**Weekly Progress Report (Sept 25 – 29): Tilly, Victoria, Lalla-Aicha**

**Progress/Accomplishments:**

* After aggregating and separating the data into distinct regions, Lalla then conducted an EDA on the raw data. She discovered the number of missing data value points in our dataset, and then further analyzed the aggregated data. This included building boxplots and discovering potential outliers.
* Tilly first found the percentage of missing values in the raw data. Then she proceeded to find county names with features with missing values over all the dates and remove them. Using SimpleImputer for categorical features and KNNImputer for numerical features (n\_neighbors=5), she then imputed the missing values.
* Victoria aggregated the COVID-19 data by month and categorized it by seven regions: Northwest, Southwest, Midwest, South, Great Lakes, Southwest, and New England. She then conducted an EDA on the imputed data to discover any disparities. She then created rough ARIMA models to find some plausible parameters for further model building.

**Next Steps:**

* Plan and assign each member a model to build for our final analysis
  + ARIMA
  + Time Series
  + Neural Network