Justin Golay
ID #010176782

WGU Email: jgolay@wgu.edu February 5, 2023

C950 Data Structures and Algorithms II

#### Introduction

The purpose of this project is to create an algorithm using Python programming language to develop an effective and efficient solution to deliver packages. There are three delivery trucks, two drivers, and 40 packages. Each delivery truck can only hold 16 packages at one time. Limitations to the delivery of these packages come in the form of delivery deadlines, special instructions for various packages, and packages that arrive late to the mailing hub and must then depart later in the day than other packages.

This program utilizes the Nearest Neighbor Algorithm to simulate package delivery. The end user can then enter a specific time and determine the delivery status of an individual package or all packages at that time.

## A. Algorithm Identification

This program utilizes the Nearest Neighbor Algorithm to order package deliveries.

## **B1.** Logic Comments

The Nearest Neighbor Algorithm is created in the following manner:

- 1. Create an empty list and assign all package objects to it.
- 2. Utilize a while loop to continue until this list is empty
- 3. Initiate a variable referring to the package's next address as a number greater than any of the distances to be compared in the delivery simulation.
- 4. Initiate a variable referring to the next package to be compared and set it to none.
- 5. Utilize a for loop to compare each package's destination with the current location and update the variables to reflect the shortest distance.
- 6. Place the package with the destination that is the closest to its closest location onto the truck.
- 7. Remove the package from the initial package list.
- 8. Repeat the process until the initial list of packages contains no packages.

## Pseudocode:

function nearestNeighborSearchAlgorithm(truck):

unsortedPackages = empty list

FOR packages on truck:

APPEND packages to unsortedPackages

## CLEAR packages from truck

```
WHILE length of unsorted packages > 0:
      nextAddress = infinity
      nextPackage = None
      FOR packages in unsortedPackages:
             truckAddress = getAddress (truck address)
             eachPackageAddress = getAddress(eachPackage address)
             IF distanceBetweenCities(truckAddress, eachPackageAddress) <=
             nextAddress:
                    nextAddress = distanceBetweenCities(truckAddress,
                    eachPackageAddress)
                    nextPackage = eachPackage
      APPEND nextPackage to truck
      REMOVE nextPackage from unsortedPackages
      UPDATE truck mileage += nextAddress
      UPDATE truck address += nextPackage address
      UPDATE truck time += nextAddress/18
      UPDATE nextPackage deliveryTime = truck time
      UPDATE nextPackage departTime = truck departTime
```

Methods used in this algorithm include getAddress() to find the integer index of a location based on its address as a string and distancesBetweenCities() takes two address indexes and locates them in a list of distances.

```
function getAddress(string address):
```

FOR each row in a list of addresses:

IF address is in row

RETURN integer index number

```
function distanceBetweenCities(xValue, yValue):

distanceBetweenCities = distance list indexed at xValue, yValue

IF distanceBetweenCities == '':
```

# $\label{eq:distance} distance Between Cities = distance \ list \ indexed \ at \ yValue, \ xValue$ RETURN float distance Between Cities

## B2. Development Environment

The program was developed utilizing PyCharm Community Edition and Python version 3.11. The operating system was Windows 11 Home version 22H2. The system uses an AMD Ryzen 5 processor and 64-bit operating system.

## B3. Space-Time and Big-O

The Nearest Neighbor Algorithm of this program has a time complexity of  $O(N^2)$ . The overall program has a time complexity of  $O(N^2)$ . Further, the program creates lists utilized by the Nearest Neighbor Algorithm that have a space complexity of O(N). The overall space complexity of the program is O(N).

## B4. Scalability and Adaptability

This program utilizes a hash table to organize packages by ID number. This results in a package look up time complexity of O(1). However, this hash table was developed with 40 buckets, which is efficient for this program but may be less efficient should the program take a much larger number of packages. The number of buckets in the hash table can be increased to improve efficiency with larger inputs.

Also, given the time complexity of this Nearest Neighbor Algorithm is O(N^2), a larger number of packages to deliver would take longer (quadratic growth) for the program compute than if it were given a smaller number of packages.

If the delivery service were to expand operations, additional servers may need to be purchased to store package information and address lists, which would grow linearly. However, delivery trucks are limited in the number of packages that they can haul due to the size of the truck. This limitation of the delivery truck means that it is unlikely the company would need to purchase faster servers to calculate delivery routes.

## B5. Software Efficiency and Maintainability

This software is efficient and easy to maintain because it is based on a commonly known algorithm called the Nearest Neighbor Algorithm with a time complexity of O(N^2) ("Nearest Neighbour Algorithm"). Developers can reference pseudocode to ensure that the algorithm maintains function even if the company utilizing the software is required to switch to another programming language.

The Nearest Neighbor Algorithm is more time efficient than other approaches to similar problems such as the Brute Force approach, which would utilize a time complexity of O(N!) (Chase et al.).

## B6. Self-Adjusting Data Structures

The self-adjusting data structure in this program includes the hash table. The primary benefit of utilizing a hash table for this type of problem is that it searches and updates an item in constant time. However, the primary weakness of this hash table is that its space complexity is O(N), which means that it will grow linearly as its input grows. This could result in significantly more memory being utilized should the program simulate a greater number of packages being delivered.

## C. Original Code

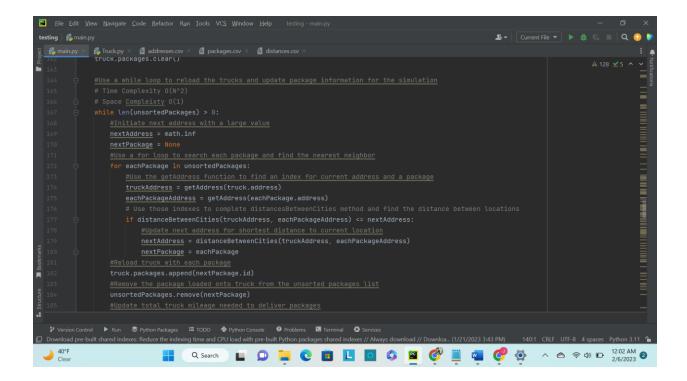
The code for this program was completed and includes the following files to execute correctly: main.py, Truck.py, addresses.csv, packages.csv, distances.csv.

## C1. Identification Information

Identification information was included on Line 1 of main.py including: Justin Golay ID #010176782.

## C2. Process and Flow Comments

Comments were added throughout the program to ensure that other software developers could read and understand the process and flow of this program. Below is an example:



#### D. Data Structure

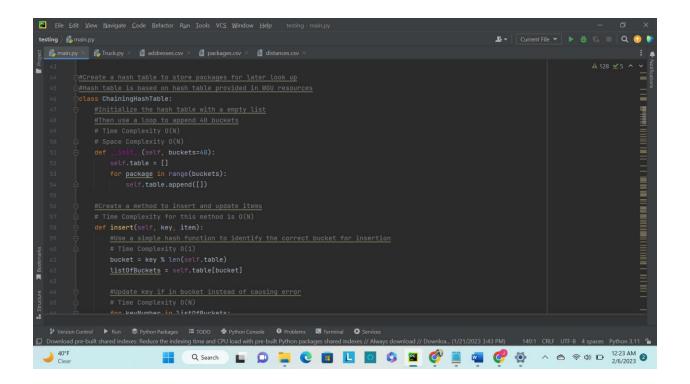
The self-adjusting data structure in this project is the hash table utilized to store package objects that store package information. They can be searched in the hash table utilizing each package's ID number as the key. Should a function call for insertion of two items with the same ID number, a collision is avoided by updating the item with the second insertion.

## D1. Explanation of Data Structure

Each package delivered in this program was read from a .csv file and placed into a package object. Each package object was stored in the hash table based on its unique package ID number. The program frequently searches for packages to execute functions required to complete the Nearest Neighbor Algorithm. This hash table allows for packages to be searched in constant time.

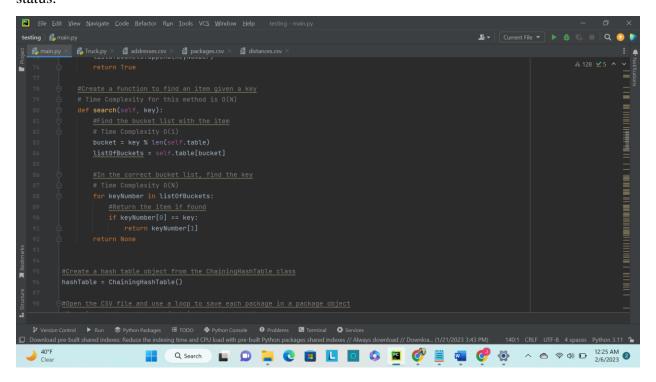
#### E. Hash Table

The hash table was utilized in this program. It was designed based on WGU resources (C950 - Webinar-1 - Let's Go Hashing).



## F. Look-Up Function

The look up function was created in the hash table. It is called "search". Each package can be searched in the hash table. Once that package has been found, specified information for each package can be found (i.e., ID number, address, delivery deadline, city, zip code, weight, and status.



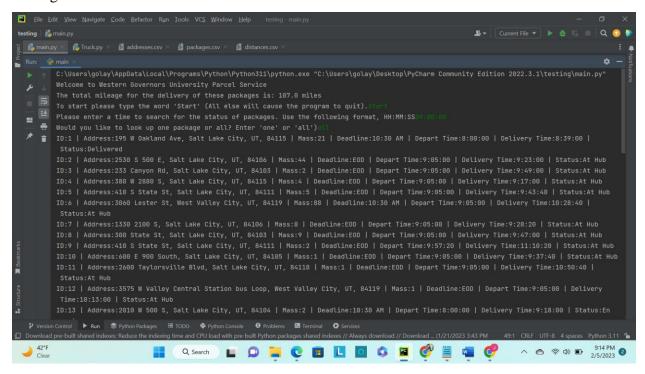
Furthermore, end users that are not familiar with coding can utilize the user interface and request to view the information for all packages at any time.

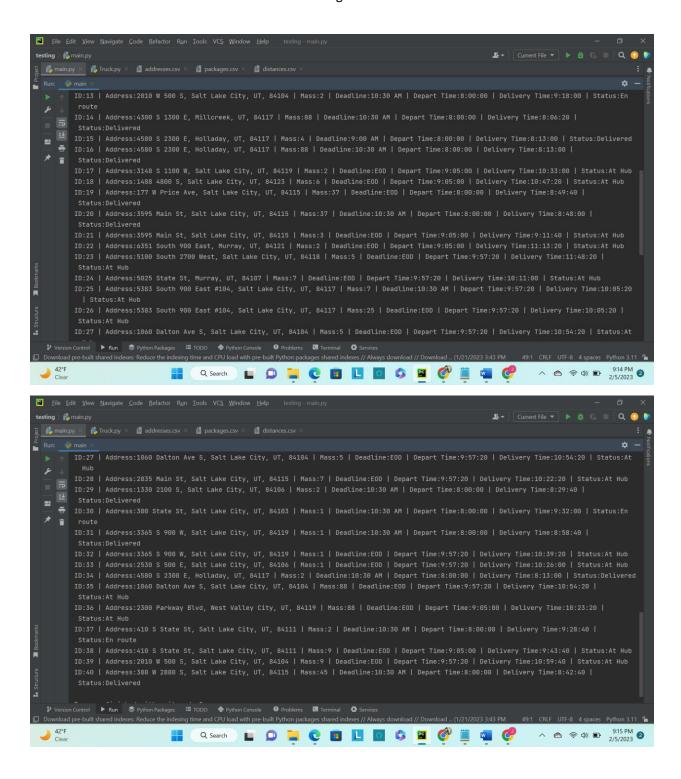
#### G. Interface

A user interface was developed for this program to allow users to look up any or all packages at any time. The total mileage for all trucks is displayed for users as well. Below are screen shots to demonstrate this functionality.

#### G1. First Status Check

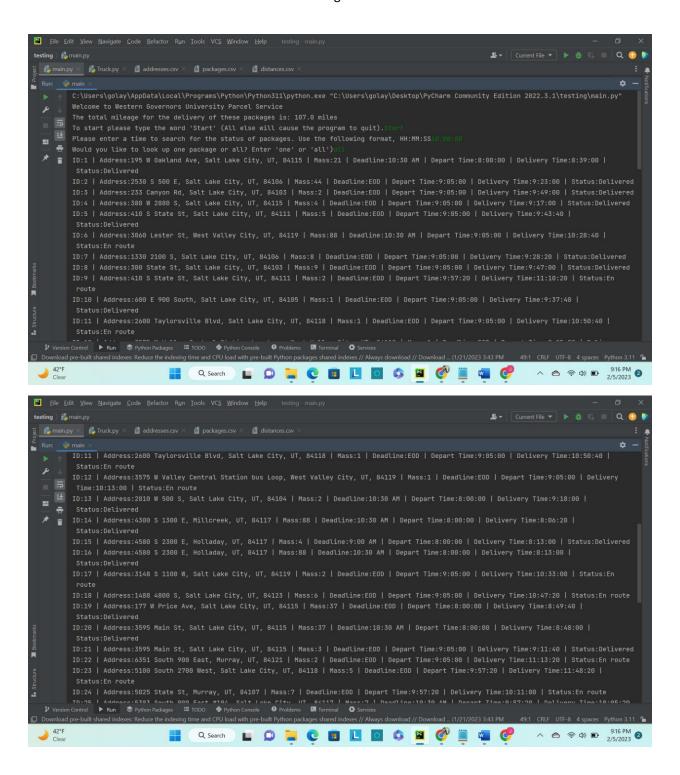
Package Information at 9:00 a.m.

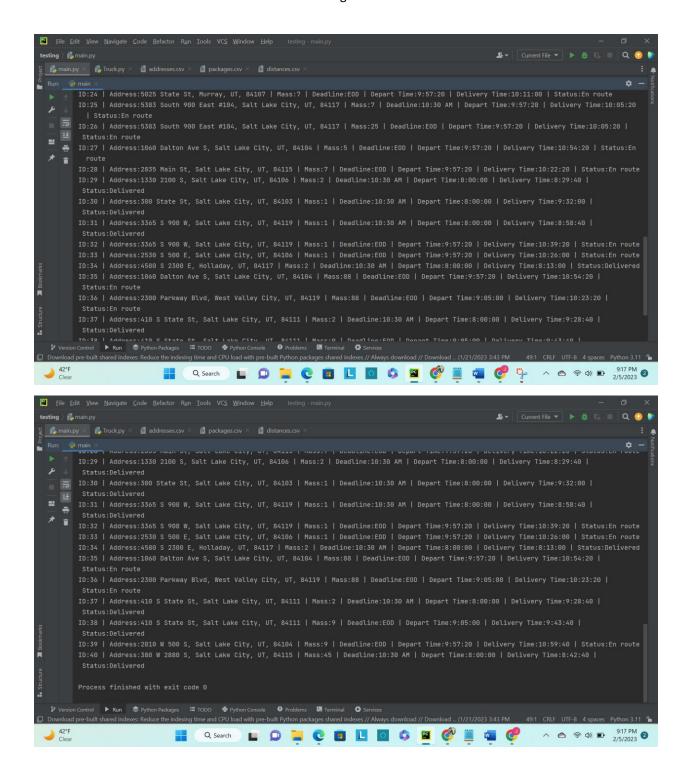




## G2. Second Status Check

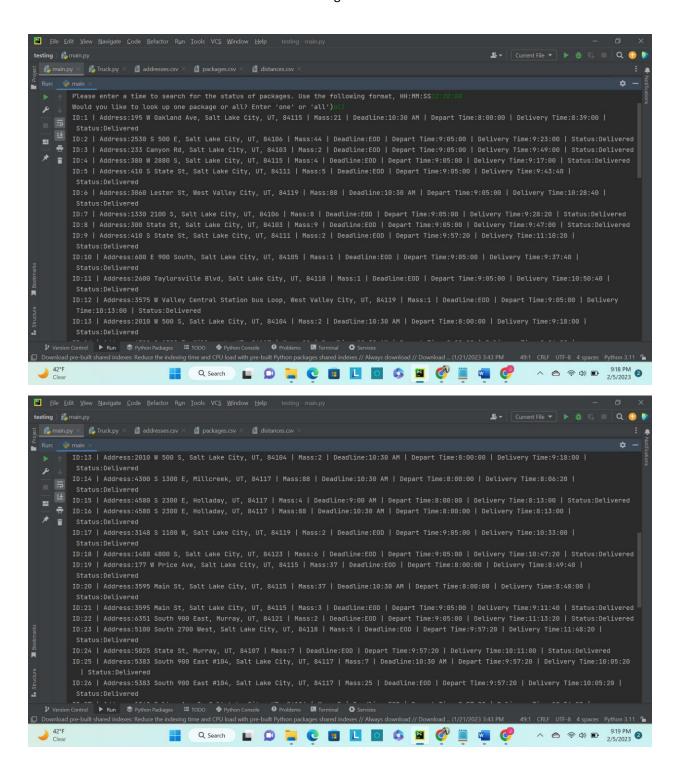
Package Information at 10:00 a.m.

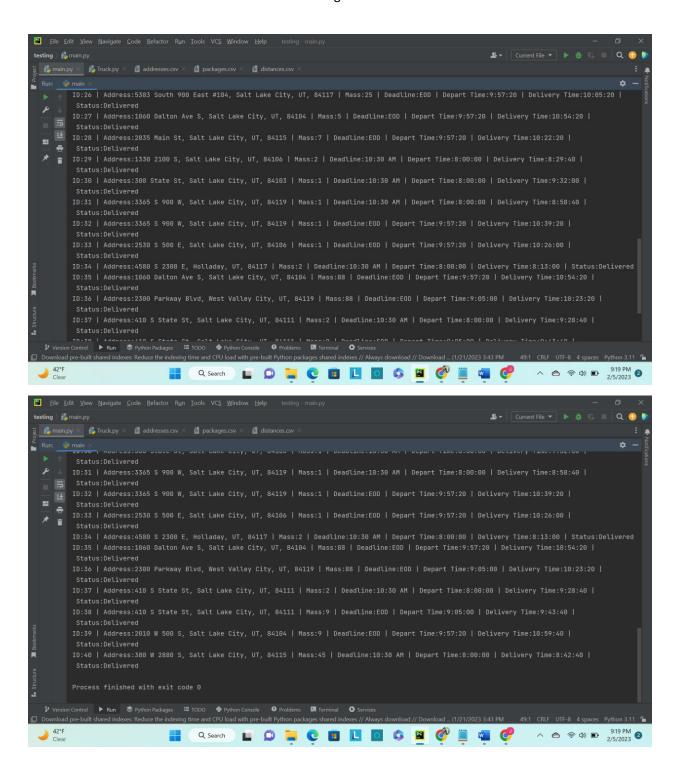




#### G3. Third Status Check

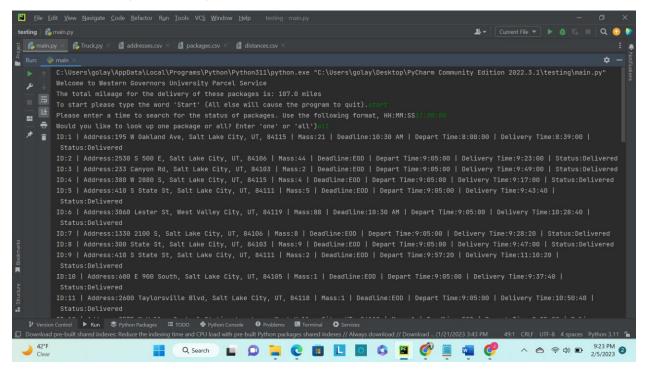
Package Information at 12:30 p.m.





#### H. Screenshots of Code Execution

The code is executed free from runtime errors and warnings as evidenced by the screen shots provided above. The following screen shot shows the total mileage of all the trucks at the top of the user interface (107.0 miles):



## I1. Strengths of Chosen Algorithm

The Nearest Neighbor Algorithm has multiple strengths and was chosen primarily because of its time complexity, scalability, and ability to adapt to changes. The time complexity of the Nearest Neighbor Algorithm is  $O(N^2)$ , which is significantly faster than the Brute Force approach to this problem, which requires a time complexity of O(N!). Furthermore, this algorithm can be scaled to larger inputs if needed for package delivery. Finally, this algorithm will execute as expected regardless of packages provided. When a delivery company must compute new routes each day, the algorithm can compute these routes so long as the addresses are correctly stored in the referenced lists.

## I2. Verification of Algorithm

The program meets all requirements for this assignment. Trucks carry no more than 16 packages, trucks travel an average of 18 miles per hour, there are no collisions, drivers do not leave the hub earlier than 8:00 a.m., special instructions are followed for packages, and all packages are delivered. All data for packages can be looked up by package ID and time in the user interface, as required.

## I3. Other possible Algorithms

A Brute Force Algorithm and Dijkstra's Algorithm ("Dijkstra's Algorithm") are both other possible algorithms that could be used to complete this assignment.

## I3A. Algorithm Differences

The Brute Force Algorithm differs from the Nearest Neighbor Algorithm in both time complexity and outcome. Brute Force has a time complexity of O(N!) but does return the optimal route for package delivery. Alternatively, Nearest Neighbor has a time complexity of O(N^2) and returns an efficient route for delivery that may not be the most efficient possible in terms of distance travelled. The Brute Force approach would require calculation of every possible route to deliver the packages prior to choosing the most efficient, whereas Nearest Neighbor only calculates the nearest delivery address based on its current location and remaining packages.

Dijkstra's Algorithm utilizes a graph to find the shortest paths of delivering packages based on their distance from the delivery hub. The difference between this and Nearest Neighbor is that Dijkstra's will find the shortest path based on the distance from a delivery truck's starting point and Nearest Neighbor will update the shortest path based on its current location.

## J. Different Approach

If I were to complete this project again following the same constraints, I would 1) adjust hashing function to be more complex and scalable, and 2) develop components of the project in more files for the purpose of simplification.

The hashing function currently utilizes modulo to hash package ID, however, other hashing approaches such as the Mid Square Method can be used to hash package IDs and maintain a fast look up rate with a larger number of packages because all of the digits in the ID assist in hash generation.

The program currently only has two programming files: main.py and Truck.py. Taking information from the package class and placing it into its own file would allow for more readable code.

## K1. Verification of Data Structure

The hash table does meet the requirements for this project. It does not utilize a built-in python dictionary.

## K1A. Efficiency

The time complexity of a look-up function slows if there are more than 40 packages or packages whose ID number hashes them into the same bucket. While an individual package in each hash bucket yields a time complexity of O(1), multiple items stored in the same hash bucket will result in a time complexity of O(N) as the look-up function must search through all IDs in that bucket.

## K1B. Overhead

The hash table in this program is self-adjusting and grows with the amount of input it receives. The current program takes 40 packages and contains a list with 27 addresses. The hash table has a space complexity of O(N) and therefore will linearly require additional memory as additional packages and addresses are added. This could require additional servers.

## K1C. Implications

Given this algorithm's design, the number of trucks would not impact the look up time or space usage for this hash table. An increased number of cities would not increase the look up time as these are stored in a list and can be quickly referenced, however, the space usage would grow linearly.

#### K2. Other Data Structures

Two other data structures that could meet the same requirements for this assignment include a dictionary and an array.

## K2a. Data Structure Differences

The hash table created in this program utilizes nested lists to store and locate items. Items are found using a search function. Alternatively, a dictionary stores key-value pairs and returns the value of an item given a key. It is possible that packages could be stored in an array for the purpose of this assignment, however, referencing items in an array would require knowledge if its index and would add complexity to correctly locating packages.

#### M. Professional Communication

Throughout this project and report, I have attempted to utilize professional communication for the purpose of explaining project requirements.

#### L. Sources - Works Cited

C950 - Webinar-1 - Let's Go Hashing. Directed by Cemal Tepe, WGU,

http://wgu.hosted.panopto.com/Panopto/Pages/Auth/Login.aspx?instance=WGU&Auth=Se

ssion View & Return Url = %2 f Panopto %2 f Pages %2 f Viewer. aspx %3 fid %3 d f 08 d 7871-d 57a-496e-a 6a1-ac7601308c71.

Chase, Connor, et al. An Evaluation of the Traveling Salesman Problem.

https://scholarworks.calstate.edu/downloads/xg94hr81q.

"Dijkstra's Algorithm." Wikipedia, 2 Feb. 2023. Wikipedia,

https://en.wikipedia.org/w/index.php?title=Dijkstra%27s\_algorithm&oldid=1137012547.

"Nearest Neighbour Algorithm." Wikipedia, 28 Feb. 2021. Wikipedia,

https://en.wikipedia.org/w/index.php?title=Nearest\_neighbour\_algorithm&oldid=10094152 97.