**MTA Signatures Project #1 – Exploratory Analysis**

**Applying the CRISP-DM Process:**

**[CRISP-DM - Basic Guide](https://www.datasciencecentral.com/profiles/blogs/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome)**

1. **Business Understanding – What does the customer want?**
   1. Gather the most email addresses so we can send them free tickets. ***Should we have a target signature goal in mind?***
      1. MTA Turnstile Dataset - (<http://web.mta.info/developers/turnstile.html>) - This will measure throughput
   2. Minimize the amount of time we need to collect the emails. Don’t want to do this 24 hours a day for 5 months. ***Should we have a target time/manpower investment goal in mind?***
      1. MTA Turnstile Dataset – (<http://web.mta.info/developers/turnstile.html>) - Hourly, Daily, Weekly Granularity
      2. Demographic Data- (<https://datausa.io/profile/geo/new-york-ny>) – Information about Gender, Age, Income, Profession
   3. The people that we send tickets to should actually be able to come.
      1. Demographic Data- (<https://datausa.io/profile/geo/new-york-ny>)
      2. District Data (No Dataset yet)– Information about the type of neighborhood
2. **Data Understanding**
3. **Data Preperation**
4. **Modeling**
5. **Evaluation**
6. **Deployment**