Analyzing Taxi Trip Patterns in NYC (2019–2021)

Team Members:

Yibei Li (yl3692), **Yuchuan Zhang** (yz2947), **Joy Wang** (zw673)

Dataset Description:

We will be using the NYC Yellow and Green Taxi trip datasets from 2019 to 2021. The dataset contains key trip information, such as pickup and dropoff times, trip distances, fare amounts, passenger counts, and geographic information (PULocationID and DOLocationID), etc. Yellow taxis primarily operate in Manhattan, while Green taxis serve outer boroughs like Brooklyn, Queens, and the Bronx, with limited access to Manhattan. The dataset also includes variables that account for the taxi fare structure, such as the RatecodeID, payment type, and congestion surcharge, which allow us to investigate pricing dynamics based on taxi regulations.

NYC Taxi Fare: https://www.nyc.gov/site/tlc/passengers/taxi-fare.page

NYC Taxi Zones: https://data.cityofnewyork.us/Transportation/NYC-Taxi-Zones/d3c5-ddgc

Yellow and Green Taxi Trip Records: https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page

Taxi Trip Data Dictionary:
https://drive.google.com/drive/folders/1_uNv3ptBu1VkD8bK_MRVZNeMFaaYjGBv?usp=drive_link

Research Questions:

1. ***How did the COVID-19 pandemic impact taxi usage patterns for both Yellow and Green taxis in terms of ridership and revenue?*** *(after considering raising prices)*
   This question aims to explore how the pandemic affected the taxi industry by analyzing changes in the number of trips, fare revenue, and trip distances over time.
2. ***What are the geographic trends in pickup and dropoff locations for Yellow and Green taxis, and how did these trends evolve during and after the pandemic?***
   This question will help uncover spatial changes in taxi trips. Using geospatial analysis, we can identify popular pickup and dropoff zone shifts across different periods.
3. ***Optional： How far can a passenger travel from Cornell Tech on Roosevelt Island using a taxi, given a budget equivalent to a typical lunch (~$20)?***
   We will calculate the maximum trip distance a passenger can cover with a set budget, accounting for fare components such as base fare, surcharges, and taxes. We'll use 'trip_distance', 'fare_amount', and related columns to estimate how far a passenger can travel within budget from Cornell Tech.
4. ***Is there a relationship between taxi trip patterns (e.g., trip distance, fare amount) and the average income of the pickup/dropoff areas?*** In this question, we will explore whether there is a correlation between the income levels of different neighborhoods and taxi trip characteristics like fare amount, trip distance, and passenger count. We will investigate if higher-income areas see longer trips, higher fares, or any other noticeable patterns in taxi usage compared to lower-income areas.

Statistical Methods:

- **Descriptive statistics:** To summarize the distribution, mean, average, and media of the taxi dataset.
- **Time Series Analysis:** To analyze trip counts and revenue changes over time, particularly focusing on the impact of the COVID-19 pandemic.
- **Geospatial Analysis:** Using ‘PULocationID’ and ‘DOLocationID’ to map and analyze pickup and dropoff locations across the city’s boroughs and identify trends over time.
- **Budget-based Distance Estimation:** To estimate the maximum travel distance possible with a $20 budget from Cornell Tech, accounting for all fare-related components such as base fare, taxes, and surcharges.
- **Correlation and Regression Analysis:** To examine the relationship between trip patterns (e.g., fare amount, trip distance) and the average income of neighborhoods, identifying whether regional income levels influence taxi usage characteristics.