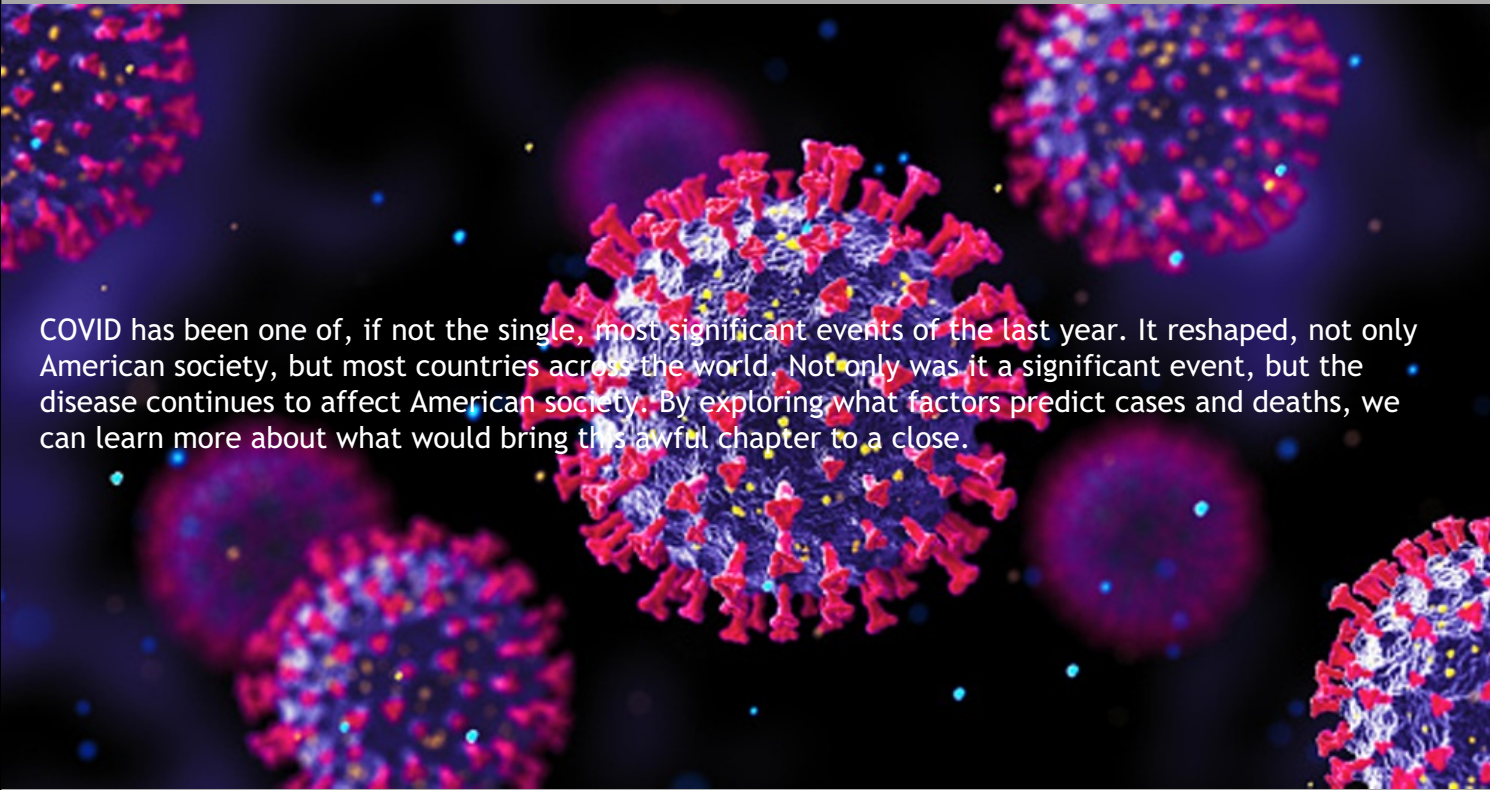


# COVID-19 Cases and Death Prediction Modeling

Topic choice	Data Source	Database Pipeline	Vaccinations table	Cases_Deaths table	ERD	Model Optimization
--------------	-------------	-------------------	--------------------	--------------------	-----	--------------------



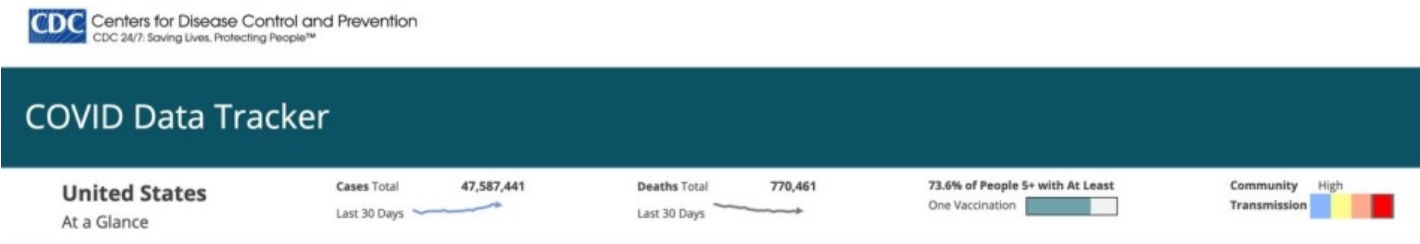
COVID has been one of, if not the single, most significant events of the last year. It reshaped, not only American society, but most countries across the world. Not only was it a significant event, but the disease continues to affect American society. By exploring what factors predict cases and deaths, we can learn more about what would bring this awful chapter to a close.

# COVID-19 Cases and Death Prediction Modeling

Topic choice	Data Source	Database Pipeline	Vaccinations table	Cases_Deaths table	ERD	Model Optimization
--------------	-------------	-------------------	--------------------	--------------------	-----	--------------------

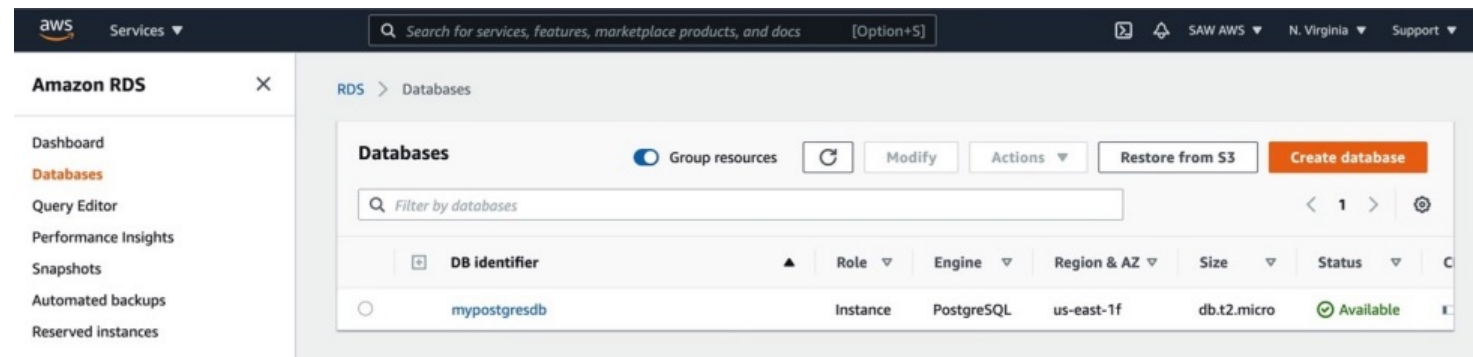
## Description of Data Source

<https://covid.cdc.gov/covid-data-tracker/#datatracker-home> The CDC collects a lot of data, and organizes it into sets. Specifically the cases and deaths over time <https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36>, and the vaccination rate over time <https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-Jurisdic/unsk-b7fc> are our starter sets.



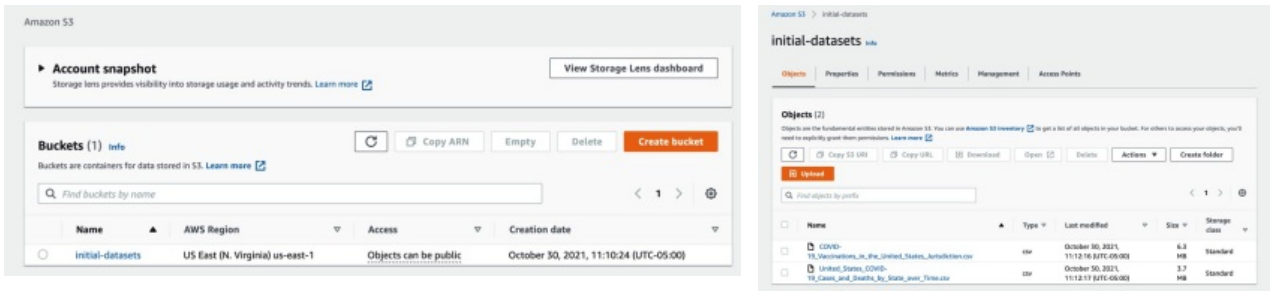
# COVID-19 Cases and Death Prediction Modeling

Topic choice	Data Source	Database Pipeline	Vaccinations table	Cases_Deaths table	ERD	Model Optimization
--------------	-------------	-------------------	--------------------	--------------------	-----	--------------------



## Database Pipeline Demystification:

The data is downloaded from the CDC website into a PostGRES database. In PostGRES, SQL code is utilized to perform joins, merges, and cleaning. The PostGRES database is connected to an Amazon Web Services database, and the csv files are stored in an S3 bucket. From there, the data cleaning script reads in those CSV files using Pandas, which are then cleaned, and fed to the machine learning modeling script.



# COVID-19 Cases and Death Prediction Modeling

Topic choice	Data Source	Database Pipeline	Vaccinations table	Cases_Deaths table	ERD	Model Optimization
--------------	-------------	-------------------	--------------------	--------------------	-----	--------------------

Creation of Vaccinations table and import of data

pgAdmin 4

File Object Tools Help

Dashboard Properties SQL Statistics my\_data\_class\_db/root@my\_aws\_postgres\_rds

Servers (3)

- Postgre
- Postgre
- my\_aws
  - Data
    - my\_data\_class\_db

Query Editor Query History

```
-- Creating tables for Project-One
CREATE TABLE vaccinations (
  date date,
  mmwr_wk int,
  year int,
  location varchar,
  concat varchar,
  distributed int,
  ...
)
```

Data Output Explain Messages Notifications

	submission_date	mmwr_wk	year	location	concat	total_cases	total_confirmed_cases	total_prob_cases	new_cases	prob
	date	integer	integer	character varying	[PK] character varying	integer	integer	integer	integer	integer
1	2020-12-19	51	2020	AK	202051AK	42659		[null]	[null]	325
2	2020-12-26	52	2020	AK	202052AK	44394		[null]	[null]	188
3	2021-01-02	53	2021	AK	202153AK	46530		[null]	[null]	289
4	2021-01-09	1	2021	AK	20211AK	48571		[null]	[null]	374
5	2021-01-16	2	2021	AK	20212AK	50264		[null]	[null]	195
6	2021-01-23	3	2021	AK	20213AK	51663		[null]	[null]	241
7	2021-01-30	4	2021	AK	20214AK	52636		[null]	[null]	129
8	2021-02-06	5	2021	AK	20215AK	53557		[null]	[null]	124
9	2021-02-13	6	2021	AK	20216AK	54488		[null]	[null]	94

# COVID-19 Cases and Death Prediction Modeling

Data Source	Database Pipeline	Vaccinations table	Cases_Deaths table	ERD	Model Optimization	Case importance on mean of 2020
-------------	-------------------	--------------------	--------------------	-----	--------------------	---------------------------------

Creation of Cases Deaths table and import of data

my\_data\_class\_db/root@my\_aws\_postgres\_rds

Query Editor

Query History

```
1 CREATE TABLE casesdeaths (
2     submission_date date,
3     mmwr_wk int,
4     year int,
5     location varchar,
6     concat varchar,
7     distributed int,
8     administered int,
9     recip_administered int,
10    admin_dose1 int,
11    series_complete int,
12    additional_doses int,
13    PRIMARY KEY (concat)
14 );
15
16
17 CREATE TABLE casesdeaths (
18     submission_date date,
19     mmwr_wk int,
20     year int,
21     location varchar,
22     concat varchar,
```

Data Output

Explain

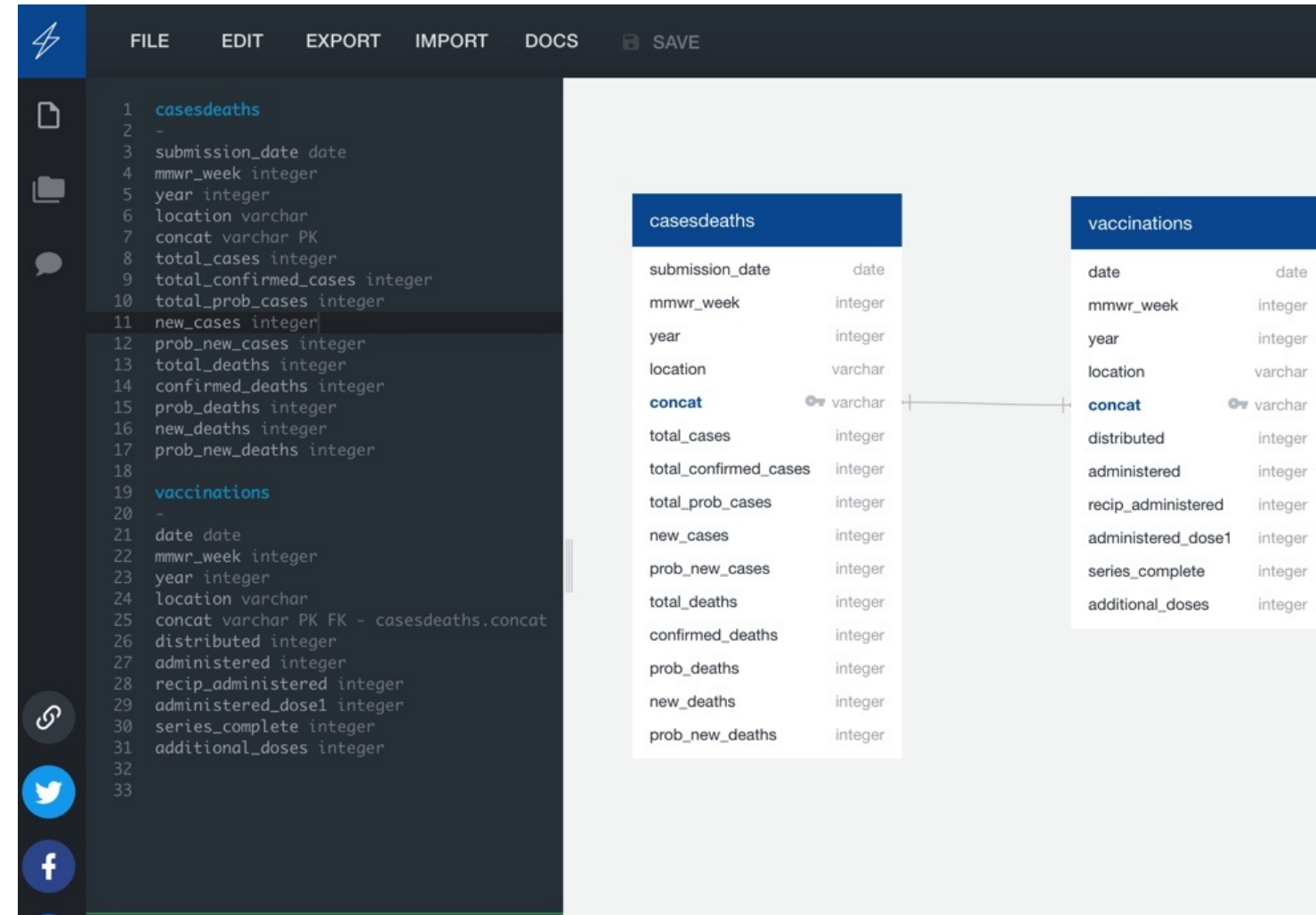
Messages

Notifications

	date	mmwr_wk	year	location	concat	distributed	administered	recip_administered	admin_dose1	series_com
	date	integer	integer	character varying	[PK] character varying	integer	integer	integer	integer	integer
1	2020-12...	51	2020	AK	202051AK	26325	1607	0	0	
2	2020-12...	52	2020	AK	202052AK	45250	11427	0	0	
3	2021-01...	53	2021	AK	202153AK	54975	18401	0	0	
4	2021-01...	1	2021	AK	20211AK	92875	28539	0	0	
5	2021-01...	2	2021	AK	20212AK	149650	54193	0	0	
6	2021-01...	3	2021	AK	20213AK	155700	88344	0	0	
7	2021-01...	4	2021	AK	20214AK	193150	115268	0	0	
8	2021-02...	5	2021	AK	20215AK	248500	139147	0	0	
9	2021-02...	6	2021	AK	20216AK	271550	175138	174962	122518	
10	2021-02...	7	2021	AK	20217AK	299935	227403	227363	148497	
11	2021-02...	8	2021	AK	20218AK	242525	250528	250706	162720	

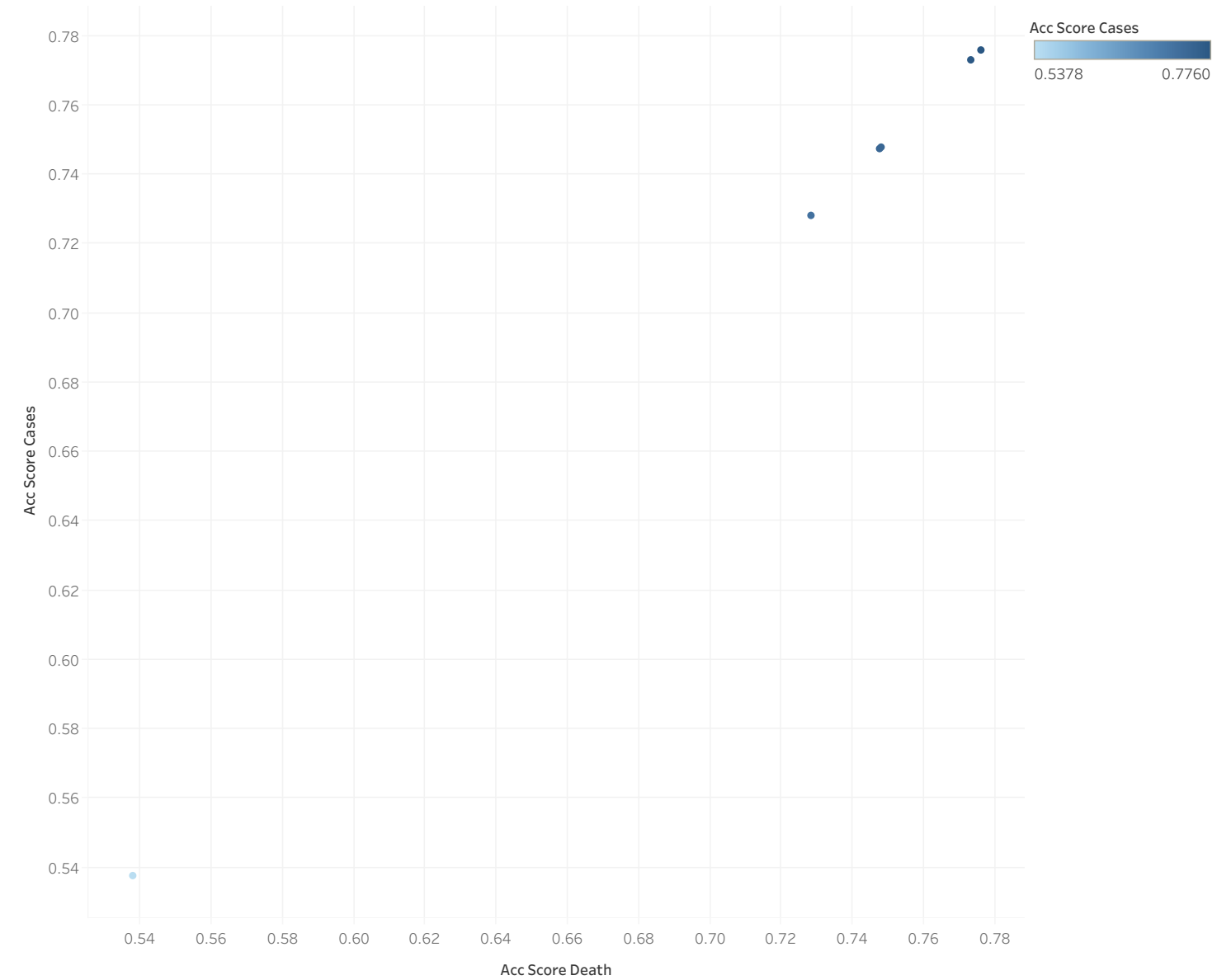
# COVID-19 Cases and Death Prediction Modeling

Database Pipeline	Vaccinations table	Cases_Deaths table	ERD	Model Optimization	Case importance on mean of 2020	Case importance on mean of 2020 and 2..
-------------------	--------------------	--------------------	-----	--------------------	---------------------------------	---



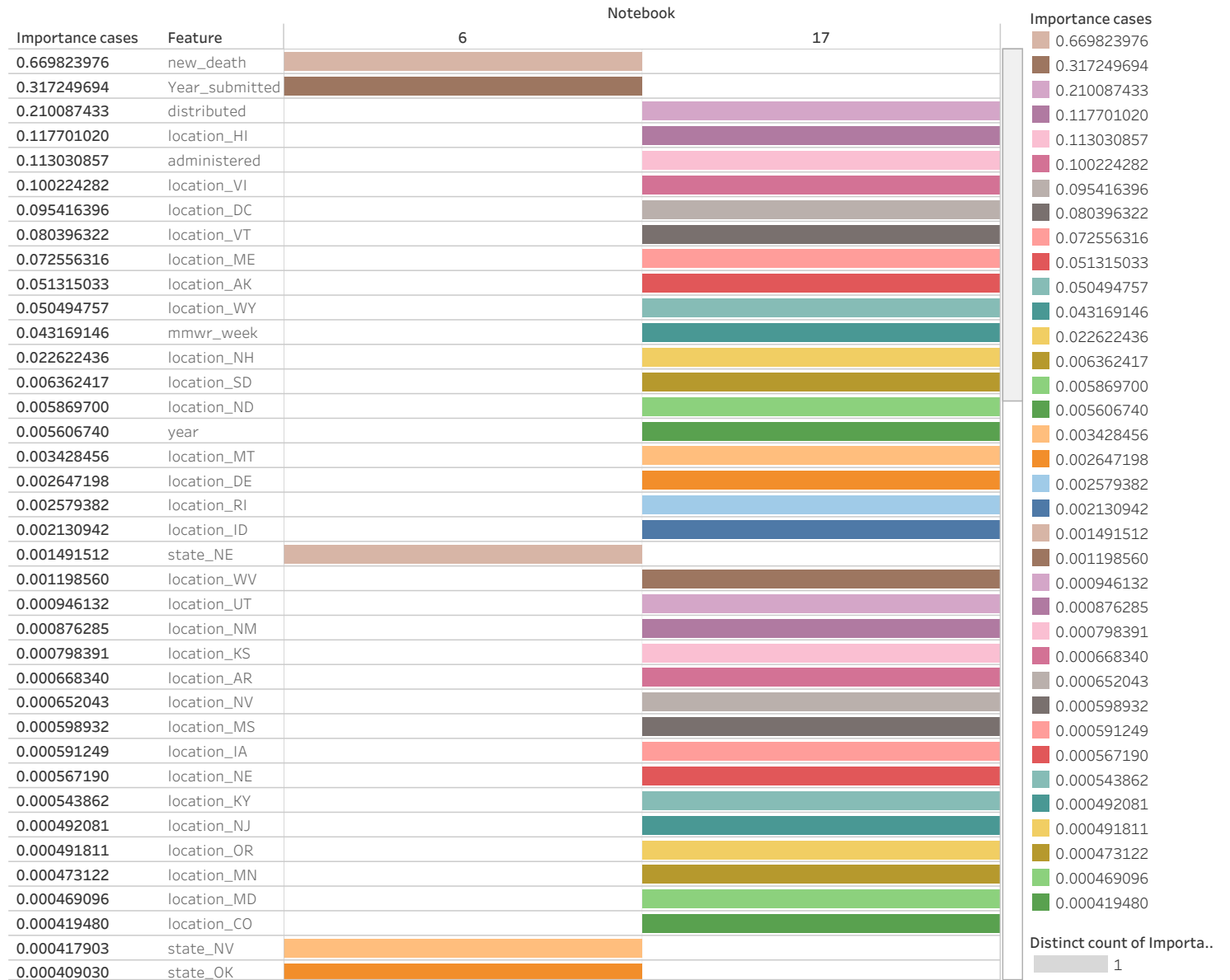
# COVID-19 Cases and Death Prediction Modeling

Vaccinations table	Cases_Deaths table	ERD	Model Optimization	Case importance on mean of 2020	Case importance on mean of 2020 and 20..	Death importance on mean of 2020
--------------------	--------------------	-----	--------------------	---------------------------------	--	----------------------------------



# COVID-19 Cases and Death Prediction Modeling

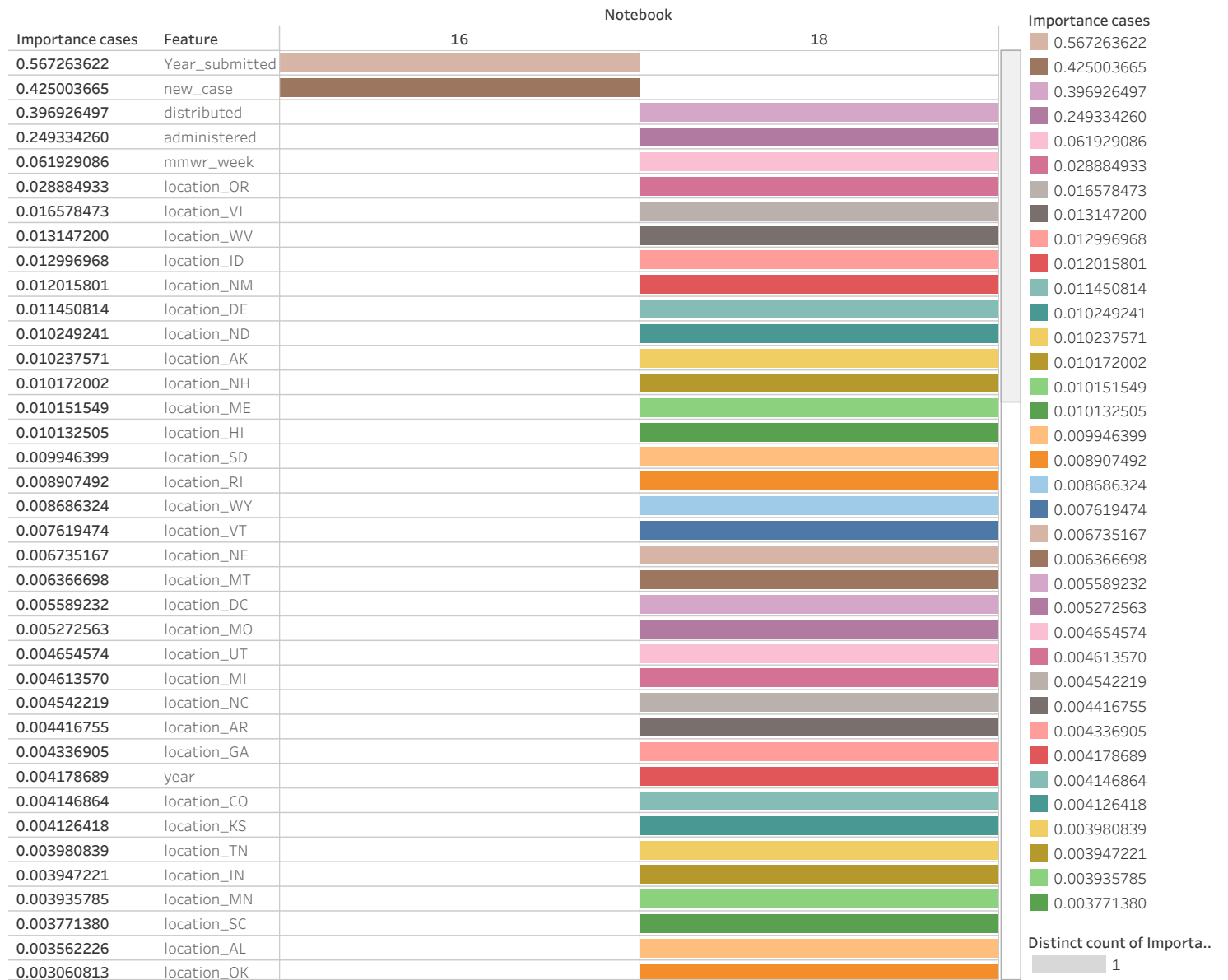
Cases_Deaths table	ERD	Model Optimization	Case importance on mean of 2020	Case importance on mean of 2020 and 20..	Death importance on mean of 2020	Death importance on mean of 2020 an..
--------------------	-----	--------------------	---------------------------------	--	----------------------------------	---------------------------------------





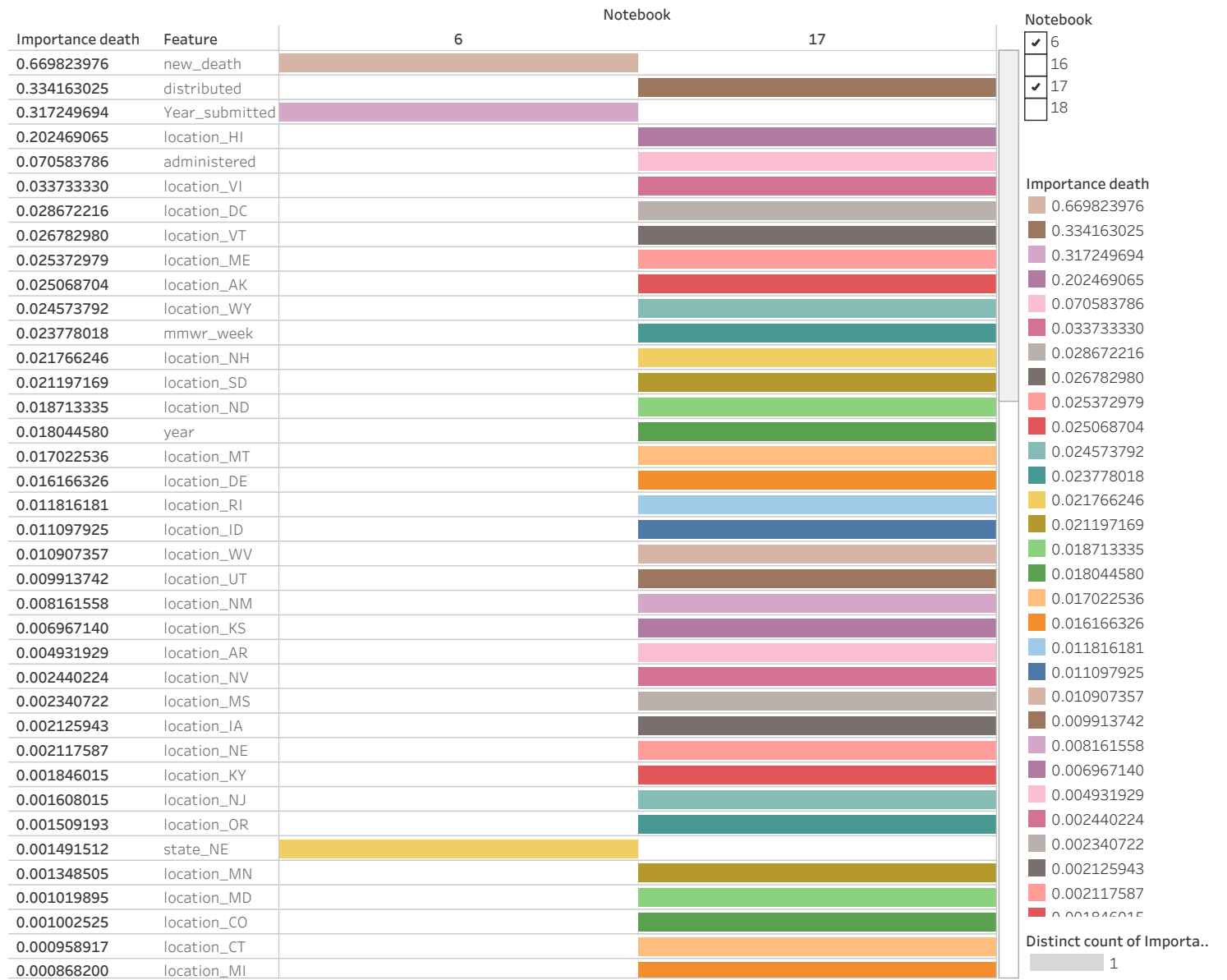
# COVID-19 Cases and Death Prediction Modeling

Cases_Deaths table	ERD	Model Optimization	Case importance on mean of 2020	Case importance on mean of 2020 and 20..	Death importance on mean of 2020	Death importance on mean of 2020 and 20..
--------------------	-----	--------------------	---------------------------------	--	----------------------------------	---



# COVID-19 Cases and Death Prediction Modeling

Cases_Deaths table	ERD	Model Optimization	Case importance on mean of 2020	Case importance on mean of 2020 and 20..	Death importance on mean of 2020	Death importance on mean of 2020 and 20..
--------------------	-----	--------------------	---------------------------------	--	----------------------------------	---



# COVID-19 Cases and Death Prediction Modeling

Cases_Deaths table	ERD	Model Optimization	Case importance on mean of 2020	Case importance on mean of 2020 and 20..	Death importance on mean of 2020	Death importance on mean of 2020 and 20..
--------------------	-----	--------------------	---------------------------------	--	----------------------------------	---

