Paranormal Distribution (Gabriel Gilling, Muna Nwana) Hackathon Submission Statement

For this Hackathon we decided to help school administrators in NYC make better informed decisions for when to reopen their schools. Using Design Thinking, we imagined the experience of a school administrator that has to weigh different factors for reopening: whether their school has structural risks that would make it a potential hotspot for the virus, and whether the school is located in an area with high infection rates to begin with.

To accomplish this, we built a model in 2 parts: first, each school in our database is attributed a unique score representing its structural risk: the ethnic makeup of the school, the amount of children that are in poverty, whether it was flagged as being in overcapacity or under capacity etc. All of our data was extracted thanks to NYC’s Open-Data API and was loaded and wrangled in R. Next, we built a dynamic time series model that calculates a day-to-day risk score with historical data with infection and death rates for each Zip-Code in NYC. The data was scraped from GitHub and then loaded in wrangled in Python and then in R. Each day, our model updates the risk score for a given zip code given the new data that comes in, and then the day’s COVID-19 Risk Score and the school-specific risk scores are aggregated.

Our project makes extensive use of IBM Cloud Assets and demonstrates their value and flexibility. First, we used custom runtime environments in Cloud Pak for Data (CP4D) in order to use the RStudio IDE as well as Jupyter Notebooks. CP4D allowed us to efficiently share the data we found online, as well as the finalized datasets we created since everything is in a single environment.

Next, we leveraged Auto-AI and Watson Machine Learning (WML) within CP4D to create optimized models to analyze and forecast the risk index scores that we built. Those features would allow a school administrator to plug in their school’s information and get two sets of data points: first, their school’s unique risk score (based on the aforementioned factors), and then a rolling forecast based on temporal COVID-19 data.

The modeling can still be improved, and more data can be added, but our product’s structure would allow school administrators to make decisions that informed by state of the art predictive algorithms, something that is currently not easily available.

We then plan on extending our work to allow higher level decision making, for example by aggregating our data at the borough and then city-level to allow policy makers to make data driven decisions as well.