TfL Bicycle Usage Project

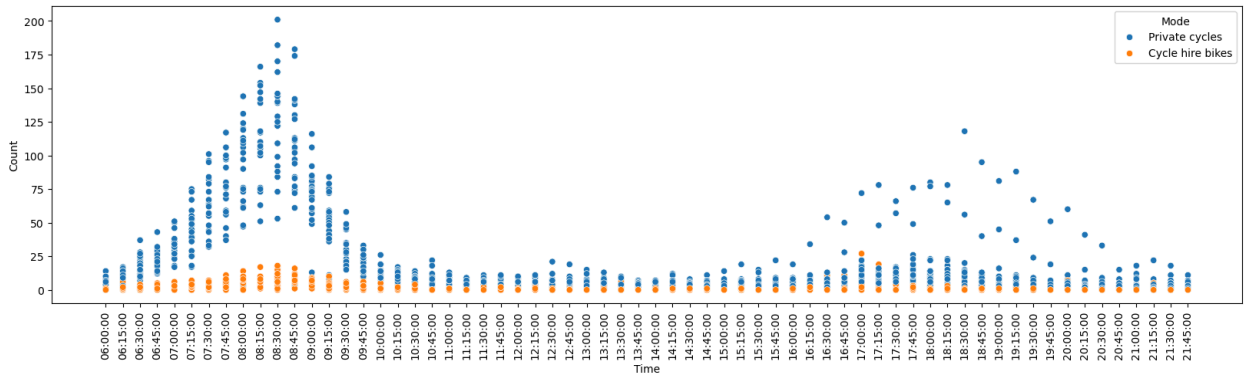
The data (obtained originally from the TfL website) consists of various files from a large number of monitoring sites across central London.

The aim of this project is to focus on the data from three different monitoring sites (ML0004, ML0021, ML0029). The goal is to research the volume of cycle traffic passing each point, with a view to introducing a charging regime for busy periods.

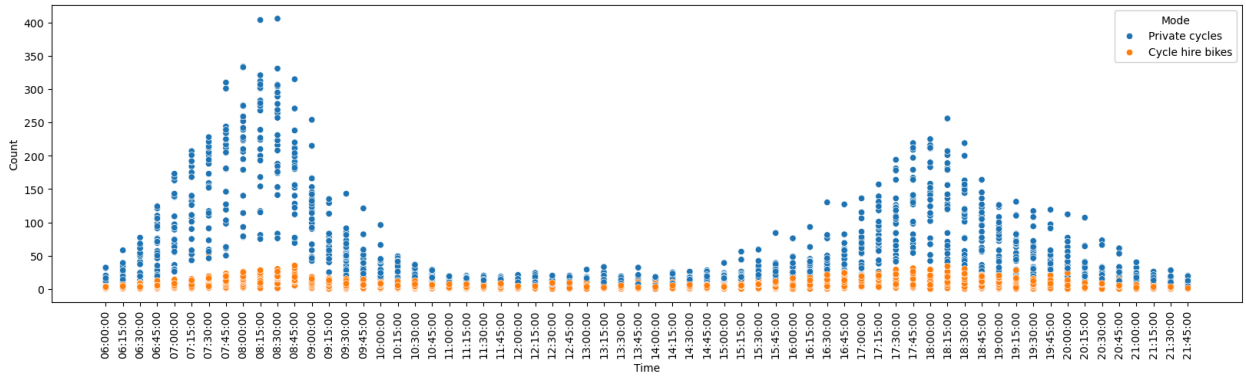
- The first task will be to aggregate the data into a single file in a suitable format, eliminating irrelevant data.
- The second task will generate various scatter plots to illustrate different aspects of the data, first over the whole time period, and second by considering only the time of day.
- The third task will consider how to aggregate and average the data in an appropriate manner and display a corresponding plot.
- The final task will apply some basic machine learning tools to identify possible charging periods.

I have shown what steps correlate with what graph alongside the user input to show how the graph has been developed.

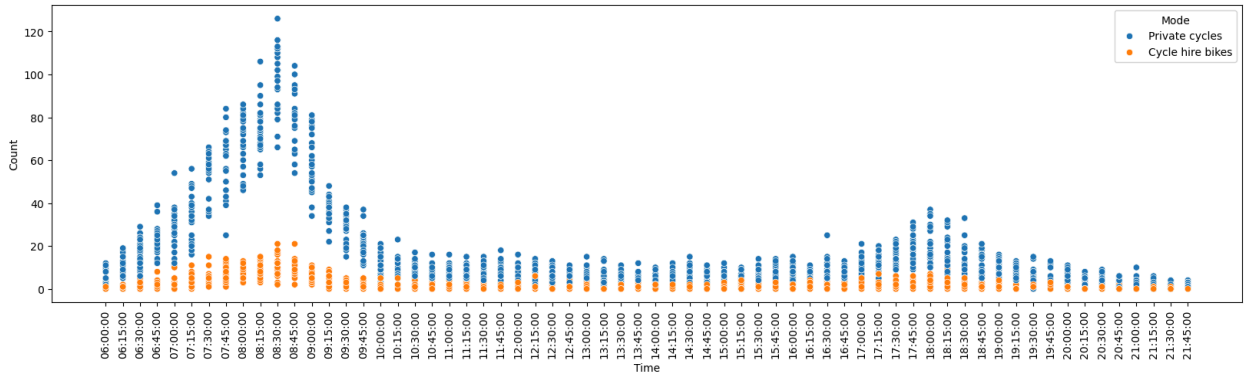
(Step 2) ML0004 -> Northbound -> N (All cycle) -> T (Time only) -> Model



ML0029 (eastbound):

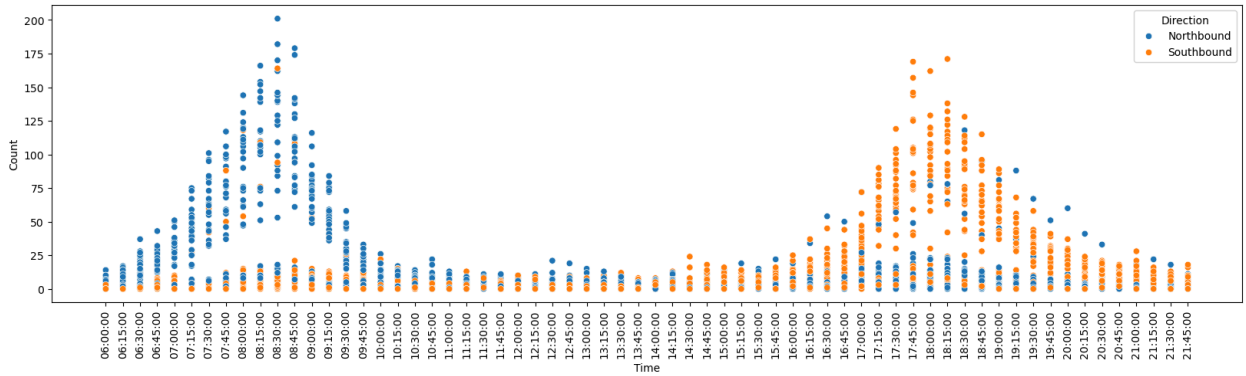


ML0021(northbound):

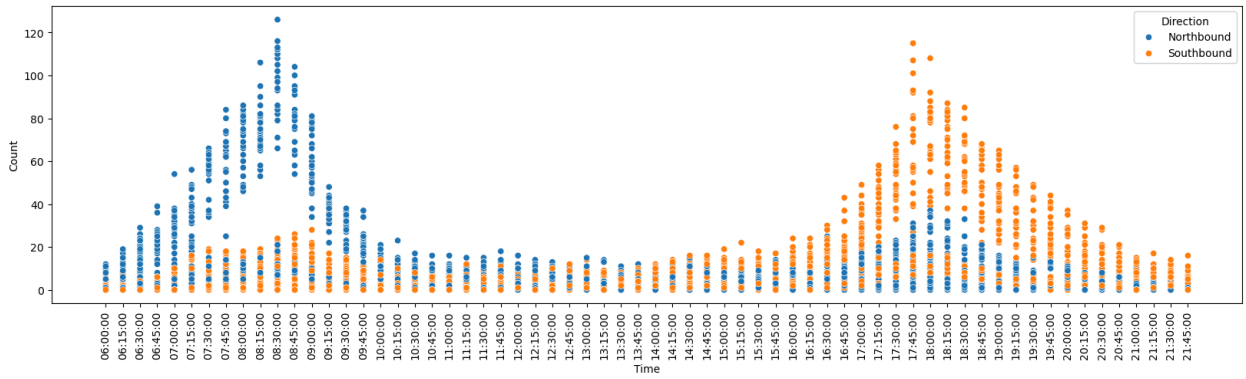


The graph shows the difference between “private cycles” and “cycle hire bikes”alongside the peak travel periods. The peak travel periods seem to be occurring around 7:00-9:00 AM while some are happening occasionally around 16:30-19:00 PM, which could be the influence through commuting to work/school from home and vice versa. Also the results are similar for the other two stations, as the other two stations have clear peak travel periods occuring at the same time.

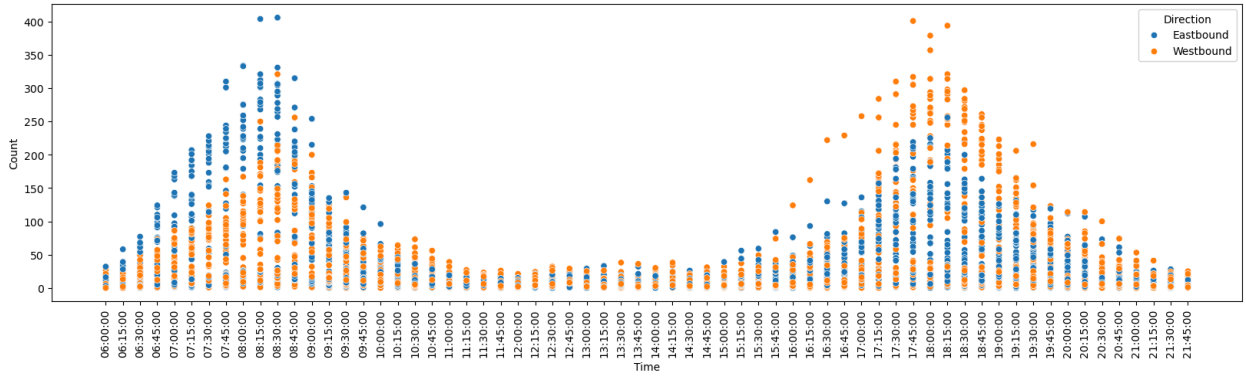
(Step 2) ML0004 -> Any (direction) -> N (All cycle) -> T (Time only) -> Direction



ML0021:

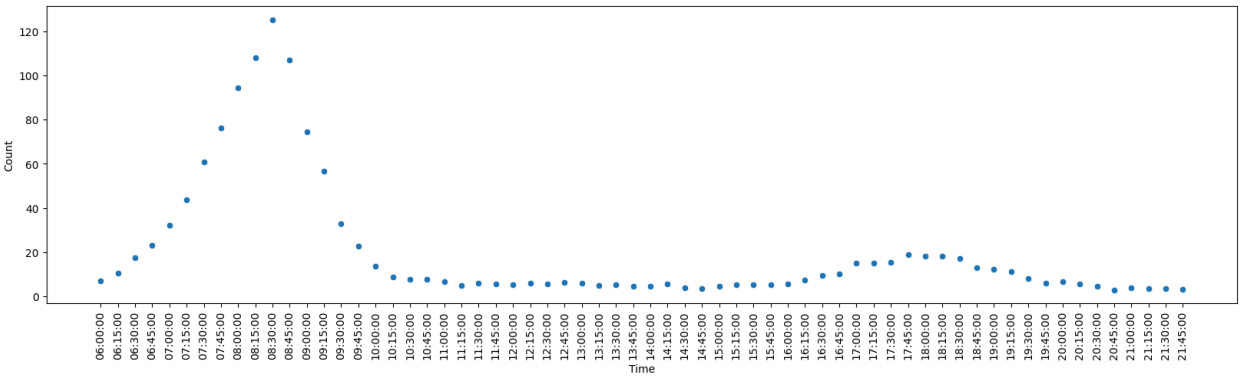


ML0029:

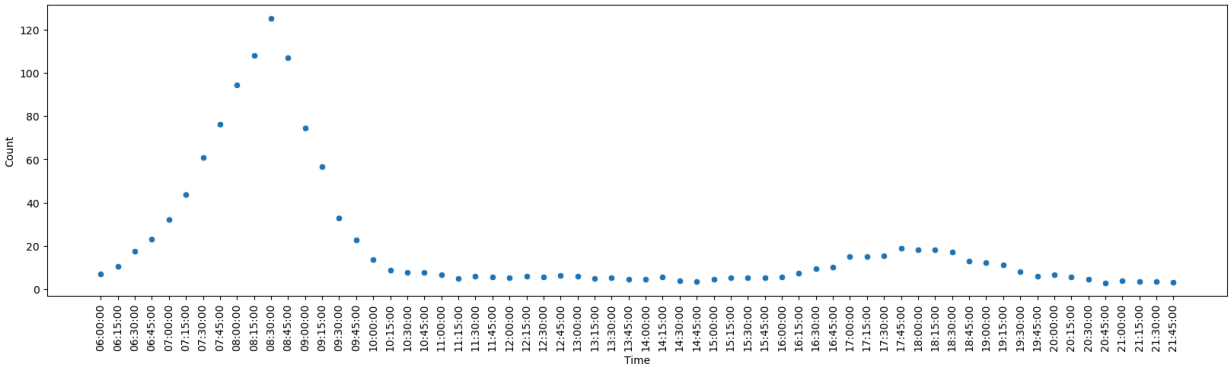


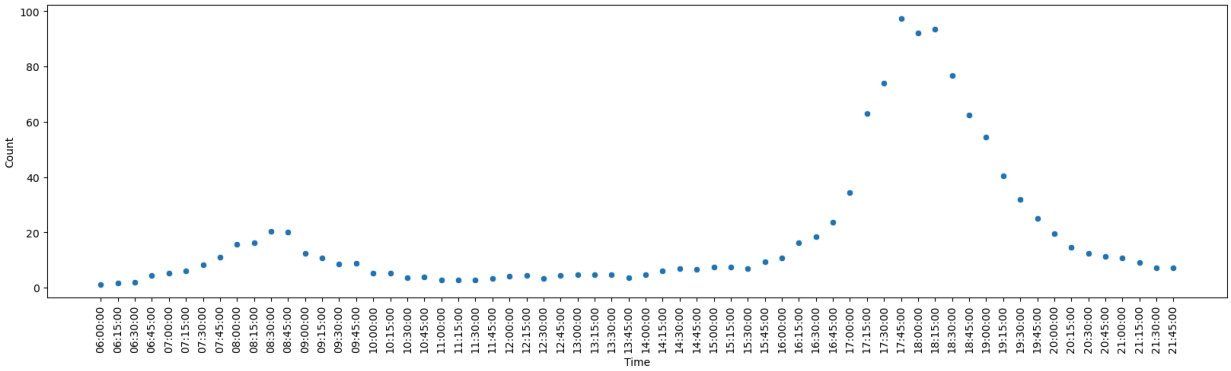
The graph suggests that the peak travel periods at the primary station in each direction appear to be higher in one direction. Due to the graph showing the difference between the Northbound and Southbound cycle. For example, in the morning it shows more count of cyclists travelling Northbound, whereas when it becomes the evening it indicates more cyclists travelling Southbound. The reasoning could be because it's more likely for the cyclists going from home to workplace and vice versa which results in one direction on North and South.

(Step 3) ML0004 -> Northbound

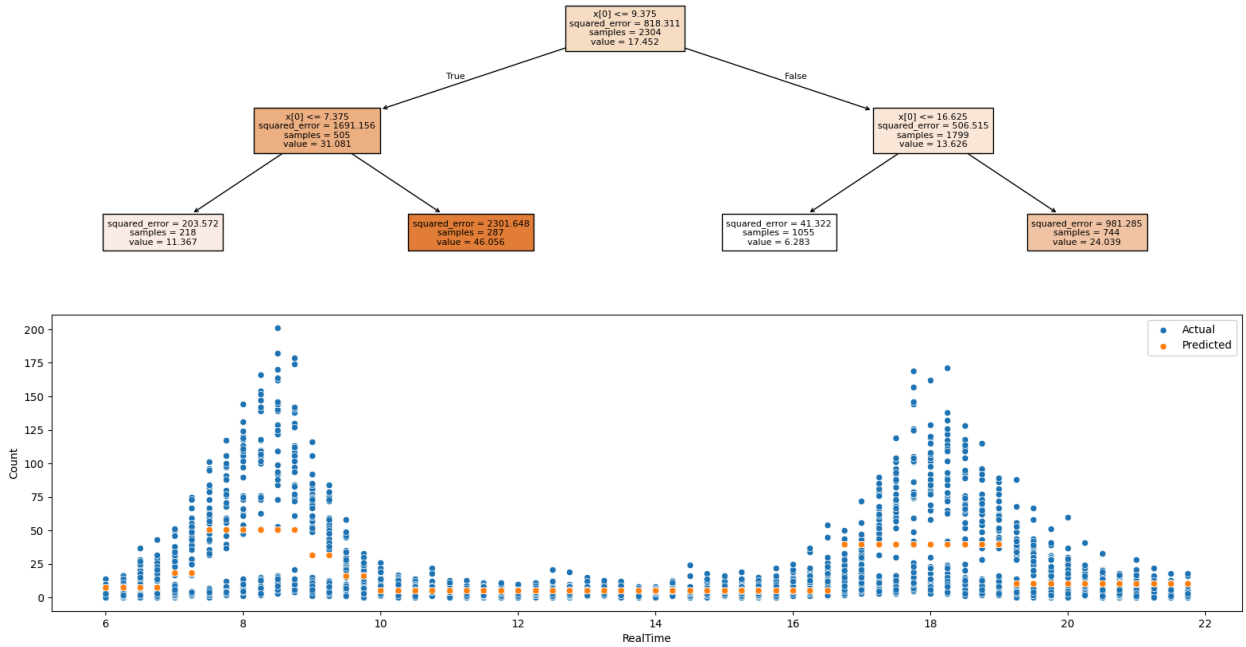


This graph helps to show the peak travel periods from the previous (3.) graph by highlighting the peak periods.
Below are two images. The top image shows the Northbound graph is simplified, reinforcing the findings that there are more cyclists in the morning (above). Alongside, the Southbound of station ML0004 (below) shows cyclists coming back from work (opposite of North). There are similar results with the other two stations as well.





ML0004 (Preferred choice of max_depth = 3):



We have chosen 3 as our value of max_depth. As we have tested with any max_depth higher than 3 may contain too specific or detailed patterns which may be irrelevant or unnecessary data to capture, like max_depth=4 will make the model too complex and there is a risk of overfitting. Also, max_depth lower than 3 may be too simplistic and not record the peak periods making it less accurate. Whereas max_depth = 3 is the perfect balance between accuracy and simplicity.

The two peak periods to be on the given plot from the tree data are roughly around:
7:30-9:30 AM and 16:00-18:00 PM and after (simply areas in which travelling to work or home is more common).

The method produced is a sensible peak period for TfL to use across London as the pattern remained consistent. However, it can be adjusted to fit more with context such as holidays and events, allowing the graph to be refined to external variables. Whilst also considering the part of London as different locations will play a huge factor towards the peak periods