Note: The page limit for the project report is 10 pages. You will be awarded negative marks if your report exceeds the page limit. If you need extra space, then you may consider including some details into Appendix. There is no page limit for Appendix.

For the course project you will work with the dataset that was collected to examine the relationship between social determinants of health and racial disparities in covid-19 vaccination at the county level in US. This study was reported here https://doi.org/10.1073/pnas.2107873118. The data (ime692_project.csv) for this project is uploaded along with this assignment. The definition of the variables used in this data are given below.

Variable	**Definition**
State	State in which the county is located
County	County name
IT_WholeRate	County level computer ownership and internet subscription data (in percentage)
HighSchool_WholeRate	County level education data. Percentage of population that qualifies as a high school graduate or higher
MedianInc_WholeAvg	County-level household median income data
republican_rate	The share of votes cast for the Republican candidate in the 2020 election
Segregation	Black-White segregation index measures. This index ranges from 0 (complete integration) to 100 complete segregation.
urban	A dummy-coded variable representing whether a county is considered urban or rural.
racial_weighted_bias	Measure indicating implicit racial bias in a county. Larger values indicate greater bias against Blacks.
hesitancy	Survey response for overall vaccine hesitancy in a county
HighSchool_Disparity	Difference in county level high school education between White and Black population
IT_Disparity	Difference in county level computer ownership and internet subscription between White and Black population
MedianInc_Disparity	Difference in county level median income between White and Black population
CvdVax_DisparityY	Difference in the Covid-19 vaccination rate (first dose) between White and Black residents in a County (in percentage).
vehicle	Proportion of vehicle ownership in a county
FacNumRate	Number of health care facilities per capita

CaseRate	Covid-19 cases per capita
Black_Prop	Proportion of black residents in a county
Test	Whether the observation is for test (1) or train (0)

Develop a predictive model in which the dependent variable is CvdVax_DisparityY. Your task is to examine the role of different socioeconomic and demographic variables in determining the covid-19 vaccination rate disparities. Note following points to complete this project.

(1) You are free to choose any prediction/regression model to examine the role of predictors in determining CvdVax_DisparityY. But use only data labelled as "train" for the model building purpose. Evaluate the performance of your models on the test data.
(2) Report three models that give best result on the test dataset. Mention the final model that you would select. Explain the reason for its selection.
(3) How would you assess the importance of different predictors in your model? Which predictors are most important in determining the racial disparity in covid-19 vaccinate rate?
(4) Are your findings similar to the results reported by authors in Table 1 of the article (see the hyperlink above)? If not, why?
(5) Upload Python/R code with your project report.