INVESTIGATIVE ANALYSIS ON THE DATA OBTAINED FROM THE SENSORS WITHIN M25 ROAD
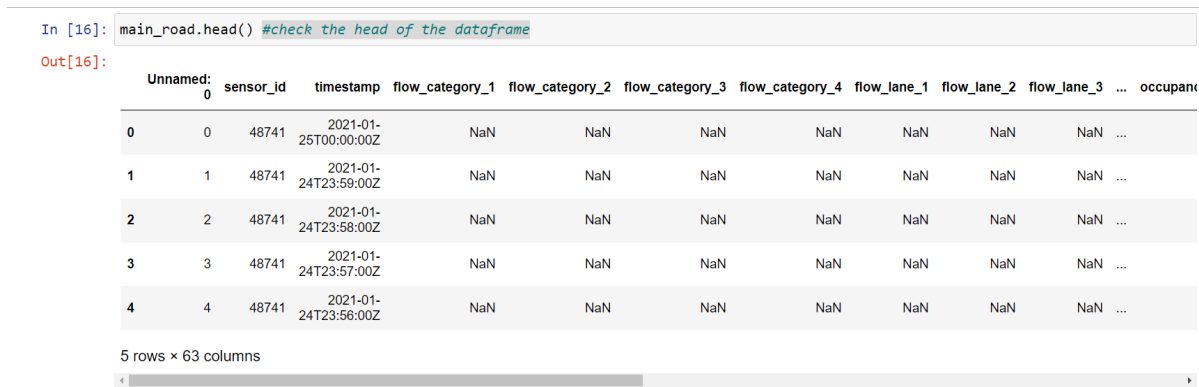
By

Anthony Adole

This report presents an investigation concerning the daily traffic performance on M25 road obtained from two different sensors positioned reasonably close to each other at the clockwise junction 10 towards A3 southbound. The data obtained was from 25th of January 2020 at 12 pm to 25th of January 2021 at 12 pm making. Therefore, the data was collated within 366 days for 8,784 hours. The sensor identification number and location on M25 road are as follows:

Sensor ID	Sensor position	Sensor location (lat,long)
49965	Slip road	51.3205395250799, -0.446298174454472
48741	main road	51.3184047579063, -0.441290074916838

The sensor positioned on the main road obtained data from three different lanes, while the sensor positioned on the slip road obtained data from two different lanes. Therefore, the total number of columns generated for the sensor data was 62 (figure 1). However, the data obtained were populated into 23 columns for the sensor positioned at the main road and its 16 columns for the sensor positioned at the slip road (.i.e *details can be obtained from the Jupyter notebook*).

Figure 1: Visual presentation of the loaded dataset in pandas dataframe.

The unit definition to enable the collation of data by the sensors are:

- Average Speed: measured in km/h
- Flow: flow rate of vehicles measured in number of vehicles per hour
- Category 1: vehicles less than 5.2m in length
- Category 2: vehicles between 5.2m and 6.6m in length
- Category 3: vehicles between 6.6m and 11.6m in length
- Category 4: vehicles more than 11.6m in length
- Headway: average time in seconds between vehicles
- Occupancy: percentage of time that the section of road is occupied by vehicles

Due to the nature of the roads at that location, as reported by highways England, the sensors could not populate a few columns. Therefore, the investigation shows a reduction in the number of columns populated with data by each sensor. Table 1 shows the list of the columns that were empty in the dataset:

Table 1: List of empty columns from both road

ROAD	EMPTY COLUMN
Main road	flow_category_1,flow_category_2,flow_category_3,flow_category_4,flow_lane_1, flow_lane_2,flow_lane_3,flow_lane_4,flow_lane_5,flow_lane_6,flow_lane_7,flow _category_1_lane_4,flow_category_1_lane_5,flow_category_1_lane_6,flow_categ ory_1_lane_7,flow_category_2_lane_4,flow_category_2_lane_5,flow_category_2_ lane_6,flow_category_2_lane_7,flow_category_3_lane_4,flow_category_3_lane_5, flow_category_3_lane_6,flow_category_3_lane_7,flow_category_4_lane_4,flow_c ategory_4_lane_5,flow_category_4_lane_6,flow_category_4_lane_7,average_spee d_lane_4,average_speed_lane_5,average_speed_lane_6,average_speed_lane_7,occ upancy_lane_4,occupancy_lane_5,occupancy_lane_6,occupancy_lane_7,headway_ lane_4,headway_lane_5,headway_lane_6,headway_lane_7
Slip road	flow_category_1,flow_category_2,flow_category_3,flow_category_4,flow_lane_1, flow_lane_2,flow_lane_3,flow_lane_4,flow_lane_5,flow_lane_6,flow_lane_7,flow _category_1_lane_4,flow_category_1_lane_5,flow_category_1_lane_6,flow_categ ory_1_lane_7,flow_category_2_lane_4,flow_category_2_lane_5,flow_category_2_ lane_6,flow_category_2_lane_7,flow_category_3_lane_4,flow_category_3_lane_5, flow_category_3_lane_6,flow_category_3_lane_7,flow_category_4_lane_4,flow_c ategory_4_lane_5,flow_category_4_lane_6,flow_category_4_lane_7,average_spee d_lane_4,average_speed_lane_5,average_speed_lane_6,average_speed_lane_7,occ upancy_lane_4,occupancy_lane_5,occupancy_lane_6,occupancy_lane_7,headway_ lane_4,headway_lane_5,headway_lane_6,headway_lane_7,flow_category_1_lane_ 3,flow_category_2_lane_3,flow_category_3_lane_3,flow_category_4_lane_3,avera ge_speed_lane_3,occupancy_lane_3,headway_lane_3

The following are some of the investigative questions used to perform the analysis of the data obtained on the two roads. The questions being posed are as follows**:**

1) How many HGVs (Heavy Goods Vehicles) turn left onto the slip road and how does their use of each lane vary throughout the day?
2) Do vehicle speeds vary from the slip road and the main road?
3) How does traffic flow vary for the two roads at 10pm for different days or perhaps the same day on different weeks?
4) Is the usage of the road lanes consistent at all times on the main road (e.g. does lane 1 always experience the most flows)?
5) What are typical flows between 10pm and 5am on the main road and what is the relationship between total flows and HGV flows (or HGV percentage)?

Question 1: How many HGVs (Heavy Goods Vehicles) turn left onto the slip road and how does their use of each lane vary throughout the day?

Analysis: from the data obtained for the slip road, HGV flow within the two lanes. Therefore making 12,144,720 as the total number of HGV on the slip road for the timestamp for the 366 days of data collection. However, from the graph presented in figure 2, lane 1 tends to be the most used lane within each hour of the day for 366 days of the data obtained.
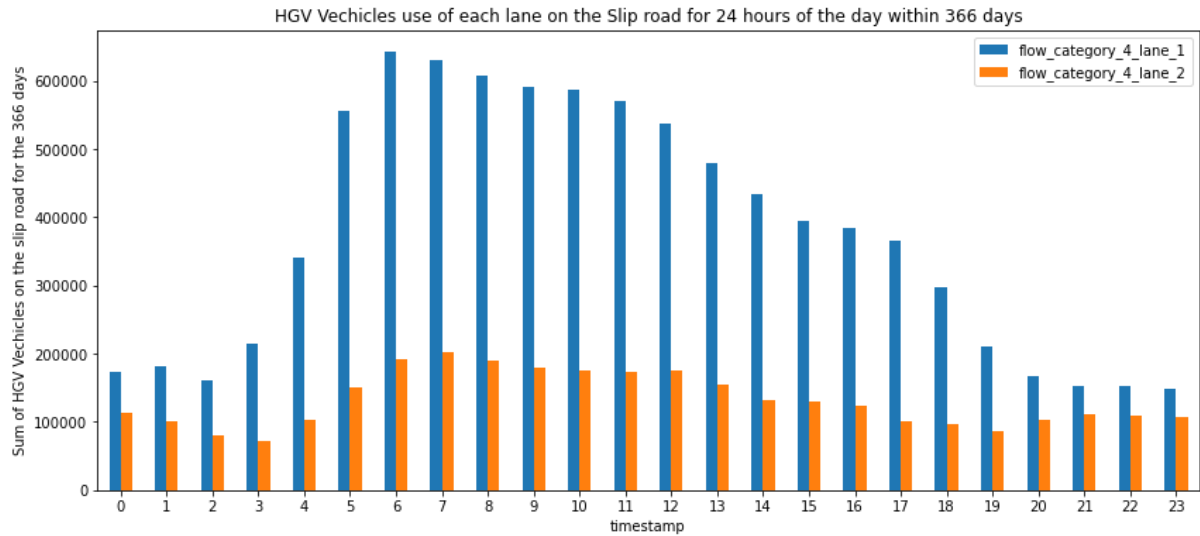
Figure 2: Use of Slip road by HGV for each Lane per hour for 366 days

However, from the data obtained it has been observed that lane 1 tends to have the most flow concerning the use of HGV for both roads within the hours of the day for the 366 days. Figure 3 shows the flow of HGV on the main road for the data obtain.
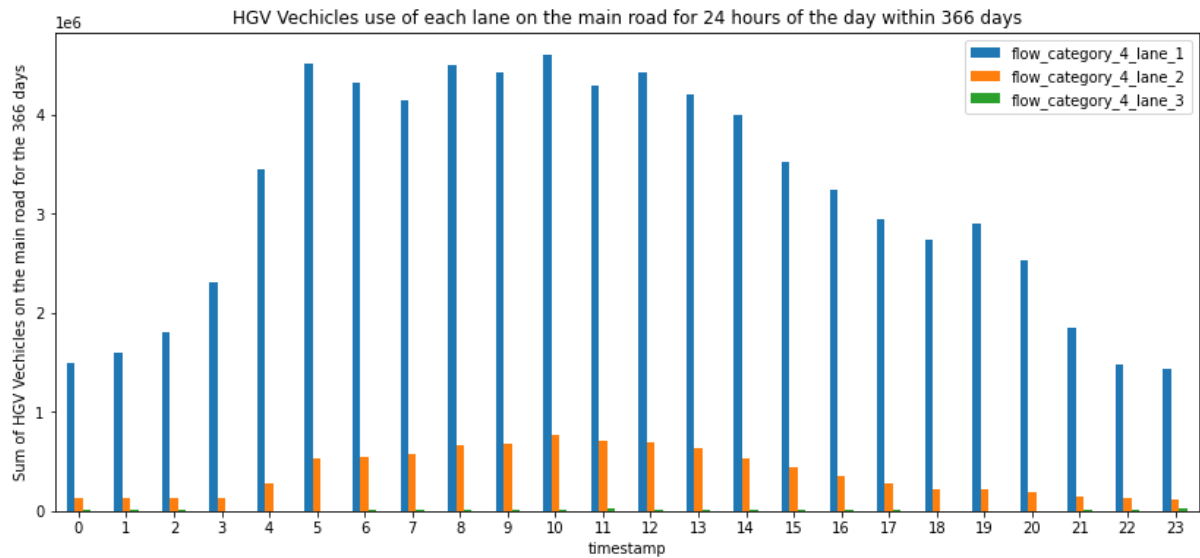

Figure 3: Use of main road by HGV for each Lane per hour for 366 days

Therefore, from figure 2, lane 1 of the slip road records the highest use of lane compared to lane 2 through out data collection time frame

Question 2: Do vehicle speeds vary from the slip road and the main road?

Analysis: From the data analysed concerning the speed for each road, the graph presented in figure 4 shows the total average speed for each road per lane usage for each hour of days the data was obtained. Therefore, the main road records the most speed compared to the slip road in figure 4. The graph shows that from 10p.m to 3a.m , lane 1 of the main road tends to have the most flow and from 4a.m to 11a.m, lane 2 shows the most flow. However, Lane 3 of the main road has the most flow from 12 noon to 4p.m of each day of the data collection.

Figure 4: Varying speed between each road per lane for each hour per day

Another analysis to confirm the finds from figure 4, was the varying speed of the days of week. However, the analysis presented in figure 5 shows that the main road still records most speed compared to the slip road. It also shows that lane 2 of the main road has most speed for each day of the week of data collection.

Figure 5: Varying speed for both roads per days of the week

Therefore, due to the graphs obtained from this analysis, it shows that the main road has the most speed compared to the slip road.

Question 3: How does traffic flow vary for the two roads at 10pm for different days or perhaps the same day on different weeks?

Analysis: For the different date of the data collection time frame, the traffic flow by 10p.m for each road per lane is presented in figure 6. This graphs shows that December 19th 2020, records the highest flow on the main road for category 1 on lane 2 while category 1 on lane 1 of the main road was the next with the most record of traffic flow on same day.
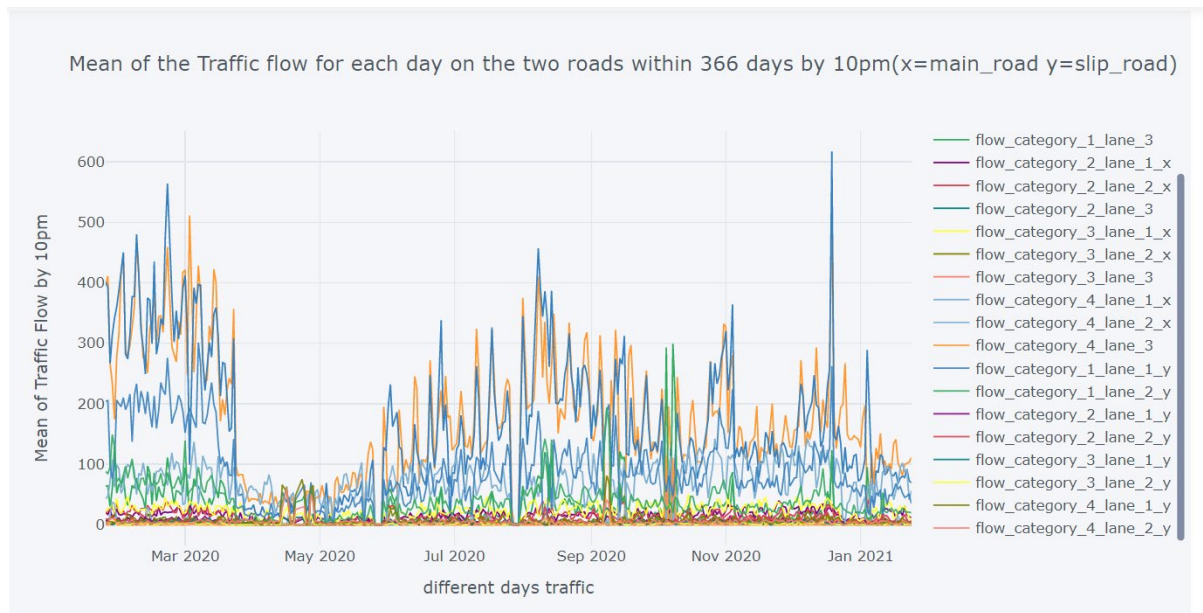
Figure 6: Traffic flow for different date within the data collection time frame

However, for the traffic flow for days of the week by 10p.m presented in figure 7 as a graph, Saturdays records the highest flow of traffic on the main road compared to the slip road. The traffic flow on Saturdays by 10p.m for the vehicles in category 1 on lane 1 and lane 2 of the main road appears to be the most busy lanes. However, vehicles in category 1 lane 1 tends to have the highest traffic flow based on the average traffic flow within the data collection time frame.
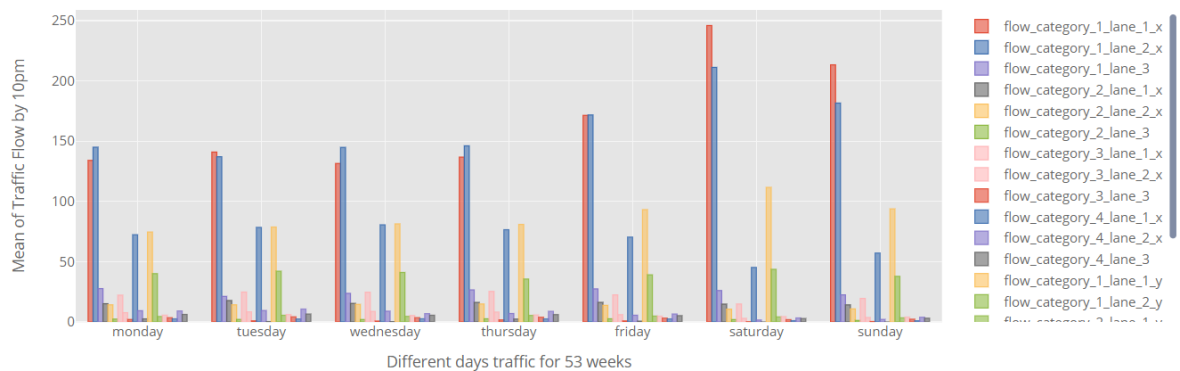

Figure 7: Traffic flow per days of week within the data collection time frame

However, figure 8 presents a line graph showing the traffic flow by 10p.m for Sunday of each week within the time frame of data collection.
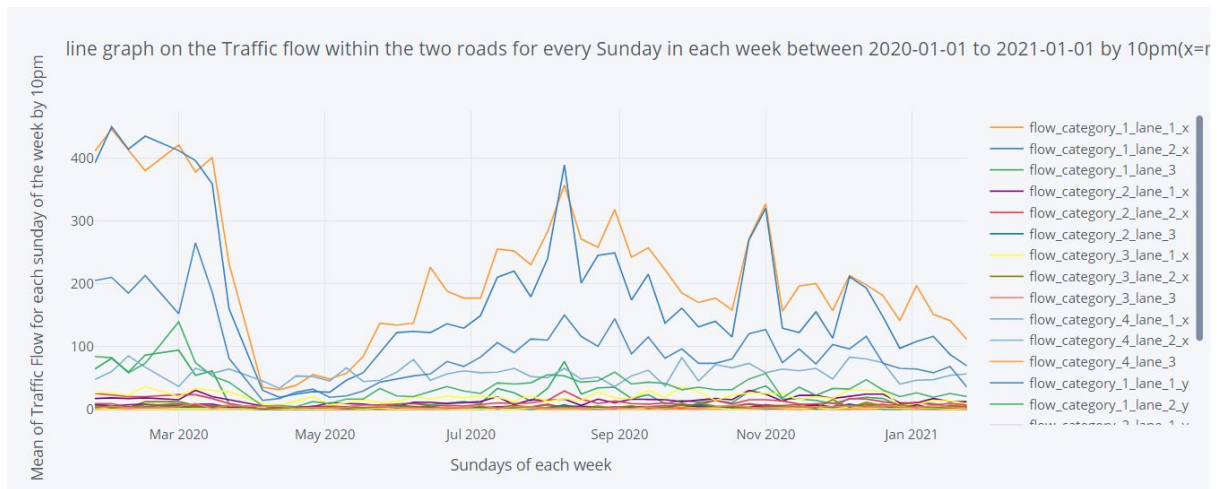
Figure 8: Sunday traffic flow by 10p.m within 366 days of data collection

Question 4: Is the usage of the road lanes consistent at all times on the main road (e.g. does lane 1 always experience the most flows)?

Analysis: this analysis is performed on both the occupancy of the road and the categories of vehicles that uses the road within the hour of each days for the 366 days of data collection.

From the graph presented in figure 9, the usage of the main road per occupancy of vehicles on each lane shows that lane 1 of the main road records the highest usage within the time frame of data collection.
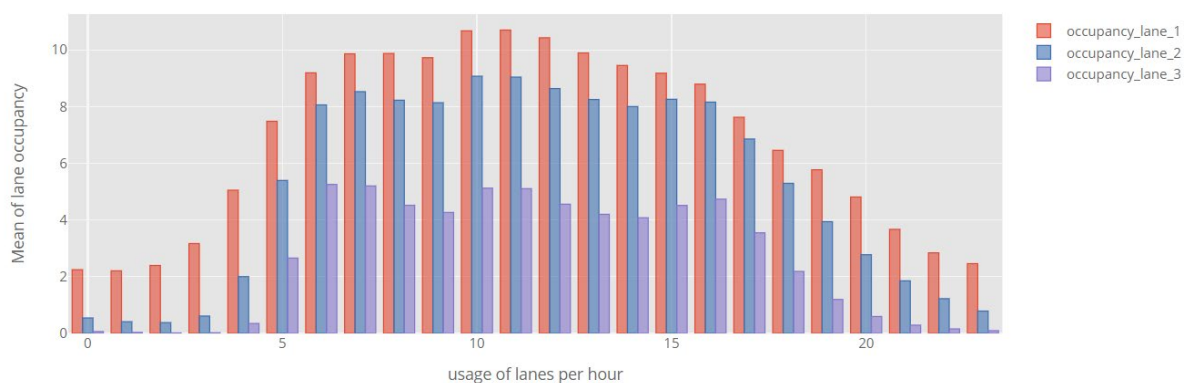

Figure 9: Occupancy of main road lanes throughout the hours of the day within the time frame of data collection.

Figure 10 shows that the Vehicles in category 1 records the highest usage of the main road within the hour of each days for the 366 days of data collection
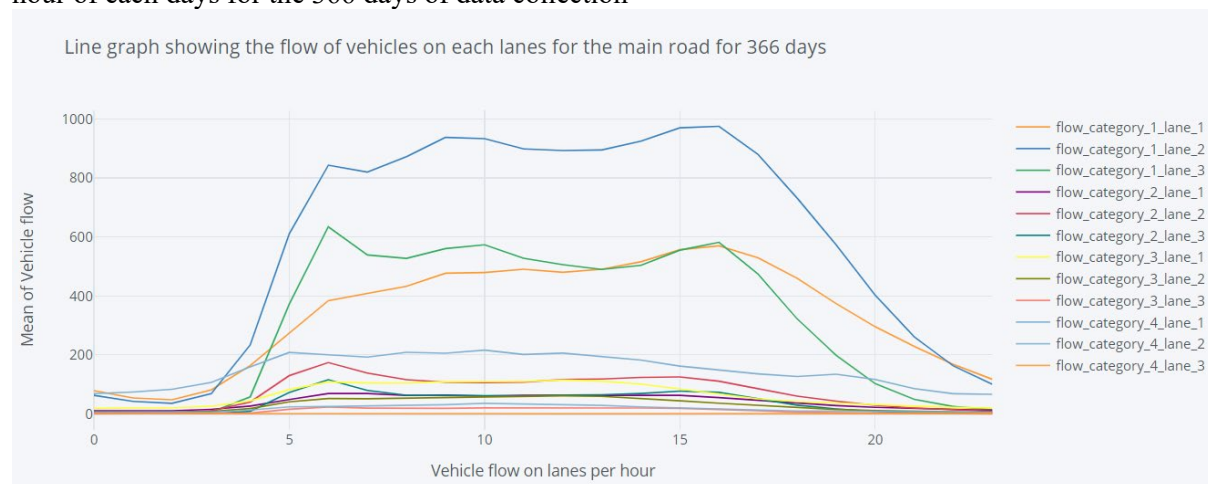

Figure 10: vehicles on each lanes of the main road for 366 days per hours of the day

Question 5: What are typical flows between 10pm and 5am on the main road and what is the relationship between total flows and HGV flows (or HGV percentage)?

Analysis: the graph presented in figure 11 shows that the typical flow between 10p.m and 5a.m on the main road records category 1 as the vehicle with the highest flow. However, lane 2 of the main road used by vehicles in category 1 tends to have the highest flow by 5a.m. The graph also informs that vehicles in category 4 on lane 1 tends to be busy at 5a.m.
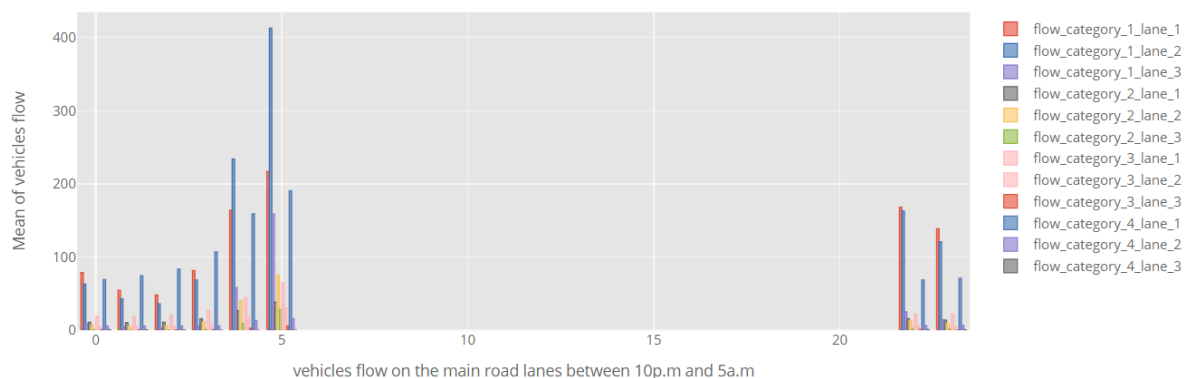

Figure 11: vehicles on each lanes for the main road by 10p.m to 5a.m for 366 days

The relationship between the total flow and HGV flows is calculated using Pearson correlation matrix. Figure 12 displays a heat map of the correlation matrix. Note that the rows and columns are sorted in the order suggested by the hierarchical clustering. The heatmap plot in this figure allows you to discover various subsets of the variables that seem to be highly correlated within the subset.
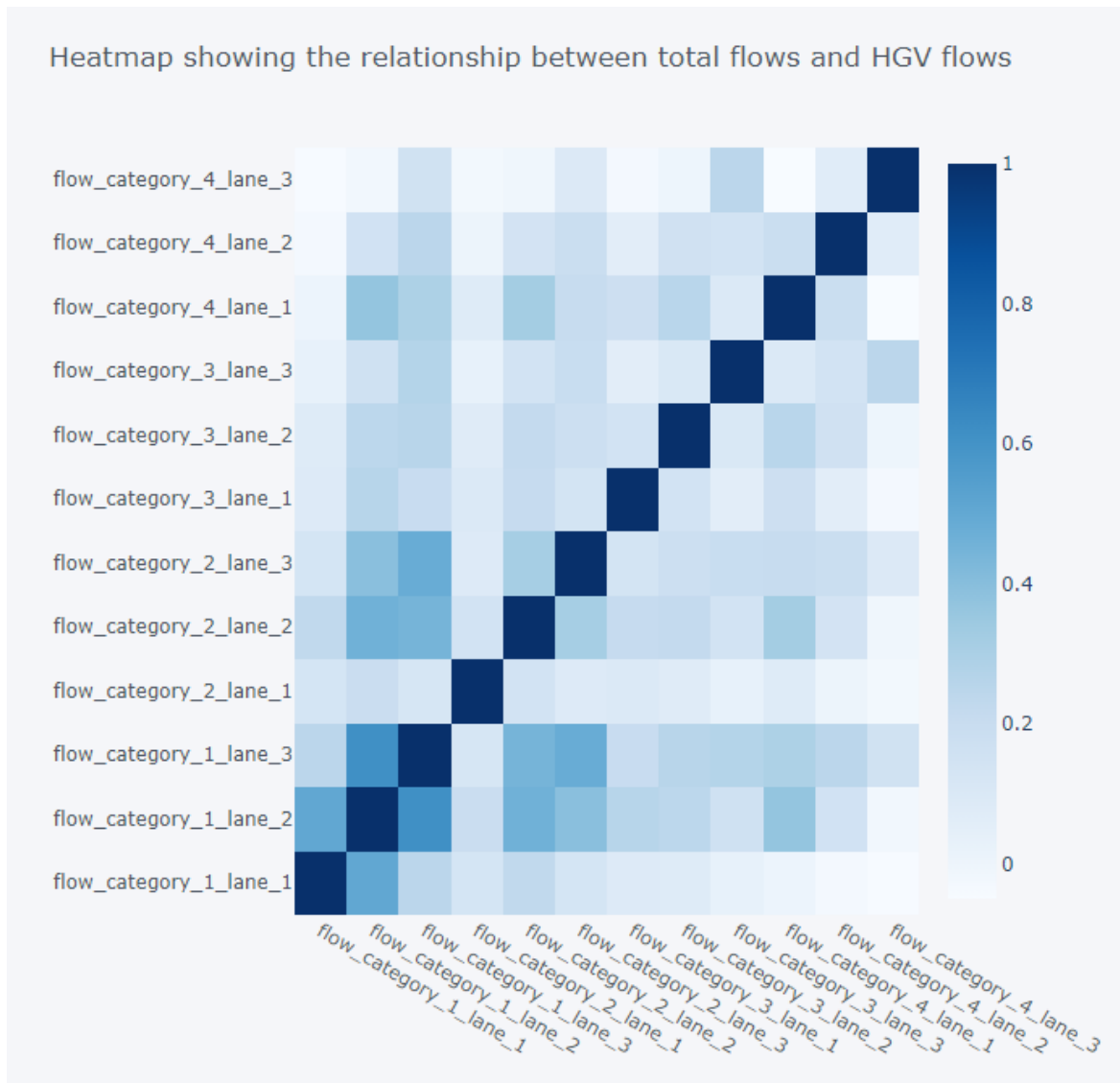
Figure 12: Heatmap showing the relationship between the total flows and HGV flows