**Slalom Qlik Hack Challenge 2018**

C40 Cities Boston BERDO Data

Project Brief

Team Slalom (Data Models)

# Project Overview

The goal of this project was to create a solution that gave the C40 Cities and Greenovate City of Boston organizations a way to view performance of key emissions and energy consumption metrics. The solution intended to create varying levels of aggregation, and give the user a way to drill down to the individual building level to view performance against peers.

# Approach

When initially analyzing the BERDO data from the city of Boston, we noticed several data problems:

* Missing/Not Available Tax Parcels year to year from BERDO Data
* Tax Parcels containing foreign characters such as hyphens, semicolons, spaces, etc.
* Inconsistently recorded measures -- sometimes properties would be rolled up into one row, other times they would be broken out into multiple rows
* Missing/Not Available Property Types
* Missing/changing Lat/Lon information year to year (buildings probably don’t move)
* 100% Steam buildings greatly exaggerated GHG Emissions & Power Use
* Wang Theatre greatly exaggerated GHG Emissions & Power Use
* Many:Many relationship between buildings and tax parcels
* Slowly Changing Dimensions

We decided that in order to build an application that was useful it was best to address these data challenges by making manipulations to improve the accuracy. In the cases of missing or changing dimensions, we were able to retrieve these values from a more complete record of the same building from a different year. We were also able to extrapolate a list of PIDs (building identifiers) from each Tax Parcel, and break out our data then by PID, which proved to be a lower level of granularity than Tax Parcel. Although this process was quite complex and time consuming in the end it paid dividends by enhancing our reporting capability. This approach gave us a more accurate representation of buildings from a geographical perspective, as well as a complete picture of the data itself. We were then able to build subsets of the data such as Multi-Year Reporters and Property Type, Neighborhood, and Building Use Benchmarks.

In addition to the data pulled from BERDO and Tax Assessments, we gathered varying layers of geospatial data for Boston neighborhoods, city council districts, social vulnerability areas, water conservation areas, and areas impacted by sea level rise. We found the intersections of buildings to each layer to create associations between these layers then added these associations back into the data model in Qlik.

The focal point for our application was the map of the city of Boston and the buildings of interest. The building layers hat we’re gathered from the Boston data portal allowed us to turn layers on and off, as well as click layers to see the buildings that are associated to that layer. We also included options to choose other metrics to color by and filters to reduce the data, should the user care about a specific subset of buildings. We built supporting charts to assist with an overview of the data and provide context. The web application also displays additional building level information and a comparison to peers relevant peers when a single building is selected.

# Technology Used

Beyond building a solid foundation from the data in Qlik Sense, we utilized [React.js](https://reactjs.org/) for the project organization, and [Material UI](https://material-ui-next.com/) as a design framework. [Enigma.js](https://github.com/qlik-oss/enigma.js/) provides a connection to the Qlik Sense Engine API. [Mapbox GL JS](https://www.mapbox.com/mapbox-gl-js/api/) library is used to create the maps and [Turf.js](http://turfjs.org/) is used for some of the geospatial calculations. [C3.js](http://c3js.org/) is used to create the supporting charts.

# Topline Findings from C40/BERDO Data

* Buildings that reported emissions data each year between 2014 and 2016 – as required –appear to be making progress on emissions reduction, having reduced total emissions by 16 percent in that period.
* There were 339 such buildings in 2016, clustered in the downtown area, Back Bay, Allston and Roxbury.
* The three-year reduction in total energy usage was led by buildings with some of the largest footprints, including hospitals, medical centers and hotels. For example, 200 Clarendon (also known as the John Hancock Tower) is among the city’s largest and most prolific buildings, yet it outperforms its peers in Back Bay in Site EUI reduction between 2014 and 2016.
* West Roxbury stands out as being the largest average emitter and consumer of energy and pollutants among all Boston neighborhoods, led by the Village at Chestnut Hill. There is another pocket of high GHG emission intensity in Allston, led by the Harvard Business School campus.

# Recommendations for the Future

In the future, we would like to see more emphasis placed on the data collection of the BERDO data. Investing in the data is a crucial, yet often overlooked piece of analytical projects, due to the challenges it poses.

Additionally, we would like to augment this analysis with more data such as similar data from other cities, and renewable energy credits data. We believe that by supplementing this information with RECs, it could potentially become a tool used to target large polluters and inform them of potential financial benefits to the reduction in their GHG Emissions and energy usage, and provide them with detailed information about how they can lead by example, and be an active participant in going green.

# Screenshots





