|  |
| --- |
| ***Big Data Analytics - Team Project Proposal*** |
| ***Part 1. General Information*** |
| **Name:** Anuj Bora and Gowtham Goli  **Project Title:** Factors affecting health in New York City  **Project Description: *(Write one paragraph to describe what this analytic will do.)***  The analytic will extract health related requests from the NYC 311 service request dataset. Then, the data will be partitioned according to boroughs and it will be compared with water quality, air quality and garbage related complains obtained from NYC’s air quality dataset, water quality dataset and OATH ECB Hearings Case Status (which will be used to extract garbage complaints) to gauge the impact of those factors on health in boroughs of NYC.  **Describe who will benefit from your analytic:**  The project will help to conclude whether the air quality, water quality and undisposed garbage is deteriorating health conditions in the examined boroughs.  **Describe how you will prove the goodness of your analytic, in other words, how will you verify that it is correct:**  We are expecting to observe a direct correlation between the water quality complaints, air quality complaints and garbage related complaints to the health related issues that are potentially caused due to food poisoning, indoor air quality and increased rodents in the respective neighborhoods. A direct proportionality, if could be established between these datasets would serve as a proof to the goodness of our analytic |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Big Data Analytics - Team Project Proposal*** | | | | |
| ***Part 2. General Data Source Information*** | | | | |
| **Data Source** | **Data Source Description** | **Data Collection** | **Data Size** | **Data Frequency** |
| NYC OpenData: 311 Service Request Dataset | This data set contains historic data from 2010-present of all service requests made by calling to 311. The data will be filtered to keep records related to health. | ☐ Realtime  ☐ Batch  ☐ Static | GB | The data is historic and so will be loaded once. |
| NYC OpenData: Water Quality Complaints Dataset | This data set contains records of water complaints made from 2010-present | ☐ Realtime  ☐ Batch  ☐ Static | MB | The data is historic and so will be loaded once. |
| NYC OpenData:  Air Quality Dataset | His data set contains air quality values of various boroughs of New York from the year 2005 to 2013. | ☐ Realtime  ☐ Batch  ☐ Static | MB | The data is historic and so will be loaded once. |
| NYC OpenData:  OATH ECB Hearings Case Status | This dataset will be used to extract records of complaints made for uncollected garbage which can be used as a one of the reason for the growth of rodents leading to rodent complaints in 311 Service Request dataset. | ☐ Realtime  ☐ Batch  ☐ Static | GB | The data is historic and so will be loaded once. |