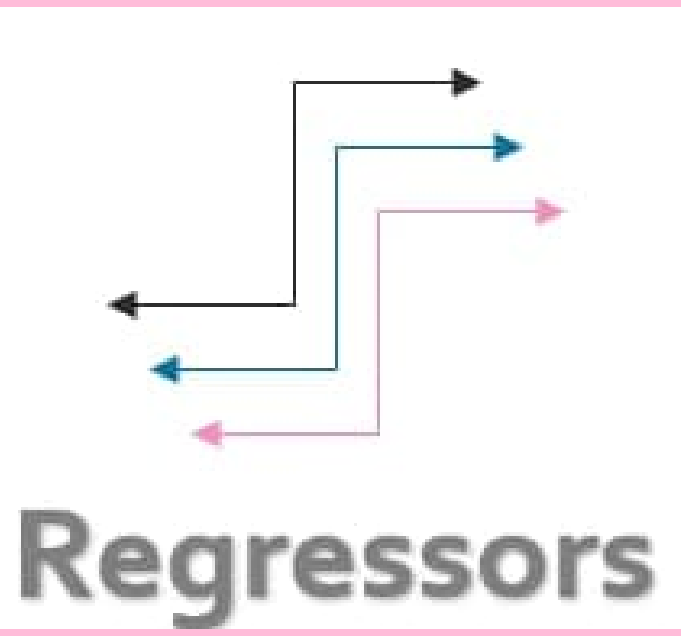
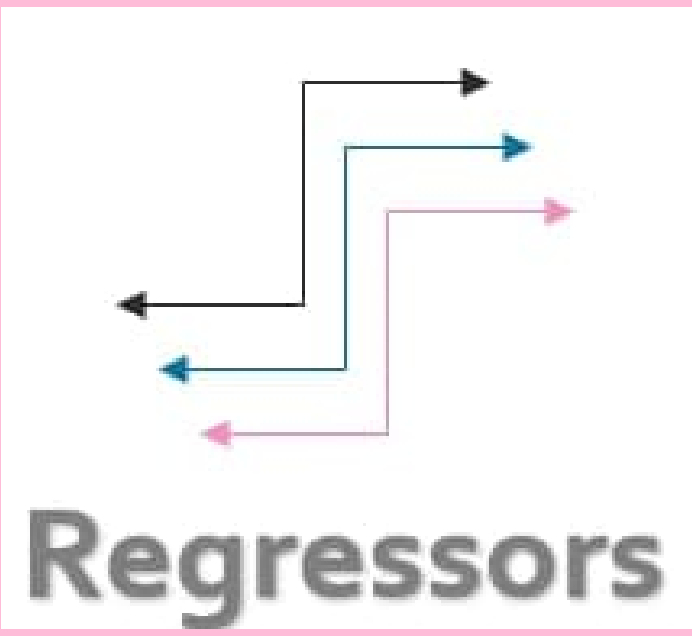


Regressors - Geographically Weighted Regression on Jakarta COVID-19 cases

Kwek Yi Chen¹, Ngah Xin Yan¹, Toh Jun Long¹

¹ School of Computing and Information Systems, Singapore Management University



Issues and Problems

Our project addresses the pressing need to curb the spread of COVID-19 in Jakarta by identifying what are the independent variables that are responsible for the spread of the virus.

Motivation

COVID-19 has become an indisputable part of our daily life ever since the virus spread to the majority of the world. Some countries are able to keep the situation under control, while some suffer the devastating effects from it. Among Asia countries, Indonesia has the highest COVID related mortality and positive rates for COVID-19 cases (Worldometers, n.d.), mainly in Jakarta, the main capital. This is likely due to certain underlying common factors within the countries. Researchers claimed that Jakarta could have as many as 4.7 million people who are possibly infected by the virus in March 2021 (Sood, 2021). This is alarming as this number constitutes to “nearly half” of Jakarta’s population. After our group was made aware about the seriousness of this matter, we decided to come up with Regressors, a Geographically Weighted Regression (GWR) application, to investigate the impacts of various variables (independent variable) on the mortality and positive rates (dependent variable).

The insufficient amount of GWR applications to collate insights of the various variables effectively and having an user-friendly application to run a wide range of GWR models with different configurations is what drives our research and application developing process.

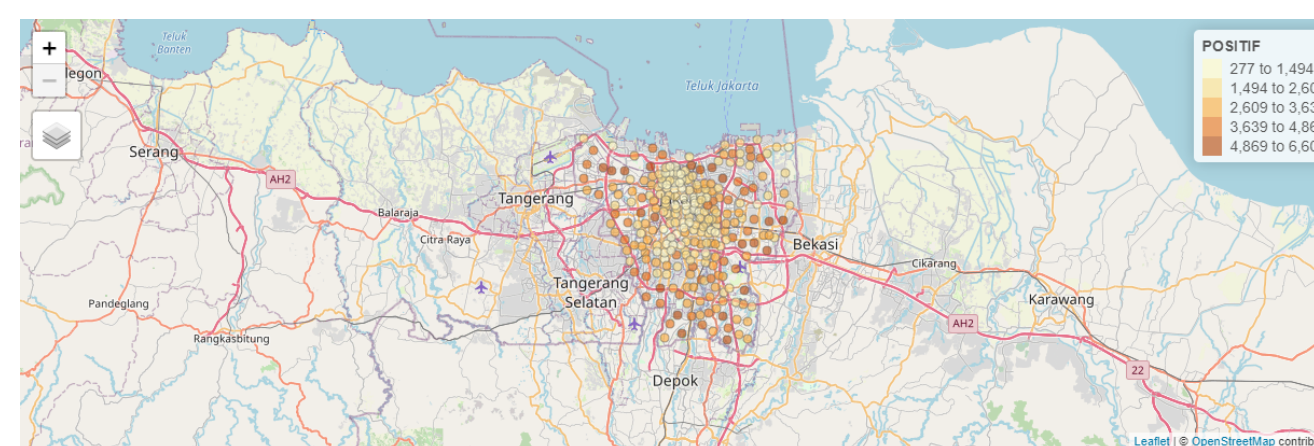
Approach

1. Data Preparation
 - o Data Sourcing: Look for data from website such as The Humanitarian Data Exchange.
 - o Data Wrangling: Import the data, change the projection of the data, filter out points of interest from the dataset, calculate proximity to each point of interest, and combine the data extracted into one dataset.
2. Exploratory Data Analysis (EDA)
 - o Spatial Point Map: To show the distribution of the selected variable geographically on an interactive map using tmap.
 - o Histogram: To show the distribution of the selected variable.
3. Geographically Weighted Regression (GWR) (Base Model/Prediction Model)
 - o Correlation plot is used to determine the most relevant explanatory variables for the regression models
 - o GWR Modelling
 - Ordinary least squares (OLS)
 - `ols_vif_tol` is used to identify multicollinearity among the variables
 - `ols_plot_resid_fit` is used to check linearity assumption
 - `ols_plot_resid_hist` is used to check for normality assumption
 - Bandwidth
 - `bw.gwr` (Fixed/Adaptive) is used to obtain the bandwidths selections for the GWR model
 - Model
 - `gwr.basic` of GWmodel is used to build a basic model

- `gwr.predict` of GWmodel is used to build a prediction model
- Approach
 - Cross Validation (CV) Score is used to identify a window size to get different subsets of the data to be parsed into the model. (Farber & Antonio, 2007)
 - Akaike Information Criteria (AIC) score is used to evaluate the model generated from the data and see how well the model can fit other data. It also determines the best evaluated model that best fits data. (Mennis & Jeremy, 2006)
- o Visualization plots the Local_R2/prediction geographically on an interactive map using tmap.

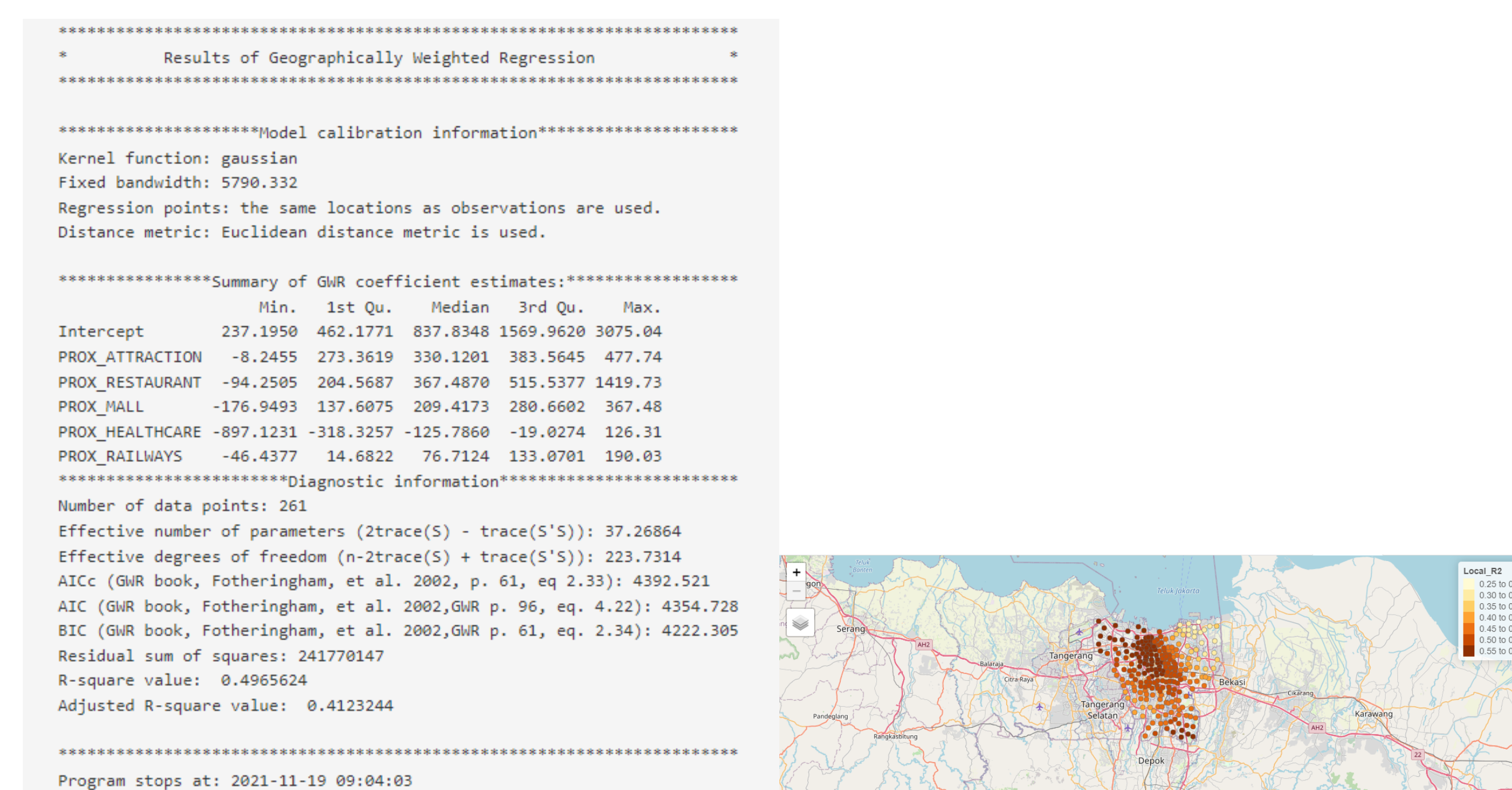
Results

Exploratory Data Analysis (EDA)



It is observed from the above that there are generally more positive cases around the outer boundary of Jakarta, whereas there are fewer positive cases in the central area of Jakarta. This indicates that the spread of the virus could have come from neighbouring provinces near the boundary of Jakarta which resulted in the high number of positive cases near the outer boundaries in Jakarta.

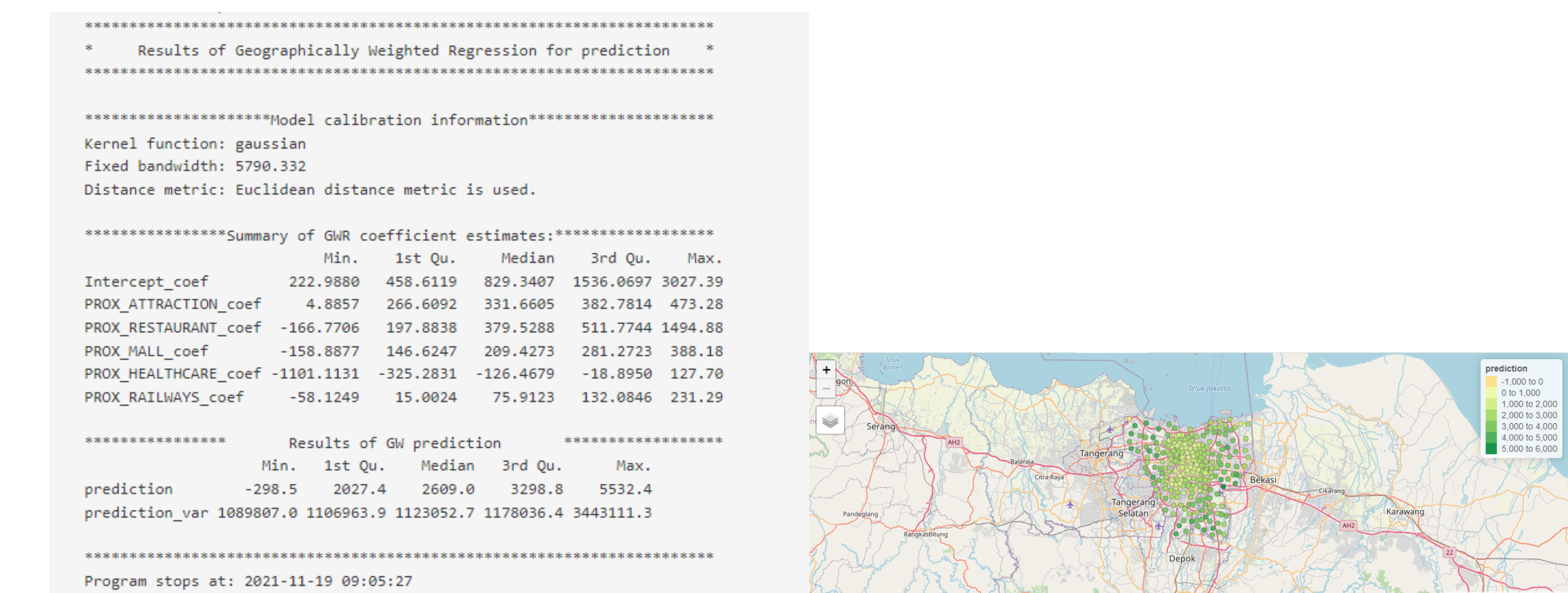
Geographically Weighted Regression (GWR)



The GWR model above is calibrated with significant independent variables after many iterations. The GWR model has an adjusted R-square value of 0.41 which means that the model is able to explain 41% of the number of positive cases in Jakarta. From the interactive map above, North and South-East regions of Jakarta have a higher Local_R2 value as compared to other regions. This indicates that the calibrated model better explains the number of positive cases in these regions, which also means that the independent variables, such as the proximity to

attractions, restaurants, malls, healthcare facilities and railways, have a strong relation to the number of positive cases.

GWR Prediction



The GWR prediction model predicts that there are a minimum of zero positive cases as the min is below zero. The predicted maximum number of positive cases is 5532.4. The median predicted number of positive cases is 2609. The interactive map above shows that East, South and North-West region of Jakarta generally have higher predicted positive cases as represented by the darker green spatial points. Closer observation can be done on these regions by the governments in order to minimise the spread of positive COVID cases.

Future Work

1. Cater for a diverse pool of user so that the barrier of entry to use the application is lower.
2. Allow deeper level of detail such that user can import their geospatial layers.

References

- Farber, Steven & Páez, Antonio. (2007). A systematic investigation of cross-validation in GWR model estimation: Empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems*. 9. 371-396. 10.1007/s10109-007-0051-3.
- Mennis, Jeremy (2006) “Mapping the Results of Geographically Weighted Regression,” *The Cartographic Journal*, Vol.43 (2), p.171-179.
- Sood, A. S. (2021, July 14). Indonesia Covid-19: Almost half of Jakarta’s population may have caught the virus, survey finds. Retrieved October 9, 2021, from <https://edition.cnn.com/2021/07/13/asia/indonesia-antibody-study-covid-intl-hnk-scli/index.html>
- Worldometers. (n.d.). COVID Live Update: 237,632,869 Cases and 4,851,284 Deaths from the Coronavirus - Worldometer. Retrieved October 8, 2021, from <https://www.worldometers.info/coronavirus/#countries>