RBDA Project Proposal
Liveability Assessment Using Big Data Analytics

Team
Animesh Ramesh(ar8006), Efe Kalyoncu(ek2608), Suryakiran Sureshkumar(ss16030), Prasanna A P(pa2490)

Abstract
As one of the most populous and diverse cities in the world, New York City presents a wide array of neighborhoods for residents to call home. However, quantifying the livability of different areas can be challenging due to the many factors that contribute to quality of life.
This project aims to leverage open data sources to generate a data-driven livability score for neighborhoods across New York City's five boroughs. Datasets spanning transportation access, food availability, safety, public health, and local amenities will be integrated and analyzed. Key data sources include NYC open data on subway stations, bus stops, Citi Bike, restaurant inspections, recognized shop locations, 311 complaints, and health facility information. Geospatial analysis will be conducted to calculate metrics such as distance to transit, food access, green space, health resources, and more for each neighborhood.
A weighted algorithm will be developed to generate an overall livability score on a scale from 1-100 for each neighborhood based on these underlying metrics. This work will demonstrate the power of unifying disparate open datasets to provide novel and actionable insights for public benefit. The methodology could be extended to other cities to generate similar livability indices.

Introduction
Liveability is fundamentally shaped by the availability and accessibility of essential services and amenities that impact the daily lives of residents. This project aims to assess the liveability of neighborhoods through a focused examination of four critical dimensions: transportation, food availability, safety, and health facilities. By harnessing the power of Big Data Analytics, we will analyze comprehensive datasets from public sources to construct a multi-faceted liveability index that reflects the true quality of life.

Transportation: A cornerstone of liveability, efficient transportation systems facilitate mobility and access to jobs, education, and services. We will evaluate the density and distribution of bus stops, subway stations, and Citi Bike stations to gauge the convenience and accessibility of public transportation options for residents across different neighborhoods.

Food Availability: Access to nutritious food is essential for health and well-being. By analyzing the location of restaurants and recognized shop healthy stores, including groceries and convenience stores, we aim to identify areas of food abundance and deserts, offering insights into the nutritional landscape of the city.

Safety: A safe environment is crucial for a high quality of life. Utilizing historic complaint and violent crimes data from the NYPD, we will assess the safety of neighborhoods, identifying

areas with higher incidences of crime that may require targeted interventions to enhance resident security.

Health Facilities: The accessibility of health services is a key indicator of a neighborhood's liveability. Through the analysis of health facility locations, we will determine the adequacy of healthcare provision and its distribution across the city, highlighting areas with potential service gaps.

Incorporating these datasets, our project seeks not only to identify and analyze current states of liveability but also to uncover patterns and correlations that can inform future planning and policy decisions. By creating a comprehensive liveability index, we aim to provide a holistic view of the environment, empowering residents with actionable insights to enhance the quality of life in their communities.

Datasets:

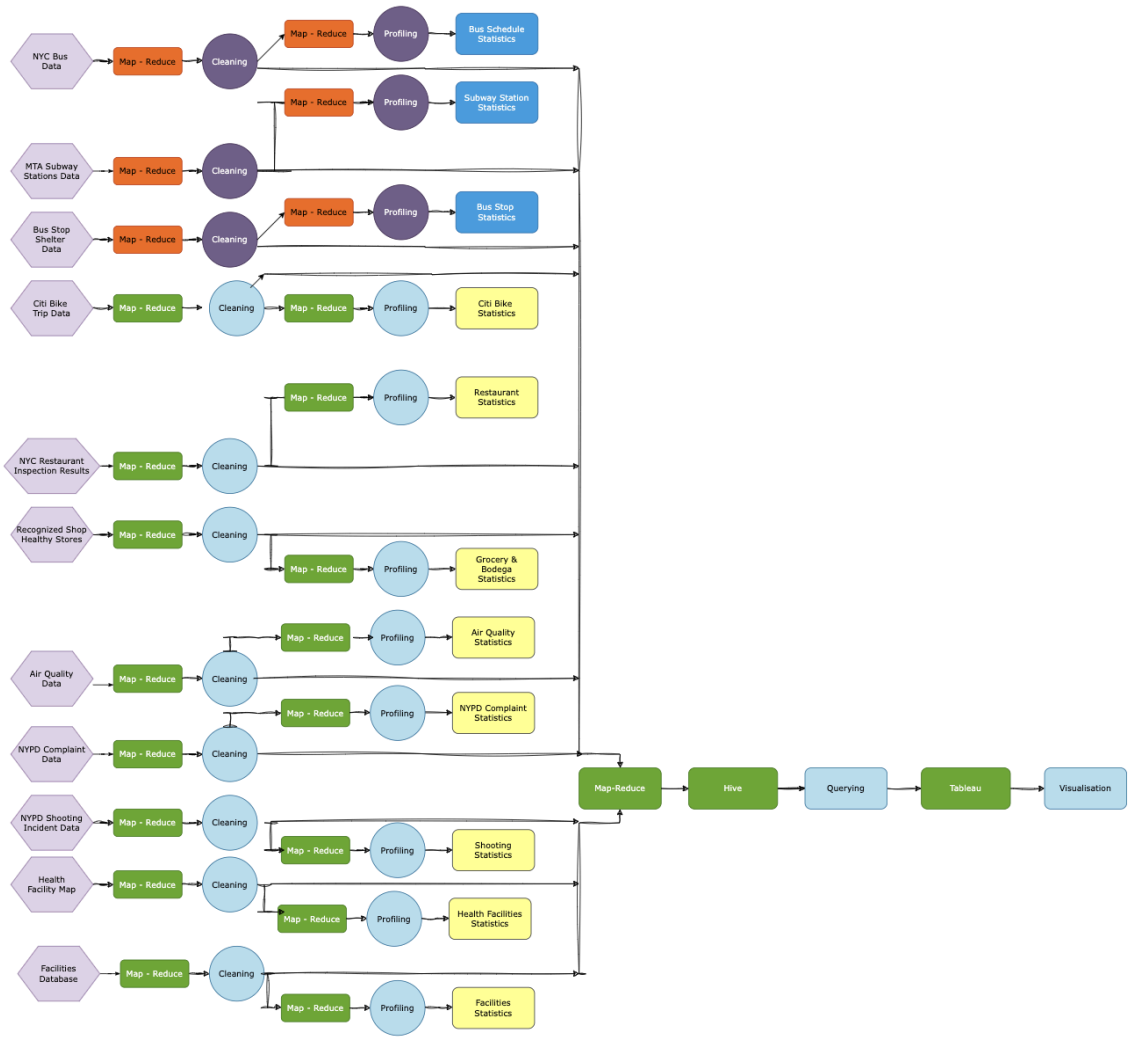
- **Bus Stop Shelters data -** This dataset contains the location of Bus Stop Shelters in New York City provided by NYC Department of Transportation.
- **MTA Subway Stations data -** A dataset listing all subway and Staten Island Railway stations, with information on their locations, Station Master Reference Number (MRN), Complex MRN, GTFS Stop ID, the services that stop there, the type of structure the station is on or in, and their ADA-accessibility status.
- **Citi Bike Trip data -** This dataset contains the Citi bike trip data and station location which is publicly available on Lyft City bike site.
- **New York City Bus Data -** This dataset originates from the New York City Metropolitan Transportation Authority (NYC MTA) bus data streaming service. It updates approximately every 10 minutes and includes information on bus locations, routes, stops, and more. Additionally, it features the scheduled arrival times from the bus timetable to indicate whether buses are running late, on schedule, or ahead of time.
- **New York City Restaurant Inspection Results data -** This dataset contains food safety inspections of restaurants in NYC provided by the state from its Open Data NYC program.
- **Recognized Shop Healthy Stores data -** This dataset contains information on bodegas and grocery stores recognized by the Shop Healthy NYC program for promoting healthier food options, including details on store names, addresses, zip codes, recognition years, and program waves. Managed by the NYC Department of Health, it highlights efforts to make healthy choices more accessible in specific neighborhoods, with annual updates on stores maintaining program standards.
- **NYPD Complaint Data Historic -** This dataset contains NYPD Complaint Incident Level Data, providing an up-to-date snapshot of criminal complaints with details on offenses, locations, times, and participant demographics. It employs specific geocoding for accuracy and excludes unfounded or voided complaints, ensuring a focus on valid criminal incidents reported in New York City.

- **NYPD Shooting Incident data -** This dataset contains NYPD Shooting Incident Level Data, capturing real-time approximations and updates of shooting incidents. It details handling multiple victims per incident, geocoding practices, and inclusion criteria focused on injury-resulting incidents. It also details the shooting's borough, precinct, jurisdiction, and includes demographic information about perpetrators and victims, alongside geographic coordinates in specific coordinate systems.
- **Health Facility Map -** This dataset from NYC Open Data features up-to-date inspection results for active NYC restaurants and college cafeterias, covering violations, grades, and types of inspections over the past three years. It reflects the adjudication outcomes where scores may be revised and notes a temporary pause in letter grading due to COVID-19 from March 2020 to July 2021.
- **Air Quality data -** This dataset, managed by the NYC Department of Health and Mental Hygiene, provides comprehensive information on air quality indicators in New York City, including pollutants like PM2.5, NO2, SO2, and O3. It offers insights into neighborhood-level air quality variations and the associated health burdens, including estimated deaths, hospitalizations, and emergency visits due to air pollution.
- **Facilities Database -** This dataset contains information on over 30,000 facilities and program sites related to City, State, or Federal services in New York City, compiled by the Department of City Planning into the City Planning Facilities Database (FacDB). It supports planning activities and allows New Yorkers to explore government resources in their neighborhoods.

Our Datasets:

Dataset	Size	Assigned To
Bus Stop Shelters data	589 KB	Efe Kalyoncu
MTA Subway Stations	65 KB	Animesh Ramesh
Citi Bike Trip Data	6.8 GB	Prasanna A P
New York City Bus Data	5.4 GB	Efe Kalyoncu
New York City Restaurant Inspection Results	98.4 MB	Animesh Ramesh
Recognized Shop Healthy Stores data	79 KB	Animesh Ramesh
NYPD Complaint Data Historic	2.84 GB	Suryakiran Sureshkumar
NYPD Shooting Incident data	5.8 MB	Animesh Ramesh
Health Facility Map	2.4 MB	Prasanna A P
Air Quality data	1.9 MB	Suryakiran Sureshkumar
Facilities Database	16.8 MB	Suryakiran Sureshkumar

Design


NYC Bus Data
Map - Reduce
Cleaning
Map - Reduce
Profiling
Bus Schedule Statistics
Map - Reduce
Profiling
Subway Station Statistics
MTA Subway Stations Data
Map - Reduce
Cleaning
Bus Stop Shelter Data
Map - Reduce
Cleaning
Map - Reduce
Profiling
Bus Stop Statistics
Citi Bike Trip Data
Map - Reduce
Cleaning
Map - Reduce
Profiling
Citi Bike Statistics
Map - Reduce
Profiling
Restaurant Statistics
NYC Restaurant Inspection Results
Map - Reduce
Cleaning
Recognized Shop Healthy Stores
Map - Reduce
Cleaning
Map - Reduce
Profiling
Grocery & Bodega Statistics
Map - Reduce
Profiling
Air Quality Statistics
Air Quality Data
Map - Reduce
Cleaning
Map - Reduce
Profiling
NYPD Complaint Statistics
NYPD Complaint Data
Map - Reduce
Cleaning
NYPD Shooting Incident Data
Map - Reduce
Cleaning
Map - Reduce
Profiling
Shooting Statistics
Health Facility Map
Map - Reduce
Cleaning
Map - Reduce
Profiling
Health Facilities Statistics
Facilities Database
Map - Reduce
Cleaning
Map - Reduce
Profiling
Facilities Statistics
Map-Reduce
Hive
Querying
Tableau
Visualisation