CONSUMER DATA RESEARCH CENTRE

USERS AGREEMENT

THIS AGREEMENT IS MADE BETWEEN

Tara Morovatdar, registered Student at **University of Koblenz-Landau** ("Student");

University of Koblenz-Landau, Germany-Koblenz-Robert Koch str 7-56073, ("Institution"); and

UNIVERSITY COLLEGE LONDON, represented by the Consumer Data Research Centre of Gower Street, London, WC1E 6BT ("CDRC").

INTRODUCTION

The Consumer Data Research Centre is an ESRC Big Data initiative, bringing together world-class researchers at University College London, and the Universities of Leeds, Liverpool and Oxford to harness the potential of consumer-related data to benefit social science researchers, businesses, government and society at large.

The CDRC requires potential users wishing to access safeguarded and controlled data services for research purposes to submit project proposals for approval by the Research Approvals Group (as defined below).

Once a Student is notified that their Project has been approved by receiving an Approved Project Notification, prior to being issued download instructions and password or site specific instructions for the assigned Secure Lab, the Student and their Institution must sign this CDRC User Agreement, including all Schedules hereto (the "Agreement"). The Agreement demonstrates that the Student and their Institution understand the seriousness of the undertaking, and that the Student and Institution understand the penalties that may be imposed for breaches of security or confidentiality, or any other breach of this Agreement.

DEFINITIONS

"Academic Supervisor" means the academic supervisor named in the Project application.

"Approved Project Notification" means the notification from the CDRC to the Student that their application to CDRC to carry out the Project has been approved, and shall set out any additional Stipulations that CDRC and/or a Data Partner may require the Student and their Institution to comply with under this Agreement, and which is annexed to this Agreement at Schedule 2, and may be updated from time to time by written notice from CDRC to the Student.

"Confidential Information" means information belonging to a Data Partner that is identified by the Data Partner as confidential in the Approved Project Notification and may include CDRC Data or Controlled Data.

"CDRC Data" means any combination of data which may be either Controlled or Safeguarded Data, and which may be subject to Stipulations.

"Controlled Data" means data which must be held under the most secure conditions with stringent access restrictions, including data which is Personal Information, and/or data that is considered commercially sensitive, and may also be subject to Stipulations.

"Data Partner" the organisation with whom the CDRC has entered into an agreement to provide Safeguarded and/or Controlled Data to the CDRC.

"**DPA**" means the Data Protection Act 1998 and any subsequent legislation or regulations.

"**End Date**" means the end date of the Project as set out in the Approved Project Notification.

"Institutional Signatory" means a signatory who is validly authorised to enter into contracts on behalf of the Institution.

"**Output**" means a pre-approved Project output as described in the Project.

"**Output Request and Release Process**" means the multi-stage Output approval whereby CDRC data scientists may check that Outputs to be released from the Secure Lab are compliant with the Approved Project Notification and are adequately non-disclosive.

"Personal Information" means personal information as defined by the DPA.

"Project" The particular research project proposal approved by the CDRC Research Approvals Group and approved in the Approved Project Notification, and which is annexed to this Agreement as Schedule 1.

"Research Approvals Group (RAG)" means the independent research approvals group whose role is to approve proposed project applications to use CDRC Data.

"Safeguarded Data" means data to which access is restricted due to licence conditions, but where data are not considered "personally-identifiable" or otherwise sensitive.

"Secure Lab" means the secure laboratory facilities within the CDRC, which comprises three secure laboratory facilities at UCL, the University of Leeds and the University of Liverpool. Each lab has its own site specific terms and conditions which Students and their Institutions may also be required to sign in order to obtain access to CDRC Data through each respective Secure Lab location.

"**Start Date**" means the start date of the Project as set out in the Approved Project Notification.

"**Stipulations**" means Project specific terms placed by CDRC and/or a Data Partner, as notified to the Student and their Institution under the Approved Project Notification with which the Student and Institution must comply in their use of the CDRC Data under this Agreement.

IT IS HEREBY AGREED:

1. Use of Data

A. If required within the Approved Project Notification, the Student must, and their Institutions must procure that the Student has satisfactorily completed a CDRC approved accredited safe researcher training course within the specified time period.

B. The Student shall be responsible for processing all CDRC Data in accordance with all applicable laws and all regulatory standards applicable to such CDRC Data.

C. The Student will ensure that any Stipulations as specified in the Approved Project Notification are adhered to.

D. Access to the CDRC Data is being provided for the statistical analysis and research as detailed in the Project and as further outlined in the Approved Project Notification. The

CDRC Data shall not be used or processed for any other purpose without the prior written consent of the CDRC and the Data Partner(s).

E. The Student shall not disclose nor compromise any of the CDRC Data from the individual records obtained or produced from the CDRC Data to anyone other than (i) those approved for the same Project and (ii) CDRC staff involved in the review of the Outputs.

F. The Student shall ensure that no attempts are made to link the CDRC Data to any other files in order to relate the particulars to any identifiable individual person, business or organisation unless such data linkage exercise has been explicitly approved by the CDRC as part of the Project.

G. On termination of the Agreement for whatever reason, all access to the CDRC Data related to the Project shall cease forthwith, and electronic access be denied. The Student must permanently destroy/delete or erase the Safeguarded Data that is in their possession within seven (7) days of the end of this Agreement, together with all hard or soft copies of the same and certify such destruction in writing (which may include email) to the CDRC.

H. The CDRC reserves the right to monitor, record, and audit, or to request a written report from the Student regarding, the use and activities relating to the use, of the CDRC Data by the Student during the lifetime of this Agreement.

I. During the lifetime of this Agreement and six months thereafter, the Student agrees that, and the Institution shall procure that, the CDRC will have right of entry to any premises where the CDRC Data is accessed in order to monitor, record or audit the use and activities relating to the use of the CDRC Data, upon the provision of reasonable notice to the Student.

J. Any incidents of unauthorised access to, processing of, or disclosing of, the CDRC Data must be reported immediately to the CDRC.

K. The Agreement is subject to review and without limitation whenever a change in the law, contracts for services with third parties, other procedures or other relevant circumstances takes place.

2. Output Release for CDRC Data

A. The Student must not reproduce, to any extent whatsoever, any CDRC Data, original datasets or copies or subsets of any CDRC Data.

B. Students are responsible for, and their Institution must procure that the Student ensures that rules and regulations for disclosure risk analysis as specified by statistical disclosure guidelines (explained during accredited safe researcher training) are applied prior to submission of analytical outputs for clearance and release.

C. No Outputs may be removed from the CDRC Secure Lab by the Student or before the Student satisfactorily completes the follow the Output Request and Release process.

D. The CDRC reserves the right to release in whole or in part, an amended version or not to release at all, as the CDRC deems appropriate, the proposed Output produced by the Student pursuant to this Agreement.

E. The Student, and the Institution agrees to work with the CDRC to meet the requirements of the Output Request and Release process. In the event that the CDRC decides not to release the proposed Output, if feasible the Student may, at CDRC's sole discretion, be allowed to revisit the Secure Lab to rectify the problem. The final decision to release an Output rests with the CDRC.

F. No CDRC Data shall be used for any commercial purpose whatsoever.

3. CDRC Data, Copyright and Publication Protocol

A. The CDRC Data and related documentation shall at all times be and remain the sole and exclusive property of the CDRC and/or the Data Partner, as the case may be. This Agreement pertains to the use of the CDRC Data and related documentation to produce a pre-approved Output for research purposes and nothing contained herein shall be deemed to convey any title or ownership interest in the CDRC Data or the related documentation to the Student.

B. Copyright in Outputs may be held singly or jointly by the Student(s) that created them, their Institution(s) or their funder(s) according to the Student's funding and Institution's agreements.

C. Any publications in connection with this Project (including but not limited to draft versions, final versions, abstracts, presentations, meetings or seminars) may only be published upon receiving written approval from CDRC. The procedure to obtain approval from CDRC shall be that the Student shall submit any proposed publication to CDRC via publications@cdrc.ac.uk strictly in accordance with the timescales specified in the Approved Project Notification. If CDRC and/or the Data Partner, at their sole discretion, determine that the proposed publication contains any Confidential Information, it shall notify the Student to ensure that such Confidential Information is deleted. Following any such deletion, CDRC shall notify the Student in writing that they may proceed with the proposed publication. CDRC and/or the Data Partner may also request that the Student delay the publication if the CDRC and/or Data Partner reasonably believes that the CDRC Data has been used outside the scope of the Project or the Approved Project Notification and the Student will make reasonable attempts to amend the publication to the satisfaction of the CDRC and/or Data Provider.

D. The Student shall comply with any further Stipulations in connection with publications as may be specified in the Approved Project Notification.

4. General

A. This Agreement shall commence on the Start Date and shall end on the End Date, subject to any earlier termination in accordance with clause 4.B or clause 4.C. Where the Agreement is due to terminate under this clause 4.A and provided the CDRC is free and able to do so, the parties may agree in writing for the continuing use of the CDRC Data in relation to the Project for a specific agreed additional period of time. The terms of this Agreement shall continue to apply to any such continuing use.

B. Without prejudice to any other right or remedy, if the Student and/or the Institution shall commit any breach of, or fail to comply with, any of their obligations under this Agreement, become bankrupt or any judgment is made against either the Student or the Institution and remains unsatisfied for seven (7) days, CDRC shall be entitled to terminate this Agreement forthwith on written notice.

C. The CDRC reserves the right to terminate this Agreement with immediate effect on written notice in the event that the Data Partner terminates the agreement to provide CDRC Data to the CDRC.

D. The Student and Institution acknowledge that any disclosure of CRDC Data other than in accordance with this Agreement may cause irreparable harm to the CDRC and/or a Data Partner. The Student, and Institution hereby agree that CDRC shall, in accordance to any

other rights or remedies to which it may be entitled, have the right to seek and obtain equitable and or injunctive relief.

E. CDRC does not warrant the accuracy or completeness of the CDRC Data, nor its fitness for any particular purpose. The CDRC Data is provided to Student "as is" and CDRC does not warrant that the CDRC Data will enable the Student to achieve the objectives of the Project, or that the CDRC Data is suitable for use in any report or for academic or research purposes.

F. The Student and/or Institution as the case may be are required to bring directly to the attention of CDRC any matters or events that may affect any of the obligations under this Agreement immediately as awareness of the matters or events may arise.

G. No failure or delay on the part of CDRC to exercise any right or remedy under this Agreement shall be construed or operate as a waiver thereof, nor shall any single or partial exercise of any right or remedy preclude the further exercise of such right or remedy.

H. The Student, and the Institution acknowledge that the CDRC may hold and process information submitted in the Project for validation and statistical purposes, and for the purposes of the management of the service and may also pass such information to other parties such as Data Partners.

I This Agreement is made under English law and the parties submit to the exclusive jurisdiction of the English courts. Notwithstanding the preceding sentence, CDRC may at any time bring proceedings for an injunction in any court of competent jurisdiction.

J. By signing this Agreement, the Student and Institution represent and undertake that: the information supplied in the Project and any supporting documentation is accurate to the best of their knowledge, and that they have read and understand and shall obey the conditions specified in this Agreement.

[SIGNATURES ON FOLLOWING PAGE]

Signed by:

STUDENT

Signature 

Name (PRINT):Tara Morovatdar

Date: 11.12.2019

INSTITUTION:

Signature: 

Universität Koblenz-Landau
Institut für Wirtschafts- und Verwaltungsinformatik
Postfach 20 16 02
D-56016 Koblenz
Tel.: (0261)287-2646
http://www.uni-koblenz.de/~iwvi

Name (print): Prof.Dr. Maria Wimmer

University of Koblenz-Landau, Dean of Faculty of Computer Science

Date: 2nd Dec, 2019

UNIVERSITY COLLEGE LONDON

Signature:

Name (print):

[Institution Name]

Date:

READ AND ACKNOWLEDGED BY THE ACADEMIC SUPERVISOR:

Signature:  

Name (print): Dr. Johann Schaible, JProf. Dr. Claudia Wagner

[Institution Name] GESIS - LEIBNIZ INSTITUTE FOR THE SOCIAL SCIENCES, UNIVERSITY OF KOBLENZ-LANDAU

Date: 12.12.2019

Consumer Data Research Centre

An ESRC Data Investment

Schedule 1: Approved Project Proposal

CONSUMER DATA RESEARCH CENTRE

Safeguarded Data
Project Proposal Form

For office use: Ref 297-07

Consumer Data Research Centre

An ESRC Data Investment

INTRODUCTION

This form should be completed by those wishing to access the Consumer Data Research Centre's (CDRC) safeguarded data collections. You should consult the ***CDRC Data Service User Guide*** before completing the form. A CDRC data scientist will be assigned to you who can provide support in the application process.

Once submitted your proposal will be forwarded to the CDRC Research Approvals Group (RAG) for independent assessment. Projects will be assessed based on the criteria listed in Appendix 1.

Approval to access data will not be granted without evidence that the applicant has acquired ethical approval for the research through their institution, or supplied evidence that this is not applicable. For non-academic projects, where there is no approval process in place, the CDRC will assist the applicant in acquiring this.

PART A. PROJECT DETAILS

1. Contact Details

1.1. Lead applicant: Tara Morovatdar
Department and Institution: University of Koblenz-Landau
Address: Germany-Koblenz-Robert Koch str 7-56073
Email: tmorovatdar@uni-koblenz.de
Telephone: +4917630174409

1.2. Initial Proposal Form Reference: Project 297

1.3. Title of Project

A fine grained station usage prediction for the Santander Bike-Sharing System in London

1.4. Access Details

Access requested from[1]: 1.09.2019
Access requested to (Date): 1.09.2020

2. Project Proposal Details

2.1. Abstract. Appropriate for a general audience. This may be used by the CDRC for reporting and publicity purposes.

This study aims at analyzing the usage of shared bike stations across London. We analyze the capacity of each station (also called "dock"), i.e, the number of used and free bicycle parking racks of a station, at different times during day and night. This allows for predicting

[1] Please note the RAG review process takes approximately 3 weeks.

how many free bike racks will a station have at a specific time point. Typically, customers of a bike-sharing system take a bike at one location to go to some other location in the city. Arriving at the desired location, they will need to return the bike at a near station. However, the nearest station might not have any more free bike racks, such that the customer will need to search for another station with free bike racks. Predicting the number of free bike racks might alleviate the situation and help the customers to go directly to a station with free bike racks. Also, it might be helpful for the bike-sharing provider in better redistributing the bikes.

2.2. **Project description.** A detailed description of the project, documenting the motivation, scope and aims of the intended research as well as the methods you will use in the proposed research.

People all around the world are using more and more bikes from Bike-Sharing Systems (BSS) and this usage is growing on a daily basis. According to Shaheen et al.[2] bike-sharing systems are good for the environment, because of the decreased usage of cars and because it is good for public health as more people are being active using bicycles. As Midgely[3] has mentioned "smart bike-sharing systems provide the missing link between existing points of public transportation and desired destinations". He also mentioned that an increase in mobility choices improves air quality and the reduction of congestion, which are general objectives of bike-sharing systems. Etienne and Latifa[4] introduce walking and biking as a soft mode transport which is a different solution to organize passenger mobility. They point out that BSS -- as a cheap way of transport -- are very suitable to reduce traffic in the cities. They also mention that because bikes do not produce green-house gases and noise, BSS can help the environment. Purnama[5] mentioned that people have more flexibility to choose their origin, route, destination and travel time based on their needs when they are using BSS.

Because of all these features, bike-sharing systems are growing very fast in the world, so it becomes more and more necessary to improve these systems through analysing and understanding their underlying structure.

Bike-sharing systems have some common issues that can be addressed. In station-based bike-sharing systems, an unbalanced bike distribution in the stations (also called docks) is one of the most important issues. At each station the number of bike racks is limited, such that it is not possible to return a bike to a full station or pick up a bike from an empty one. The bikes in a bike sharing system need to be re-balanced frequently because at different stations in different periods of time the check-in and check-out of the bikes are unbalanced. In other words, the system usage is asymmetric and fluctuates during the day. Some other factors like altitude differences, weather conditions or events in the city may cause irregular

[2] Susan Shaheen, Stacey Guzman, and Hua Zhang. Bike-sharing in Europe, the Americas, and Asia: past, present, and future. Transportation Research Record: Journal of the Transportation Research Board, (2143):159–167, 2010.

[3] Peter Midgley. The role of smart bike-sharing systems in urban mobility.Journeys,2(1):23–31, 2009.

[4] Côme Etienne and Oukhellou Latifa. Model-based count series clustering for bike-sharing system usage mining: a case study with the vélib'system of paris.ACM Transactions on Intelligent Systems and Technology (TIST), 5(3):39, 2014.s

[5] Ida Bagus Irawan Purnama. Spatiotemporal analysis of urban mobility dynamics: a case study of bicycle sharing system. 2018.

or asymmetric demands and flow of the bikes. Bike-sharing providers know that the availability of free bikes and bike racks in the stations is one of the most important aspects in BSS, so they have tried different solutions, i.e., redistributing the bikes with trucks or giving incentive to users to keep the travels short(i.e. Usually the first 30 minutes of a ride are often for free and after that incur charges[6] or the fees increase progressively[7]) so the bikes are more available, to solve the unbalanced distribution of the bikes.

The users benefit from the flexibility of bike-sharing systems but it poses a major challenge and recurrent problem for the operators to satisfy the forthcoming demand with the limit in the resources. Because if a user attempts to drop off a bike at a full station or tries to rent a bike from an empty station the user would experience a bad quality of service when trying to use the BSS[8].

Reallocating the bikes takes too much time so it is too late to decide to do that when you detect an imbalance, even if it is you always monitor the stations and detect the imbalance instantly after it occurs.

Liu et al[9] conducted a study on NYC CitiBike to develop a model that helps the bike-sharing system to avoid expensive re-balancing and to get better user experience. They tried to predict station demand which is defined as average pick-up/hour with neural network and also station balance which is the situation if the overall pick-up (or drop-off) for a period of time is greater than a threshold with the use of the neural network. Nair et al.[10] conducted an empirical analysis on Paris bike-sharing system Vélib, regarding the station's location they found that a close coupling of transit(metro stations) and bike station can increase utilization, then they developed a mixed-integer program. The model outputs where and how many bicycles to rearrange. This helps operators to redistribute bikes so that they meet a reliability level. In 2016 Yang, Zidong, et al.[11] proposed a Spatio-temporal bicycle mobility model with the use of historical data from the Chinese city Hangzhou which has the world's largest BSS. For the first time, they predicted check-in and check-outs on a station level.Their goal was to advance the bike-rebalancing algorithm designs. Using their mobility model with a time-varying parameter for the check-in prediction and random forest for check-out prediction they reached outperforming results.

---

[6] Nair, R., Miller-Hooks, E., Hampshire, R. C. and Bušı ́c, A. [2013], 'Large-scale vehicle sharing systems: analysis of vélib",International Journal of Sustainable Transportation7(1), 85–106

[7] ogel, P., Greiser, T. and Mattfeld, D. C. [2011], 'Understanding bike-sharing systems using data mining: Exploring activity patterns',Procedia-Social and Behavioral Sciences20, 514–523.

[8] Adish Singla, Marco Santoni, Gábor Bartók, Pratik Mukerji, Moritz Meenen, and Andreas Krause. Incentivizing users for balancing bike sharing systems. InAAAI, pages723–729, 2015.

[9] Junming Liu, Qiao Li, Meng Qu, Weiwei Chen, Jingyuan Yang, Hui Xiong, Hao Zhong, and Yanjie Fu. Station site optimization in bike sharing systems. In2015 IEEE International Conference on Data Mining pages 883–888. IEEE, 2015

[10] Nair, R., Miller-Hooks, E., Hampshire, R. C., & Bušić, A. (2013). Large-scale vehicle sharing systems: analysis of Vélib'. International Journal of Sustainable Transportation, 7(1), 85-106.

[11] Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J., & Moscibroda, T. (2016, June). Mobility modeling and prediction in bike-sharing systems. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (pp. 165-178). ACM.

In Our research, to alleviate the unbalanced situation, we propose to predict the bike usage of each station, so planning ahead will be easier, one example is that one can use our model results to reduce the costs of redistribution. Two different dependencies affect the imbalance in bike distribution: a temporal one, which means at different times of the day and week the use of the bikes is different. The other dependency is spatial due to short one-way trips. As mentioned above some bike-sharing providers are using trucks to move bikes from full stations to empty ones and some promote users to rent bikes for short times so the bikes are available for everyone[12]. For example, Santander uses both ways to make sure that bikes are available for more people. Santander gives users incentive to keep the trips short by the policy they have about the trips which says the first 30 minutes are free of charge and any minute above that (up to 30 minutes) it costs £2. They also use trucks to relocate the bikes so the bikes are available everywhere.

In this work, we aim to design a model that predicts the capacity of a specific station at a given time. In order to develop a model capable of predicting the odds of a station having an empty dock in a time window of 30 minutes, (non)linear regression is conducted. When a customer hires a bike, the Santander app can recommend the 6 closest stations to the user destination sorted by the probability of having an available empty dock during the next 30 minutes.

Using the Docking Station Locations data we tend to observe how much the number of bikes and stations has been increased during the years. Using regression analysis can help us to understand the pattern that the Bike-sharing system has been changed during the years so the system administrators can plan for the future better.

2.3. Research Category:

2.3.1.Is this request for an Undergraduate, Masters project? U'grad ☐ Masters ☒

2.3.2.Is this request for a PhD project? PhD ☐

2.3.3. Is your project funded, commissioned or sponsored by a funding body or any other organisation? Yes ☐ No ☒ Funding application in progress ☐

Please include the name, postal and web address of your current or prospective funder, and your grant/project reference number (if applicable).
Click here to enter text.

2.4. Project Impact. Please describe the anticipated scientific and societal benefits of the project and the ways in which you intend to maximise those benefits.

This master thesis will help researchers to better understand human mobility in the context of bike-sharing systems, to identify the underlying patterns. It also can help the authorities to

[12] Patrick Vogel, Torsten Greiser, and Dirk Christian Mattfeld. Understanding bike-sharing systems using data mining: Exploring activity patterns. Procedia-Social and Behavioural Sciences, 20:514–523, 2011.

plan for the future better. Nevertheless, the aim of this research is to provide a recommendation system which gives the users a bike station recommendation list based on their destination so it gives a better feeling to users if they don't face a situation that the station does not have an empty dock.

2.5. End Users. Who are the main end users of this research (academic research, central government, consultancy, industry, local government, NHS, public sector, third sector? List all that are applicable.)

Academic research is the main end user for this work. Naturally, the results can also be used by others, such as the local government (e.g., improve bike lane situations in specific areas) or industry (e.g. better plan redistribution plans for bikes).

2.6. Outputs and Publications. What are the intended outputs or publications arising from the use of these data? (For example, journal articles, PhD thesis, report for government department, policy documents for a local authority, White Papers, new software or other tools, etc.)

This data is used for a Master thesis. We plan to publish a comprised version of it as an article in a journal or a conference

2.7. Research Team

Please list the names, affiliation and email addresses of all known members of your research team who are requesting access to the CDRC safeguarded data. If you are a student please include your academic supervisor.

To access some of the CDRC safeguarded data products it is necessary for users to have successfully completed a safe researcher training course. Please see the terms of use information for the particular dataset.

Consumer
Data
Research
Centre

An ESRC Data
Investment

For office use: Ref 297-07

Consumer Data Research Centre

An ESRC Data Investment

Research Team *(including lead applicant, add more rows if required)*

Title, Name	**Department/ Institution**	**Institutional email address**	**Comp traini speci comp**
JProf.Dr.Claudia Wagner	Assistant professor in Computer Science at University Koblenz-Landau and the interim Scientific Director of the department Computational Social Science at GESIS - Leibniz Institute for the Social Sciences.	clwagner@uni-koblenz.de	
Dr.Johann Schaible	Postdoctoral researcher and Team Lead of team "Knowledge Discovery" of the Computational Social Science department at GESIS - Leibniz Institute for the Social Sciences	johann.schaible@gesis.org	

2.8. Ethical Approval

2.8.1. Have you sought or are you seeking **ethical approval from an institutional ethical approvals panel** or any other appropriate body?

Yes ☐ No ☒

2.8.2. If *Yes*, please provide **evidence** of the status of the application or the outcome of the ruling issued. Please list what evidence you are enclosing below and return it as a separate attachment in PDF format when you return this application form. Feel free to add any comments below.

Click here to enter text.

2.8.3. If *No*, please bring this to the attention of the CDRC as soon as possible so that routes for ethical approval may be discussed.

Click here to enter text.

2.8.4. If you believe that ethical approval is not required for your research, please provide justification for this below.

There is no data concerning user information. Our research is simply using information about bike stations and bike-sharing patterns. That is why we believe no ethical approval is required.

Consumer Data Research Centre

An ESRC Data Investment

PART B. DATA REQUEST

3.1. Data Required. Please provide the following information for each dataset requested. Please add more lines if required.

Data Partner	Data Set	Access to Full Data Set requested or specific variables (list)	Geographic Extent
	London Bicycle Sharing System Docking Station Individual Observations	All Observations - all the columns	*London*
	London Bicycle Sharing System Docking Station Locations	All the columns	*London*

PART C: DECLARATION

By completing this declaration I hereby declare that the information included in this application form is true and correct to the best of my knowledge. I understand that any false or misleading information given by me in connection with my application may result in termination of the application process and/or other sanctions.

I also agree that I will be the single point of contact for progress updates and communication regarding the progress of the application.

I agree for my personal information to be used for the purposes of processing this application in accordance with the relevant data laws of the UK.

☑ I consent to my contact details being added to the CDRC contacts database so that the CDRC can send me notifications of CDRC related activities.

☑I understand that forwarding this form by email constitutes an electronic signature.

☑ I understand that final approval for this project may require the additional submission of project approval forms.

Name: Tara Morovatdar

Date: 20/08/2019

4. Data Linkage

4.1. Data Linkage. If your project will be linking more than one data source, describe which data sources will be linked and how the linkage will be done, including any specific variables that need to be linked (if known). If any of the data to be linked has identifying information as defined in the Data Protection Act 1998 or General Data Protection Regulations please provide details. Please note that no project that has the potential to re-identify individuals through data linkage will be approved.

In this work, we want to use Santander Cycle Hire journeys dataset which is provided by Transport for London (TfL) and is an open dataset. As I have asked Oliver O'Brien, the station id in Santander Cycle Hire journeys is the same as ucl_id in Bike location dataset and operator_id in London Bicycle Sharing System Docking Station Individual Observations. So I can use these columns to link the datasets.

APPENDIX 1: RAG CRITERIA FOR ASSESSMENT

The role of the RAG is to provide independent and transparent assessments of applications by researchers for access to data through both the CDRC Safeguarded and CDRC Secure services based on a set of standard evaluation criteria. RAG is independent to the CDRC and will include representation from the academic, big data, industrial sectors as well as the data partners concerned. For full Terms of Reference and membership see https://www.cdrc.ac.uk/data-services/using-our-data/

Criteria for Approval

- **Scientific advancement –** how the project has the potential to advance scientific knowledge, understanding and/or methods using consumer data;
- **Public good** – how the project has the potential to provide insight and/or solutions that could benefit society;
- **Privacy and ethics –** the potential privacy impacts or risks, and wider ethical considerations relating to the project
- **Project Design and Methods –** how the project will be conducted and who will be involved with a focus on demonstrating project feasibility.
- **Cost and resources issues –** what impact the project is likely to have on CDRC resources, including CDRC staff time and use of infrastructure, as well as any data acquisition costs. Resource requirements should be justified.

Schedule 2: Approved Project Notification

Approved Project Notification

Project Reference Number: 297-01
Project Title: A fine grained station usage prediction for the Santander Bike-Sharing System in London
CDRC User(s): Tara Morovatdar, Prof.Dr.Claudia Wagner, Dr.Johann Schaible
Institution: University of Koblenz-Landau
Proposal Accepted: 23/09/2019

Stipulations imposed by CDRC, including those from RAG members and Data Partner(s)

1. CDRC

1.1. The CDRC User(s) must only use the data for the purpose of research as described in the approved Project Proposal 297-01.
1.2. Approval for use of the data for this purpose is from the Start Date 23/09/2019 until the End Date 01/09/2020. If the CDRC User(s) requires access to this data beyond the End Date a request for an extension may be submitted to the CDRC with an explanation on why this is required.
1.3. By the end date the data and all copies of the data must be deleted. Notification must be sent to the CDRC by email to info@cdrc.ac.uk within 7 days of this date to confirm that the deletion has been completed.
1.4. There is no requirement to complete safe researcher training to access safeguarded Bicycle Sharing System Docking Stations data.
1.5. The academic supervisors of student CDRC Users are required to ensure that their students adhere to all the stated stipulations.
1.6. There is no requirement when accessing this dataset to submit for review a copy of any proposed publication (including reports, abstracts, or presentation to a journal, editor, meeting, seminar or other third party) to the CDRC prior to submission for publication or presentation.
1.7. Users are required to deposit copies of working papers, peer-reviewed journal articles, logs of impact and other publications for access with the CDRC site wherever copyright permits. Where a proposed publication has been produced using more than one dataset, the User should adhere to the partner stipulation which has the longest review period. Where this is not possible, full references to research outputs are required for CDRC audit purposes. Please email publications@cdrc.ac.uk when publications are ready for deposit or logging.
1.8. Published outputs must include an acknowledgement stating "The data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC 297-01, ES/L011840/1; ES/L011891/1".
1.9. Users are requested to provide a brief case study of the supported work to be included in the Research Outputs section of the CDRC website https://www.cdrc.ac.uk/research/

On behalf of CDRC: Nick Bearman

NE Bearman

Date: 23/09/2019

CONSUMER DATA RESEARCH CENTRE

USERS AGREEMENT

THIS AGREEMENT IS MADE BETWEEN

Dr. Johann Schaible, at **GESIS – Leibniz Institute for the Social Sciences** ("User");

GESIS – Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, 50667 Köln, Germany, ("Institution"); and

University College London, represented by the Consumer Data Research Centre of Gower Street, London, WC1E 6BT ("CDRC").

INTRODUCTION

The Consumer Data Research Centre is an ESRC Big Data initiative, bringing together world-class researchers at University College London, and the Universities of Leeds, Liverpool and Oxford to harness the potential of consumer-related data to benefit social science researchers, businesses, government and society at large.

The CDRC requires users wishing to access safeguarded and controlled data services for research purposes to submit project proposals for approval by the Research Approvals Group (as defined below).

Once a User is notified that their Project has been approved by receiving an Approved Project Notification, prior to being issued download instructions and password or site specific instructions for the assigned Secure Lab, the User and their Institution must sign this CDRC User Agreement, including all Schedules hereto (the "Agreement"). The Agreement demonstrates that the User and their Institution understand the seriousness of the undertaking, and that the User and Institution understand the penalties that may be imposed for breaches of security or confidentiality, or any other breach of this Agreement.

DEFINITIONS

"Approved Project Notification" means the notification from the CDRC to the User that their application to CDRC to carry out the Project has been approved, and shall set out any additional Stipulations that CDRC and/or a Data Partner may require the User and their Institution to comply with under this Agreement, and which is annexed to this Agreement at Schedule 2, and may be updated from time to time by written notice from CDRC to the User.

"Confidential Information" means information belonging to a Data Partner that is identified by the Data Partner as confidential in the Approved Project Notification and may include CDRC Data or Controlled Data.

"**CDRC Data**" means any combination of data which may be either Controlled or Safeguarded Data, and which may be subject to Stipulations.

"Controlled Data" means data which must be held under the most secure conditions with stringent access restrictions, including data which is Personal Information, and/or data that is considered commercially sensitive, and may also be subject to Stipulations.

"Data Partner" the organisation with whom the CDRC has entered into an agreement to provide Safeguarded and/or Controlled Data to the CDRC.

"**DPA**" means the Data Protection Act 1998 and any subsequent legislation or regulations.

"**End Date**" means the end date of the Project as set out in the Approved Project Notification.

"Institutional Signatory" means a signatory who is validly authorised to enter into contracts on behalf of the Institution.

"**Output**" means a pre-approved Project output as described in the Project.

"**Output Request and Release Process**" means the multi-stage Output approval whereby CDRC data scientists may check that Outputs to be released from the Secure Lab are compliant with the Approved Project Notification and are adequately non-disclosive.

"Personal Information" means personal information as defined by the DPA.

"Project" The particular research project proposal approved by the CDRC Research Approvals Group and approved in the Approved Project Notification, and which is annexed to this Agreement as Schedule 1.

"Research Approvals Group (RAG)" means the independent research approvals group whose role is to approve proposed project applications to use CDRC Data.

"Safeguarded Data" means data to which access is restricted due to licence conditions, but where data are not considered "personally-identifiable" or otherwise sensitive.

"Secure Lab" means the secure laboratory facilities within the CDRC, which comprises three secure laboratory facilities at UCL, the University of Leeds and the University of Liverpool. Each lab has its own site specific terms and conditions which Users and their Institutions may also be required to sign in order to obtain access to CDRC Data through each respective Secure Lab location.

"**Start Date**" means the start date of the Project as set out in the Approved Project Notification.

"**Stipulations**" means Project specific terms placed by CDRC and/or a Data Partner, as notified to the User and their Institution under the Approved Project Notification with which the User and Institution must comply in their use of the CDRC Data under this Agreement.

IT IS HEREBY AGREED:

1. Use of Data

A. If required within the Approved Project Notification, the User must, and their Institutions must procure that the User has satisfactorily completed a CDRC approved accredited safe researcher training course within the specified time period.

B. The User shall be responsible for processing all CDRC Data in accordance with all applicable laws and all regulatory standards applicable to such CDRC Data.

C. The User will ensure that any Stipulations as specified in the Approved Project Notification are adhered to.

D. Access to the CDRC Data is being provided for the statistical analysis and research as detailed in the Project and as further outlined in the Approved Project Notification. The CDRC Data shall not be used or processed for any other purpose without the prior written consent of the CDRC and the Data Partner(s).

E. The User shall not disclose nor compromise any of the CDRC Data from the individual records obtained or produced from the CDRC Data to anyone other than (i) those approved for the same Project and (ii) CDRC staff involved in the review of the Outputs.

F. The User shall ensure that no attempts are made to link the CDRC Data to any other files in order to relate the particulars to any identifiable individual person, business or organisation unless such data linkage exercise has been explicitly approved by the CDRC as part of the Project.

G. On termination of the Agreement for whatever reason, all access to the CDRC Data related to the Project shall cease forthwith, and electronic access be denied. The User must permanently destroy/delete or erase the Safeguarded Data that is in their possession within seven (7) days of the end of this Agreement, together with all hard or soft copies of the same and certify such destruction in writing (which may include email) to the CDRC.

H. The CDRC reserves the right to monitor, record, and audit, or to request a written report from the User regarding, the use and activities relating to the use, of the CDRC Data by the User during the lifetime of this Agreement.

I. During the lifetime of this Agreement and six months thereafter, the User agrees that, and the Institution shall procure that, the CDRC will have right of entry to any premises where the CDRC Data is accessed in order to monitor, record or audit the use and activities relating to the use of the CDRC Data, upon the provision of reasonable notice to the User.

J. Any incidents of unauthorised access to, processing of, or disclosing of, the CDRC Data must be reported immediately to the CDRC.

K. The Agreement is subject to review and without limitation whenever a change in the law, contracts for services with third parties, other procedures or other relevant circumstances takes place.

2. Output Release for CDRC Data

A. The User must not reproduce, to any extent whatsoever, any CDRC Data, original datasets or copies or subsets of any CDRC Data.

B. Users are responsible for, and their Institution must procure that the User ensures that rules and regulations for disclosure risk analysis as specified by statistical disclosure guidelines (explained during accredited safe researcher training) are applied prior to submission of analytical outputs for clearance and release.

C. No Outputs may be removed from the CDRC Secure Lab by the User or before the User satisfactorily completes the follow the Output Request and Release process.

D. The CDRC reserves the right to release in whole or in part, an amended version or not to release at all, as the CDRC deems appropriate, the proposed Output produced by the User pursuant to this Agreement.

E. The User and the Institution agrees to work with the CDRC to meet the requirements of the Output Request and Release process. In the event that the CDRC decides not to release the proposed Output, if feasible the User may, at CDRC's sole discretion, be allowed to revisit the Secure Lab to rectify the problem. The final decision to release an Output rests with the CDRC.

F. No CDRC Data shall be used for any commercial purpose whatsoever.

3. CDRC Data, Copyright and Publication Protocol

A. The CDRC Data and related documentation shall at all times be and remain the sole and exclusive property of the CDRC and/or the Data Partner, as the case may be. This Agreement pertains to the use of the CDRC Data and related documentation to produce a

pre-approved Output for research purposes and nothing contained herein shall be deemed to convey any title or ownership interest in the CDRC Data or the related documentation to the User.

B. Copyright in Outputs may be held singly or jointly by the User(s) that created them, their Institution(s) or their funder(s) according to the User's funding and Institution's agreements.

C. Any publications in connection with this Project (including but not limited to draft versions, final versions, abstracts, presentations, meetings or seminars) may only be published upon receiving written approval from CDRC. The procedure to obtain approval from CDRC shall be that the User shall submit any proposed publication to CDRC via publications@cdrc.ac.uk strictly in accordance with the timescales specified in the Approved Project Notification. If CDRC and/or the Data Partner, at their sole discretion, determine that the proposed publication contains any Confidential Information, it shall notify the User to ensure that such Confidential Information is deleted. Following any such deletion, CDRC shall notify the User in writing that they may proceed with the proposed publication. CDRC and/or the Data Partner may also request that the User delay the publication if the CDRC and/or Data Partner reasonably believes that the CDRC Data has been used outside the scope of the Project or the Approved Project Notification and the User will make reasonable attempts to amend the publication to the satisfaction of the CDRC and/or Data Provider.

D. The User shall comply with any further Stipulations in connection with publications as may be specified in the Approved Project Notification.

4. General

A. This Agreement shall commence on the Start Date and shall end on the End Date, subject to any earlier termination in accordance with clause 4.B or clause 4.C. Where the Agreement is due to terminate under this clause 4.A and provided the CDRC is free and able to do so, the parties may agree in writing for the continuing use of the CDRC Data in relation to the Project for a specific agreed additional period of time. The terms of this Agreement shall continue to apply to any such continuing use.

B. Without prejudice to any other right or remedy, if the User and/or the Institution shall commit any breach of, or fail to comply with, any of their obligations under this Agreement, become bankrupt or any judgment is made against either the User or the Institution and remains unsatisfied for seven (7) days, CDRC shall be entitled to terminate this Agreement forthwith on written notice.

C. The CDRC reserves the right to terminate this Agreement with immediate effect on written notice in the event that the Data Partner terminates the agreement to provide CDRC Data to the CDRC.

D. The User and Institution acknowledge that any disclosure of CRDC Data other than in accordance with this Agreement may cause irreparable harm to the CDRC and/or a Data Partner. The User and Institution hereby agree that CDRC shall, in accordance to any other rights or remedies to which it may be entitled, have the right to seek and obtain equitable and or injunctive relief.

E. CDRC does not warrant the accuracy or completeness of the CDRC Data, nor its fitness for any particular purpose. The CDRC Data is provided to User "as is" and CDRC does not warrant that the CDRC Data will enable the User to achieve the objectives of the Project, or that the CDRC Data is suitable for use in any report or for academic or research purposes.

F. The User and/or Institution as the case may be are required to bring directly to the attention of CDRC any matters or events that may affect any of the obligations under this Agreement immediately as awareness of the matters or events may arise.

G. No failure or delay on the part of CDRC to exercise any right or remedy under this Agreement shall be construed or operate as a waiver thereof, nor shall any single or partial exercise of any right or remedy preclude the further exercise of such right or remedy.

H. The User and the Institution acknowledge that the CDRC may hold and process information submitted in the Project for validation and statistical purposes, and for the purposes of the management of the service and may also pass such information to other parties such as Data Partners.

I This Agreement is made under English law and the parties submit to the exclusive jurisdiction of the English courts. Notwithstanding the preceding sentence, CDRC may at any time bring proceedings for an injunction in any court of competent jurisdiction.

J. By signing this Agreement, the User and Institution represent and undertake that: the information supplied in the Project and any supporting documentation is accurate to the best of their knowledge, and that they have read and understand and shall obey the conditions specified in this Agreement.

[SIGNATURES ON FOLLOWING PAGE]

Signed by:

USER

Signature 

Name (PRINT): JOHANN SCHAIBLE

Date: Nov. 19th, 2019

INSTITUTION:

Signature: 

Name (print): PROF. DR. CHRISTOF WOLF

[Institution Name] GESIS Leibniz Institute for the Social Sciences

Date: Dec. 20th, 2019

UNIVERSITY COLLEGE LONDON

Signature:

Name (print):

[Institution Name]

Date:

Consumer Data Research Centre

An ESRC Data Investment

Schedule 1: Approved Project Proposal

CONSUMER DATA RESEARCH CENTRE

Safeguarded Data
Project Proposal Form

For office use: Ref 297-07

Consumer Data Research Centre

An ESRC Data Investment

INTRODUCTION

This form should be completed by those wishing to access the Consumer Data Research Centre's (CDRC) safeguarded data collections. You should consult the ***CDRC Data Service User Guide*** before completing the form. A CDRC data scientist will be assigned to you who can provide support in the application process.

Once submitted your proposal will be forwarded to the CDRC Research Approvals Group (RAG) for independent assessment. Projects will be assessed based on the criteria listed in Appendix 1.

Approval to access data will not be granted without evidence that the applicant has acquired ethical approval for the research through their institution, or supplied evidence that this is not applicable. For non-academic projects, where there is no approval process in place, the CDRC will assist the applicant in acquiring this.

PART A. PROJECT DETAILS

1. Contact Details

1.1. **Lead applicant:** Tara Morovatdar
Department and Institution: University of Koblenz-Landau
Address: Germany-Koblenz-Robert Koch str 7-56073
Email: tmorovatdar@uni-koblenz.de
Telephone: +4917630174409

1.2. **Initial Proposal Form Reference:** Project 297

1.3. **Title of Project**

A fine grained station usage prediction for the Santander Bike-Sharing System in London

1.4. **Access Details**
Access requested from[1]: 1.09.2019
Access requested to (Date): 1.09.2020

2. Project Proposal Details

2.1. **Abstract**. Appropriate for a general audience. This may be used by the CDRC for reporting and publicity purposes.

This study aims at analyzing the usage of shared bike stations across London. We analyze the capacity of each station (also called "dock"), i.e, the number of used and free bicycle parking racks of a station, at different times during day and night. This allows for predicting

[1] Please note the RAG review process takes approximately 3 weeks.

how many free bike racks will a station have at a specific time point. Typically, customers of a bike-sharing system take a bike at one location to go to some other location in the city. Arriving at the desired location, they will need to return the bike at a near station. However, the nearest station might not have any more free bike racks, such that the customer will need to search for another station with free bike racks. Predicting the number of free bike racks might alleviate the situation and help the customers to go directly to a station with free bike racks. Also, it might be helpful for the bike-sharing provider in better redistributing the bikes.

2.2. **Project description.** A detailed description of the project, documenting the motivation, scope and aims of the intended research as well as the methods you will use in the proposed research.

People all around the world are using more and more bikes from Bike-Sharing Systems (BSS) and this usage is growing on a daily basis. According to Shaheen et al.[2] bike-sharing systems are good for the environment, because of the decreased usage of cars and because it is good for public health as more people are being active using bicycles. As Midgely[3] has mentioned "smart bike-sharing systems provide the missing link between existing points of public transportation and desired destinations". He also mentioned that an increase in mobility choices improves air quality and the reduction of congestion, which are general objectives of bike-sharing systems. Etienne and Latifa[4] introduce walking and biking as a soft mode transport which is a different solution to organize passenger mobility. They point out that BSS -- as a cheap way of transport -- are very suitable to reduce traffic in the cities. They also mention that because bikes do not produce green-house gases and noise, BSS can help the environment. Purnama[5] mentioned that people have more flexibility to choose their origin, route, destination and travel time based on their needs when they are using BSS.

Because of all these features, bike-sharing systems are growing very fast in the world, so it becomes more and more necessary to improve these systems through analysing and understanding their underlying structure.

Bike-sharing systems have some common issues that can be addressed. In station-based bike-sharing systems, an unbalanced bike distribution in the stations (also called docks) is one of the most important issues. At each station the number of bike racks is limited, such that it is not possible to return a bike to a full station or pick up a bike from an empty one. The bikes in a bike sharing system need to be re-balanced frequently because at different stations in different periods of time the check-in and check-out of the bikes are unbalanced. In other words, the system usage is asymmetric and fluctuates during the day. Some other factors like altitude differences, weather conditions or events in the city may cause irregular

[2] Susan Shaheen, Stacey Guzman, and Hua Zhang. Bike-sharing in Europe, the Americas, and Asia: past, present, and future. Transportation Research Record: Journal of the Transportation Research Board, (2143):159–167, 2010.

[3] Peter Midgley. The role of smart bike-sharing systems in urban mobility.Journeys,2(1):23–31, 2009.

[4] Côme Etienne and Oukhellou Latifa. Model-based count series clustering for bike-sharing system usage mining: a case study with the vélib'system of paris.ACM Transactions on Intelligent Systems and Technology (TIST), 5(3):39, 2014.s

[5] Ida Bagus Irawan Purnama. Spatiotemporal analysis of urban mobility dynamics: a case study of bicycle sharing system. 2018.

or asymmetric demands and flow of the bikes. Bike-sharing providers know that the availability of free bikes and bike racks in the stations is one of the most important aspects in BSS, so they have tried different solutions, i.e., redistributing the bikes with trucks or giving incentive to users to keep the travels short(i.e. Usually the first 30 minutes of a ride are often for free and after that incur charges[6] or the fees increase progressively[7]) so the bikes are more available, to solve the unbalanced distribution of the bikes.

The users benefit from the flexibility of bike-sharing systems but it poses a major challenge and recurrent problem for the operators to satisfy the forthcoming demand with the limit in the resources. Because if a user attempts to drop off a bike at a full station or tries to rent a bike from an empty station the user would experience a bad quality of service when trying to use the BSS[8].

Reallocating the bikes takes too much time so it is too late to decide to do that when you detect an imbalance, even if it is you always monitor the stations and detect the imbalance instantly after it occurs.

Liu et al[9] conducted a study on NYC CitiBike to develop a model that helps the bike-sharing system to avoid expensive re-balancing and to get better user experience. They tried to predict station demand which is defined as average pick-up/hour with neural network and also station balance which is the situation if the overall pick-up (or drop-off) for a period of time is greater than a threshold with the use of the neural network. Nair et al.[10] conducted an empirical analysis on Paris bike-sharing system Vélib, regarding the station's location they found that a close coupling of transit(metro stations) and bike station can increase utilization, then they developed a mixed-integer program. The model outputs where and how many bicycles to rearrange. This helps operators to redistribute bikes so that they meet a reliability level. In 2016 Yang, Zidong, et al.[11] proposed a Spatio-temporal bicycle mobility model with the use of historical data from the Chinese city Hangzhou which has the world's largest BSS. For the first time, they predicted check-in and check-outs on a station level.Their goal was to advance the bike-rebalancing algorithm designs. Using their mobility model with a time-varying parameter for the check-in prediction and random forest for check-out prediction they reached outperforming results.

---

[6] Nair, R., Miller-Hooks, E., Hampshire, R. C. and Bušı ́c, A. [2013], 'Large-scale vehicle sharing systems: analysis of vélib",International Journal of Sustainable Transportation7(1), 85–106

[7] ogel, P., Greiser, T. and Mattfeld, D. C. [2011], 'Understanding bike-sharing systems using data mining: Exploring activity patterns',Procedia-Social and Behavioral Sciences20, 514–523.

[8] Adish Singla, Marco Santoni, Gábor Bartók, Pratik Mukerji, Moritz Meenen, and Andreas Krause. Incentivizing users for balancing bike sharing systems. InAAAI, pages723–729, 2015.

[9] Junming Liu, Qiao Li, Meng Qu, Weiwei Chen, Jingyuan Yang, Hui Xiong, Hao Zhong, and Yanjie Fu. Station site optimization in bike sharing systems. In2015 IEEE International Conference on Data Mining pages 883–888. IEEE, 2015

[10] Nair, R., Miller-Hooks, E., Hampshire, R. C., & Bušić, A. (2013). Large-scale vehicle sharing systems: analysis of Vélib'. International Journal of Sustainable Transportation, 7(1), 85-106.

[11] Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J., & Moscibroda, T. (2016, June). Mobility modeling and prediction in bike-sharing systems. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (pp. 165-178). ACM.

In Our research, to alleviate the unbalanced situation, we propose to predict the bike usage of each station, so planning ahead will be easier, one example is that one can use our model results to reduce the costs of redistribution. Two different dependencies affect the imbalance in bike distribution: a temporal one, which means at different times of the day and week the use of the bikes is different. The other dependency is spatial due to short one-way trips. As mentioned above some bike-sharing providers are using trucks to move bikes from full stations to empty ones and some promote users to rent bikes for short times so the bikes are available for everyone[12]. For example, Santander uses both ways to make sure that bikes are available for more people. Santander gives users incentive to keep the trips short by the policy they have about the trips which says the first 30 minutes are free of charge and any minute above that (up to 30 minutes) it costs £2. They also use trucks to relocate the bikes so the bikes are available everywhere.

In this work, we aim to design a model that predicts the capacity of a specific station at a given time. In order to develop a model capable of predicting the odds of a station having an empty dock in a time window of 30 minutes, (non)linear regression is conducted. When a customer hires a bike, the Santander app can recommend the 6 closest stations to the user destination sorted by the probability of having an available empty dock during the next 30 minutes.

Using the Docking Station Locations data we tend to observe how much the number of bikes and stations has been increased during the years. Using regression analysis can help us to understand the pattern that the Bike-sharing system has been changed during the years so the system administrators can plan for the future better.

2.3. Research Category:

2.3.1.Is this request for an Undergraduate, Masters project? U'grad ☐ Masters ☒

2.3.2.Is this request for a PhD project? PhD ☐

2.3.3. Is your project funded, commissioned or sponsored by a funding body or any other organisation? Yes ☐ No ☒ Funding application in progress ☐

Please include the name, postal and web address of your current or prospective funder, and your grant/project reference number (if applicable).
Click here to enter text.

2.4. Project Impact. Please describe the anticipated scientific and societal benefits of the project and the ways in which you intend to maximise those benefits.

This master thesis will help researchers to better understand human mobility in the context of bike-sharing systems, to identify the underlying patterns. It also can help the authorities to

[12] Patrick Vogel, Torsten Greiser, and Dirk Christian Mattfeld. Understanding bike-sharing systems using data mining: Exploring activity patterns. Procedia-Social and Behavioural Sciences, 20:514–523, 2011.

plan for the future better. Nevertheless, the aim of this research is to provide a recommendation system which gives the users a bike station recommendation list based on their destination so it gives a better feeling to users if they don't face a situation that the station does not have an empty dock.

2.5. End Users. Who are the main end users of this research (academic research, central government, consultancy, industry, local government, NHS, public sector, third sector? List all that are applicable.)

Academic research is the main end user for this work. Naturally, the results can also be used by others, such as the local government (e.g., improve bike lane situations in specific areas) or industry (e.g. better plan redistribution plans for bikes).

2.6. Outputs and Publications. What are the intended outputs or publications arising from the use of these data? (For example, journal articles, PhD thesis, report for government department, policy documents for a local authority, White Papers, new software or other tools, etc.)

This data is used for a Master thesis. We plan to publish a comprised version of it as an article in a journal or a conference

2.7. Research Team

Please list the names, affiliation and email addresses of all known members of your research team who are requesting access to the CDRC safeguarded data. If you are a student please include your academic supervisor.

To access some of the CDRC safeguarded data products it is necessary for users to have successfully completed a safe researcher training course. Please see the terms of use information for the particular dataset.

Consumer
Data
Research
Centre

An ESRC Data
Investment

For office use: Ref 297-07

Consumer Data Research Centre

An ESRC Data Investment

Research Team *(including lead applicant, add more rows if required)*

Title, Name	**Department/ Institution**	**Institutional email address**	**Completed a safe researcher training course. If yes, please specify course and date of completion.**
JProf.Dr.Claudia Wagner	Assistant professor in Computer Science at University Koblenz-Landau and the interim Scientific Director of the department Computational Social Science at GESIS - Leibniz Institute for the Social Sciences.	clwagner@uni-koblenz.de	
Dr.Johann Schaible	Postdoctoral researcher and Team Lead of team "Knowledge Discovery" of the Computational Social Science department at GESIS - Leibniz Institute for the Social Sciences	johann.schaible@gesis.org	

For office use: Ref 297-07

Consumer Data Research Centre

An ESRC Data Investment

For office use: Ref 297-07

2.8. Ethical Approval

2.8.1. Have you sought or are you seeking **ethical approval from an institutional ethical approvals panel** or any other appropriate body?

Yes ☐ No ☒

2.8.2. If *Yes*, please provide **evidence** of the status of the application or the outcome of the ruling issued. Please list what evidence you are enclosing below and return it as a separate attachment in PDF format when you return this application form. Feel free to add any comments below.

Click here to enter text.

2.8.3. If *No*, please bring this to the attention of the CDRC as soon as possible so that routes for ethical approval may be discussed.

Click here to enter text.

2.8.4. If you believe that ethical approval is not required for your research, please provide justification for this below.

There is no data concerning user information. Our research is simply using information about bike stations and bike-sharing patterns. That is why we believe no ethical approval is required.

Consumer Data Research Centre

An ESRC Data Investment

PART B. DATA REQUEST

3.1. Data Required. Please provide the following information for each dataset requested. Please add more lines if required.

Data Partner	Data Set	Access to Full Data Set requested or specific variables (list)	Geographic Extent	Temporal Extent
	London Bicycle Sharing System Docking Station Individual Observations	All Observations - all the columns	*London*	*2010-present*
	London Bicycle Sharing System Docking Station Locations	All the columns	*London*	*2010-present*

For office use: Ref 297-07

4. Data Linkage

4.1. Data Linkage. If your project will be linking more than one data source, describe which data sources will be linked and how the linkage will be done, including any specific variables that need to be linked (if known). If any of the data to be linked has identifying information as defined in the Data Protection Act 1998 or General Data Protection Regulations please provide details. Please note that no project that has the potential to re-identify individuals through data linkage will be approved.

In this work, we want to use Santander Cycle Hire journeys dataset which is provided by Transport for London (TfL) and is an open dataset. As I have asked Oliver O'Brien, the station id in Santander Cycle Hire journeys is the same as ucl_id in Bike location dataset and operator_id in London Bicycle Sharing System Docking Station Individual Observations. So I can use these columns to link the datasets.

PART C: DECLARATION

By completing this declaration I hereby declare that the information included in this application form is true and correct to the best of my knowledge. I understand that any false or misleading information given by me in connection with my application may result in termination of the application process and/or other sanctions.

I also agree that I will be the single point of contact for progress updates and communication regarding the progress of the application.

I agree for my personal information to be used for the purposes of processing this application in accordance with the relevant data laws of the UK.

☑ I consent to my contact details being added to the CDRC contacts database so that the CDRC can send me notifications of CDRC related activities.

☑I understand that forwarding this form by email constitutes an electronic signature.

☑ I understand that final approval for this project may require the additional submission of project approval forms.

Name: Tara Morovatdar

Date: 20/08/2019

APPENDIX 1: RAG CRITERIA FOR ASSESSMENT

The role of the RAG is to provide independent and transparent assessments of applications by researchers for access to data through both the CDRC Safeguarded and CDRC Secure services based on a set of standard evaluation criteria. RAG is independent to the CDRC and will include representation from the academic, big data, industrial sectors as well as the data partners concerned. For full Terms of Reference and membership see https://www.cdrc.ac.uk/data-services/using-our-data/

Criteria for Approval

- **Scientific advancement –** how the project has the potential to advance scientific knowledge, understanding and/or methods using consumer data;
- **Public good** – how the project has the potential to provide insight and/or solutions that could benefit society;
- **Privacy and ethics –** the potential privacy impacts or risks, and wider ethical considerations relating to the project
- **Project Design and Methods –** how the project will be conducted and who will be involved with a focus on demonstrating project feasibility.
- **Cost and resources issues –** what impact the project is likely to have on CDRC resources, including CDRC staff time and use of infrastructure, as well as any data acquisition costs. Resource requirements should be justified.

Schedule 2: Approved Project Notification

Approved Project Notification

Project Reference Number: 297-01
Project Title: A fine grained station usage prediction for the Santander Bike-Sharing System in London
CDRC User(s): Tara Morovatdar, Prof.Dr.Claudia Wagner, Dr.Johann Schaible
Institution: University of Koblenz-Landau
Proposal Accepted: 23/09/2019

Stipulations imposed by CDRC, including those from RAG members and Data Partner(s)

1. CDRC

1.1. The CDRC User(s) must only use the data for the purpose of research as described in the approved Project Proposal 297-01.
1.2. Approval for use of the data for this purpose is from the Start Date 23/09/2019 until the End Date 01/09/2020. If the CDRC User(s) requires access to this data beyond the End Date a request for an extension may be submitted to the CDRC with an explanation on why this is required.
1.3. By the end date the data and all copies of the data must be deleted. Notification must be sent to the CDRC by email to info@cdrc.ac.uk within 7 days of this date to confirm that the deletion has been completed.
1.4. There is no requirement to complete safe researcher training to access safeguarded Bicycle Sharing System Docking Stations data.
1.5. The academic supervisors of student CDRC Users are required to ensure that their students adhere to all the stated stipulations.
1.6. There is no requirement when accessing this dataset to submit for review a copy of any proposed publication (including reports, abstracts, or presentation to a journal, editor, meeting, seminar or other third party) to the CDRC prior to submission for publication or presentation.
1.7. Users are required to deposit copies of working papers, peer-reviewed journal articles, logs of impact and other publications for access with the CDRC site wherever copyright permits. Where a proposed publication has been produced using more than one dataset, the User should adhere to the partner stipulation which has the longest review period. Where this is not possible, full references to research outputs are required for CDRC audit purposes. Please email publications@cdrc.ac.uk when publications are ready for deposit or logging.
1.8. Published outputs must include an acknowledgement stating "The data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC 297-01, ES/L011840/1; ES/L011891/1".
1.9. Users are requested to provide a brief case study of the supported work to be included in the Research Outputs section of the CDRC website https://www.cdrc.ac.uk/research/

On behalf of CDRC: Nick Bearman

N E Bearman

Date: 23/09/2019

CONSUMER DATA RESEARCH CENTRE

USERS AGREEMENT

THIS AGREEMENT IS MADE BETWEEN

Prof.Dr.Claudia Wagner, at **University Koblenz-Landau** ("User");

University of Koblenz-Landau, Germany-Koblenz-Robert Koch str 7-56073, ("Institution"); and

University College London, represented by the Consumer Data Research Centre of Gower Street, London, WC1E 6BT ("CDRC").

INTRODUCTION

The Consumer Data Research Centre is an ESRC Big Data initiative, bringing together world-class researchers at University College London, and the Universities of Leeds, Liverpool and Oxford to harness the potential of consumer-related data to benefit social science researchers, businesses, government and society at large.

The CDRC requires users wishing to access safeguarded and controlled data services for research purposes to submit project proposals for approval by the Research Approvals Group (as defined below).

Once a User is notified that their Project has been approved by receiving an Approved Project Notification, prior to being issued download instructions and password or site specific instructions for the assigned Secure Lab, the User and their Institution must sign this CDRC User Agreement, including all Schedules hereto (the "Agreement"). The Agreement demonstrates that the User and their Institution understand the seriousness of the undertaking, and that the User and Institution understand the penalties that may be imposed for breaches of security or confidentiality, or any other breach of this Agreement.

DEFINITIONS

"Approved Project Notification" means the notification from the CDRC to the User that their application to CDRC to carry out the Project has been approved, and shall set out any additional Stipulations that CDRC and/or a Data Partner may require the User and their Institution to comply with under this Agreement, and which is annexed to this Agreement at Schedule 2, and may be updated from time to time by written notice from CDRC to the User.

"Confidential Information" means information belonging to a Data Partner that is identified by the Data Partner as confidential in the Approved Project Notification and may include CDRC Data or Controlled Data.

"CDRC Data" means any combination of data which may be either Controlled or Safeguarded Data, and which may be subject to Stipulations.

"Controlled Data" means data which must be held under the most secure conditions with stringent access restrictions, including data which is Personal Information, and/or data that is considered commercially sensitive, and may also be subject to Stipulations.

"Data Partner" the organisation with whom the CDRC has entered into an agreement to provide Safeguarded and/or Controlled Data to the CDRC.

"**DPA**" means the Data Protection Act 1998 and any subsequent legislation or regulations.

"**End Date**" means the end date of the Project as set out in the Approved Project Notification.

"Institutional Signatory" means a signatory who is validly authorised to enter into contracts on behalf of the Institution.

"**Output"** means a pre-approved Project output as described in the Project.

"**Output Request and Release Process**" means the multi-stage Output approval whereby CDRC data scientists may check that Outputs to be released from the Secure Lab are compliant with the Approved Project Notification and are adequately non-disclosive.

"Personal Information" means personal information as defined by the DPA.

"Project" The particular research project proposal approved by the CDRC Research Approvals Group and approved in the Approved Project Notification, and which is annexed to this Agreement as Schedule 1.

"Research Approvals Group (RAG)" means the independent research approvals group whose role is to approve proposed project applications to use CDRC Data.

"Safeguarded Data" means data to which access is restricted due to licence conditions, but where data are not considered "personally-identifiable" or otherwise sensitive.

"Secure Lab" means the secure laboratory facilities within the CDRC, which comprises three secure laboratory facilities at UCL, the University of Leeds and the University of Liverpool. Each lab has its own site specific terms and conditions which Users and their Institutions may also be required to sign in order to obtain access to CDRC Data through each respective Secure Lab location.

"**Start Date**" means the start date of the Project as set out in the Approved Project Notification.

"**Stipulations**" means Project specific terms placed by CDRC and/or a Data Partner, as notified to the User and their Institution under the Approved Project Notification with which the User and Institution must comply in their use of the CDRC Data under this Agreement.

IT IS HEREBY AGREED:

1. Use of Data

A. If required within the Approved Project Notification, the User must, and their Institutions must procure that the User has satisfactorily completed a CDRC approved accredited safe researcher training course within the specified time period.

B. The User shall be responsible for processing all CDRC Data in accordance with all applicable laws and all regulatory standards applicable to such CDRC Data.

C. The User will ensure that any Stipulations as specified in the Approved Project Notification are adhered to.

D. Access to the CDRC Data is being provided for the statistical analysis and research as detailed in the Project and as further outlined in the Approved Project Notification. The CDRC Data shall not be used or processed for any other purpose without the prior written consent of the CDRC and the Data Partner(s).

E. The User shall not disclose nor compromise any of the CDRC Data from the individual records obtained or produced from the CDRC Data to anyone other than (i) those approved for the same Project and (ii) CDRC staff involved in the review of the Outputs.

F. The User shall ensure that no attempts are made to link the CDRC Data to any other files in order to relate the particulars to any identifiable individual person, business or organisation unless such data linkage exercise has been explicitly approved by the CDRC as part of the Project.

G. On termination of the Agreement for whatever reason, all access to the CDRC Data related to the Project shall cease forthwith, and electronic access be denied. The User must permanently destroy/delete or erase the Safeguarded Data that is in their possession within seven (7) days of the end of this Agreement, together with all hard or soft copies of the same and certify such destruction in writing (which may include email) to the CDRC.

H. The CDRC reserves the right to monitor, record, and audit, or to request a written report from the User regarding, the use and activities relating to the use, of the CDRC Data by the User during the lifetime of this Agreement.

I. During the lifetime of this Agreement and six months thereafter, the User agrees that, and the Institution shall procure that, the CDRC will have right of entry to any premises where the CDRC Data is accessed in order to monitor, record or audit the use and activities relating to the use of the CDRC Data, upon the provision of reasonable notice to the User.

J. Any incidents of unauthorised access to, processing of, or disclosing of, the CDRC Data must be reported immediately to the CDRC.

K. The Agreement is subject to review and without limitation whenever a change in the law, contracts for services with third parties, other procedures or other relevant circumstances takes place.

2. Output Release for CDRC Data

A. The User must not reproduce, to any extent whatsoever, any CDRC Data, original datasets or copies or subsets of any CDRC Data.

B. Users are responsible for, and their Institution must procure that the User ensures that rules and regulations for disclosure risk analysis as specified by statistical disclosure guidelines (explained during accredited safe researcher training) are applied prior to submission of analytical outputs for clearance and release.

C. No Outputs may be removed from the CDRC Secure Lab by the User or before the User satisfactorily completes the follow the Output Request and Release process.

D. The CDRC reserves the right to release in whole or in part, an amended version or not to release at all, as the CDRC deems appropriate, the proposed Output produced by the User pursuant to this Agreement.

E. The User and the Institution agrees to work with the CDRC to meet the requirements of the Output Request and Release process. In the event that the CDRC decides not to release the proposed Output, if feasible the User may, at CDRC's sole discretion, be allowed to revisit the Secure Lab to rectify the problem. The final decision to release an Output rests with the CDRC.

F. No CDRC Data shall be used for any commercial purpose whatsoever.

3. CDRC Data, Copyright and Publication Protocol

A. The CDRC Data and related documentation shall at all times be and remain the sole and exclusive property of the CDRC and/or the Data Partner, as the case may be. This Agreement pertains to the use of the CDRC Data and related documentation to produce a

pre-approved Output for research purposes and nothing contained herein shall be deemed to convey any title or ownership interest in the CDRC Data or the related documentation to the User.

B. Copyright in Outputs may be held singly or jointly by the User(s) that created them, their Institution(s) or their funder(s) according to the User's funding and Institution's agreements.

C. Any publications in connection with this Project (including but not limited to draft versions, final versions, abstracts, presentations, meetings or seminars) may only be published upon receiving written approval from CDRC. The procedure to obtain approval from CDRC shall be that the User shall submit any proposed publication to CDRC via publications@cdrc.ac.uk strictly in accordance with the timescales specified in the Approved Project Notification. If CDRC and/or the Data Partner, at their sole discretion, determine that the proposed publication contains any Confidential Information, it shall notify the User to ensure that such Confidential Information is deleted. Following any such deletion, CDRC shall notify the User in writing that they may proceed with the proposed publication. CDRC and/or the Data Partner may also request that the User delay the publication if the CDRC and/or Data Partner reasonably believes that the CDRC Data has been used outside the scope of the Project or the Approved Project Notification and the User will make reasonable attempts to amend the publication to the satisfaction of the CDRC and/or Data Provider.

D. The User shall comply with any further Stipulations in connection with publications as may be specified in the Approved Project Notification.

4. General

A. This Agreement shall commence on the Start Date and shall end on the End Date, subject to any earlier termination in accordance with clause 4.B or clause 4.C. Where the Agreement is due to terminate under this clause 4.A and provided the CDRC is free and able to do so, the parties may agree in writing for the continuing use of the CDRC Data in relation to the Project for a specific agreed additional period of time. The terms of this Agreement shall continue to apply to any such continuing use.

B. Without prejudice to any other right or remedy, if the User and/or the Institution shall commit any breach of, or fail to comply with, any of their obligations under this Agreement, become bankrupt or any judgment is made against either the User or the Institution and remains unsatisfied for seven (7) days, CDRC shall be entitled to terminate this Agreement forthwith on written notice.

C. The CDRC reserves the right to terminate this Agreement with immediate effect on written notice in the event that the Data Partner terminates the agreement to provide CDRC Data to the CDRC.

D. The User and Institution acknowledge that any disclosure of CRDC Data other than in accordance with this Agreement may cause irreparable harm to the CDRC and/or a Data Partner. The User and Institution hereby agree that CDRC shall, in accordance to any other rights or remedies to which it may be entitled, have the right to seek and obtain equitable and or injunctive relief.

E. CDRC does not warrant the accuracy or completeness of the CDRC Data, nor its fitness for any particular purpose. The CDRC Data is provided to User "as is" and CDRC does not warrant that the CDRC Data will enable the User to achieve the objectives of the Project, or that the CDRC Data is suitable for use in any report or for academic or research purposes.

F. The User and/or Institution as the case may be are required to bring directly to the attention of CDRC any matters or events that may affect any of the obligations under this Agreement immediately as awareness of the matters or events may arise.

G. No failure or delay on the part of CDRC to exercise any right or remedy under this Agreement shall be construed or operate as a waiver thereof, nor shall any single or partial exercise of any right or remedy preclude the further exercise of such right or remedy.

H. The User and the Institution acknowledge that the CDRC may hold and process information submitted in the Project for validation and statistical purposes, and for the purposes of the management of the service and may also pass such information to other parties such as Data Partners.

I This Agreement is made under English law and the parties submit to the exclusive jurisdiction of the English courts. Notwithstanding the preceding sentence, CDRC may at any time bring proceedings for an injunction in any court of competent jurisdiction.

J. By signing this Agreement, the User and Institution represent and undertake that: the information supplied in the Project and any supporting documentation is accurate to the best of their knowledge, and that they have read and understand and shall obey the conditions specified in this Agreement.

[SIGNATURES ON FOLLOWING PAGE]

Signed by:

USER

Signature

Name (PRINT): CLAUDIA WAGNER

Date: Nov. 21st, 2019

INSTITUTION:

Signature:

Name (print): Prof. Dr. Maria A. Wimmer

[Institution Name] University of Koblenz-Landau, Dean of Faculty of computer science

Date: 2nd Dec, 2019

Dekanin
Fachbereich Informatik

UNIVERSITY COLLEGE LONDON

Signature:

Name (print):

[Institution Name]

Date:

Consumer Data Research Centre

An ESRC Data Investment

Schedule 1: Approved Project Proposal

CONSUMER DATA RESEARCH CENTRE

Safeguarded Data
Project Proposal Form

For office use: Ref 297-07

Consumer Data Research Centre

An ESRC Data Investment

INTRODUCTION

This form should be completed by those wishing to access the Consumer Data Research Centre's (CDRC) safeguarded data collections. You should consult the ***CDRC Data Service User Guide*** before completing the form. A CDRC data scientist will be assigned to you who can provide support in the application process.

Once submitted your proposal will be forwarded to the CDRC Research Approvals Group (RAG) for independent assessment. Projects will be assessed based on the criteria listed in Appendix 1.

Approval to access data will not be granted without evidence that the applicant has acquired ethical approval for the research through their institution, or supplied evidence that this is not applicable. For non-academic projects, where there is no approval process in place, the CDRC will assist the applicant in acquiring this.

PART A. PROJECT DETAILS

1. Contact Details

1.1. Lead applicant: Tara Morovatdar
Department and Institution: University of Koblenz-Landau
Address: Germany-Koblenz-Robert Koch str 7-56073
Email: tmorovatdar@uni-koblenz.de
Telephone: +4917630174409

1.2. Initial Proposal Form Reference: Project 297

1.3. Title of Project

A fine grained station usage prediction for the Santander Bike-Sharing System in London

1.4. Access Details

Access requested from[1]: 1.09.2019
Access requested to (Date): 1.09.2020

2. Project Proposal Details

2.1. Abstract. Appropriate for a general audience. This may be used by the CDRC for reporting and publicity purposes.

This study aims at analyzing the usage of shared bike stations across London. We analyze the capacity of each station (also called "dock"), i.e, the number of used and free bicycle parking racks of a station, at different times during day and night. This allows for predicting

[1] Please note the RAG review process takes approximately 3 weeks.

how many free bike racks will a station have at a specific time point. Typically, customers of a bike-sharing system take a bike at one location to go to some other location in the city. Arriving at the desired location, they will need to return the bike at a near station. However, the nearest station might not have any more free bike racks, such that the customer will need to search for another station with free bike racks. Predicting the number of free bike racks might alleviate the situation and help the customers to go directly to a station with free bike racks. Also, it might be helpful for the bike-sharing provider in better redistributing the bikes.

2.2. **Project description.** A detailed description of the project, documenting the motivation, scope and aims of the intended research as well as the methods you will use in the proposed research.

People all around the world are using more and more bikes from Bike-Sharing Systems (BSS) and this usage is growing on a daily basis. According to Shaheen et al.[2] bike-sharing systems are good for the environment, because of the decreased usage of cars and because it is good for public health as more people are being active using bicycles. As Midgely[3] has mentioned "smart bike-sharing systems provide the missing link between existing points of public transportation and desired destinations". He also mentioned that an increase in mobility choices improves air quality and the reduction of congestion, which are general objectives of bike-sharing systems. Etienne and Latifa[4] introduce walking and biking as a soft mode transport which is a different solution to organize passenger mobility. They point out that BSS -- as a cheap way of transport -- are very suitable to reduce traffic in the cities. They also mention that because bikes do not produce green-house gases and noise, BSS can help the environment. Purnama[5] mentioned that people have more flexibility to choose their origin, route, destination and travel time based on their needs when they are using BSS.

Because of all these features, bike-sharing systems are growing very fast in the world, so it becomes more and more necessary to improve these systems through analysing and understanding their underlying structure.

Bike-sharing systems have some common issues that can be addressed. In station-based bike-sharing systems, an unbalanced bike distribution in the stations (also called docks) is one of the most important issues. At each station the number of bike racks is limited, such that it is not possible to return a bike to a full station or pick up a bike from an empty one. The bikes in a bike sharing system need to be re-balanced frequently because at different stations in different periods of time the check-in and check-out of the bikes are unbalanced. In other words, the system usage is asymmetric and fluctuates during the day. Some other factors like altitude differences, weather conditions or events in the city may cause irregular

[2] Susan Shaheen, Stacey Guzman, and Hua Zhang. Bike-sharing in Europe, the Americas, and Asia: past, present, and future. Transportation Research Record: Journal of the Transportation Research Board, (2143):159–167, 2010.

[3] Peter Midgley. The role of smart bike-sharing systems in urban mobility.Journeys,2(1):23–31, 2009.

[4] Côme Etienne and Oukhellou Latifa. Model-based count series clustering for bike-sharing system usage mining: a case study with the vélib'system of paris.ACM Transactions on Intelligent Systems and Technology (TIST), 5(3):39, 2014.s

[5] Ida Bagus Irawan Purnama. Spatiotemporal analysis of urban mobility dynamics: a case study of bicycle sharing system. 2018.

or asymmetric demands and flow of the bikes. Bike-sharing providers know that the availability of free bikes and bike racks in the stations is one of the most important aspects in BSS, so they have tried different solutions, i.e., redistributing the bikes with trucks or giving incentive to users to keep the travels short(i.e. Usually the first 30 minutes of a ride are often for free and after that incur charges[6] or the fees increase progressively[7]) so the bikes are more available, to solve the unbalanced distribution of the bikes.

The users benefit from the flexibility of bike-sharing systems but it poses a major challenge and recurrent problem for the operators to satisfy the forthcoming demand with the limit in the resources. Because if a user attempts to drop off a bike at a full station or tries to rent a bike from an empty station the user would experience a bad quality of service when trying to use the BSS[8].

Reallocating the bikes takes too much time so it is too late to decide to do that when you detect an imbalance, even if it is you always monitor the stations and detect the imbalance instantly after it occurs.

Liu et al[9] conducted a study on NYC CitiBike to develop a model that helps the bike-sharing system to avoid expensive re-balancing and to get better user experience. They tried to predict station demand which is defined as average pick-up/hour with neural network and also station balance which is the situation if the overall pick-up (or drop-off) for a period of time is greater than a threshold with the use of the neural network. Nair et al.[10] conducted an empirical analysis on Paris bike-sharing system Vélib, regarding the station's location they found that a close coupling of transit(metro stations) and bike station can increase utilization, then they developed a mixed-integer program. The model outputs where and how many bicycles to rearrange. This helps operators to redistribute bikes so that they meet a reliability level. In 2016 Yang, Zidong, et al.[11] proposed a Spatio-temporal bicycle mobility model with the use of historical data from the Chinese city Hangzhou which has the world's largest BSS. For the first time, they predicted check-in and check-outs on a station level.Their goal was to advance the bike-rebalancing algorithm designs. Using their mobility model with a time-varying parameter for the check-in prediction and random forest for check-out prediction they reached outperforming results.

---

[6] Nair, R., Miller-Hooks, E., Hampshire, R. C. and Bušĭ c, A. [2013], 'Large-scale vehicle sharing systems: analysis of vélib",International Journal of Sustainable Transportation7(1), 85–106

[7] ogel, P., Greiser, T. and Mattfeld, D. C. [2011], 'Understanding bike-sharing systems using data mining: Exploring activity patterns',Procedia-Social and Behavioral Sciences20, 514–523.

[8] Adish Singla, Marco Santoni, Gábor Bartók, Pratik Mukerji, Moritz Meenen, and Andreas Krause. Incentivizing users for balancing bike sharing systems. InAAAI, pages723–729, 2015.

[9] Junming Liu, Qiao Li, Meng Qu, Weiwei Chen, Jingyuan Yang, Hui Xiong, Hao Zhong, and Yanjie Fu. Station site optimization in bike sharing systems. In2015 IEEE International Conference on Data Mining pages 883–888. IEEE, 2015

[10] Nair, R., Miller-Hooks, E., Hampshire, R. C., & Bušić, A. (2013). Large-scale vehicle sharing systems: analysis of Vélib'. International Journal of Sustainable Transportation, 7(1), 85-106.

[11] Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J., & Moscibroda, T. (2016, June). Mobility modeling and prediction in bike-sharing systems. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (pp. 165-178). ACM.

In Our research, to alleviate the unbalanced situation, we propose to predict the bike usage of each station, so planning ahead will be easier, one example is that one can use our model results to reduce the costs of redistribution. Two different dependencies affect the imbalance in bike distribution: a temporal one, which means at different times of the day and week the use of the bikes is different. The other dependency is spatial due to short one-way trips. As mentioned above some bike-sharing providers are using trucks to move bikes from full stations to empty ones and some promote users to rent bikes for short times so the bikes are available for everyone[12]. For example, Santander uses both ways to make sure that bikes are available for more people. Santander gives users incentive to keep the trips short by the policy they have about the trips which says the first 30 minutes are free of charge and any minute above that (up to 30 minutes) it costs £2. They also use trucks to relocate the bikes so the bikes are available everywhere.

In this work, we aim to design a model that predicts the capacity of a specific station at a given time. In order to develop a model capable of predicting the odds of a station having an empty dock in a time window of 30 minutes, (non)linear regression is conducted. When a customer hires a bike, the Santander app can recommend the 6 closest stations to the user destination sorted by the probability of having an available empty dock during the next 30 minutes.

Using the Docking Station Locations data we tend to observe how much the number of bikes and stations has been increased during the years. Using regression analysis can help us to understand the pattern that the Bike-sharing system has been changed during the years so the system administrators can plan for the future better.

2.3. Research Category:

2.3.1.Is this request for an Undergraduate, Masters project? U'grad ☐ Masters ☒

2.3.2.Is this request for a PhD project? PhD ☐

2.3.3. Is your project funded, commissioned or sponsored by a funding body or any other organisation? Yes ☐ No ☒ Funding application in progress ☐

Please include the name, postal and web address of your current or prospective funder, and your grant/project reference number (if applicable).
Click here to enter text.

2.4. Project Impact. Please describe the anticipated scientific and societal benefits of the project and the ways in which you intend to maximise those benefits.

This master thesis will help researchers to better understand human mobility in the context of bike-sharing systems, to identify the underlying patterns. It also can help the authorities to

[12] Patrick Vogel, Torsten Greiser, and Dirk Christian Mattfeld. Understanding bike-sharing systems using data mining: Exploring activity patterns. Procedia-Social and Behavioural Sciences, 20:514–523, 2011.

plan for the future better. Nevertheless, the aim of this research is to provide a recommendation system which gives the users a bike station recommendation list based on their destination so it gives a better feeling to users if they don't face a situation that the station does not have an empty dock.

2.5. **End Users.** Who are the main end users of this research (academic research, central government, consultancy, industry, local government, NHS, public sector, third sector? List all that are applicable.)

Academic research is the main end user for this work. Naturally, the results can also be used by others, such as the local government (e.g., improve bike lane situations in specific areas) or industry (e.g. better plan redistribution plans for bikes).

2.6. **Outputs and Publications.** What are the intended outputs or publications arising from the use of these data? (For example, journal articles, PhD thesis, report for government department, policy documents for a local authority, White Papers, new software or other tools, etc.)

This data is used for a Master thesis. We plan to publish a comprised version of it as an article in a journal or a conference

2.7. **Research Team**

Please list the names, affiliation and email addresses of all known members of your research team who are requesting access to the CDRC safeguarded data. If you are a student please include your academic supervisor.

To access some of the CDRC safeguarded data products it is necessary for users to have successfully completed a safe researcher training course. Please see the terms of use information for the particular dataset.

Consumer
Data
Research
Centre

An ESRC Data
Investment

For office use: Ref 297-07

Consumer Data Research Centre

An ESRC Data Investment

Research Team *(including lead applicant, add more rows if required)*

Title, Name	**Department/ Institution**	**Institutional email address**	**Comp traini speci comp**
JProf.Dr.Claudia Wagner	Assistant professor in Computer Science at University Koblenz-Landau and the interim Scientific Director of the department Computational Social Science at GESIS - Leibniz Institute for the Social Sciences.	clwagner@uni-koblenz.de	
Dr.Johann Schaible	Postdoctoral researcher and Team Lead of team "Knowledge Discovery" of the Computational Social Science department at GESIS - Leibniz Institute for the Social Sciences	johann.schaible@gesis.org	

2.8. Ethical Approval

2.8.1. Have you sought or are you seeking **ethical approval from an institutional ethical approvals panel** or any other appropriate body?

Yes ☐ No ☒

2.8.2. If *Yes*, please provide **evidence** of the status of the application or the outcome of the ruling issued. Please list what evidence you are enclosing below and return it as a separate attachment in PDF format when you return this application form. Feel free to add any comments below.

Click here to enter text.

2.8.3. If *No*, please bring this to the attention of the CDRC as soon as possible so that routes for ethical approval may be discussed.

Click here to enter text.

2.8.4. If you believe that ethical approval is not required for your research, please provide justification for this below.

There is no data concerning user information. Our research is simply using information about bike stations and bike-sharing patterns. That is why we believe no ethical approval is required.

Consumer Data Research Centre

An ESRC Data Investment

PART B. DATA REQUEST

3.1. Data Required. Please provide the following information for each dataset requested. Please add more lines if required.

Data Partner	Data Set	Access to Full Data Set requested or specific variables (list)	Geographic Extent
	London Bicycle Sharing System Docking Station Individual Observations	All Observations - all the columns	*London*
	London Bicycle Sharing System Docking Station Locations	All the columns	*London*

4. Data Linkage

4.1. Data Linkage. If your project will be linking more than one data source, describe which data sources will be linked and how the linkage will be done, including any specific variables that need to be linked (if known). If any of the data to be linked has identifying information as defined in the Data Protection Act 1998 or General Data Protection Regulations please provide details. Please note that no project that has the potential to re-identify individuals through data linkage will be approved.

In this work, we want to use Santander Cycle Hire journeys dataset which is provided by Transport for London (TfL) and is an open dataset. As I have asked Oliver O'Brien, the station id in Santander Cycle Hire journeys is the same as ucl_id in Bike location dataset and operator_id in London Bicycle Sharing System Docking Station Individual Observations. So I can use these columns to link the datasets.

PART C: DECLARATION

By completing this declaration I hereby declare that the information included in this application form is true and correct to the best of my knowledge. I understand that any false or misleading information given by me in connection with my application may result in termination of the application process and/or other sanctions.

I also agree that I will be the single point of contact for progress updates and communication regarding the progress of the application.

I agree for my personal information to be used for the purposes of processing this application in accordance with the relevant data laws of the UK.

☑ I consent to my contact details being added to the CDRC contacts database so that the CDRC can send me notifications of CDRC related activities.

☑I understand that forwarding this form by email constitutes an electronic signature.

☑ I understand that final approval for this project may require the additional submission of project approval forms.

Name: Tara Morovatdar

Date: 20/08/2019

APPENDIX 1: RAG CRITERIA FOR ASSESSMENT

The role of the RAG is to provide independent and transparent assessments of applications by researchers for access to data through both the CDRC Safeguarded and CDRC Secure services based on a set of standard evaluation criteria. RAG is independent to the CDRC and will include representation from the academic, big data, industrial sectors as well as the data partners concerned. For full Terms of Reference and membership see https://www.cdrc.ac.uk/data-services/using-our-data/

Criteria for Approval

- **Scientific advancement –** how the project has the potential to advance scientific knowledge, understanding and/or methods using consumer data;
- **Public good** – how the project has the potential to provide insight and/or solutions that could benefit society;
- **Privacy and ethics –** the potential privacy impacts or risks, and wider ethical considerations relating to the project
- **Project Design and Methods –** how the project will be conducted and who will be involved with a focus on demonstrating project feasibility.
- **Cost and resources issues –** what impact the project is likely to have on CDRC resources, including CDRC staff time and use of infrastructure, as well as any data acquisition costs. Resource requirements should be justified.

Schedule 2: Approved Project Notification

Approved Project Notification

Project Reference Number: 297-01
Project Title: A fine grained station usage prediction for the Santander Bike-Sharing System in London
CDRC User(s): Tara Morovatdar, Prof.Dr.Claudia Wagner, Dr.Johann Schaible
Institution: University of Koblenz-Landau
Proposal Accepted: 23/09/2019

Stipulations imposed by CDRC, including those from RAG members and Data Partner(s)

1. CDRC

1.1. The CDRC User(s) must only use the data for the purpose of research as described in the approved Project Proposal 297-01.
1.2. Approval for use of the data for this purpose is from the Start Date 23/09/2019 until the End Date 01/09/2020. If the CDRC User(s) requires access to this data beyond the End Date a request for an extension may be submitted to the CDRC with an explanation on why this is required.
1.3. By the end date the data and all copies of the data must be deleted. Notification must be sent to the CDRC by email to info@cdrc.ac.uk within 7 days of this date to confirm that the deletion has been completed.
1.4. There is no requirement to complete safe researcher training to access safeguarded Bicycle Sharing System Docking Stations data.
1.5. The academic supervisors of student CDRC Users are required to ensure that their students adhere to all the stated stipulations.
1.6. There is no requirement when accessing this dataset to submit for review a copy of any proposed publication (including reports, abstracts, or presentation to a journal, editor, meeting, seminar or other third party) to the CDRC prior to submission for publication or presentation.
1.7. Users are required to deposit copies of working papers, peer-reviewed journal articles, logs of impact and other publications for access with the CDRC site wherever copyright permits. Where a proposed publication has been produced using more than one dataset, the User should adhere to the partner stipulation which has the longest review period. Where this is not possible, full references to research outputs are required for CDRC audit purposes. Please email publications@cdrc.ac.uk when publications are ready for deposit or logging.
1.8. Published outputs must include an acknowledgement stating "The data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC 297-01, ES/L011840/1; ES/L011891/1".
1.9. Users are requested to provide a brief case study of the supported work to be included in the Research Outputs section of the CDRC website https://www.cdrc.ac.uk/research/

On behalf of CDRC: Nick Bearman

NE Bearman

Date: 23/09/2019