P^3 Big Project Proposal

Team Members: SA, JL, LM, NT
Project Topic: To what extent has bicycle theft in Toronto become more prevalent since 2023?

Introduction
Toronto residents are accustomed to the preponderance of cyclists sharing the busy streets with automobiles. Amongst large North American cities, Toronto boasts a lengthy bike route network with connections to most subway stations [1]. Many inhabitants utilize this infrastructure as a popular mode of transport throughout the year [2]. While the city's Transportation Services department made significant bike lane upgrades last year [3], this success was largely overshadowed by news articles highlighting a series of thefts and stabbings [4], contributing to an overall negative public sentiment towards public transportation security. It remains to be known if this negative perception lingers today.

Our motivation for pursuing this project stems from a shared appreciation for biking and a desire for the city of Toronto to improve its bikeshare user safety. During our second brainstorming session, two group members highlighted their passion for biking as a hobby and as a necessity for commuting. In addition, everyone was aware of occasional bicycle thefts occurring around downtown, including the University of Toronto St. George campus [5]. Based on these experiences, we aim to use our data science skills to understand and explain why biking within the city has become more dangerous and less enjoyable for users. This project is interesting as it will help us demystify the sensation of "feeling unsafe" when moving around Toronto. Given the exploratory nature of this topic and time constraints, our team will not incorporate machine learning as a component of this venture.

Background
Toronto has a long history of building and maintaining public bike infrastructure. To address bicycle theft, the city has implemented several strategies aimed at enhancing bike security for its citizens:

1. **Bicycle Registration**: Toronto residents are encouraged to register their bicycles with the Toronto Police Service. This registration helps in tracking and recovering stolen bicycles and acts as a deterrent to theft. By registering, you provide the police with essential information like make, model, and serial number which can be helpful to find a stolen bike [6].
2. **StreetSmartTO**: An initiative to help support Toronto's growing community of people that cycle. These workshops aim to educate people on effective strategies for securing their bicycles and enhance ride safety in the city [7].
3. **Secure Parking Location**: Cyclists are advised to park their bicycles in visible, well-lit areas and to use designated bike racks that are securely fastened to the ground. These parking racks are common in the Toronto area, especially around the university campus [8].

Dataset

The primary dataset for our project is publicly available from the City of Toronto [9][10] and the Toronto Police Service [11]. Our datasets consist of multiple different data types, including geospatial (GIS shapefiles) [12], HTML tables [13], .csv files, text files, and HTTPS requests for web scraping. To collect data, we will first navigate to the dataset through a URL. For data that is more difficult to access, we will utilize APIs. All data will be cleaned and organized into readable forms prior to analysis and visualization.

Project Plan

Our team is committed towards producing quality deliverables expected of aspiring data scientists. To achieve this goal, we will establish a regular time during the week to meet and work together towards fulfilling project milestones. Discussions will be in-person, unless circumstances warrant a virtual meetup. We will also endeavour to schedule time with our project mentor, Ali, for guidance on tackling difficult obstacles and resolving team conflicts when appropriate. When working remotely, team members will maintain communication via WhatsApp and utilize a shared GitHub repository for code version control throughout the project. All code will be written using the Python programming language and submissions will be organized using Jupyter notebooks. The division of tasks is summarized below.

Component	Deadline	SA	JL	LM	NT
Proposal Submission (PS)	Oct. 20	Risk assessment, planning, editing, submission, references	Dataset sourcing, motivation, inspiration, editing	Background, dataset sourcing, references, editing, formatting	Introduction, dataset sourcing, motivation, background, editing
Data Wrangling (DW)	Nov. 25	GIS file clean-up, writing code, visuals	Clean-up, file parsing, writing code, data analysis	Clean-up, file parsing, writing code, data analysis	Clean-up, writing code, data analysis/visuals
Medium Article (MED)	Dec. 4	First draft writing, creation of visuals, editing	Second draft writing, editing, and image uploads	First draft writing, editing, formatting, submission	Second draft writing, editing, formatting
Final Presentation (PRES)	Dec. 4	Content development, speaking, and answering questions	Visual edits, writing, speaking, answering questions, submission	Creation of shared presentation document, content development, speaking, answering questions	Visual edits, proofreading, content development, speaking, answering questions
GitHub Repository (GIT)	Dec. 4	Upload of GIS files, code clean-up, commits, pushes	Proofreading code, commits, pushes, code clean-up, testing, Markdown file	Upload of files, code cloning, commits, and pushes, code clean-up, testing	Proofreading code, commits, pushes, code clean-up, testing, submission

The foreseeable project timeline after proposal submission is also presented below.

Name	Intervals					
Dates	**10/20-10/26**	**10/27-11/02**	**11/03-11/09**	**11/10-11/16**	**11/17-11/23**	**11/24-12/04**
DW	Download files, setup notebooks, and absorb the nature of the data.	Create a data pipeline, then parse, filter, clean data. Create visualizations to help communicate data trends. Source other datasets if needed. Analyze results.		Further refinement of data to converge upon a compelling story.	Final analysis and visualizations. Make code efficient, polish notebooks and organize files. Finish DW by 11/25 or earlier.	
MED	Set up an account and read the beginner's guide.	Create an outline for the article. Incrementally add content based on results obtained from data analysis. Write rough drafts.			Finalize draft and proofread article. Fix formatting and communication issues prior to submission.	
PRES	Initialize a presentation template. Work on rough outline iterations alongside DW. Decide on an aesthetic as development progresses.		Update and refine content quality and layout. Summarize and synthesize key results. Communicate significance. Incorporate (some) use of multimedia.		Prepare script. Refine content. Update plots. Check for errors.	Add and finalize content, rehearse, time, practice for Q&A.
GIT	Initialize and clone repository.	Clone, add, commit, and push minor and major updates to the repository as needed. Check version control.			Clone, commit, and push last revisions. Upload all files, including a README.	

Risk Assessment

As with all team projects, risks exist and must be acknowledged. Within the scope of this data science project, the most significant risks are described as follows:

Risk	Likelihood	Mitigation Strategy
Schedule conflicts with term work.	**High:** Team members have maximum course loads and an overall busy semester.	Planning early in the term is instrumental. This enables the identification of potential conflicts so that our team can strategize accordingly when assigning roles. Flexible working hours allow members to contribute according to their individual schedules.
Group member(s) not fulfilling work expectations.	**Moderate:** Although we have heavy course loads, we have distributed work according to strengths and should facilitate fulfilling work expectations.	Our team will check-in through the WhatsApp group chat. We will schedule weekly meetings to ensure all members are aligned and any concerns are addressed promptly. Assign clear roles and responsibilities for the project so we can work efficiently.
Data quality, cohesiveness, and integration challenges.	**High:** We will rely on many different types of public data, which may be incomplete.	Our team will continue to explore additional data sources, in case there are shortcomings in a particular dataset. We will also be wary of possible formatting and naming convention inconsistencies across the different types of data.
Late submissions of project deliverables.	**Low:** We have internal and official deadlines recognized in this document and/or our personal calendars.	Our team has a clear and robust project plan that helps to track deadlines, tasks, and milestones. Through the project plan, we can monitor progress against the timeline, enabling early detection of potential delays.

References

[1] PBSC Urban Solutions, “Station Information HTTPS Request,” [Online]. Available: https://tor.publicbikesystem.net/ube/gbfs/v1/en/station_information. [Accessed: Oct. 20, 2024].

[2] City of Toronto, “Cycling in Toronto,” Toronto.ca. [Online]. Available: https://www.toronto.ca/services-payments/streets-parking-transportation/cycling-in-toronto/ [Accessed: Oct. 20, 2024].

[3] City of Toronto, “2023 Cycling Year in Review,” Toronto.ca. [Online]. Available: https://www.toronto.ca/wp-content/uploads/2024/05/8fea-2023-Toronto-Cycling-Year-in-Review.pdf [Accessed: Oct. 20, 2024].

[4] CityNews, “Timeline: Disturbing trend of violent attacks on the TTC,” [Online]. Available: https://toronto.citynews.ca/2023/01/26/ttc-violence-timeline-attacks/ [Accessed: Oct. 20, 2024].

[5] Siew, Angelina, “Riding into spring: safety, awareness, and eco-friendly transportation,” [Online]. Available: https://www.fs.utoronto.ca/news/riding-into-spring-safety-awareness-and-eco-friendly-transportation/ [Accessed: Oct. 20, 2024].

[6] Toronto Police Service, “Bicycle Registration,” TPS.ca. [Online]. Available: https://www.tps.ca/services/bicycle-registration/ [Accessed: Oct. 20, 2024].

[7] City of Toronto, “Cycling Safety Education & Theft Prevention,” Toronto.ca. [Online]. Available: https://www.toronto.ca/services-payments/streets-parking-transportation/cycling-in-toronto/cycling-safety-education-theft-prevention/ [Accessed: Oct. 20, 2024].

[8] Toronto Police Service, “Bicycle Safety and Theft Prevention,” TPS.ca. [Online]. Available: https://www.tps.ca/crime-prevention/bicycle-safety-and-theft-prevention/ [Accessed: Oct. 20, 2024].

[9] Toronto Police Service, “Bicycle Thefts,” Toronto Police Service Open Data. [Online]. Available: https://data.torontopolice.on.ca/datasets/TorontoPS::bicycle-thefts-open-data/about [Accessed: Oct. 20, 2024].

[10] City of Toronto, “Transit and Community Safety Data,” Toronto.ca. [Online]. Available: https://www.toronto.ca/city-government/data-research-maps/transit-community-safety-data/ [Accessed: Oct. 20, 2024].

[11] City of Toronto, “Budget 2024,” [Online]. Available: https://data.torontopolice.on.ca/datasets/584b12967d214bb9a673505d97295eea_0/explore [Accessed: Oct. 20, 2024].

[12] University of Toronto, “Toronto Transit Commission (TTC) Geospatial Data,” MDL Library. [Online]. Available: https://mdl.library.utoronto.ca/collections/geospatial-data/toronto-transit-commission-ttc [Accessed: Oct. 20, 2024].

[13] Statistics Canada, “Tax filers and dependants with income by total income, sex and age,” [Online]. Available: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1110000801 [Accessed: Oct. 20, 2024].