Politecnico di Torino

ICT for Smart Mobility
Lab Report
Lab 3

Prof. Marco Mellia
Prof. Luca Vassio

Group 11

Seyed Mohammad Mehdi Hosseini - S301769
Seyedeh Sadaf Ekram - S301178
Elaheh Lotfimahyari - S301125

January 14, 2022

Contents

3.1 Introduction

This laboratory report aims to build upon the data analyzed in the first lab, which focused on car rentals of the "Car2Go" and "Enjoy" car-sharing services in Torino city. The goal of this laboratory is to determine the most likely class of users of car sharing systems. To achieve this, we will compare the Origin-Destination (OD) matrices of car-sharing users with the IMQ data. We will use the IMQ dataset metadata, such as gender, age, and goals of the trip, to filter the data and obtain the most similar OD matrix as the one from car-sharing users. This analysis will provide valuable insights into the demographic characteristics and transportation patterns of car-sharing users in Torino, and can help with the development and expansion of car-sharing services and facilities in the future.

3.2 Origin-Destination (OD) Matrices Computation

3.2.1 IMQ Data

IMQ 2013 is a dataset that contains information collected through phone interviews. The dataset comprises of 52,120 completed interviews, which resulted in the reporting of 105,099 trips. The trips reported in the dataset were undertaken from Monday to Friday and cover 185 zones of Piedmont, with 23 of those zones located in Torino. The interviews were stratified by several demographic factors, including gender, age, and motivation for the trip. Gender data is divided into male and female, age data is divided into four categories (11-19, 20-49, 50-64, 65+), and motivation for the trip is divided into categories such as going to work, going to school, visiting parents, etc. It provides valuable information on the transportation patterns and preferences of Torino residents during the weekdays, specifically in Torino.

Thank to the previous data prepossessing of the raw IMQ data (provided in *"spostamentiTorino.csv"*), each instance of the data includes the origin and destination of each interview (trip) by different zones in Torino city. By creating a tuple pair of origin and destination, we can count the number of each pair occurrence and assign it as the correspondence element of the IMQ data OD matrix which is a trip between two zones. Table 3.1 is the list of the zones ids and names in Torino city provided in *"TorinoZonescol.geojson"* file. Moreover, Figure 3.4 (In Appendix) shows the OD matrix for all the 16568 trips recorded in *"spostamentiTorino.csv"* file without any filtering.

3.2.2 Car-sharing Rentals Data

The data for the rentals is available on MongoDB in the *"ictts_PermanentBookings"* collection and the *"ictts_enjoy_PermanentBookings"* collection. In order to facilitate efficient data retrieval and analysis, there are three indexes that are important for this study. The first and second indexes are *"init_loc"* and *"final_loc"* which include 2D coordinates, to support geospatial queries, allowing us to easily retrieve data based on geographic location. The second index is *"init_time"* which allows for filtering data over time, enabling us to retrieve data for specific time periods including days and hours.

In order to obtain the OD matrix for all car-sharing data in Torino, we employed a pipeline to derive the OD matrix from each of the services (Car2Go, Enjoy) that fall within a predefined time period and geographic zones. Subsequently, both results were consolidated into a unified matrix referred to as the car-sharing OD matrix. To accomplish this task, we implemented a pipeline utilizing MongoDB to filter the data and compute the total rentals within a specified time frame between two zones. The pipeline consisted of the following stages:

• **Project:** In this stage, new fields were added to the documents in the collection. The first field was *"hour"* which extracted the hour from the *"init_date"* field. The second field was *"day"* which

Q001	Centro	Q007	Aurora	Q013	Pozzo Strada	Q019	Falchera
Q002	S.Salvario	Q008	Vanchiglia	Q014	Parella	Q020	Regio Parco
Q003	Crocetta	Q009	Nizza-Millefonti	Q015	Vallette	Q021	Madonna Pilone
Q004	S.Paolo	Q010	Lingotto	Q016	Madonna Campagna	Q022	Cavoretto
Q005	Cenisia	Q011	S.Rita	Q017	B.Ta Vittoria	Q023	Mirafiori Sud
Q006	S.Donato	Q012	Mirafiori Nord	Q018	B.Ra Milano		

Table 3.1: Zones ids and names

extracted the day of the week from the *"init_date"* field. The *"init_loc"*, *"final_loc"*, and *"init_time"* fields were also included in the **Project** stage.

• **Match:** In this stage, certain conditions were applied to the documents in order to filter the data. The day (derived from *"init_date"*) should be greater than or equal to the *"startDay"* and less than *"endDay"*. The hour ((derived from *"init_date"*)) should be greater than or equal to *"startHour"* and less than or equal to *"endHour"*. The *"init_loc"* should be within the geospatial boundaries specified by the *"orig_zone"* and *"final_loc"* should be within the geospatial boundaries specified by the *"dest_zone"* in *"TorinoZonescol.geojson"* file (See Table 3.1).

• **Count:** In this stage, we counted the number of documents that passed through the previous stages, and created a new field *"tot"* with the total count for each "Enjoy" and "Car2Go" services and at the end, we'll combine these values to form the total trips from *"orig_zone"* to *"dest_zone"*.

Through the execution of this pipeline, with a temporal constraint and varying origins and destinations, we are able to obtain elements of the OD matrix for car-sharing rentals. By specifying the *"startDay"* as 2 and *"endDay"* as 7 (corresponding to Monday through Friday, during which the IMQ data is also available) and *"startHour"* as 0 and *"endHour"* as 23, we can obtain the complete OD matrix of car-sharing rentals in the city of Torino, as shown in Figure 3.5 (In Appendix).

3.3 Data Filtering and Similarity Measurement Criteria

3.3.1 Car-sharing Data Filtering Criteria

The car-sharing data was filtered according to several scenarios outlined in Table 4. It is important to note that the IMQ data primarily pertains to trips made Monday through Friday, thus it is not appropriate to compare them with the car-sharing rentals on the weekends (Saturday and Sunday). Table 3.2 illustrates all the feasible filters for the car-sharing rentals data, in accordance with the criteria previously mentioned.

Time of Day	Day of Week
Monday-Friday from 0AM-5AM	Monday 24h
Monday-Friday from 5AM-9AM	Tuesday 24h
Monday-Friday from 9AM-2PM	Wednesday 24h
Monday-Friday from 2PM-7PM	Thursday 24h
Monday-Friday from 7PM-12PM	Friday 24h

Table 3.2: All considered filters for car-sharing rentals data.

Gender	Age	Trip Reason
(1) Male	(3) From 11-19	(8) Working Reason
(2) Female	(4) From 20-49	(9) Study
-	(5) From 50-64	(10) Shopping
-	(6) +65	(11) Bring Someone
-	-	(12) Medical Visits
-	-	(13) Sport or Leisure
-	-	(14) Going Back Home
-	-	(15) Visiting Relatives/Friends
-	-	(16) Other
-	-	(17) Going Back Home (Interview Day)

Table 3.3: All considered filters for IMQ data

3.3.2 IMQ Data Filtering Criteria

The IMQ data contains various attributes that can be used for analysis, however, the most crucial ones were selected, including "Gender", "Age", and "Trip Reason". Table 3.3 presents all the possible filters for the IMQ data based on the aforementioned criteria.

3.4 OD Matrices Similarity Comparison Method

Prior to the computation of similarity and distance, it is essential to normalize the elements of the OD matrix that contribute to the measurement, so as to enable comparability with another OD matrix, which is also normalized. To achieve this, we employed the Min-Max normalization technique to scale each individual matrix element between 0 and 1. This normalization method ensures that all the values of the matrix lie within a specific range, making it useful for handling variables with different scales and for avoiding the dominance of one variable over the others. In order to determine the similarity between two matrices, we employed the Euclidean distance (L2 Norm) measure. This measure quantifies the degree of similarity between two matrices. The Euclidean distance is defined as the square root of the sum of the squared differences between the corresponding elements of the two matrices, as shown in Equation 3.1. In this equation, C represents the car-sharing OD matrix and I represents the IMQ OD matrix to be compared. If two matrices are identical (which means the maximum possible similarity), the distance between them would be zero (distance score).

$$d(i,j) = \sqrt{\Sigma_{i=1}^{n}\Sigma_{j=1}^{n}(C_{ij} - I_{ij})^2} \tag{3.1}$$

3.5 OD Matrices with Single Filter Comparison

In this section, we aim to compare car-sharing rentals and IMQ data by applying a single filter to the IMQ data. Figure illustrates all the Euclidean distance scores related to all the single filters presented in Table 3.2 and Table 3.3. The vertical rows in the figure correspond to the filters associated with the car-sharing rentals data (See Table 3.2) and the horizontal columns correspond to the filters associated with the IMQ data (See Table 3.3).

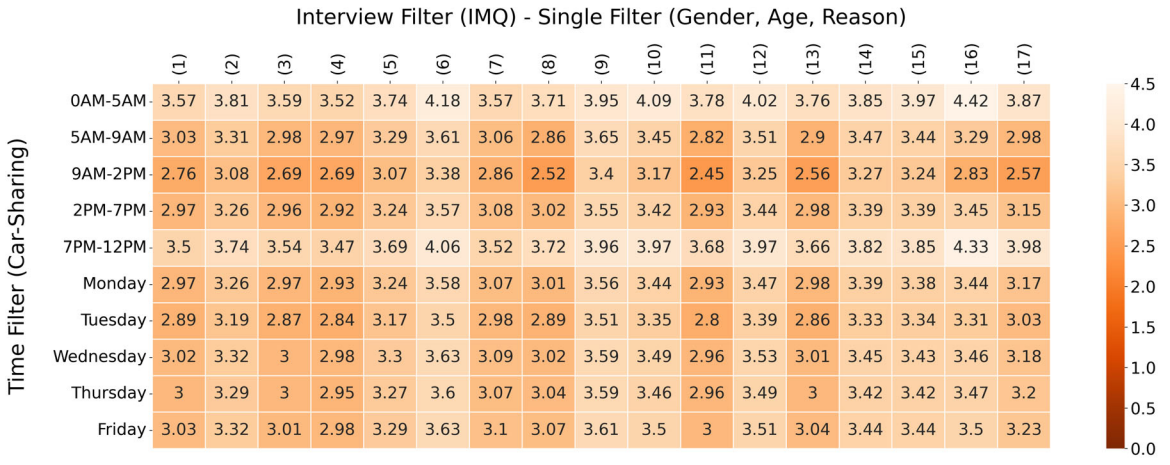

Figure 3.1: Distance score between car-sharing rentals (Rows) and IMQ (Columns) data using only a single filter - Columns are the numbers indicated in Table 3.3

Figure 3.1 presents the results of a comparison between the car-sharing rentals and IMQ data, considering a single filter at a time. The results indicate that, when considering the Gender filter alone, the behavior of males is more similar to the car-sharing rentals pattern than "Females". Furthermore, when the Age filter is applied, the OD matrix of individuals aged 20-49 is found to be more similar to the car-sharing rentals pattern. Lastly, when analyzing the Trip Reason filter, it is observed that the OD matrix of trips taken to "Bring Someone" is more closely aligned with the car-sharing rentals pattern. All these facts applies to all days and hours (See Table 3.2.

3.6 OD Matrices with Two Filters Comparison

3.6.1 Filtering on Both Gender and Age

The results of comparing car-sharing rentals and IMQ data, considering two filters at a time (Gender and Age), are presented in Figure 3.2. The analysis indicates that OD matrix of females aged 11-19 years old and males aged 20-49 have the most similar pattern to the car-sharing rentals pattern across the majority of time periods. It is noteworthy that the most similar OD matrix is found to be for females aged 11-19 years old between 9AM-2PM, with a distance score of 2.41.

3.6.2 Filtering on Both Gender and Trip Reason

The results of comparing car-sharing rentals and IMQ data, considering two filters at a time (Gender and Trip Reason), are presented in Figure 3.3. The analysis indicates that OD matrix of females who mentioned "Bring Someone" and "Sport and Leisure" as their trip reason, have the most similar pattern to the car-sharing rentals pattern across the majority of time periods. It is noteworthy that the most similar OD matrix is found to be for females who wanted to bring someone between 9AM-2PM, with a distance score of 2.43.

3.6.3 Conclusion

To compare by time filters, the most similarities happens between 9AM-2PM and between the day weeks (Monday-Friday), Tuesday has the most similarity to the IMQ data (which is also valid for aforementioned paragraphs).

In conclusion, among all comparisons, Females aged 11-19 years old have the most similarity (lowest distance) with the car-sharing rentals data between 9Am-2PM with distance score 2.41 in total.

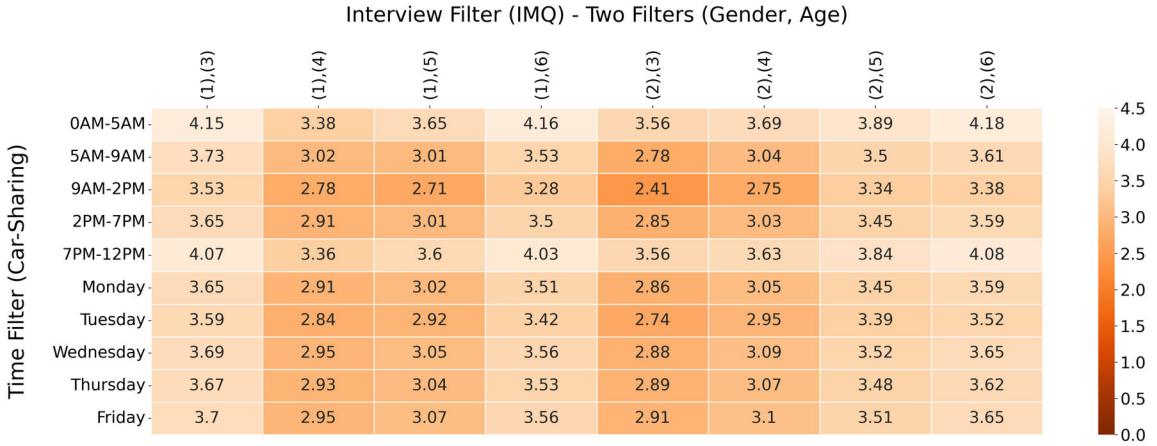

Figure 3.2: Distance score between car-sharing rentals (Rows) and IMQ (Columns) data using two filters on Gender and Age - Columns are the numbers indicated in Table 3.3

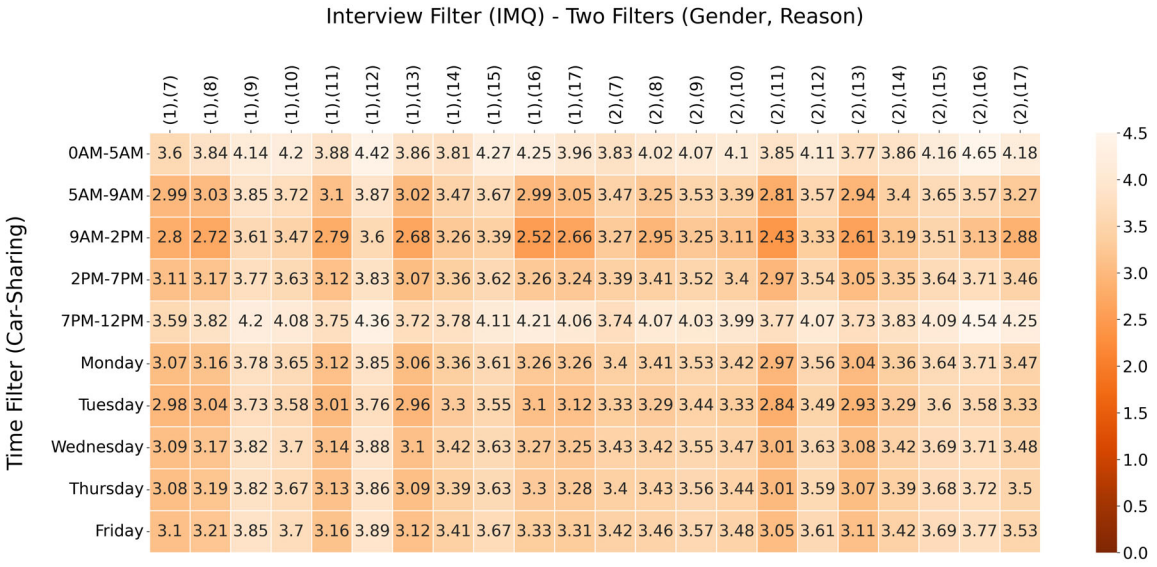

Figure 3.3: Distance score between car-sharing rentals (Rows) and IMQ (Columns) data using two filters on Gender and Trip Reason

3.7 Appendix: Additional Figures and Source Code

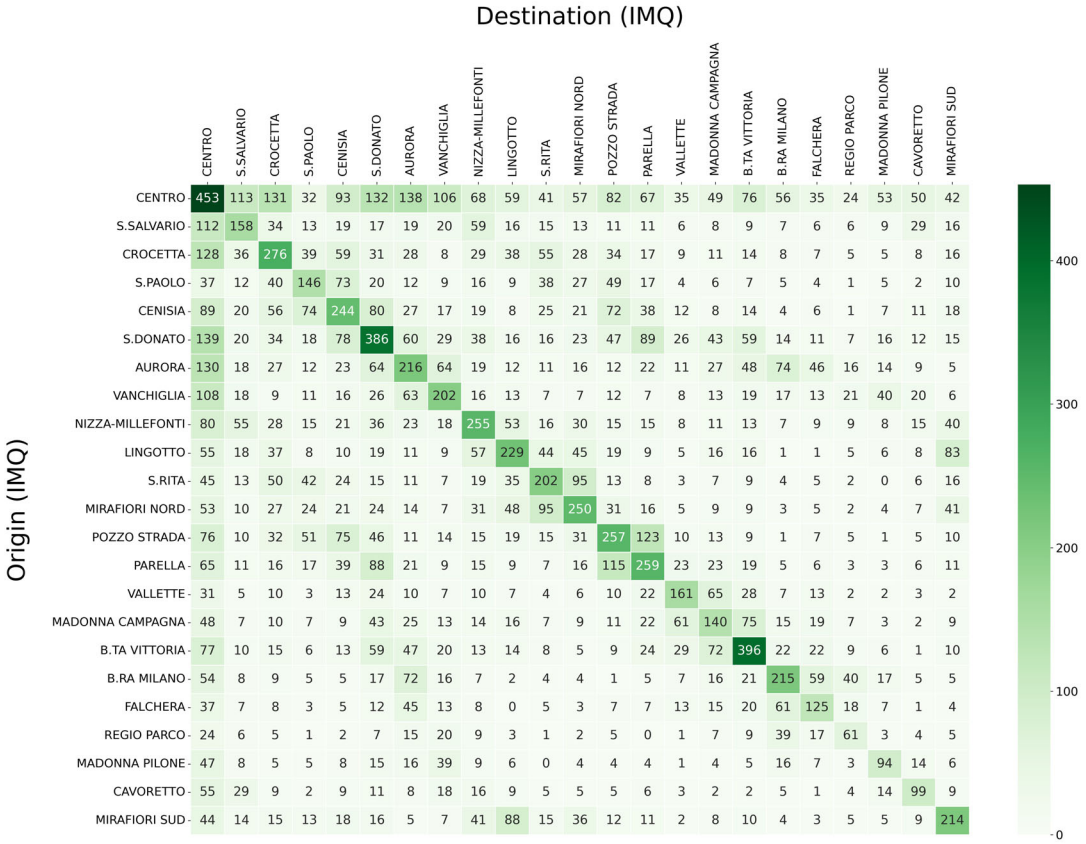

Figure 3.4: Origin-Destination Matrix for all IMQ Data

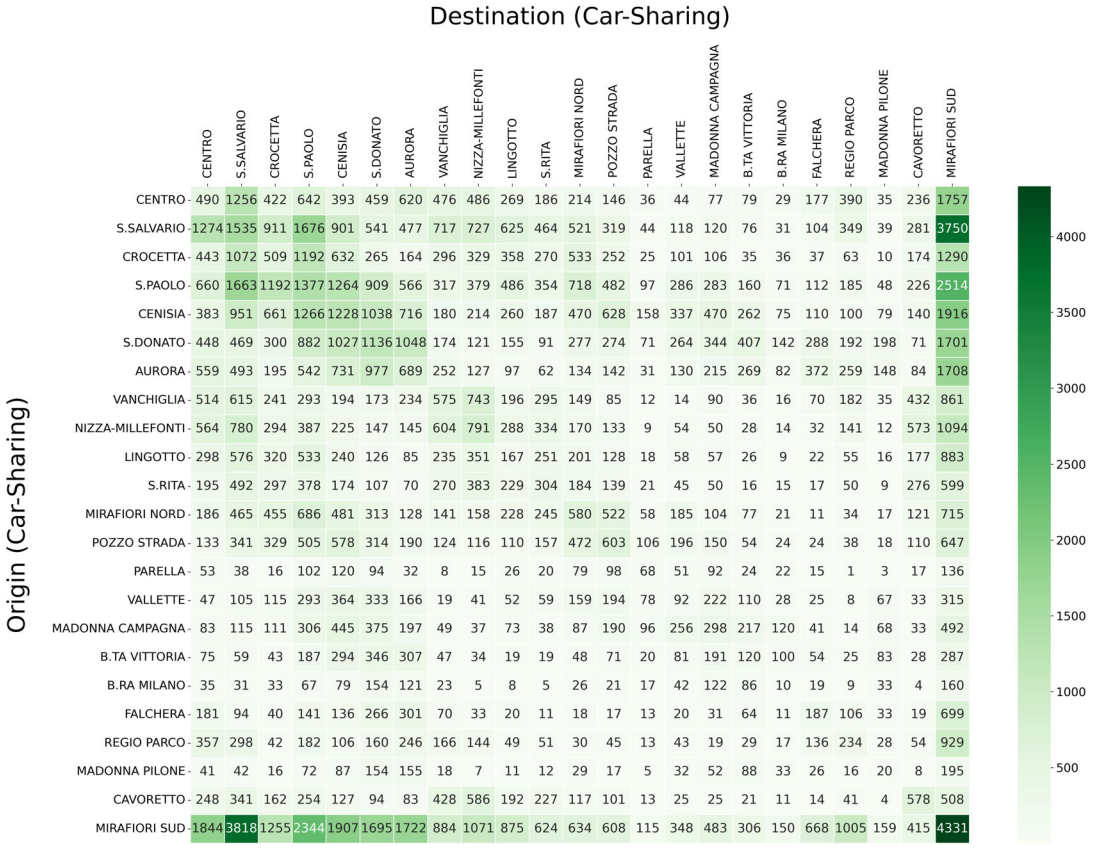

Figure 3.5: Origin-Destination Matrix for all car-sharing rentals Data