U.S. Covid-19 Vaccinations Analysis at County Level

Department of Computer Science, Stony Brook

1 Abstract

The main aim of our project is to study the COVID vaccination data in the US at county level and provide insights as to what factors affect the vaccination. We also aim to provide suggestions about how the vaccination rate can be improved.

2 Introduction

The Covid-19 pandemic has affected the entire world in an unforeseen manner and the only way to overcome its effects is to have a global and safe implementation of a vaccination program that has broad clinical and socioeconomic benefits. There have been multiple studies which have shown that getting vaccinated, regardless of whether you already had Covid-19, provides you better protection. But recently, the vaccination rates have dipped and the infection rates have surged. So, we wanted to study the factors that affected the vaccination rates in the U.S. and studying that at a county level will give us granular insights. This can be achieved by understanding how certain factors like economic, social, cultural and scientific, affect the vaccination rates of these counties. Apart from the vaccination alone, we also wanted to study if the vaccination is actually causing cases and deaths to reduce.

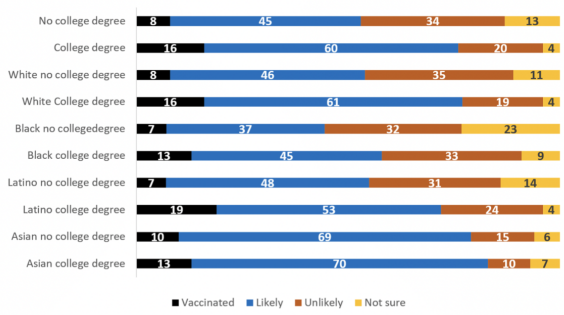
3 Background

We looked at some already existing visualisations [1] to see related county-level analysis. We understood that the vaccination rates are diverse across counties and we can actually analyze the factors that affect this. One such example [2] was that of higher poverty rates which caused lack of insurance coverage and other hurdles in accessing the vaccines. This gave us an initial understanding of the various factors that can be taken into consideration for our study. Moreover, based on the prioritization given for the early access to vaccines for healthcare workers, long term residents, older adults and people with different medical conditions, counties with higher shares of people aged 65 or older have higher vaccination rates(31.4%) than those of aged 65 below (29.8%). Counties with people having high-risk medical conditions are seen to have lower vaccination average compared to healthier people. Next, we found many studies analyzing personal factors that influence vaccination rates. Some of these factors are listed below:

1. Patriotism: A survey [3] showed that there is almost 10% decrease in acceptance of a vaccine that did not originate in the U.S. vs a vaccine developed in the U.S.

2. Education: A more educated person will make an informed decision (based on efficacy, location of vaccine development, and risk of adverse effects, etc.) while an uneducated person will go by word of mouth and/or intuition. A survey [4] showed that a college degree is associated with a 43% increase in willingness to get vaccinated. A detailed view of how education is related to vaccination can be seen in Figure 1 below.

Figure 1: Relation of education level with vaccination acceptance



3. Risk perceptions: The way a person understands the fatal effects of Covid-19 is also associated with vaccine uptake. A survey [5]

showed that people who relied on "conservative" news outlets, Republicans, and who had low confidence in scientists are least likely to vaccinate themselves or children.

4. Race: A survey [6] showed that white people account for the largest share of people who remain unvaccinated, but, overall, Black and Hispanic people are less likely than their White counterparts to have received a vaccine, leaving them at increased risk.

Finally, we wanted to research how vaccination rates affect the infection rates. On looking at data we observed that vaccines are not 100% effective. As the number of fully vaccinated people goes up, the number of vaccine breakthrough infections also increases. Vaccine breakthrough infections are monitored by COVID-NET system [7] which is a population based surveillance system that collects reports of lab-confirmed COVID-19 and the related hospitalization. Delta variant is one such breakthrough infection. However, the risk of infection remains much higher for unvaccinated people. We also observed that vaccinated people are less likely to be infected and less likely to experience hospitalization or death. A study [8] also showed that the community transmission of COVID-19 at county-level which are considered as 'high' transmission levels have the vaccination rate of 30.5% compared to those counties with 'low' community transmission levels of 27.2

4 Experimental setup

Our approach was to first collect the data, then do analysis on the data for different factors and finally form a model which can be used to predict the vaccination rates of counties in the future.

4.1 Data

We have considered using data from the following sources -

1. This is our primary data source which gives us the vaccination information at county level.

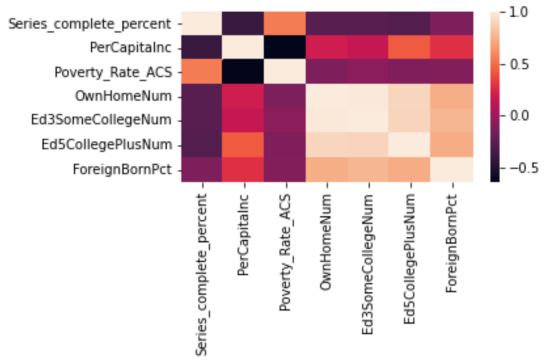
2. This dataset gives us information like population, age, ace, immigrants, etc. at county level.

3. This is the data source for hesitancy information regarding vaccination.

4.2 Exploratory data analysis

We did multiple explorations to see how different factors are related to the vaccination rate. Initially, we looked at the correlation of different features with the vaccination rate. The heatmap can be seen in Figure 2 below.

Figure 2: Spearman correlation of different features with vaccination rates



As can be seen from the heatmap, some features have a strong positive correlation with vaccination rate, some features have a strong negative correlation with the vaccination rate while some have no correlation. We then picked certain features to individual analysis which we describe below:

1. **Age vs Vaccination Rates** : We plotted vaccination rates over time for different age groups to see if the age group affects the vaccination rate. It can be seen from Figure 3 that *older people tend to have a larger vaccination rate*. This can also be due to older people getting vaccinated first.

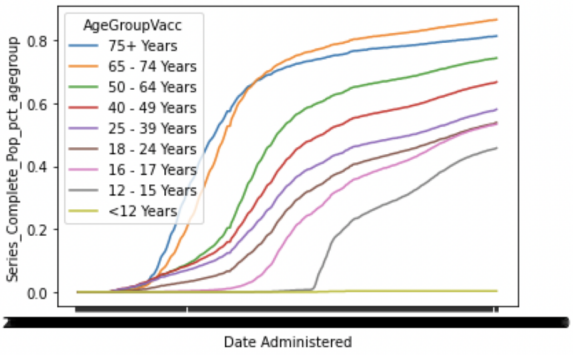


Figure 3: Age vs Vaccination rates

2. **Poverty rates vs Vaccination rates** : We first found that the maximum poverty rate across all counties was 55%. So, we binned the poverty rate into 6 bins (of 10 size/bin) and analyzed the vaccination trend according to the poverty rates. We found that the *vaccination rate is inversely correlated to the poverty rate*. This seems to be in sync with what we know about poverty and availability of health care facilities.

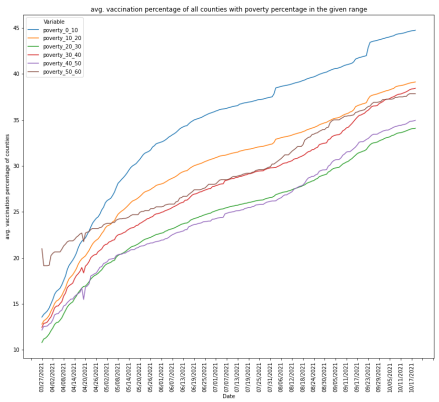


Figure 4: Poverty rate vs Vaccination rate

3. **Metro/Non-metro vs Vaccination rates** : We categorized and observed the vaccination trend in "Metro" and "Non-metro" counties, and we can observe that the vaccination rates are lagging behind in "Non-metro" counties. This tells us that *more developed counties have access to better facilities and hence better vaccination rates*. Moreover, this points to one way of improving vaccination rates would be to have non-metro county reaching health camps.

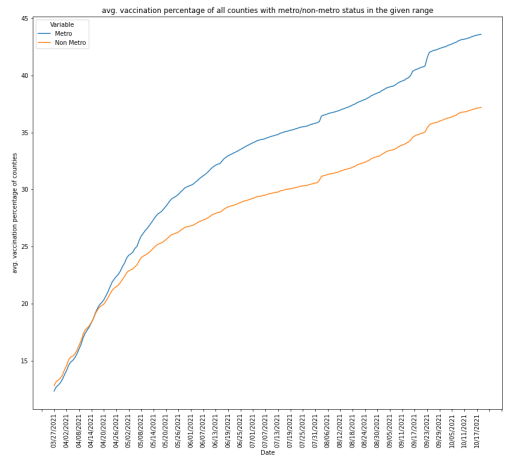


Figure 5: Metro/Non-metro vs Vaccination rate

4. **Vaccine hesitancy vs Vaccination rates** : We also looked at the relationship between vaccine hesitancy and the vaccination rates. Vaccine hesitancy is a measure defined by the SVI index of each county. We observed that the above two are *inversely correlated*. This signifies that we can have programs to spread awareness about the vaccine among more hesitant people and counties.

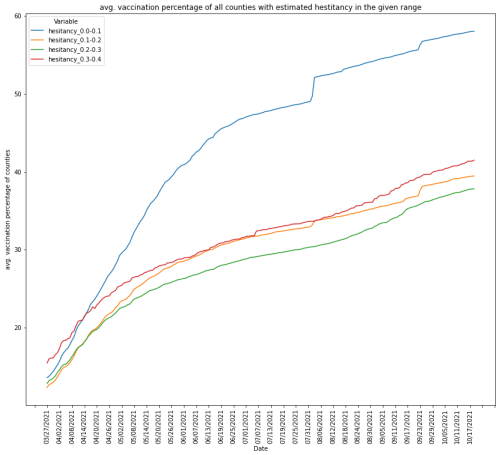


Figure 6: Vaccine hesitancy vs Vaccination rate

5. **Political standing vs Vaccination rates** : We looked at the political standings (Republican/Democratic) for each county. We observed that till a particular time, the vaccination rates were constant for either standings, but it gradually got *separated with republic supporters vaccine rates falling behind democrats supporting county* [9].

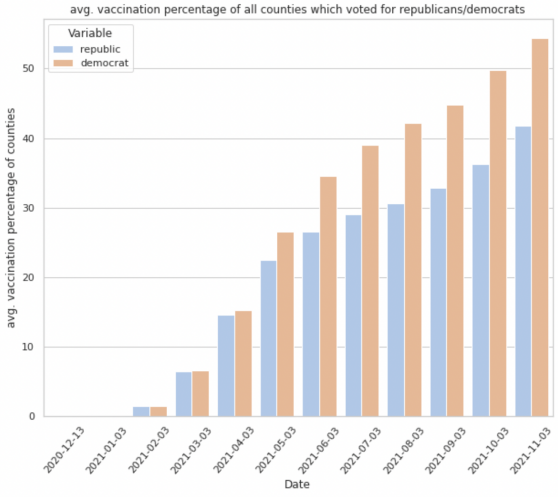


Figure 7: Political standing vs Vaccination rate

6. **Race & Ethnicity vs Vaccination rates** : Considering race and ethnicity, 63% of people had received at least one dose of the vaccine. Where, two thirds were White (61%), 11% were Black, 17% were Hispanic, 6% were Asian, 1% were American Indian or Alaska Native (AIAN), and <1% were other Pacific Islander (NHOPI), while 5% were reported multiple or other race.The share of recent vaccinations going to the Black and Hispanic population is smaller than their share of total people who have received at least one dose.

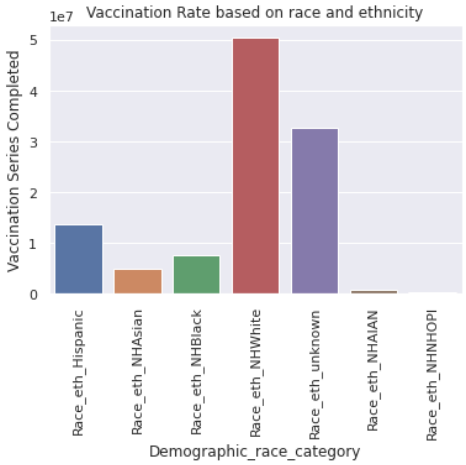


Figure 8: Race and Ethnicity vs Vaccination rate

7. **Gender vs Vaccination rates** : Data on gender differences in vaccination rates veered in an unexpected direction, where a large group of unvaccinated Americans were largely men. Nearly 9.5 million more women than men have been vaccinated in the U.S. where a greater share of women are getting the vaccine. Among older Americans, who had early access to the vaccine, women outnumber men. The U.S. Census Bureau estimate that women make up about 55 percent of all adults age 65 and over and hence are more vaccinated.

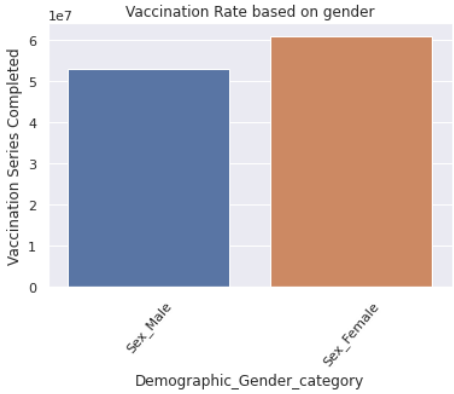


Figure 9: Gender vs Vaccination rate

In this manner, we explored different features that affected the vaccination rates across different counties. We also used this information to suggest certain improvements to the vaccine campaign, which we summarize in the findings section.

5 Model training and prediction

Once we had a gist of which and how different features can be used to predict vaccination rates, we went ahead to try to predict vaccination rates of certain counties. We took two approaches:

5.0.1 Time-series based prediction using Prophet

On exploring how to forecast using an existing time series data we came across the Prophet [10] model created by Facebook which seemed like a decent choice for the following reasons:

1. Producing accurate predictions is not an easy task and requires much work which can be easily done by Prophet with comparable results.

2. Prophet is robust to missing data and shifts in trends and handles outliers as well.

3. Moreover, it has different layers which can be added if we want to account for yearly, weekly effects as well as on relevant factors.

We formed the time series data using the following steps:

1. We took the data as discussed above in Section 4.1 given by CDC as of November first week.

2. This data had a field named "Series_Complete_Pop_Pct" which we used as the value for our time series data. It basically gives the percentage of people who are fully vaccinated in a county.

3. The "Date" field served as the timestamp.

4. Finally, we grouped the above time series data per county.

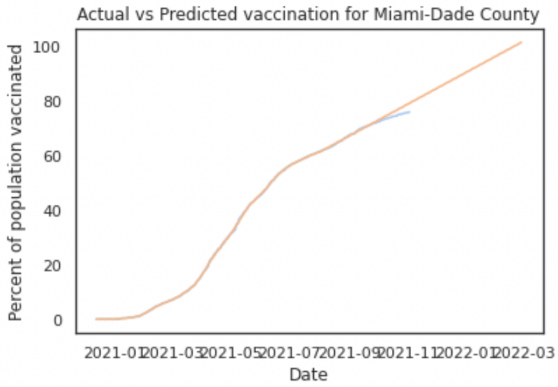
We then used the above data to train the Prophet model and predict vaccination percentages for some of the top counties. The model was trained on data till October 1, 2021 and the predictions were made for the next 150 days or 5 months after that. The dates for which the predicted vaccination

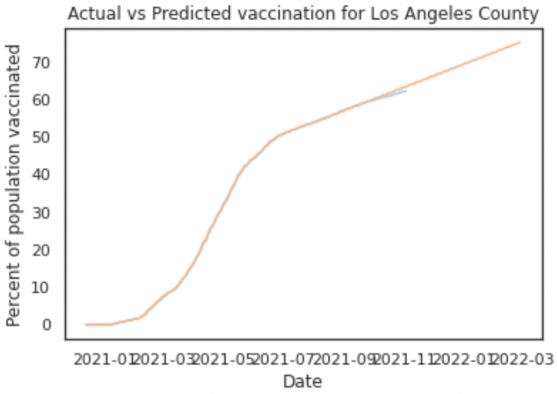
County	Date to reach 80% vaccination
Miami-Dade County	2021-11-10
Queens County	2021-11-29
Santa Clara County	2021-12-23
Broward County	2022-01-28
Los Angeles County	2022-04-24
Riverside County	2022-07-03
San Diego County	2022-08-05
Maricopa County	2022-08-29

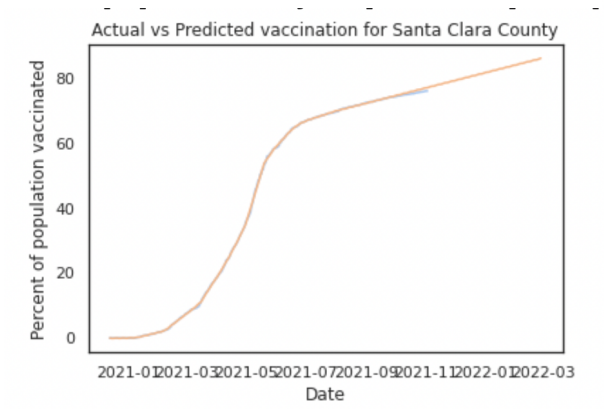
Table 1: Prediction for some of the top counties

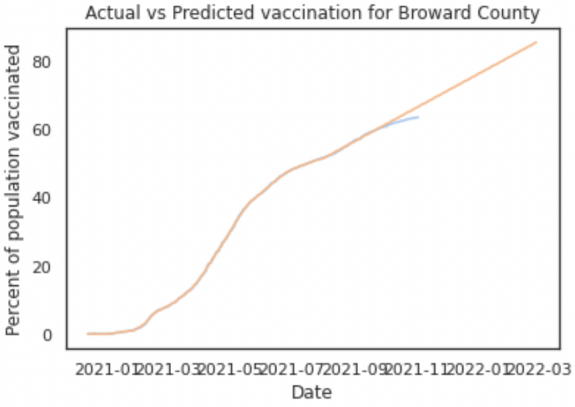
percentage reached 80% is given below for some of the top counties:

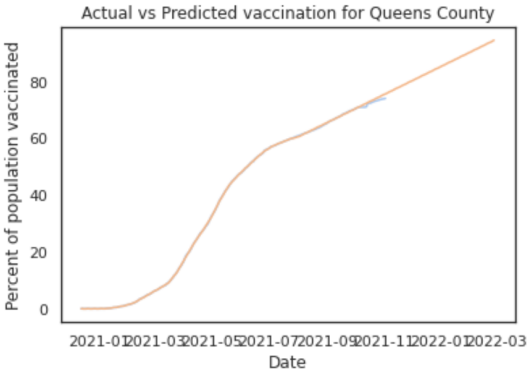
The plots for predicted percentage vs the actual percentage can be seen below for some of the counties which may reach 80% vaccination in the coming days. The blue line represents the actual vaccination percentage whereas the orange line in each plot represents the predicted percentage for the next 150 days (5 months).











5.0.2 Time-series based prediction using LSTM based architecture

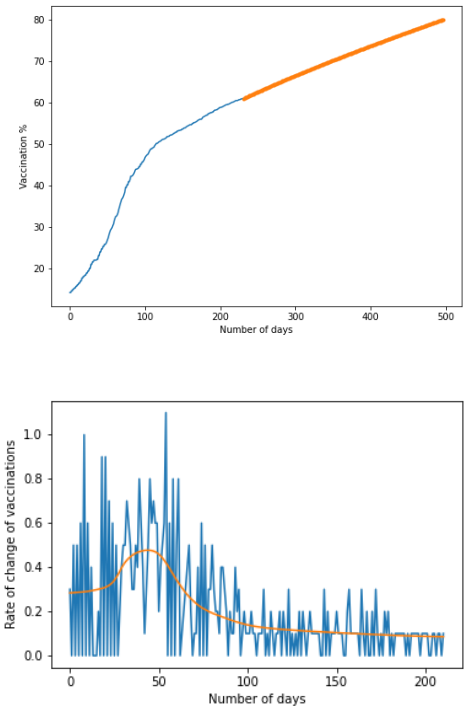
From our analysis, we realized that vaccination rate would depend on previous vaccination rates significantly.

So, we decided to built a simplified model using LSTM, which trains on the timeseries data where the labels are the rate of change between successive entries in the time series. Precisely for a window of length L, the label is Y[i] - Y[i-1] where Y represented the rate of change in the vaccination rates, and the training data we split the entire timeseries data generated by samplying every L samples using a window length of L. And we build this for each county.

The model learns the shape of the change in vaccination rates over time. It gives an insight on approximately how much time it might take to reach a vaccination of over 80% for each county. In the plots, the blue line represents the data we have till now and the orange line shows the forecasted rate for the following days.

Below we showcase our observations for three counties -

Figure 10: Forecast vaccination rates and rate of change ground truth vs predicted data in Los Angeles county



We see the below results for three counties we predicted for:

Figure 11: Forecast vaccination rates and rate of change ground truth vs predicted data in Maricopa county

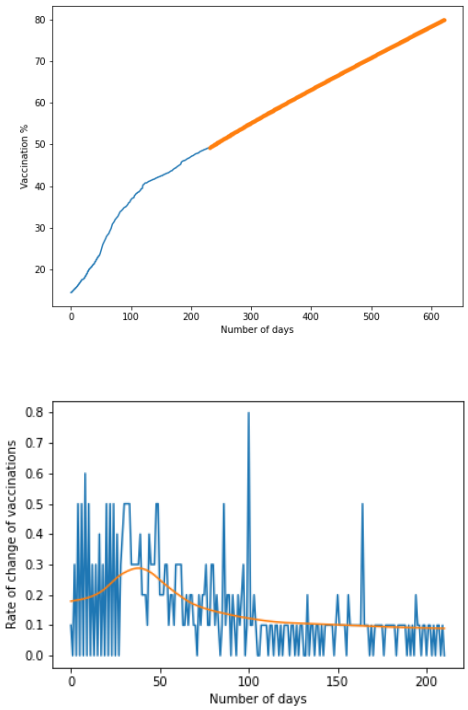
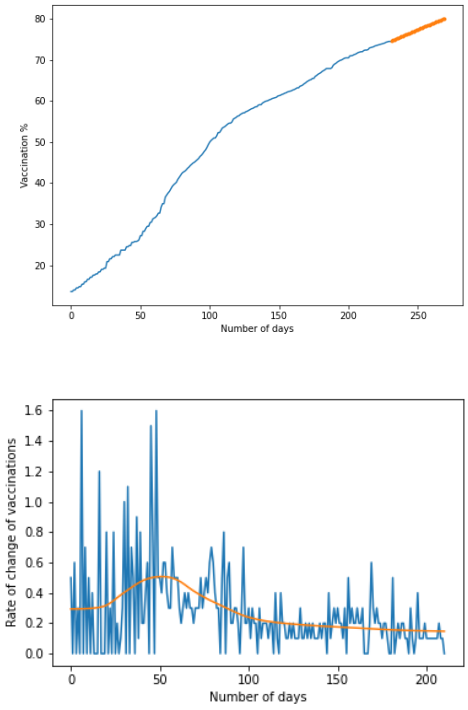


Figure 12: Forecast vaccination rates and rate of change ground truth vs predicted data in Miami-Dade county



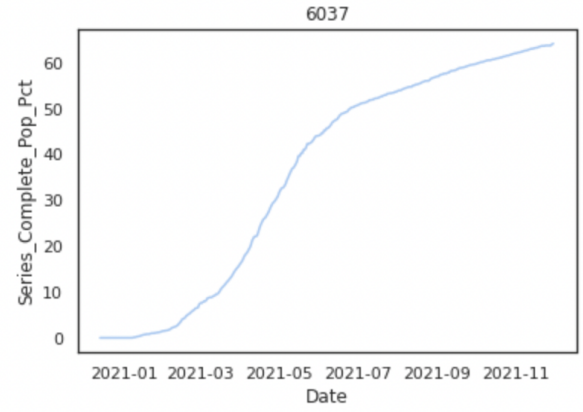
County	Days to reach 80% vaccination
Los Angeles county	270
Miami-Dade county	50
Maricopa county	480

Table 2: Prediction for some of the top counties using LSTM model

6 Vaccination vs Cases and Deaths

We as humans have done a very good job in thinking of, making and distributing the COVID vaccine across the world. Although, the vaccine has only been tested on a small subset of people. So, we also wanted to study if the vaccine is actually effective. We plotted the vaccination rates against the number of cases for some of the top counties. We also plotted the vaccination rates against the deaths. For this, we merged the vaccination data with the COVID cases and deaths data to see the trend. The below figures describe the COVID vaccination and cases trends for Los Angeles county (FIPS 6037).

Figure 13: Trend of vaccination rates in Los Angeles county



It can be observed from the above two plots that as the number of vaccinated people increased, the number of cases grew at a slow pace and and again picked up pace (around 2021-08) once the vaccination rate declined. It can also be observed from the third plot above, that as the vaccinations increased the rate at which people died decreased and the death plot plateaued.

7 Summary of findings

Through this project, we found multiple factors that affect the vaccination rates in the US by doing a county level data analysis. We list below some of our main findings:

Figure 14: Trend of cases in Los Angeles county

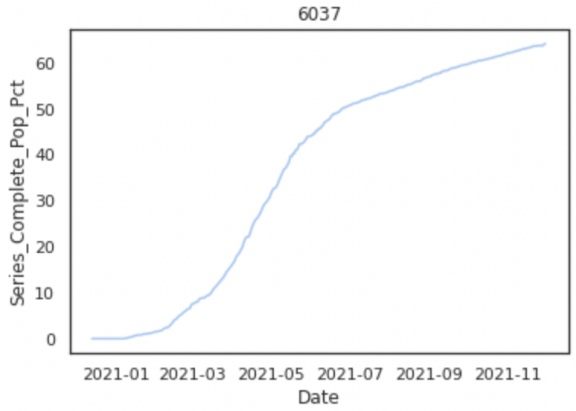
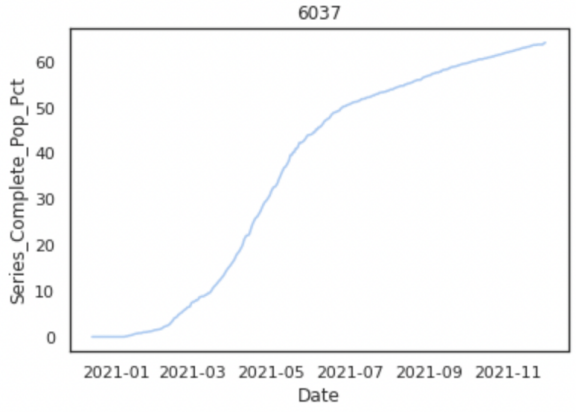


Figure 15: Trend of deaths in Los Angeles county



1. It is important to spread the vaccination campaign to poorer counties where the health facilities may not be very good.

2. It is also essential to spread awareness about the vaccine in less educated areas and areas where the population is hesitant to take the vaccine.

3. Even if not fully effective, the vaccine is our best bet currently and data shows that it has reduced infection rate and deaths to a certain extent.

4. Changing the views of those who are highly sceptical about vaccines is very difficult to achieve, hence by focusing on giving positive and accurate information about vaccines and building people's resilience to false information is important.

5. It is also important for the local health departments to work on providing food and other services at vaccination drives to help accelerate the vaccination process. This will also help in creating a supportive environment for community engagement.

6. We use LSTM and Prophet based algorithms to predict the time required to reach 80% vaccination.

8 References

1. https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-total-admin-rate-total/

2. https://www.kff.org/coronavirus-covid-19/issue-brief/vaccination-is-local-covid-19-vaccination-rates-vary-by-county-and-key-characteristics/

3. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2771872

4. https://healthpolicy.usc.edu/evidence-base/education-is-now-a-bigger-factor-than-race-in-desire-for-covid-19-vaccine/

5. https://bmcpublichealth.biomedcentral.com/track/pdf/10.1186/s12889-021-10862-1.pdf

6. https://www.kff.org/coronavirus-covid-19/issue-brief/latest-data-on-covid-19-vaccinations-by-race-ethnicity/

7. https://www.cdc.gov/mmwr/volumes/69/wr/mm695152e2.htm

8. https://www.cdc.gov/mmwr/volumes/70/wr/mm7032e3.htm

9. https://www.kff.org/policy-watch/the-red-blue-divide-in-covid-19-vaccination-rates/

10. https://towardsdatascience.com/a-quick-start-of-time-series-forecasting-with-a-practical-example-using-fb-prophet-31c4447a2274