

# Mapping of Tourism Destinations to Travel Behavioural Patterns

Mete Sertkan<sup>(✉)</sup>, Julia Neidhardt, and Hannes Werthner

E-Commerce Group, TU Wien, Vienna, Austria  
{mete.sertkan, julia.neidhardt, hannes.werthner}  
@ec.tuwien.ac.at

**Abstract.** Tourism is an information intensive domain, where recommender systems have become an essential tool to guide customers to the right products. However, they are facing major challenges, since tourism products are considered as complex and emotional. It has been shown that the seven-factor model is a legitimate way to counter some of these challenges. However, in order to recommend an item, it has also to be described in terms of this model. This work's aim is to find a scalable way to map tourism destinations, defined by their attributes, to the seven-factor model. Through statistical analysis and learning methods it is shown that there is a significant relationship between particular destination features and the seven-factors and that destinations can be grouped in a meaningful way using their attributes.

**Keywords:** User modelling · Tourism recommender systems  
Tourism destinations · Statistical analysis · Cluster analysis  
Seven-factor model

## 1 Introduction

The rapid development of Information & Communication Technologies (ICT), and especially the emergence of the World Wide Web, lead to fundamental changes in the tourism industry, both on the supply and the demand side. Consumers now have ubiquitous access to vast amounts of information at a very low cost, allowing them to compose and compare products and optimize their trips. However, increasing cognitive costs to process the amount and variety of information could lead to the problem of information overload. This shows the necessity of new techniques and tools to analyse, categorize and visualize information in a proper way (Hwang, Gretzel, Xiang, & Fesenmaier, 2006). On the other side, the Web also allows a massive “informatization” of the whole tourism value chain, resulting in many novel value-generating strategies, to satisfy new consumer needs (Werthner & Ricci, 2004; Werthner et al., 2015).

According to a recent study (Ipsos MediaCT, 2014) people rely on online sources such as social media, photo- and video-sites and search engines to get inspired where to go or how to travel. It is shown that 65% of the leisure travellers start researching online before a travel decision. Particularly in this early phase of decision making a considerable amount of people has difficulties to explicitly express their preferences and needs (Zins, 2007). Recommender Systems (RSs) are facilitating this

decision-making. Ricci, Rokach, and Shapira (2015) define RSs as “software tools and techniques providing users with suggestions for items a user may wish to utilize”. Particularly, profiling and personalization might help in such cases, where preferences and needs are unknown or hard to express. Especially in tourism this is a big challenge, since tourism products are considered as very complex (i.e., they typically combine accommodation, transportation, activities, food, etc.), mostly intangible and highly associated with emotional experiences (Werthner & Ricci, 2004).

Neidhardt, Seyfang, Schuster and Werthner (2014, 2015) introduced a picture based approach to elicit the preferences of a user and a seven-factor model to capture the respective user’s profile within a travel recommender system. These factors form the basis of a seven-dimensional vector space and are referring to travel behavioural patterns summarized as *Sun & Chill-Out*, *Knowledge & Travel*, *Independence & History*, *Culture & Indulgence*, *Social & Sport*, *Action & Fun*, and *Nature & Recreation*. Based on a picture-selection process, a user’s profile is determined. This profile comprises a score for each of the factors and thus can be seen as a point in the seven-dimensional vector space. In order to provide recommendations to a user, those items have to be determined that are closest to him/her. Thus, also the items have to be mapped into the vector space, i.e., represented with respect to the travel behavioural patterns. In order to build up a reasonable recommendation base more than 10,000 tourism products were initially mapped manually by experts. Obviously, this approach does not scale.

The aim of this work is to find an automated way of relating tourism products to travel behavioural patterns. In particular, the relationships between the seven-factors and attributes of destinations will be examined in order to map the destinations onto the seven-factors and to group similar destinations for a better understanding and generalization.

The rest of the paper is organized as follows. In Sect. 2 the state of the art is presented, focusing on tourist roles, the seven-factor model, and tourism recommender systems. In Sect. 3, the data sample is described. In Sect. 4 regression analyses based on expert mappings are conducted and discussed. In Sect. 5 a cluster analysis is conducted and presented. In Sect. 6 the main insights are summarized and conclusions are drawn.

## 2 State-of-the-Art

### 2.1 Tourist Roles & The Seven-Factor Model

Much research has been conducted in order to identify and categorize tourist roles, describing the relation between a person’s travel behaviour and his or her preferences, interest, and needs. In this context, Gibson and Yiannakis (2002) introduced a well-established classification framework, distinguishing 17 different tourist roles, capturing short-term behaviour. Gretzel, Mitsche, Hwang, and Fesenmaier (2006) demonstrated that tourist roles can be used in order to recommend touristic activities and, in turn, destinations. It has also been shown that tourist roles can be related to personality traits. Delić, Neidhardt and Werthner (2016) are providing significant evidence to the relation between the well-established Big-Five personality traits

(Goldberg, 1990) and the 17 tourist roles (Gibson & Yiannakis, 2002). Personality traits tend to be stable over time and can be considered as long-term preferences of a person (Woszczyński, Roth, & Segars, 2002). Taking into account both the Big-Five personality traits (long-term behaviour) and 17 tourist roles (short-term behaviour) Neidhardt et al. (2014, 2015) developed a seven-factor model by conducting factor analysis. Each factor is referring to a travel behavioural pattern, which are summarized in Table 1. These factors are easier to process computationally as well as cognitively compared to the original 22 dimensions. Neidhardt and Werthner (2017) showed that based on different demographic characteristics different user groups can be well distinguished within this model.

**Table 1.** Seven-factor model

Factor	Description
<i>Sun &amp; Chill-Out</i>	A neurotic sun lover, who likes warm weather and sun bathing and does not like cold, rainy or crowded places
<i>Knowledge &amp; Travel</i>	An open minded, educational and well-organized mass tourist, who likes traveling in groups and gaining knowledge, rather than being lazy
<i>Independence &amp; History</i>	An independent mass tourist, who is searching for the meaning of life, is interested in history and tradition, and likes to travel independently, rather than organized tours and travels
<i>Culture &amp; Indulgence</i>	An extroverted, culture and history loving high-class tourist, who is also a connoisseur of good food and wine
<i>Social &amp; Sports</i>	An open minded sportive traveller, who loves to socialize with locals and does not likes areas of intense tourism
<i>Action &amp; Fun</i>	A jet setting thrill seeker, who loves action, party, and exclusiveness and avoids quiet and peaceful places
<i>Nature &amp; Recreation</i>	A nature and silence lover, who wants to escape from everyday life and avoids crowded places and large cities

## 2.2 Tourism Recommender Systems

The relationship between tourism and ICT is described as a symbiosis by Gretzel (2011) and travel & tourism have always been major application domains for Web-related services (Werthner & Klein, 1999). As the amount of information on the Web started to rise the call for techniques to cope with information overload began to grow. One answer to that are RSs. There are different types of recommendation techniques and hybrid solutions, but according to Burke and Ramezani (2011) the most appropriate ones in the matter of tourism are either content-based (recommending items similar to the ones the user already liked in the past) and/or knowledge-based (recommending items using domain knowledge about user preferences and needs and utility of items to the user). This work is based on a picture-based approach to RSs (Neidhardt et al. 2014, 2015), which can be considered as a content- and knowledge-based approach. Preferences and needs of a user are determined via a simple picture selection process. In this way, the so-called cold-start problem of learning-based techniques (collaborative, content-based, and demographic) (Burke, 2007) as well as

tedious questionnaires for preference elicitation are avoided. Thus, the users are addressed on an emotional, implicit level and do not have to state their preferences explicitly. According to Garcia, Sebastia, and Onaindia (2011) tourism RSs can be distinguished into two types: one focusing on destination selection the other on activities that can be performed at a certain destination. This work belongs to the first mentioned group, in contrast to Neidhardt et al., (2014, 2015), where the focus lies on the Point of Interests (POIs), e.g., activities, events, restaurants, sights. Much research has already been conducted targeting destination recommender systems (Fesenmaier, Werthner, & Wöber, 2006; Borràs, Moreno, & Valls, 2014), but most of them are mainly focused on a distinct region or POIs in a destination. There are few that are focussed on personality traits and motifs of a user (see for example Braunhofer, Elahi, & Ricci, 2014).

### 3 The Data Sample

The data is provided as a SQL-dump by a German eTourism company<sup>1</sup> and consists of more than 30,000 destinations all around the world. Almost all countries are represented in the database, but the majority of destinations are located in the USA, Germany, France, Italy, Spain, Great Britain, Austria, Greece, Switzerland, and Sweden (65%). Beirman (2003) defines tourism destinations as “a country, state, region, city or town which is marketed or markets itself as a place for tourists to visit”. In this work destinations are defined in a similar way, except the range goes from a hamlet (population <100) to a metropolis (population >1M). Destinations are described via destination features, which can be separated into two groups:

1. **Motivational ratings** in an interval of 0–1, describing the degree of appropriation for a particular motif. Following 27 motivational ratings are listed: *nightlife, wellness, shopping, nature & landscape, image & flair, culture, sightseeing, entertainment, mobility, price level, accommodations, gastronomy, beach & swimming, golf, scuba diving, kite & windsurfing, hiking, cycling, horseback riding, winter sports, sports, family, quietness, surfing, sailing, gays, and mountain biking*. The motivational ratings are determined considering factors such as infrastructure, climate, geographical nature, user opinions, number of services, image, and marketing.
2. **Geographical attributes** are presented in binary format, describing the presence or absence of a particular attribute. Following 14 attributes are listed: *sea, mountain, lake, island, sandy beach, metropolis, forest, river, desert, old town, pebble beach, sand & pebble beach, hill, swamp, volcano, fjord, flat decaying sand beach, beach promenade, wine-growing, heath, health resort, and winter sports resort*.

For this study, 561 destinations were chosen randomly and mapped manually by experts by assigning a score for each factor using the scale 0–0.25–0.50–0.75–1. The higher the score the more suitable the destination for that specific factor. The consulted

---

<sup>1</sup> More details and the data are provided upon request.

experts are members of an Austrian eTourism company that has already been using an implementation of the picture based approach introduced by Neidhardt et al. (2014, 2015). Hence, they have expertise in both the seven-factor model and the travel destinations. Three experts analyzed 561 destinations first individually and then determined a final mapping in a joint-discussion. Table 2 shows the resulting distributions.

**Table 2.** Factor distribution of the expert mapping

	% of destinations scored with				
	0	0.25	0.50	0.75	1
<i>Sun &amp; Chill-Out</i>	39.93	19.07	13.90	8.74	18.36
<i>Knowledge &amp; Travel</i>	36.72	29.41	17.65	9.63	6.59
<i>Independence &amp; History</i>	13.90	24.60	35.29	18.90	7.31
<i>Culture &amp; Indulgence</i>	28.70	21.75	23.53	13.19	12.83
<i>Social &amp; Sport</i>	1.78	8.20	53.83	32.09	4.10
<i>Action &amp; Travel</i>	50.09	21.03	12.66	6.77	9.45
<i>Nature &amp; Recreation</i>	16.58	15.15	22.82	25.85	19.61

For example, in case of the factor *Sun & Chill-Out* 39.93% of the destinations scores were with 0, 19.07% with 0.25, 13.9% with 0.5, 8.74% with 0.75, and 18.36% with 1. A similar distribution can also be observed for factors *Knowledge & Travel*, *Culture & Indulgence*, and *Action & Fun*, where the majority of destinations either score with 0 or 0.25. On the other hand, for the factors *Independence & History* and *Social & Sports* most of the destinations score with 0.5 and only very few with 0 or 1. The only factor where the differences are not that pronounced is *Nature & Recreation*.

#### 4 Mapping of Destination Features to Seven Factors

The aim of the work is not only to project destinations into the seven-dimensional vector space of travel behavioural patterns using their features, but more importantly to understand the relationship between the seven-factors and destination features. In James, Witten, Hastie, and Tibshirani (2013a) it is suggested to choose linear models over more complex ones if inference and interpretability is the goal. Taking this into account, a multiple linear regression model (James, Witten, Hastie, & Tibshirani, 2013b) with step-wise variable selection (James, Witten, Hastie, & Tibshirani, 2013c) is applied. All seven factors are considered as independent from each other, since they are obtained from factor analysis. Therefore, they can be treated separately by fitting a model for each travel behavioural pattern, which takes the features of a destination as input and returns the factor score (0–1) as output. The data sample is split into a training and test set in a ratio of 80/20. Model performance is assessed by  $R^2$  (proportion of variance explained) and root mean square error (standard deviation of the residuals/prediction errors) of training ( $RMSE_{train}$ ) and test set ( $RMSE_{test}$ ).

**Table 3.** Performance of the resulting multiple linear regression models

	R <sup>2</sup>	RMSE <sub>train</sub>	RMSE <sub>test</sub>	F-statistic	Signif.
<i>Sun &amp; Chill-Out</i>	0.70	0.21	0.25	148.6	***
<i>Knowledge &amp; Travel</i>	0.66	0.18	0.20	145.6	***
<i>Independence &amp; History</i>	0.56	0.19	0.20	98.7	***
<i>Culture &amp; Indulgence</i>	0.58	0.22	0.25	108.1	***
<i>Social &amp; Sports</i>	0.19	0.16	0.18	23.1	***
<i>Action &amp; Fun</i>	0.76	0.16	0.19	121.7	***
<i>Nature &amp; Recreation</i>	0.52	0.20	0.23	105.0	***

Table 3 shows the performance of each model. There is a statistically significant relationship between each travel behavioural pattern and destination features (used in the corresponding models), with  $p < 0.001$  (\*\*\*). RMSE<sub>train</sub> and RMSE<sub>test</sub> are close, indicating that the resulting models will perform similar out of sample. Overall, all travel behavioural patterns are well described by the resulting models, except *Social & Sports*, where only 19% of the variance is explained. This is caused by the uneven distribution of the expert mapping of *Social & Sports*, where 53.83% of the destination scored with 0.5 and only 1.78% scored with 0 and 4.10% with 1 respectively. Nevertheless, there is significant evidence of a relation between destination features and the factor *Social & Sports*. *Sun & Chill-Out* and *Action-Fun* have the best fitted models, where 70 and 76% of the variance can be explained.

The resulting multiple linear regression models comprise both motivational ratings and geographical attributes. After the variable selection 18 out of 27 motivational ratings and 7 out of 14 geographical attributes in total are used. Table 4 summarizes the outcomes of the regression analysis. Motivational ratings *sightseeing*, *culture*, *entertainment*, *family*, *quietness*, *gastronomy*, and *image & flair* appear in more than one model. Also, geographical attributes *sea*, *health resort*, and *winter sports resort* are in several models.

**Model A—*Sun & Chill-Out*.** The geographical attributes *sea*, *health resort*, and

**Table 4.** Results of the regression analysis

Coefficients	Model A	Model B	Model C	Model D	Model E	Model F	Model G
(Intercept)	0.23***	0.07***	0.27***	0.11***	0.46***	-0.03*	0.72***
<i>Sightseeing</i>	-0.23***	0.43***	0.18***	0.20***	-0.14***	0.12***	-0.17***
<i>Sea</i>	0.43***	-0.09***	-0.12***	-0.07**	-	0.14***	-
<i>Culture</i>	-	0.28***	0.45***	0.38***	-	-	-
<i>Entertainment</i>	-	0.33***	-	-	-0.21*	0.63***	-
<i>Family</i>	-	-0.16***	-	-0.14**	-	-0.36***	-
<i>Quietness</i>	-	-	-0.09***	-	-	-0.11***	0.44***
<i>Gastronomy</i>	-	-	0.14**	0.28***	-	-	-0.35***
<i>Image &amp; flair</i>	-	-	-	0.17***	-	-	-0.09**
<i>Health resort</i>	0.27***	-	-	-	-	-	0.14**
<i>Winter sports resort</i>	-	-	-0.14**	-	-	0.14*	-
<i>Beach &amp; swim</i>	0.30***	-	-	-	-	-	-
<i>Nature &amp; landscape</i>	0.15***	-	-	-	-	-	-
<i>Nightlife</i>	-0.25***	-	-	-	-	-	-
<i>Sports</i>	-	-	-	-	0.43***	-	-
<i>Hiking</i>	-	-	-	-	0.08***	-	-
<i>Winter sports</i>	-	-	-	-	-	0.35***	-
<i>Sailing</i>	-	-	-	-	-	0.18***	-
<i>Mobility</i>	-	-	-	-	-	0.16**	-
<i>Mountain biking</i>	-	-	-	-	-	0.08**	-
<i>Golf</i>	-	-	-	-	-	0.11**	-
<i>Shopping</i>	-	-	-	-	-	-	-0.34***
<i>Lake</i>	0.18***	-	-	-	-	-	-
<i>Old town</i>	-	-	-	-	-0.09*	-	-
<i>Metropolis</i>	-	-	-	-	-	0.17***	-
<i>Mountains</i>	-	-	-	-	-	-	0.14***

Note Significance level is coded as follows: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

*lake* and motivational ratings *beach & swim* and *nature & landscape* are significantly, positively related to this factor. Those features can be interpreted as indicators for relaxation and indirectly also for sun. On the other side, motivational ratings *sightseeing* and *nightlife* have a significant negative impact, which can be associated with crowded places and mass tourism.

**Model B—Knowledge & Travel.** The motivational ratings *sightseeing*, *culture*, and *entertainment* are significantly, positively related to this factor. Those features can be seen as the main motivation of travellers with the ambition to learn something on their trip. On the contrary, motivational rating *family* and geographical attribute *sea* are significantly, negatively associated with this factor. The negative impact of motivational rating *family* can be explained by the fact that most destinations in the data sample, which have an increased motivational rating *family* (>0.5), show typical characteristics of recreational travel destinations. Namely, they are quietly situated and

in the nature. The majority of destinations in the data sample, which are located at the *sea*, have sea resort characteristics, where cultural and educational interests are short comings. This explains the negative sign of the geographical attribute *sea*.

**Model C—Independence & History.** The motivational ratings *culture*, *sightseeing*, and *gastronomy* are significantly, positively related to this factor. Those features can be seen as the main motivation of travellers with interests in history and tradition. Whereas motivational rating *quietness* and geographical attributes *sea* and *winter sports resort* are significantly, negatively related. The negative impact of the geographical attribute *winter sports resort* is not surprising. The majority of destinations in the database, which are located in a *winter sport resort*, are mostly dedicated to winter sports and après-ski only. Thus, interests of an independent traveller with passion for history and tradition is short coming. The negative impact of the geographical attribute *sea* can be interpreted in the same way as for the factor *Knowledge & Travel* (see Model B—*Knowledge & Travel*). The motivational rating *quietness* is an indicator for low population and tourism density, where high rated places are considered more appropriate for recreational travellers than for people with interest in history and tradition. By this means, the negative sign can be explained.

**Model D—Culture & Indulgence.** The motivational ratings *culture*, *sightseeing*, *gastronomy* and *image & flair* are significantly, positively related to this factor. Those ratings can be interpreted as the main motivation of a culture and history interested high class tourist, who is also a connoisseur of good food and wine. On the other hand, the motivational rating *family* and the geographical attribute *sea* have a significant negative impact on the factor, which has the same cause as in *Knowledge & Travel* (see Model B—*Knowledge & Travel*).

**Model E—Social & Sports.** The motivational ratings *sports* and *hiking* are significantly, positively related to this factor, which is obvious. Whereas, the motivational ratings *entertainment* and *sightseeing* and the geographical attribute *old town* are significantly, negatively related to the factor. Those features can be seen as indicators for crowded places and mass tourism.

**Model F—Action & Fun.** The geographical attributes *metropolis*, *sea*, and *winter sports resort* and motivational ratings *winter sports*, *sailing*, *entertainment*, *mobility*, *sightseeing*, *mountain biking*, and *golf* are significantly, positively related to this factor. *Metropolis*, *sea*, *entertainment*, *mobility*, and *sightseeing* can be seen as indicators for vibrant places with party and fun. Whereas *winter sports*, *mountain biking*, and *winter sports resort* can be interpreted as motivators for thrill seeking and après-ski loving tourists. Finally, *golf* and *sailing* can be interpreted as indicators for exclusiveness and VIP. On the other hand, the motivational ratings *family* and *quietness* have a negative impact on the factor *Action & Fun*, which is not surprising and self-explanatory.

**Model G—Nature & Recreation.** The motivational rating *quietness* and geographical attributes *mountains* and *health resort* are significantly, positively related to this factor. Those features can be seen as the main interests of tourist, whose goal is recreation, finding silence and peace, and escaping from everyday life. On the other hand, the motivational ratings *gastronomy*, *sightseeing*, *shopping*, and *image & flair* have a significant negative impact on this factor. Those features can be interpreted as indicators for crowded places and city life.



## 5 Clustering

Identifying conceptually meaningful groups of destinations with shared common characteristics will help to further understand the data and its structure, which may contribute to a more generalized solution. Prototype-based, partitional clustering techniques are considered, where the most prominent ones are k-means and k-medoids. Since the data comprises binary attributes, centroids are not meaningful. Therefore, k-medoids is applied. A medoid corresponds per definition to an actual data point, which is considered as the most representative point for the cluster (Pang-Ning, Steinbach, & Kumar, 2006). Specifically, Partitioning Around Medoids (PAM) (Kaufman & Rousseeuw, 1990), the most common k-medoids algorithm, is used. Since the data consists of two different data types, i.e., binary (geographical attributes) and continuous (motivational ratings), the Gower distance (appropriate for mixed data types) (Gower, 1971) is used as distance metric. In order to find an appropriate number of clusters, the internal evaluation metric silhouette width (Rousseeuw, 1987) is used for assessment. Based on the silhouette width 4, 6 and 8 cluster solutions are considered, but for the sake of interpretability a 6-cluster solution is chosen. In Table 5 average factor scores and corresponding standard deviations (SD) for each cluster are listed.

**Table 5.** Average factor scores (and SD) in different clusters

	C1	C2	C3	C4	C5	C6
<i>Sun &amp; Chill-Out</i>	0.17 (0.26)	0.15 (0.19)	0.28 (0.31)	0.29 (0.33)	0.76 (0.27)	0.90 (0.20)
<i>Knowledge &amp; Travel</i>	0.19 (0.24)	0.88 (0.19)	0.17 (0.21)	0.41 (0.27)	0.43 (0.24)	0.25 (0.23)
<i>Independence &amp; History</i>	0.38 (0.26)	0.88 (0.13)	0.34 (0.25)	0.57 (0.24)	0.54 (0.19)	0.43 (0.22)
<i>Culture &amp; Indulgence</i>	0.27 (0.30)	0.93 (0.17)	0.25 (0.27)	0.57 (0.29)	0.56 (0.27)	0.41 (0.25)
<i>Social &amp; Sport</i>	0.52 (0.20)	0.52 (0.17)	0.60 (0.17)	0.62 (0.19)	0.64 (0.18)	0.56 (0.18)
<i>Action &amp; Fun</i>	0.10 (0.21)	0.84 (0.28)	0.08 (0.16)	0.31 (0.24)	0.62 (0.31)	0.31 (0.24)
<i>Nature &amp; Recreation</i>	0.48 (0.34)	0.06 (0.16)	0.80 (0.19)	0.57 (0.27)	0.36 (0.27)	0.65 (0.22)
Cluster Size	184	47	141	57	60	72

Destinations in C1 are more or less suitable for *Social & Sports* and *Nature & Recreation*. All other factors have a low average score. Liederbach (Hessen, Germany) is in the centre of C1 and the most representative observation. It is a small, calm suburb with almost no tourism destination characteristics, but just a few offers of sport activities.

Destinations in C2 are very well suited for *Knowledge & Travel*, *Independence & History*, *Culture & Indulgence* and *Action & Fun*. However, tourists with high *Sun & Chill-Out* or *Nature & Recreation* score will avoid destinations in C2. The most representative data point of C2 is Brussels, the capital of Belgium. The centre of C2 is a vibrant metropole, with many opportunities for nightlife, party, culture, sightseeing, and gastronomy. Recreation, relaxation and peace are alien to the medoid of C2.

Destinations in C3 are very well suited for *Nature & Recreation* and they also show an increased average score in factor *Social & Sports*. Also noteworthy is that people with high *Action & Fun* score, will avoid destinations in C3 at all. The centre of the cluster is Schönberg am Kamp (Lower Austria). It is a small, peaceful town in the nature good for hiking, cycling, escaping from everyday life, and recreational tourism.

Destinations in C4 are more or less suitable for *Knowledge & Travel*, *Independence & History*, *Social & Sports*, and *Nature & Recreation*. Overall, there is no dominant factor. The medoid of C4 is Todtnau (Baden-Württemberg, Germany). It is a rural town, where 60% is covered with forest. Todtnau is good for hiking, cycling, mountain biking, and other sports, but also for recreational tourism. Nevertheless, it also offers cultural activities, gastronomy, nightlife and entertainment in a moderate level. Noteworthy to mention is that the effect of both recreation and sports indicators and culture, history, tradition and entertainment indicators are dampening each other. Hence, there is no pronounced differences in the factor scores and none of them are either satisfied or unsatisfied.

Destination in C5 have on average a high score in *Sun & Chill-Out*, an increased score in *Social & Sports* and *Action & Fun*, and a moderate level of *Knowledge & Travel*, *Independence & History*, and *Culture & Indulgence*. *Nature & Recreation* is the only factor, which can be considered as low. The centre of C5 is Grand Baie on the island of Mauritius. It is considered as a centre of beach tourism on the northern coast of the island. Besides beach tourism there are plenty of opportunities for nightlife, entertainment, gastronomy, shopping, and water sports.

Destinations in C6 are a perfect match for *Sun & Chill-Out*. Additionally, *Nature & Recreation* and *Social & Sports* have an increased score on average. The medoid of C6 is Anaxos, a small, intimate, and tranquil seaside resort, which is located on the island of Lesbos (Greece). Beside sun, beach, and recreation there are opportunities for (water) sports and get-together with locals.

In summary, it can be said that there is an underlying natural structure of the data. Thus, six conceptually meaningful groups of destinations could be identified. For a better understanding, these groups or clusters can be simplified and summarized as follows: C1—*passionless suburb*, C2—*energetic city*, C3—*peaceful village*, C4—*ordinary town*, C5—*vibrant beach resort*, C6—*tranquil seaside resort*. The identified groups can help to better understand the (dis)similarities among the destinations and may also contribute to a more generalized solution of mapping without the need of prior information in contrast to the supervised method in the previous section. Nevertheless, considering the outcomes of the regression analysis one can clearly see relationships between the used destination features and the most determinant features of the cluster solution. Especially, destination features that appear in more than one regression model are the most decisive characteristics of cluster cohesion (separation) in the presented model.

## 6 Conclusions

Primarily, this work's aim is to identify and explain associations between destination attributes and the seven-factor model to enable an automated mapping of destinations onto the seven factors. To do so, a multiple linear regression analysis with step wise variable selection was conducted. Seven models were established, one for each of the seven factors. The resulting models are providing strong evidence that there is a significant relation between selected destination features and the factors. Overall, all travel behavioural patterns are well described (52–76% of the variance) by the resulting models, except *Social & Sports*, where only 19% of the variance can be explained. This is caused by an uneven distribution of the sample. A statically sounder sample will be targeted in future work. A linear model has the benefit of interpretability, but might suffer in performance. Additionally, the chosen variable selection model reveals core attributes, but is greedy (relevant information might get lost). In this case a performance evaluation of different methods (principal component regression, ridge regression etc.) is planned. Furthermore, a cluster analysis was conducted in order to determine if there is an underlying natural structure of the data sample. Six conceptually meaningful groups were identified (*passionless suburb, energetic city, peaceful village, ordinary town, vibrant beach resort, tranquil seaside resort*). Those clusters can foster a better understanding of the (dis)similarities among destinations and also of the relation between destination features and the seven-factor model. Further, these clusters can be used for more accurate recommendations or can be targeted directly by a RSs. In future work, it is also planned to recommend accommodations in a second step. Thus, an analysis revealing the relation between hotel features and the seven factors is needed. However, this aim immediately shows a disadvantage of the followed approach, namely data source dependency. To counter this problem there is an ongoing study, which aims to build up a comprehensive data model of tourism products. This data model will serve as an “intermediary” layer between the respective data source and the seven-factor model and can therefore be used to harmonize heterogeneous sources of data (e.g., by mapping different sources of destination data onto this layer).

## References

- Beirman, D.: Defining tourism destination in crisis. In: *Restoring Tourism Destinations in Crisis: A Strategic Marketing Approach*, pp. 3–17. CAB International (2003)
- Borràs, J., Moreno, A., Valls, A.: Intelligent tourism recommender systems: a survey. *Expert Syst. Appl.* **41**(16), 7370–7389 (2014)
- Braunhofer, M., Elahi, M., Ricci, F.: Usability assessment of a context-aware and personality-based mobile recommender system. In: *International Conference on Electronic Commerce and Web Technologies*, pp. 77–88. Springer, Cham (2014, September)
- Burke, R.: Hybrid web recommender systems. In: *The Adaptive Web*, pp. 377–408. Springer, Berlin/Heidelberg (2007)
- Burke, R., Ramezani, M.: Matching recommendation technologies and domains. In: *Recommender Systems Handbook*, pp. 367–386. Springer, USA (2011)

- Delić, A., Neidhardt, J., Werthner, H.: Are sun lovers nervous? E-Rev. Tourism Res. (eRTR), ENTER 2016 Special issue (2016)
- Fesenmaier, D.R., Wöber, K.W., Werthner, H. (eds.): Destination recommendation systems: behavioral foundations and applications. CAB International (2006)
- Garcia, I., Sebastia, L., Onaindia, E.: On the design of individual and group recommender systems for tourism. *Expert Syst. Appl.* **38**(6), 7683–7692 (2011)
- Gibson, H., Yiannakis, A.: Tourist roles: needs and the lifecourse. *Ann. Tourism Res.* **29**(2), 358–383 (2002)
- Goldberg, L.R.: An alternative “description of personality”: the big-five factor structure. *J. Pers. Soc. Psychol.* **59**(6), 1216 (1990)
- Gower, J.C.: A general coefficient of similarity and some of its properties. *Biometrics* 857–871 (1971)
- Gretzel, U.: Intelligent systems in tourism: a social science perspective. *Ann. Tourism Res.* **38**(3), 757–779 (2011)
- Gretzel, U., Mitsche, N., Hwang, Y., Fesenmaier, D.R.: Travel personality testing for destination recommendation systems. In: *Destination Recommendation Systems: Behavioral Foundations and Applications*, vol. 42(4), pp. 121–136. CAB International (2006)
- Hwang, Y., Gretzel, U., Xiang, Z., Fesenmaier, D.R.: Travel De. In: *Destination Recommendation Systems: Behavioral Foundations and Applications*, vol. 42(4), pp. 357–371. CAB International (2006)
- Ipso MediaCT: The 2014 traveler’s road to decision. Retrieved 8, 2017 from <https://www.thinkwithgoogle.com/consumer-insights/2014-travelers-road-to-decision/> (2014)
- James, G., Witten, D., Hastie, T., Tibshirani, R.: Statistical learning. In: *An Introduction to Statistical Learning*, vol. 112, pp. 15–57. Springer, New York (2013a)
- James, G., Witten, D., Hastie, T., Tibshirani, R.: Linear regression. In: *An Introduction to Statistical Learning*, vol. 112, pp. 59–126. Springer, New York (2013b)
- James, G., Witten, D., Hastie, T., Tibshirani, R.: Linear model selection and regularization. In: *An Introduction to Statistical Learning*, vol. 112, pp. 203–264. Springer, New York (2013c)
- Kaufman, L., Rousseeuw, P.J.: Partitioning around medoids (program PAM). In: *Finding Groups in Data: An Introduction to Clustering Analysis*, pp. 68–125. Wiley, Hoboken, NJ (1990)
- Neidhardt, J., Schuster, R., Seyfang, L., Werthner, H.: Eliciting the users’ unknown preferences. In: *Proceedings of the 8th ACM Conference on Recommender Systems*, pp. 309–312 (2014)
- Neidhardt, J., Seyfang, L., Schuster, R., Werthner, H.: A picture-based approach to recommender systems. *Inf. Technol. Tourism* **15**(1), 49–69 (2015)
- Neidhardt, J., Werthner, H.: Travellers and their joint characteristics within the seven-factor model. In: *Information and Communication Technologies in Tourism 2017*, pp. 503–515. Springer (2017)
- Pang-Ning, T., Steinbach, M., Kumar, V.: Cluster analysis: basic concepts and algorithms. In: *Introduction to Data Mining*, pp. 487–568. Pearson Addison-Wesley (2006)
- Ricci, F., Rokach, L., Shapira, B.: Recommender systems: introduction and challenges. In: *Recommender Systems Handbook*, pp. 1–34. Springer, USA (2015)
- Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987)
- Werthner, H., Alzua-Sorzabal, A., Cantoni, L., Dickinger, A., Gretzel, U., Jannach, D., Neidhardt, J., Pröll, B., Ricci, F., Scaglione, M., Stangl, B., Stock, O., Zanker, M.: Future research issues in it and tourism. *Inf. Technol. Tourism* **15**(1), 1–15 (2015)
- Werthner, H., Klein, S.: *Information technology and tourism—a challenging relationship*. Springer-Verlag, Wien, New York (1999)
- Werthner, H., Ricci, F.: E-commerce and tourism. *Commun. ACM* **47**(12), 101–105 (2004)

- Woszczyński, A., Roth, P., Segars, A.: Exploring the theoretical foundations of playfulness in computer interactions. *Comput. Hum. Behav.* **18**(4), 369–388 (2002)
- Zins, A.H.: Exploring travel information search behavior beyond common frontiers. *Inf. Technol. Tourism* **9**(3–1), 149–164 (2007)