**Outline:**

Business Problem and Questions

Data Sources

Model

Database & Dashboard we selected

Validate the problem

**Business Problem and Questions (Need to add information here)**

**Data Sources**

Big Belly Trash Bin Usage:

<https://metadata.phila.gov/#home/datasetdetails/5543866e20583086178c4f1e/representationdetails/55438ab49b989a05172d0d55/>

Big Belly Waste Baskets (Trash Bins):

<https://metadata.phila.gov/#home/datasetdetails/555f8139f15fcb6c6ed4414f/representationdetails/556de53bcf0e0dca19464e91/>

**Model (Still Undecided)**

Classification

RBM – unsupervised learning, monitor a system, building a binary recommendation system, or working with a specific set of data

* Monitor population to gauge frequency of trash pickup

Arima – time series model, categorizes moving averages and auto-regression

* Predict how much garbage is going to be generated, could be accurate for the next 4-5 years
* Moving average – if it hits a certain point, this would be the impact of trash

**Database**

SQL pgAdmin

We will access the data by using two CSV and creating tables for each file in pgAdmin. We then merged the files by creating a new table.

We have 99,632 observations

* The only change we had to make in the CSV was to rename one of the columns

We have 9 features

* Objectid
* serialnum
  + The serial number of the BigBelly on the street. Each BigBelly has a unique SN
* description
  + Intersection where the BigBelly is located
* Recycler
  + When yes the bigbelly has an attached recycler. When No, it is trash only.
* lat
  + Latitude of BigBelly location
* lng
  + Longitude of BigBelly location
* streamtype
  + The type of material the BigBelly holds. Typically in Philly, they are all Trash, but other customers have Single Stream, Bottles/Cans, Paper, etc.
* timestamp\_
  + The Date/Time of the collection
* level\_
  + The fullness of the bin when it was collected at the timestamp. We use a GREEN/YELLOW/RED system. GREEN is fairly empty (about 30 gallons of trash), YELLOW is full (about 90 Gallons) and RED is the highest (about 150 gallons).

**Dashboard**

Tableau

**Legend**

GREEN/YELLOW/RED system

GREEN is fairly empty (about 30 gallons of trash)

YELLOW is full (about 90 Gallons)

RED is the highest (about 150 gallons)

**Communication Protocols**

* Weekly meeting – Wednesday’s 8:30-9:30p
* Agenda for weekly meeting

**Next Steps**

Create a heatmap in Tableau showing the locations of the trash pickup based on weight

Finalize the Machine Learning Model