

## **Project Proposal**

### **Which MTA stations are optimal for street teams?**

At first this scenario is unrealistic, a nonprofit organization of breast cancer awareness solicited our engagement to use MTA subway data to help them optimize the placement of their street teams, such that they can gather the most signatures, ideally from those who will attend the gala and contribute to our cause.

If we assumed that the higher traffic of pedestrians would mean more signature, but this analysis would be more precise if leveraged by the proportion of individuals more interested in the cause of breast cancer awareness. We assumed, therefore, that female with high income or that work for health centers would be more prone to attend to the gala. For this purpose, we collected demographics data from the Census and identified the areas in NYC with more concentration of health centers.

### **Assumptions**

- Station traffic
- Women population
- Income level
- Close to health centers

### **The data**

the NYC MTA data - which was a good source for extracting the volume of foot passers nearby the subway stations, we collected the NYC Census Data for demographics by Census Block.

### **Tools**

Pandas – Python - SQL