

For office use only

Team Control Number

For office use only

T1 \_\_\_\_\_

**72973**

F1 \_\_\_\_\_

T2 \_\_\_\_\_

F2 \_\_\_\_\_

T3 \_\_\_\_\_

Problem Chosen

F3 \_\_\_\_\_

T4 \_\_\_\_\_

**B**

F4 \_\_\_\_\_

---

**2018**  
**MCM/ICM**  
**Summary Sheet**

## **PTPs: E-Guide Helping Plan a Tour**

### **Summary**

When travelling in a new city, it is important but challenging to choose Points of Interests (POIs) to visit and design the route. The quality of a tour is highly dependent on tour planning. However, planning a tour takes much time and effort.

To free the tourist from tour planning, our paper provides a personalized tour planning system (PTPs), which can offer the user a tour plan based on his/her preference, budget and time constraint. Moreover, we revise our PTPs to deal with “super POIs”, namely, ones whose scale and structure cannot be omitted.

We first set up a criterion to characterize each POI’s type and popularity. Then we build our PTPs model and examine its effectiveness by applying it to New York City (NYC).

PTPs consists of three sub-models: Evaluation Model, Route Planning Model, and Hotel & Restaurant Recommendation Model. There is a figure showing the connection between all parts.

Evaluation Model gives each POI a “satisfaction score”. It measures the closeness between the type of a POI and the user’s preference, meanwhile considers the POI’s popularity, and produce a score to describe how satisfactory a POI is for the user.

Route Planning Model generates an optimal touring route under the constraints of time and budget. We apply a genetic algorithm to find a route with highest total satisfaction score.

Hotel & Restaurant Recommendation Model

To acquire the data we need,

**Keywords:** tour planning, route design,

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Background . . . . .	2
1.2	Restatement of problem . . . . .	2
1.3	Literature review . . . . .	2
<b>2</b>	<b>Assumptions</b>	<b>3</b>
<b>3</b>	<b>Nomenclature</b>	<b>3</b>
<b>4</b>	<b>Model I: Personalized Tour Planning System (PTPs)</b>	<b>4</b>
4.1	Rate the POIs: Evaluation Model . . . . .	4
4.2	Construction of the graph . . . . .	4
4.3	Plan the Tour Route . . . . .	5
4.4	Restaurant & Hotel Recommendation . . . . .	6
<b>5</b>	<b>Model II: PTPs with Super POIs</b>	<b>7</b>
<b>6</b>	<b>Implementation</b>	<b>7</b>
6.1	Data . . . . .	7
6.2	Analysis & Results . . . . .	8
<b>7</b>	<b>Model Analysis</b>	<b>9</b>
7.1	Sensitivity Analysis . . . . .	9
7.2	Strengths & Weaknesses . . . . .	9
7.2.1	Strengths . . . . .	9
7.2.2	Weaknesses . . . . .	9
<b>8</b>	<b>Conclusion</b>	<b>10</b>
	<b>Appendices</b>	<b>12</b>

# 1 Introduction

## 1.1 Background

Travelling to another city is a delightful experience for most people. There are so many places of interest to visit, such as museums, parks and so on. With limited time, tourists have to choose among those places and design their own route. Therefore, tour planning is an important task for tourists. However, it takes a lot of time and massive information to make a perfect plan, and several constraints(e.g. cost, transportation time, etc.) should be considered.



Figure 1: View of a city.

## 1.2 Restatement of problem

As required by the question, we are supposed to determine a tour route among Points of Interests (POIs), which include spots, hotels and restaurants. The route we provide should cover the popular POIs, meanwhile cater for the user's personal preference, namely, maximize the user's satisfaction, and meet the constraints.

Furthermore, for "super POIs" (those with large scales) we should consider their inner structures, positions of entrances and exits and give a more subtle tour plan.

## 1.3 Literature review

Existing study on tourism products recommendation applies one of the following four methods[1]:

- **Content-based recommendation**, which recommends tourism products similar to products chosen by the user.

- **Collaborative filtering recommendation**, which recommends the choice of users with similar preference.
- **Knowledge-based recommendation**, including constraint-based[2] and case-based, makes recommendation according to Knowledge-based rules.
- **Social-based recommendation**, which applies users' relationship in social media.

Besides, in other works, factors such as distance[3], user interests[4], and popularity[5] are taken into consideration. However, some of these methods are dependent on a large amount of historical datas, therefore have difficulties in cold-start. Moreover, datas on travelling are sparse, and getting these datas from users might be viewed as an invasion of privacy.

There is also study on travelling route planning[6]. The majority of current works need users to input the origin and destination of travelling, and cannot deal with accommodation arrangement, which limits their application. Therefore, we combine spot recommendation and route planning in our model, and get rid of reliance on historical datas.

## 2 Assumptions

To simplify the real life situation, we will make the following assumptions at the start of constructing our models.

- **Sightseeing is more important than accommodation.** User's preference of spots are considered in higher priority than restaurants and hotels.
- **We further assume that a tourist always stays in the same hotel during a trip.** The user doesn't have to bother carrying.
- **User's satisfaction on POIs solely depends on his/her preference and the popularity**, while the impact of other factors(e.g. mood, weather, etc.) is neglected. Because the quality of POIs is vague and subjective, thus hard to quantify, we assume that generally the more popular a POI is, the higher quality it is of.
- **Road between any two spots is straight, and its length is approximately equal to the great-circle distance.**
- **Ticket and traffic fares are ignorable compared to prices of hotels and restaurants.** Therefore, the cost of visiting a spot and travelling between spots can be viewed as zero.
- **The spots' scales are relatively small**, including super POIs. In the scale of the city, a super POI can still be viewed as a point. In the scale of a super POI, the subspots in it can be viewed as points. Therefore, the time taken moving from one spot to another within a super POI is ignorable.

## 3 Nomenclature

Notation	Definition	Unit
$\vec{cat}$	Vector that describes a spot's category	
$\vec{prf}$	Vector that describes the user's preference of spots' category	
$Intr$	Rate of user's interest to a certain spot	
$Pop$	Rate of POI's popularity among the public	
$Sat$	Score of satisfaction of a spot	
$t_{dur}$	Time duration of touring in a spot	hour
$t_{tra}$	Time duration of traffic between two spots	hour

## 4 Model I: Personalized Tour Planning System (PTPs)

In this section, we will discuss our models. Before diving into details, briefly explain the whole idea (as figure 2 shows). Firstly, we give each spot a score based on its popularity and closeness to user's preference. Then we construct a graph, containing scores, traffic time and tour duration. We apply a genetic algorithm to find a way maximizing the score under the constraints of time. Furthermore, we analysis the relation between final satisfaction score and some properties.

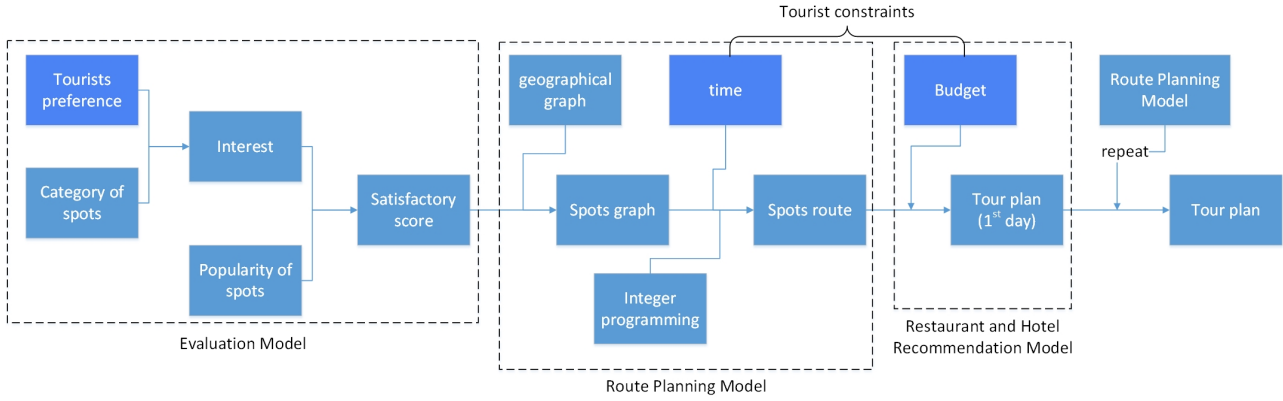


Figure 2: Overview of Personalized Tour Planning System

### 4.1 Rate the POIs: Evaluation Model

We define a series of dimensions to classify the POIs (e.g. landmark, museum, historical site, etc.) A spot is assigned a value ranging from 0 to 1 for each dimension. Thus we build a category vector  $\vec{cat} = [r_1, r_2, \dots, r_n]$ , where  $r_i$  denotes the rate of  $i^{th}$  category, to classify each spot. Through a user interface, the user is asked to input the following values:

- His/her "interest rate" to each category of POI;
- Total travelling time limit;
- Highest budget acceptable.

We normalize the data, thus getting the user's preference vector  $\vec{prf}$ , which is a unit vector. Then the rate of user's interest to a certain spot() can be calculated:

$$Intr(s1) = \cos < \vec{cat}(s1), \vec{prf} >$$

To quantify the each POI's popularity  $Pop$ , we acquire the datas of search volume and rating. We normalize the datas using a linear mapping, in which the maximum and minimum are converted to 1 and 0, respectively. Then we calculate the POI's satisfaction score by the following fomula:

$$Sat = \alpha Intr + (1 - \alpha) Pop$$

where  $\alpha$  is a positive weighting coefficient.

### 4.2 Construction of the graph

#### Data acquirement

After giving each POI a score, we calculate the distance between each two POIs by their latitudes and longitudes using Haversine fomula:

$$hav\left(\frac{d}{r}\right) = hav(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)hav(\lambda_2 - \lambda_1)$$

where:

- $hav$  is the haversine function:  $hav(\theta) = \sin^2(\frac{\theta}{2})$ ;
- $d$  is the distance between two spots along the great circle of the sphere;
- $r$  is the radius of the sphere, namely the radius of earth in this case;
- $\varphi$  and  $\lambda$  represent the latitude and longitude respectively.

We estimate the speed of each kind of transportation, and with the distance between spots  $d$  assigned, we can calculate the traffic time taken between each two spots.

**Construct the graph** With the datas prepared, we construct a weighted directed graph (figure 3), of which the nodes are spots. Every node is weighted by its score, and has touring duration as another property. The edge, which denotes road between the two spots it connects, is weighted by the traffic time.

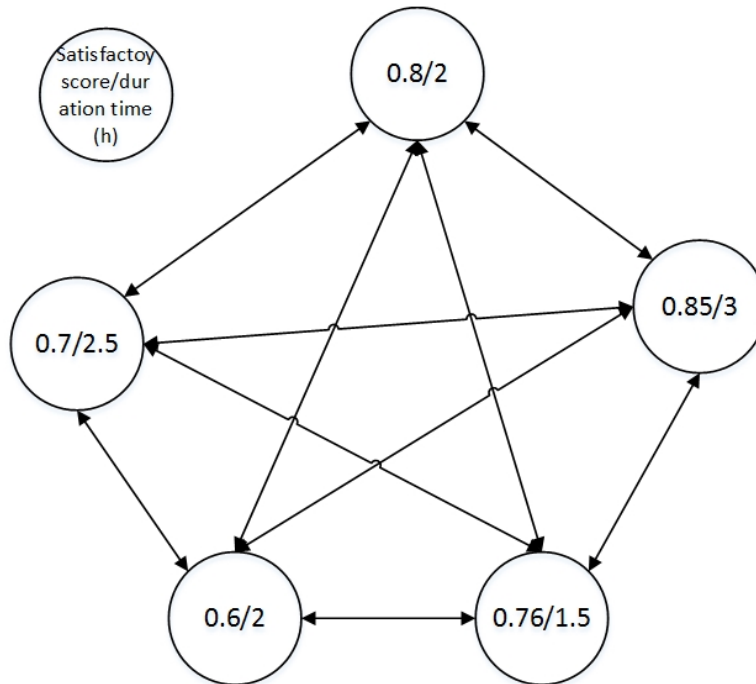


Figure 3: Structure of Weighted Graph Model

### 4.3 Plan the Tour Route

With all the spots rated and a weighted graph constructed, we apply a non-linear programming to design the route. Since the exact optimum is hard to reach, we use genetic algorithm (as figure 4

shows) with non-linear constraints to generate a near-optimal route:

$$\max \sum_{i=1}^n Sat(s_i) \text{ (fitness function)} \quad (1)$$

$$\text{s.t. } \sum_{i=1}^n t_{dur}(s_i) < t_{max} \quad (2)$$

$$s_i \text{ is only toured one time, } (i = 1, 2, \dots, n) \quad (3)$$

where:

- $s_i$  is the  $i^{th}$  spot,  $i = 1, 2, \dots, n$ ,  $n$  is the total amount of spots;
- $t_{max}$  is the longest total travelling time duration that the tourist can accept.

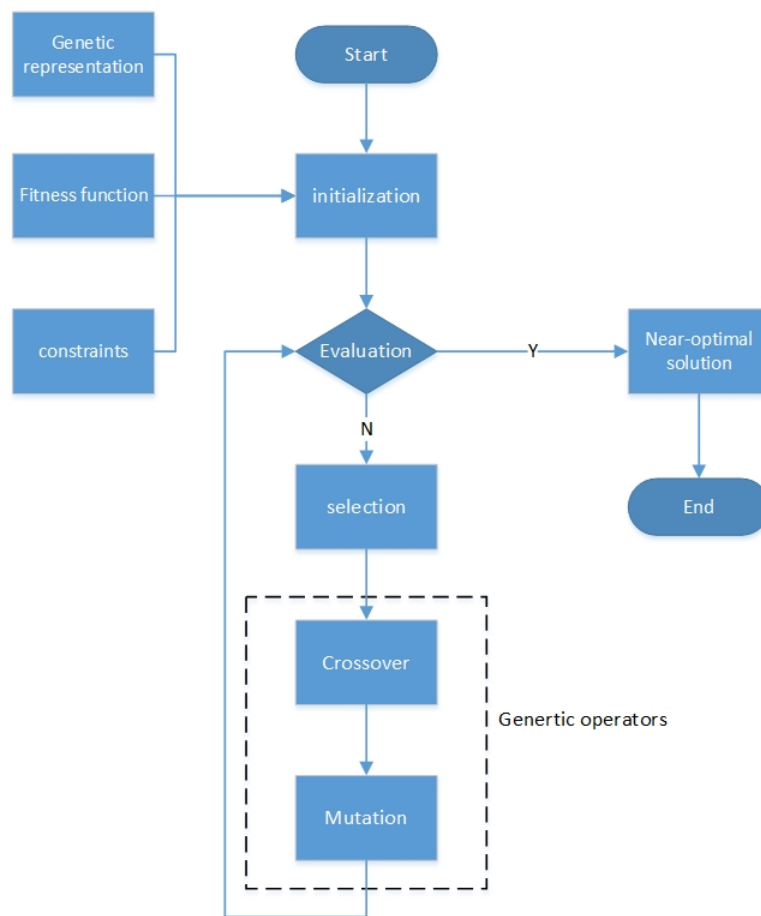


Figure 4: Sketch of genetic algorithm

we use path coding method, represent a path starting from spot  $s_{k0}$  via  $s_{k1}, s_{k2}, \dots, s_{km}$  with  $k0 - k1 - \dots - km$ . Applying the *ga* function in MATLAB, we find out a near-optimal route and the corresponding value of fitness function.

#### 4.4 Restaurant & Hotel Recommendation

After providing a route to travel around spots, we work on ways to recommend hotels and restaurant to consummate our tour planning system. Since hotels and restaurants are different in many ways, we use different methods in recommendation.

## Restaurants

From the route planning process, a travelling sequence  $\mathcal{S} = [s_1, s_2, \dots, s_n]$  is generated. We can also generate a sequence of starting times and ending times of touring the spots,  $\mathcal{T} = [t_1, t_2, \dots, t_{2n}]$ . If we set a meal time (e.g. 12am for lunch and 6pm for supper), we can find a  $t_m$  in  $\mathcal{T}$  that is closest to the meal time, and find the corresponding  $s_k$  of  $t_m$ .

Then we seek for restaurants that suits the constraints in a circle with  $s_k$  as center and a radius of 500 meters (might be adjusted due to the user's preference). With the similar method we apply in rating the spots, we give each restaurant a score according to its category, popularity and the user's preference. The one with highest score is what we recommend to the user.

## Hotels

As stated in 2, the tourist stay in the same hotel during the whole tour. In our model, we first list the spots in the first day's schedule, and select from the hotels economically suitable in a 500m-radius-circle around each spot. The hotel with best popularity among those selected ones will be our recommendation.

## 5 Model II: PTPs with Super POIs

In 4, we regard the POIs as points without shape, inner structure and volume. For the Super POIs, inner information should be take into consideration, therefore we revise our model (figure 5). To begin with, we evaluate the POIs and generate a optimal route using PTPs model. Then we transform each super POI to a weighted graph, in which the nodes are gates and subspots to view. Each subspot has a tour duration as its property.

In the tour route generated, each super POI has two adjacent nodes (two spots). We select two gates with least distance to the adjacent nodes and decide the entrance and exit according to the tour direction. After the selection of gates is done, we design the tour route inside the super POI. Using the evaluation model we give each subspot a satisfaction score. And because the traffic time between subspots is ignorable, we just need to maximize the total satisfaction score under the constraint of time.

To choose among all subspots in a super POI, we apply a greedy algorithm: Pick the highest-scored subspot, and then pick the highest-scored one from the left ones  $\dots$  until total tour duration reaches the time limit. Planning a route to tour around the chosen subspots starting from the entrance and ending at the exit is a travelling salesman problem (TSP).

## 6 Implementation

To test our PTPs model, we first apply it to New York City (NYC). Using the information we gathered, we apply the models and design a tour plan in NYC.

### 6.1 Data

Because there is no existing dataset about NYC's POIs, in this paper we ourselves analyze many datas from reliable sources and experiment with them. We acquire a list of POIs in NYC from the *Trip Advisor* website[7]. Then we gather detailed information from Google's Places API Web Service[8], including:

- The position of each POI, namely its latitude and longitude.



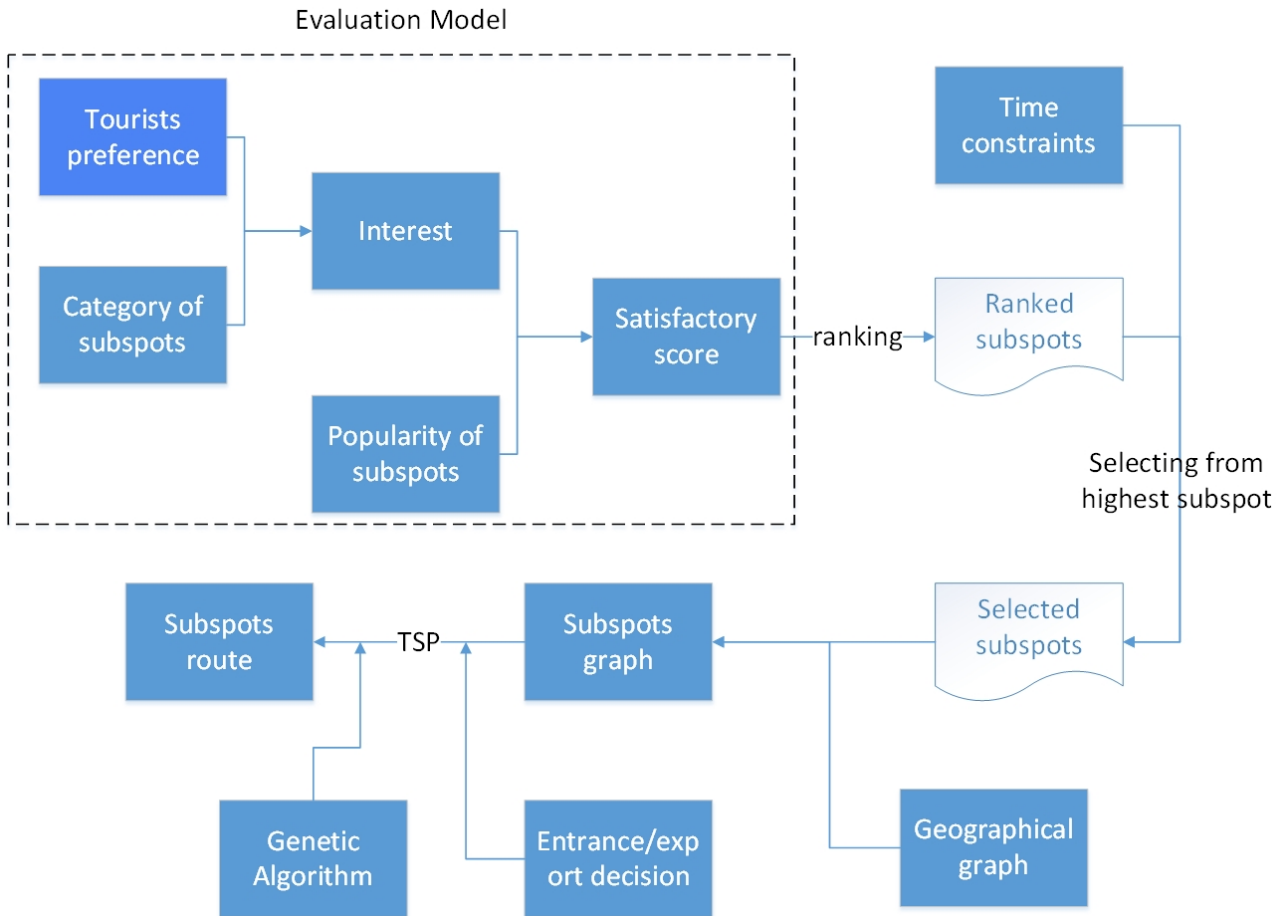


Figure 5: Overview of the revised model

- The ticket price of each POI;
- Search volume and public rate of each POI.

Applying the methods mentioned in 4, we get the distance between each two POIs in NYC, and each spot's price and satisfaction score.

## 6.2 Analysis & Results

### Tour plan

We simulate some “users” and input their requests (available time, preference of POIs, etc.) to our PTPs and generate personalized tour plans for them. Figure 6 shows the plan of route.

### Analysis of result

To examine the effectiveness of our PTPs, for each “user” we generate a series of alternative tour plans using a randomized algorithm. We calculate the total satisfaction score for each tour plan, including the PTPs-recommended one and the randomized ones. We find out that the tour plan PTPs generates has great advantages in comparison to other plans, that is to say, has the highest satisfaction score among the plans.

For further discussion, we calculate the satisfaction scores (fitness values) under different  $t_{max}$ , and visualize the result in figure 7.



Figure 6: The personalized tour route PTPs generates.

We then set different duration (days) of tour and respectively calculate the total satisfaction score  $Sat$ . Figure 8 shows the relationship between  $Sat$  and tour duration.

## 7 Model Analysis

### 7.1 Sensitivity Analysis

### 7.2 Strengths & Weaknesses

#### 7.2.1 Strengths

1. Have the ability of cold-start. Our model can give personalized tour plan without the user's historical data.
2. Combine the processes of choosing which POIs to go and designing the route in one app/web-site.
3. Use massive accurate data of New York City. Only with these dependable data can we validate that our model is built appropriately.
- 4.

#### 7.2.2 Weaknesses

1. Complex traffic conditions in cities are omitted. We simplify the traffic network to weighted graph. However, in the real condition there are much more factors influencing the traffic time.
2. Tourist's satisfaction is assumed only related to his/her interest and the POI's popularity. The impact of complex subjective and emotional factors during the tour are not considered.
- 3.

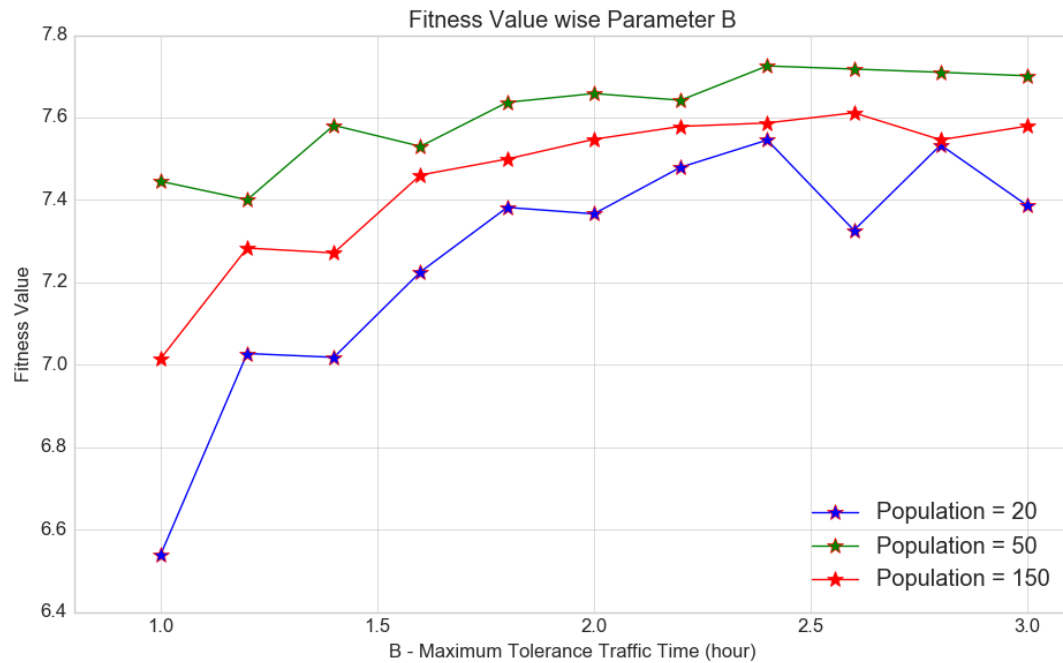


Figure 7: Relationship between satisfaction score and traffic time limit.

## 8 Conclusion

## References

- [1] Y. Wen, Y. Shi, G. Cai, Y. Miao, and G. Long, "A survey of personalized travel recommendation," *Computer Science*, no. Z11, 2014.
- [2] X. Wang, M. Chen, and X. Li, "Research and design of constraint-based travel recommendation system," *Computer Technology and Development*, pp. 83–125, 2012.
- [3] W. Zhang, L. Han, D. Zhang, P. Ren, J. Ma, and Z. Chen, "Geopmf: a distance-aware tour recommendation model," *Journal of Computer Research & Development*, 2017.
- [4] Q. Wu, Y. Zhou, D. Wen, and Z. He, "Personalized trip itinerary recommendation based on user interests and points of interest popularity," *Journal of Computer Applications*, 2016.
- [5] Q. Li, W. Xu, and H. Jiang, "Personalized recommendation for traveling planning based on online word-of-mouth," *Business Review*, vol. 28, no. 6, pp. 113–118, 2016.
- [6] B. Hao, Y. Tan, W. Wang, and J. Jia, "Intellegent tourism route planning system based on ant colony optimization," *Journal of Jinggangshan University (Natural Science)*, no. 1, pp. 8–13, 2015.
- [7] "New york city travel guide," Trip Advisor. [Online]. Available: [https://www.tripadvisor.com/Travel\\_Guide-g60763-New\\_York\\_City\\_New\\_York.html](https://www.tripadvisor.com/Travel_Guide-g60763-New_York_City_New_York.html)
- [8] "Places api web service," Google. [Online]. Available: <https://maps.google.com>

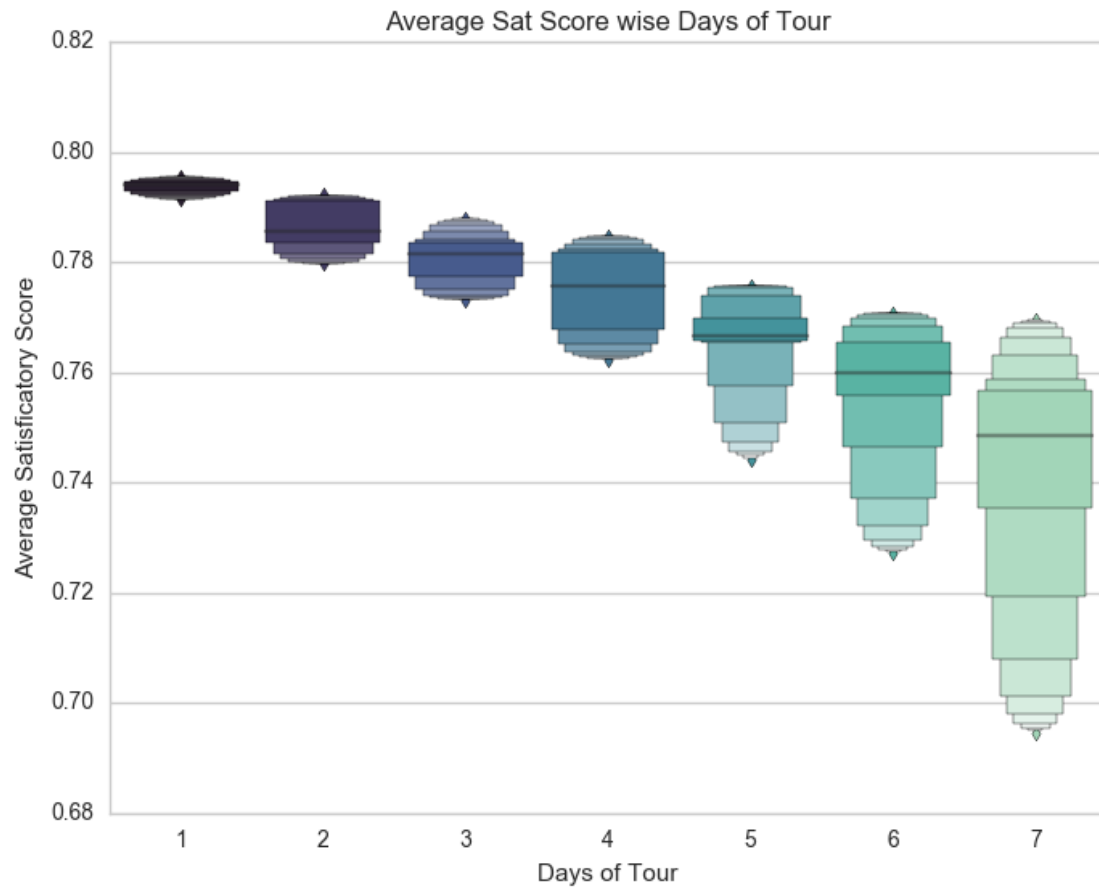


Figure 8: *Sat*-tour duration relationship

## Get FREE from Tour Planning!

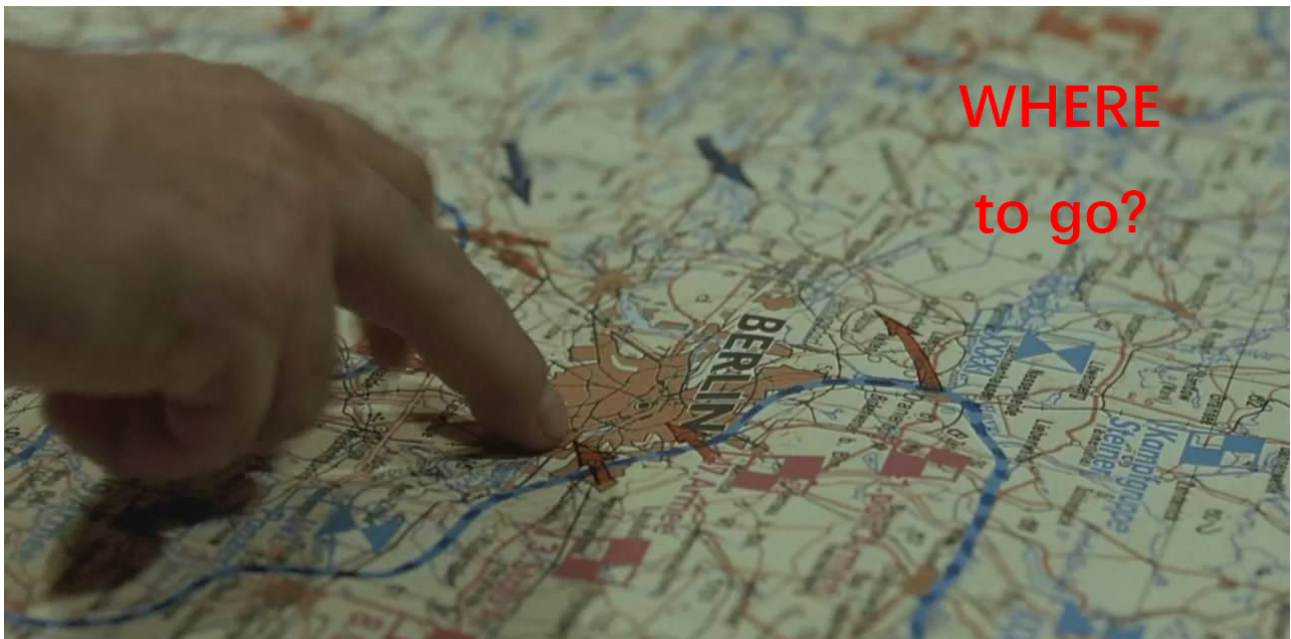
Want to explore a new city? Hard to decide where to go?

Ooooh, a tour should be relaxed. Don't let tour planning burden you!

Why not experience the convenience *Where to GO* provides!

Personalized tour planning, a touch and all done.





## Appendices