OR6205 13611 Deterministic Ops Research SEC 09 Fall 2019

Boston Sightseeing – A Mathematical Approach

Submitted by

Abhilash Janardhanan

Shashank Shet

Pratik Nagare

NORTHEASTERN UNIVERSITY

2019-2020

**ABSTRACT**

Tourists visiting new cities are often faced with the decision problem of choosing the places to sightsee and the sequence of visits themselves all while sticking to a budget. This problem falls under a class of problems known as the Tourist Trip Design Problem (TTDP). There are multiple variants for the TTDP, but in this report we attempt to solve an optimization model where the traveller wishes to maximize the total number of places visited over multiple days on top of satisfying several adjoining constraints.

For this project, we utilize data of the tourist places within the city of Boston, obtained from Trip advisor web pages. The aim of this project is to provide students based in Boston an optimal trip plan based on their total travel days and budget criteria. From the problem - variables, constraints and objective function were formulated and modelled in AMPL utilizing concepts from both Integer optimization and network models. Results meeting the constraints provided are generated to obtain the largest number of places to visit. Additionally, sensitivity analysis of the problem was carried out by changing parameters to determine the variation in the obtained results.

Keywords: Tourist Trip design problem (TTDP), Tour planning, Trip Optimization

**ACKNOWLEDGEMENT**

We wish to express our sincere thanks to Prof. Muhammad Noor E Alam, course instructor, for his inspiring lectures which elevated our interest in Operational Research. We would also like to thank Mr. Tasnim Ibn Faiz, course TA, for providing us with all the necessary guidance for this project.

We also place on record, our sense of gratitude to one and all, who directly or indirectly have lent their hand in this project.

Thank You.

**CONTENTS**

[**I.** **INTRODUCTION** 1](#_Toc26150087)

[**i.** **PROBLEM DEFINITION** 1](#_Toc26150088)

[**ii.** **OBJECTIVE** 2](#_Toc26150089)

[**II.** **LITERATURE REVIEW** 3](#_Toc26150090)

[**III.** **METHODOLOGY** 6](#_Toc26150091)

[**IV.** **ALGEBRAIC FORMULATION** 11](#_Toc26150092)

[**V.** **FINDINGS** 13](#_Toc26150093)

[**VI.** **SENSITIVITY ANALYSIS** 15](#_Toc26150094)

[**i.** **SENSITIVE ANALYSIS BASED ON CHANGE IN BUDGET** 15](#_Toc26150095)

[**ii.** **SENSITIVE ANALYSIS BASED ON CHANGE IN HOURS IN A DAY** 18](#_Toc26150096)

[**iii.** **SENSITIVE ANALYSIS BASED ON CHANGE IN NUMBER OF DAYS** 20](#_Toc26150097)

[**VII.** **FUTURE SCOPE OF THE PROJECT** 22](#_Toc26150098)

[**VIII.** **CONCLUSION** 24](#_Toc26150099)

[**IX.** **REFERENCES** 25](#_Toc26150100)

**LIST OF FIGURES**

[Figure 1: Trip Advisor 6](#_Toc26135553)

[Figure 2: Information of location 7](#_Toc26135554)

[Figure 3: All locations plotted on map 8](#_Toc26135555)

[Figure 4: Distance of a location from every other location 8](#_Toc26135556)

[Figure 5: Sensitivity Analysis W.R.T to Budget (a) 15](#_Toc26135557)

[Figure 6: Sensitivity Analysis W.R.T to Budget (b) 16](#_Toc26135558)

[Figure 7: Sensitivity Analysis W.R.T to Budget (c) 17](#_Toc26135559)

[Figure 8: Sensitivity Analysis W.R.T to Number of Days (a) 18](#_Toc26135560)

[Figure 9: Sensitivity Analysis W.R.T to Number of Days (b) 19](#_Toc26135561)

[Figure 10: Sensitivity Analysis W.R.T to Hours in a Day (a) 20](#_Toc26135562)

[Figure 11: Sensitivity Analysis W.R.T to Hours in a Day (b) 21](#_Toc26135563)

**LIST OF TABLES**

[Table 1: Algebraic Expressions 11](#_Toc26135574)

[Table 2: Travelled.tab output 13](#_Toc26135575)

# **INTRODUCTION**

Whenever a traveller wishes to explore a city, they have multiple aspects to consider. The chief among them being, the duration of the trip, the places they wish to visit as well as the amount of money they are willing to expend. It becomes a hassle-filled process when the traveller must make active decisions to trade-off between the places to visit vs the amount of travel time and the value for the money spend. They often resort to enlisting the help of a local guide or go through multiple travel websites, peruse the reviews and ratings of each place and make an informed decision to utilize the time available optimally.

In order to aid in the process of choosing the set of places to visit each day, this very problem can be formulated as a route optimization model. There have been multiple advances in the field of Operation research with respect to route/trip optimization as they form the backbone of modern-day supply chain networks, core algorithms that power the rise sharing economy and providing consumers with navigation systems. Solving this toy problem of sightseeing trip optimization could help us understand better the nuances of modelling these types of problems as well as use this as the starting point for a personalized trip plan.

## **PROBLEM DEFINITION**

For this problem, we don the persona of a graduate student who has just moved into the city of Boston. The life of a graduate student can be very tricky with the youthful and experience seeking nature nudging towards exploring the places around them but faced with the reality of being in the lower end of the income chains. In order to provide the best value for their money while still being able to travel to as many places as possible, we aim to formulate this as an optimization problem.

Typical optimization problem we have encountered so far in the curriculum with respect to shortest path can be solved using the Travelling salesman approaches. In a travelling salesman case, however, all nodes must be visited. In our case, the traveller has the option to visit a subset of the places from the entire set of places that Boston has to offer. Moreover, it would not be prudent to visit the same place multiple times, as that would reduce the diversity the explorer is exposed to. The traveller also has the option of choosing how many days of travel they wish to undertake. This formulation falls under the class of problems called as the Tourist Trip Design Problem (TTDP).

## **OBJECTIVE**

Based on the inputs from a person in the form of budget (in dollars), hours in a day that is to be spent traveling and number of days that is to be spent traveling, number of miles covered per gallon and cost per gallon of gas, the objective is to find out:

* The maximum number of places that can be visited.
* The order of the places that the person should visit.
* Cost of the entire trip

# **LITERATURE REVIEW**

1. **The Travelling Salesman Problem the Function Of Transport Network Optimization** (Slavomir Vukmirović, Drago Pupavac)

Investigating various alternative routes for transport one can easily find larger number of optimal transport relations enables management to get greater flexibility of companies and also, faster and easier decision making. So managers can consider different optimal alternative solutions and select the most optimum solution from the point of view with various relevant criteria. Empirical research affirm that even suboptimal solutions (within 1% deviation) can be used for effective problem solving.

1. **Local Search for Vehicle Routing and Scheduling Problems: Review and Conceptual Integration (**Birger Funke, Tore Grunert)

The aim of this paper is to select a local best neighbor and to reach a local optimum as fast as possible. This can be achieved by search methods, which do not scan all neighbor possibilities explicitly. Analysis allow to perceive how the properties of the partial moves and the constraints of the VRSP affect the selection of best possible search technique.

1. **Disruptive Transportation: The Adoption, Utilization and Impacts of Ride-Hailing in the United States** (Regina R. Clewlow, Gouri Shankar Mishra)

By gathering data through a representative panel in seven major U.S. metropolitan areas, this

paper presents initial investigation on the adoption of ride-hailing services and their potential effects on travel behavior, including vehicle ownership, trip generation, mode substitution, and vehicle miles traveled. One cannot supposed to have the travel behavior effects associated with ride-hailing transfer to other shared modes, or vice versa. That is, the results are specific to ride-hailing, and do not necessarily apply to car sharing, bike sharing, or micro transit services.

1. **A Column Generation Algorithm for Vehicle Scheduling and Routing Problems**

(Tasnim Ibn Faiz, Chrysafis Vogiatz, Md. Noor-E-Alam)

This study find out a variant of the open vehicle pickup and delivery problem from the perspective of resource supplier. The suppler should pick up and deliver rentals vehicle in given window of time. Author represent two models for vehicle acquisition scheduling and routing decision problem: an arc based model and path based model. The arc-based MIP model is solved with the exact method; the model gives the number of vehicles to rent, their starting times, and finishing times. The path-based integer model, contradictory, is solved with the help of column generation algorithm. Paths (combination of tasks) are generated in the pre-processing steps. Researcher presented two path generation algorithms: the first algorithm generates all possible paths (compatible task combinations), while the second algorithm provides a set of paths with least idle times for starting the initial restricted master problem. Both algorithms work very fast and the path generation times are very low near to negligible. Researcher gave complete numerical example to show how the proposed models can be used to develop schedules and routes for vehicles to deliver relief materials to the affected people in a post-disaster period. Author compared the outcomes and solution times given by the two models for five different size problems. The path-based model and column generation algorithm perform better than the commercial solver. The reduction in solution time by the column generation algorithm is more remarkable as problem size increases. The reduction in solution time achieved by the column generation algorithm and the fast path generation algorithms show that the model can be re-solved again and again with updated task list to imitate a dynamic vehicle scheduling and routing problem. During humanitarian crises, when the damage status and demand rates in the affected region are revealed intermittently over time, these algorithm will be most effective in repeatedly solving the decision model with the most updated data to get optimal solution very quickly.

1. **GA Based Traveling Salesman Problem Solution and its Application to Transport Routes Optimization** (U. Hacizade, I. Kaya)

The goal of this study is to develop an effective method for the field audit team which scrutinize certain points (bus stops) to speed up the process and decrease the time required. Furthermore, it is also aimed to generate optimum software which is based on Genetic Algorithms approach. In this context, the most well-known traveling salesman problem with relational optimization problem is defined as a method of solution used in genetic algorithms. Traveling Salesman Problem is a problem about finding the shortest route starting from one city and turning back to the same city while considering only one pass through each city (points, nodes or components) where the distances between each city are given. There are different methods for solving the Travelling Salesman Problem. TSP algorithms to solve problems of medium size required for very long periods are not very satisfactory. Despite, heuristic methods results in the shortest period of time are near optimum.

1. **Route Optimization Techniques: An overview** (Bale, D. L. T .Nwachukwu, E. O. Ugwu, C.)

The dynamic nature of road transport system and its associated uncertainty has caused high level of emotional apprehension among road users today. Most times people embark on journey without initial knowledge of what will befall them in terms of traffic conditions on the selected route. This study reveals the real nature of road transport and different routing problems in road transportation. Different route optimization techniques both non-intelligent and intelligent were discussed in this given paper. Their general strength and weakness were shown and from the survey the intelligent approach proffer more proper solutions to the route optimization problems when faced with the reality of intricate and multi-criteria scenario.

# **METHODOLOGY**

An extensive literature review was carried out to find papers or articles that were on similar lines that of our project to know what research was already carried out.

Subsequently, data was gathered from Trip Advisor by web scraping the individual pages of each tourist spot that appeared in the ‘Top Places to visit in Boston’ list. The most popular 50 locations based on ratings were shortlisted for the solution phase. The information gathered through this process included: the coordinates of the location, the working hours of the location, individual ratings, number of reviews, cost of admission/Fare and package deals and the recommended amount of time to be spent at a location.

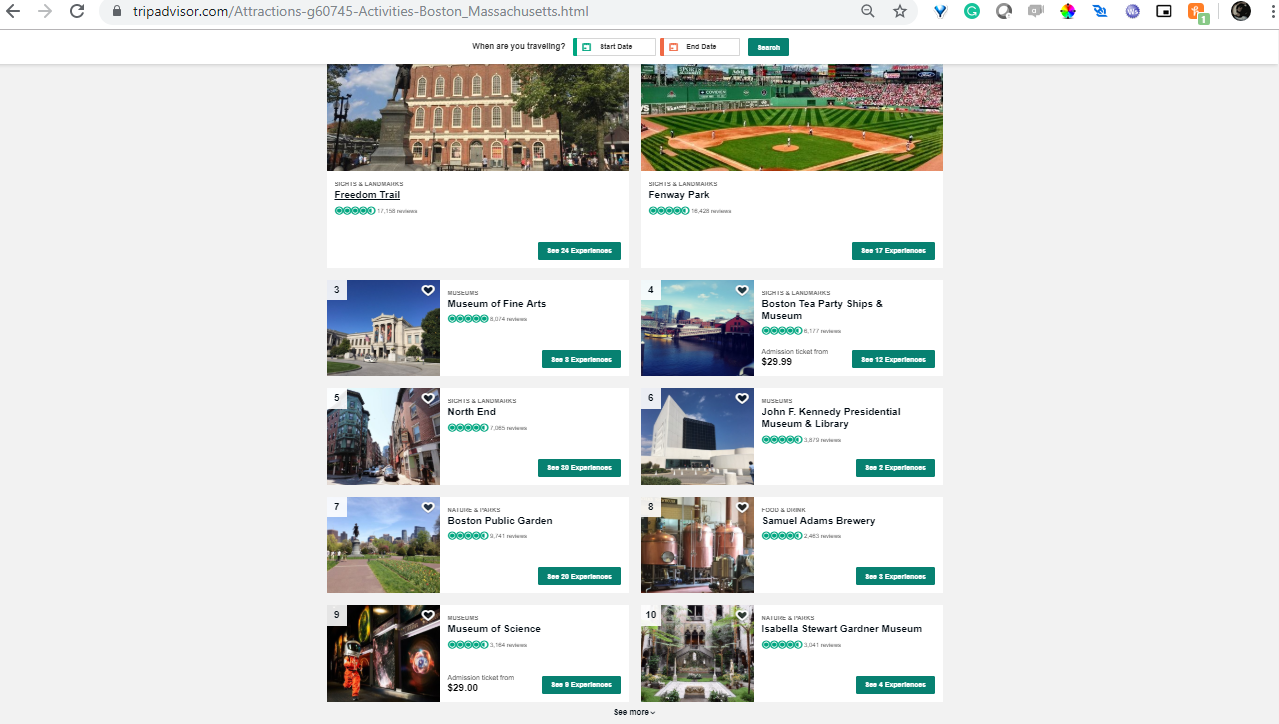


Figure : Trip Advisor

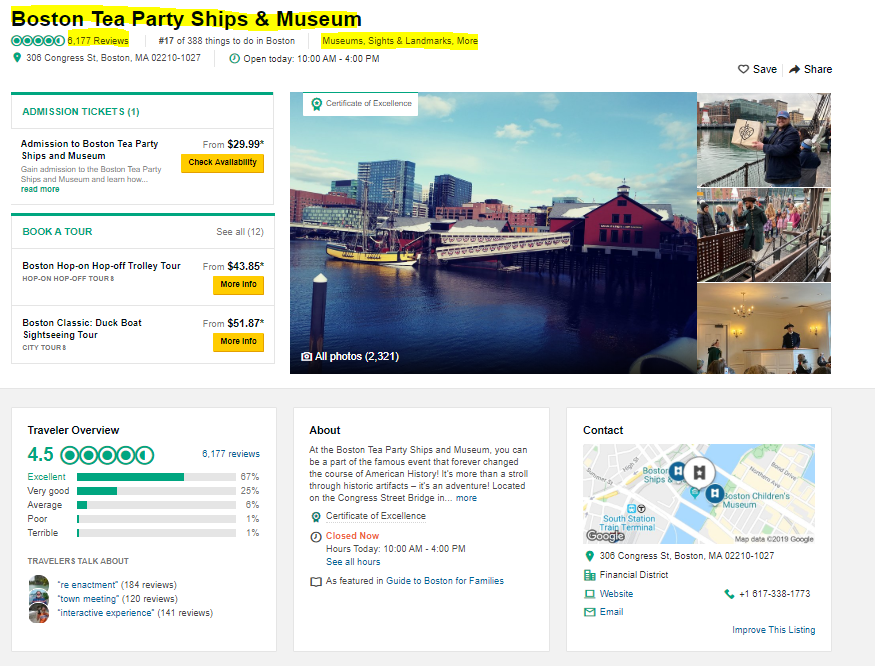
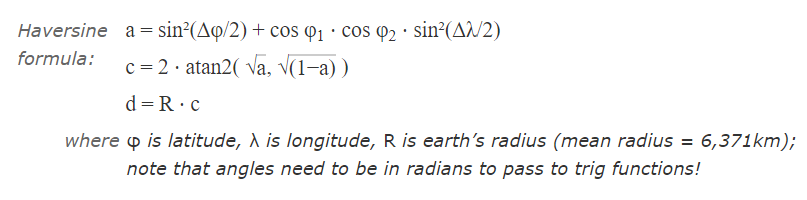


Figure : Information of location

The coordinates that were obtained from the map were used to find out the exact location of the spot in Boston as well as calculating distance between each spot to other sightseeing locations. This was carried out by using the Haversine formula from an online calculator by providing the coordinates of the locations in order to obtain the distance as outputs.



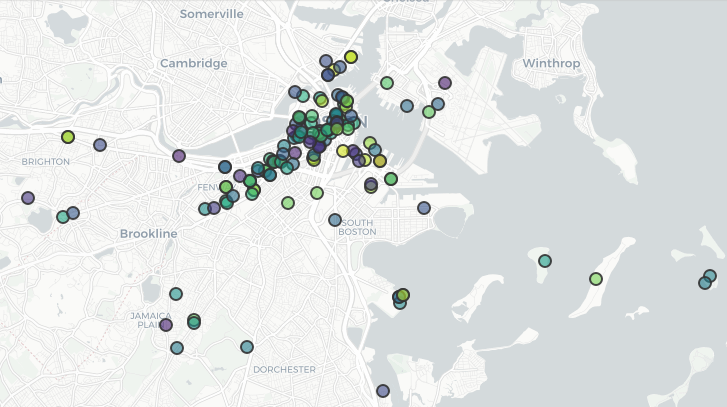


Figure : All locations plotted on map

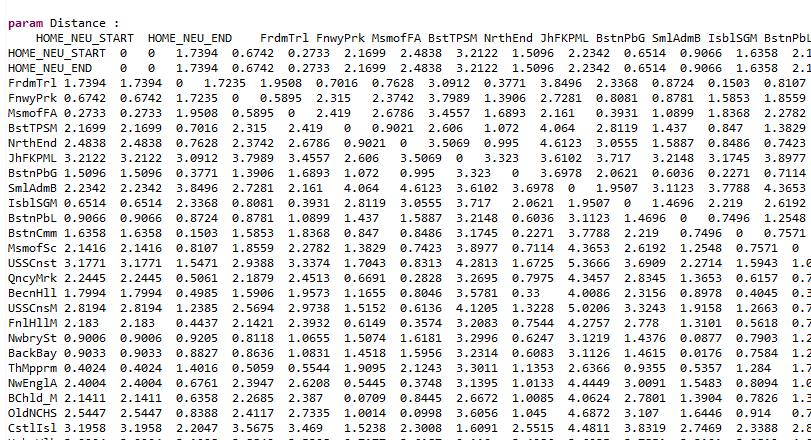


Figure : Distance of a location from every other location

The primary decision variables of the given problem which are to be provided by the user are:

* Budget spent on the tour (in dollars)
* Number of hours in a day (in hours)
* Number of days (in days)

Certain secondary decision variables have been considered constant throughout the problem but can be varied depending upon the input provided by scientists and engineers. They are:

* Mileage obtained from the vehicle used for traveling between location (in miles per gallon)
* Cost of gas or fuel for the vehicle (dollars per gallon)

For computing the desired solutions, AMPL IDE (Version 3.5.0.201802211250) was used. Certain situations have been made to make the calculation easier (algebraic and on AMPL) and for better understanding while making the AMPL data and mod file. They are as follows:

* A Homebase is set from where the day begins and ends. This is done so that the person has a place to start from and to return to. The coordinates put for this location are of the Northeastern University’s address (360 Huntington Ave, Boston, MA 02115). These dummy places are named HOMEBASE\_START and HOMEBASE\_END. A person can only leave from HOMEBASE\_START and not go out from, similarly, a person can only reach and not leave from HOMEBASE\_END.
* The number of days which can be given for traveling can be equal or greater to one. They will be selected manually.
* The person has to leave each place that is visited (except for HOMEBASE\_END).
* The speed of vehicle used for traveling between the locations is 25 Miles per hour.
* To account for traffic and non-straight path between the locations, a time factor of 1.75 is taken.
* The time spent at the sightseeing locations is taken in hours and is within the hours per day.

The values of primary deciding variable for the given problem are taken as:

* Budget spent on the tour = 200 dollars
* Number of hours in a day = 10 hours
* Number of days = 8 days

These values are taken as the base values of the project as they can be considered to be the values of the respective parameter for an average student coming to Boston.

The results of AMPL will include:

* Distance covered per day (in miles)
* Total number of places visited out of the given number of locations per day.
* Total cost incurred (which includes all types of cost while sightseeing like entry fees, expenditure for gas, etc.)

For finding out the computational statistics, certain parameters were also found out to understand the prowess possessed by the computer in carrying out the calculations. The parameters are:

* Total solve elapsed time
* Total solve system time
* Total solve user time
* Total solve time

Doing this helps to understand the amount of time the computer is taking to calculate the solutions for the given problem and whether there needs to be an improvement in the formulation or the system on which the solutions to the problem are being found.

# **ALGEBRAIC FORMULATION**

The algebraic expressions of the problem can be given as shown in the table below:

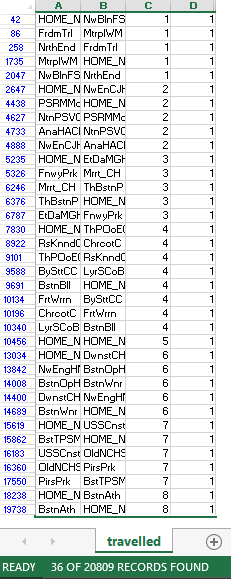
Table : Algebraic Expressions

|  |  |
| --- | --- |
| Decision variable of Travel | where,  (dim 1) s = index of place at which travel starts,  (dim 2) e = index of place at which travel ends,  (dim 3) d = day of travel |
| Homebase Starting index | 0 |
| Homebase Ending index | N+1 |
| Indices of travel places | 1, 2, …., N |
| Total number of Days |  |
| Cost of Admission at a place ‘s’ |  |
| Time Factor |  |
| Cost Factor | = cost per gallon /miles per gallon |
| Distance between place ‘s’ and ‘e’ in miles |  |
| Time to spend at place ‘s’ in hrs |  |
| Time value of entering a place ‘s’ on a day ‘d’ |  |
| Minimum miles per hour for traveling |  |
| Objective: Total Visits |  |
| Budget constraint |  |
| In degree out-degree match constraint |  |
| Place singularity constraint |  |
| Self-loop constraint |  |
| No path to Source Base or from End Base  (No Invalid Path constraint) |  |
| Time spent Constraint |  |
| Time value continuity Constraint |  |
| Constraint to Start at Homebase |  |
| Constraint to End at Homebase |  |

# **FINDINGS**

Once the computation is completed on AMPL, we get a tab file by the name "traveled" in the selected directory of AMPL. It can be viewed using any software suitable for the viewer, however, for this project, Microsoft Excel was used. The tab file when viewed, will give you the sequence of places to be visited each day, starting from the first to the last day, assuming that more than one day was used for sightseeing. The data on the tab file being viewed can be filtered for better visualization. The data found in the excel file is as follows:

Table : Travelled.tab output



The results obtained on AMPL are:

* Distance covered per day = 76.5116 miles (16.7108 + 4.6659 + 7.9915 + 21.4188 + 0 + 13.7736 + 8.2482 + 3.7028)
* Total number of places visited out of the given number of locations per day = 28 places (4+4+4+7+0+4+4+1)
* Total cost incurred = $192.012 (which includes all types of cost while sightseeing like entry fees, expenditure for gas, etc.)

The various amount of time taken (in seconds) by the system to find the solution is found to be as follows:

* Total solve elapsed time: 33.86
* Total solve system time: 6.35938
* Total solve user time: 57.875
* Total solve time: 64.2344

# **SENSITIVITY ANALYSIS**

Sensitivity analysis, also known as what-if analysis, is carried out to determine how independent variable values will have an impact on the dependent variable, as well as the objective function under a given set of assumptions. For the given problem, we will carry out a sensitivity analysis by changing the values of budget, hours in a day and number of days.

## **SENSITIVE ANALYSIS BASED ON CHANGE IN BUDGET**

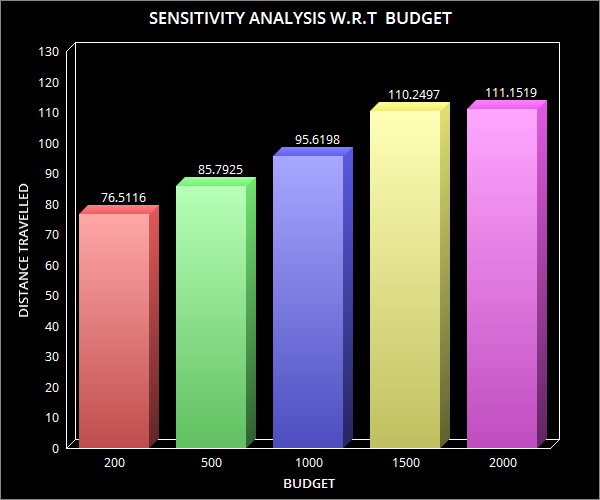


Figure : Sensitivity Analysis W.R.T to Budget (a)

The graph here represents distance travelled by the traveller based on the budget. A growth in distance travelled can be observed as there is an increase in budget. This is because as there is more monetary capability, the traveller can travel more and hence the distance travelled is more.

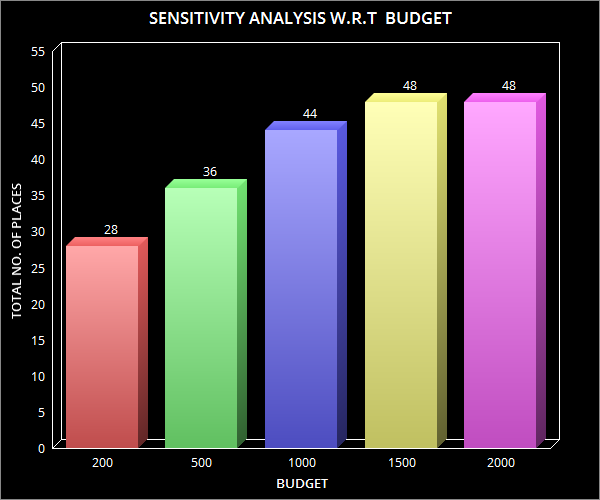


Figure : Sensitivity Analysis W.R.T to Budget (b)

The graph here represents the total number of places travelled by the traveller based on the budget. A growth in total number of places can be visited is observed as there is an increase in budget. This is because as there is more monetary capability, the traveller can travel more and hence the number of places travelled is more.

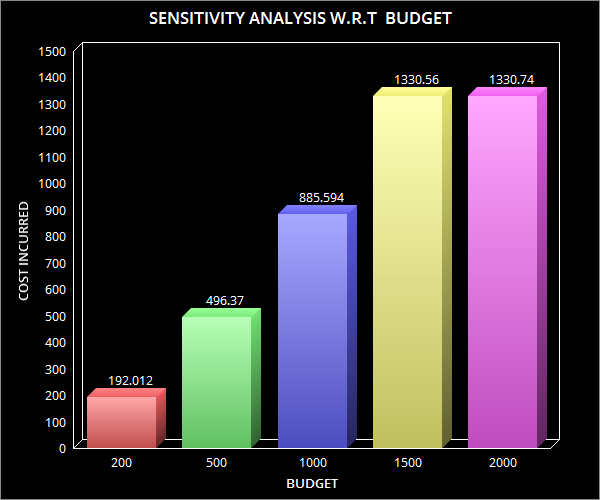


Figure : Sensitivity Analysis W.R.T to Budget (c)

The graph here represents cost incurred to the traveller based on the budget. A growth in cost incurred can be seen as there is an increase in budget. This is because as there is more monetary capability, the traveller can visit more places and hence the spending done on the places is more.

## **SENSITIVE ANALYSIS BASED ON CHANGE IN HOURS IN A DAY**

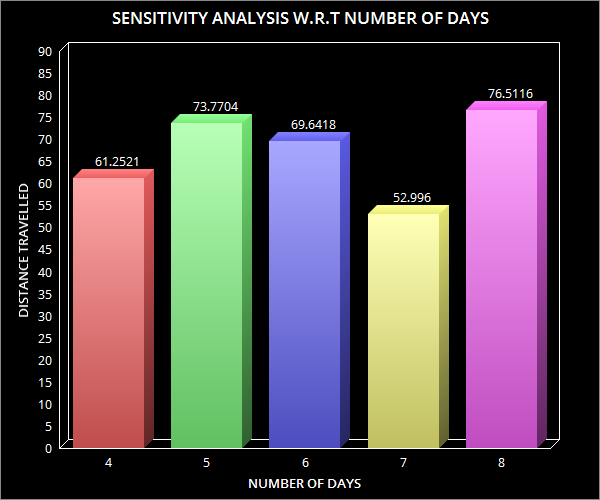


Figure : Sensitivity Analysis W.R.T to Number of Days (a)

The graph here represents distance travelled by the traveller based on number of days. A growth in distance travelled can be seen, then there is a fall and then again a rise in distance travelled takes place.

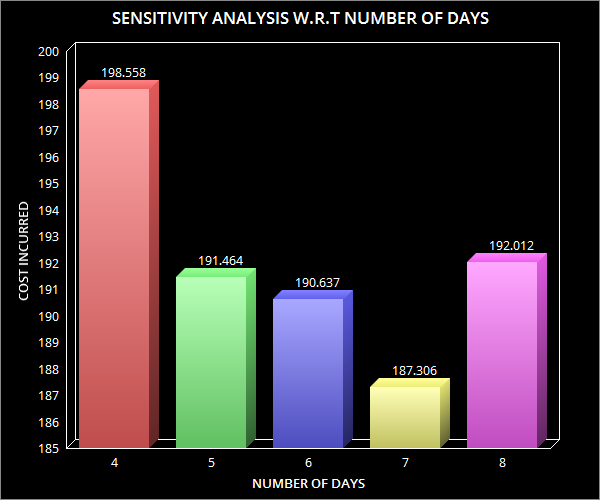


Figure : Sensitivity Analysis W.R.T to Number of Days (b)

The graph here represents cost incurred to the traveller based on number of days. A decline in cost incurred can be seen and then there is a rise in cost incurred as the number of days increases.

## **SENSITIVE ANALYSIS BASED ON CHANGE IN NUMBER OF DAYS**

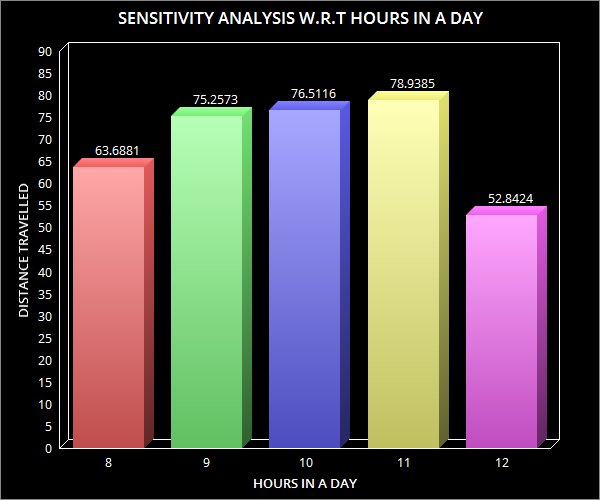


Figure : Sensitivity Analysis W.R.T to Hours in a Day (a)

The graph here represents distance travelled by the traveller based on hours in a day. A growth in distance travelled can be seen and then there is a fall in distance travelled as the number of hours in a day is increased.

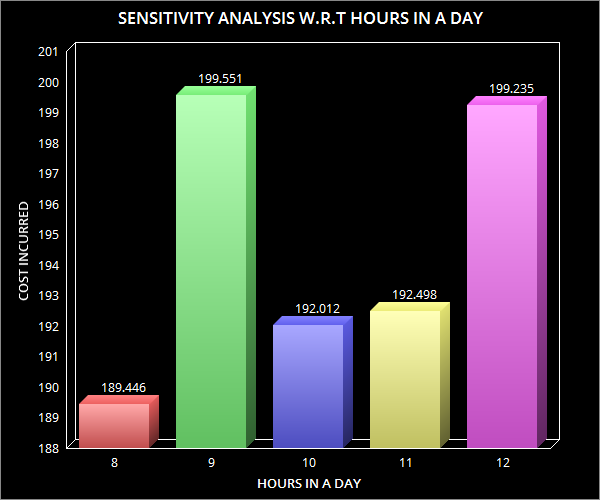
****

Figure : Sensitivity Analysis W.R.T to Hours in a Day (b)

The graph here represents cost incurred to the traveller based on hours in a day. A growth in the cost incurred to the traveller can be seen, then there is a decline and then again a rise in cost incurred takes place as the number of hours in a day increase.

# **FUTURE SCOPE OF THE PROJECT**

While the project covers most of the constraints, there are certain aspects of research yet to be explored and improved upon. They are as discussed below.

1. Accurate distance:

While calculating the distance between the places, the linear distance was considered between the locations, multiplied by a scaling factor to account for the traveling distance. However, for a more realistic calculation, it would be better to calculate the traveling distance between each location. This is because the path taken by a vehicle is based on the roads between the locations.

1. Accessibility of the sightseeing locations:

The assumption is taken that the sightseeing locations are open at the time the traveler is going to the location. It could go granular to the stage that research could be split into an hour's time slot to get more realistic results.

1. Long-time for actual data processing:

When the computation of 150 locations was attempted, we had a problem to get a result on AMPL. On a computer with an i5 processor with 8 GB ram, the computation was not finished even after 12 hours (747 minutes to be accurate). The better result may be obtained with a better processing power computer.

1. The jumbled sequence of paths:

In the tab file, although we did find the places that should be visited on each day, on few occasions we get a jumbled path and then we need to manually make the sequence of locations to be visited. This aspect can be worked upon in the future to get the exact sequence of locations.

1. Incorporating personal preferences:

The data web scrapped from Trip Advisor had locations ranked based on the ratings that they received. This helps the traveller to decide which locations to visit based on the ratings. However, the traveller cannot make a sequence of places to visit based on personal preferences. If the traveller has to use the mathematical approach done in the project, the traveller will have to follow the sequence of the locations given by the AMPL solution.

1. Incorporating Mandated breaks/Restaurant travel for eating times during the trip:

The solution obtained in the project assumes that the traveller will be continuously travelling to visit the sightseeing locations and not take breaks, which may include using the restroom or going to a restaurant to lunch/dine. The number of hours in a day that the traveller spends sightseeing should also include breaks as per the traveller’s wishes.

1. Optimization results provided at a more granular level, say, hour-level

# **CONCLUSION**

The objective of finding out which locations should a person (graduate student) visit based on the budget, number of days and number of hours per day, was successfully achieved. The 'traveled .tab' file gave the user the division of places to be visited per day. At the same time, solutions of AMPL helped us find out the distance that the user will be covering per day sightseeing at the locations, along with the places visited per day and the total amount that the user will have to pay in order to visit those places, with surety that the expenditure will never go above the budget. The sensitivity analysis carried out at the later stage will help the user understand different scenarios by changing the deciding parameters and to make a decision on to which parameter can be varied and how much the change in parameter can affective the objective function on the basis of change in budget, number of days and number of hours of days.

The solution obtained can be improved upon with a lot of potential in making the solution even more accurate. There are certain aspects of the project that can be worked upon to make it even more efficient and the solution even more realistic.

# **REFERENCES**

1. **U. Hacizade, I. Kaya.**  GA Based Traveling Salesman Problem Solution and its Application to Transport Routes Optimization, *Science Direct. 2018.*
2. **Tasnim Ibn Faiz, Chrysafis Vogiatz, Md. Noor-E-Alam.** A column generation algorithm for vehicle scheduling and routing problems, *Science Direct. 2019.*
3. **Bale, D. L. T. Ugwu, C. Nwachukwu, E. O.** Route Optimization Techniques: An Overview *International Journal of Scientific & Engineering Research, Volume 7, Issue 11, November-2016.*
4. **Naixue Xiong, Wenliang Wu, Chunxue Wu,** An Improved Routing Optimization Algorithm Based on Travelling Salesman Problem for Social Networks, *MDPI June 2017.*
5. **Slavomir Vukmirović, Ph.D, Drago Pupavac, Ph.D.** The Travelling Salesman Problem in the Function of Transport Network Optimization.
6. **Patrick Jaillet, Xin Lu** Online Traveling Salesman Problems with Service Flexibility, *August 2011.*
7. **Trip Advisor [Online]** [**https://www.tripadvisor.com/Attractions-g60745-Activities-Boston\_Massachusetts.html**](https://www.tripadvisor.com/Attractions-g60745-Activities-Boston_Massachusetts.html)**.**
8. **Movable Type scripts [Online]** [**https://www.movable-type.co.uk/scripts/latlong.html**](https://www.movable-type.co.uk/scripts/latlong.html)**.**