KEY FINDINGS

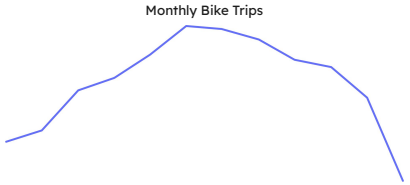
GENERAL TRENDS AND INFORMATION THAT COULD
HELP WITH KEY DECISIONS OR FUTURE PROJECTS

SCHEDULE

Bike demand peaks at 17:00 but also spikes at 08:00 suggesting that people commuting to and from work are a significant audience for the product.

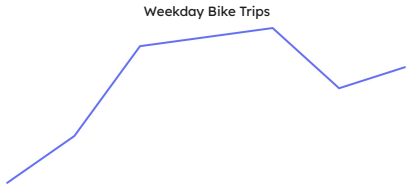
SEASONALITY

The summer increases demand for bikes significantly with June and July being the most popular months and December and January being the least popular.



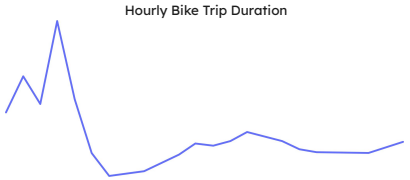
ROUTINE

Wednesday, Thursday and Friday seem to be the most popular days of the week. Interestingly Monday seems to be the least popular.



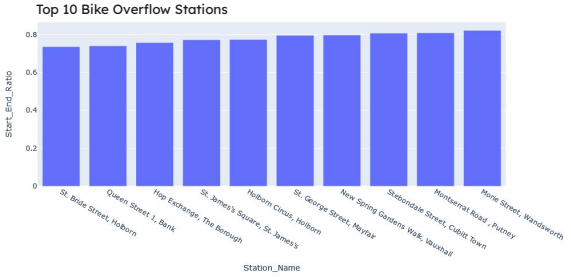
DURATION

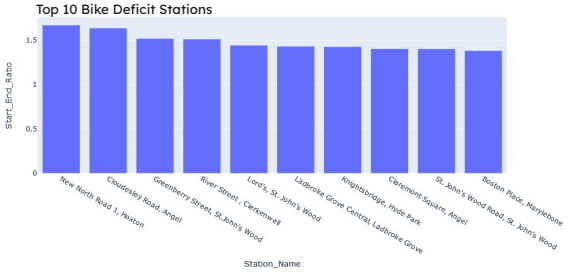
The longer trips are often taken before 5AM, and the shortest trips taken during similar hours to the peak trip volume, this could be due to the decreased proportion of trips taken for pleasure.



RESOURCES

There are some stations that might require resource reallocation as such as New North Road 1, Hoxton, which gets around twice as many bikes starting there as it does leaving.
On the other hand New Spring Gardens Walk, Vauxhaul seems to experience an overflow of bikes as for every 8 bikes that leave 10 arrive.





USE CASES

POTENTIAL PROJECTS THAT COULD HELP THE EXPANSION
INTO RENTAL-CYCLES BASED ON PROVIDED DATA

DEMAND FORECASTING

This data could be used to forecast the demand at specific stations over the next week or month. It would aid with resource allocation as well as help plan for events.

This project could be accomplished using a time series model and an aggregation of the provided data by days or even hours depending on required accuracy. Then the day of the week, hour of the day, station name and even more data can be calculated to give historical data for the model to use.

STATION PLACEMENT

The start and end stations of the trips could be used to find common trips and routes that many cyclists use. This information would be very useful when opening new stations as it would help identify key areas that are often cycled through but lack station resources.

However, this sort of project would greatly benefit from additional data such as user data and actual route data, rather than just start and end station locations.

MAINTENANCE DETECTION

The data could also be used to identify potentially faulty bikes that might need repairing. This would save a lot of resources as checking every bike manually could be avoided entirely if the project is successful. Even if a fault reporting system is implemented into an application for the user, it is unlikely that all faults will be reported and many faulty bikes might get avoided entirely by the user.

This could be done by aggregating the data by bikes and getting historical data on them. A model can then be developed to look at the bikes history and detect anomalies that might suggest a bike is not working correctly.

I've selected this project as the prototype as I believe it is very good for demonstrating the utility of the provided data and could be very beneficial to the company.

PROTOTYPE

SHOWCASE OF THE PROTOTYPE PROJECT
WITH EXPLANATION AND EVALUATION

DATA

The data was processed to aggregate by bikes. This gives us historical data [1] for each individual bike over its lifetime.

- Unusual Trips were trips that lasted under 3 minutes, this was chosen as it would be an extremely low duration for a trip and could indicate something went wrong with the bike.
- Recent Average Trip Duration Change shows the change in average trip duration when going from looking at all trips to looking at only trips made after 2023-01-01. A significant decrease could indicate a recent fault with the bike.

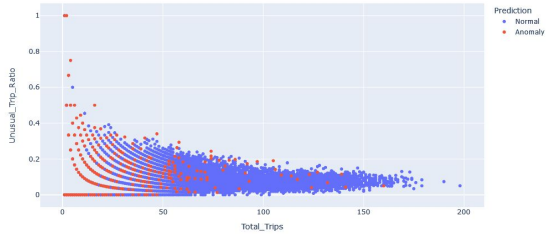
MODEL

As we do not have access to any fault history on the bikes, it is difficult to make a model that can predict exactly this. However, because we have data that could be indicative of a faulty bike, we can use an unsupervised model for anomaly detection of the data. This identifies the bikes that have unusual usage data and most likely suggests something wrong with the bike.

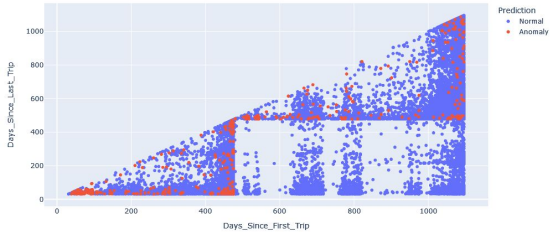
EVALUATION

The model works fairly well, and identifies a small volume of bikes that seem to be anomalous based on the averages [2] in comparison to the normal data [3]. The unusual trip graph [4] also suggests the model is working as bikes with low total trips and high unusual trip ratios are labeled as anomalous at a higher rate. Furthermore, the trip dates graph [5] suggests the model is identifying faulty bikes as at some point in 2022 many bikes were replaced with new ones and those that weren't replaced (likely operational bikes) were not labelled as anomalous.

Unusual Trip Graph [4]



Trip Dates [5]



OUTPUTS

Data Features [1]

- Total Trips
- Days Since Last Trip
- Days Since First Trip
- Unusual Trips
- Unusual Trip Ratio
- Average Daily Trips
- Average Monthly Trips
- Average Trip Duration
- Recent Average Trip Duration Change

Anomalous Data Averages [2]

Total_Trips	30.365517
Days_Since_First_Trip	571.882759
Days_Since_Last_Trip	336.963218
Unusual_Trips	3.878161
Unusual_Trip_Ratio	0.155761
Average_Daily_Trips	1.218846
Average_Monthly_Trips	4.734026
Average_Trip_Duration	165.190039
Recent_Average_Trip_Duration_Change	-10.237164

Normal Data Averages [3]

Total_Trips	56.738820
Days_Since_First_Trip	774.674433
Days_Since_Last_Trip	301.265262
Unusual_Trips	4.860402
Unusual_Trip_Ratio	0.090372
Average_Daily_Trips	1.109344
Average_Monthly_Trips	3.896840
Average_Trip_Duration	22.134146
Recent_Average_Trip_Duration_Change	3.391612

CONCLUSIONS

CONCLUSION OF WORK DONE AND RECOMMENDATIONS
FOR SUCCESSFUL EXPANSION

FINDINGS

The historical data describes the way users interact with rental bikes and what work might be required in terms or resource management.

- A significant audience takes short journeys to commute to and from work. This could be used to promote your bike rentals to this sort of audience and potentially develop app features that this audience would benefit from especially.
- The time and seasonality that people take trips could be used to manage resources and bike inventory and could also be used to promote the service during these times of high demand.
- The Station data shown that resource reallocation would be required for certain stations especially to avoid issues of bike overflow of deficit.

WORK DONE

A simple model prototype was developed to show the utility of this data when looking at the maintenance of bikes. The model is not perfect, mainly because it is hard to verify if an anonymous bike actually has a fault or if there is some other reason for the unusual usage data.

However, it does show the proof of concept and paints a clear picture of the of the data that would be very beneficial to a project like this once the expansion has begun.

RECOMMENDATION ONE

The two recommendations I would give to make the expansion into this new industry successful would be to gather data on bike trips similar to the provided data as this would be extremely helpful in developing projects to streamline many aspects of the company. Furthermore, gather as much additional data as possible, such as:

- Data of faulty bikes to aid with a supervised fault detection model.
- GPS data if possible to identify common routes taken by users.
- User data, this would be very beneficial to many aspects of the company, but especially so for planning new stations.

RECOMMENDATION TWO

I also recommend focusing on resource allocation. Having an efficient and reliable system to move bikes between stations as well as store bikes in inventory at times of low demand is extremely important for several reasons; bikes being left out or not being used for a long time is bad for the longevity of the bike and most importantly, not meeting demand means users can't access bikes when they need them, therefore the company loses out on potential income.

The data clearly shows that having a deficit at a popular station on a popular day during the summer could have a significant impact on profit and must be avoided.