

# MGT 6203 Group Project Proposal Template

Please edit the following template to record your responses and provide details on your project plan.

## TEAM INFORMATION (1 point)

**Team #:**

**Team Members:**

1. Team Member Nick Cunningham; GT Id (OMSA) or EdX username (MM)

[Insert background information: Name, professional background, education background, previous analytics related projects you have worked on]

2. Team Member 2 Name; GT Id or EdX username
3. Team Member 3 Name; GT Id or EdX username
4. Team Member 4 Name; GT Id or EdX username
5. Team Member 5 Name; GT Id or EdX username

## OBJECTIVE/PROBLEM (5 points)

**Project Title:** Evaluating Risks Associated with e-bike and e-scooter ridership in NYC

**Background Information on chosen project topic:**

Evaluating Risks associated with e-bike and e-scooter ridership in NYC

**Problem Statement (clear and concise statement explaining purpose of your analysis and investigation):**

With the increase in alternative/electric modes of transportation there are also unknown risks to both the users of these technologies and to the vehicles and passengers sharing the road.

**State your Primary Research Question (RQ):**

What is the relationship between e-bike and e-scooters crashes in NYC?

**Add some possible Supporting Research Questions (2-4 RQs that support problem statement):**

1. Does time of day impact the survival rate of e-bike and e-scooter accidents?
2. Do certain boroughs or areas have a higher propensity for e-bike accidents?
3. Are men or women more likely to be involved in an e-bike or e-scooter accident?

#### 4. What was the pandemic's impact on e-bike/e-scooter accidents?

**Business Justification: (Why is this problem interesting to solve from a business viewpoint? Try to quantify the financial, marketing or operational aspects and implications of this problem, as if you were running a company, non-profit organization, city or government that is encountering this problem.)**

Businesses and cities that provide or allow these forms of transportation need to be able to weigh the costs and benefits. If there is a decrease in traffic congestion and increase in convenience, does this outweigh the potential for property damage and personal injury. This has clear financial implications for the city of New York, as well as the viability of the business model for companies that either directly sell electric products or provide them as a service.

### **DATASET/PLAN FOR DATA (4 points)**

**Data Sources (links, attachments, etc.):**

<https://catalog.data.gov/dataset/motor-vehicle-collisions-crashes>

<https://catalog.data.gov/dataset/motor-vehicle-collisions-person>

<https://catalog.data.gov/dataset/motor-vehicle-collisions-vehicles>

**Data Description (describe each of your data sources, include screenshots of a few rows of data):**

The crash table contains details on the crash event. Each row represents a crash event.

The person table contains details for people involved in the crash. Each row represents a person involved.

The vehicle table contains details on each vehicle involved in the crash. Each row represents a motor vehicle involved.

**Key Variables: (which ones will be considered independent and dependent? Are you going to create new variables? What variables do you hypothesize beforehand to be most important?)**

Collision\_id is an important variable because it is the join key for the 3 tables.

Dependent variables – Number of persons injured, number of persons killed, etc.

Could combine this into one "injury" variable

Independent variables – Vehicle type, borough, crash time, contributing factor, person\_sex,

## **APPROACH/METHODOLOGY (8 points)**

**Planned Approach (In paragraph(s), describe the approach you will take and what are the models you will try to use? Mention any data transformations that would need to happen. How do you plan to compare your models? How do you plan to train and optimize your model hyper-parameters?))**

First, we will join the tables. To create a reasonable comparison, we will then remove crashes that are vehicle on vehicle or single vehicle. This will be more manageable to work with. We will then perform data cleansing to remove missing/null values and any other transformations. Using the person\_injury as the dependent variable and transforming the vehicle\_type column into dummy variables we will use a logistic regression model to calculate the comparable probability of injury for different electric transports versus other non-vehicle crashes (i.e. bicycle, pedestrian, etc.). We will also include additional dependent variable, i.e. gender to see how this impacts the performance of the model and if there are other more important contributing factor to injury. If e-transport is found to be significantly more dangerous we can layer in additional variables to see if location (borough), time of day, etc. also have an impact on safety.

Lastly, we'd like to do explore treatment effects and use the pandemic as a natural experiment to explore if there were any changes in correlations before, during, and after the pandemic shutdown. If crashes went down but other elements increased (i.e. severity of injury, death, etc.)

**Anticipated Conclusions/Hypothesis (what results do you expect, how will you approach lead you to determining the final conclusion of your analysis) Note: At the end of the project, you do not have to be correct or have acceptable accuracy, the purpose is to walk us through an analysis that gives the reader insight into the conclusion regarding your objective/problem statement**

We anticipate being able to prove that e-bikes/scooters are significantly more dangerous than other forms of non-vehicular transportation and that the risks is increased during certain times and in certain areas. Additionally, we anticipate proving that while the pandemic decreased absolute number of crashes, it increased the likelihood of death. Potentially due to other factors like collective stress, carelessness.

**What business decisions will be impacted by the results of your analysis? What could be some benefits?**

Government policies could definitely be impacted by the results. If the analysis is able to prove high risk of injury the city of New York could ban e-bikes/scooters. If certain times or areas show to be particularly dangerous, it could also institute a partial ban (no e-bikes from 4-7pm, no e-bikes in Manhattan, etc.)

## **PROJECT TIMELINE/PLANNING (2 points)**

**Project Timeline/Mention key dates you hope to achieve certain milestones by:**

TBD

**Appendix (any preliminary figures or charts that you would like to include):**