POLITECNICO DI TORINO

ICT for Smart Mobility

Labratory Reports - Group 12

Report 3 - OD Matrix Similarity

Authors
Arman MohammadiGilani s301000
Sara Gholamhosseinzadeh s301689
Nikoo Arjang s300262

Supervisors
Marco Mellia
Luca Vassio

Academic Year 2022/23

1 Introduction

The objective of the third laboratory is to delve deeper into the analysis of the data collected in the first two laboratories in order to identify the most probable user classes of the car sharing system.

To achieve this, we conducted a comparison between the Origin-Destination (OD) matrices obtained from the Rental database and the information gathered from user interviews (IMQ dataset). An OD matrix serves as a comprehensive representation of movement within a specific area and plays a vital role in assessing transportation demand.

In the OD matrix, each row corresponds to the origins of trips, while the columns represent the destinations. Each cell within the matrix denotes the intersection of a trip originating from a particular origin and terminating at a specific destination. The numerical values within these cells indicate the frequency of trips along those routes, thereby providing valuable insights into the demand for different transportation connections.

By analyzing the OD matrices derived from both the Rental database and IMQ dataset, we aim to identify patterns and trends that can help us determine the most likely classes or categories of car sharing system users. These findings will contribute to a better understanding of user preferences, travel behavior, and the specific needs and demands of different user segments within the car sharing system.

2 Data Extraction

As previously mentioned, our study involves working with two distinct datasets: IMQ and Rental. To facilitate data extraction from these datasets, we have employed different platforms tailored to each specific dataset.

For the Rental dataset, we utilized MongoDB as a robust tool for efficient data extraction. MongoDB offers flexible querying capabilities, allowing us to retrieve relevant information from the Rental dataset seamlessly.

On the other hand, for the IMQ dataset, we employed the Pandas library, a powerful data manipulation and analysis tool in Python. Pandas provided us with an extensive range of functions and methods to extract and process the data from the IMQ dataset effectively.

2.1 Data Extraction from Rental dataset

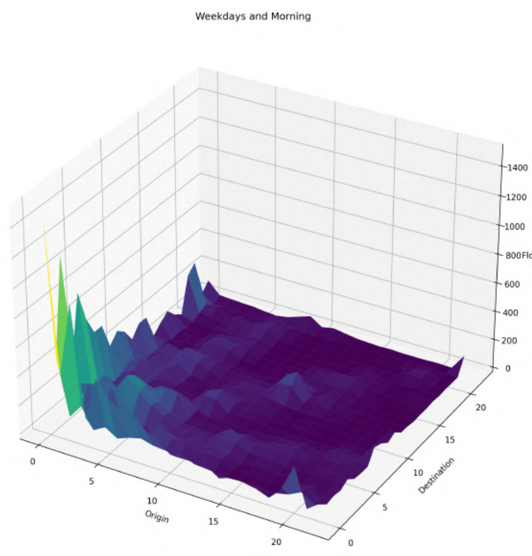
To extract the Origin-Destination (OD) matrices from the Rental dataset, we utilized the ictts-PermanentBookings dataset available in MongoDB. This dataset provides information on the initial location, final location, and initial time of car rentals.

In order to create the OD matrices, we defined zones boundaries using GeoJSON data specifically tailored for Turin. By mapping out these zones, we were able to classify the initial and final locations of the car rentals accordingly.

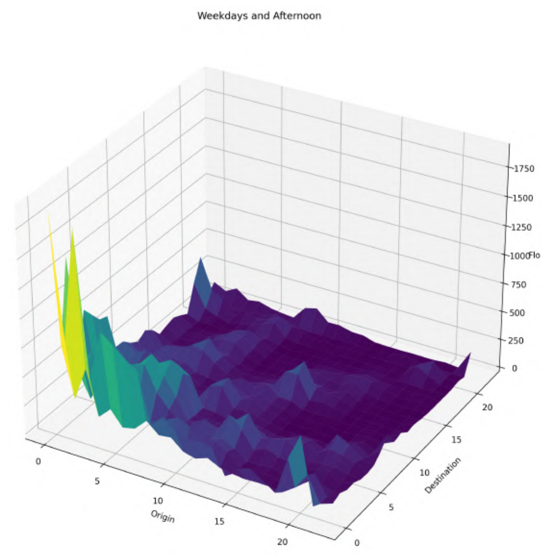
To generate comprehensive OD matrices, we followed specific filtering criteria based on various factors such as weekends or weekdays, as well as morning or afternoon rentals. This allowed us to extract four distinct OD matrices representing different scenarios.

For example, by filtering the data based on weekends, we focused on rentals that occurred specifically on Saturdays and Sundays. Similarly, by filtering based on weekdays, we analyzed rentals that took place on Mondays to Fridays. This segregation allowed us to observe any variations in the travel patterns and destinations between weekends and weekdays.

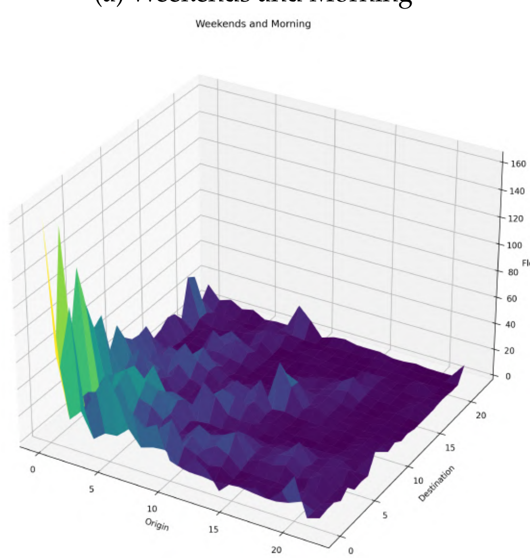
By applying these filters and extracting the corresponding data, we derived four separate OD matrices, each providing valuable insights into the spatial distribution and flow of car-sharing users based on specific time periods and days of the week.



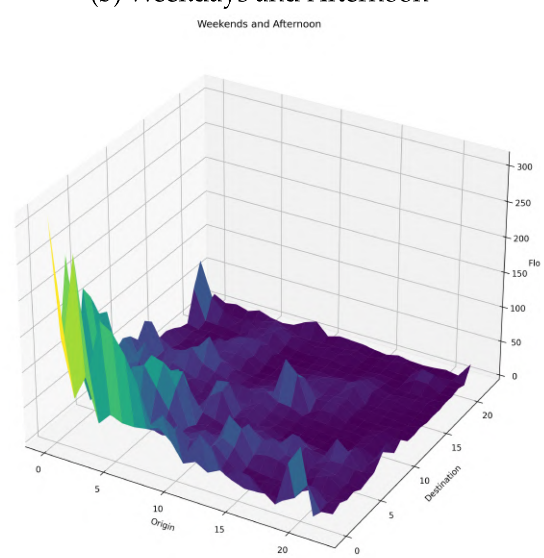
(a) Weekends and Morning



(b) Weekdays and Afternoon



(c) Weekends and Morning



(d) Weekends and Afternoon

Figure 1: Graphs of OD matrices extracted from Rental dataset

Figure 1 displays the four distinct Origin-Destination (OD) matrices that were extracted from the Rental dataset. Each graph within Figure 1 provides a visual representation of the origin and destination points, along with the corresponding number of trips recorded during specific time periods of the day.

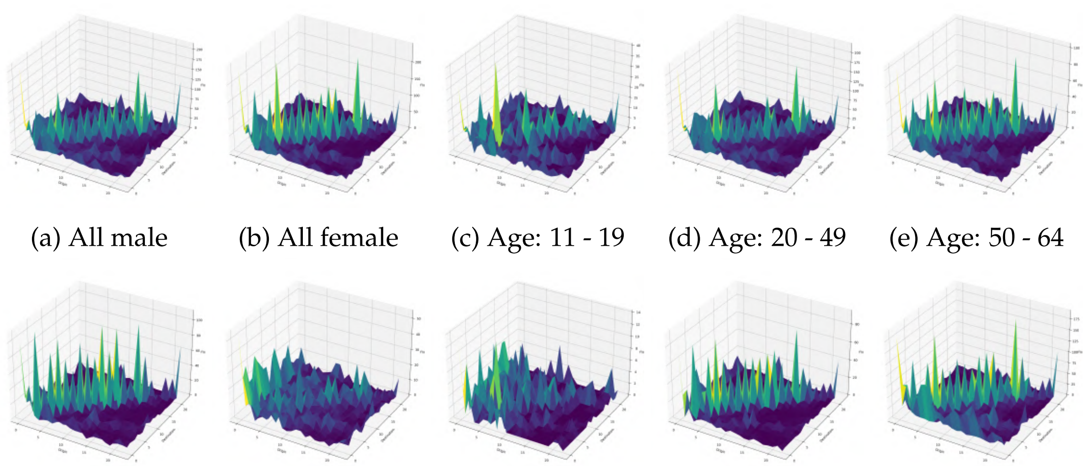
The depicted graphs from the Rental OD matrices suggest a concentrated distribution of trips within the central part of the city. This observation implies that a significant proportion of car-sharing users tend to travel to and from zones located in the central area of the city.

The central part of the city often serves as a hub of various activities, such as commercial, cultural, and entertainment centers, making it a popular destination for individuals utilizing car-sharing services. This concentration of trips in the central zones may indicate a higher demand for transportation options within this region, likely influenced by factors like proximity to workplaces, shopping areas, tourist attractions, or social events.

2.2 Data Extraction from IMQ dataset

The IMQ dataset, collected through phone interviews, contains valuable information gathered from 105,099 trips conducted from Monday to Friday. The dataset covers a total of 185 zones within the Piedmont region, with 23 zones specifically focused on Turin. The interviews were structured to capture data on various demographic variables, including gender (male/female), age groups (11-19, 20-49, 50-64, 65+), and trip motivations (such as going to work, going to school, visiting parents, and more).

To extract the Origin-Destination (OD) matrices from the IMQ dataset, we employed the *pivot table* method in Pandas. This method facilitated the extraction process, offering a straightforward approach to derive the OD matrices based on zone of arrival and departure. By filtering the data according to gender, age, and the purpose of the trip, we were able to generate customized OD matrices that highlight the travel patterns and preferences of different demographic segments.



(a) All male (b) All female (c) Age: 11 - 19 (d) Age: 20 - 49 (e) Age: 50 - 64

(f) Age: 65+ (g) Aim: Work (h) Aim: Study (i) Aim: Shopping (j) Aim: Home

Figure 2: Graphs of OD matrices extracted from IMQ dataset

The observed graphs from the OD matrices indicate that the majority of trips tend to occur within the same or nearby zones, with a few exceptions. Specifically, there are notable variations when considering specific age groups, such as individuals aged 11 to 19, or when the purpose of the trip is to go to work or study.

For the age group of 11 to 19, the graphs may show a wider distribution of trips across different zones compared to other age groups. This suggests that individuals within this age range might engage in more diverse activities or have different transportation needs, resulting in a broader range of trip destinations.

Similarly, when the aim of the trip is to go to work or study, the graphs might reveal a more dispersed pattern of trip origins and destinations. This can be attributed to the fact that individuals commuting to work or educational institutions often have specific locations they need to travel to, which could be spread out across the city or even beyond the usual residential zones.

Understanding these variations in trip patterns is essential for urban planners, transportation authorities, and car-sharing service providers. It allows for the implementation of targeted strategies and services to accommodate the unique needs and preferences of different demographic segments. By tailoring transportation options, infrastructure development, and resource allocation, stakeholders can enhance the efficiency, accessibility, and overall effectiveness of the transportation system, contributing to a smoother and more sustainable urban mobility experience.

3 Comparison

To assess the similarity of OD matrices, we employed the L2 distance as our comparative metric. However, to ensure accurate calculations for the L2 distance, a normalization step was implemented for each OD matrix. This normalization process involved dividing each element within a row by the sum of all elements in that particular row.

By normalizing the OD matrices in this manner, we effectively accounted for varying magnitudes and overall row-wise distributions. This normalization step is crucial as it allows for a fair and unbiased comparison of the OD matrices, ensuring that the subsequent L2 distance calculations provide reliable results. This approach helps to eliminate any potential bias introduced by the absolute values or row-wise sums of the original OD matrices.

When comparing two matrices, *A=a[ij]* and *B=b[ij]*, the Euclidean distance is calculated by summing the distances between each corresponding element in the matrices. This distance computation involves taking the square root of the sum of squared differences between the elements of *A=a[ij]* and *B=b[ij]*.

The Euclidean distance metric, provides a measure of dissimilarity between the two matrices. By summing the individual element-wise distances, we capture the overall discrepancy between the matrices in terms of both magnitude and direction.

$$Euclidean\,(A,B) = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(a_{ij}-b_{ij}\right)^2} \tag{1}$$

Table 1 provides a clear overview of the similarity among the OD matrices. Among the various comparisons made, it is evident that the OD matrix showing the highest similarity is the one representing "Rental Weekdays-Afternoon" and "IMQ with the aim of work."

On the other hand, the OD matrices displaying the least similarity are "Rental Weekends-Morning" and "IMQ 65+." The dissimilarity between these matrices suggests a notable contrast in the underlying travel patterns or trends.

Table 1: Distance Matrix

	Weekends-Morning	Weekdays-Afternoon	Weekends-Morning	Weekends-Afternoon
Male	1.4	1.2	1.4	1.3
Female	1.6	1.4	1.6	1.5
11-19	1.4	1.2	1.4	1.3
20-49	1.3	1.1	1.3	1.2
50-64	1.6	1.4	1.6	1.5
65+	2.1	2	2.1	2
Work	0.94	0.91	1	0.96
Study	1.4	1.4	1.5	1.4
Shopping	2	1.8	1.9	1.9
Home	1.9	1.6	1.9	1.7

Analyzing the levels of similarity or dissimilarity between OD matrices is crucial for understanding travel patterns, identifying variations in transportation demand, and evaluating the effectiveness of specific interventions or policies. These insights can contribute to informed decision-making in transportation planning and resource allocation.