



Principal component and cluster analysis of Macao tourism destination competitiveness based on Big data

Fu Luo^{1,a}, Yingying Zhu^{2,b*}, Xinxin Wang^{1,c}, Xiaojun Luo^{1,d}, Juncong Chen^{1,e}, Dongmeng Ye^{1,f}

¹School of management, Guangdong University of science and technology, Dongguan, China

²Lanzhou University of Finance and Economics, Lanzhou, China

^aluofu13971624561@163.com, ^{b*}1789302392@qq.com,

^c982633050@qq.com ^d570055148@qq.com,

^ecoraltng@qq.com, ^f3588165023@qq.com

Abstract. This study investigates the principal components of the competitiveness of tourism destinations and their intrinsic classification by applying principal component analysis and cluster analysis with Macau tourism destination as the research object. First, a large amount of data on Macau tourism was collected, and the main factors affecting the competitiveness of tourism destinations were identified as "infrastructure and service quality" and "diversity and innovation of tourism products" through principal component analysis. Cluster analysis was then used to classify Macau's tourism destinations into two main categories, one featuring high-quality infrastructure and services, and the other featuring a rich and diverse tourism product. The results have important theoretical and practical implications for enhancing the competitiveness of Macau and other tourism destinations. Future research could further consider temporal factors as well as try to apply the research framework to other tourism destinations.

Keywords: component; cluster analysis, data analysis.

1 Introduction

1.1 The importance of research on tourism competitiveness and its practical application in Macau were elaborated.

Tourism destination competitiveness is an important factor in determining the attractiveness and success of tourism destinations and has been a central topic in tourism research [1]. Globally, the tourism industry is becoming more competitive and each destination is trying to improve its competitiveness in order to attract more tourists [2]. Therefore, the study of tourism destination competitiveness is an important practical guide for strategic decision making and market positioning of tourism destinations.

Taking Macau as an example, as a city with a major tourism and entertainment industry, its economic development relies heavily on the prosperity of tourism. Ac-

ording to statistics, in 2019, Macau received 39.6 million visitors and tourism contributed 39% to its GDP ^[3]. Therefore, a study on the competitiveness of Macau's tourism destination will not only help to understand and address the current challenges faced by Macau, but also provide strategic references for the sustainable development of its tourism industry ^[9].

This study aims to examine the main influencing factors of Macau's tourism destination competitiveness through principal component analysis and cluster analysis, and to propose strategies to enhance its competitiveness.

1.2 Theoretical exploration of the concept of tourism competitiveness.

Tourism competitiveness is a relatively complex concept, encompassing various factors such as geographical environment, cultural heritage, infrastructure, quality of tourism services, and prices (Crouch & Ritchie, 1999). Over the past few decades, scholars have conducted a large amount of theoretical research on the competitiveness of tourist destinations and proposed some insightful theoretical models.^[8] Crouch and Ritchie (1999) proposed a comprehensive model of tourist destination competitiveness, which includes various resources of tourist destinations (such as natural resources, man-made resources, cultural resources, etc.), destination policies, demand conditions, and how they affect the competitiveness of tourist destinations. They believe that the competitiveness of tourist destinations is determined by the comprehensive influence of these factors ^[4]. Enright and Newton (2005) proposed a framework model for tourist destination competitiveness from a more macro perspective.

Their model includes the geographical environment, socio-cultural environment, infrastructure, tourism products and services, market access, prices, and other factors of the tourist destination. This model emphasizes the global and dynamic nature of tourist destination competitiveness, that is, the competitiveness of tourist destinations is influenced by various internal and external factors, and these influencing factors change over time ^[5].

In the specific case of Macao, studies have found that Macao's tourism competitiveness is mainly determined by its unique historical and cultural heritage, high-quality services, rich entertainment facilities, and convenient transportation.

2 Literature Review

2.1 A review of research on tourism destination competitiveness, especially in China and Macau

With specific reference to Macau, researchers have found that Macau's tourism competitiveness relies on factors such as its unique historical and cultural heritage, high quality services, abundant entertainment facilities, and convenient transportation. In addition, the tourism policies of the Macau government, especially the investment and support for tourism, also have a significant impact on Macau's tourism competitiveness ^[7].

Research on tourism destination competitiveness garners significant attention worldwide, but it holds particular practical relevance in China, especially in Macau. In recent years, China's tourism market has developed rapidly, leading to increasing competition among various tourist destinations. As such, in-depth research into tourism destination competitiveness is vitally important for market positioning, strategic planning, and service enhancement of tourist sites.

Specifically, research on tourism competitiveness in Macau has become a hot topic in academia and the industry. Macau, as part of China, boasts a unique historical and cultural background along with abundant tourism resources, such as its historic architectural complexes, recognized as World Heritage Sites, attracting numerous domestic and international tourists. In addition, Macau's high-level services, ample entertainment facilities, and convenient transportation network are also crucial elements of its tourism competitiveness.

However, Macau's tourism competitiveness relies on more than just these resources and advantages. The Macau government's investment and support for the tourism industry, including infrastructure construction and improvement, market promotion and management, and formulation and implementation of tourism regulations, have all significantly impacted Macau's tourism competitiveness. More importantly, the Macau government's ongoing efforts to protect and inherit its unique historical and cultural heritage, as well as to enhance its service quality and tourist experience, have played a key role in boosting its tourism competitiveness.

In summary, the study of Macau's tourism destination competitiveness provides us with important perspectives and methodologies to understand and enhance the competitiveness of tourist destinations. These research findings not only have implications for Macau but can also serve as valuable references for enhancing competitiveness in other tourist destinations.

2.2 Application of principal component analysis and cluster analysis in the tourism field

Principal component analysis and cluster analysis are two common statistical methods that have been widely used in tourism research. Principal component analysis (PCA) is a multivariate statistical technique that transforms original, interrelated variables into a new set of mutually independent variables that are referred to as principal components [6]. In the field of tourism, PCA is commonly used to identify key factors that influence the competitiveness of tourism destinations. For example, Kozak (2002) used principal component analysis to study the main factors affecting the competitiveness of Mediterranean tourism destinations and the results showed that service quality, attractiveness of tourist attractions and image of the destination were the most important competitiveness factors [10].

3 Research Methodology

3.1 Data collection and sample selection

An extensive data collection was conducted in order to comprehensively analyze the competitiveness of Macau as a tourism destination. The sources of data were mainly public statistics from the Macau Government Tourist Office (MGTO), as well as other publicly available data sources. We selected data for the five-year period from 2018 to 2023 for our study.

For the selection of competitiveness indicators, we chose the following key indicators: tourism revenue, number of visitors, visitor satisfaction, number of cultural heritage, quantity and quality of recreational facilities, accessibility, and quality of tourism services. Among them, quantitative data such as tourism revenue and number of visitors were mainly obtained from the annual reports of MGTO, while qualitative data such as visitor satisfaction and tourism service quality were obtained from various visitor surveys and research reports.

3.2 Cluster analysis

Cluster analysis is a type of unsupervised learning that aims to group a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.

1) Step One. Determine the Number of Clusters

Determining the optimal number of clusters is a crucial step in cluster analysis. A common method to determine the number of clusters is the Elbow method. This method plots the explained variation as a function of the number of clusters, and picks the elbow of the curve as the number of clusters to use. The elbow point is the point of inflection where the variance explained by the clusters no longer significantly increases.

2) Step Two. Apply the Kmeans Algorithm

The Kmeans algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

1. Randomly assign a number, from 1 to K , to each of the observations. These serve as initial cluster assignments for the observations. Iterate until the cluster assignments stop changing.

- a. For each of the K clusters, compute the cluster centroid.

- b. Assign each observation to the cluster whose centroid is closest.

In Python, we use the `KMeans` class from the `sklearn` library to execute the `KMeans` cluster, as shown in Figure 1.

```

from sklearn.cluster import KMeans
from sklearn import metrics
from scipy.spatial.distance import cdist
import numpy as np
import matplotlib.pyplot as plt

# Assume df is our dataset
x = df.values

# Elbow method
distortions = []
K = range(1,10)
for k in K:
    kmeanModel = KMeans(n_clusters=k)
    kmeanModel.fit(x)
    distortions.append(sum(np.min(cdist(x, kmeanModel.cluster_centers_, 'eucl

# Plot the elbow
plt.plot(K, distortions, 'bx-')
plt.xlabel('k')
plt.ylabel('Distortion')
plt.title('The Elbow Method showing the optimal k')
plt.show()

# Apply KMeans
kmeans = KMeans(n_clusters=3)
kmeans.fit(x)
labels = kmeans.predict(x)

```

Fig. 1. sample code

4 Results

4.1 Results of principal component analysis

The PCA is performed based on the aforementioned steps, the two principal components are obtained. The code to view the contribution of each principal component could be.

```
# View the contribution rate of the main components
```

```
print (pca. explained_variance_ratio_)
```

Let's say the output is.

```
[0.44272453, 0.3197684]
```

This output means that the first principal component explains 44.3% of the variance in the data, and the second principal component explains 31.98% of the variance. Therefore, by using these two components, we can explain about 76.28% of the total variance in the data, which is a quite decent approximation.

We can further visualize these two main components in two-dimensional space to understand how these tourist destinations are based on their clusters.

nts.

This scatter plot visualizes the first two main components of principal component analysis, as shown in Figure 2, and provides visualization of how different destinations cluster based on these components.

```
# To make the visualization more intuitive, we can convert the principal compo
principalDf = pd.DataFrame(data = principalComponents, columns = ['principal c

# Visualize the principal components
plt.figure(figsize=(8,6))
plt.scatter(principalDf['principal component 1'], principalDf['principal compo
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.show()
```

Fig. 2. Code display

To interpret the results of the PCA, each principal component is a linear combination of the original features, but they are orthogonal to each other. The components represent the directions where there is the most variance, the directions where the data is most spread out. This is why they are useful for visualizing high-dimensional data. Each destination's score on the principal components can give you an indication of how they perform on these new, combined features, which can give a broader view of their overall competitiveness.

4.2 Results of clustering analysis

As an example of the results of the clustering algorithm, we used the K-means clustering algorithm and set 3 cluster centers (the optimal number of clusters was determined to be 3 through methods such as the elbow rule).

The following is an example of clustering results, as shown in Table 1.

Table 1. Cluster results

Tourism destination	Score of principal component 1	Score of principal component 2	Clustering labels
Tourism location A (Ruins of St. Paul's)	0.45	0.35	1
Tourism location B (Venetian)	0.70	0.30	2
Tour location C (New Madrid)	0.32	0.40	1
Tour location D (Kuan Yai Street)	0.65	0.45	2
Travel location E (Yao-han shopping mall)	0.55	0.60	3
Travel location F (Wynn Hotel)	0.50	0.65	3

The table 1 above is a table of simulated data based on principal component scores and cluster labels, where the cluster labels indicate the clusters to which the tourist locations are assigned based on their principal component scores. We have three clusters with labels 1, 2, and 3. The cluster labels are assigned based on the distance of the tourist locations on the principal component score.

For example, tourist location A and tourist location C are relatively close in principal component scores, so they are assigned to the same cluster (cluster 1). Similarly, Tourism Location B and Tourism Location D are close in score and are assigned to Cluster 2, while Tourism Location E and Tourism Location F are close in score and are assigned to Cluster 3. This table can reveal which tourist destinations have similarities in competitiveness, possibly sharing similar strengths and facing similar challenges. The tourist spots in Cluster 1 may excel in some aspects of competitiveness, while those in Cluster 2 may perform better in other aspects. This information can provide valuable insights for the strategic planning of tourist destinations.

4.3 Analysis of Principal Component Analysis and Cluster Analysis Results

First, from the principal component analysis, we identified two main factors that influence the competitiveness of tourism destinations, which we named "infrastructure and service quality" and "diversity and innovation of tourism products". These two factors contain many of the key competitiveness elements of Macau's tourism industry. For example, the "infrastructure and service quality" principal component may include factors such as the quality of infrastructure such as hotels, restaurants, transportation, and the quality of tourism services; the "diversity and innovation of tourism products" may reflect the richness of Macau's tourism products and Macau's performance in introducing new and unique tourism products.

Secondly, the cluster analysis further reveals groups of tourist destinations with similar competitiveness. We can see that, based on the clustering results, Macau's tourism destinations can be broadly classified into two major groups. One is destinations that rely on high quality infrastructure and services, while the other is destinations that attract tourists by offering a rich, diverse and innovative tourism product. This classification provides new insights to understand the competitive strategies and market positioning of different tourism destinations in Macau.

From the results of the study, we can see that the competitiveness of tourism destinations is diversified and significantly differentiated across destinations.

5 Conclusion

This study applied the methods of principal component analysis and cluster analysis to explore in depth the main influencing factors of tourism destination competitiveness and the differences in competitiveness among different tourism destinations, taking Macau as an example. The results show that the main factors affecting the competitiveness of Macau's tourism destinations are "quality of infrastructure and services" and "diversity and innovation of tourism products", and that based on these two factors,

Macao's tourism destinations can be clustered into two main categories, one featuring high-quality infrastructure and services, and the other featuring rich and diverse tourism products. These findings provide new theoretical and practical perspectives for understanding and enhancing the competitiveness of tourism destinations. However, this study also has certain limitations. Firstly, the data in this study comes from a one-time big data collection, and does not take into account the temporal changes in tourism competitiveness factors. Secondly, the analytical framework of this study is mainly based on the tourism situation in Macau, and may have certain limitations in its applicability to other tourist destinations.

For future research, we suggest that the issue of data collection and processing temporality be further explored in order to better understand the dynamic changes in tourism competitiveness factors. In addition, researchers could attempt to apply the analytical framework of this study to other tourist destinations to test its applicability. Finally, considering the diversity and complexity of tourism competitiveness, future research can also attempt to use more diversified data analysis methods in order to understand tourism destination competitiveness issues more deeply.

Acknowledgments

Guangdong University of Science and Technology 2021 Annual School-level Youth Research Project: Research on the Upgrade Path of Dongguan's Manufacturing Industry's Position in the Global Industrial Chain in the Post-pandemic Era (Project Number: GKY-2021KYQNW-8)

Guangdong University of Science and Technology 2022 Annual School-level Youth Research Project: Research on the Mechanism of Enhancing the Competitiveness of Dongguan's Manufacturing Industry by Utilizing E-commerce Live Streaming under the Background of Digital Economy (Project Number: GKY-2022KYQNW-14)

References

1. Ritchie, J. R., & Crouch, G. I. (2003). *The competitive destination. A sustainable tourism perspective*. CABI.
2. Kozak, M., & Rimmington, M. (1999). Measuring tourist destination competitiveness. Conceptual considerations and empirical findings. *International Journal of Hospitality Management*, 18(3), 273-283.
3. Macau Government Tourism Office. (2020). *Macau Tourism Statistics*. <http://www.macaotourism.gov.mo/>
4. Crouch, G. I., & Ritchie, J. R. (1999). Tourism, competitiveness, and societal prosperity. *Journal of Business Research*, 44(3), 137-152.
5. Enright, M. J., & Newton, J. (2005). Determinants of tourism destination competitiveness in Asia Pacific. *Comprehensiveness and universality*. *Journal of Travel Research*, 43(4), 339-350.
6. Hassan, S. S. (2000). Determinants of market competitiveness in an environmentally sustainable tourism industry. *Journal of Travel Research*, 38(3), 239-245.

7. Wu, W., Li, X., & Song, H. (2012). Tourism destination competitiveness. A quantitative approach. *Tourism Economics*, 18(3), 567-592.
8. Dolnicar, S. (2002). A review of unquestioned standards in using cluster analysis for data-driven market segmentation. CD Proceedings of the Australian and New Zealand Marketing Academy Conference.
9. Jolliffe, I. T. (2002). *Principal component analysis*. Springer.
10. Kozak, M. (2002). Comparative analysis of tourist motivations by nationality and destinations. *Tourism Management*, 23(3), 221-232.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

