveries are re-loaded back into the previous module then, they are sorted in this module where they can perform one of the following sorts:

- i. Merge Sort
- ii. Quick Sort
- iii. Quick Sort (Pivot is a Median of Three Values)
- iv. Quick Sort (Pivot is a Random Value)

When sorted, the data is presented in a sorted order based on shortest delivery time to longest delivery time as well as how long it took to sort the data. Finally, CityDrop Logistics has the ability to do sorts on super large sets of sorted data which is done by loading a CSV file which has nine different test cases for 100, 500 and 1000 delivery records.

# Data Structures Used

## 1. Why Graphs?

Graphs were the most obvious choice for module one. This is because the nodes of a graph can be used to represent each hub/intersection, and the routes can easily be represented by edges.

Additionally, the edges of the graph can be weighted, which means that the travel time can be assigned between each hub/intersection/destination. The weighting is particularly useful for calculating the shortest path from one point to another using Dijkstra's Shortest Path Algorithm.

An important aspect of this first module was that the shortest path from one hub to another had to be determined. Dijkstra's shortest path algorithm was the best choice for this, and this method works with graphs which meant that graphs was a good choice for module one.

Further requirements meant that any cycles in the network and disconnected hubs had to be determined, and this can be done using Breadth-First and Depth First searches which work with the graph data structure.

Hence, for module one, the graph data structure was determined to be the most suitable data structure and hence which is why it has been used in the project.

### 2. Why Hash Tables?

Hash tables were the most obvious choice for module two. This is because hash tables offer fast lookups and efficient data management. Hash tables are efficient in managing data because as the number of customers grows, the hash table has the ability to resize depending on the load factor (increasing when >0.7 and decreasing when <0.3).

The implementation of a "good" hashing function allows the hash table to minimise collisions which allows for a good spread/distribution of entries. Linear probing in the hash tables ensures that in the event of a collision, the hash table will deal with the collision by moving the entry to the next available entry that has not been used.

Hash tables were also a good choice based on the implementation of the hash entry object itself which is able to store all of the necessary information which allows all of the required information to be stored without having to strip the information and go about substituting it back after hashing is complete.

Hence, for module two, the hash table was determined to be the most suitable data structure and hence which is why it has been used in the project.

# 3. Why Heaps?

Heaps were the most obvious choice for module three. This is because it offers efficient management of parcels based on calculated priority. Furthermore, as the module is frequently requiring access to the highest priority item, it is able to efficiently access that information.

The use of a *max* heap in the project ensures that the value of each parent node is greater than or equal to the values of its children, this ensures that the root element is always the highest priority element.

In essence, the *max* heap which has been implemented in this project is a pseudo priority queue.

Hence, for module three, the heap was determined to be the most suitable data structure and hence why it has been used in this project.

## 4. Why use Sorting Algorithms?

The use of sorting algorithms was essential as it was a requirement to sort the delivery records at the end of the day based on delivery times. Due to nature of the application; a situation where there could be potentially hundreds if not thousands of deliveries it only made sense to choose efficient and quick sorting algorithms. In this case, that was the use of *Quick* and *Merge* sorts.

The *quick* sort was chosen as one of the sorting algorithms for a few reasons. Firstly, it is an *in-place* sort. This means that data is efficiently sorted within the original array and does not need additional overheads or other temporary arrays making it ideal in low memory situations, this also means that more data can be added to an array without having to factor in overheads.

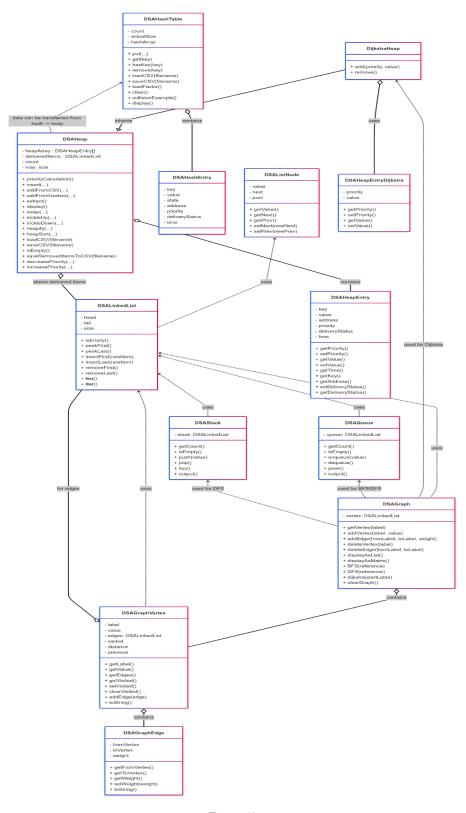
However, this sorting algorithm is not always the best choice, particularly when presented with completely reversed data where it slows down a lot. In this case, it is better to use *Merge* sort.

Merge sort has also been implemented for several reasons. First of all, merge sort is a very stable sort where in all cases for all types or ordered data, (reversed, random etc.) it performs identically. This particular case makes it far better than the quick sort.

However, the downside to *Merge* sort is that it is not an in-place sort which means that it requires an additional array to keep temporary storage of the sorting array which is a trade off for stability and the guaranteed sorting speed when compared to *quick* sort.

Hence, for module two, *Quick* and *Merge* sorts have been implemented into the project due to there good performance with both having their own pros and cons allowing CityDrop Logistics to determine what should be utilised in each situation they face.

# Module Integration



Page 5

# Algorithm Complexity Analysis

Module	Algorithm/Operation	Time Complexity	Space Complexity	Reason
1 Graphs	Add Vertex	O(V)	O(1)	Traverses for duplicates and adds O(1)
	Get Vertex	O(V)	O(1)	Traverses through linked list
	Add Edge	O(E)	O(1)	Checks for duplicate edges then
	naa Eage	(L)		adds edge O(1)
	Delete Edge	O(E)	O(1)	Removes edge from both
				vertices (2 x O(E))
	Display as List	O(V + E)	O(1)	Iterates all vertices and edges
	Display as Matrix	$O(V + E)$ $O(V^2)$	$O(V^2)$	Fills and prints matrix it does a
			, ,	double loop over vertices
	BFS w/ Depth	O(V + E)	O(V)	Standard BFS over all vertices
	Detection			and edges. Space complexity for
				the depth list for depth of
				vertices.
	DFS w/ Cycle	O(V + E)	O(V)	Standard DFS over all vertices
	Detection			and edges. Space complexity for
				the visited list for the cycles.
	Dijkstra	O((V + E)  Log  V)	O(V)	Each vertex added to heap once
				and heap operations are O(log
				V).
	Clear Graph	O(V)	O(1)	Traverses list of vertices and
				resets values .
	Sort Adjacent List	$O(E^2)$	O(E)	Double loop over the list of
				edges.
2 Hashes	Put	O(1)	O(N)	Linear probing is $O(1)$ and hash
	Get			table has O(N) space where N is
	Has Key			number of entries.
	Remove			
	Hash	O(N)	O(1)	Hashes over the number of
				characters in the string
	Find Slot	O(1)	O(1)	Linear Probing
	Resize	O(N)	O(N)	Rehashes N number of entries
				and allocates new array of N
				size
	Next Prime	O(N  sqrt(N))	O(1)	Primes are checked up to the
				square root of the new size N
	Save CSV	O(N)	O(1)	Iterates over all entries to write
	Load CSV	O(N)	O(N)	Iterates over N rows of the CSV
				file and does one "Put" per row
	Load Factor	O(1)	O(1)	Divides two numbers
	Clear	O(N)	O(1)	Empties the array over N
				number of elements and creates
				a new array with self.actualsize
	D: 1	0(31)	0(1)	= 0, hence 2 elements
0.11	Display	O(N)	O(1)	Prints each slot
3 Heaps	Priority Calculation	O(1)	O(1)	Calculates the priority using an the equation
	Insert	O(Log N)	O(1)	Insert at end and Trickle Up
	Add From CSV			
	Add from Hashes			

3 Heaps Cont.	Extract	O(Log N)	O(1)	Remove the root and trickle down
	Display	O(N)	O(1)	Iterate through heap elements
	Swap	O(1)	O(1)	Swap two elements
	Trickle Up	O(Log N)	O(1)	Standard heap operation
	Trickle Down			
	Heapify	O(N)	O(1)	Iterates from last non-leaf node
				and calls Trickle Down for each
				node from the last non-leaf node
				to root node.
	Heap Sort	O(N Log N)	O(1)	Standard heap sort.
	Decrease Priority	O(N Log N)	O(1)	Update Trickle Down
	Increase Priority			Update Trickle Up
	Load CSV	O(N Log N)	O(N)	For N number of entries there is
				and then N inserts. Space
				complexity of N number of
				entries in heap.
	Save CSV	O(N)	O(1)	Write all heap entries
	Is Empty	O(1)	O(1)	Checks heap count
	Save Removed to CSV	O(N)	O(1)	For N number od delivered
				items write to a CSV.
4 Sorts	Merge Sort	O(N Log N)	O(N)	Recursively split and merge,
		Best/Average /		space complexity comes from
		Worst		temp arrays.
	Quick Sort	O(N Log N)	O(Log N)	Standard quick sort that selects
		Best/Average and		a pivot and partition list into
		O(N <sup>2</sup> ) Worst		two halves, one with elements
				less than pivot and one with
				elements greater than pivot and
				recursively sorts each half.
				Space complexity is due to
				number of recursive calls.

# Sample Output

Please note that due to the sheer size of the test cases for the sorting algorithms, I have provided examples for all of the 100 element sorts. 500 and 1000 element sort test can be run using the program with instructions given in the README.md or README.pdf.

```
tions: A, B, C, D, E, F, G, M
A-D, A-C, B-E, E-F, E-G, F-G, D-F, D-C
utce): A-B: 18, A-D: 1, A-C: 2, B-E: 1, E-F: 5, E-G: 12, F-G: 2, D-F: 8, D-C: 7
   oth List (Hops Between Each Hub Intersection):
4 | 0 , B | 1 , C | 1 , D | 1 , E | 2 , F | 2 , G | 3
AB, BE, EF, FD, DC, FG
 -> B -> E -> F -> D -> A
-> B -> E -> F -> D -> C -> A
```

Fig 1 – Loading of Test Graph with Output

Fig 2 – Loading CSV File with Test Customers into Hash Table with Output

Fig 3 – Raw Heap Output (Not a Sorted List Based on Priorities)

Fig 4 – Removing Items from Heap Until Empty Message Shows Up Followed by Exporting Removed Heap Items as CSV then Reimporting Them for Sorting

Fig 5 – Returning to Main Menu and Performing Merge Sort with Time Output

Fig 6 – Performing Quick Sort with Time Output

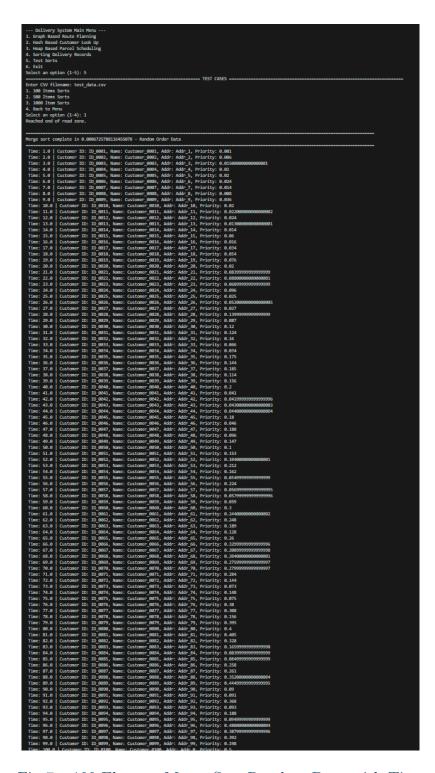


Fig 7-100 Element Merge Sort Random Data with Time

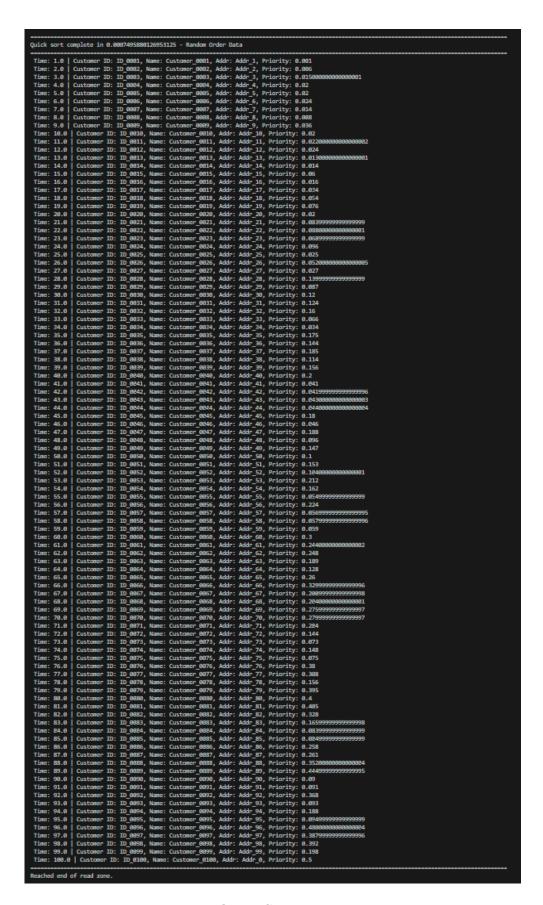


Fig 8 - 100 Element Quick Sort Random Data with Time

Fig 9 – 100 Element Merge Sort Nearly Sorted Data with Time

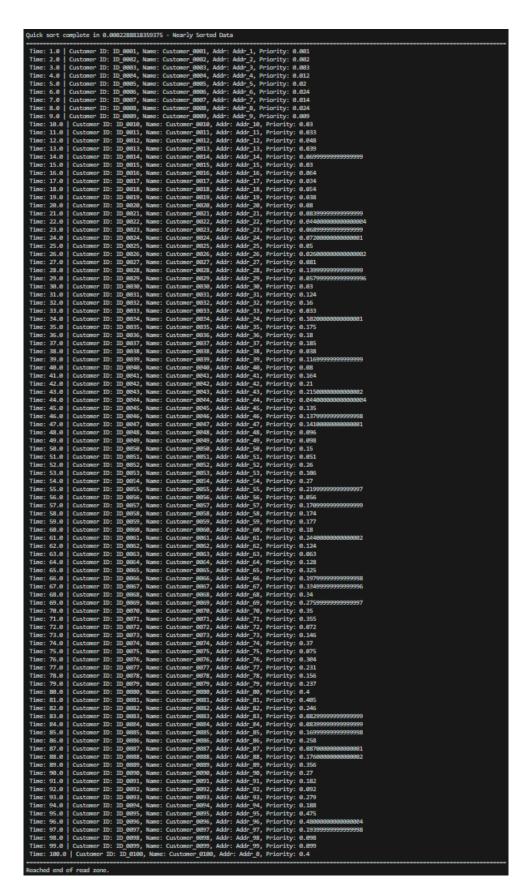


Fig 10 – 100 Element Quick Sort Nearly Sorted Data with Time

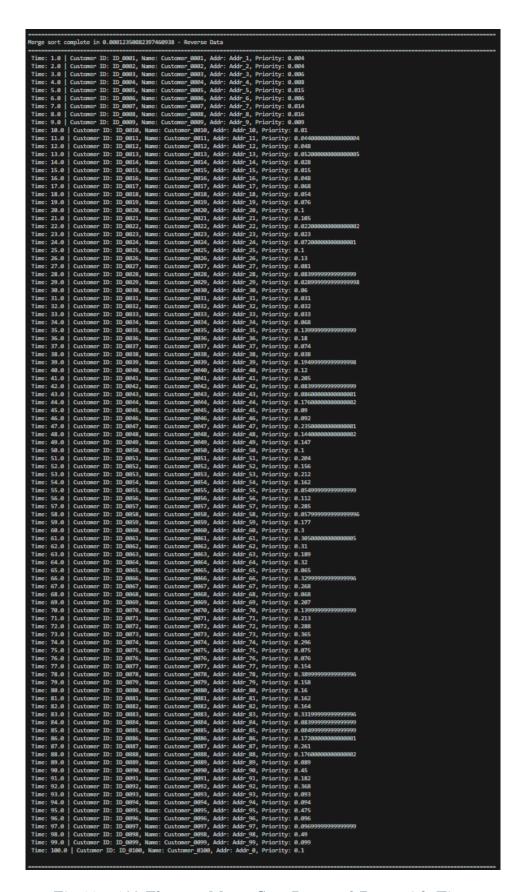


Fig 11 – 100 Element Merge Sort Reversed Data with Time

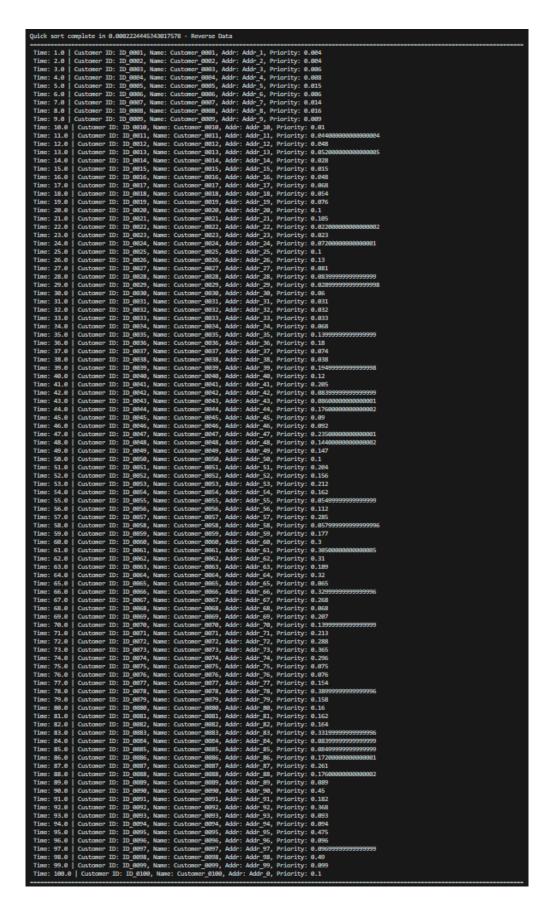


Fig 12-100 Element Quick Sort Reversed Data with Time

# Reflection

Over the course of the past few weeks, completing this assignment has personally given me a lot of insight into programming a complex interconnected system of software modules that effectively and efficiently process data using self-made data structures and algorithms. I am quite pleased with the outcome of my assignment, it gave me a much deeper understandings of the inner workings of the data structures and algorithms, and it really allows me to appreciate the fact that we can simply call all of this in a programming language using a single simple function instead of having to do all of this from scratch.

I am particularly pleased with how feature-rich I have made my program allowing for multiple ways of moving data to and from different modules, in particular I am impressed with how I have been able to use a heap object to keep essential delivery information to sort it in the last module to ensure that I do not have to go and match data to the sorted times. I would have never thought that would ever work before this assignment. I have also managed to learn a lot about how the data structures are loaded into memory and how I can access their relative information. This was particularly the case when I was trying to sort delivery data as I would constantly get errors until I printed the outputs to realise that I was trying to sort heap objects and not the actual times. This also widened my understanding of how computers effectively manage these data structures in memory.

Aside from all of the positives, there are also some areas of improvement. Particularly, with the efficiency of the data structures and algorithms I have implemented. Given that these have been developed from scratch there is obviously room to make them more efficient in terms of space efficiency in memory and how much processing power they need to perform their desired tasks.

To conclude, I have learnt a lot about the data structures and algorithms we have learnt over the course of the semester from hashes and heaps to graphs and the stacks, queues and linked lists that underpin these data structures and algorithms. I have learnt how to effectively implement them all together to make a complex, yet easy to operate software package which is able to effectively demonstrate my learnt knowledge from COMP1002. I have also widened my understanding of potential areas of improvement to my program if I was to attempt it again.

# Limitations and Assumptions

Some limitations include:

1. When loading processed data from the heaps back into the hash table, sometimes the system throws an error, but the data still loads into the hash table.

2. With the test cases module, if you do not load CSV file and proceed to run a sort it still outputs a time. I assume that is the time it takes for it to realise the CSV file is not loaded or there is no data to sort.