ICT for Smart Mobility

01DSABH

Laboratory 1 Report

Group:

12

Members:

Mohammadreza Taheri - s327795
Lena Kazemahvazi - s327796
Saeed Amiri - s326315

Academic Year 2024/2025

1 Preliminary Data Analysis

1.1 Data Exploration

- **Number of Documents in Each Collection:**
  `ActiveBookings`: 8,743, `ActiveParkings`: 4,790,
  `PermanentBookings`: 28,180,508, `PermanentParkings`: 28,312,676,
  `enjoy_ActiveBookings`: 0, `enjoy_ActiveParkings`: 0,
  `enjoy_PermanentBookings`: 6,653,472, `enjoy_PermanentParkings`: 6,689,979.

- **Similarity Between `PermanentParkings` and `PermanentBookings`:**
  `PermanentBookings` records when cars are booked, while `PermanentParkings` marks when cars are available again. Differences in counts arise from temporary service interruptions.

- **Cities in the Dataset:**
  `PermanentBookings/PermanentParkings`: Amsterdam, Austin, Berlin, Calgary, Columbus, Denver, Firenze, Frankfurt, Hamburg, Madrid, Milano, Montreal, Munchen, New York City, Portland, Rheinland, Roma, Seattle, Stuttgart, Torino, Toronto, Vancouver, Washington DC, Wien.
  `ActiveBookings/ActiveParkings`: Subset of these cities, excluding Amsterdam.
  `enjoy_PermanentBookings/enjoy_PermanentParkings`: Bologna, Catania, Firenze, Milano, Roma, Torino.
  `enjoy_ActiveBookings/enjoy_ActiveParkings`: No data.

- **Date Range of Data Collection:** Collection periods are shown in Figure 1, converted from UNIX time for readability.

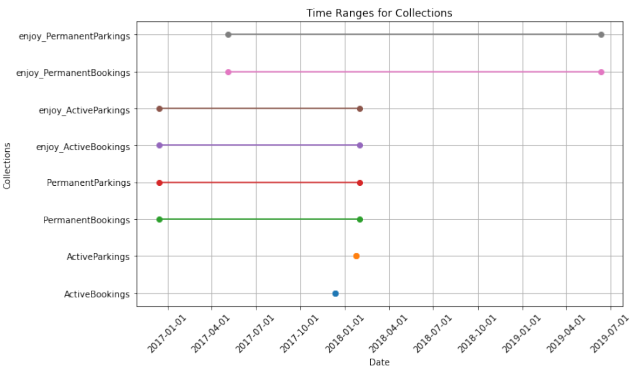


Figure 1: New York City

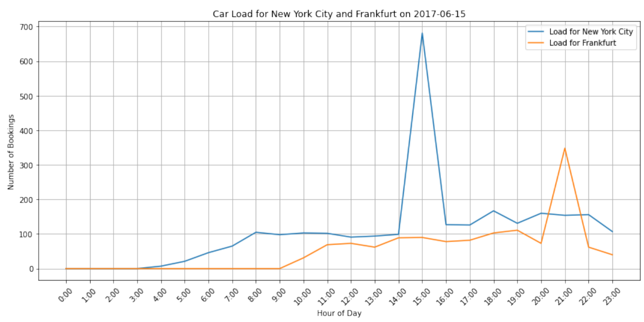


Figure 2: Hamburg

- **Time Zone of Timestamps:** Hourly booking demand in New York City and Frankfurt shows peaks separated by five hours, matching their timezone difference. This suggests timestamps are recorded in UTC (see Figure 2).

1.2 City Specific Analysis

For team G, the assigned cities are: New York City, Hamburg, Madrid.
For each assigned city:

- **Total Unique Cars:** New York City: 968, Hamburg: 1012, Madrid: 475.

- **Estimated Weekly Fleet Size (Week of 2017-07-01):** NYC: 506, Hamburg: 785, Madrid: 451. This represents 52.27%, 77.57%, and 94.95% of total unique cars, respectively, highlighting high turnover in NYC and low turnover in Madrid (see Figure 3).

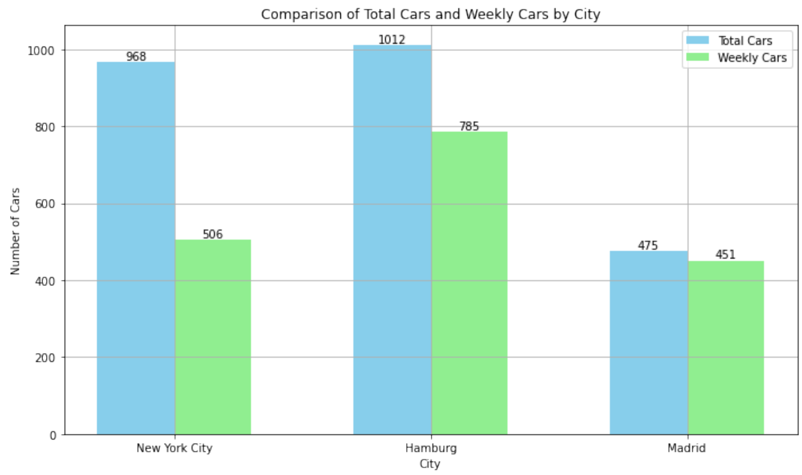
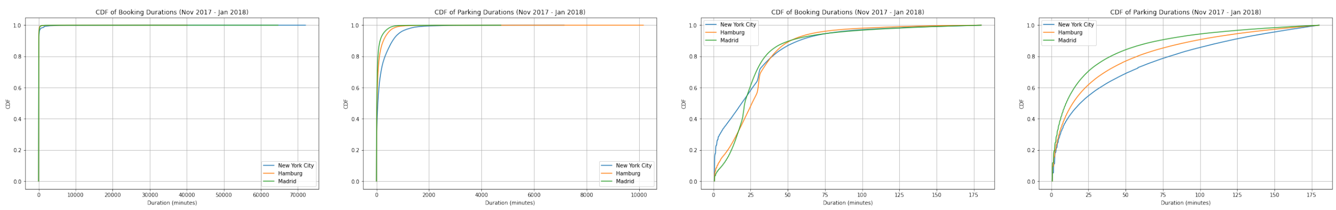


Figure 3: Weekly Fleet Size vs. Total Unique Cars

- **December 2017 Bookings:** NYC: 74,957, Hamburg: 257,019, Madrid: 171,057.
- **Bookings with Alternative Transport Modes:** Only Torino and Milano have such data: Torino: 296,398, Milano: 727,059. All other cities lack alternative transport data.

2 Car Sharing Usage Characterization

2.1 Distribution of Booking/Parking Duration (CDF Analysis)



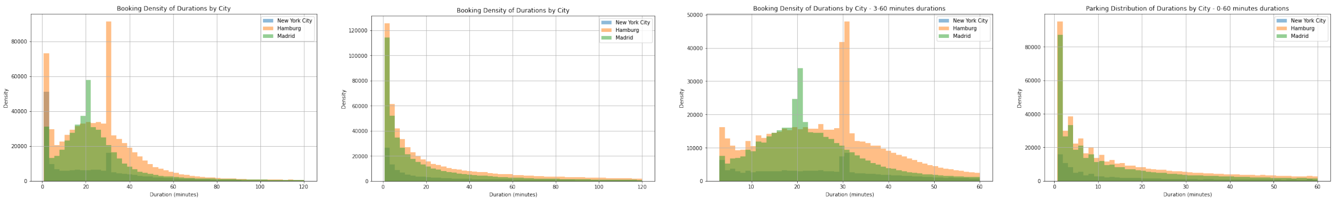
(a) Bookings (Unfiltered) (b) Parkings (Unfiltered) (c) Bookings (Filtered) (d) Parkings (Filtered)

Figure 4: CDF of Bookings and Parkings Duration (Unfiltered and Filtered) for Each City.

Observations: Unusually long booking/parking durations, exceeding 3 hours, suggest potential data anomalies or system issues. The CDF analysis (Figure 4a and 4b) highlights this pattern for all trips, while the filtered CDF (Figure 4c and 4d) focuses on trips under 3 hours.

a. Which city has more density for larger values of duration? Is this expected? Does the CDF suggest the presence of some outliers?

After filtering durations under 2 hours, booking densities (Figure 5a) show longer durations are more common in Hamburg. Parking densities (Figure 5b) appear similar across cities, potentially reflecting shared fleet dynamics or car distribution patterns.



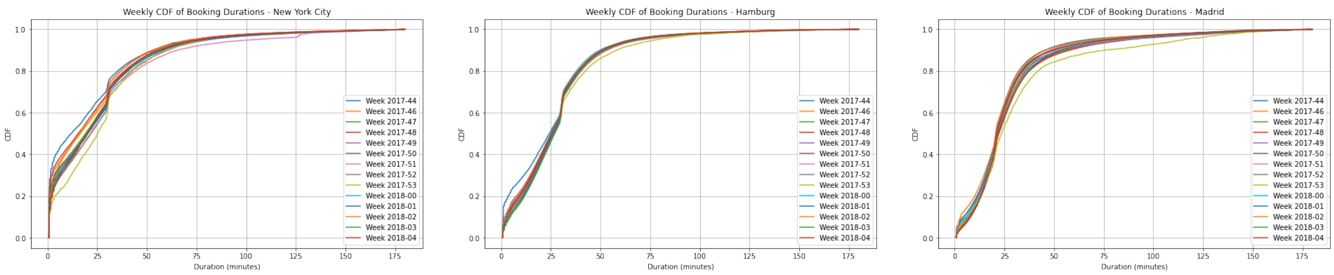
(a) Booking (partfiltered) (b) Parking (partfiltered) (c) Booking (Filtered) (d) Parking (Filtered)

Figure 5: CDF of Bookings and Parkings Densities (Unfiltered and Filtered).

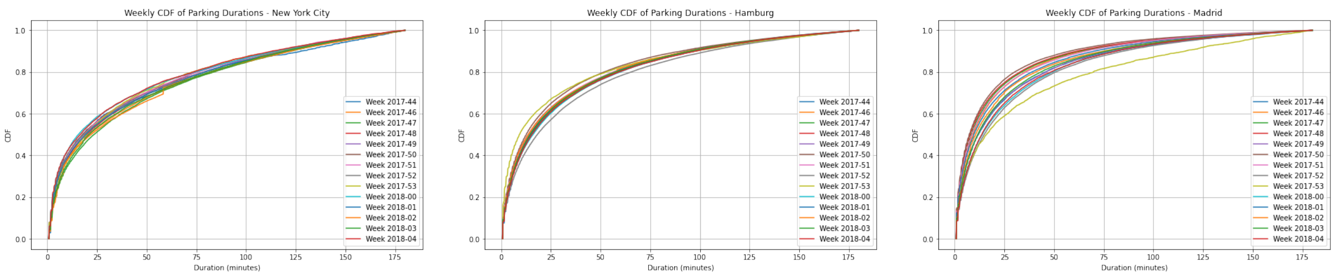
b. Interpretation of CDF Differences

The booking CDFs reveal city-specific patterns in user behavior and trip distribution. In *New York City*, shorter trips dominate, likely reflecting urban density and abundant alternative transport. Conversely, *Hamburg* shows a higher density of longer trips, potentially due to geographic layout or user preferences for extended rentals.

c. Time-based CDF Changes



Bookings: NYC **Bookings: Hamburg** **Bookings: Madrid**



Parking: NYC **Parking: Hamburg** **Parking: Madrid**

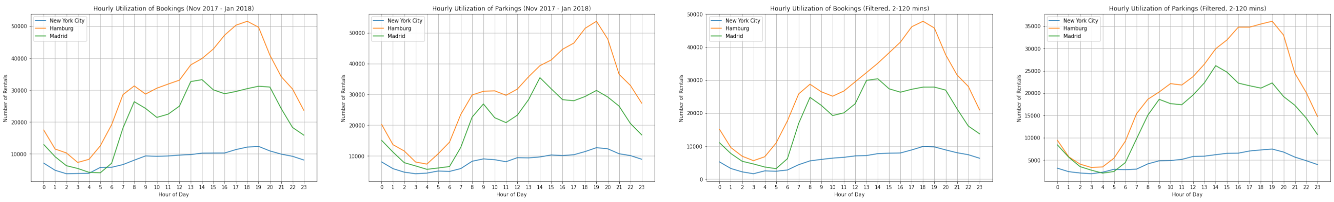
Figure 6: Comparison of Bookings and Parking for Three Cities in Different Weeks.

As generally shown in figure 6, We see that generally the shape of the curve of the CDF for each city reamins the same regardless of the week of choice. This can be considered as a characteristic. Maybe people tend to have a certain preference or maybe the routes and the size of the city is playing a role here.

2. System Utilization over Time

Different usage patterns are observed throughout the day, with peaks in the morning, midday, and afternoon, likely reflecting commuting times (Figure 7a and 7b).

Each city exhibits unique trends. For example, Hamburg shows higher usage at night, while New York City maintains a consistent usage rate throughout the day without sharp peaks.



(a) Bookings (Unfiltered) (b) Parkings (Unfiltered) (c) Bookings (Filtered) (d) Parkings (Filtered)

Figure 7: System Utilization for Bookings and Parkings (Unfiltered and Filtered).

Observations: The dataset shows differences in usage across cities, with fewer cars and documents in New York City compared to Hamburg and Madrid. This may reflect the dominance of car-sharing services in Europe, while ride-sharing services like Uber are more popular in North America.

3. Criterion for Outlier Filtering

We filtered **too long** rentals (2–3 hours) and now apply a filter for **too short** trips (less than a minute). For parking, only an upper bound is applied since very short durations are plausible.

- **Bookings:** 1–2 minutes to 2–3 hours.
- **Parkings:** 0 minutes to 2–3 hours.

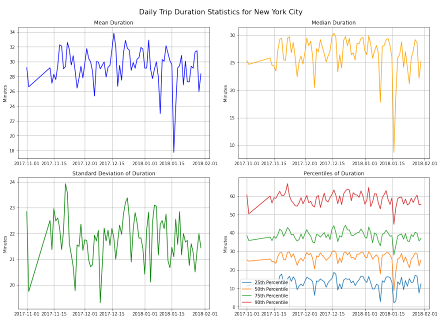
4. System Utilization with Filtered Data

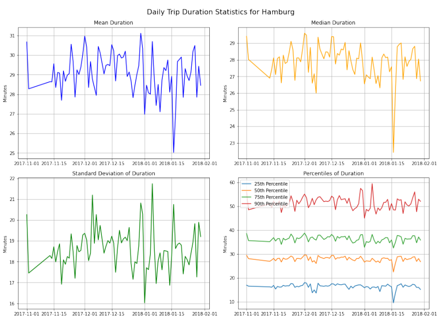
Applying filters to bookings reduces rentals but maintains a similar utilization curve across cities (figure 7c and 7d). Hamburg and Madrid still show three peaks, while New York City has a smoother curve. Filtering "too short" and "too long" rentals results in more normal density behavior.

Since booking and parking distributions differ, lower bounds should vary (see 2.1). Trips shorter than 1–2 minutes are rare, but brief parking is plausible, as cars can be rented again immediately. Figures 5c and 5d shows filtered booking and parking durations under stricter upper bounds.

Observation: A peak at 30 minutes appears for New York City and Hamburg, while Madrid peaks at 20 minutes. These sharp peaks warrant further investigation.

5. Descriptive Statistics of Booking Duration





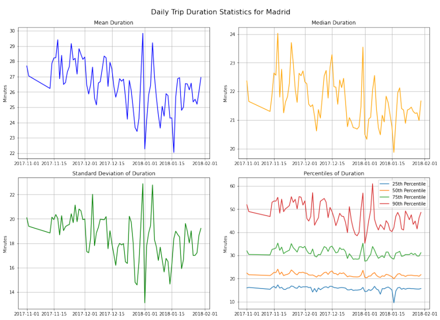


Figure 8: New York City Figure 9: Hamburg Figure 10: Madrid

a. Do these figures change over time?

Mean durations vary over time, with peaks observed on weekends and dips during holidays.

b. Is it possible to spot any periodicity (e.g., weekends vs. weekdays, holidays vs. working periods)?

Clear periodicity exists, mainly week-dependent. Patterns repeat every 7 days due to fixed dates like weekends and the start of weekdays.

c. Is it possible to spot any trend (e.g., increasing, decreasing, holiday periods)?

Periodic trends persist, with notable peaks and dips for specific dates. For instance, New Year's Eve in New York City (figure 8) shows a large peak, followed by a dip the next morning, reflecting its festive significance. Similar peak-dip behavior is evident on notable dates in European cities (figures 9, 10).

6. Correlation with Alternative Transport (Milano)

a. Valid Rentals with Alternative Transport Data

Valid alternative transport modes are identified by filtering documents with public transport durations and distances not equal to -1. This provides a dataset for further analysis.

b. Rental Probability vs. Transport Duration

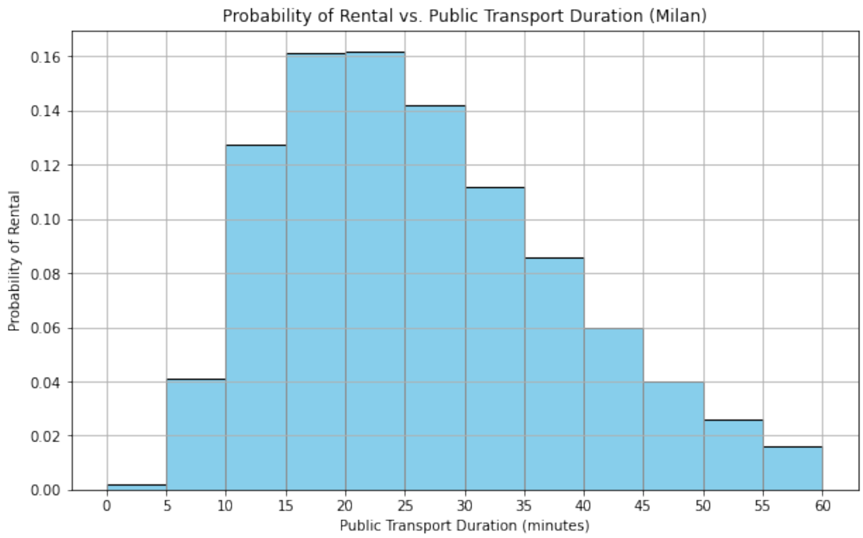
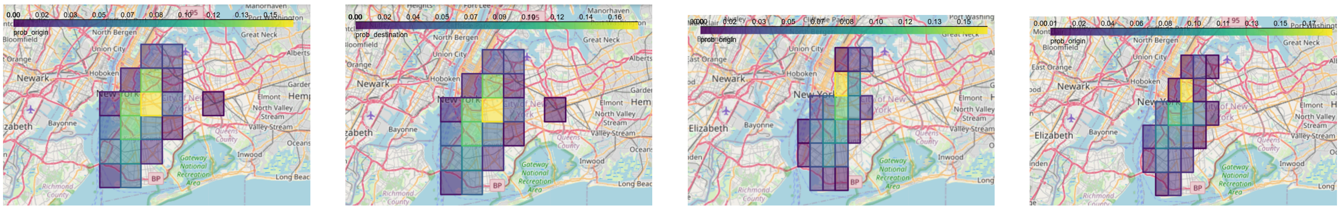


Figure 11: Histogram of Rental Probabilities by Public Transport Duration Bins.

Figure 11 shows car rental probabilities against public transport durations. Rentals peak for trip times of 15–30 minutes, while shorter or longer public transport durations reduce car-sharing usage.

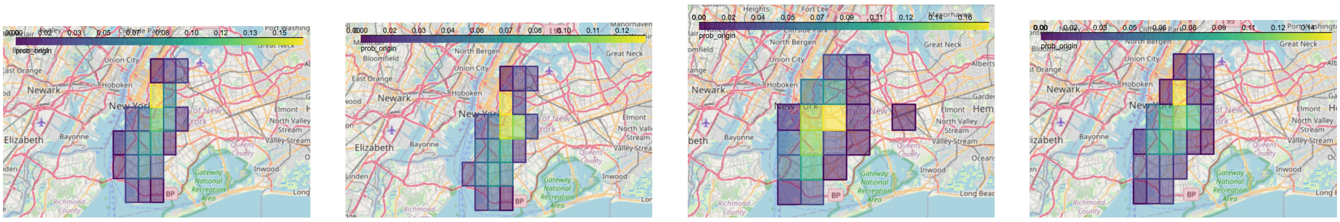
7. Car Density Analysis (Location of Booked and Returned Cars)

We considered New York City from the PermanentBookings collection, sampling 10% from the 500,000 filtered trips for visualization. The heatmaps below represent different analyses of origins and destinations.



(a) Origins (b) Destinations (c) Mon 8-10 A.M. (d) Mon 6-8 P.M.

Figure 12: Heatmaps: Origins/Destinations & Monday Time Variations (New York City).



(a) Mon. 8-10 A.M. (b) Wed. 8-10 A.M. (c) Weekday (d) Weekend

Figure 13: Heatmaps: Different Days & Weekday/Weekend Comparisons (New York City).

The analysis of car booking probabilities at various times was conducted using a sampled dataset derived from the entire set of trips, applying appropriate time-based criteria. A zoning system consisting of 42 square zones was utilized, with each trip being mapped to one of these zones. This zoning approach allowed for visualization and density computation of the bookings.

Figures 12a and 12b show the densities of car bookings for different origins and destinations, respectively. Due to the large number of data points, the true essence of the density in some areas may not be fully captured. The data is normalized, and the dense areas in the visualizations appear similar, indicating general trends rather than precise details.

The differences in the booking probabilities for the origins are most noticeable when comparing Monday morning bookings at different times. Figure 12c reveals more pronounced booking densities in certain areas compared to figure 12d, which reflects a different time window (6-8 A.M.).

Further comparison of bookings during fixed time intervals (8-10 A.M.) on two different days, Monday and Wednesday, shows that the overall heatmap patterns are quite similar. Figures 13a and 13b highlight this similarity, with both days exhibiting comparable density distributions for the origins.

Additionally, when comparing weekday bookings to weekend bookings, significant differences in the heated areas are observed. Figures 13c and 13d show that the density of bookings is notably higher during weekdays, likely due to work-related travel patterns, while weekend bookings follow a different distribution, with varying peak areas.