Assignment 3: Final report

Team: Cloud 9

Promoting Cycling in London: Data-informed Recommendations

Submission date: 27th November 2023

1. **Background/ Context of the Business**

Cycling is a cornerstone of sustainable urban transport and a solution to urban congestion. ThoughtWorks has entrusted Cloud 9 with a mission: to enhance cycling culture in London. This report outlines our completed and future steps to unravel the complexities of cycling behaviour in London and formulate data-informed strategies. Our research hinges upon two central priorities, how can cycling uptake be increased? And, what factors drive people to cycle?

Our exploration dives into elements such as access to cycling infrastructure, road safety, trip dynamics, weather, and demographics. Delving into historical impacts and Transport for London's (TfL) progress, our collaborative journey aims to propel London towards a more sustainable and cyclist-friendly future.

2. **Project Development Process**

Research Rationale

Our work drew from a comprehensive array of sources to inform our research hypotheses and recommendations to ThoughtWorks. TfL's 2016 "Attitudes towards cycling" report [1] provided insights into public perceptions, influencing strategic decisions. The Department of Transport's 2020 publication, "A Moment of Change," [2] offered a broader context on increasing cycling uptake. Heinen and Buehler's systemic review [3] deepened our understanding of bicycle parking behaviour, vital for enhancing cycling infrastructure. The Census 2021 data on sex distribution [4] aided demographic analysis and London TravelWatch's report on personal security [5] and 2CV's report for TfL on "Cycling potential in London's diverse communities" [6] enriched our safety considerations. Additionally, insights from The Guardian's article on repurposing abandoned tube tunnels [7] prompted unconventional but forward-thinking suggestions. The culmination of these diverse perspectives substantiates our list of recommendations, presented comprehensively in the report's appendix, offering ThoughtWorks a nuanced and data-driven approach to enhance London's cycling landscape.

Data Collection and Cleaning

For the analysis, we used datasets provided by the client and publicly accessible databases published by TfL and the Office for National Statistics (ONS). The client's datasets covered patterns of cycling by gender in Outer London, bicycle ownership in Inner and Central London and weather data about all three regions between the years of 2015 and 2021. And the chief source of data for researching cycling infrastructure was the Cycling Infrastructure Database (CID) published by TfL. Our analyses also made use of scientific studies and reports published by experts and authorities in the field of urban mobility and sustainability.

The team used Excel, Python and R to thoroughly clean the collected data and create a shared repository of CSV files. The cleaning process involved removing blank entries or those missing crucial information, removing duplicate entries as well as standardising information about categories such as weather and time periods.

Client and Audience Considerations

- Our visualisations were generated using Python, R, Tableau and Microsoft Excel, and aimed to embody our commitment to clarity and accessibility. We also extended similar considerations to our writing material that aims to be as clear and simple as possible.
- We chose styles such as line charts, correlation lines and bar graphs as these convey information in clear terms while being easy to interpret for non-business users as well. These styles are also well-suited to the team's pace, since they could

be generated quickly in response to new research demands while being easy to integrate into presentation material (eg.: for stakeholders after video calls).
- For the infrastructural analysis charts we chose bold, dark text and clear labels and lines along with a muted colour palette to enhance visual focus. And colour palettes such as blue-orange were chosen for charts in the demographics and weather sections to make sure our charts were accessible to colour-blind audiences.

Recommendations

From this research we have formulated a list of recommendations. For the full overview please see Appendices A and B.

Infrastructure
- To address the shortage of bike parking by collaborating with public transport authorities citywide to strategically improve bicycle parking facilities.
- Regularly assess the usage and satisfaction levels of newly enhanced bike parking spaces, with the goal of achieving higher daily utilisation and fostering positive feedback from users.

Cycling Safety
- Invest in cycling safety by continuing to develop safety features like turning exceptions and except cycle signposts.
- Track trends in total cycle counts and the number of 'cycle except' arrangements to gauge the effectiveness of these safety measures.
- Gather feedback from cyclists to ensure continual improvements in safety.

Demographics
- Begin to address extreme gender disparities in cycling by conducting further research into why women are more reluctant to cycle. Starting points include the personal security of cyclists.

Weather
- Promote weather-resilient cycling infrastructure and introduce incentive programs to encourage cycling in adverse conditions.
- Monitor the usage of covered bike parking spaces and track cyclist participation in incentive programs.
- Conduct research into repurposing disused Tube Stations as Cycle Routes

3. **Technical Overview of the Code**

Choice of Analysis Tools and Techniques

For this project, we chose to utilise a combination of Python, R, and Excel, each with its unique strengths. Python's extensive data manipulation and visualisation libraries, along with its statsmodels.api for regression analysis, proved ideal for handling large and complex datasets. The interactive Jupyter Notebook environment facilitated data exploration and visualisation. R's specialised data analysis packages were particularly useful for extracting data from the Cycling Infrastructure Database (CID). Excel's user-friendly interface and data management capabilities played a pivotal role in data cleaning, data wrangling, and initial data exploration.

Data Preprocessing

Prior to conducting our analysis, we undertook several data preprocessing steps to ensure the quality and integrity of our datasets. This included handling missing values, identifying and removing outliers, transforming data into appropriate formats and removing information not useful for analysis. For example, columns without useful information for our analysis, such as the weekday on which the cycle count was recorded, were removed to obtain smaller datasets for handier analysis. Information about signages needed for our analysis were also extracted from JSON files into appropriate data frames.

Data Analysis Process

Our data analysis process involved several key stages:

1. Data Loading and Manipulation: We loaded the data into both Python and R using Pandas and dplyr, respectively, and performed necessary manipulations, such as data type conversions and data structure adjustments.

2. Exploratory Data Analysis (EDA): We conducted EDA in Python, R, Tableau and Excel to gain a comprehensive understanding of the data distribution, identify patterns and trends, and detect any potential anomalies. This involved techniques like descriptive statistics, visualisations, and correlation and regression analyses. Specifically, datasets with common columns were joined to enable further EDA. For example, the Biking site dataset was joined with inner, outer and central London cycle count datasets to enable EDA in terms of total cycle count by borough.

3. Statistical Tests: To further explore the relationships between variables and identify potential predictors of cycling patterns, we employed correlation and regression analysis. We utilised p-values to assess the statistical significance of our findings and R-squared values to evaluate the strength of the relationships between variables. The predictors were chosen according to various aspects of the cycling infrastructure in our research, namely bike parking capacities, cycle

safety in terms of number of related road signposts and access to dedicated cycling lanes such as Cycle Superhighways and Quietways.

4. Visualisations: Throughout the analysis, we created various visualisations in Python, R, Tableau and Excel to effectively communicate our findings. These included line graphs, bar charts and scatter plots.

Challenges and Limitations

Data analysis challenges in a team setting necessitate clear communication and alignment regarding data formats and types to avoid further data wrangling and ensure seamless integration.

We were limited by the availability of gender information to the Outer London dataset, which restricted our ability to conduct more city-wide analyses.

Code Organisation and Modularity

We organised our code in a modular fashion, utilising functions and classes, and documented the code with comments to promote reusability and maintainability. This approach enhanced our codes' readability and understandability, making it easier for future modifications and collaborations.

Conclusion

The choice of appropriate analysis tools and techniques, careful data preprocessing, and structured code organisation played a critical role in ensuring the success of this data analysis project. These aspects contributed to the reliability of our findings and the overall efficiency of our work.

4. **Patterns, Trends, and Insights**

Infrastructure

Bicycle parking facilities

Our research revealed a strong and statistically significant correlation between the availability of bicycle parking facilities and an increased propensity to cycle. We strongly recommend that workplaces, education establishments, TfL, and local councils prioritise the provision of adequate bicycle parking infrastructure.

Road signage

There was an evident correlation between Cycle exception signs (i.e. ‘No left turn’ and *No entry* ‘Except cycle’ signs) and cycling uptake, which we believed would be due to cyclists feeling safer on the roads where these measures were in place, and therefore encouraged to cycle more in these areas. Specifically for ‘Banned Left Turn with Cycle Exception’ signage, there was an R^2 value of 48%.

Cycle Superhighways and Quietways

We investigated whether there is a positive correlation between cycling uptake and more cycle superhighways and quietways and found hardly any correlation with very low R^2 values. We believe this is due to the fact that superhighways and quietways are not present in every borough but in around half of London boroughs. However, we would recommend further assessment of the effectiveness of superhighways & quietways per location in helping increase cycling before further introducing them to other London boroughs.

Demographics

Gender disparity in cycling

Analysing the Outer London dataset revealed extreme gender disparities with women consistently forming the lower share of cyclists between 2015 and 2021. In 2021, female cyclists formed 16.7% of the recorded rides while male cyclists formed 82.9%. However, in the same year, London’s population saw a ratio of about 1:1 among men and women with women outnumbering men. This confirmed our initial hypothesis that women were less likely to cycle than men. This disparity calls for urgent attention and research which we have outlined in the Further Exploration section (refer to Appendix A).

Weather

The data showed a clear preference to cycle in dry weather with at least 85% of trips being made when the weather is either sunny or dry each year. This confirms our hypothesis and therefore the link between favourable conditions and increased cycling. As a result we have recommended to TfL that they promote weather-resilient cycling infrastructure such as sheltered bike parking, as well as implement incentive programs to reward cyclists for choosing sustainable transportation during bad weather conditions.

References

1. Attitudes towards cycling (2016). Transport for London. https://content.tfl.gov.uk/attitudes-to-cycling-2016.pdf

2. A Moment of Change: Increasing Cycling Uptake By Department of Transport (2020): https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1005208/moment-of-change-increasing-cycling-uptake.pdf

3. Heinen, E., & Buehler, R. (2019). Bicycle parking: a systematic review of scientific literature on parking behaviour, parking preferences, and their influence on cycling and travel behaviour. Transport Reviews, 39(2), 198-220 https://eprints.whiterose.ac.uk/143013/1/Bicycle%20parking%20paper%20revision3%20final.pdf

4. Sex - Census 2021. (2021). Office for National Statistics. https://www.ons.gov.uk/datasets/TS008/editions/2021/versions/4/filter-outputs/9beb3fac-22bb-46ce-81b3-01e50a8e969e#get-data

5. Personal Security on London's Transport Network. (2022). London TravelWatch https://www.londontravelwatch.org.uk/wp-content/uploads/2022/01/Personal-security-on-Londons-Transport-network.pdf

6. Cycling potential in London's diverse communities. (2021). Transport for London. https://content.tfl.gov.uk/cycling-potential-in-londons-diverse-communities-2021.pdf

7. Bike paths in abandoned tube tunnels: is the London Underline serious: https://www.theguardian.com/cities/2015/feb/05/bike-paths-abandoned-tube-tunnels-london-underline

Appendices

Appendix A: Further exploration

- *Cycling Infrastructure:*

  As we recommended further investigation into the effectiveness of Cycle Superhighways and Quietways in increasing cycling uptake by borough, it would be beneficial to review existing research to inform the direction of analysis. Detailed research about the 'connected bikeability' concept has been identified: "Beecham, R., Tait, C., Lovelace, R. and Yang, Y. 2022. Connected bikeability in London: which localities are better connected by bike and does this matter?" In the publication, the term 'connected bikeability' is the extent to which a route network enables cycling in the context of selected origin-destination pairs. It has depicted a model in calculating the bikeability score. Taking reference of the model may be helpful in developing the methodology of assessing the effectiveness of Cycle Superhighways and Quietways.

- *Demographics:*

  In reports by London TravelWatch and TfL, the top cited reason by women for a reluctance to cycle surrounded personal safety and security. These reports were based on surveys conducted among women and minorities in London, offering a tangible direction in which to start thinking about alleviating the gender disparity in cycling.

  Our next step would be to verify through data how widespread the issue of women's personal security is, and if there are significant correlations between the counts of women cycling in different boroughs/regions and the rates of personal security failures such as Bicycle Theft. For this we will be making use of databases such as the London Metropolitan Police Service's Monthly Crime Dashboard in combination with available datasets on the gender of cyclists in an area.

- *Converting Disused Tube Stations into Dedicated Cycle Routes*

  With over 40 stations available for repurposing, converting disused tube stations into dedicated cycle routes could prove a viable method to address safety concerns, weather-related challenges, and promote transport efficiency while cycling.

  To assess the feasibility of repurposing these tube tunnels, future steps would include analysing historical records and providing cost estimates. In addition, steps should be taken to assess demographic data and conduct public opinion surveys to help identify potential users of the tunnels, and gauge public sentiment towards them. Integration with existing infrastructure is crucial and therefore geographic data could inform research about suitable connection points and accessibility.

- *City Comparisons*

  Refining and verifying the core factors influencing cycling uptake would involve adopting a global perspective. There can be important takeaways from the experiences of other major cities comparable to London, both within the UK, and in other countries. Studies such as "Making Cycling Irresistible: Lessons from The Netherlands, Denmark and Germany" by Buehler and Pucher (2007) and "Cycling behaviour in 17 countries…for what purpose, and how far?" by Goel et al. (2021) provide helpful pointers about the process by which to compare cycling behaviours and infrastructure across countries.

  Our next step in this case would be to structure our analytical approach, formulate our hypotheses, and make use of the New York and Sydney datasets provided by the client to compare cycling infrastructure and patterns in London, and to acquire other major city datasets if required.

Appendix B

Research / Data insights	Recommendations	How to Measure Success
Lack of bike parking discourages cycling / Statistically significant correlation between bicycle parking capacity and average daily cycle counts with a p-value close to zero	**Enhance the capacity of bicycle parking facilities** strategically across the city to accommodate growing demand. **Collaborate with public transport authorities** to integrate bike parking facilities seamlessly with transit hubs	**Monitor the frequency and capacity usage** or enhanced bike parking facilities, and regularly **collect feedback** from cyclists regarding the satisfaction with the new bike parking spaces. A successful outcome would involve an **increase in the number of bikes parked daily** and **a positive sentiment** reflected from users.
People must feel safe cycling / cycle exception signage counts' statistically significant correlation with Total cycles per head	Continue to **invest in safety features of the cycling infrastructure,** e.g. Turning exception signposts, Except cycle signposts.	**Track total cycle counts' trend & number of 'cycle except' arrangement to facilitate cycling safety** and **gather feedback from cyclists.**
Only 16% of Londoners in Sept 2016 have used Cycle Superhighways / Presence of Superhighways does not correlate with more cycling	**Keep monitoring utilisation and assess effectiveness of Cycle Superhighway & Quietway per location in helping increase cycling before further introducing them to other London boroughs**	**Track cycle counts' trend along each Superhighway** and **gather feedback from cyclists.**
Growth trend in Santander Bike hire	Launch **targeted Santander Bike campaigns** to rekindle interest and promote their benefits. Foster partnerships and **engagement with local communities** to understand preferences and barriers. Introduce **incentive programs** such as discounts or loyalty rewards for frequent Santander bike users.	**Measure the frequency of Santander bike rides over time**. **Evaluate community engagement** levels through event participation and survey responses. **Track the uptake of incentive programs.** Success will be reflected by **increased ride frequency**, **positive sentiment from the community**, and **increased participation in incentive programs**.