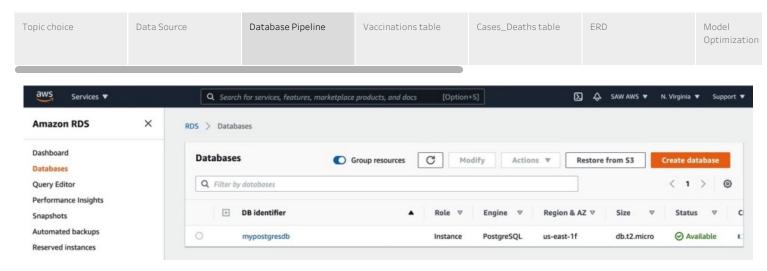
Topic choice	Data Source	Database Pipeline	Vaccinations table	Cases_Deaths table	ERD	Model Optimization
4 . A.A.						
A STATE OF THE STA		. 44				
COVID has be American soo	een one of, if not ciety, but most co	the single, most untries acress th	significant events e world. Not only	s of the last year. Was it a significa	It reshaped, not ant event, but the	only
disease conti	inues to affect Am ore about what wo	nerican society. E	By exploring what	factors predict of	cases and deaths,	we
_ 16	1				4	
	2730 A		_		. 🕭	



Description of Data Source

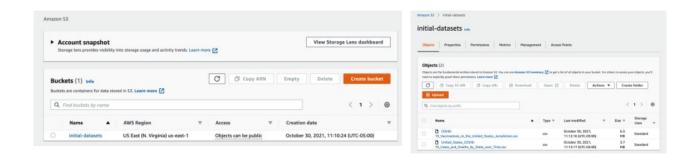
https://covid.cdc.gov/covid-data-tracker/#datatracker-home The CDC collects a lot of data, and organizes it into sets. Specifically the cases and deaths over time https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36, and the vaccination rate over time https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-Un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





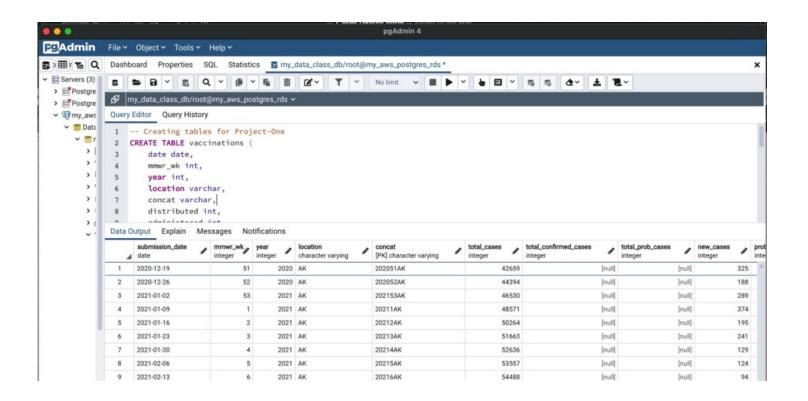
Database Pipeline Demystification:

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.



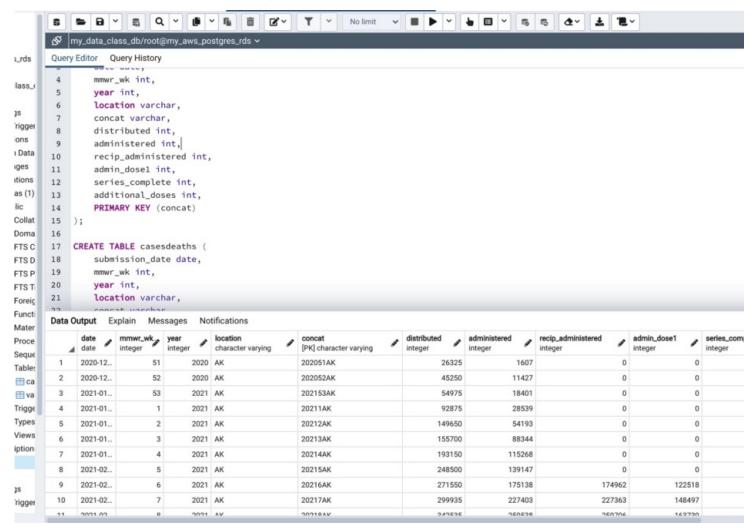
Topic choice	Data Source	Database Pipeline	Vaccinations table	Cases_Deaths table	ERD	Model Optimization

Creation of Vaccinations table and import of data

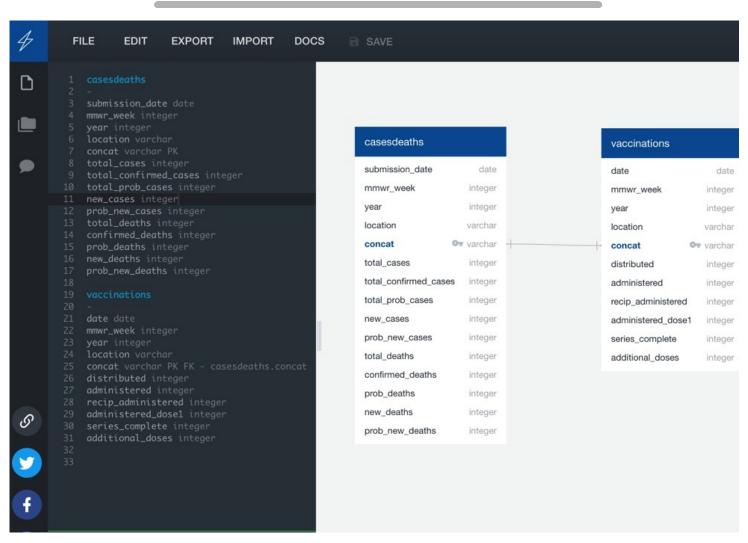


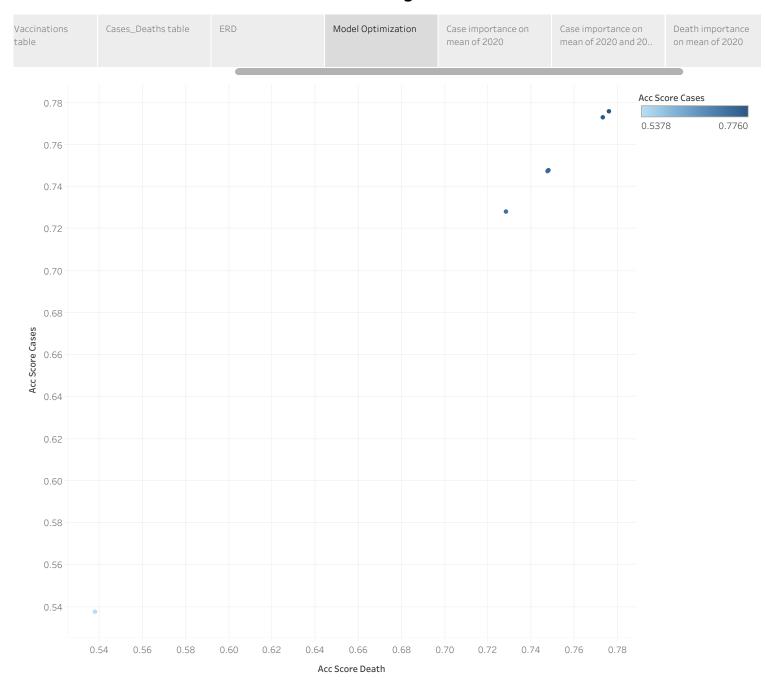
Data Source Database Pipeline Vaccinations table Cases_Deaths table ERD Model Optimization Case importance on mean of 2020

Creation of Cases Deaths table and import of data



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





Cases_Deaths table	ERD	Model Optimization	Case importance mean of 2020	on	Case importance on mean of 2020 and 20	Death importance mean of 2020	on Death importance on mean of 2020 an
Importance cases	s Feature	6	Note	book	17		Importance cases 0.669823976
0.669823976	new_death	,					0.317249694
0.317249694	Year_submitted						0.210087433
0.210087433	distributed						0.117701020
0.117701020	location_HI						0.113030857
0.113030857	administered						0.100224282
0.100224282	location_VI						0.095416396
0.095416396	location_DC						0.080396322
0.080396322	location_VT						0.072556316
0.072556316	location_ME						0.051315033
0.051315033	location_AK						0.050494757
0.050494757	location_WY						0.043169146
0.043169146	mmwr_week						0.022622436
0.022622436	location_NH						0.006362417
0.006362417	location_SD						0.005869700
0.005869700	location_ND						0.005606740
0.005606740	year						0.003428456
0.003428456	location_MT						0.002647198
0.002647198	location_DE						0.002579382
0.002579382	location_RI						0.002130942
0.002130942	location_ID						0.001491512
0.001491512	state_NE						0.001198560
0.001198560	location_WV						0.000946132
0.000946132	location_UT						0.000876285
0.000876285	location_NM						0.000798391
0.000798391	location_KS						0.000668340
0.000668340	location_AR						0.000652043
0.000652043	location_NV						0.000598932
0.000598932	location_MS						0.000591249
0.000591249	location_IA						0.000567190
0.000567190	location_NE						0.000543862
0.000543862	location_KY						0.000492081
0.000492081	location_NJ						0.000491811
0.000491811	location_OR						0.000473122
0.000473122	location_MN						0.000469096
0.000469096	location_MD						0.000419480
0.000419480	location_CO						Distinct count of Importa
0.000417903	state_NV						1
0.000409030	state_OK						Τ.

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
	ı		Noteb			Importance cases
Importance case		16		18		0.567263622
0.567263622	Year_submitted					0.425003665
0.425003665	new_case					0.396926497
0.396926497	distributed					0.249334260
0.249334260	administered					0.061929086
0.061929086	mmwr_week					0.028884933
0.028884933	location_OR					0.016578473
0.016578473	location_VI					0.013147200
0.013147200	location_WV					0.012996968
0.012996968	location_ID					0.012015801
0.012015801	location_NM					0.011450814
0.011450814	location_DE					0.010249241
0.010249241	location_ND					0.010237571
0.010237571	location_AK					0.010172002
0.010172002	location_NH					0.010151549
0.010151549	location_ME					0.010132505
0.010132505	location_HI					0.009946399
0.009946399	location_SD					0.008907492
0.008907492	location_RI					0.008686324
0.008686324	location_WY					0.007619474
0.007619474	location_VT					0.006735167
0.006735167	location_NE					0.006366698
0.006366698	location_MT					0.005589232
	location_DC					0.005272563
0.005272563 0.004654574	location_MO location_UT					0.004654574
0.004654574	location_MI					0.004613570
0.004513570	location_NC					0.004542219
0.004342219	location_AR					0.004416755
0.004416733	location_GA					0.004336905
0.004330903	year					0.004178689
0.004176864	location_CO					0.004146864
0.004126418	location_KS					0.004126418
0.003980839	location_TN					0.003980839
0.003947221	location_IN					0.003947221
0.003935785	location_MN					0.003935785
0.003771380	location_SC					0.003//1300
0.003562226	location_AL					Distinct count of Importa
0.003060813	location_OK					1

Cases_ I 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
		_	Notebo		. !	Notebook
Importance dea		6		17		6
0.669823976	new_death					16
0.334163025	distributed					<u>✓</u> 17 18
0.317249694	Year_submitted					
0.202469065	location_HI					
0.070583786	administered					
0.033733330	location_VI					mportance death 0.669823976
0.028672216	location_DC					0.334163025
0.026782980	location_VT					0.317249694
0.025372979	location_ME					0.202469065
0.025068704	location_AK					0.070583786
0.024573792	location_WY					0.033733330
0.023778018	mmwr_week					0.028672216
0.021766246	location_NH					0.026782980
0.021197169	location_SD					0.025372979
0.018713335	location_ND					0.025068704
0.018044580	year					0.024573792
0.017022536	location_MT					0.023778018
0.016166326	location_DE					0.021766246
0.011816181	location_RI					0.021197169
0.011097925 0.010907357	location_ID location_WV					0.018713335
0.010907357						0.018044580
0.009913742	location_UT location_NM					0.017022536
						0.016166326
0.006967140 0.004931929	location_KS location_AR					0.011816181
0.004931929	location_NV					0.011097925
0.002440224	location_MS					0.010907357
0.002340722	location_IA					0.009913742
0.002123343	location_NE					0.008161558
0.002117387	location_KY					0.006967140
0.001646015	location_NJ					0.004931929
0.001509193	location_OR					0.002440224
0.001303133	state_NE					0.002340722
0.001431512	location_MN					0.002125943
0.001348303	location_MD					0.002117587
0.001013635	location_CO					0 00104601E
0.000958917	location_CT					Distinct count of Importa
0.000938917	location_MI					1

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
Importance d	e Feature	16	Notebook	18		mportance death 0.525764485
0.525764485	Year_submitted					0.460965546
0.460965546	new_case					0.404784725
0.404784725	distributed					0.239541711
0.239541711	administered					0.049987957
0.049987957	mmwr_week					0.034665211
0.034665211	location_OR					0.019331157
0.019331157	location_VI					0.013288699
0.013288699	location_WV					0.012667376
0.012667376	location_ID					0.012427632
0.012427632	location_NM					0.011357690
0.011357690	location_DE					0.009483568
0.009483568	location_ND					0.008268094
0.008268094	location_AK					0.007896774
0.007896774	location_NH					0.007824348
0.007824348	location_ME					0.007383429
0.007383429	location_HI					0.007202593
0.007202593	location_SD					0.006928209
0.006928209	location_RI					0.006907804
0.006907804	location_WY					0.006715775
0.006715775	location_VT					0.006641137
0.006641137	location_NE					0.006309778
0.006309778	location_MT					0.006259000
0.006259000	location_DC					0.006255131
0.006255131	location_MO					0.006206658
0.006206658						0.005535065
0.005535065						0.005415000
0.005415000	_					0.005195093
0.005195093						0.005082759
0.005082759						0.005076655
0.005076655						0.005048010
0.005048010						0.004999321
0.004999321						0.004920178
0.004920178						0.004818630
0.004818630						0.004517528
0.004517528						0.004272009
0.004272009	_					Distinct count of Importa
0.004188340					L	1
0.003975259	location_OK					±