Analyzing Bike Trip Dynamics: Temporal Patterns and Day-Specific Variations

Luis Baroja

December 14, 2023

Introduction

Bike trips have emerged as a popular and environmentally friendly transportation alternative, particularly in large cities. Understanding bike trip dynamics, and ultimately the user behavior behind them, is crucial to constantly improve the efficiency of this green transportation alternative.

With the goal of understanding the dynamics of this mode of transportation, this study takes a close look at the inter-arrival times of trip accumulation as a primary step in comprehending bike trips. Describing how frequently users engage with the bikes is a crucial initial step in paving the way for a more comprehensive account of the inner workings of bike-sharing systems in metropolitan areas.

In the present work we propose three variations of a Poisson process core model to describe observed oscillations in the frequency of bike trips throughout the day. Our results reveal interesting patterns at specific hours and meaningful differences between weekdays and weekends.

Data

For this report we utilized the Metro Bike Share public data spanning the third quarter of 2023, from July 1 to September 30.[1] This data set, the most recent available, includes a total of 130,753 bike trips within the LA metropolitan area, each characterized by 15 distinct variables. No missing observations were detected and, in general, the variables presented meaningful information.[2]

[1] The data set `metro-trips-2023-q3.csv` is available as a .zip file in `https://bikeshare.metro.net/about/data/`.

[2] An important exception is the variable indicating trip `duration`. Even though the trip `start` and trip `end times` seem to be accurate, the maximum duration is recorded as 1440 minutes, or 24 hours, even for trips extending over multiple days according to the previous two variables (structured as dates). Even though such variable is not relevant for the analyses presented next, it is worth reporting its anomalies.

We computed the inter-arrival times between each trip, that is the temporal distances between successive bike trip starts. Additionally, we extracted the hour and minute of each trip start as separate numerical variables and identified the day of the week (e.g., Monday, Tuesday, etc.) to explore potential variations and latent differences across different days.[3]

Figure 1 visually represents the inter-arrival times of each trip. Specifically, each point marks the time since the start of the last trip against the time of start of the current trip. The most immediate finding is the non-homogeneous distribution of inter-arrival times throughout the day: during the early morning hours inter-arrival times are long, but they gradually decrease as the day progresses, only to start increasing again during the early night hours. Modeling this general pattern is the main objective of the present analyses.

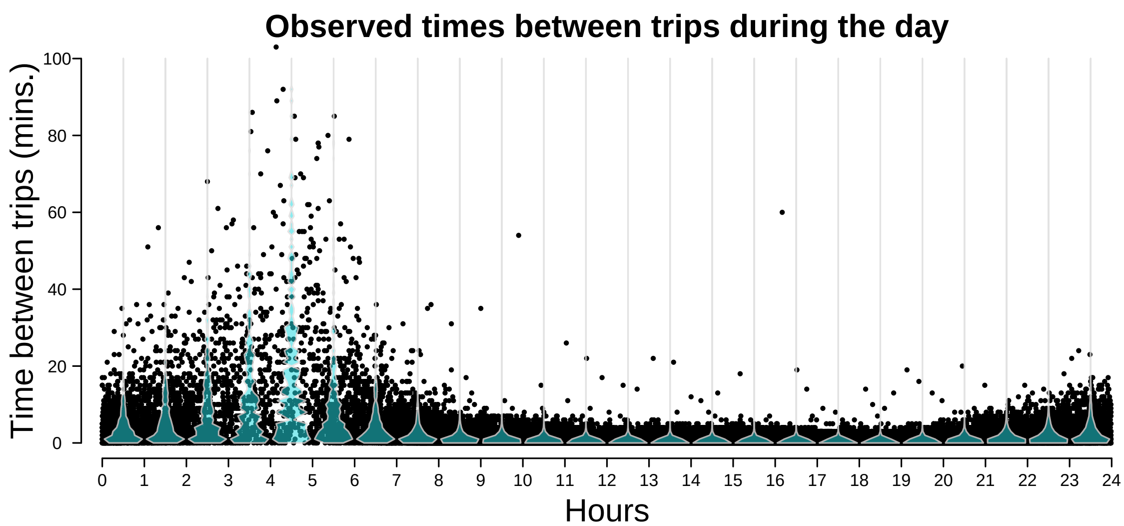

Figure 1: Inter-arrival times of each trip. During the early hours in the morning trips are infrequent and there are long gaps between them, but during business hours the opposite pattern arises. The blue histograms suggest the inter-arrival times are approximately Exponential within each hour.

After a more detailed analysis by hour an interesting characteristic of the inter-arrival distribution becomes evident. The blue histograms in Figure 1 suggest that the inter-arrival times approximately follow an Exponential distribution within each hour of the day. This finding suggests a possible direction to model bike trips using Poisson processes, to be described below.

[3] `R` code for data handling procedures, plots, and statistical analyses can be found in the GitHub repository of this project: `https://github.com/JLBaroja/BikeTrips`.

Methods

To account for the findings from our exploratory analyses we devised three models to capture the non-homogeneous nature of inter-arrival times of bike trips throughout the day. These three models share the assumption that bike trips accumulate accordingly to an homogeneous Poisson process within each hour, and thus their inter-arrival times can be modeled using hour-specific Exponential distributions. We start with a very simple model and then move forward adding more complexity each time.

Model $\mathcal{H}$ ($\mathcal{M}_{\mathcal{H}}$) - Inter-arrival Rates *by Hour*. The simplest model calculates 24 Exponential rates, each corresponding to a distinct hour of the day under the assumption of homogeneity within each hour. This approach serves as a foundation for understanding the variability in inter-arrival trip times throughout the day and provides the basis for the models that follow.

$$\begin{aligned} &\mathcal{M}_{\mathcal{H}} \\ \lambda_h &\sim Uniform(0, 3) \\ I_{j,h} &\sim Exponential(\lambda_h), \end{aligned}$$

where λ_h is the rate of inter-arrivals during hour h and $I_{j,h}$ is the inter-arrival of the jth trip that occurred within hour h.

Model $\mathcal{D}$ ($\mathcal{M}_{\mathcal{D}}$) - Rates by *Hour and Day*. Building on $\mathcal{M}_H$, the second model extends it by incorporating variability specific to each day of the week. Specifically, this model calculates 24×7 exponential rates, one for each hour of each day. By considering weekly variations we aim to uncover potential changes in trip behavior between business days and weekends, amid other possibilities.

$$\begin{aligned} &\mathcal{M}_{\mathcal{D}} \\ \lambda_{h,d} &\sim Uniform(0, 3) \\ I_{j,h,d} &\sim Exponential(\lambda_{h,d}) \end{aligned}$$

Model $\mathcal{S}$ ($\mathcal{M}_{\mathcal{S}}$) - Time-Dependent *Sinusoidal Rates*. $\mathcal{M}_{\mathcal{S}}$ introduces a tentative approach by explicitly modeling the rate of each hour as a function of time, again respecting variability between days. The core of this model is a double sinusoidal function to capture the intricate changes in rates throughout the day. The idea is that the first sinusoidal wave may capture the largest oscillations observed in the data and identified by models $\mathcal{M}_{\mathcal{H}}$ and $\mathcal{M}_{\mathcal{D}}$ (described in detail below), while the second wave attempts to characterize more subtle differences between weekdays and weekends that are evident when analyzing the results of the first two models. Even though the complexity of this model makes the interpretation of its parameters very challenging, our hope is that it may offer valuable insights into the intricate dynamics of the inter-arrival times of bike trips.

$$
\begin{aligned}
&\mathcal{M}_{\mathcal{S}}\\
\alpha_1 &= -1\\
\omega_1 &= 1/24\\
\omega_2 &\sim Uniform(1/24, 4/24)\\
\alpha_{2_d} &\sim Uniform(-0.1, 0)\\
\delta_{1_d} &\sim Uniform(0, 1)\\
\delta_{2_d} &\sim Uniform(1, 3)\\
\beta_d &\sim Uniform(0, 1)\\
\lambda_{h,d} &= \alpha_1 \sin(2\pi w_1 h + \delta_{1_d}) +\\
&+ \alpha_{2_d} \sin(2\pi w_2 h + \delta_{2_d}) +\\
&+ \beta_d\\
I_{j,h,d} &\sim Exponential(\lambda_{h,d}),
\end{aligned}
$$

where the constants across days, α_1 and ω_1, set the scale of the amplitude of the first, big oscillation of inter-arrival rates, and the assumption that such oscillation occurs once every 24 hours, respectively. ω_2 is also shared across days, but unobserved. This parameter represents the frequency of the second, smallest oscillation, which can take values between 1 and 4 times a day (to account for breaks from work and meals, for example). α_{2_d} corresponds to the amplitude of the second oscillation relative to the first, while δ_{1_d} and δ_{2_d} are measures of displacement of both waves relative to each other. Finally, β_d can be interpreted as the average inter-arrival rate of each day. Note that $\mathcal{M}_{\mathcal{S}}$ includes some parameters common to all days but also day-specific parameters with the objective of characterizing the differences between days exposed by the first two models in terms of latent, unobserved variables that control the hourly dynamics of inter-arrival times.

All three models were implemented within the Bayesian modeling framework and computationally calculated using JAGS (Just Another Gibbs Sampler) (Plummer, 2003). By employing Bayesian methodology we benefit from its flexibility in informing the complex models under consideration while properly handling their uncertainty with respect to the data. The details of each implementation, such as measures of convergence and MCMC sampling parameters, are omitted in this report for brevity and clarity, but are publicly available in the GitHub repository referred above. In general, the results reported in this work meet standard criteria regarding chain convergence.

Results

Findings of $\mathcal{M}_{\mathcal{H}}$ and $\mathcal{M}_{\mathcal{D}}$

Figure 2 presents the main result of models $\mathcal{M}_{\mathcal{H}}$ and $\mathcal{M}_{\mathcal{D}}$. The dashed lines connect the percentiles $P_{0.025}$ and $P_{0.975}$ of each of the 24 posterior inter-arrival rates according to $\mathcal{M}_{\mathcal{H}}$. The dynamic pattern of these rates is consistent with the data: during the early hours of the day the inter-arrival rates are low but they increase during the afternoon. $\mathcal{M}_{\mathcal{D}}$ infers the same rates for each day separately. The colored polygons in Figure 2 present these results. Weekdays (Mon-Fri, represented by semi transparent polygons) are hardly distinguishable from one another as most of them present similar rate dynamics. Weekends, however, are clearly different in two important respects. First, while inter-arrival rates begin increasing around 7am during weekdays, they start increasing about one and a half hours later on Saturdays and Sundays. This temporal shift likely aligns with variations in commuting patterns and the start of workdays.

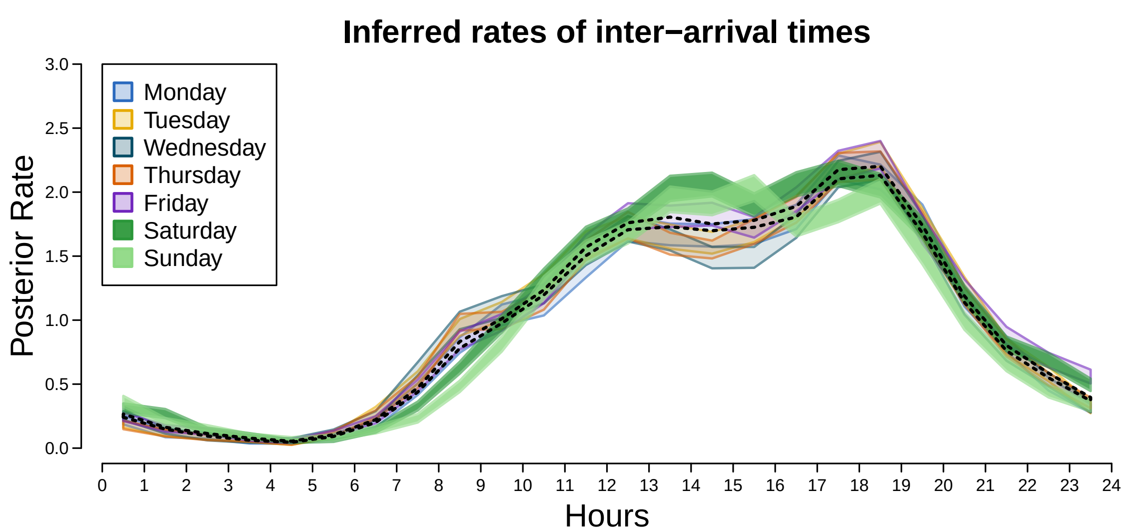

Figure 2: Inferred exponential rates λ_h according to models $\mathcal{M}_{\mathcal{H}}$ (dashed lines) and $\mathcal{M}_{\mathcal{D}}$ (colored polygons).

Secondly, during business days a distinct pattern emerges in the afternoon characterized by two “rest” periods between 8 and 11, in which the increase of inferred rates slows down, and then between 13 and 16 hours, when it actually reverses. These intervals potentially correspond to the commencement of work and lunch hours for many individuals. In contrast, weekends lack these discernible patterns, or at least they are attenuated, hinting at a greater diversity in the purposes behind bike trips during these days. Taken together, these results highlight the added information provided by $\mathcal{M}_{\mathcal{D}}$: by treating each day separately this model is able to identify subtle differences between weekends and business days not readily apparent in the raw data.

Findings of $\mathcal{M}_{\mathcal{S}}$

Figure 3 provides a summary of the sinusoidal model results through histograms of some of its posterior distributions.

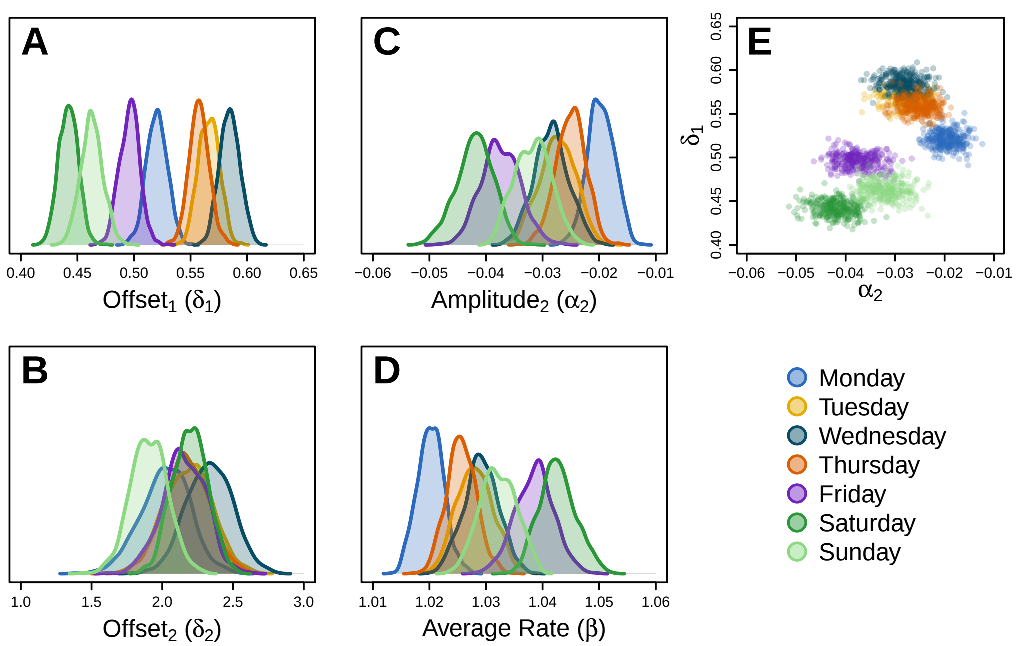

Figure 3: Results of the sinusoidal model $\mathcal{M}_{\mathcal{S}}$. Weekends and Fridays are grouped together in the joint posterior distribution of most day-specific parameters. This grouping is consistent with the results of model $\mathcal{M}_{\mathcal{D}}$ and with the raw data.

In Panel A, posterior distributions depict the offset of the first sine component, revealing clear groupings between weekdays and weekends. Panel B displays posterior distributions over the offset of the second sine, showing no discernible differences between days. Panel C illustrates the amplitude of the second sine, while Panel D presents the average rates. Both panels C and D reaffirm the distinction between weekends and Fridays compared to the rest of the week. This observed grouping is further evidenced in Panel E, which depicts a joint posterior distribution between two example nodes. Weekends and Fridays occupy distinct regions in the joint posterior compared to the rest of the week.

It's worth noting that the findings in Panel D of Figure 3 align with Table 1, which enumerates the total number of trips recorded each day of the week. Notably, Fridays and Saturdays consistently record significantly more trips than other days, a trend consistent with the $\mathcal{M}_{\mathcal{S}}$'s identification of them as days with the largest average inter-arrival rates.

In summary, even though interpreting the parameters in model $\mathcal{M}_{\mathcal{S}}$, along with their scales, is not immediate, the fact that this complex model clearly separates weekdays from weekends supports the idea of fundamental differences in the processes controlling bike trip dynamics between those types of days.

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**
Trips	17954	18752	18084	18320	19303	20833	17507

Table 1: Total number of trips each day of the week. Fridays and Saturdays have approximately 5% and 10% more trips than other days, respectively.

Discussion

The proposed models successfully captured the main pattern in inter-arrival times and identified subtle distinctions across days of the week. This information could be valuable for users, aiding in trip planning, and for the service provider, for example in optimizing infrastructure maintenance in specific days and hours.

The key limitation in the models discussed above is the loss of information due to the assumption of homogeneity within each hour. More refined approaches could model smoother changes in rates, possibly at minute or even second level.

While the third model's conclusions seem sensitive, its complex (and even chaotic) nature and arbitrary prior choices highlight the challenge of implementing this type of models with Bayesian tools and the importance of looking for theoretically motivated models and priors for more meaningful interpretations.

A possible extension to this set of analyses would involve exploring patterns at individual bike stations just as we did with respect to days of the week. This could reveal station-specific dynamics since, for example, the successive trip starts may behave differently in distant areas of the city. Understanding station-level variations could also be useful in better planning user trips and maintaining Metro Bike Share's infrastructure. Needless to mention, applying the current analyses to data sets from different bike trip providers could benefit both the general conclusions about trip dynamics and the specific needs of each provider.

Furthermore, the data set at hand is well suited for diverse analyses and extensions. Variables such as trip duration and bike trajectories between stations, can offer additional layers of information. For example, modeling bike routes and transitions between stations as discrete finite space Markov chains could provide valuable information on bike traffic patterns, aiding in optimizing station maintenance and understanding the flow of bikes across the city.

Conclusion

In summary, while the present analyses have explored trip frequency dynamics and their variations throughout the day and throughout the week, more intricate analyses could build on them and provide crucial information for future research, paving the way for a deeper understanding of bike-sharing behavior.

References

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing. March 20–22, Vienna, Austria. ISSN 1609-395X*.