VAX-CHI-NATION

Made by : Josiah Chung, Kshitij Mittal, Danil Meresenschi, Yash Verma

Date : 12/7/2022

Overview

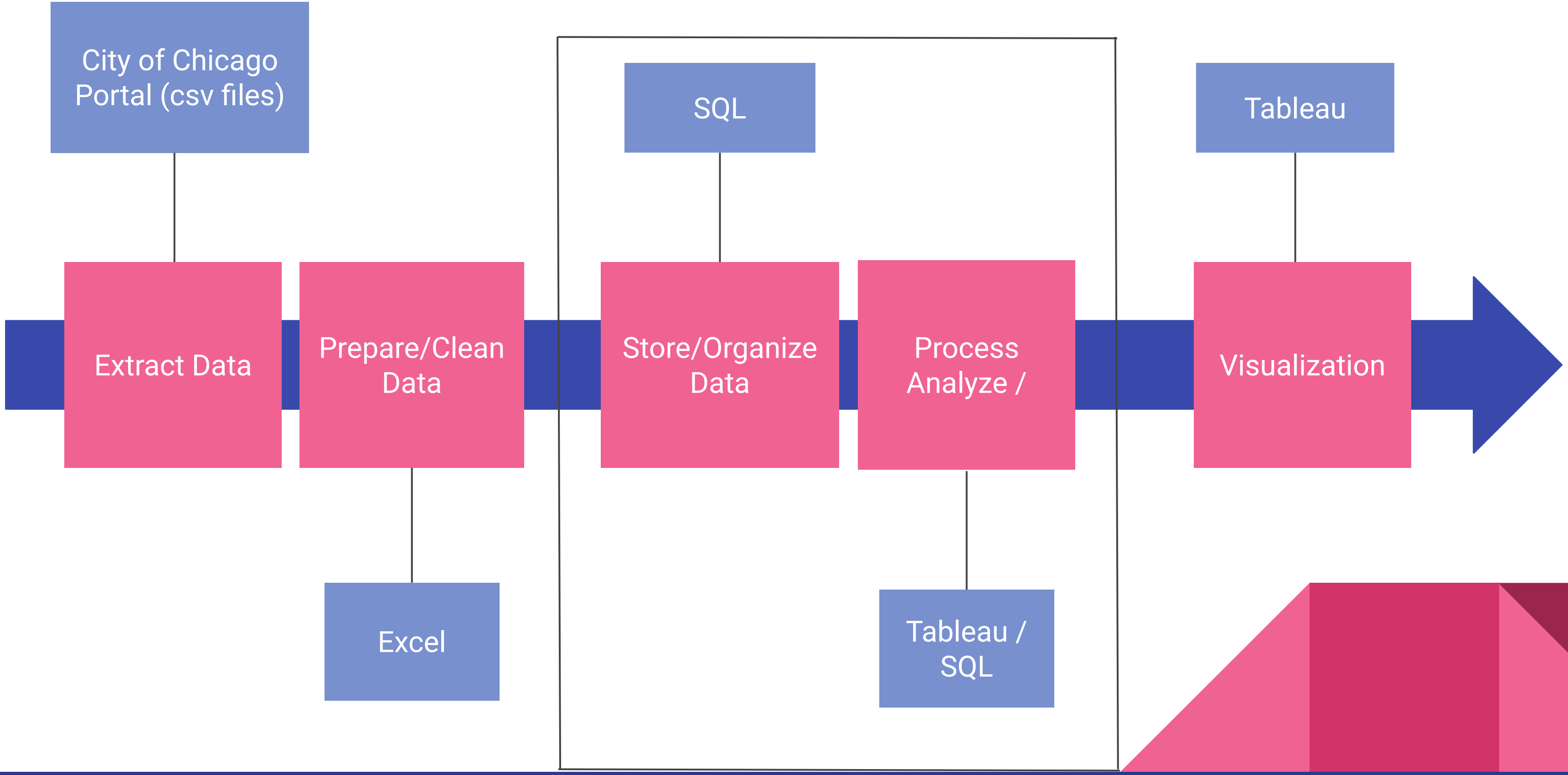
- Introduction
- Staging Process
- Assumptions & Limitations
- Business Questions
- Visualizations
- Intended vs Actual Outcomes
- Recommendations
- Lessons Learned

Introduction

The city of Chicago has a robust vaccination program named - Vax-Chi-Nation. Our aim is to understand if this investment has yielded significant improvements in the public health scenario in Chicago. With a focus on descriptive statistics, we will analyze the vaccination rates and other relevant COVID-19 data metrics and visualize our findings on a dashboard.

To measure the success of our campaign, we will compare our findings and measurements to the US National average and/or figures from similar cities in the Midwest. By analyzing our own vaccination efforts and comparing them to others, we will evaluate what methodologies we believe have worked and what areas we can further optimize and improve for future campaigns.

Date Pipeline Tech Stack



Pre-Processing. Cleaning

Removed redundant columns to improve interpretability.

community_area_or_zip_code	community_area_name	ccvi_score	ccvi_category
70	Ashburn	45.1	MEDIUM
60625		25.5	LOW
1	Rogers Park	30.9	LOW
60612		31.7	MEDIUM

ccvi_score	ccvi_category	rank_socioeconomic_status
45.1	MEDIUM	34
25.5	LOW	
30.9	LOW	32
31.7	MEDIUM	

ccvi_id	zipcode_id	ccvi_score	ccvi_category
1	4	25.5	LOW
2	5	31.7	MEDIUM
3	6	36.1	MEDIUM
4	7	24.6	LOW
5	8	62.5	HIGH
6	9	34.9	MEDIUM
7	10	47	MEDIUM
8	11	8.3	LOW

- Remove: Community Zip Code & Name
- Remove: Ranks

Narrowed the scope of time-series data and aggregated entries by month to create a more comprehensive model.

Zip Code	Date	Total Doses - Daily
60603	12/15/2020	0
60603	12/16/2020	0
60603	12/17/2020	7
60601	2/19/2021	112
60601	2/20/2021	40
60601	2/21/2021	19

date_id	full_date	month	year
1	20-Dec	Decembei	2020
2	21-Jan	January	2021
3	21-Feb	February	2021
4	21-Mar	March	2021
5	21-Apr	April	2021
6	21-May	May	2021
7	21-Jun	June	2021
8	21-Jul	July	2021
9	21-Aug	August	2021
10	21-Sep	Septembe	2021
11	21-Oct	October	2021
12	21-Nov	Novembe	2021
13	21-Dec	Decembei	2021
14	22-Jan	January	2022
15	22-Feb	February	2022
16	22-Mar	March	2022
17	22-Apr	April	2022
18	22-May	May	2022
19	22-Jun	June	2022

- Less missing observations.
- Greater analysis depth across 19 months.

Pre-Processing. Wrangling

Aggregated data on the premise of location and date.

Aggregated several tables into one in accordance with respective dependencies.

week_end	age_group	1st_dose
12/19/2020	18-29	255
12/19/2020	18-29	95
12/19/2020	18-29	1627
12/19/2020	18-29	409

\+

Zip Code	Date	Total Doses - Daily
60603	12/15/2020	0
60603	12/16/2020	0
60603	12/17/2020	7
60601	2/19/2021	112
60601	2/20/2021	40
60601	2/21/2021	19

\+

ZIP Code	Week End	Cases - Weekly	Tests - Weekly
60604	5/23/2020	3	16
60604	5/30/2020	2	20
60604	6/6/2020	1	13
60604	6/13/2020	1	9

vaxchi_id	zipcodeID	date_id	sum_second_dose_18_64	sum_second_dose_65_plus	sum_covid_cases
1	4	1	34	1	166
2	4	2	386	38	145
3	4	3	231	332	32
4	4	4	383	225	87

Normalized the data tables

address
1713 S Ashland Ave Chicago, IL 60608
1645 A West School St Chicago, IL 60657
4326 W Montrose Ave Chicago, IL 60641
9718 S Halsted St, Chicago, IL 60628

address_test	city	state	zipcode_id
1713 S Ashland Ave	Chicago	IL	11
1645 A West School St	Chicago	IL	56
4326 W Montrose Ave	Chicago	IL	42
9718 S Halsted St,	Chicago	IL	30

- The address column is partitioned into respective columns.

Inserted surrogate keys to index transformed data sets.

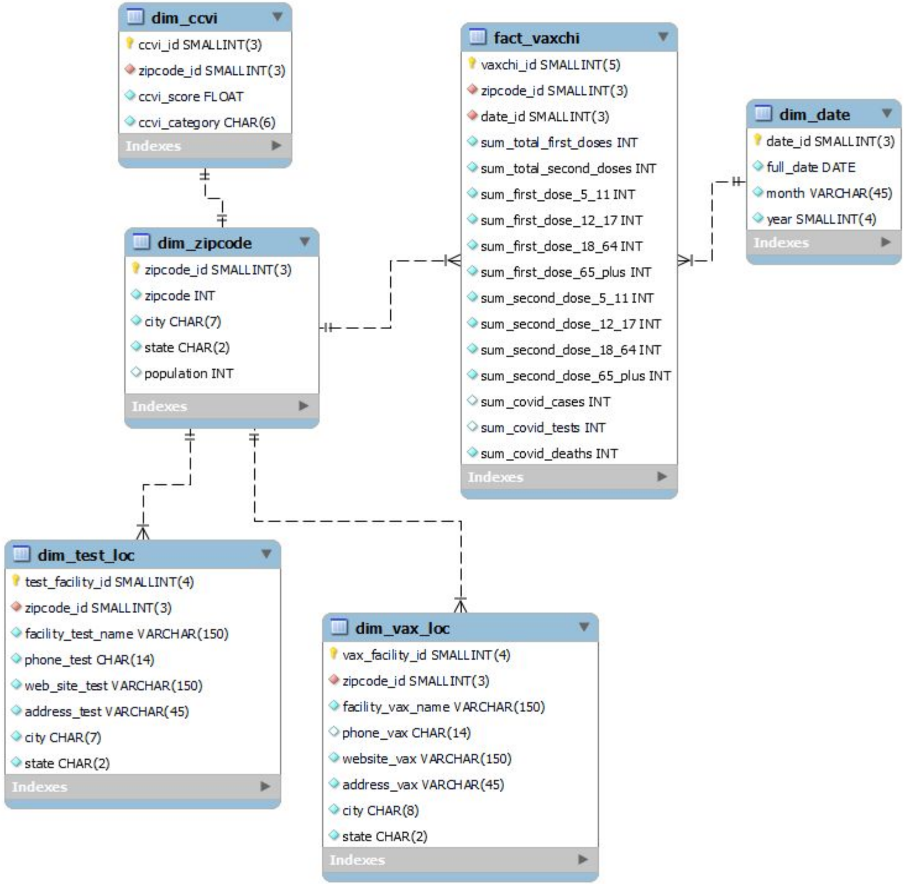
facility_id	facility_name
1	ACCESS Center for Discovery and Learning
16	ACCESS Madison
25	ACHN at Jorge Prieto Health Center
26	Alivio - Western Avenue

vax_facility_id	facility_name_vax
1	ACCESS Center for Discovery and Learning
2	ACCESS Madison
3	ACHN at Jorge Prieto Health Center
4	Alivio - Western Avenue

- Adjusted the numeration of primary keys.

Data Model

- dim_ccvi - community vulnerability index, depicts the severeness of the pandemic's impact in zip codes.
- dim_zipcode - the store of all zip codes, a lot of analysis is done on the premise of location.
  - Population can be nullable since not all zip codes had a record of populations in the datasets.
  - City is fixed char, since there are only 7 consistent characters(Chicago)
  - State is fixed char, since there are only 2 consistent characters(IL)
- dim_test_loc - the store of all test locations
- dim_vax_loc - the store of all vaccination locations
  - Phone can be nullable, since some entries had no record of it.
- fact_vaxchi - the store of vaccination, cases tests, and deaths records.
  - Cases and Tests can be nullable, since some zip codes had no record of those attributes.
- dim_date - the store of all dates with the respective normalized columns.



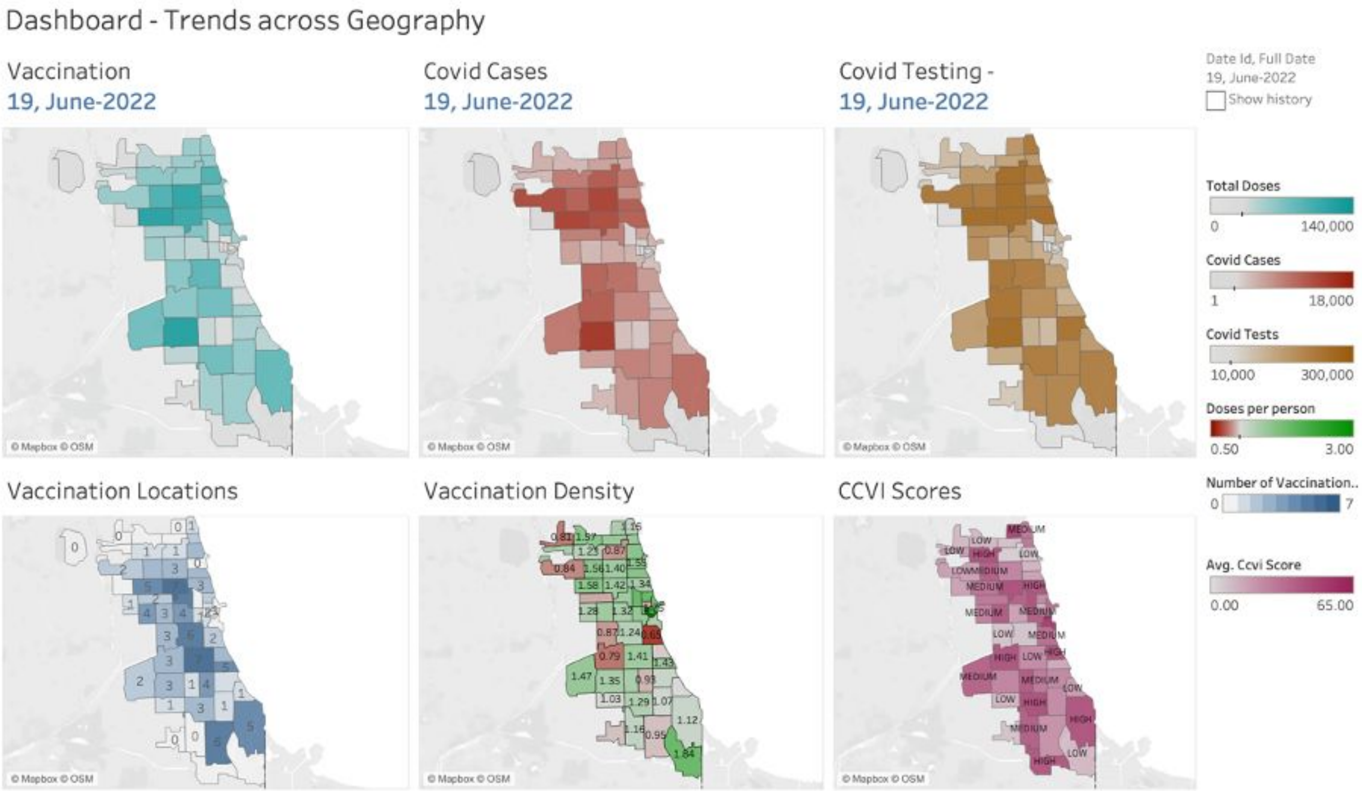
Assumptions and Limitations

- Assumption that vaccination reporting was comprehensively covered by our data
- Assumption that our data on testing and vaccination locations were comprehensive
- Limitation that vaccinations could have been obtained in different zip codes between the first and second doses.
- Limitation of one of our original data sets prevented us from analyzing relationship between race and vaccination rates

Business Questions

- Evaluating vaccination rates based on zip code
- Evaluating vaccination rates based on age group
- Interaction between COVID-19 cases and testing numbers
- Evaluating the effect of the availability of vaccination and testing sites on vaccination rates
- Relationship between CCVI score and vaccination rates
- The interaction between CCVI scores and available vaccination and test sites

Visualizations



Visualizations

SQL Sample 1

```sql
# Which zip code and month had the highest number of total vaccinations(first and second doses) in the 18-64 age category?
SELECT
    fact_vaxchi.zipcode_id,
    fact_vaxchi.date_id,
    dim_date.month,
    dim_date.year,
    dim_zipcode.zipcode,
    (fact_vaxchi.sum_first_dose_18_64 + fact_vaxchi.sum_second_dose_18_64) as `18_64_Total`
FROM fact_vaxchi
    INNER JOIN dim_zipcode on dim_zipcode.zipcode_id=fact_vaxchi.zipcode_id
    INNER JOIN dim_date on dim_date.date_id=fact_vaxchi.date_id
GROUP BY fact_vaxchi.zipcode_id, fact_vaxchi.date_id
HAVING `18_64_Total`
ORDER BY `18_64_Total` DESC
LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

zipcode_id	date_id	month	year	zipcode	18_64_Total
21	14	January	2022	60618	35560

SQL Sample 2

```sql
# During what month the lowest number of cases was recorded?
SELECT
    fact_vaxchi.date_id,
    dim_date.month,
    dim_date.year,
    SUM(fact_vaxchi.sum_covid_cases) as 'Sum_Covid'
FROM fact_vaxchi
    INNER JOIN dim_date on dim_date.date_id=fact_vaxchi.date_id
GROUP BY fact_vaxchi.date_id
ORDER BY fact_vaxchi.sum_covid_cases ASC
LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

date_id	month	year	Sum_Covid
7	June	2021	1504

SQL Sample 3

```sql
# What zipcodes have 5 or more vacination locations?
SELECT
    dim_vax_loc.zipcode_id,
    dim_zipcode.zipcode,
    count(dim_vax_loc.facility_vax_name) as 'Number of Facilities'
FROM dim_vax_loc
    INNER JOIN dim_zipcode ON dim_vax_loc.zipcode_id=dim_zipcode.zipcode_id
GROUP BY dim_vax_loc.zipcode_id
HAVING count(dim_vax_loc.facility_vax_name)>=5
ORDER BY count(dim_vax_loc.facility_vax_name) DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

zipcode_id	zipcode	Number of Facilities
12	60609	7
48	60647	7
11	60608	6
25	60622	6
30	60628	6
18	60615	5
20	60617	5
40	60639	5

Recommendation

- Establish a better data infrastructure to avoid delays in trend analytics for future epidemics.
- During pandemics, developing infrastructure that prepares for the possibility of a variant could help reduce negative health outcomes.
- Prioritize increasing availability of vaccination centers in highly vulnerable communities.
- Promote more health education initiatives to encourage remaining individuals to receive vaccination.

Intended vs Actual Outcomes

- Gathering and cleaning process of reliable datasets would be easy
- Working with covid related data, the ideas about visualizations were not so vivid
- Be able to show different types of analysis on all the information related to our covid program
- We intended to make a solid relation between the CCVI scores and the vaccination rates

- Reliable data gathering process and cleaning it for the model was not easy
- Zip Code as a data value allowed us to make great map visualizations
- Had to reduce demographics for analysis because of data limitations
- Couldn’t really establish a solid relation between the CCVI scores/category with vaccination rates

Lessons Learned

- What could have been better?
  - More time could have been spent on gathering datasets; there may have been datasets that worked better with the rest of our model, required less cleaning, or simply contained more information.
  - Many iterations onto our model and data slowed down progress throughout the quarter; although this helped us polish our results, more planning in the future before jumping into the project could help us optimize our time

- What went well?
  - Strong communication and commitment to meeting deadlines.
  - The particularity when investigating the data allowed for interpreting anomalies. Example: Airport zip code.
  - We found that ease of implementation and interpretability is frequently more important than complexity and excessive optimization.
  - Use of zip code in the model greatly benefitted our visualizations.

Thank You