

COVID-19 PROJECT



**Hannah**

Tableau Dashboard  
Designer & Project  
Report speedy  
typewriter



**Dikshant**

Expert data cleaner &  
Twitter sentiment  
analyst



**Lisa**

Cheerful Motivator &  
Project Manager



**Rajashekar**

Twitter scrapping  
solutions specialist &  
Power BI  
visualizations  
generator

## THE TEAM



# AGENDA

## **1. Twitter Scraping**

- Results and process
- Replicating the report

## **2. Dashboard presentation**

- Global Vaccine Distribution dashboard
- US State Vaccination Progress dashboard
- US Vaccine Hesitancy dashboard

## **3. Data Cleaning and Sources Overview**

## **4. Summary of findings**

## **5. Challenges & Current State of Project**

# TWITTER SCRAPPING

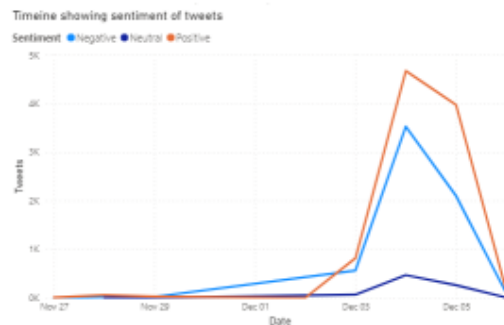
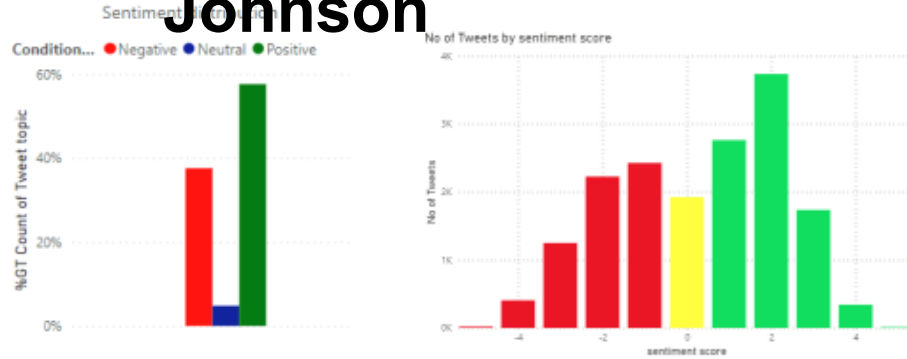


186,452 Tweets Scraped - 123,915 After Sentiment Analysis| Nov 27<sup>th</sup> - Dec 6<sup>th</sup>

Vaccine	Keywords	Total Collected Tweets	Total Cleaned Tweets <small>*after sentiment analysis</small>
<b>Pfizer</b>	pfizer, Pfizer-BioNTech, BioNTechpfizer	10614	7611
<b>Moderna</b>	Moderna, moderna_tx, Moderna-NIAID, NIAID, NIAID-Moderna	8373	6111
<b>Johnson &amp; Johnson</b>	Johnson & Johnson, Johnson and Johnson, Janssen, Janssen Pharmaceutical, J&J	27798	16829
<b>Oxford-AstraZeneca</b>	OXFORDVACCINE, Oxford-Astraeneca, OxfordAstraZeneca, AstraZeneca, Vaxzevria, Covishield	1738	1145
<b>SputnikV</b>	Sputnik V, sputnikv, sputnikvaccine	61	38
<b>Covaxin</b>	covaxin, BharatBiotech	310	205
<b>Sinovac</b>	coronavac, sinovac	865	65

Tweet Topic	Keywords	Total Collected Tweets	Total Cleaned Tweets <small>*after sentiment analysis</small>
<b>Hygiene</b>	hand sanitizer, sanitizer, wash hands, wash face, soap, soap water, hand soap, sanitize	23258	14139
<b>Wear Mask</b>	mask, wearamask, masking, N95, face cover, face covering, face covered, mouth cover, mouth covering, mouth covered, nose cover, nose covering, nose covered, cover your face, coveryourface	47438	34965
<b>Travel</b>	travel, outing, camping, air-travel	27661	16593
<b>Social Distancing</b>	social distancing, physical distancing, 6 feet, social distance, physical distance	15474	9350
<b>Social Gathering</b>	social gathering, gathering, party, restaurant	24617	16864

# Johnson & Johnson



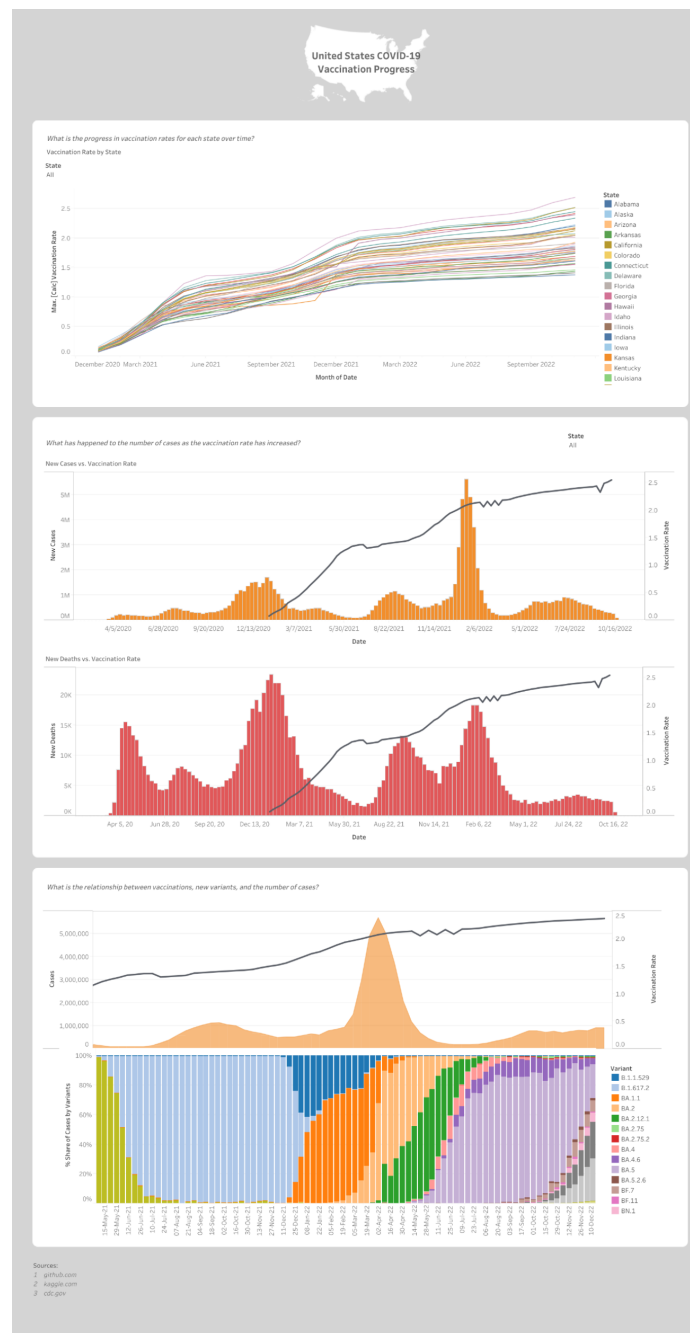
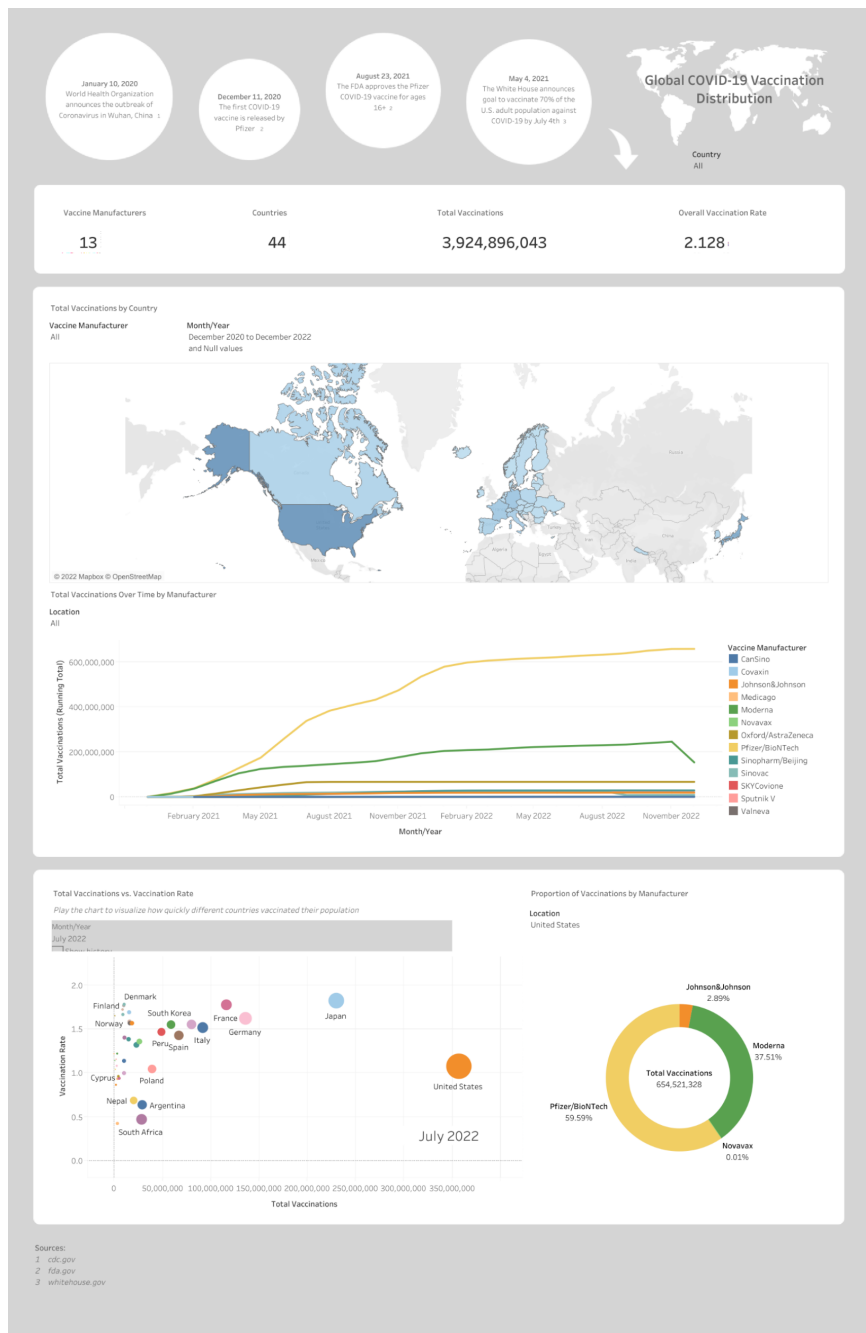
## REPLICATING THE REPORT

### Differences in Methodology

- Using R
- 186,452 tweets vs. 1.2 million tweets
- 11-day collection range vs. 5 weeks
- afinn sentiment analysis package
- No time series forecasting models



# DASHBOARDS



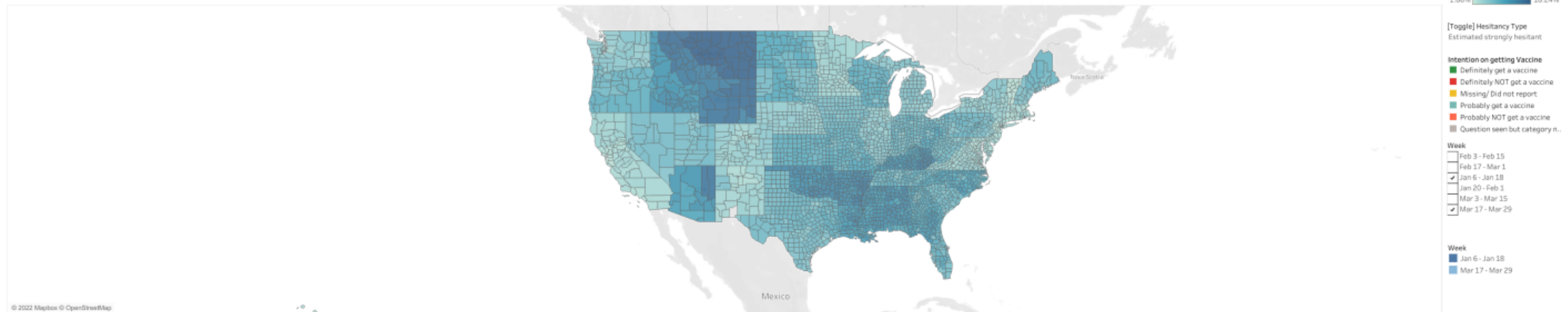
# COVID-19 Vaccination, Cases

1. What has the diffusion of vaccines across the world looked like?
2. What is the progress in vaccination rates for each state over time?
3. What has happened to the number of cases as the vaccination rate has increased?
4. What is the relationship between vaccinations, new variants, and the number of cases?

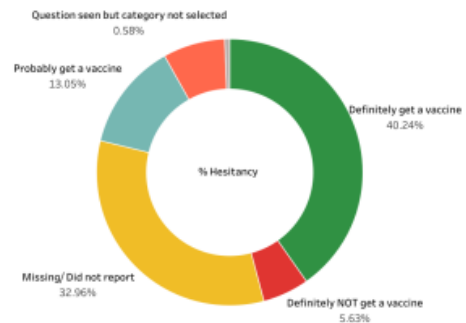
# U.S Vaccination Hesitancy Analysis

5. How has COVID-19 vaccine hesitancy evolved across the US from early 2021?

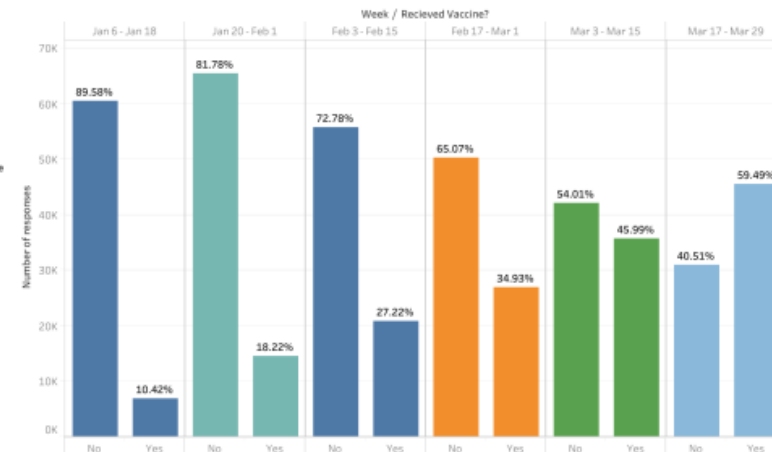
Covid 19 Hesitancy across the US



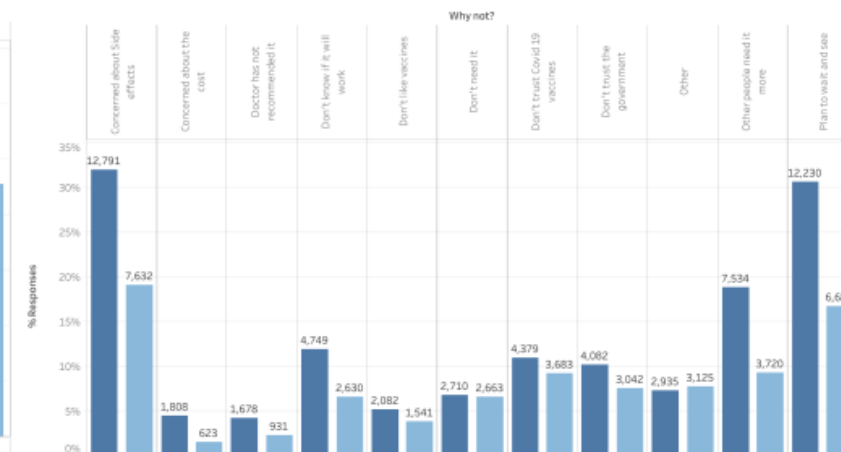
Hesitancy Rate



Received Vaccines?



Why not vaccines?



Source: Housing Pulse Survey & CDC



# DATA CLEANING & SOURCES

## Converting running total into Daily

- Successful for daily vaccinations by manufacturer
- Unsuccessful for Daily cases and deaths. (Tried Python, excel, and R)
  - Problem with a large file and running total grouped by multiple fields.
- Found a new dataset for daily deaths, cleaned and combined it with the given original dataset.

## Cleaning the Hesitancy data

- Used data dictionary and converted encoded columns into corresponding texts.

## Cleaned Twitter data

- Removed punctuations, hashtags, unwanted space, etc.

## Additional Data Sources

- **CDC**
  - Daily deaths and cases
  - Vaccine Hesitancy level by state and county
- **Housing Pulse Survey**
  - Hesitancy Analysis & reasons for hesitation
- **Kaggle**
  - Relationship between Vaccinations and Variants



# CHALLENGES

---

Calculating daily values from running totals

---

Identifying datasets for hesitancy data

---

Replicating report methodology

---

Creating donut charts :)



# SUMMARY OF FINDINGS

- Differences between report and our analysis – *why?*
- Future work / revised methodology
- COVID-19 data analysis comparable to global and national estimates

**CURRENT STATE OF PROJECT: Almost complete**

**QUESTIONS?**