Given the rising COVID-19 cases the Singaporean government will always take the necessary precautions to prevent an unprecedented rise in COVID-19 transmission so this increases the currency indexing frequently as a workplace and school starts to close so in such a case right what would our surgency index be in Singapore So what we try to predict here is the strains index given the conditions that Singapore is in so we used para meters such as new cases total cases new devs new test for the difference and people fully vaccinated so firstly I will talk about our data so why is this data used is that the sources are very clearly documented and for at least one Singaporean data all of it is from MOH directly and make sure that their land this ensures that the data is accurate and not tweaked or it's false and also it contains very little missing values and is updated daily so this ensures that the data is is current and is ready to errors and for the for the for the day today is missing we filled it with a group trendline alesis using excel and also for some of the the values that are very obviously just filled it up manually so for example the vaccination rates before the vaccine came out is is 0 provide sources for very large variety of para meters that we can explore with so actually we were given our other we found the data set that has a lot of features and so we were able to narrow it down to six month that we think is the best SO4 model we actually considered 6 features and the first one is the new cases moved by million sorry corresponds to that community transmission of COVID-19 and gives a clear indication of prevalence of the virus in the community so correct so high higher of new cases are transmission read index Disney did very strong indicator of what the students index will be because right now I Singapore is moving towards the endemic policy higher number of cases but also give us an idea of who has a natural immunity so this will bring us closer towards herd immunity default does a lower she can see index the new devs also because December government is very concerned about the number of deaths they always say that they want to move towards the endemic face with as little casualties as possible so it would be obvious that if that the new devs increase then the next to go higher to lower than about text yeah so then a new test as well so usually when they have higher number of testing then we will be able to control the community transmission better so that we do not have to use lockdown sensor method of controlling the virus so we think the positive rate is also a very good indicator because any kids are I roughly like what is the reservoir of undetected cases so we we roughly know that how many of the community transmissions going on beyond what we tested positive so the last one is the people fully vaccinated 400 yeah so this one is is is very important as a line with our governments and any policy because when you are fully vaccinated yes your chances of dying so OK so we start by implementing the dependence my policy but not the packages and then we read the data and afterwards we bought all the functions from our cohort 9 so then from the hotline we actually made a new function called make model so this this creates a function for do one single function be able to train the the model as is required so the arguments are they offer value the better value is the number of iterations the starting number of rooms so are we are talking about in our data set we start from which value we start from with an average value so it would it is corresponding to the dates in on the excel file if you want to look at it so the starting one we yeah sorry so yeah so yeah misidentified movie starts at 46 which is where is Singapore actually started to have COVID-19 cases and then we just ended at the last value which is the the data we extracted the values from from all data source so if you just parameters be Simply put the features that we wanted to evaluate and we are also free to put other features maybe require them so far initial model me actually just beast artwork features using intuition guess which one is which features would most likely predict well all of the pretty distros index well and the initial model users the alpha values better values and a number of iterations that we calculated actually below that victim of the motor so for our initial model so this is our first Test we actually quite good value so I'll square values 0.5 Ms is 11.28 and mean absolute error is 0.23 so we try to improve upon the model by using five different conditions so the first one is we use the entire data set versus the selective data set so for initial model where she used to selective data set which is restart from roll 46 so it's about yeah when the first order to hit Singapore in February in 2020 so redeem that it's better to use a more target data set because stop to us actually started at 2020 at the start of the pandemic in China which are at the time was still relatively is in Singapore then secondly test using alpha values so we tried 0.010 point 013.05 and 0.1 so we actually I mean I knew this for all the three metrics and then appended values and then we we just plotted the graph and some data tables to compare the values so we found that with the upper range of 0.001 RNA 2000 children iterations the MSE value is actually 19.26 40.01 is 11.28 and for 0.05 onwards it's actually just 11.28 so since we are actually looking at seriously index at two decimal places we decided that our 0.01 is is good enough for us to predict the value so we decided that the best of our value is actually 0.01 so for better values we actually are so sorry that the starting data values so it's restart with zeros ones or fives or tents so yeah basically because of the number of iterations being so high and the grid and dissent is learning very well actually so that is there's a very insignificant difference but we still look at the values and we saw that the mean square error is lowest when we use zeros as our starting constants so we decided that zeros is our best better value to start with so to test the optimal number of iterations we actually created range 103 hundred all the way to 5500 five 1600 I think 5000 plus and then we again compare all the values so at about 100 iterations we get a 407 MSE values and about 700 regrettable 12 MSE value and we decided that at 3300 iterations is when that we actually get up to two decimal places in accuracy so we decided that 3300 would be the the least number of iterations that we need to obtain are good change the index up to two decimal places great so so lastly we considered the optimal features that we should consider so to do this right actually we once again we started from scratch and then we tried every single feature so it's not is is beyond those that are just that we consider above by order features that we have in our data set so similarly we we created graphs and tables and so we created one table and then we once again we added every single value in so how this works is that firstly we made our model with only total cases which is our first first feature and then we calculated our R-squared values MSE values and me values and after that we created another model with the first and second features so total cases with the addition of new cases then again we can be computed our metrics that she metrics are square MSCN every values and then same for the third one that we do total cases new cases new cases moved so 123 features so with that we calculated all our metrics and then we calculated the difference in the metrics and how much the successive or the additional feature would help all model optimize the the MSE values are values that every value so for absolute error is actually so small that it's pretty much no difference but for the mean squared error we identified five different features that actually made a huge difference to our our models so it's actually the total cases new cases new cases smoothed total deaths and a new case is smoothed you can see on the the table here CMSE\_value\_difference you see is a new difference there's no difference negative 27 point 394.8 negative 4.54 then you have NSRNSR NSR which stands for no significant reduction so from 4 from actually features fall too shelf right there's actually it actually helps to minimize our mean squared error just it is so small that is pretty much insignificant so we can see that our mean square error the 4th value is 16.99 while at the trove one is 16.18 so essentially fair little difference so we decided not to consider it so we redeemed at the five features that make a significant difference listen this year envy set up data design only features that we want to consider in our you know new model because that we we decided that the rest of the values would make a difference but you would make a very insignificant difference then not much point to include it especially when the datasets get very huge and the time complexity would just increase only be very very little difference in in optimizing our mean squared error and R2 squared value and mean absolute error so we sort of take them up and only include these five features