Tourism Analytics: A Recommendation Engine for Itinerary Planning

A. Downward[1], C. Chen[1], S. Chowdhury[2]
[1]Department of Engineering Science
[2]Department of Civil and Environmental Engineering
University of Auckland
New Zealand
a.downward@aucklanduni.ac.nz

Abstract

Tourism is one of New Zealand's largest export industries, surpassing the dairy industry in 2016. In the year to March 2016, total tourism expenditure was $34.7 billion. Despite this significant revenue, planning for tourists is not an easy task, since information about tourism activities is not centralised. Moreover, information about tourists themselves is not available for tourism providers. Thus the potential benefits of tourism are not being maximized.

In this project we seek to address these issues by creating a planning tool for tourists, accounting for both individual preferences and demographic information. This tool will provide an optimised itinerary which maximises the enjoyment of the tourist, subject to their time and budgetary restrictions. This is modelled as a variant of a vehicle routing problem.

The additional benefit of this tool is that it can become a source of real-time information on tourists and provide the tourism sector with detailed information on the types of tourists visiting New Zealand.

Key words: tourism, mixed-integer programming, recommendation engine.

1 Introduction

In this section we will provide some context around the current state of the tourism sector in New Zealand and discuss several opportunities for using data and optimisation to improve outcomes within the sector. This may lead to strategies for growth whereby both the tourism sector and the tourists have improved information upon which to base their decisions.

1.1 Background

The tourism industry in New Zealand is currently experiencing significant growth, with international visitor numbers having increased by 11% and international visitor expenditure by 18% in the last year alone (Ministry for Business, Innovation and Employment 2016). In fact, in 2016 tourism became New Zealand's largest export sector, overtaking the dairy industry (Cook, F. 2016).

With this sector's current growth, issues are developing in several of the main tourism hotspots (e.g. Rotorua and Queenstown), where, at the height of the tourism season, accommodation is extremely tight (Bradley, G. 2016). If we simply consider this problem on a local scale, the economically rational response may be to build more accommodation in these regions - however, any additional accommodation built to meet the demand in the peak season, may lead to underutilisation in the off-season. For this reason, a better solution may be attainable if we consider this issue from the perspective of New Zealand as a whole. There are many other towns and regions which would benefit from additional tourism in the peak season, and being able to more evenly tourists over the country will additionally help to develop the economies of these regions.

In standard economic theory for an efficient market, price signals ought to deliver efficient behaviour. For example, if accommodation is scarce then prices rise, and demand is curtailed. However, in a market with incomplete information, these same price signals may lead to inefficient behaviour. This is because tourists may reduce their time in New Zealand or not visit at all due to the high prices in specific locations, when, in fact, due to a lack of information, other cheaper regions were not considered (even though they may have delivered equally enjoyable experiences).

1.2 Opportunities

In this work, we seek to develop a model which is designed to be the back-end of a planning tool for tourists. This tool will bridge the gap between the tourism sector and tourists, providing a two-way flow of information.

For the tourists, they will be able to access information about tourism activities throughout the country, and be provided with daily itineraries that account for the tourists' preferences and budgets, while also considering transportation requirements. This not only benefits the individual tourists, but also the sector as a whole, due to the increased enjoyment and positive word of mouth from the tourists.

The other main benefit of such a tool is in its ability to collect data. This tool will be tailored to the individual tourist, and thus will need to collect demographic information as well as encourage tourists to rate the activities they've undertaken. This kind of data will be extremely useful to the tourism sector, since it will provide a real-time view of the types of tourists who are visiting and what they enjoy. This can assist in guiding advertising campaigns as well as long-term planning decisions for the industry.

The key step to achieving the benefits stated above is creating an app that is easy to use and provides benefits to the tourists. There are two main advantages for the tourists: (1) providing user-tailored itineraries, and (2) making the planning of their trip easier. This paper will focus on how to construct a model that can achieve (1).

2 Itinerary Planning Model

In this section, we will outline a model for itinerary planning for tourists, accounting for individual preferences and budgetary constraints. Given a tourist's gender, age, interests, budget, and the number of days that they are visiting for, the model will output a set of daily itineraries that maximise the enjoyment of the tourist, while complying will all constraints. This problem is similar to a vehicle routing problem (Laporte 1992), in that we are seeking a set of non-overlapping subtours through a set of activity nodes, however, there are a number of side-constraints that make this problem more complex. We present the notation for this model below.

This model requires a database at the backend with information about

the activities that are available; for each activity, this database stores:

- GPS coordinates;
- entry price;
- opening / closing time;
- type of activity (e.g. culture, adventure, etc.);
- minimum and maximum duration; and
- utility as a function of duration.

The model also queries Google Maps to estimate travel times between activities. We will now present a mixed-integer program for a simplified version of the model where we allow only fixed durations and utilities for all activities. The full model allows for flexible durations, with corresponding utility functions for each activity (Chen, C. 2016).

Sets

$i/j \in \mathcal{A}$	Activities, indexed by i or j;
$(i,j) \in \mathcal{R}$	Routes (from activity i to activity j);
$k \in \mathcal{L}$	Layers (the kth layer pertains to the kth activity visited on a given day); and
$d \in \mathcal{D}$	Days, indexed by d.

Parameters

U_i	Utility acquired by a tourist who does activity i;
$\tau_{ij}^{\text{travel}}$	Travel time from activity i to activity j;
τ_i^{activity}	Time spent at activity i;
t_i^{open}	Opening time for activity i;
t_i^{close}	Closing time for activity i;
h_d^{start}	Tourist's start time on day d;
h_d^{finish}	Tourist's finish time on day d;
$c_{i,j}^{\text{travel}}$	Cost of travel from i to j;
c_i^{activity}	Cost of doing activity i;
B	Budget.

Variables

$y_{i,d}$	If the tourist does activity i on day d this is set to 1, otherwise it is 0;
$x_{ij,k,d}$	If activity j is the kth activity and follows activity i then this is set to 1, otherwise it is 0;
$t_{k,d}^{\text{arrive}}$	Time on day d when the tourist arrives at the kth activity; and
$t_{k,d}^{\text{depart}}$	Time on day d when the tourist leaves the kth activity.

Objective Function

The objective function of this model is to maximize the total utility accrued by the tourist.

$$\max \quad \sum_{d \in \mathcal{D}} \sum_{i \in \mathcal{A}} U_i \times y_{i,d}.$$

Constraints

This model has a number of constraints linking the activities undertaken to the order of activities within the subtours for each day. In addition, we must account for timings and other feasibility considerations. To simplify the presentation of the model we will deal with the constraints in blocks.

Logical Constraints We first note that the sets of variables x and y are binary.

$$\begin{aligned} &y_{i,d} \in \{0,1\}, && \forall i \in \mathcal{A},\ \forall d \in \mathcal{D}, \\ &x_{i,j,k,d} \in \{0,1\}, && \forall i \in \mathcal{A},\ \forall j \in \mathcal{A},\ \forall k \in \mathcal{L},\ \forall d \in \mathcal{D}. \end{aligned}$$

Linking activities to the subtour: Constraint (1) ensures that each activity is only undertaken at most once over the course of the trip (except for activity 1). Equation (2) ensures that if an activity is undertaken on a given day, it is represented within the subtour.

$$\sum_{d \in \mathcal{D}} y_{i,d} \leq 1, \qquad \forall i \in \mathcal{A} \backslash \{1\}, \tag{1}$$

$$\sum_{j \in \mathcal{A}} \sum_{k \in \mathcal{L}} x_{i,j,k,d} = y_{i,d}, \qquad \forall i \in \mathcal{A},\ \forall d \in \mathcal{D}. \tag{2}$$

Hotel / accommodation: We assume that the first and last locations are both activity 1 on each day; this could be the hotel the tourist is staying at. Equation (3) simply ensures that you must leave activity 1 in layer 1 on each day, and equation (4) ensures that you return to this location at some point. Since you can only leave each activity once per day, due to equation (2), this means that activity 1 will be the final destination each day.

$$\sum_{j \in \mathcal{A}} x_{1,j,1,d} = 1, \qquad \forall d \in \mathcal{D}, \tag{3}$$

$$\sum_{j \in \mathcal{A}} \sum_{k \in \mathcal{L}} x_{j,1,k,d} = 1, \qquad \forall d \in \mathcal{D}. \tag{4}$$

Subtour feasibility To ensure that we construct a single subtour for each day, we find a single path through the activities, which are replicated over multiple layers. Constraint (5) ensures that there is at most one arc within each layer. Constraint (6) ensures that in order for there to be an arc in the path from activity j to any other activity in layer k, there must be an arc from some activity to activity j in layer $k-1$. Note that constraints (3), (4), (5) and (6) together ensure that there will be a single subtour each day that must start and end at activity 1.

$$\sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{A}} x_{i,j,k,d} \leq 1, \qquad \forall k \in \mathcal{L},\ \forall d \in \mathcal{D}, \tag{5}$$

$$\sum_{i \in \mathcal{A}} x_{i,j,k-1,d} \geq \sum_{i \in \mathcal{A}} x_{j,i,k,d}, \qquad \forall j \in \mathcal{A},\ \forall k \in \mathcal{L} \backslash \{1\},\ \forall d \in \mathcal{D}. \tag{6}$$

Arrival and departure times In order to ensure that the subtours that we construct are feasible, and can be completed in the available time we must keep track of the arrival and departure times for each activity within the tour. To achieve this, the arrival and departure times correspond the position of the activity, k, in the subtour (i.e., what layer it pertains to). We first must initialise the arrival for the first activity to simply be equal to the tourist's start time for each day; this is achieved through equation (7). Constraint (8) ensures that you have arrived back at activity 1 before the finish time for each day.

$$t_{1,d}^{\text{arrive}} = h_d^{\text{start}}, \qquad \forall d \in \mathcal{D}, \tag{7}$$

$$t_{n,d}^{\text{arrive}} \leq h_d^{\text{finish}}, \qquad \forall d \in \mathcal{D}, \tag{8}$$

where n is the last element in $\mathcal{L}$.

Constraint (9) accounts for the travel time between activities, whereas equation (10) accounts for the time spent at an activity.

$$t_{k,d}^{\text{arrive}} \geq t_{k-1,d}^{\text{depart}} + \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{A}} \tau_{i,j}^{\text{travel}} x_{i,j,k-1,d}, \qquad \forall k \in \mathcal{L} \backslash \{1\},\ \forall d \in \mathcal{D}, \tag{9}$$

$$t_{k,d}^{\text{depart}} = t_{k,d}^{\text{arrive}} + \sum_{i \in A} \sum_{j \in A} \tau_{i}^{\text{activity}} x_{i,j,k,d}, \qquad \forall k \in \mathcal{L},\ \forall d \in \mathcal{D}. \tag{10}$$

We must also ensure that we take note of the opening and closing times of individual activities. These requirements are enforced through constraints (11) and (12).

$$t_{k,d}^{\text{arrive}} \geq \sum_{i \in A} \sum_{j \in A} t_{i}^{\text{open}} x_{i,j,k,d}, \qquad \forall k \in \mathcal{L} \backslash \{1\},\ \forall d \in \mathcal{D}, \tag{11}$$

$$t_{k,d}^{\text{depart}} \leq \sum_{i \in A} \sum_{j \in A} t_{j}^{\text{close}} x_{i,j,k,d}, \qquad \forall k \in \mathcal{L} \backslash \{1\},\ \forall d \in \mathcal{D}. \tag{12}$$

Budget Finally, we need to comply with the budget of the tourist; constraint (13) ensures that the total expenditure is within the budget that they specify.

$$\sum_{d \in \mathcal{D}} \sum_{k \in \mathcal{L}} \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{A}} x_{i,j,k,d} \left(c_{i,j}^{\text{travel}} + c_{i}^{\text{activity}} \right) \leq B. \tag{13}$$

In the next section we will detail how the enjoyment of activities can be personalised using a recommendation engine.

3 Recommendation Engine

In order to provide the model with the right input about the utility associated with different activities, we have designed a recommendation engine based on collaborative filtering. This initially updates the utility for activities based simply on demographic information, and preferences around the types of activities the tourist wishes to undertake. However, we also permit users to submit a rating (0–10) for each activity they do; this will help to further refine not only the recommendations they receive in the future, but also the recommendations for all other users who are deemed to be *similar* to them.

3.0.1 Collaborative Filtering

Implementing the collaborative filtering recommendation engine involves calculating similarity scores for each pair of users. This can then be used to compute updated utility scores for each activity. We consider a set of users $p\ in\mathcal{P}$, each having provided an associated vector $\mathbf{r}_p$ whose components are the demographic information of the user as well as ratings for undertaken activities. We can therefore compute a similarity score $\sigma_{p,q}$ for two users p and q, as follows:

$$\sigma_{p,q} = \frac{\mathbf{r}_p \cdot \mathbf{r}_q}{|\mathbf{r}_p|\,|\mathbf{r}_q|}.$$

This similarity score is essentially the cosine of the angle between the two vectors $\mathbf{r}_p$ and $\mathbf{r}_q$, so it will equal 1 if the vectors are identical, and 0 if the vectors are orthogonal. In the final step, this similarity score is used to provide an estimate of a target user's ratings for activities that they have not yet undertaken. Equation (14) estimates user p's rating for activity i, $\hat{r}_{p,i}$, given a set of precomputed similarity scores.

$$\hat{r}_{p,i} = \frac{\sum\limits_{q \in P} r_{q,i}\sigma_{p,q}}{\sum\limits_{q \in P} z_{q,i}\sigma_{p,q}}, \tag{14}$$

note that $z_{q,i} = 1$ if user q has rated activity i, and 0 otherwise; and $r_{q,i}$ is the rating that user q have given activity i. Essentially the estimated rating of activity i for user p is a weighted average of ratings from other users, with a higher rating applied to users that are deemed to be more similar.

Having such a recommendation engine in place gives value to the user in providing accurate feedback on their interests and ratings of specific activities. This means the engine with be continually updated with new data that can also assist the industry as a whole make better planning decisions to provide for the interests of the tourists.

4 Results

To demonstrate how this model works, we will show how the itineraries that are produced adapt to the individual tourist. We will consider a 22-year-old female tourist (A) and a 63-year-old male tourist (B), each based in Mt

Eden, and spending one day in Auckland (beginning at 10am and returning by 5pm). For each tourist, we will consider a budget of either $100 or $300 for the day and see how the itineraries adapt. Note the model we used to get these results is slightly more complex than that introduced in Section 2; here we allow for the durations spent at activities to be optimized, assuming that the marginal utility decreases with the duration (the details of this enhanced model can be found in Chen, C. 2016).

Tourist A was defined to like adventure tourism, with less emphasis on cultural or social activities. However, with a limited budget we see that the tourist is restricted in their options, first heading to One Tree Hill, then the Art Gallery for just over 2 hours, heading up to the top of Sky Tower before going to the Auckland Zoo for almost three hours. This route is depicted in Figure 1. For each figure, Mt Eden is marked in red, and the activities visited are labelled alphabetically from B to E; the grey circles denote the locations of all possible activities.

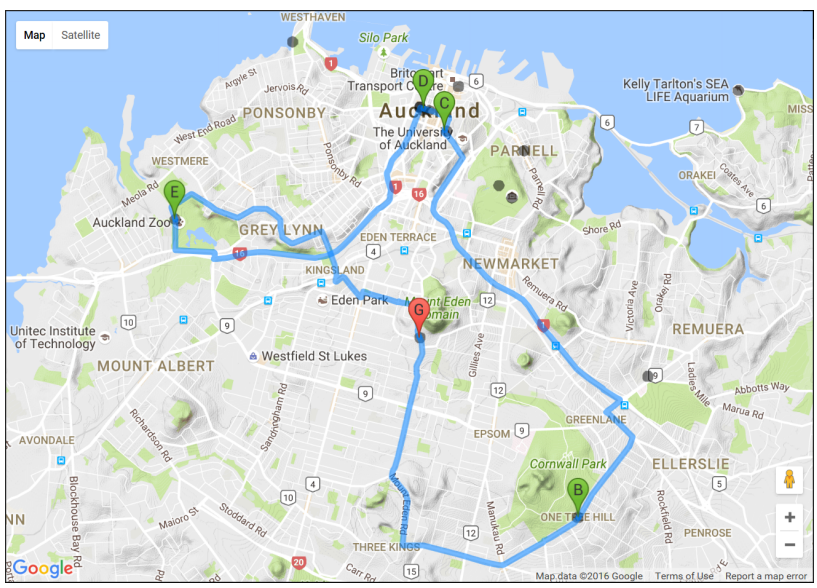

Figure 1: Optimal schedule for tourist A, with a budget of $100.

However, with a budget of $300 for the day, in the optimal itinerary tourist A heads to the Auckland Bridge Climb and Bungy (costing $160) and the Museum ($25), instead of One Tree Hill and Sky Tower, which were free; this route is shown in Figure 2.

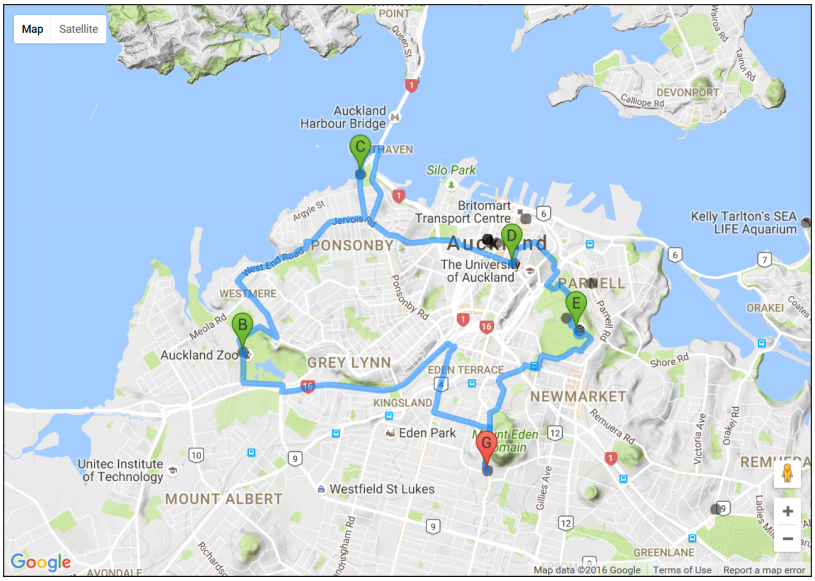

Figure 2: Optimal schedule for tourist A, with a budget of $300.

We can compare this with the itinerary for tourist B, who prefers cultural and social activities. This tourist with a $300 dollar budget first goes to the War Memorial Museum, then the Auckland Art Gallery, the Sky City Casino and finally the Sky Tower. The corresponding route is depicted in Figure 3.

We break down the specific details of the itinerary depicted in Figure 3 in Table 1 below.

Activity	Arrival Time	Departure Time	Duration
Mt Eden	–	10:00	–
War Memorial Museum	10:10	12:30	2h 20m
Auckland Art Gallery	12:40	15:00	2h 20m
Sky City Casino	15:20	16:20	1h
Sky Tower	16:20	16:40	20m
Mt Eden	17:00	–	–

Table 1: Optimal itinerary for Tourist B, with a budget of $300.

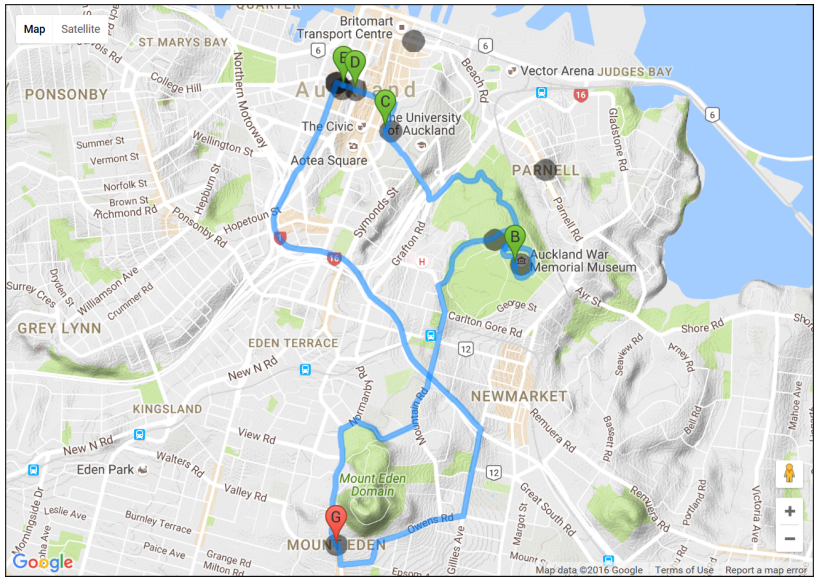

Figure 3: Optimal schedule for tourist B, with a budget of $300.

5 Conclusions

In this project we have developed a preliminary model which automatically constructs itineraries for tourists who are spending several days in a location. This model accounts for the interests of the tourist and provides feasible itineraries for each day. The prototype model has been written in Python, using PuLP as the modelling language, and solved using Gurobi; the SolverStudio plugin for Excel is used to manage the data.

This model will form the back-end of a smartphone app that both provides itineraries to tourists, and accepts feedback, ensuring that the data behind the model is constantly improving, and so too are the recommendations made. It is this two-way flow of information that makes this system valuable, not only to tourists, but to the tourism operators, who can leverage this information to make better decisions for their businesses.

This prototype provides a foundation upon which to incorporate additional features that can provide additional value to tourists. Specifically, we want to consider how we can plan itineraries for groups of tourists, who may not all share the same interests. Furthermore, since things do not always go to plan, the model should incorporate robustness. Finally, the impact of

weather on activities should be considered, and forecasts will be used to provide better itineraries, accounting for the weather. With these additions the model will become a multi-criteria, stochastic, mixed-integer programming problem. This model can then be expanded to deal with longer time periods where different amounts of time are spent in various towns and cities.

In terms of the recommendation engine, more work is required to tune the algorithm to ensure that the right weights are put on different components of the $\mathbf{r}$ vectors when determining the similarity score of different users.

With these enhancements, this tool will be extremely flexible and able to adapt to the individual preferences of users. The challenge will be in marketing this tool to foster a sufficient initial user base to provide reliable information to the tourism sector on which to base future investment decisions.

References

Bradley, G. 2016. "Hotel shortage critical as tourist boom rolls on." *New Zealand Herald*. http://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=11643974 (visited: 2016-10-28).

Chen, C. 2016. "A Tourism Planning Tool." *Engineering Science Part IV Projects*.

Cook, F. 2016. "NZ Tourism Overtakes Dairy as Highest Earner." *New Zealand Herald*. http://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=11736164 (visited: 2016-10-28).

Laporte, Gilbert. 1992. "The vehicle routing problem: An overview of exact and approximate algorithms." *European Journal of Operational Research* 59 (3): 345–358.

Ministry for Business, Innovation and Employment. 2016. "Key Tourism Statistics." Technical Report. http://www.mbie.govt.nz/info-services/sectors-industries/tourism/documents-image-library/key-tourism-statistics.pdf (visited: 2016-10-30).