Understanding the Complex Interactions in New York City: Weather and Taxi

Introduction

Governments and decision makers can significantly benefit from the increasing amount of data collected by ubiquitous sensor networks. Among those are taxi and weather datasets.

In this project we developed a framework consisting in data aggregation and a web based interactive visualization, which enabled us to identify interesting patterns in New Yorkers' behavior.

Datasets

	Weather	Taxi
Attributes of interest	Temperature and weather	Pickup and Drop-off locations
Temporal resolution	Hours	Seconds
Time zone	GMT	EST

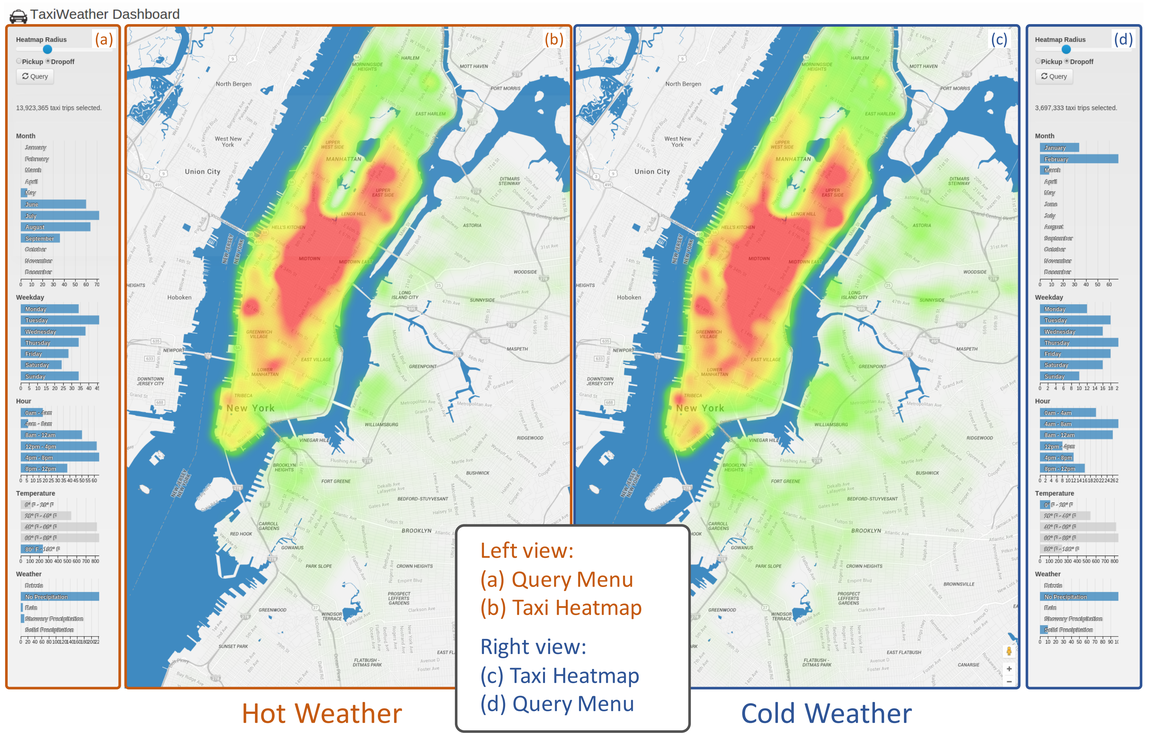
Data Processing

Data cleaning
- Remove taxi trips outside NYC
- Remove records with missing weather data

Join datasets
- Align Temporal Resolution
- Same number of records as Taxi

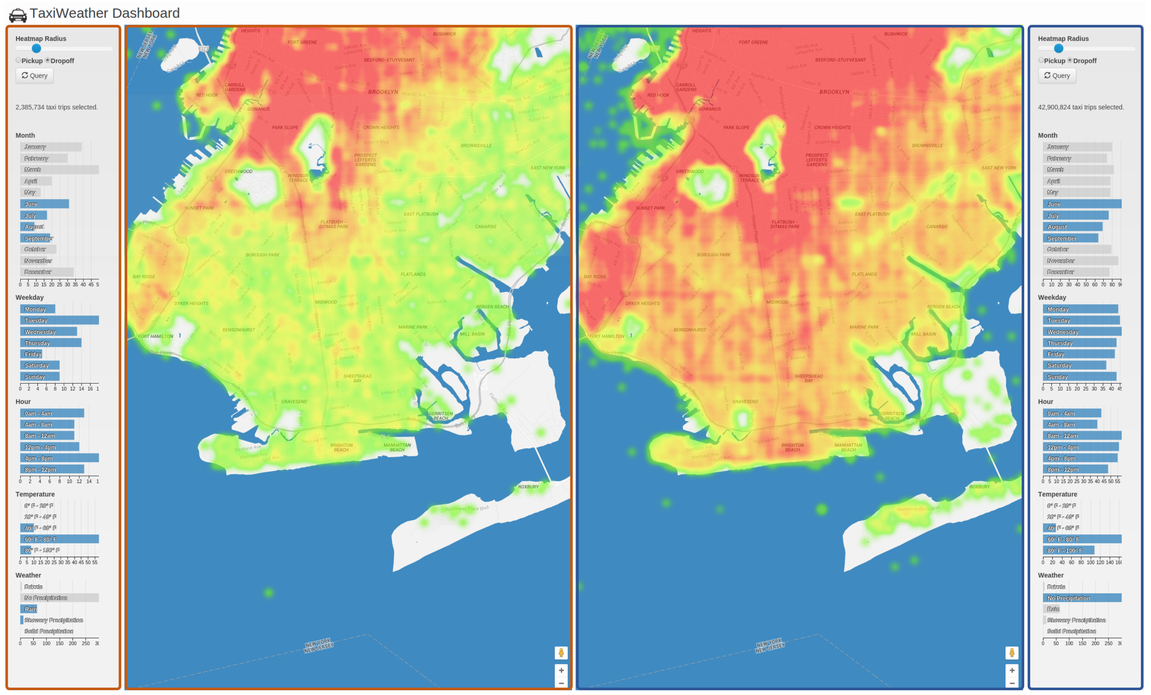
Aggregate data
- Mapper: ((month, weekday, hour, weather, temperature, bin_latitude, bin_longitude, pickup/drop-off) , 1)
- Reducer: Counter

TaxiWeather Dashboard



Hot Weather — Cold Weather

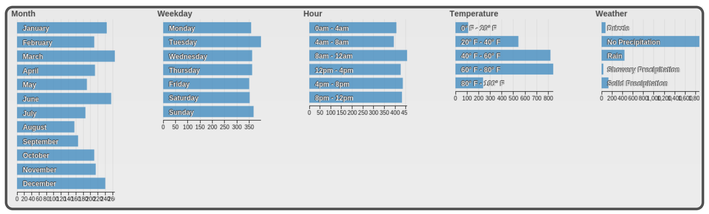
Taxi density increases in regions of Brooklyn and Queens when temperatures drop.



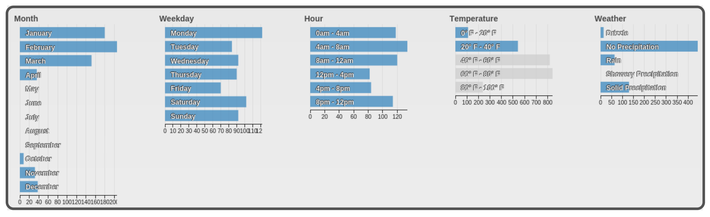
Precipitation — No Precipitation

During summer, the density of taxi trips at the beach is greater when there is no precipitation.

Querying the dataset



Bar charts without dimension filtering



Bar charts with dimension filtering

Conclusions

In this project, we implemented techniques for the exploration New York City's taxi and weather datasets. We used Spark and MapReduce to clean, join and aggregate our data. Furthermore, we developed an interactive visualization that enabled us to discover and formulate hypothesis about taxi behavior.

With our framework, a domain expert is able to perform a thorough analysis of the data, finding potential problems and proposing solutions to improve the quality of life in the city.

Future Work:

- Larger datasets, space and time span.
- Support for external databases, e.g. MongoDB.

Technologies

- Spark
- Google Maps API
- D3.js
- DC.js
- Crossfilter
- jQuery
- Bootstrap

NYU

https://github.com/jorgehpo/WeatherTaxiExploration

Raoni Lourenco, Jorge Ono