Proposal to Increase Cycling Uptake in London

Prepared for Transport for London
By Team InLondon (Adeline, Rohit, Suki, Tamas)

Background

London Mayor's 2018 Transport Strategy's overarching objective of making London city more liveable by improving public transport systems, getting more people to walk and cycle and reducing dependency on cars. Specifically for cycling, the goal is to increase the number of trips by 400% by 2041. Currently, cycling is an under-utilized mode of transport, accounting for less than 4% of total trips per person per year, and only 0.8 million daily trips are made by cycling.

Research & Hypothesis

As London is a large city, our immediate priority was to identify which area of London presents the greatest opportunity. Outer London came out on top, ahead of Central and Inner London because it is home to 64% of London's population but represents only 4% of bicycle trips in 2021.

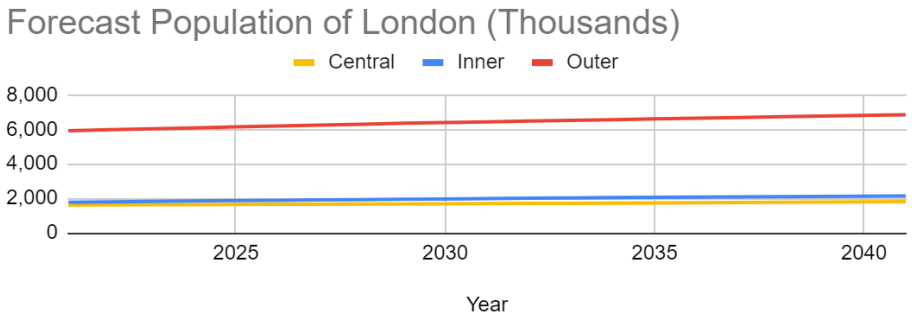


Fig 1: Forecast of London Population from 2020

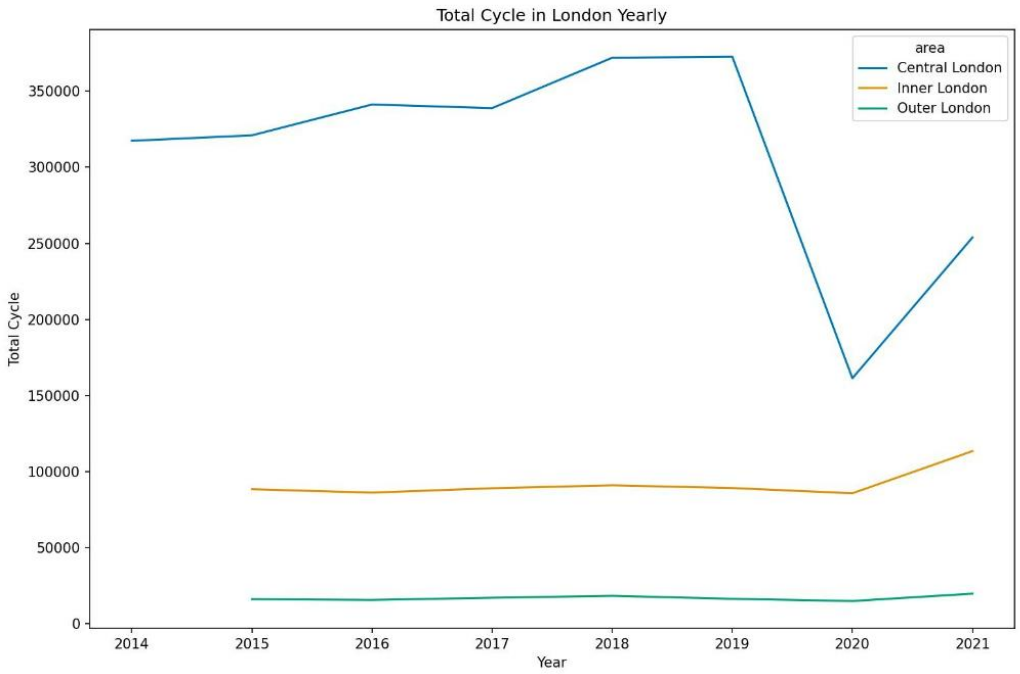


Fig 2: Total No. of Cycles in London per Annum

Next, we wanted to understand what are the key considerations and barriers towards cycling. Through secondary research, we identified 4 key factors that we believe we can influence to be more pro-cycling. The key factors are:

1. Bicycle theft - 75% of users who are open to cycling are concerned
2. Lack of parking - 65% of users who are open to cycling are concerned
3. Cost of bicycle ownership - 50% of users who are open to cycling are concerned
4. Access to car - People who do not own a car are 2x more likely to cycle than those who do

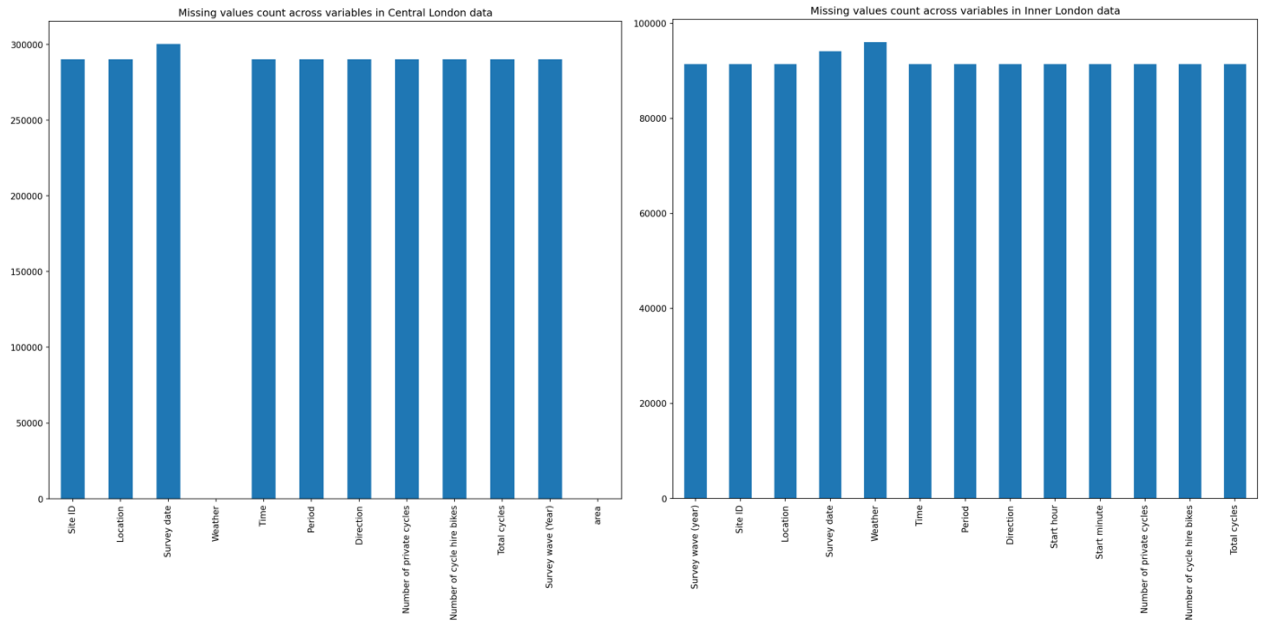
Recommendation 1: Increase Bike Parking Availability

Dataset Used

Our study on potential recommendations began with exploratory data analysis over the London parking infrastructure dataset and London bike traffic count datasets. Additionally, a dataset for the London parking infrastructure was extracted using the 'CycleInfraLnd' package from GitHub using the 'devtools' package.

Data Cleaning Approach

From the figures, it can be seen that the three datasets contain a lot of NaN values. The Central & Inner London NaN values were dropped as it was found that the entire observation in rows was NaN. Outer London NaN values were filled with the most common date, and weather pattern, as the percentage of missing values was negligible when compared to the total observation.



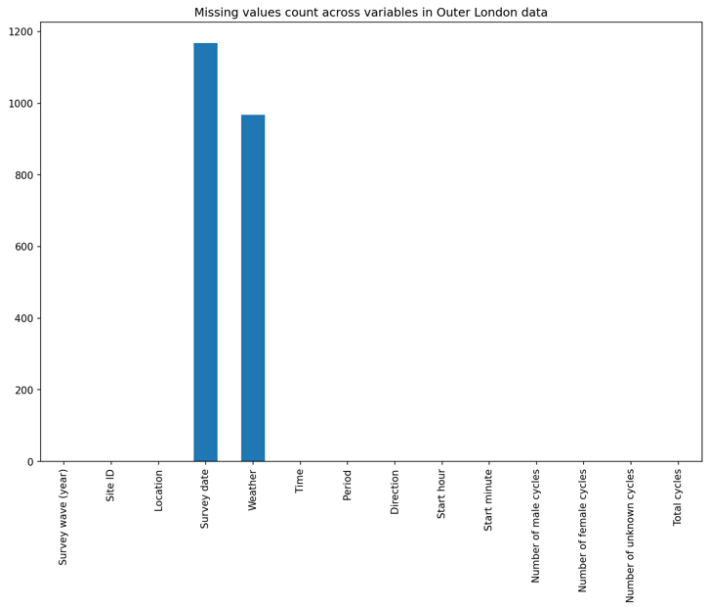


Fig 3 to 5: Count of Missing Data in the London Bike Traffic Datasets

The column ‘Survey date’ was converted to datetime format and was then split into three columns for day, month & year. The Weather column was categorized into Wet, Dry, Windy, and Other and stored in the column ‘Category.’ Initially, all the values in the 'Category' column are set to 'Other.' Then, using the 'loc' function, the 'Weather' column is checked for specific weather conditions using string matching. If the condition is satisfied, the value in the 'Category' column is updated accordingly.

An additional column, ‘Area,’ was inserted for each data frame, and then the three data frames were concatenated. This concatenated data frame was then merged with the Biking Sites Dataset. The 'Borough' column from the 'Biking Site' dataset is merged to the 'merged_cycle_df' dataset based on a common column 'UnqID' (unique identifier), which exists in both datasets. This is achieved using the merge() function, with the 'left' join option specified. The resulting dataset will have all the columns from 'merged_cycle_df' and the 'Borough' column from 'Biking Site.'

Visualization & Insight

Visualized by a bar chart, it came to our notice that Outer London has 40% less parking than Inner London despite its population. With parking availability being a key concerning factor on cycling uptake, parking coverage expansion in Outer London is crucial in driving the growth.

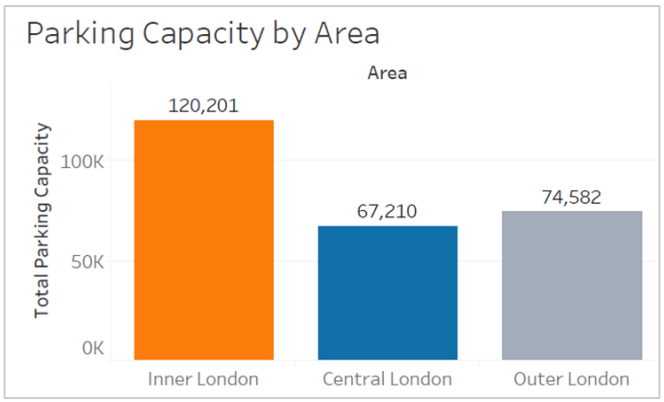


Fig 6: Bike Parking Capacity by London Areas

A heatmap was also created on Python to reflect the cycling directions by hour to study the London bike traffic flow along the day. The heatmap enabled scaled comparison of the bike trip volume, where the highest volume directions and timeslots were displayed in red to visualize their intensity. We would

recommend prioritizing the parking expansion in the Northern and Southern parts of Outer London, as it was validated visually that the mainstream of London bike trips were north and south-bound respectively during peak hours in the morning and evening time.

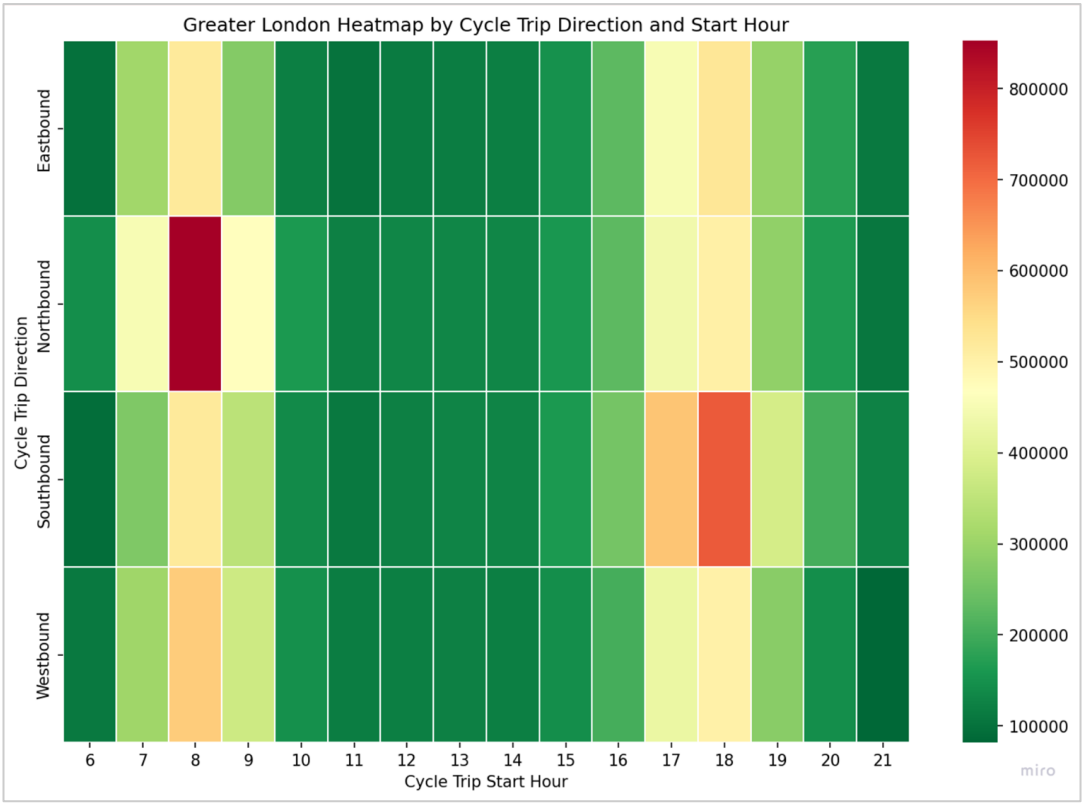


Fig 7: London Cycle Trip Directions by Hour of Day

We also looked into the types of parking to add. The visualization was created in Tableau by using the relationship between the 'Boroughs' of the 'merged_cycle_df' and the parking infrastructure dataset. From the parking infrastructure capacity bar charts, we learnt that secure and locker parking were seriously lacking in Outer London when compared against other parts of London. Since cycle theft was the primary friction for cycling uptake, these infrastructure should be considered as priority in the expansion plan.

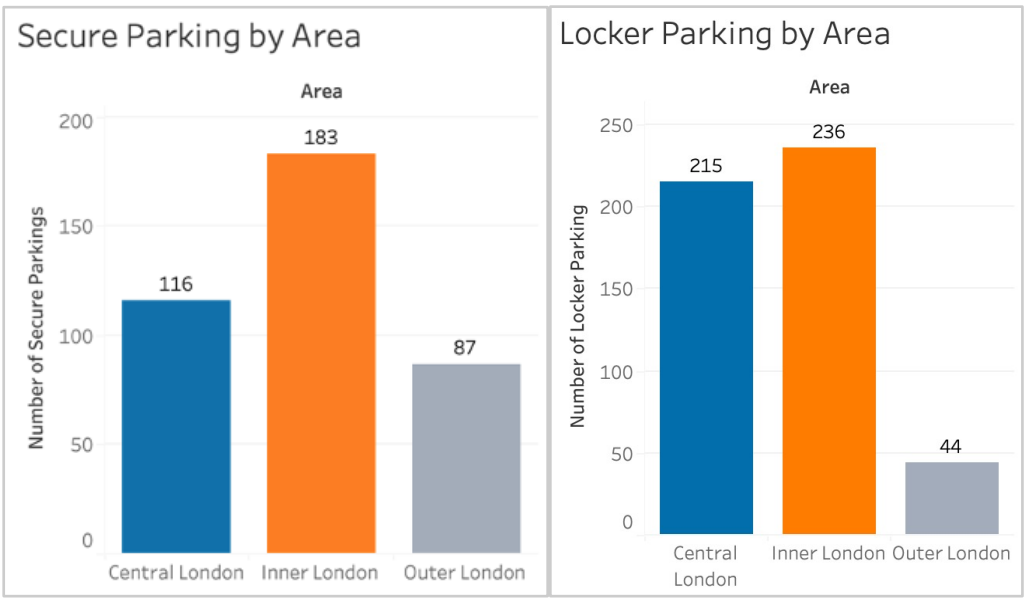


Fig 8 & 9: No. of Secure & Locker Parking by London Area

Recommendation 2: Increase Hire Bike Coverage and Expand Assortment of Bicycle Types

Dataset Used

We started our analysis from the bike traffic dataset provided by Thoughtworks. Due to limitations of the dataset we had explored additional data, including the Santander Bike traffic data from TFL, and the New York Citibike cycle trip data from the New York City open data.

We also conducted a sentiment analysis on dockless bike related Tweets. The search query for Twitter used the 'snscrape' library to retrieve tweets in the English language that were posted between January 1st, 2018, and February 1st, 2023.

Data Cleaning Approach

Each tweet text in the dockless bike sentiment analysis was cleaned by removing URLs, mentions, hashtags, and special characters using regular expressions. The cleaned text was then appended to the 'cleaned_tweets' list. The 'snscrape' library was used instead of Twitter API to get more tweets.

Visualization & Insight

We found similar daily, monthly and annual trends as all cycling trips.We can conclude that there are several limits of usage due to the capacity of the hire bikes especially electric bikes and the area covered by docked Santander bikes, therefore we would recommend expanding Santander Bike coverage or introducing more hire bike players to the city. To prevent market fragmentation brought by the presence of various players, TFL might consider developing a unified app to consolidate all hike bike services on one platform to increase hire bike accessibility.

Meanwhile, a modest growth was observed on e-bike usage since its launch in 2022. Since e-bikes were dockless bikes that enabled more convenient parking, we also suggested increase the proportion of e-bikes among hire bikes supply in London.

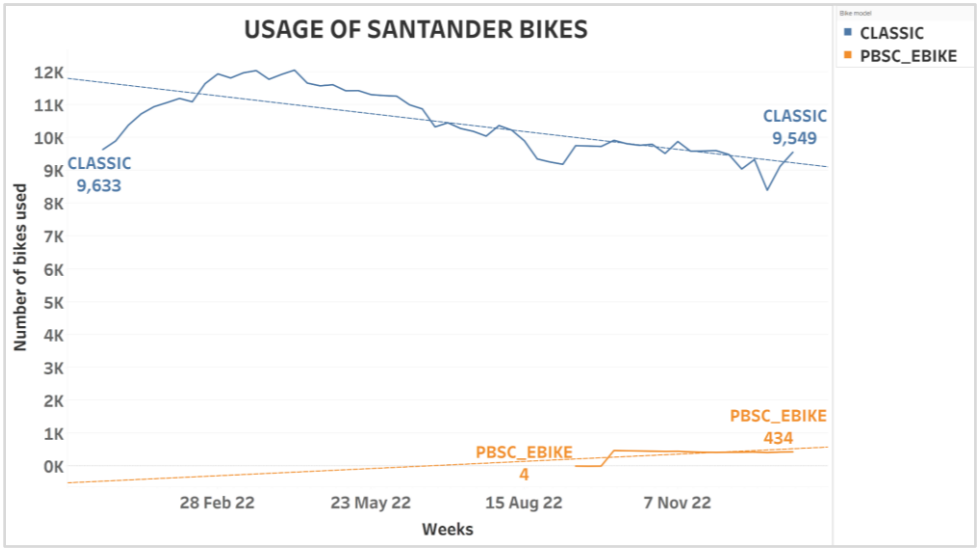


Fig 10: Locker Parking Count by Area

Given the hire bike data in London was relatively limited, we also delved into the Citibike share bike data in New York City for benchmarking. Looking at the member/ casual rider trip count split per each bike type, it was noticeable that casual riders had higher likelihood to use e-bikes than other bike types. It was an indication that e-bikes could be a preferred option for new or potential cyclists.

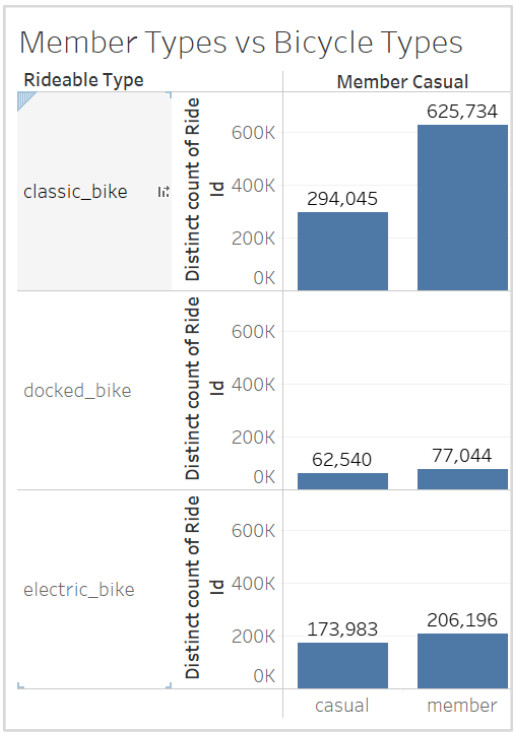


Fig 11: Member & Casual Rider Trip Count by Bike Type, New York Citibike

Additionally, we conducted a sentiment analysis on dockless bikes to understand how people felt about them. It appeared the sentiment towards dockless bikes was generally positive, with 44.71% of tweets expressing positive sentiment, 25.18% expressing negative sentiment, and 30.12% expressing neutral sentiment. This indicated that there was a mixed sentiment towards dockless bikes, with slightly more positive than negative sentiment.

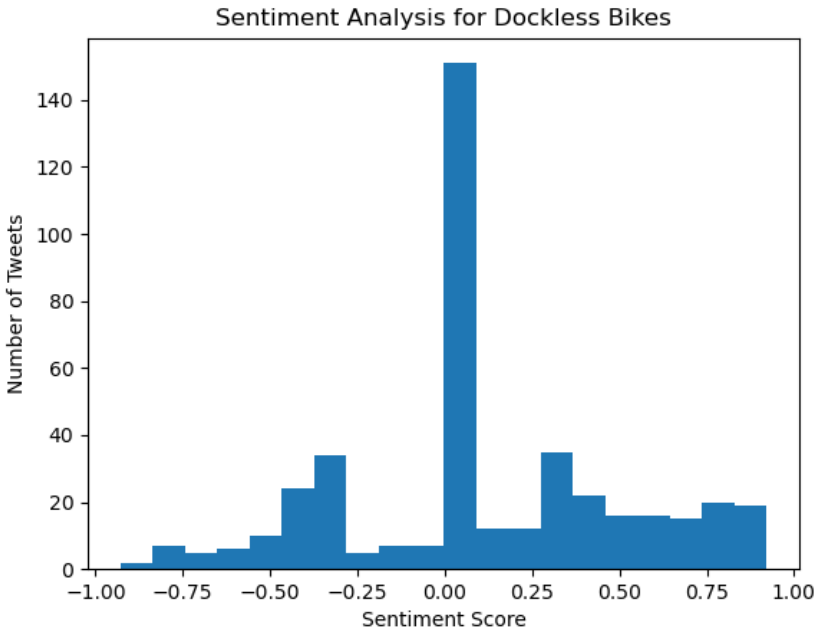
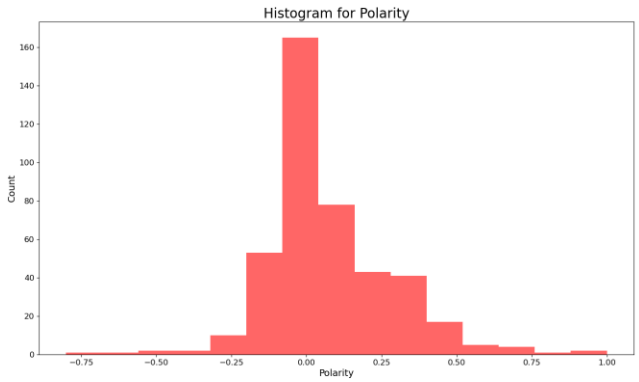


Fig 12: Sentiment Analysis for London Dockless Bikes



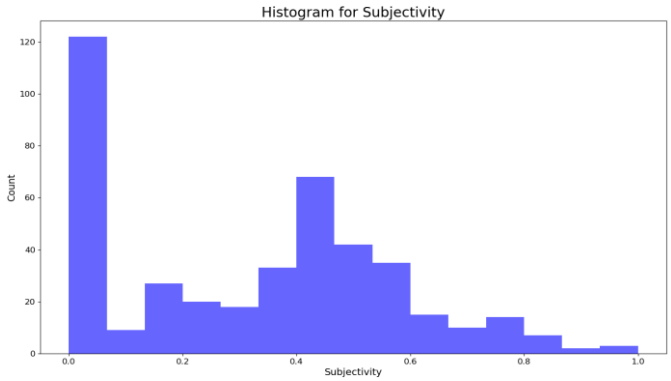


Fig 13: London Dockless Bikes Polarity Histogram Fig 14: London Dockless Bikes Subjectivity Histogram

The histogram of polarity scores suggested that most of the tweets were somewhat positive, with polarity scores clustered around 0.25-0.5. The histogram of subjectivity scores suggested that most tweets were somewhat subjective, with subjectivity scores clustered around 0.0-0.6, implying that many people had expressed their personal experiences and opinions.

It was worth noting that there were also some tweets with very negative polarity scores, which were often related to issues with careless parking causing trouble for pedestrians and those using wheelchairs. Hence we suggested designated parking space for dockless bikes could be considered to minimize pavement obstructions in the future.

Recommendation 3: Discourage Ownership and Use of Private Cars

Dataset Used

We relied on two external datasets - vehicle registrations from the Department of Transport and households tenure in London from the Office for National Statistics - to make comparisons between the average number of vehicles per household in Inner versus Outer London. A time period of 2010 to 2020 was used as that was the common time period between the two reports. Central London was not analyzed as there was no household data available on Central London.

Visualization & Insight

We used Python to multiple Excel sheets into a single dataframe before exploring data and exported an excel sheet of summarised data, before using Excel or Google Sheets to create visualisations.

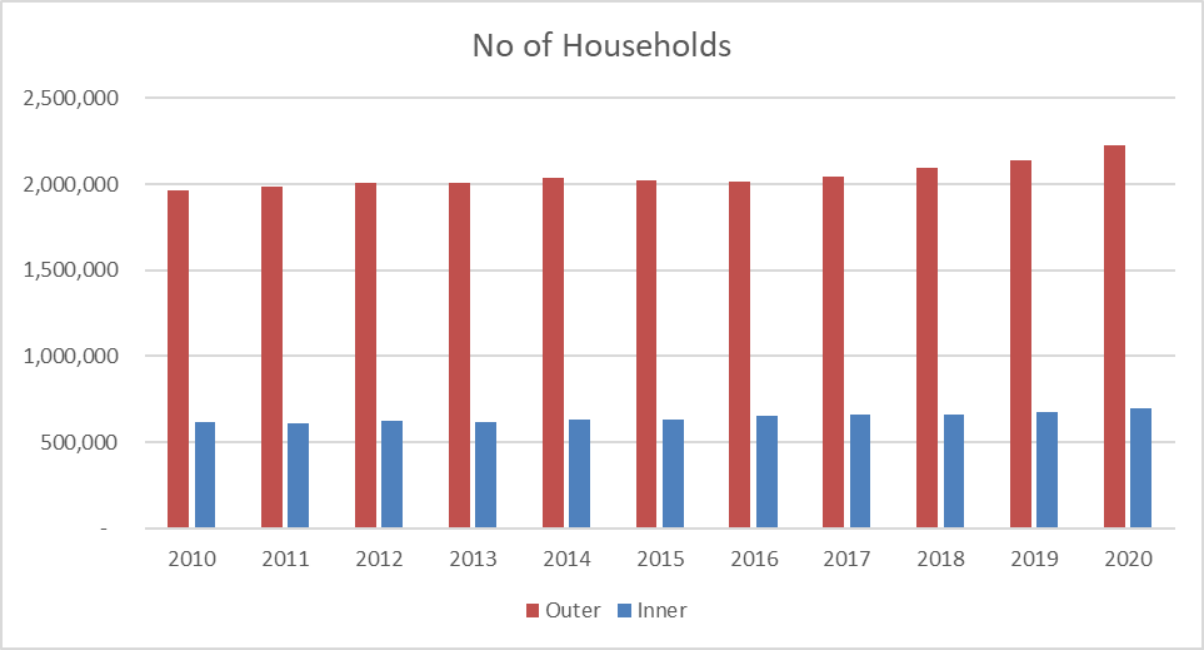


Fig 15: No. of Households in Inner and Outer London from 2010 to 2020

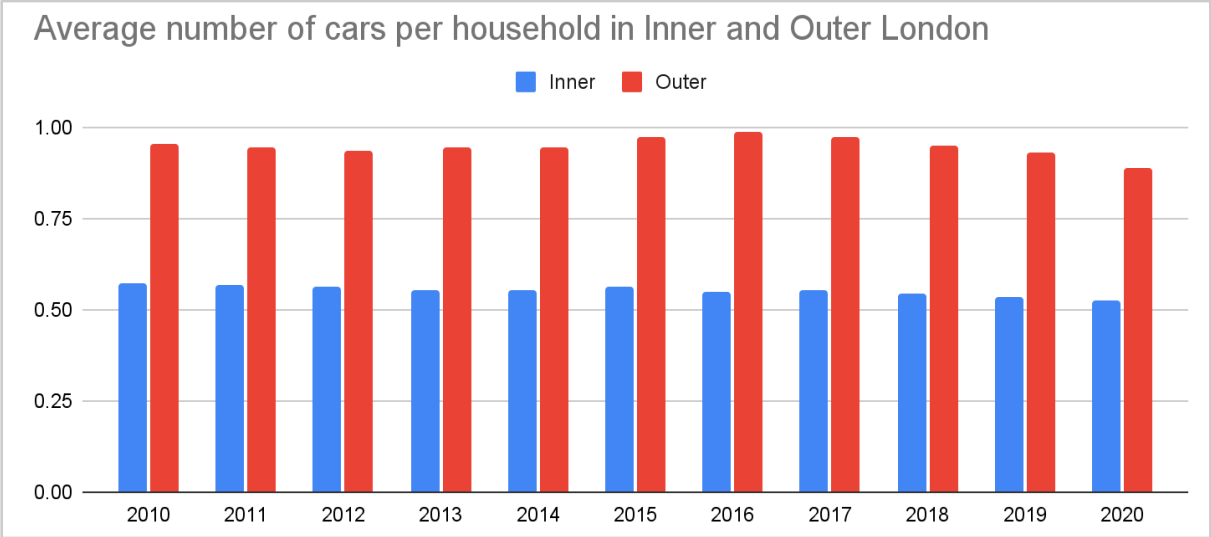


Fig 16: Average Number of Cars per Household in Inner vs Outer London from 2010 to 2020

There are 3x the number of households in Outer London vs Inner London and the average number of cars per household is 2x in Outer London compared to Inner London.

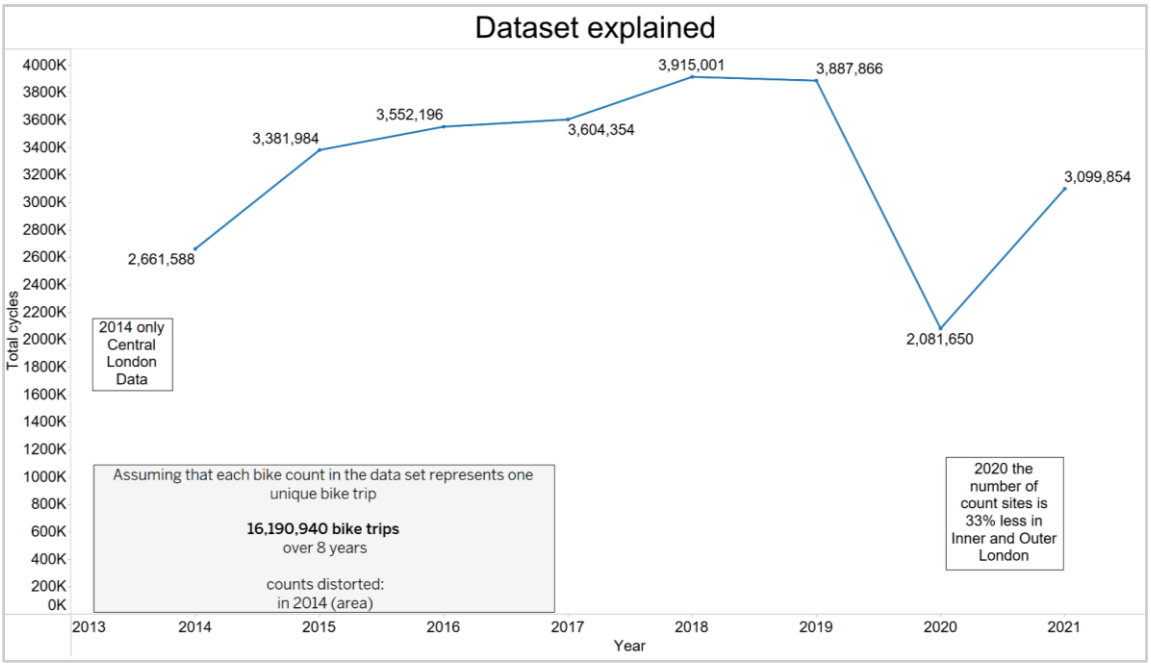
Almost every household in Outer London has a car and if car ownership was reduced, cycling uptake in Outer London would increase. Therefore, we suggest making car ownership and usage less desirable by increasing taxes and fees related to car ownership, increasing parking rates and widening the congestion zone. However, because the marginal cost of changing policies is rather low, cars are used to commute over long distances and to be fair to the entire population, we suggest that these measures are carried out London-wide and not just within Outer London.

Conclusion

Our 3 recommendations addressed the various frictions that prevented people from adopting cycling such as theft, parking availability, affordability, ease of access and availability of alternatives. In adopting these measures, we hope to change the landscape of the transportation system in London by promoting home to hub cycling trips, activating longer bike journeys and improving the overall cycling experiences.

Assumptions and Limitations

The bike data collection methodology posed constraints to our analysis. For instance we were unable to recognize the identity of each bike, leading to the assumption that every bike traffic count equaled one unique bike ride. The volume of traffic data collected could be skewed by the bike counter locations, which in reality was at large unevenly distributed across London areas.



We were also aware that cycle lanes and safety were pivotal elements in the cycling uptake growth plan, yet they would not be our focus area this time in view of the stretched project timeline.

Opportunities

<u>Data Collection</u>

The no. of bike counters for bike traffic data collection in Central London is approximately twofold of those in Outer London. This is one of the most impactful limitations in our study, and to minimize data skewness we recommend installation of additional bike counters in Outer London, ideally proportionate to where the parking infrastructure expansion is planned.

Currently TFL only collects data on the quantity and location of parking infrastructure, yet the utilization of parking has not been scrutinised. As parking utilization is highly reflective of the cycling demand in an area, we recommend TFL to install counters at major, sizable parking spaces to monitor the utilization level, especially at the transportation hubs.

<u>Further Analysis</u>

TFL can explore ways to collect cyclist demographics, attitude and behavioural data, then leverage them to conduct clustering analysis for a rider segmentation exercise. Once the rider segments are identified, a more targeted approach can be taken to address the needs of specific segments, and enhance marketing efficiency through tailored communications.

To optimize ROI of the hire bike expansion investments, TFL can again use clustering techniques to conduct a hire bike station profiling exercise, grouping stations by their respective profiles built on bike usage data. A standardized expansion approach can be applied to stations with the same profile to reduce development complexity and leadtime.

Reference

1. Department for Transport, Cycling, Motorcycling, School travel, Concessionary Travel and Road Safety (2022)
   https://www.gov.uk/government/statistical-data-sets/nts06-age-gender-and-modal-breakdown#cycling-and-motorcycling
   https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1101178/nts9903.ods

2. Transport for London, Travel in London Report 15 (2022)
   https://board.tfl.gov.uk/documents/s19181/Travel%20in%20London%2015%20Overview.pdf
   https://tfl.gov.uk/cdn/static/cms/documents/travel-in-london-report-15-data.xlsx

3. Greater London Authority, Land Area and Population Density, Ward and Borough (2011)
   https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough

4. Cycling Potential in London's Diverse Communities, Cycling Factsheet London (2021)
   https://content.tfl.gov.uk/cycling-potential-in-londons-diverse-communities-2021.pdf

5. Transport for London, Cycle Parking Implementation Plan (2019)
   https://content.tfl.gov.uk/cycle-parking-implementation-plan.pdf