



**NATIONAL CENTER FOR TRANSPORTATION
CYBERSECURITY AND RESILIENCY**
A USDOT National Transportation Center

Data Management Plan

Project Title:	Privacy-preserving data analytics using synthetic data generation
PI Name and Contact:	Murat Kantarcioglu Professor University of Texas at Dallas 972-883-6616 muratk@utdallas.edu
Submission Date:	January 31, 2025
Version:	2.0
Project Performance Period	January 1, 2024 – December 31, 2024

Project Summary

This project aims to address the security and privacy concerns arising from the sharing of public transportation data. We use New York City taxi data as a case study and conduct systematically evaluations of the performance of widely used tabular data generative models. We propose a metric tailored to transportation datasets to measure the fidelity of these models and their privacy preservation.

1. Data Description

The data used in this project are as follows:
(The data can be collected by downloading it from the links)

1. 2015 Green Taxi Trip Data
(https://data.cityofnewyork.us/Transportation/2015-Green-Taxi-Trip-Data/gi8d-wdg5/about_data)
2. New York City Green Taxi Trip Records from March 2019
(<https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page>)

Link to the implementation: <https://github.com/chengenw/transportation>.

Note: Please download the code if needed, as the link may not be permanent.

2. Data format and metadata standards

Dataset 1 noted above is in csv format and dataset 2 noted above is in the PARQUET format.

3. Access Policies

The datasets are public.

4. Policies and Provisions for Re-use, Redistribution, Derivatives

There are no ethical issues involved. Datasets are publicly available.

5. Plans for Archiving and Preservation of Access and Data Integrity

The datasets are public.

6. Change Log

Version	Date	Description
1.0	1/31/2025	Provide description
2.0	5/16/2025	Added the implementation link