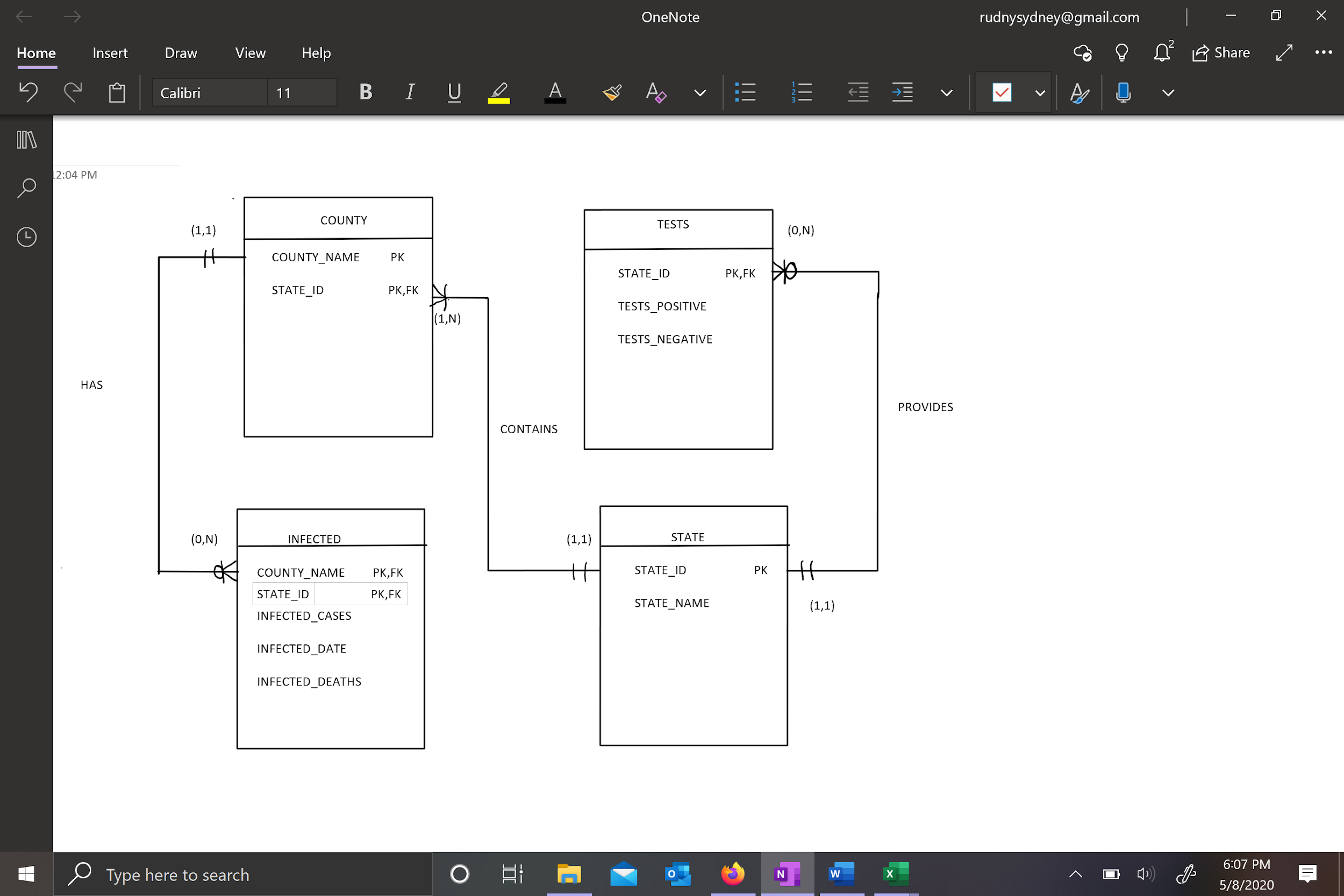
Sydney Rudny

Final Database Project

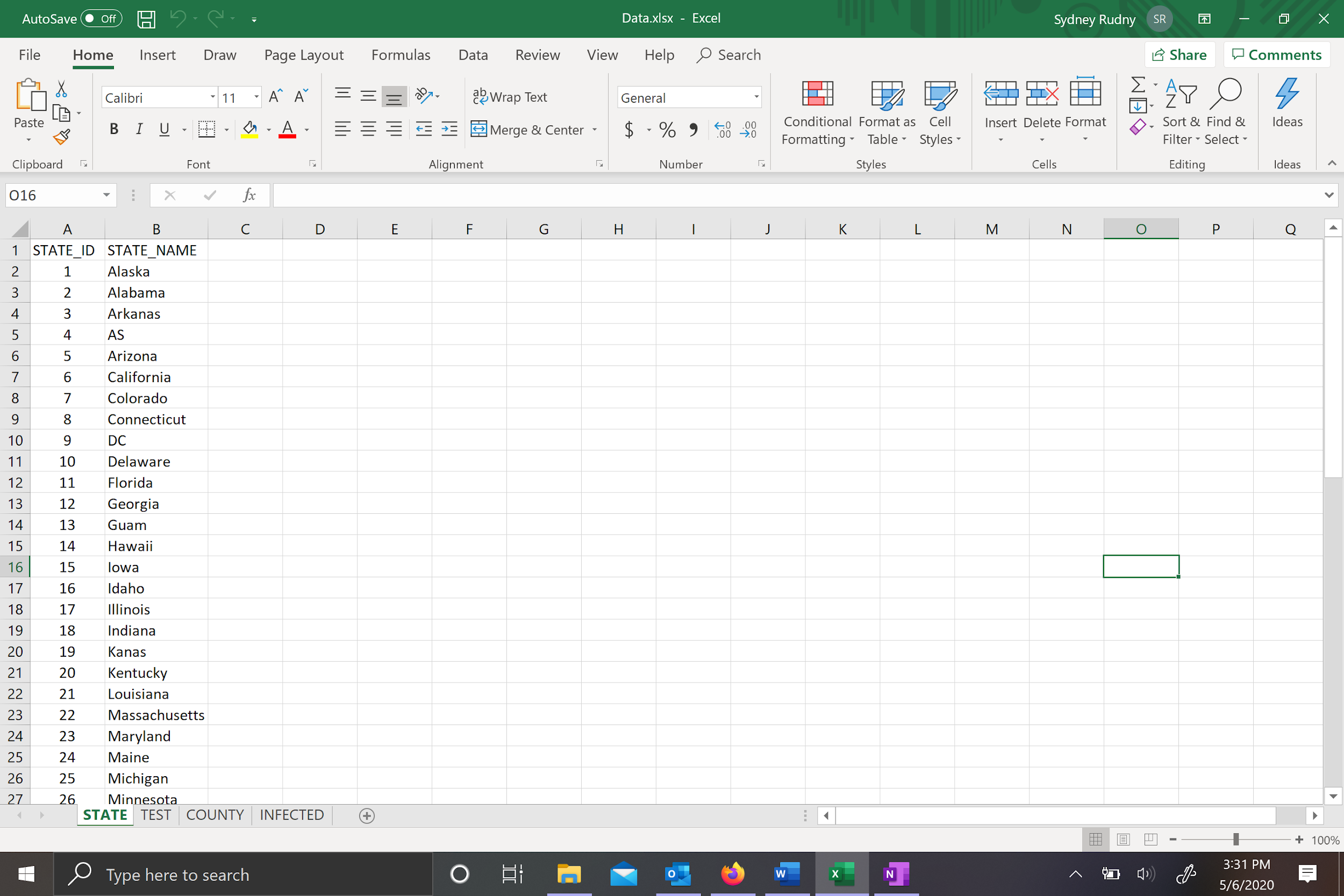
5/12/2020

**Data:**

I pulled two different data sets from Kaggle (<https://www.kaggle.com/fireballbyedimyrnmom/us-counties-covid-19-dataset>) and from Our World in Data (<https://ourworldindata.org/covid-testing>). These provided me with the best data that I could organize to create a COVID information database. I messed around with a couple different ways to do an ERD before choosing the datasets. Here is the ERD I found worked the best:



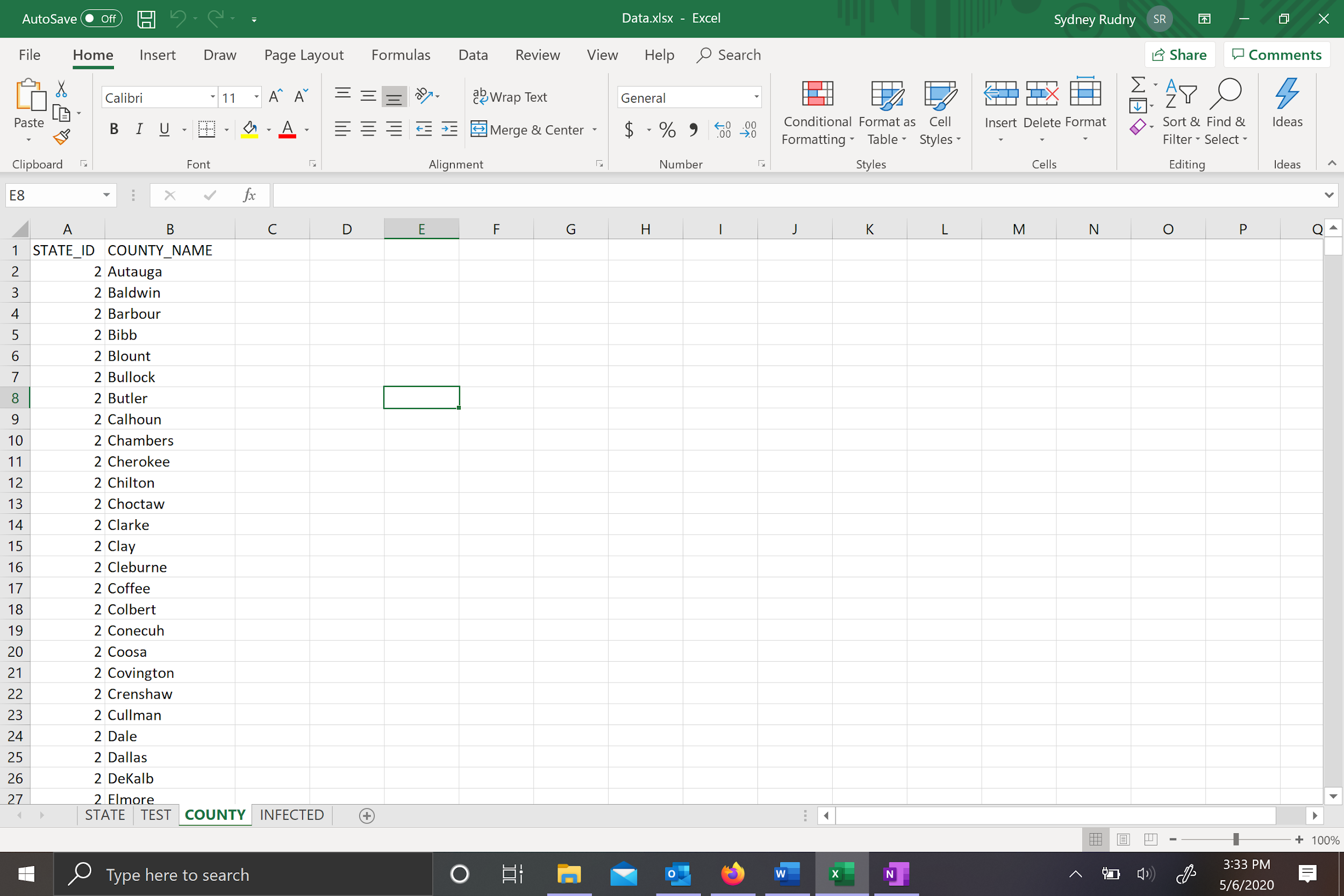
From Kaggle I found a table that displayed information about US counties and COVID data. I uploaded it into an excel sheet and did some minor edits. The original table had 6 columns and I ended up keeping only 4 of them (county, cases, deaths, and date). Next I pulled testing data for the United States from Our World in Data into excel. Like the Kaggle data it also contained a lot of columns that I didn’t really need so I only kept 3 (state, positive tests, and negative tests). From here I began building the tables for the database out into excel so it was easier for the upload. I created an excel worksheet for each of the four tables I had in mind to build: State, County, Tests, and Infected. For the state sheet I created ID’s for each state based off of the states/territories that were shown from the testing data.



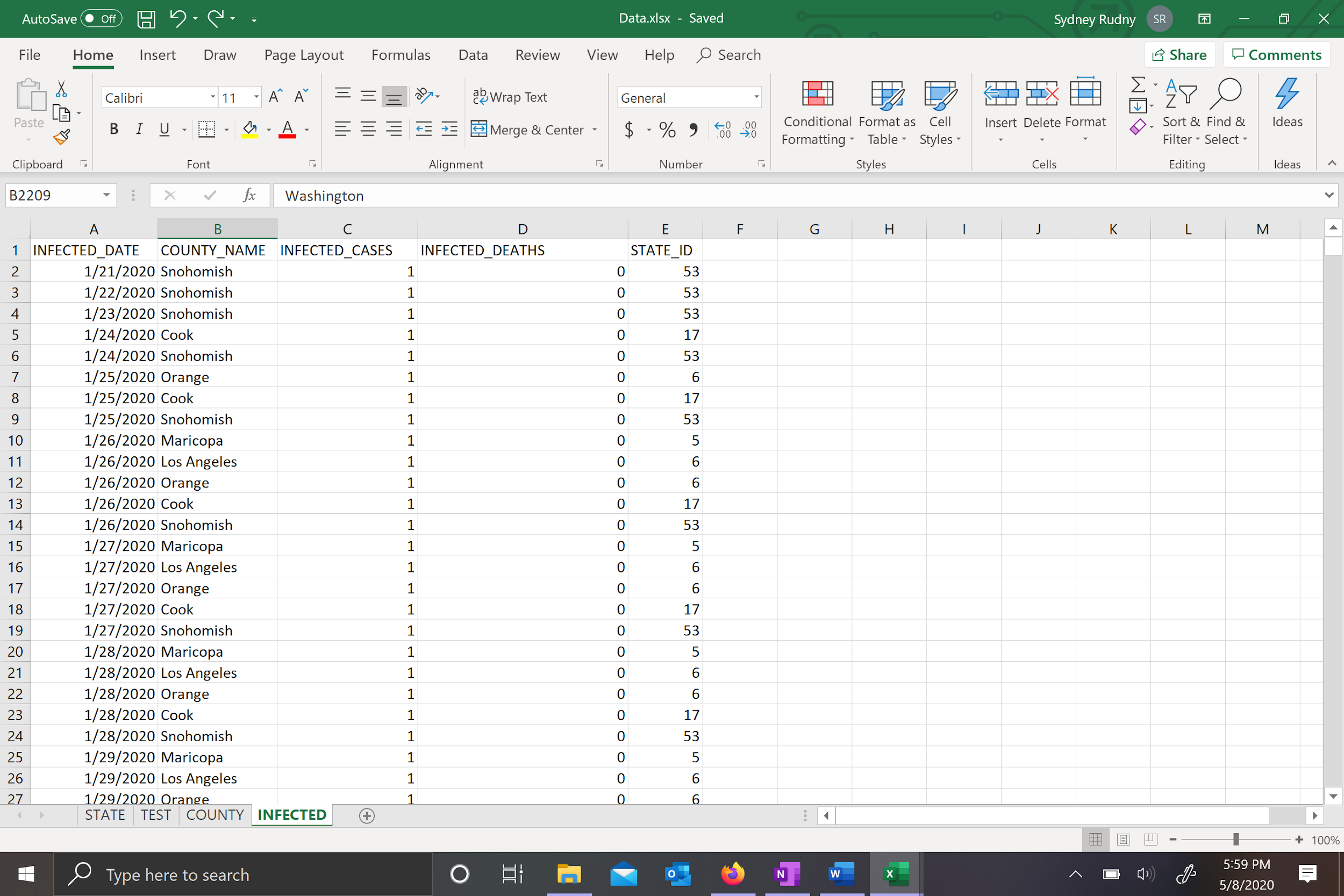
Then I created a testing worksheet for the test table and deleted the state names and replaced it with the ID’s I had created.



Next I created a county table. At first I thought it would be best to give the counties ID’s but there was over 2,000 of them and they were repeated multiple times since there was multiple dates so I decided to leave the counties to just their name and also made state ID in the county table as a primary and foreign key incase county names were used multiple times for different states.



The infected was left alone as it was from Kaggle except for the state names which I had to find and replace in excel into the corresponding state ID numbers.

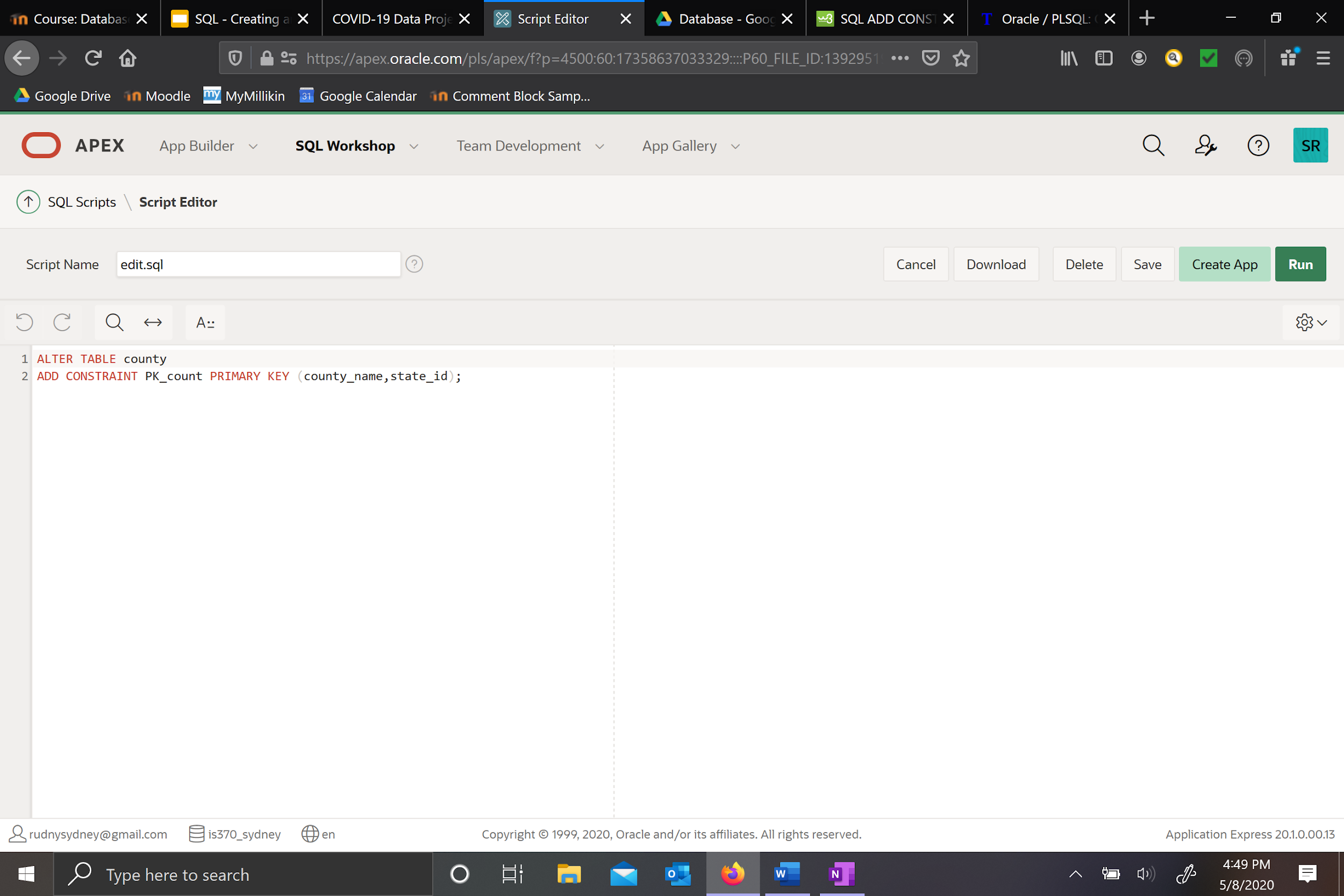


**ORACLE APEX:**

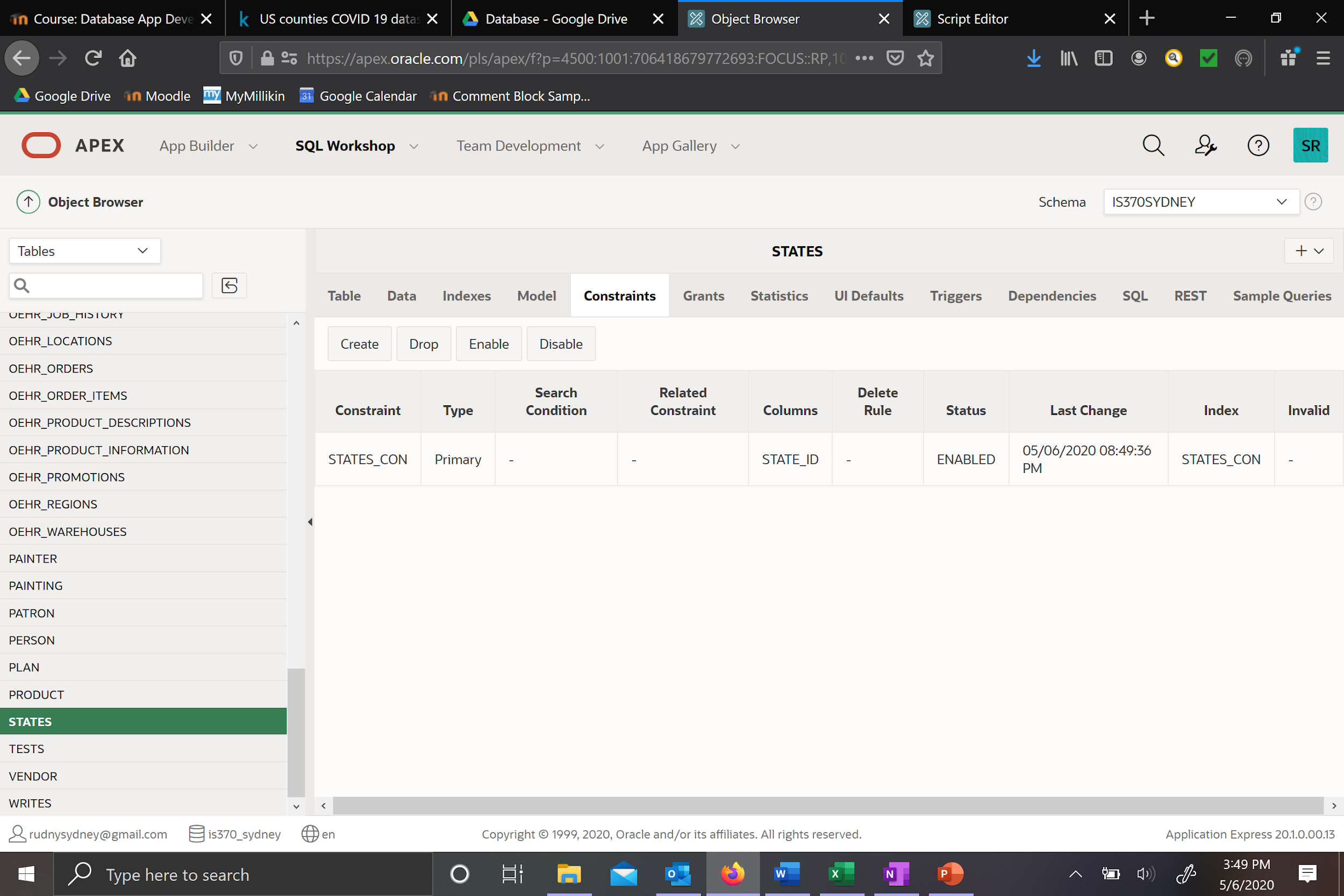
Separating the data into excel worksheets proved to very helpful when it came to uploading the data into Oracle APEX. In APEX, I used the Data Workshop to upload each of the worksheets separately to create their own tables. Here are the results:

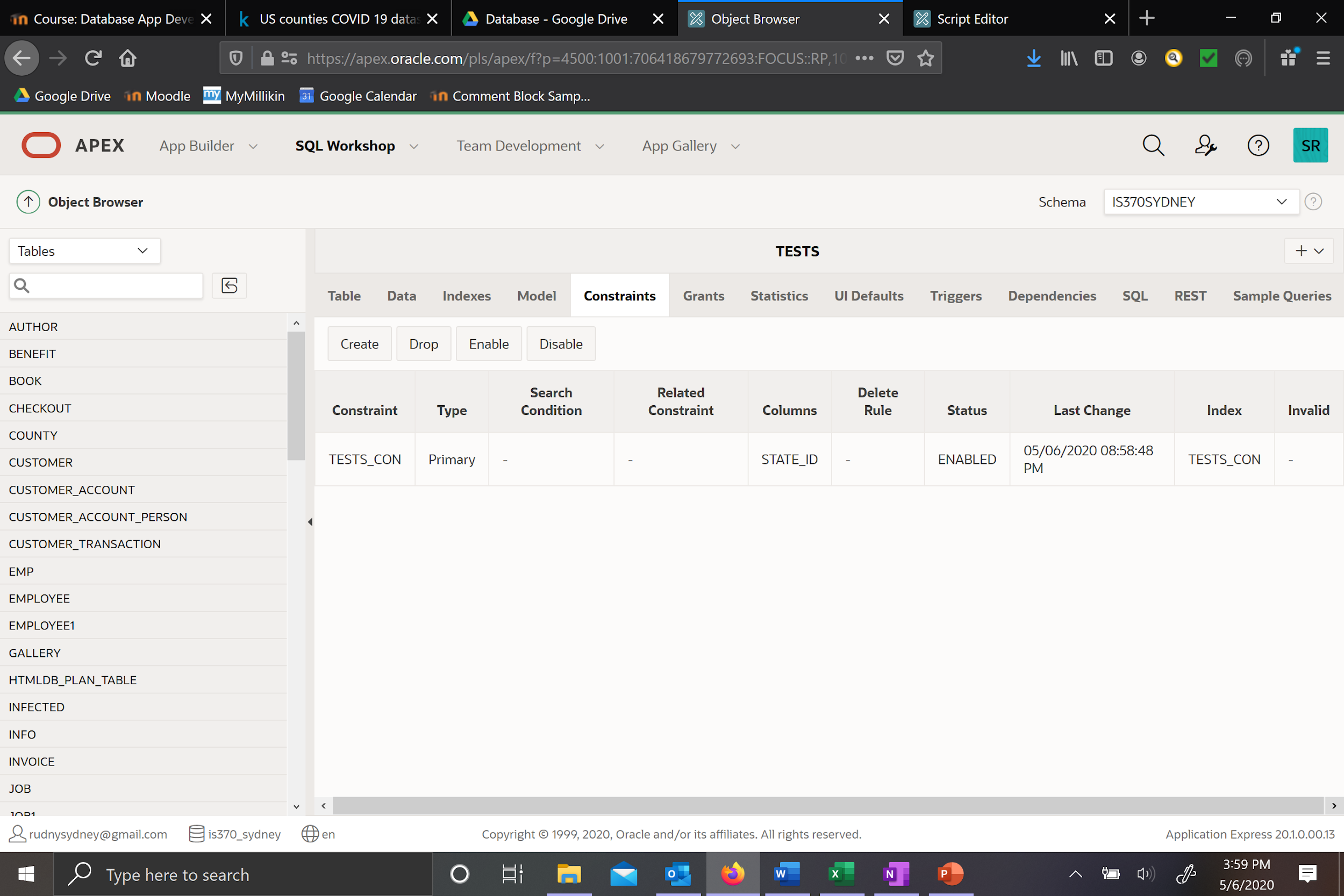
|  |  |
| --- | --- |
|  |  |
|  |  |

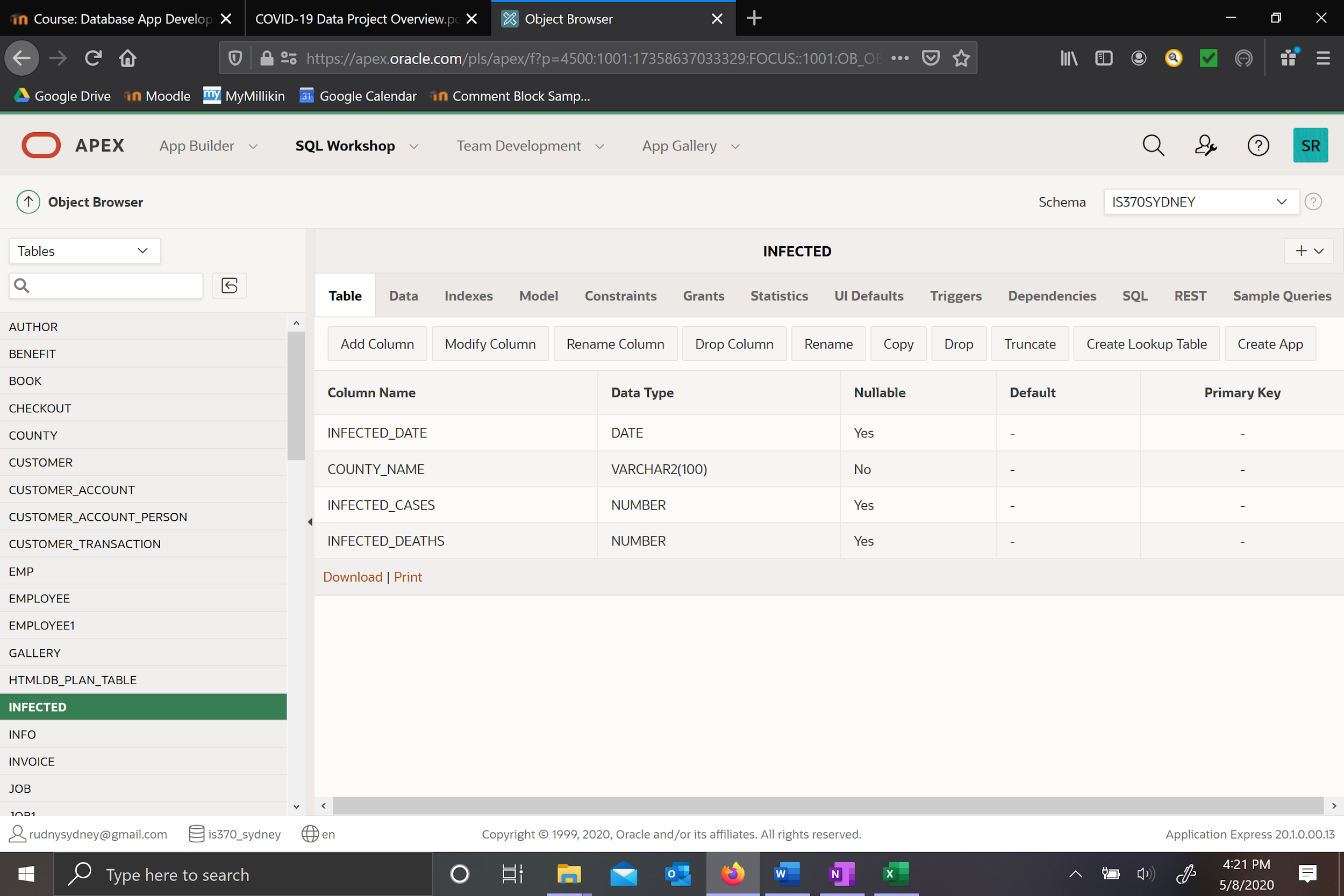
Next, I assigned the primary and foreign key restraints through the data workshop editor. The county table used two attributes to make an unique PK so I had to do that through the script editor.

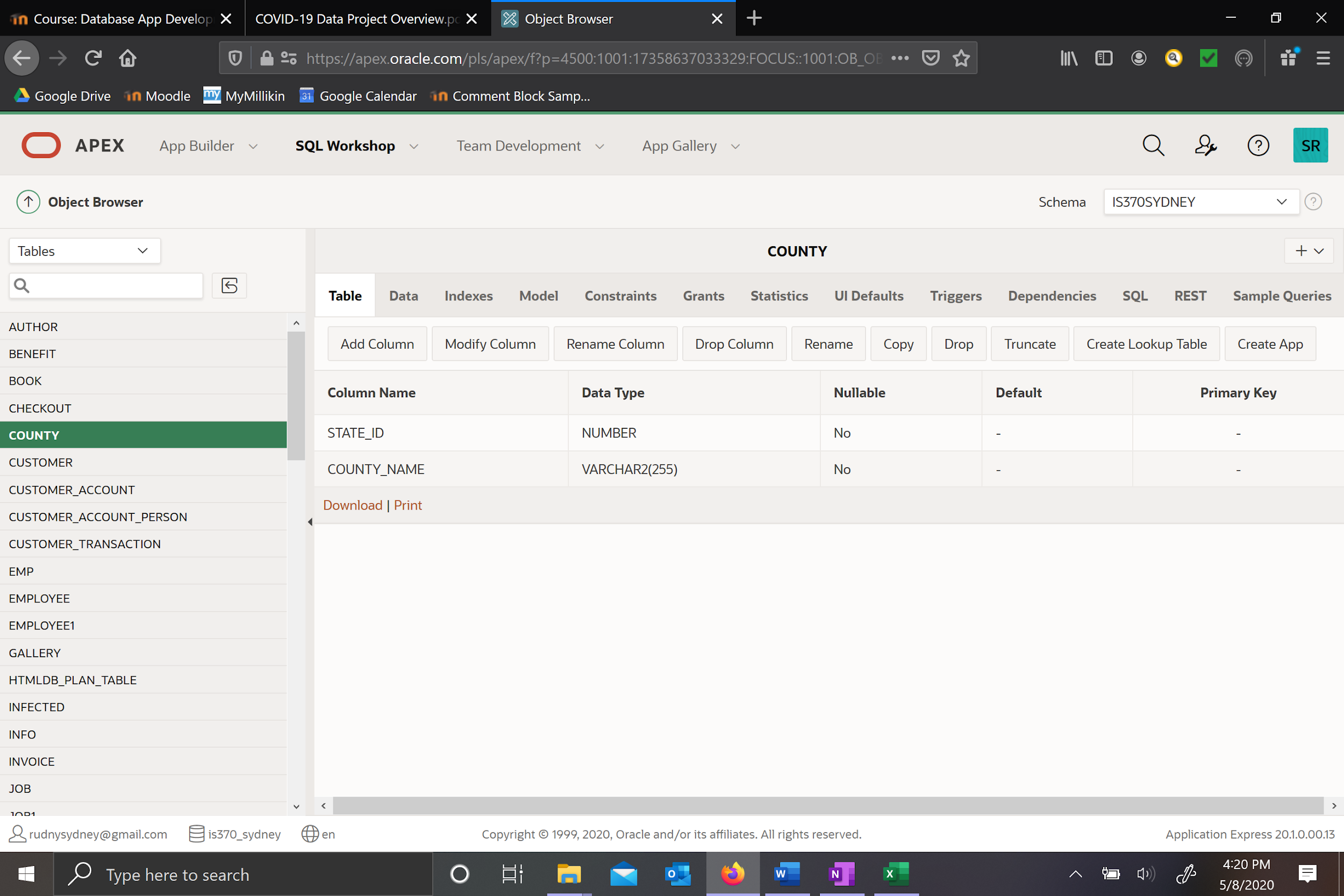


Other constraints (data types) were already in place from the upload. APEX based the data type off the first few rows of each column and assigned it which were correct when I looked back over the constraints. Screenshots below show examples of how attributes assigned data types and edits made to create PK/FK.









**Queries & Results:**

|  |  |  |  |
| --- | --- | --- | --- |
| SELECT s.state\_name, ((t.tests\_positive) + (t.tests\_negative)) AS “Total Tests”  FROM states s, tests t  WHERE s.state\_id = t.state\_id; | **Page: Total Tests by State**  This report shows the total amount of tests taken in each state. (Testing data up until May 1st) | | |
| SELECT s.state\_name, t.tests\_positive  FROM states s, tests t  WHERE s.state\_id = t.state\_id  ORDER BY t.tests\_positive DESC; | **Page: Positive Tests per State**  This graph shows the number of positive tests from each state. | | |
| SELECT t.tests\_positive, CONCAT(ROUND((t.tests\_positive/(t.tests\_positive + t.tests\_negative)\*100),2),'%') AS " Percent Postive"  FROM tests t, states s  WHERE s.state\_id = t.state\_id AND s.state\_name = (SELECT state\_name FROM states WHERE state\_name = 'Illinois'); | | **Page: Illinois Positive Testing**  This report shows the percent of people who tested positive from Illinois. | |
| SELECT infected\_date, SUM(infected\_cases) AS "US Total Known Cases"  FROM infected  GROUP BY infected\_date  ORDER BY infected\_date ASC; | | **Page: Total US Cases By Date**  This shows a linear graph representation of the total known tested cases in the US. | |
| SELECT t.tests\_positive, s.state\_name,  CASE  WHEN t.tests\_positive < 5000 THEN 'Low'  WHEN t.tests\_positive >= 5000 AND t.tests\_positive < 15000 THEN 'Mid'  WHEN t.tests\_positive >= 15000 THEN 'High'  ELSE 'Unknown'  END AS "Infected Level"  FROM tests t, states s  WHERE s.state\_id = t.state\_id; | | **Page: Infected Ranges**  This report displays a classification of which states have low,mid, or high amounts of positive COVID tests. | |
| SELECT infected\_date, infected\_deaths, county\_name  FROM infected  WHERE county\_name = 'New York City'; | | **Page: New York City Virus Deaths**  This report shows an area graph display of the death rates in New York City (where COVID has hit the worst). | |
| SELECT infected\_date, infected\_cases, county\_name  FROM infected  WHERE county\_name = 'New York City'; | | **Page: New York City Spread**  The first part of this report shows the an area graph of the infection cases in New York City. | |
| SELECT CONCAT(ROUND(((i.infected\_deaths)/(i.infected\_cases))\*100,2),'%') AS "Death Rate", i.county\_name  FROM infected i, county c, states s  WHERE i.county\_name = c.county\_name AND c.state\_id = s.state\_id  AND infected\_date = '04-24-2020'AND infected\_cases > 1000 AND s.state\_name = (SELECT state\_name FROM states WHERE state\_name = 'Illinois'); | | **Page: Illinois Death Rate 4/24/2020**  This report shows the death rate in each county in Illinois. Case data was recorded from January to April 25th, 2020 so I chose one of the last days the data was recorded to get a somewhat recent record. | |
| SELECT county\_name,infected\_date,infected\_deaths  FROM infected  WHERE county\_name = 'New York City'  ORDER BY infected\_date ASC; | | **Page: New York City Spread**  The second part of this report displays a breakdown in numbers the virus deaths by day. | |
| SELECT ((t.tests\_negative)+(t.tests\_positive)) AS "Total Tests", CONCAT(ROUND((t.tests\_positive/(t.tests\_positive + t.tests\_negative)\*100),2),'%') AS " Percent Positive",s.state\_name  FROM tests t, states s  WHERE s.state\_id = t.state\_id | | | **Page: Testing % by State**  This report shows the total tests and the % that tested positive in each state. |
| SELECT SUM(tests\_positive) AS "US Cases", CONCAT(ROUND(((SUM(tests\_positive))/328000000)\*100,2),'%') AS "% of US Population",  CONCAT(ROUND((SUM((tests\_positive))/7800000000)\*100,2),'%') AS "% of World Population"  FROM tests; | | | **Page: US case comparison**  This report shows the total tests compared to the US and World population to put the infection spread into perspective. (Population numbers were rough estimates from www.worldometers.info) |