2. Data Analysis

The main motivation for data analysis is to provide input to the simulation model and check whether the output matches the real time output of the system. The inter arrival time was used as input, and queue length and the average time spent by the vehicle in the node was calculated to validate the output.

2.1 Data Received

The data provided to us by the traffic department of Magdeburg consists of inflow and outflow of vehicles in all of the 5 arms of the node. The data was recorded on a weekday (Tuesday), from 6am to 7pm in 2009. The data contains 52 samples, each sample size being 15 minutes. Sample includes individual vehicle counts, total vehicles, car units and vehicles per hour. Vehicle counts include Cars, Trucks, Buses and other motor vehicles, but does not include any information about Trams, pedestrians or cycles. The basic node and traffic flow representation drawn from the given data is as shown in the figure below.

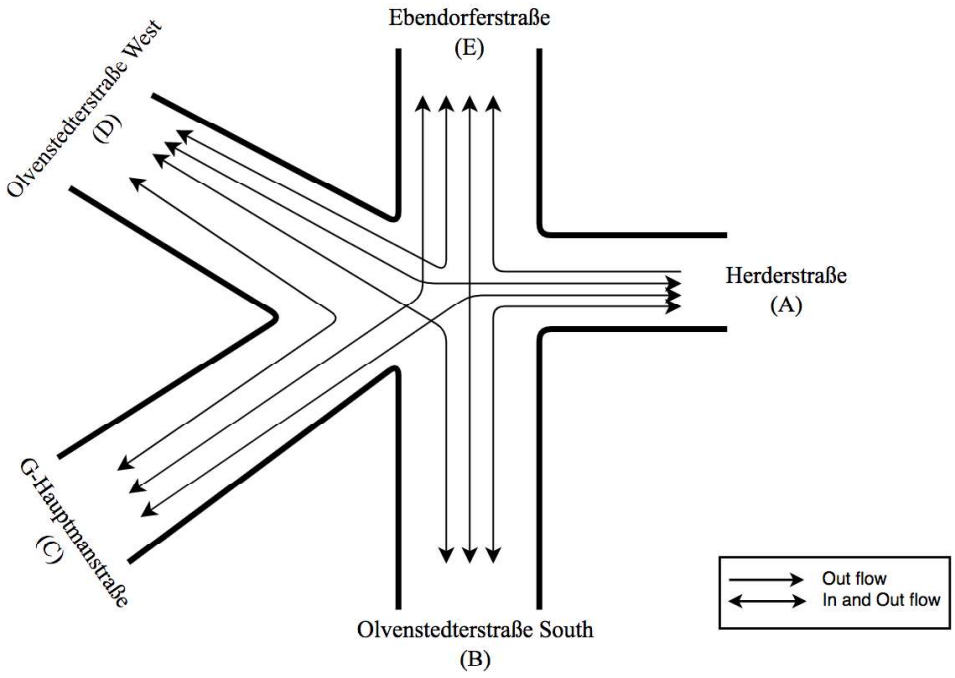

Figure 07: Basic node and Traffic flow representation

2.2 Data Quality

The Data quality will be measured using the below four metrics:

- **Completeness** - the data is complete with respect to vehicle flow in the node and accident statistics, but does not provide inter-arrival times, signal times, pedestrian data or tram frequency.
- **Uniqueness** - there is quite some amount of redundant data, such as motor vehicle count, total vehicle count, vehicle count per hour, total vehicle count per hour which are repetitions of the same data.
- **Validity** - the provided data is more than a decade old and might not be valid for the current condition of the node.
- **Consistency and Accuracy** - the provided data is assumed to be consistent and accurate when it was recorded as it is obtained from a single source.

2.3 Comparison between data collected in 2009 and 2021

There is a significant difference between the vehicle flow data given to us by the traffic department of Magdeburg, collected in the year 2009 and the current node situation. As of in 2021, two major factors affect the Wilhelmstedter platz traffic data, one of which is the ongoing pandemic condition that leads to reduction in the overall traffic density throughout the city and the second one is the Olvenstedter Straße road block due to construction, that restricts the vehicle movement in that direction. This comparison is drawn in the table below for our understanding of the scenario.

Street	Vehicle Outflow (2009)	Load Percentage (%)	Vehicle Outflow (2021)	Load Percentage (%)
Herder Str	92	8.52	67	16.54
Olvenstedter Str South	433	40.12	64	15.80
Gerhart-Hauptmann Str	131	12.14	123	30.37
Olvenstedter Str West	196	18.16	59	14.56
Ebendorfer Str	227	21.03	92	22.71
Total	1079		405	

Table 01: Comparison b/w 2009 and 2021 vehicle flow during peak hours

From the table we can see that there is a large amount of load shift among the roads and the old traffic data does not hold good for the present day scenario. Hence there is a necessity to collect new data from the node.

2.4 Data Collection

For the purpose of analysis, the data was required to be as updated and recent as possible, as significant variations were observed in the traffic behaviour due to factors like construction near the Haupt-bahnhof and the ongoing pandemic. Therefore essential data such as inter-arrival times of the vehicles and signal readings was collected by our team in real time with the help of Excel Macros feature, tram and bus readings were collected conveniently from the INSA app. To get maximum approximation, the data was collected in the morning and evening hours at the time where traffic was at its peak in the node. The signal data had no specific pattern that could be observed as they were affected by tram arrivals and pedestrian crossings at random.

2.4.1 Input Data

The data that was required as input to the simulation model are :

- Inter-arrival time
- Turn probabilities
- Signal phases

2.4.2 Output Data

The data that was output from the model and used for validating the model are :

- Queue length
- Average time spent by the vehicle in the node

The real-time values for the output variables can be found in the validation section.

2.5 Signal phases

In order to design the signals as similar to the real world as possible, we calculated the duration of the signals in each phase during data collection. The traffic phases control the respective stop lines in that particular phase.

Figure 08: Signal phases

2.6 Probability Distribution Function Estimation

In order to make a right guess of the distribution of cars, we have drawn histograms of number of cars vs inter-arrival time for every cross road in the node. This can be observed in figures 3, 6, 9, 12 and 15. By comparing the histograms obtained, with various probability distribution functions, we guessed that the distribution is exponential. This was done for all the 5 streets and we had all the distributions guessed as exponential. In order to check the correctness of our assumption quantile-quantile plots were drawn, which can be seen in figures 4, 7, 10, 13 and 16. Since the q-q plots obtained approximate to a straight line and the plots pass through the origin, the probability distribution function estimation was found to be correct. We tested our guessed distribution with a chi-square test, that of which can be seen in figures 5, 8, 11, 14 and 17 and its corresponding values in the tables 2, 4, 6, 8 and 10.

Apart from the distribution data, turn probabilities of vehicles in each street were calculated in order to design a system with traffic flow as close to the real world node as possible. These probabilities can be observed in the tables 3, 6, 9, 12 and 15.

Herderstraße (A)

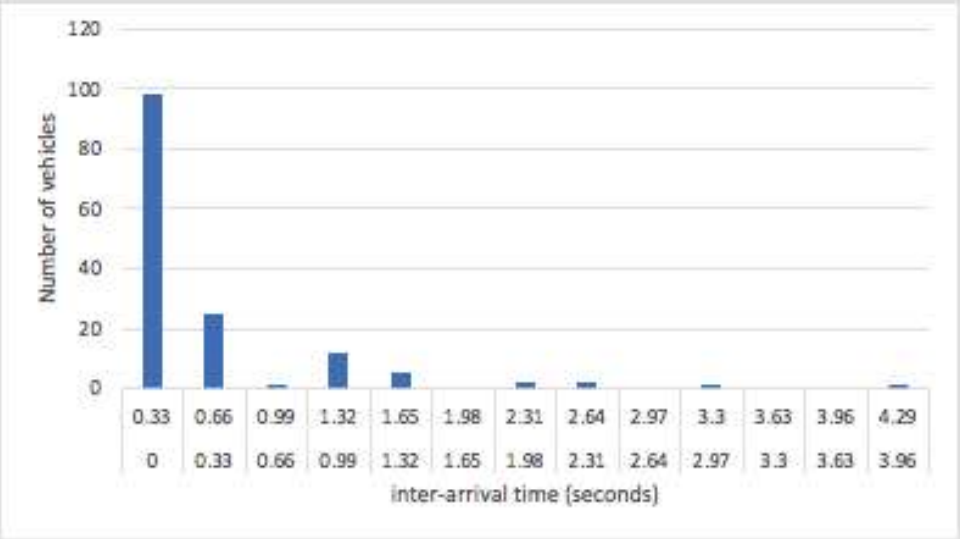

Figure 09: Histogram for vehicle inter-arrival times in Herderstraße

Distribution	Exponential distribution
Mean	Lambda
0.390	2.5644

Table 02: Distribution and parameters of Herderstraße

Olvenstedter Straße (west)	Ebendorfer Straße	Olvenstedter Straße (south)	Gerhart Hauptmann Straße
0.3214	0.25	0.0178	0.4107

Table 03: Turn probabilities of vehicles flowing from Herderstraße

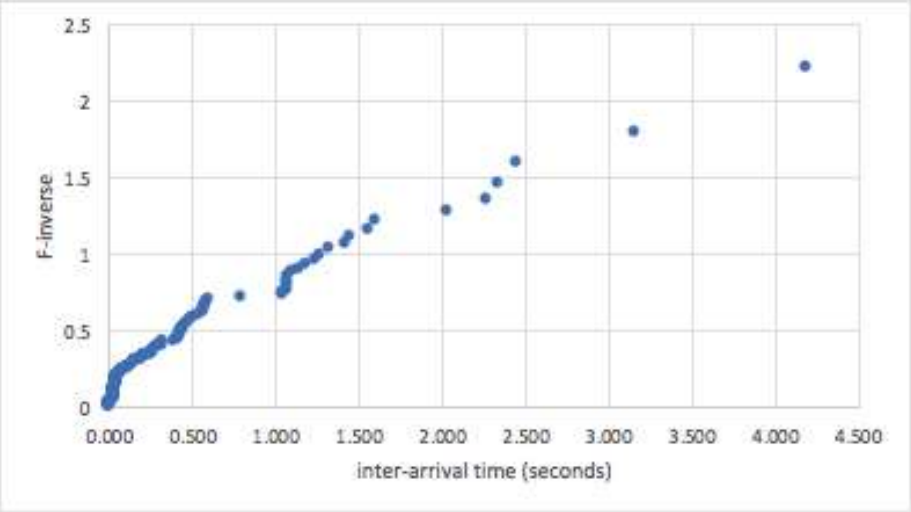

Figure 10: Q-Q plot for Herderstraße

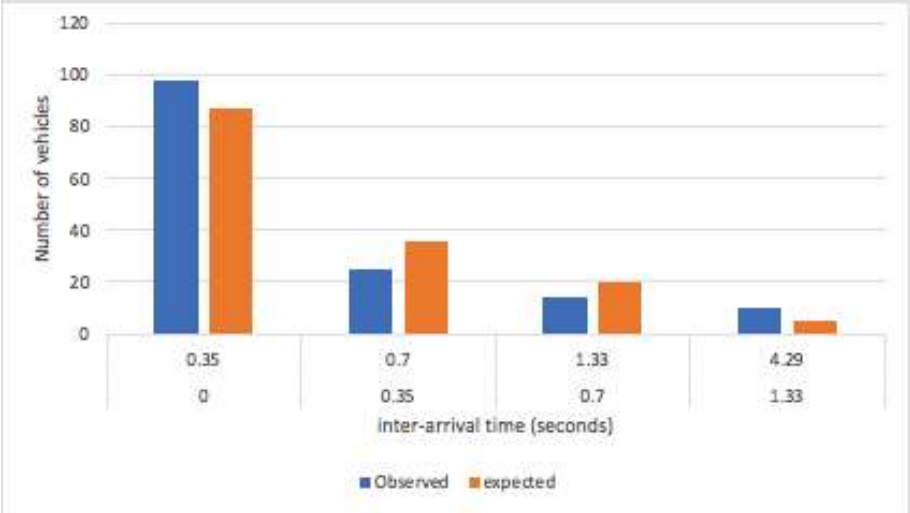

Figure 11: Chi square test for Herderstraße

Number of data points, N	147
Degrees of freedom, f	2
Alpha, a	0.99
χ^2_0	11.5169
$\chi^2_{f,a}$	9.2103

Table 04: Chi square test values for Herderstraße

Olvenstedter Straße South (B)

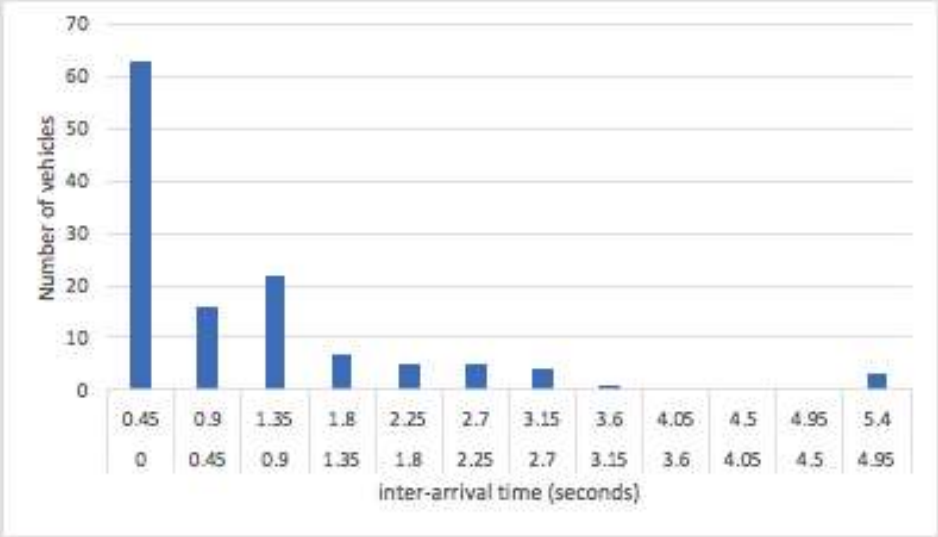

Figure 12: Histogram of vehicle inter-arrival times in Olvenstedter Straße (south)

Distribution	Exponential distribution
Mean	Lambda
0.86	1.1639

Table 05: Distribution and parameters of Olvenstedter Straße (south)

Olvenstedter Straße (west)	Ebendorfer Straße	Herderstraße	Gerhart Hauptmann Straße
0.2619	0.6031	0.1269	0.0079

Table 06: Turn probabilities of vehicles flowing from Olvenstedter Straße (south)

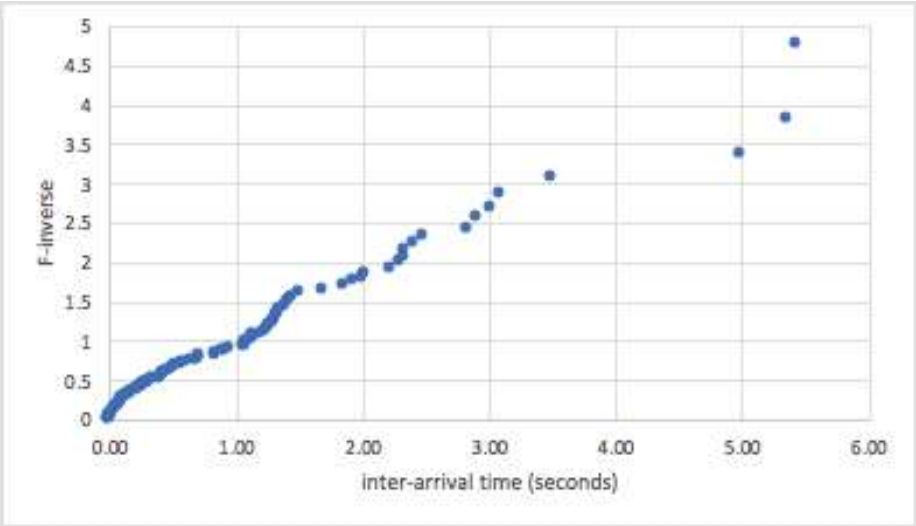

Figure 13: Q-Q plots for Olvenstedter Straße (south)

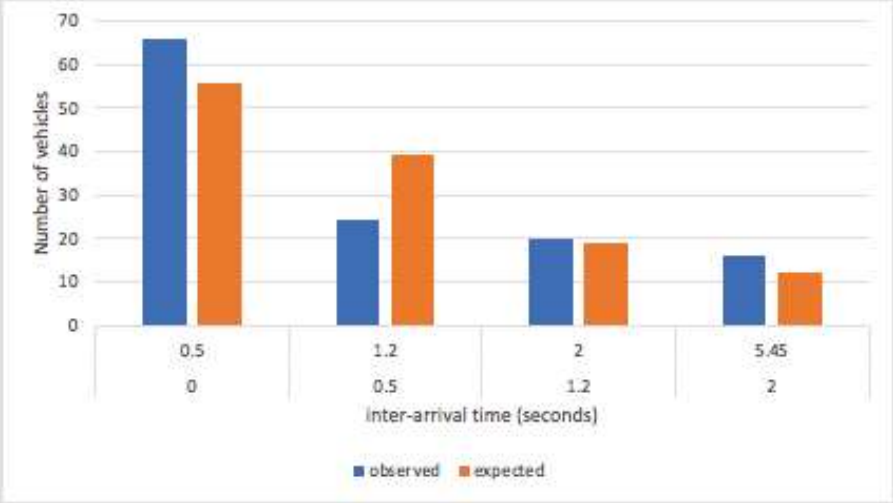

Figure 14: Chi square test for Olvenstedter Straße (south)

Number of data points, N	126
Degrees of freedom, f	2
Alpha, a	0.99
χ^2_0	9.2145
$\chi^2_{f,a}$	9.2103

Table 07: Chi square test values for Olvenstedter Straße (south)

Gerhart-Hauptmann Straße (C)

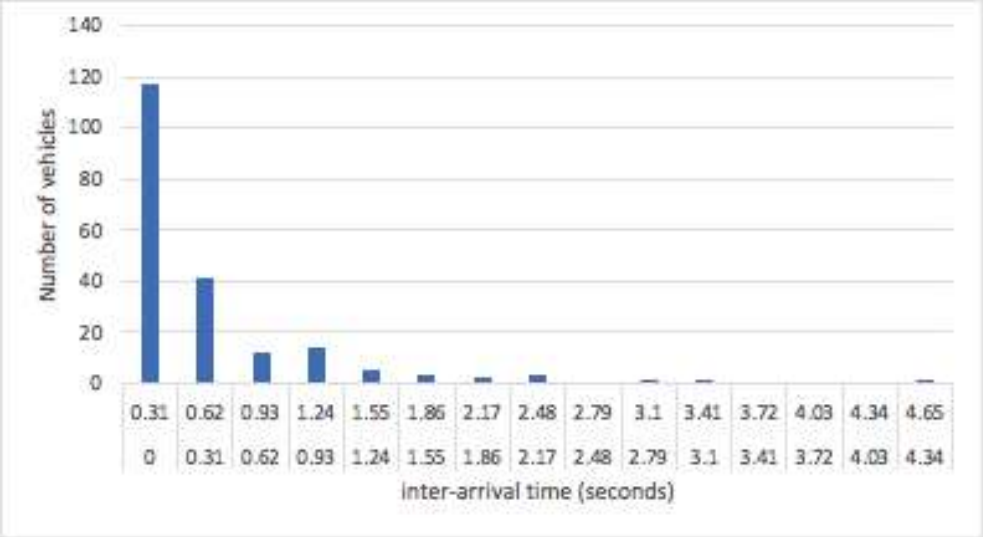

Figure 15: Histogram of vehicle inter-arrival times in Gerhart-Hauptmann Straße

Distribution	Exponential distribution
Mean	Lambda
0.4615	2.1669

Table 08: Distribution and parameters of Gerhart-Hauptmann Straße

Olvenstedter Straße (west)	Ebendorfer Straße	Herderstraße	Olvenstedter Straße (south)
0.0366	0.6788	0.2293	0.0550

Table 09: Turn probabilities of vehicles flowing from Gerhart-Hauptmann Straße

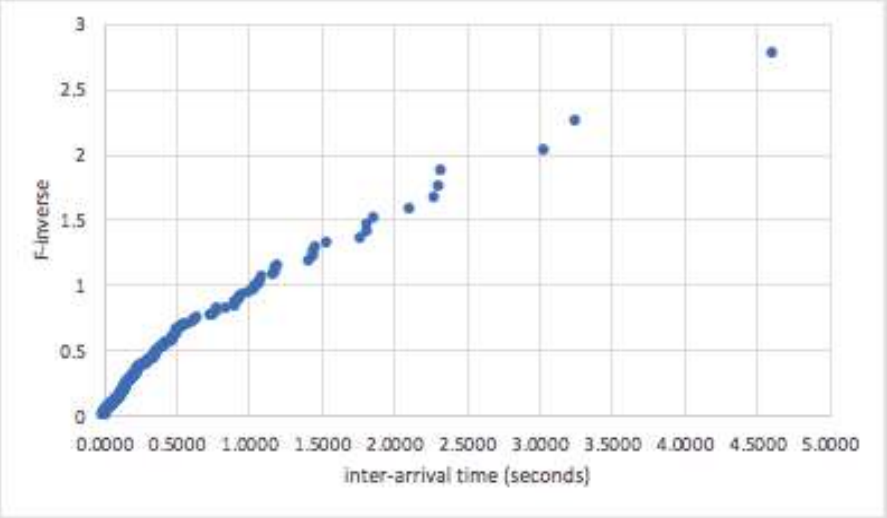

Figure 16: Q-Q plots for Gerhart-Hauptmann Straße

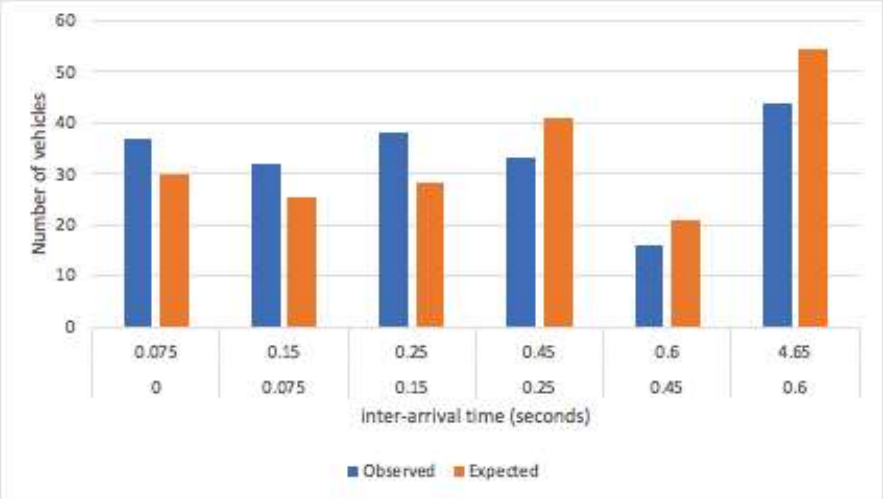

Figure 17: Chi square test for Gerhart-Hauptmann Straße

Number of data points, N	200
Degrees of freedom, f	4
Alpha, a	0.98
χ^2_0	11.4488
$\chi^2_{f,a}$	11.6678

Table 10: Chi square test values for Gerhart-Hauptmann Straße

Olvenstedter Straße West (D)

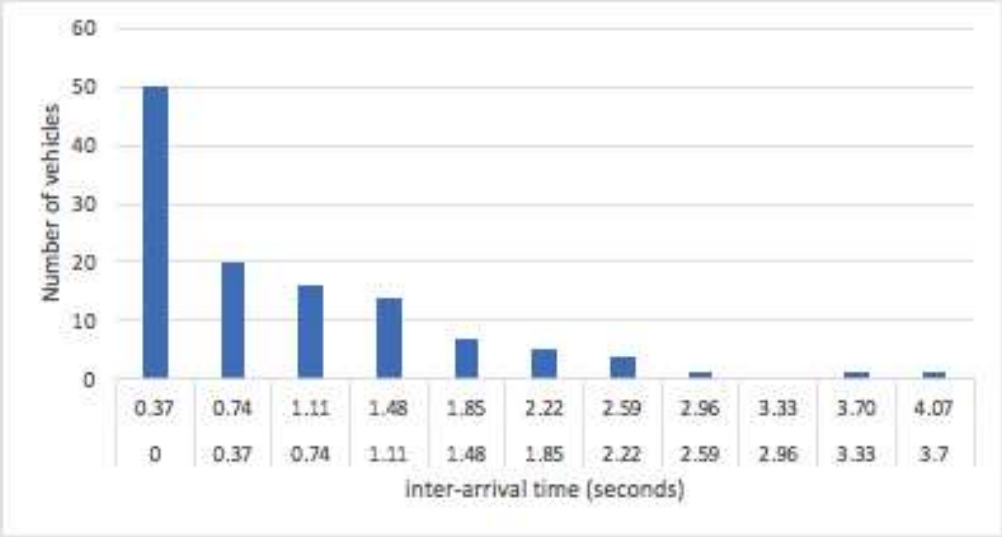

Figure 18: Histogram of vehicle inter-arrival times in Olvenstedter Straße (west)

Distribution	Exponential distribution
Mean	Lambda
0.77	1.3045

Table 11: Distribution and parameters of Olvenstedter Straße (west)

Ebendorfer Straße	Herderstraße	Olvenstedter Straße (south)	Gerhart Hauptmann Straße
0.4227	0.3658	0.1544	0.0569

Table 12: Turn probabilities of vehicles flowing from Olvenstedterstraße (west)

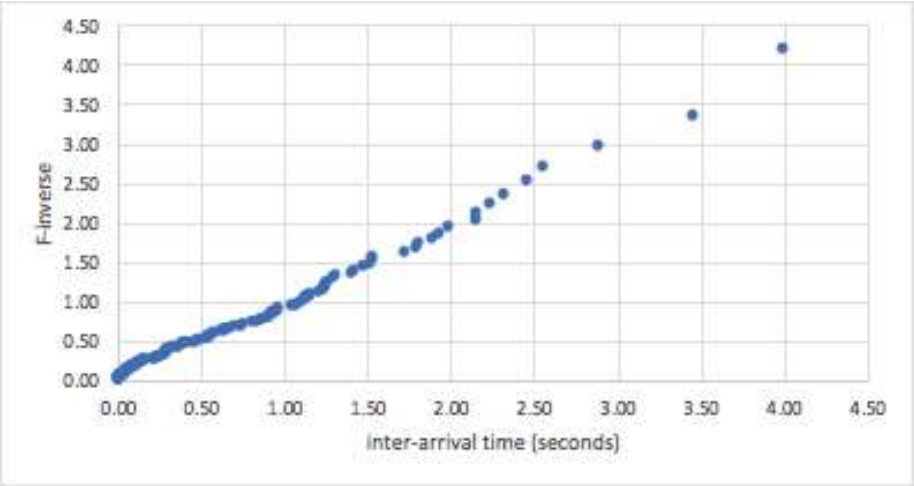

Figure 19: Q-Q plot for Olvenstedter Straße (west)

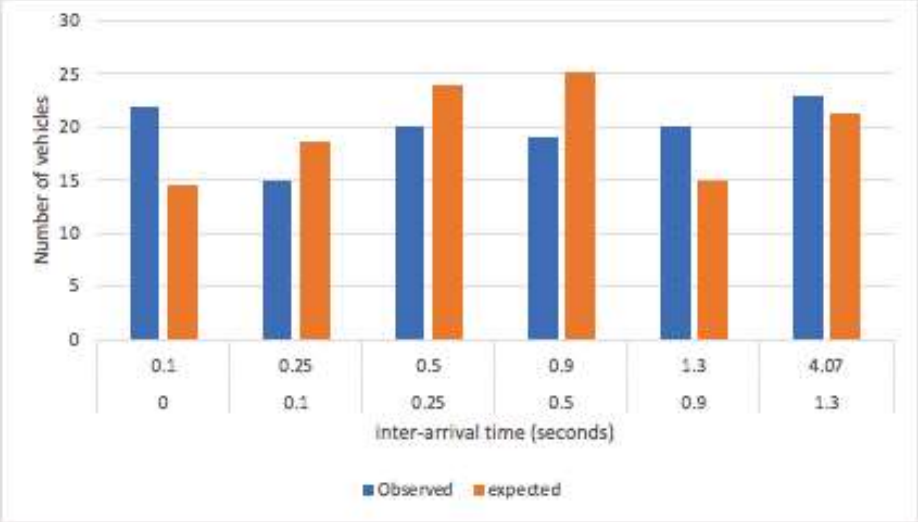

Figure 20: Chi square test for Olvenstedter Straße (west)

Number of data points, N	119
Degrees of freedom, f	4
Alpha, a	0.93
χ^2_0	8.5031
$\chi^2_{f,a}$	8.6664

Table 13: Chi square test values for Olvenstedter Straße (west)

Ebendorfer Straße (E)

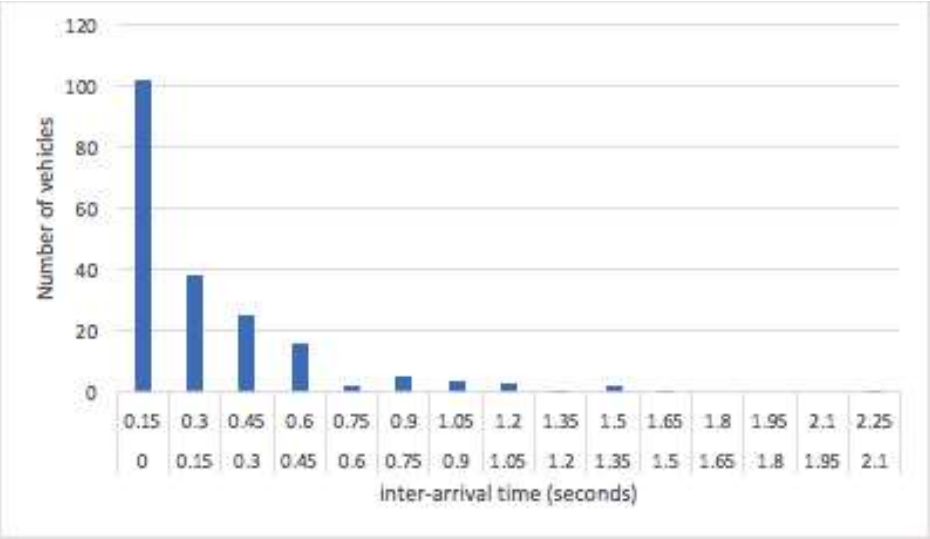

Figure 21: Histogram of vehicle inter-arrival times in Ebendorfer Straße

Distribution	Exponential distribution
Mean	Lambda
0.2562	3.9025

Table 14: Distribution and parameters of Ebendorfer Straße

Olvenstedter Straße (west)	Herderstraße	Olvenstedter Straße (south)	Gerhart Hauptmann Straße
0.2405	0.0168	0.3037	0.4388

Table 15: Turn probabilities of vehicles flowing from Ebendorfer Straße

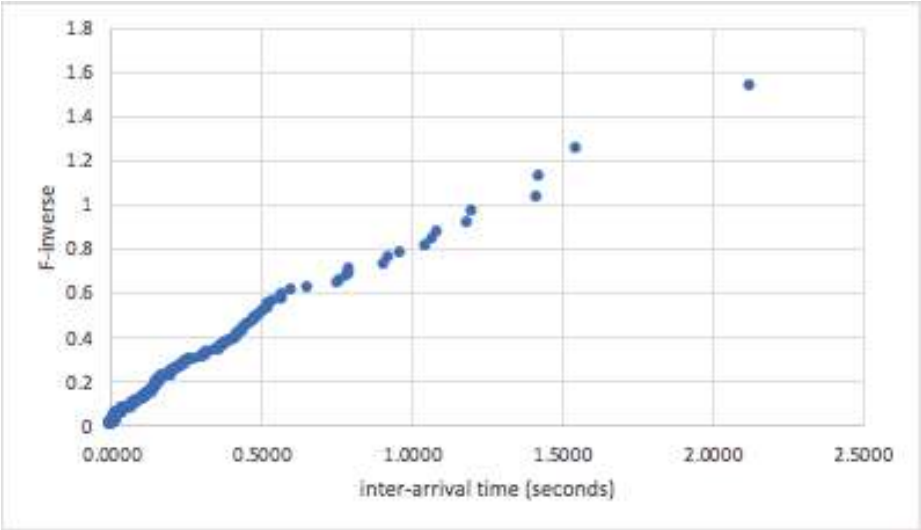

Figure 22: Q-Q plot for Ebendorfer Straße

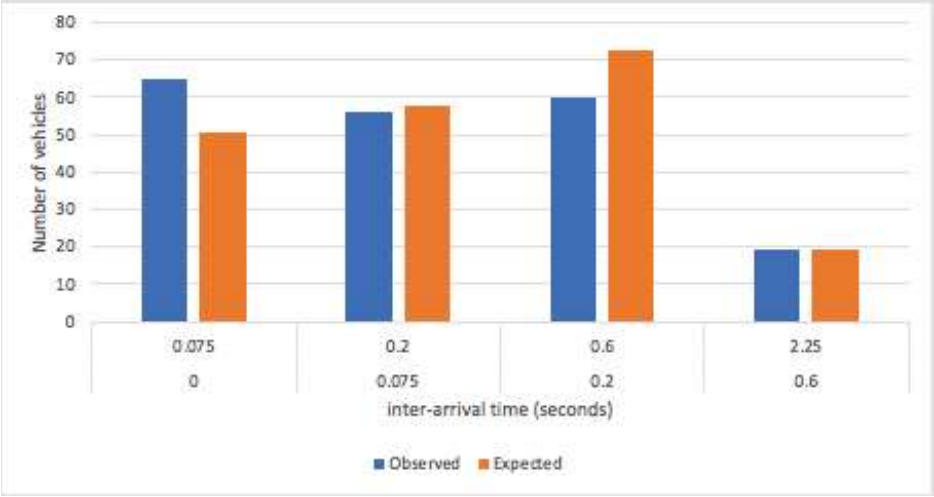

Figure 23: Chi square test for Ebendorfer Straße

Number of data points, N	200
Degrees of freedom, f	2
Alpha, a	0.96
χ^2_0	6.1721
$\chi^2_{f,a}$	6.4377

Table 16: Chi square test values for Ebendorfer Straße

2.7 Difficulties encountered while obtaining the data

- Bringing the team together to agree upon a common time for data collection.
- The team did not anticipate the frequent weather changes that led to delay in the data collection and the members had to spend additional hours into it.
- Recollecting of the data due to inconsistencies in the time of the day.
- Change of recording methods from stopwatch to Excel Macros for more accuracy and precision led to repeated data collection.
- Due to unavailability of all the members and five different cross arms, there was an increase of load on the members that led to small inaccuracies while recording.
- Deciding exactly what type of data was necessary to build the model and how to record the same.