ICT for Smart Mobility Laboratory Report Lab 1

Prof. Marco Mellia

Group 11

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

October 30, 2022

Contents

1.1 Step1 - Lab1 - Preliminary Data Analysis

1.1.1 Number of Present Documents in Each Collection

Collection	Active Bookings	Active Parkings	Permanent Bookings	Permanent Parkings
Car2Go	8743	4790	28180508	28312676
Enjoy	0	0	6653472	6689979

Table 1.1: Number of present documents in each collection

1.1.2 Similarities of Documents in Permanent Parkings and Permanent Bookings

Both *Permanent Bookings* and *Permanent Parkings* have the same number of documents, differing at a negligible number of documents. As long as there is a booking in the service, there would be a parking in advance, so these two collections are expected to have similar number of documents except the cases in which there is a system issue or maybe a maintenance is undergoing.

1.1.3 Cities in the Collecting Data

Here is the List of 26 Cities in *Car2Go* Data Collection:
Amsterdam, Austin, Berlin, Calgary, Columbus, Denver, Firenze, Frankfurt, Hamburg, Madrid, Milano, Montreal, Munchen, New York City, Portland, Rheinland, Roma, San Diego, Seattle, Stuttgart, Torino, Toronto, Twin Cities, Vancouver, Washington DC, Wien

Also, the list of *Enjoy* Data Collection including 6 cities can be observed :
Bologna, Catania, Firenze, Milano, Roma, Torino

1.1.4 Start and End Time of the Collection

Car2Go: The first document for *Car2Go* data collections has been recorded at Stuttgart on 2016-12-13 18:38:23 local time zone (2016-12-13 18:38:23 GMT+1), and respectively for *Enjoy* data collections on 2017-05-05 17:06:21 local time zone (2017-05-05 17:06:21 at GMT+1) at. The most newest document for *Car2Go* has been added to the database on 2018-01-31 08:11:33 local time zone in Washington DC(2018-01-31 14:11:33 GMT+1) and about *Enjoy*, it's on 2019-06-10 19:16:20 local time zone in Milano. (2019-06-10 19:16:20 GMT+1)

1.1.5 Timestamps for different Time zones

The Unix time is referred to Greenwich Mean Time (GMT), while the human readable format for the dates is referred to the local time zones.

1.1.6 Available Cars in Each City

In Seattle, there have been reported 1473 total vehicles in the whole period of *Car2Go* service and now there are 870 active cars giving service in Seattle. In Milano, 1153 *Car2Go* cars and 1870 *Enjoy* cars have been recorded in total.However, *Enjoy* has got no active cars but *Car2Go* still is active with 818 cars all across Milano.

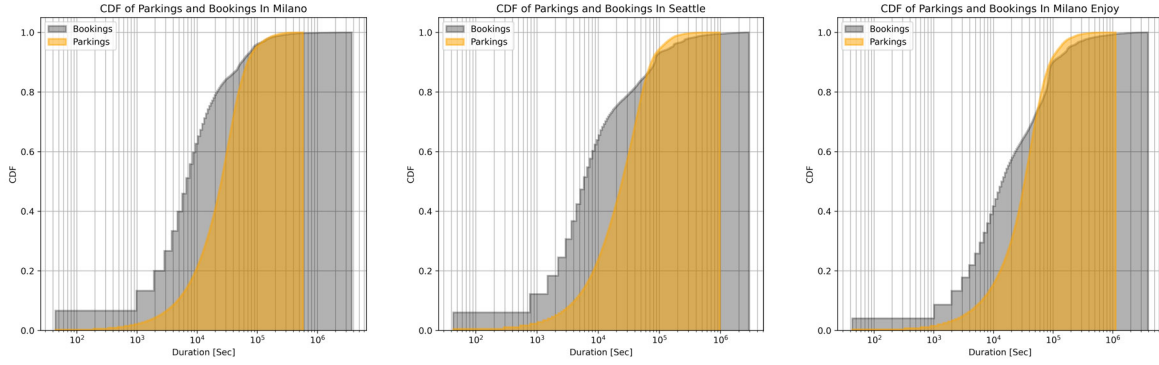| Alternative transportation	Walking	Public Transport	Driving
Seattle	0	0	0
Milano	727793	648718	824891
Milano Enjoy	824411	648718	824891

Table 1.2: Alternative transportation options

Figure 1.1: CDF of RAW Bookings/Parkings data for each city

1.1.7 Recorded Booking on the January 2018 in Each City

In January 2018, there are 107,437 bookings in Seattle, 242,608 bookings by *Car2Go* users in Milano and 214,188 *Enjoy* users in Milano using these companies.

1.1.8 Alternative Transportation Mode

There are three transportation alternatives mentioned on table 1.2. In general, there were no recorded alternative transportation for Seattle city.

1.2 Step2 - Lab1 - Analysis of Data

The data for the three cities is the analytics regarding the period of 1st of December 2017 until 28th February 2018.

1.2.1 Cumulative Distribution Function Derivation

According to the Figure 1.1, it can be observed that the CDF of booking is longer than parking for all three studying cities which is due to the unfiltered data. For example, if a car has been taken for a maintenance, it will be considered as booking. The unusual time of maintenance will cause an outlier into our database which is the reason of longer CDF of booking. Booking time in Milano city for both *Car2Go* and *Enjoy* companies are more considerable than in Seattle city. This difference can be caused by different size of the land area or traffic situations.

Considering Figure 1.2, the booking trend in Mondays and Tuesdays are considerably different from the other days but the parking trend is not significantly different from the other days in a week..
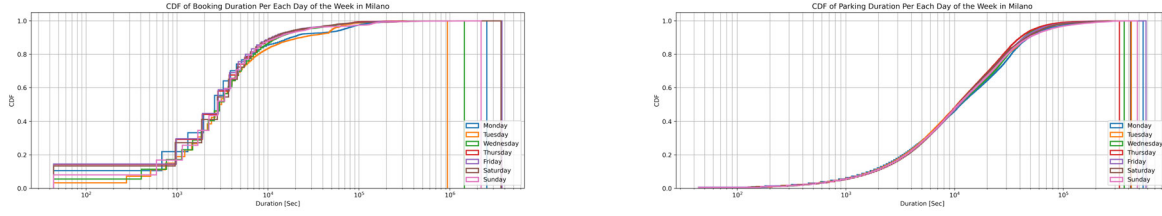
Figure 1.2: CDF of RAW Bookings/Parkings data per each day of the week in Milano-Enjoy

1.2.2 System Utilization Over Time

In this phase, the number of rentals on hourly basis are evaluated so as to get a sense of system utilisation. No filtering is applied on the data in order to remove outliers (the major picks in figures) or misleading measurements but is still possible to see a pattern in the data with respect to the day and night car usage. By looking at Figure 1.3, it can be observed that a portion of data is missing from the database, because the system failed in obtaining data from 1st of February to 28 of February. When it comes to Seattle, the number of booked/parked cars were significantly lower with respect to Milano. It can be explained by the fact that Milano may be more hectic and demand is much higher.

1.2.3 Filtering Criteria

To filter possible outliers, some different filtering criteria have been applied based on duration of the usage. The following filtering criteria have been applied:
• No booking that lasts for less than 4 minutes: A car is usually booked for more than a few minutes (likely a system error or cancelled booking can be less than 4 minutes)
• No bookings that last for more than 2 hours, so that too long: booking cars are removed since it can be a system error or a car in maintenance,
• The start and end coordinates of a booking event have to be different otherwise it is a booking due to system error.
• No parking lasts for less than 4 minutes, since, when a user parks a car and someone else wants to pick it up, there should be at least an amount of time for the car to reappear at the system to be booked again by someone else.

1.2.4 Filtering Applied

The results of the filtering criteria applied to the data are shown in Figure 1.4. The booking series has a smoother behaviour meaning that the filtering is somehow successful. This cannot be said for the parking series. Parkings are difficult to filter since a car can be parked for a long time. A way to approach the problem might be to look for those parkings whose previous booking has been filtered out, meaning that also those parking events have to be filtered.
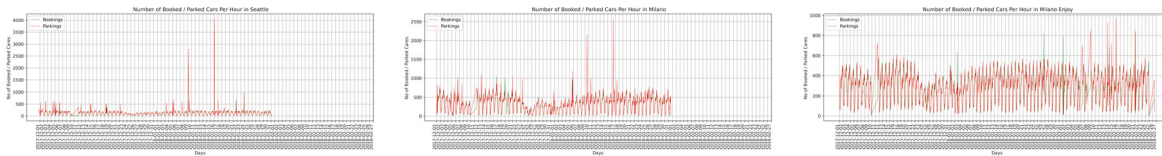

Figure 1.3: Number of RAW Booked/Parked cars data per hour along the whole period
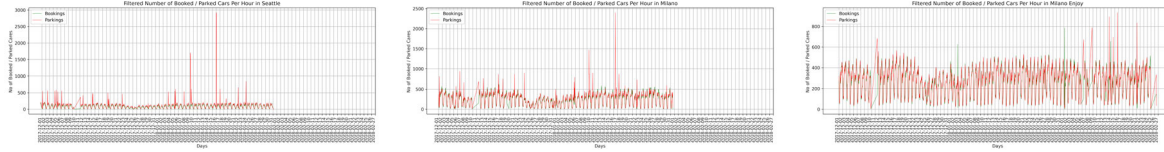
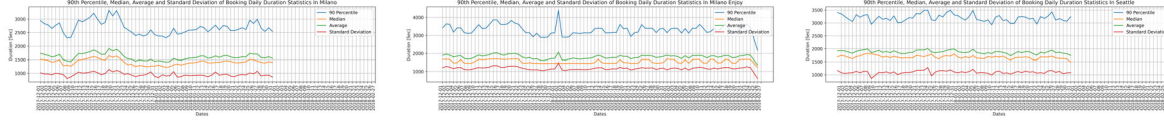Figure 1.4: Number of Filtered Booked/Parked cars data per hour along the month

Figure 1.5: Statistics of Bookings

1.2.5 Data Statistics

Figure 1.5 and Figure 1.6 show the average, median, standard deviation, and 90 percentiles. Since we evaluated these operators on the filtered data, the average value and median show a similar behavior. However, The median is under the average because it is influenced by high values. The standard deviation is always higher for the parkings than for the bookings. Parkings has a larger variance with respect to the bookings which has a more compact distribution. This means that the values of booking duration evaluated over time are closer to the average, while the values of parking duration are more widely spread.

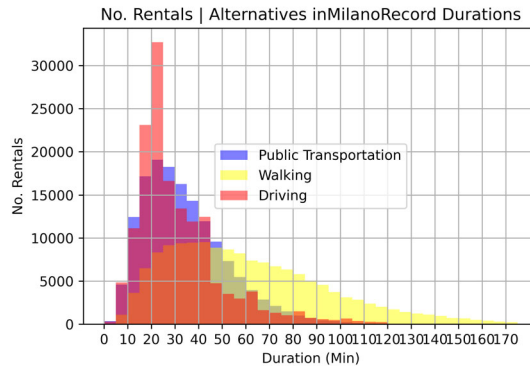Figure 1.6: Statistics of Parkings

Figure 1.7: Histogram of Rentals vs Alternative Transportation Durations

1.2.6 OPTIONAL TASK: Correlation with other transport options

The purpose behind this task is to assess the number of rentals given the availability of other transport modes. The data of Milano Enjoy has been filtered to get the ones with alternative transport modes.For example figure 1.7 demonstrates if public transport takes 20 to 25 minutes, people prefer to use Enjoy services to reach their destination. When it takes more than 80 minutes to use public transport or

driving options, customers tend less to rent a car. However, for walking option, the graph is more distributed. The tendency for renting a car still exists even after 2 hours of walking.

One the probability side, it's possible to derive from histogram:

$$P\{AlternativeTransport > x | TotalRentals\}$$

but as long as we don't know the probability of total rentals, we can't expect to have the probability of rentals considering the alternative transports using Bayes Theorem:

$$P\{TotalRentals | AlternativeTransport > x\} =$$

$$P\{AlternativeTransport > x | TotalRentals\} P\{Rentals\} / P\{AlternativeTransport > x\}$$