**NYC Taxi Trip Data – A Data Science Project**

Aviv Lahat, aviv.lahat@mail.huji.ac.il, a\_lahat1

Haya Riesel, haya.riesel@mail.huji.ac.il, hrizel

Miriam Goldstein, Miriam.goldstein1@mail.huji.ac.il, miriamgo

**The Idea:** In NYC, it’s very likely that one’s form of transportation will be in a yellow cab. Observing taxi traffic can provide city planners, citizens and taxi companies with information in diverse aspects, such as pricing, hourly and season patters, insights on where to direct taxi drivers and more.

We aim to analyze these aspects and draft a recommendation for city planners based on NYC taxi trip data that spans over several years.

**The Data:** From the NYC OpenData (<https://opendata.cityofnewyork.us/>)

1. 2011-2021 Taxi Trips   
   From each year we queried the necessary rows and columns for the specific questions. Each year is ~8 GB of data, about 114 M taxi rides, where each row is a single ride and the corresponding columns are pick-up and drop-off time, passenger count, trip distance, etc.
2. NYC Taxi Zones – 4 KB GeoJSON file

New York is divided into 263 zones, each for which the json file gives information on area, id, borough, geometry, etc.

**The Solution:**

Part 1: Hey Party People!

When planning new neighborhoods in the city, city planners and different groups of citizens like families, young couples, students, etc. want to know where they should hunt for housing. Having information on the night life is a crucial consideration. To visualize this question, we mapped out taxi pick-up hot spots from 10pm to 5am, which we label as night-life hours. In addition to knowing where the night life is happening, we looked at frequent pairs of pick-up and drop-off spots. These could be, for example, popular student housing locations slightly out of the center, that have cheaper rent and the corresponding student bars to go out to.

IMAGES

If you want to live in the night life center, from this data the recommendation would be that you should lean towards **area 1** or **area 2** where **high percentage** of taxi rides were around here during the night-life hours. Otherwise, for example, if you’re a family with children, you’d want something quieter and the recommended areas are **area 3** or **area 4** where only **low percentage** of taxi rides were there throughout the night-life hours. To evaluate these findings, we focused on Manhattan and checked two different aspects: the locations of schools for children K-12 grade, and the locations of the top-rated bars and clubs.

**Evaluation:**

**Future Work:**

**Conclusion:**