CitiBike Dataset and NYPD Traffic Accident Dataset Analysis

The NYPD public accident dataset is accessed from the website, https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95/about_data using the api along with an access token. The token is obtained by registering at the site https://data.cityofnewyork.us/signup . The CitiBike trip datasets are available from the website https://s3.amazonaws.com/tripdata/index.html .

All instances of accidents involving bikes were retrieved and saved in the csv file bikeAccidentsNY.csv. One can easily perform this step by running the python script "create_BikeAccidentData.py" from the terminal using the command:
python create_BikeAccidentData.py --token Your_Token --output Your_OuputFilename
Download CitiBike trip data zip files from the website mentioned above, unzip them, and save the csv files inside the data folder named "citibike-tripdata". Due to hardware limitations, I have only used trip data csv files from January 2023 which were already 1795412 entries. However, the code can handle more entries on a suitable hardware.
Both datasets, bikeAccidentsNY.csv and citibike dataset are jointly processed by running the python script "analyze_CitiBike_and_NYPDbikeAccidents_data.py" from the terminal using the command:
python analyze_CitiBike_and_NYPDbikeAccidents_data.py

We found that there were 61140 bike accident instances in total out of which in 4286 instances the cyclist was unharmed, in 56609 instances the cyclist was injured and in 245 instances the cyclist died. We broke down these total figures by the five districts in New York, i.e., Bronx, Brooklyn, Manhattan, Queens and Staten Island. At the same time, we found out the number of trips that originated from these five districts from the CitiBike dataset, which were, 41063, 393563, 1272488, 88298, 0 corresponding to Bronx, Brooklyn, Manhattan, Queens and Staten Island respectively. We contrasted these numbers against the population of these five districts and found that the number of Citibikes rented per 100k inhabitants is far larger than the deaths of cyclists per 100k and injuries to cyclists per 100k in each of the five districts mentioned. So, in my opinion it is beneficial for both CitiBike and an Insurance company to get into a collaboration. The insurance company can have a steady revenue with very less number of serious claims resulting due to accidents. At the same time, CitiBike can grow their business more by assuring their customers that they provide insurance coverage in the event of unlikely accidents.

Instructions for running the code

1. Create a python virtual env, install pandas, numpy, matplotlib, sodapy, glob inside it.
2. Create an api token by registering at the site https://data.cityofnewyork.us/signup
3. Activate the virtual env and run:
 python create_BikeAccidentData.py --token Your_Token --output Your_OuputFilename
This will create the bike accident data.
4. Download and place CitiBike trip datasets, .csv files, inside the folder citibike-tripdata
5. Run python analyze_CitiBike_and_NYPDbikeAccidents_data.py