rezAWARE
ANALYTICS & DATA SCIENCE
3RD DECEMBER 2021

# Flattening the Curve: When do we dip below the 5% line?

***Clark County, Nevada***

## Abstract

There have been a total 354,924 reported Covid-19 cases in Clark County, Nevada with a daily 7-day average of 419 cases, to date. The famous tour destination: Vegas, was initially hit with a drastic drop in visitors in April 2020 with only 106,900 compared to 3,545,100 at the beginning of the year. They managed to rejuvenate to 3,390,200 visitors by October 2021. The World Health Organization (WHO) states that if a region's transmission rates fall below their recommended 5% safety threshold, that region would be able to loosen their preventive restrictions. It was within our interest to determine when Clark County can bounce back to the pre-COVID19 ways, i.e. drop below the 5% threshold. We applied machine learning algorithms to predict Nevada's test positivity rates - an important deciding factor that identifies if an area falls into a low, moderate, or high transmission zone. After performing five regression models, we concluded that the Random Forest Regressor performed the best with a mean absolute error (MAE) of 0.587 and a R-squared value of 95.1%. The results show that the transmission rates of Nevada, which are currently at 7.3%, will not fall below the 5% safety threshold by the end of January 2022. However, further predictions through our Gradient Boosting Regressor show a steady drop below the 5% mark during the month of June 2022.

# Introduction

With the community of Las Vegas getting out and about and adjusting to this new norm, the State of Nevada along with the World Health Organisation (WHO) needs to take responsible actions to safeguard citizens whilst considering how they can be allowed to resume daily activities. The safety threshold recommended by the WHO states that counties that are currently above the 5% test positivity rate will be required to follow Centers for Disease Control and Prevention (CDC) guidelines like wearing masks in indoor settings and crowded outdoor settings. A County that falls below this threshold and continues at a steady or lower rate for a period of 14 days will be allowed to loosen their restrictions.

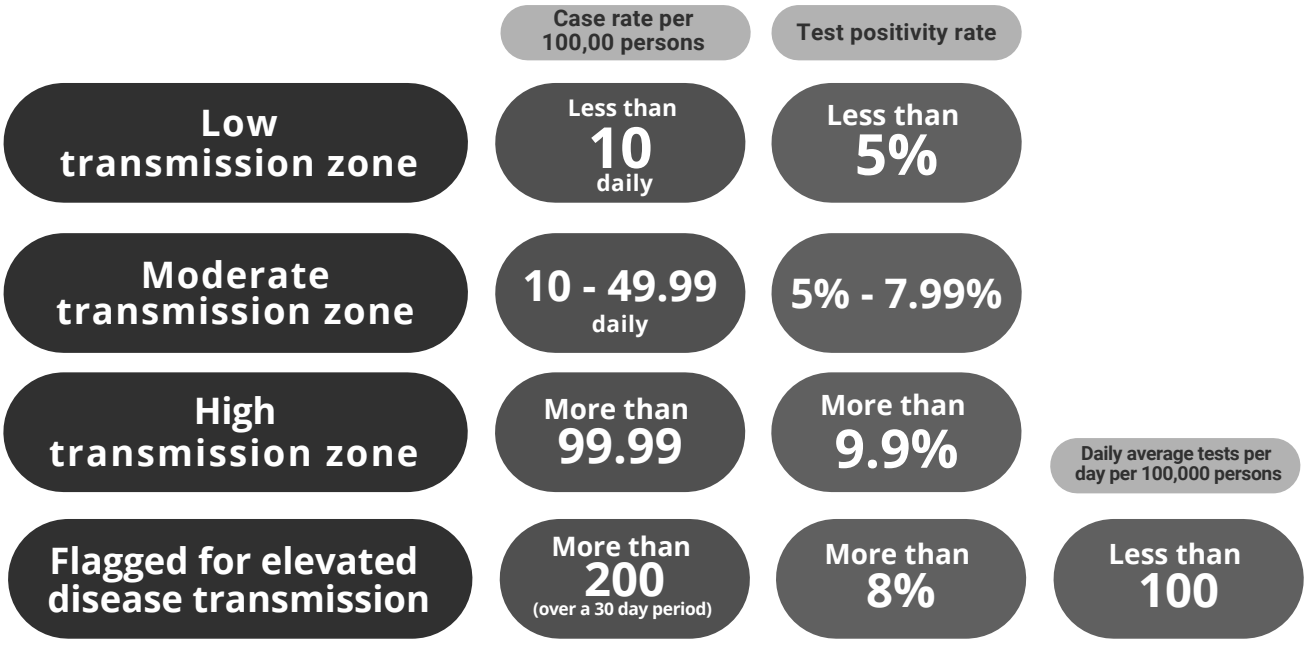



**Table 1 :** Criteria for tansmission zones

# Nevada: County Overview for Disease Transmission

There is a minimum seven day delay in collecting granular United States county-wise transmission data. Therefore, we have analysed Nevada as a whole because those statistics are published relatively faster than the granular county-level data. Given that 73% of the residents of Nevada live in Clark Country and that we had previously seen a resembling trend between the state and the county, it made sense to simply use the available daily data of the State of Nevada.

As of 2nd December 2021, 7 out of the 17 Counties, including Clark County, have only met one of the above criteria: more than 200 case rates per 100,000 persons, whilst the remaining counties have met 2 or more criteria and have been flagged for elevated disease transmission.

| | Nevada | Clark County |
|---|---|---|
| Daily case rate per 100,000 persons | 21 | 19 |
| Test positivity rate | 7.3% | 6.8% |
| Daily average tests per 100,000 persons | 285 | 294 |

**Table 2 :** Current Covid-19 rates of Nevada vs. Clark County

We analyzed the daily data of Nevada through our regression models to predict future test positivity rates to examine if or not these rates fall below the 5% safety threshold as recommended by the WHO, thereby travelers can be well aware of the situation so they know what guidelines to follow and expect, which will help make their decisions as to if and when to take a trip or promote a trip to Vegas.

# Methodology

For our predictive analysis on the transmission rates of Nevada up until end January 2022, we commenced our research with the most common, Linear Polynomial Regression. For model comparisons and hoping to further improve on accuracy by reducing any errors or overfitting caused by multicollinearity, we considered Ridge Regression as well. The dataset was split into a 70% training set and 30% testing set and we ran a 10-fold cross validation so the models can better identify trends and patterns of the data provided.

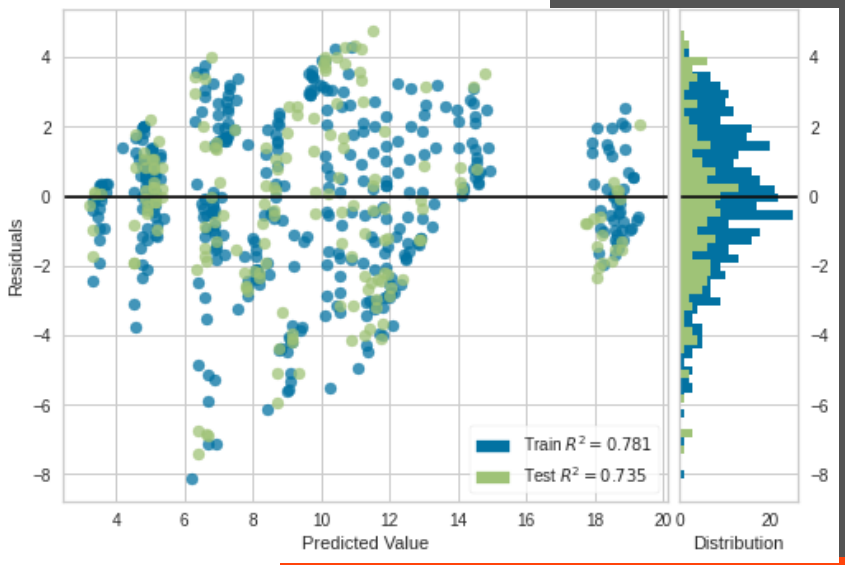



**Figure 1 :** Residual Plot for Linear Regression model

However , as illustrated in **Figure 1**, by analyzing the R-squared values we see that only 78.1% and 73.5% of the variance of the training set and testing set respectively, are explained by the linear regression model. And similarly, for ridge regression, less than 80% of the dataset is being explained by regression model.

Due to the low accuracy of these two models, we carried on our research to find a more accurate model, which led us to an easy to use, low code machine learning library: PyCaret.

By considering the test positivity rate as our dependent variable and testing how it changes over time, the PyCaret library was used to run our data set across all regression models and rate them according to their model scores: in this case we looked at the mean absolute error score (MAE) and the R-Squared score. We then went ahead with the top 3 regression models picked for us: the Random Forest Regressor, the K Nearest Regressor and the Gradient Boosting Regressor. After choosing the models, we got each model to plot the data over the actual data plot to visually see the performance of the model. Then we continued to use the model to predict data up until 31st January 2022.

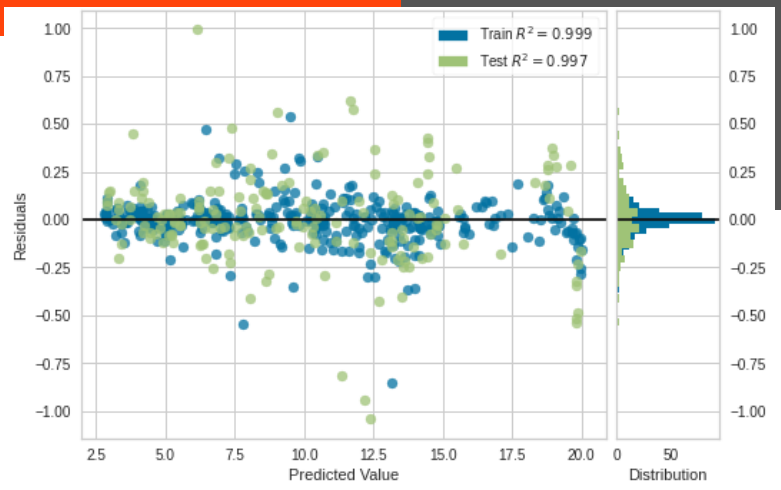



**Figure 2 :** Residual Plot for Random Forest Regression model

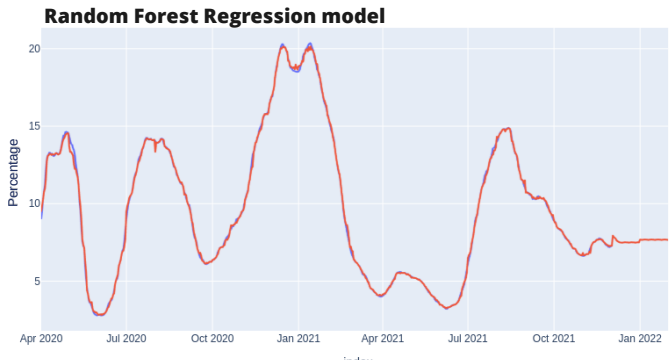
**Figure 2**: From the Random Forest Regression model 99.9% of the training data and 99.7% of the testing data fit our regression model, proving high accuracy in the model predicting data since the model explains almost all changes in the test positivity rates.

All three models were able to fit more than 95% of the Nevada daily test positivity rates accurately. By running all three of these models, besides the quantitative comparisons calculated, we were also able to visualize the difference and compare the models through the fitted model line and prediction line.

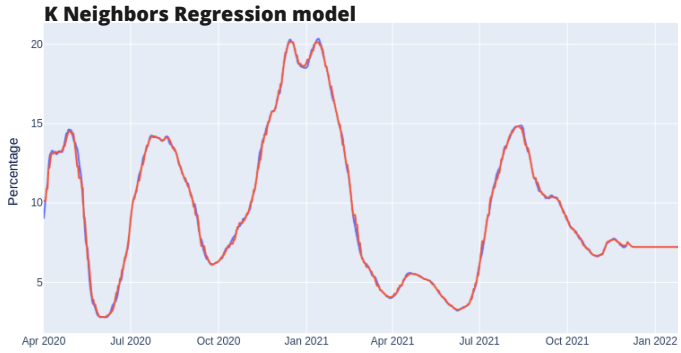
In this paper we analyse the results from the three final regressors to discuss their prediction accuracies.

# Results and Discussion

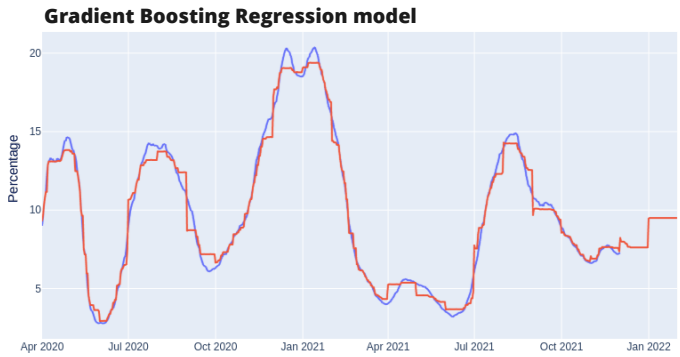
## Nevada Transmission Rates




**A** The Random Forest Regression model has predicted the daily transmission rates from April 2020 - 31st January 2022 (red line) on top of the actual daily transmission rates (blue line) which has actual data up until 2nd December 2021




**B** The K-Neighbors Regression model has predicted the daily transmission rates from April 2020 - 31st January 2022 (red line) on top of the actual daily transmission rates (blue line) which has actual data up until 2nd December 2021




**C** The Gradient Boosting Regression model has predicted the daily transmission rates from April 2020 - 31st January 2022 (red line) on top of the actual daily transmission rates (blue line) which has actual data up until 2nd December 2021

**Figure 3 :** Regression model predictions of Nevada transmission rates (percentage) - top 3 regressors

**Figure 3** : Visually, we notice that the Gradient Boosting Regression model's prediction line is not as accurate as the other two. The K Neighbors Regressor and the Random Forest Regressor show a better model fit as the predicted values (red line) fit the actual values (blue line) well with only a small error. This explains the similar trend in the forecasted values.

## Nevada Transmission Rates ( April 2020 - January 2022)

**Random Forest Regression Model**

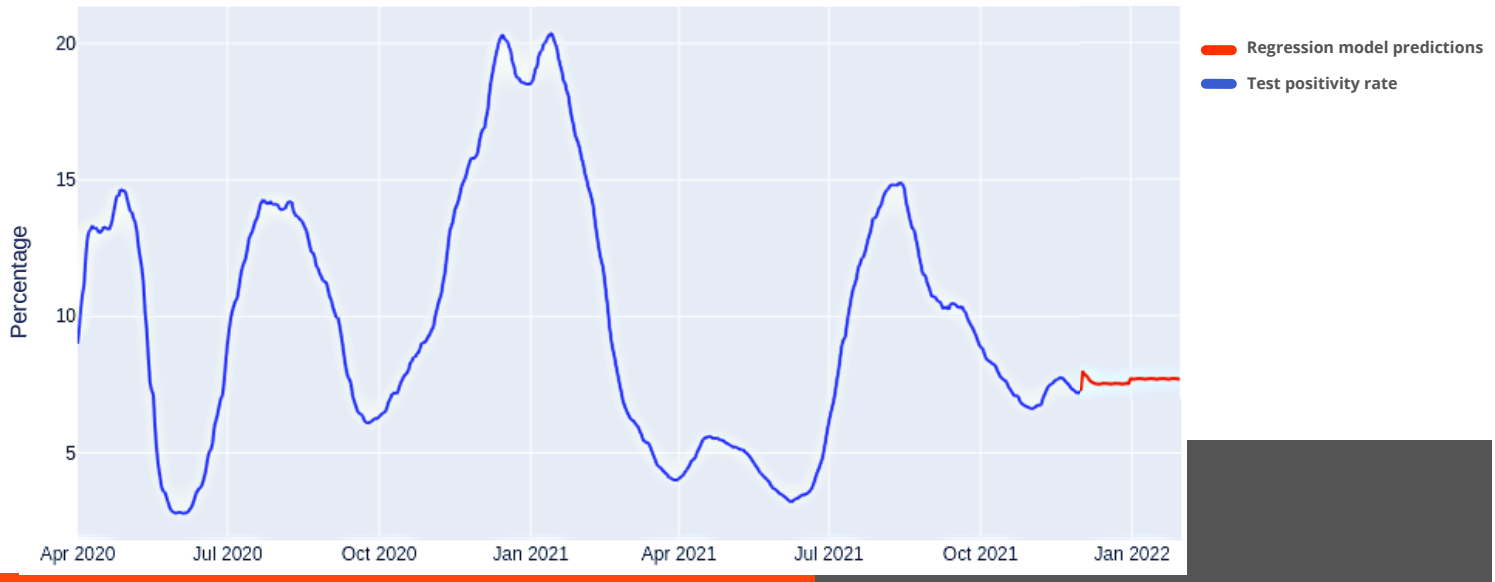



**Figure 4 :** Nevada Transmission rates (percentage) with predictions

**Figure 4:** As predicted by the model, the transmission rates of Nevada will not fall below the 5% safety threshold recommended by the WHO, by 31st January 2022. We then assume that Clark County would follow a similar trend (due to 73% of the population of Nevada being concentrated in Clark County). This implies that since Clark County test positivity rates are currently at a 6.8% and will not drop below the 5% safety threshold by 31st January 2022, that the County will continue to follow the CDC recommended guidelines, that everyone, including fully vaccinated individuals, must wear masks in public indoor settings and crowded outdoor settings .

Further, to analyze the stabilized trend that start with a small climb, the model has examined the existing data and the rise in test positivity rates from October 2020 – December 2020 (the first detection of the B.1.1.7 UK Variant was in December), and predicted a similar trend (rise) this year as well. This could be due to people getting out and about these days or the season bringing in crowds into specific areas (shopping areas, coffee shops etc.)

Our finding is that the Random Forest Regression model provided us with a prediction for the transmission rates of Nevada, with a 0.9508 R-squared score which means that the model represents 95.08% of the variation in the test positivity rates. As our model looks at previous Covid-19 transmission rates and predicts how they will perform in the future, given that Nevada is not hit unexpectedly by a new/worse variant, and the State's ability and resources to perform tests or obtain results do not change, then our predictions up until 31st January 2022 should follow.

To further our predictions we ran our regression models up until the 31st of July 2022. The Random Forest and K Neighbors Regression models show that the transmission rates of Nevada will still not fall below the 5% mark by this time. However, the **Gradient Boosting Regression model shows a dip below the 5% mark from June 1st to June 30th 2022**. Given that the Gradient Boosting Regressor has a 0.625 MAE score and an approximate 95% R-squared score, Nevada could possibly face a removal of the mask mandate during the month of June -- subject to no new variants that could affect the State, causing the effectiveness of the Vaccinations to decrease and thereby increase the number of positive tests, or any other factors that may affect the transmission rates of the County.

# Conclusion

Our intention was to find out when Clark County (in this case Nevada data) transmission rates will fall below the 5% safety threshold recommended by the WHO. We predict this will not happen by the end January 2022, but according to our Gradient Boosting Regression Model this mark will be achieved in the month of June 2022 following another rise above the 5% threshold. Our limitations when predicting these results is that we have only considered one variable: the transmission rates instead of comparing them with other variables like Community Movement in the area, which greatly affect the spread of the virus. However we will be running our models at least 3 times in the future to confirm accuracy of the model and the PyCaret library, and also to analyse how well our predictions have done once further actual data has been released.