**Problem Statement Presentation**

Weining Hu

The George Washington University

Whu369@gwu.edu

**Source of Data**

* The New York City taxi and Limousine Commission

**Dataset**

* Green taxi data from June 2019 (to be determined)

**Background**

Hi everyone, I decided to spend the following two months in analyzing New York City green taxi data. As you may or may not have known, there are two types of taxis in New York. The most well-known type is yellow taxi. It is meant to serve passengers inside of the densely populated areas of Manhattan. As opposed to the yellow taxi, green taxis were meant to serve only the outer boroughs. Prior to 2013, most of the outer areas were underserved by yellow cabs and dominated by unlicensed drivers. Consequently, the New York State government launched the Green Taxi program in 2013 to tackle those problems. Many freelancing studies have been done to analyze the green taxi data collected in 2015. In my capstone project, I plan to analyze the most recent data in 2019 and compare my results with the previous research findings, and further evaluate how much better or how much worse the green taxis were performing in recent time.

**Project Plans**

I plan to dedicate most of my time to data cleaning, spatial data study, and GIS visualization. The final report will be complemented with a modeling section, in which I plan to build models to predict taxi tip as a percentage of the total fare. That’s all of my presentation. Any questions?