Introduction to Database Systems

Final Project Proposal

10811020 陳光悦

10801128 陳俊鴻

0816091 荊姿芸

chatGPT

2.1 **Data**

- The description of your data
  - Introduction to the data

    The datasets for yellow and green taxi trips contain fields that record the date/time of pick-up and drop-off, pick-up and drop-off locations, trip distance, itemized fares, rate types, payment types, and driver-reported passenger counts. These data were gathered by technology providers authorized under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP) and provided to the NYC Taxi and Limousine Commission (TLC). The TLC did not create the trip data and therefore does not guarantee its accuracy. Meanwhile, For-Hire Vehicle (“FHV”) trip records capture the dispatching base license number and the pick-up date, time, and taxi zone location ID. These records are submitted by bases and published by the TLC as is, without confirming their accuracy or completeness. Hence, they may not represent the total number of trips dispatched by all TLC-licensed bases. The TLC does conduct routine reviews of the records and takes enforcement actions as necessary to ensure complete and accurate information. The shape file for taxi zone locations is included below.

  - Where is the data from

    The Official Website of the City of New York

(https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page)

- What do the columns and tables mean

Key	Explanation
VendorID	A code indicating the technology provider that provided the record.
trip_pickup_datetime	The date and time when the passenger(s) were picked up.
trip_dropoff_datetime	The date and time when the passenger(s) were dropped off.
passenger_count	The number of passengers in the vehicle. This does not include the driver.
trip_distance	The distance of the trip in miles.
RatecodeID	The rate code for the trip, indicating the type of fare.
store_and_fwd_flag	A flag indicating whether the trip record was held in the vehicle memory before sending it to the vendor.
PULocationID	The location ID where the passenger(s) were picked up, as defined in the Taxi Zone Lookup table.
DOLocationID	The location ID where the passenger(s) were dropped off, as defined in the Taxi Zone Lookup table.
payment_type	The payment method used for the trip.
fare_amount	The metered fare amount in dollars and cents.
extra	The extra charges and/or credits added to the fare, such as rush hour surcharges.
mta_tax	The amount of tax imposed on the metered fare.
tip_amount	The tip amount in dollars and cents.
tolls_amount	The total amount of tolls paid in dollars and cents.
improvement_surcharge	The amount of improvement surcharge in dollars and cents.
total_amount	The total amount charged to the passenger, including all taxes, tolls, and tips.

- Other information about your data (e.g. will it be updated in the future?)

  The data will be updated monthly, but we will only use the data of 2022..

- The source of your data

  https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page

2.2 **Application Design**

- Main idea
    - The purpose of your application
        - To know the number of taxis in different places and time periods.
- Functionality
    - What kind of information will be presented to users
        - The spatial and temporal density information of taxis in New York City.
    - What kind of interaction will be available
        - Search for different place and time period.
    - What will be the scenario when a user use your application
        - Anytime they wonder whether it's a good place for him/her to hail a taxi.
    - What kind of data will users be able to interact with
        - All kinds of data (as long as you remembered :))
- Interface
    - Expected interface look (use figure or text to explain)

      To show the number of taxis with different time periods in the map.

2.3 **Work Plan**

- Time schedule
    - 04/05 ～ 04/07：Final Project Proposal（陳光悅、陳俊鴻、荊姿芸）
    - 04/08 ～ 04/21：To be familiar with the datasets (and ChatGPT)
    - 04/21 ～ 05/05：Write PySpark.sql code
    - 05/05 ～ 05/19：Transfer to MySQL (if necessary)
    - 05/19 ～ 06/02：Creat the interface
    - 06/02：Final Project DONE !!!
- Discussion

  黃品澔 404 Not Found :)

  https://hackmd.io/@CHC1551/database

- Repo
  https://github.com/ginnyching/20203BD_Final_Project