**Project: - Physician Compare Dataset**

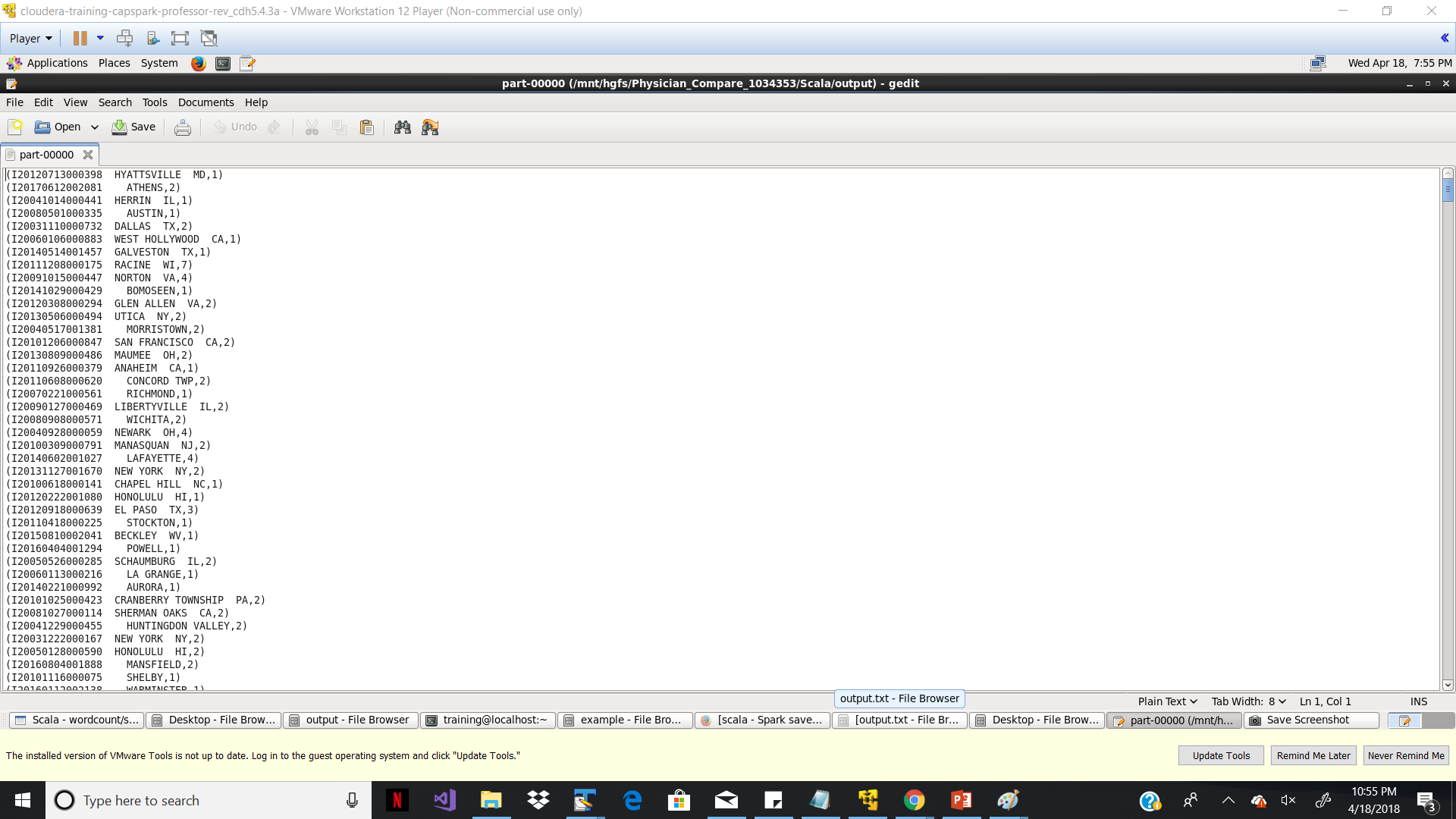
**Name: - Bhavana Vankhede UB ID: - 1034353**

**Scala Programming: -**

Report will be produce for physician details with address who are working on multiple locations and single locations separately

Here the< key,value> pair was <PAC\_ID ,(City,State) >

Please find the code attached in Scala Folder.



**Big Query: -**

Data is imported on Google storage platform. As to implement the SCD 1 (Slowly Changing Dimensions), there has been 2 tables needs to be created for SCD implementation.

Staging and target table.

1. Physician\_Compare.Test -- target table which will be updated if city, state ,flag details of any existing physician changes. Else it will insert a new record.
2. Physician\_Compare.Test\_stagging – is a staging table where all data is populated from input.

**Query** :-

merge into Physician\_Compare.Test as cm

using Physician\_Compare.Test\_stagging as cs

on (cm.pac\_id = cs.pac\_id)

when matched and cm.city <> cs.city and cm.state <> cs.state

then update set cm.city =cs.city, cm.flag='N', cm.state=cs.state

when not matched

then insert

(NPI,

PAC\_ID,

Professional\_Enrollment\_ID,

Last\_Name,

First\_Name,

Middle\_Name,

Suffix,

Gender,

Credential,

Medical\_school\_name,

Graduation\_year,

Primary\_specialty,

Secondary\_specialty\_1,

Secondary\_specialty\_2,

Secondary\_specialty\_3,

Secondary\_specialty\_4,

All\_secondary\_specialties,

Organization\_legal\_name,

Group\_Practice\_PAC\_ID,

Number\_of\_Group\_Practice\_members,

Line\_1\_Street\_Address,

Line\_2\_Street\_Address,

Marker\_of\_address\_line\_2\_suppression,

City,

State,

Zip\_Code,

Phone\_Number,

Hospital\_affiliation\_CCN\_1,

Hospital\_affiliation\_LBN\_1,

Hospital\_affiliation\_CCN\_2,

Hospital\_affiliation\_LBN\_2,

Hospital\_affiliation\_CCN\_3,

Hospital\_affiliation\_LBN\_3,

Hospital\_affiliation\_CCN\_4,

Hospital\_affiliation\_LBN\_4,

Hospital\_affiliation\_CCN\_5,

Hospital\_affiliation\_LBN\_5,

Professional\_accepts\_Medicare\_Assignment,

Reported\_Quality\_Measures,

Used\_electronic\_health\_records,

Committed\_to\_heart\_health\_through\_the\_Million\_Hearts\_\_\_initiative\_ ,

flag)

values

(

cs.NPI,

cs.PAC\_ID,

cs.Professional\_Enrollment\_ID,

cs.Last\_Name,

cs.First\_Name,

cs.Middle\_Name,

cs.Suffix,

cs.Gender,

cs.Credential,

cs.Medical\_school\_name,

cs.Graduation\_year,

cs.Primary\_specialty,

cs.Secondary\_specialty\_1,

cs.Secondary\_specialty\_2,

cs.Secondary\_specialty\_3,

cs.Secondary\_specialty\_4,

cs.All\_secondary\_specialties,

cs.Organization\_legal\_name,

cs.Group\_Practice\_PAC\_ID,

cs.Number\_of\_Group\_Practice\_members,

cs.Line\_1\_Street\_Address,

cs.Line\_2\_Street\_Address,

cs.Marker\_of\_address\_line\_2\_suppression,

cs.City,

cs.State,

cs.Zip\_Code,

cs.Phone\_Number,

cs.Hospital\_affiliation\_CCN\_1,

cs.Hospital\_affiliation\_LBN\_1,

cs.Hospital\_affiliation\_CCN\_2,

cs.Hospital\_affiliation\_LBN\_2,

cs.Hospital\_affiliation\_CCN\_3,

cs.Hospital\_affiliation\_LBN\_3,

cs.Hospital\_affiliation\_CCN\_4,

cs.Hospital\_affiliation\_LBN\_4,

cs.Hospital\_affiliation\_CCN\_5,

cs.Hospital\_affiliation\_LBN\_5,

cs.Professional\_accepts\_Medicare\_Assignment,

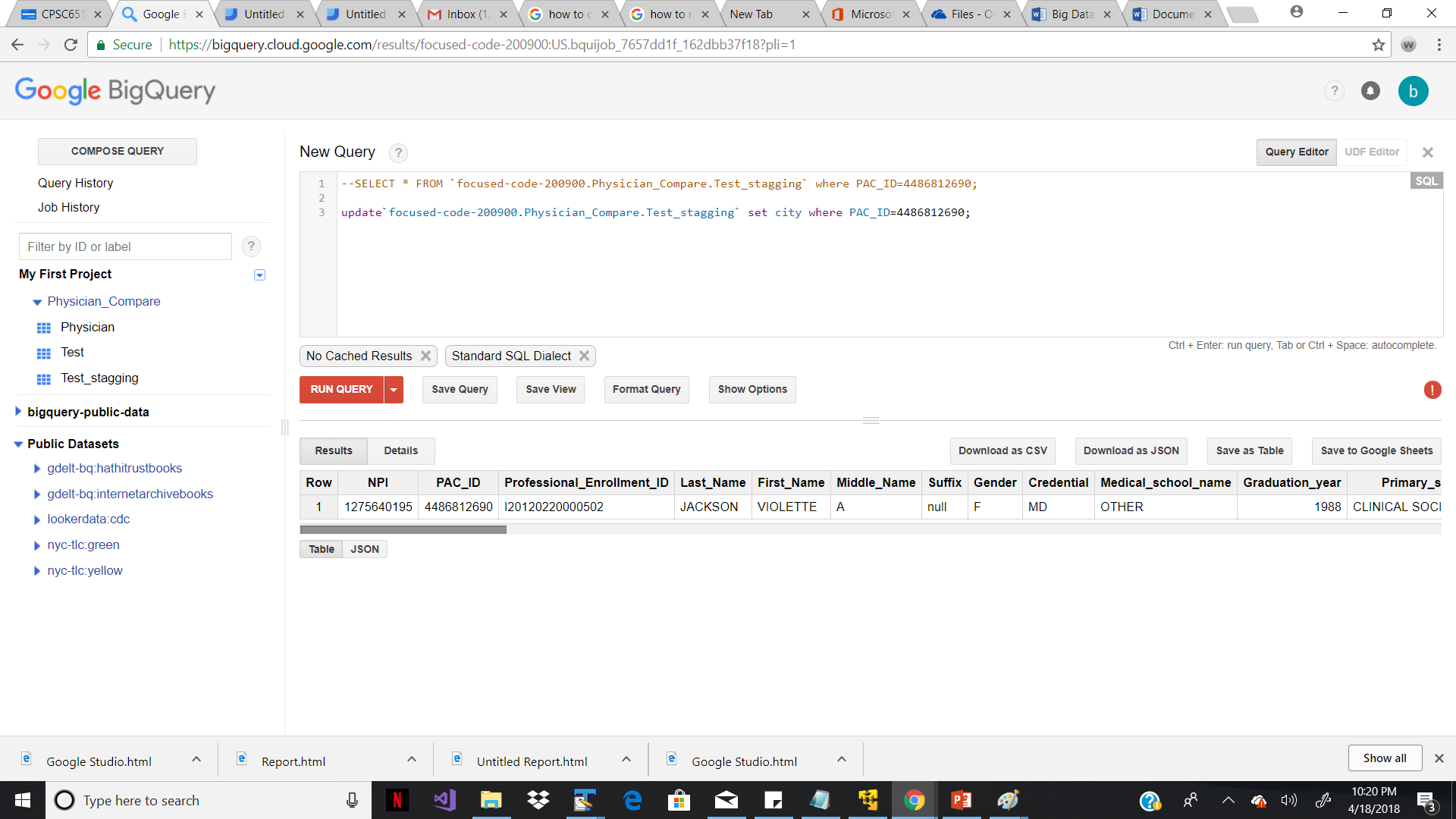
cs.Reported\_Quality\_Measures,

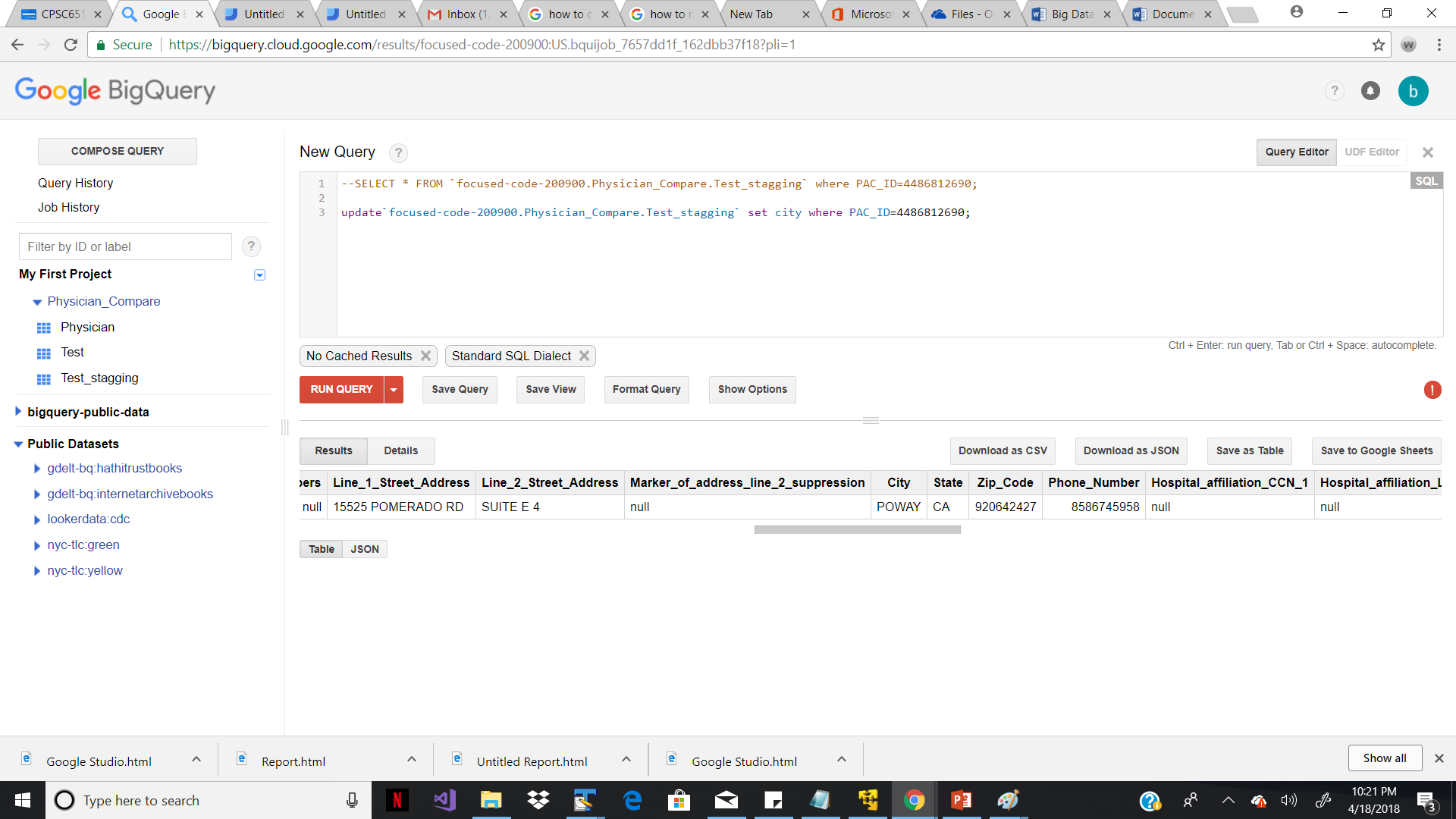
cs.Used\_electronic\_health\_records,

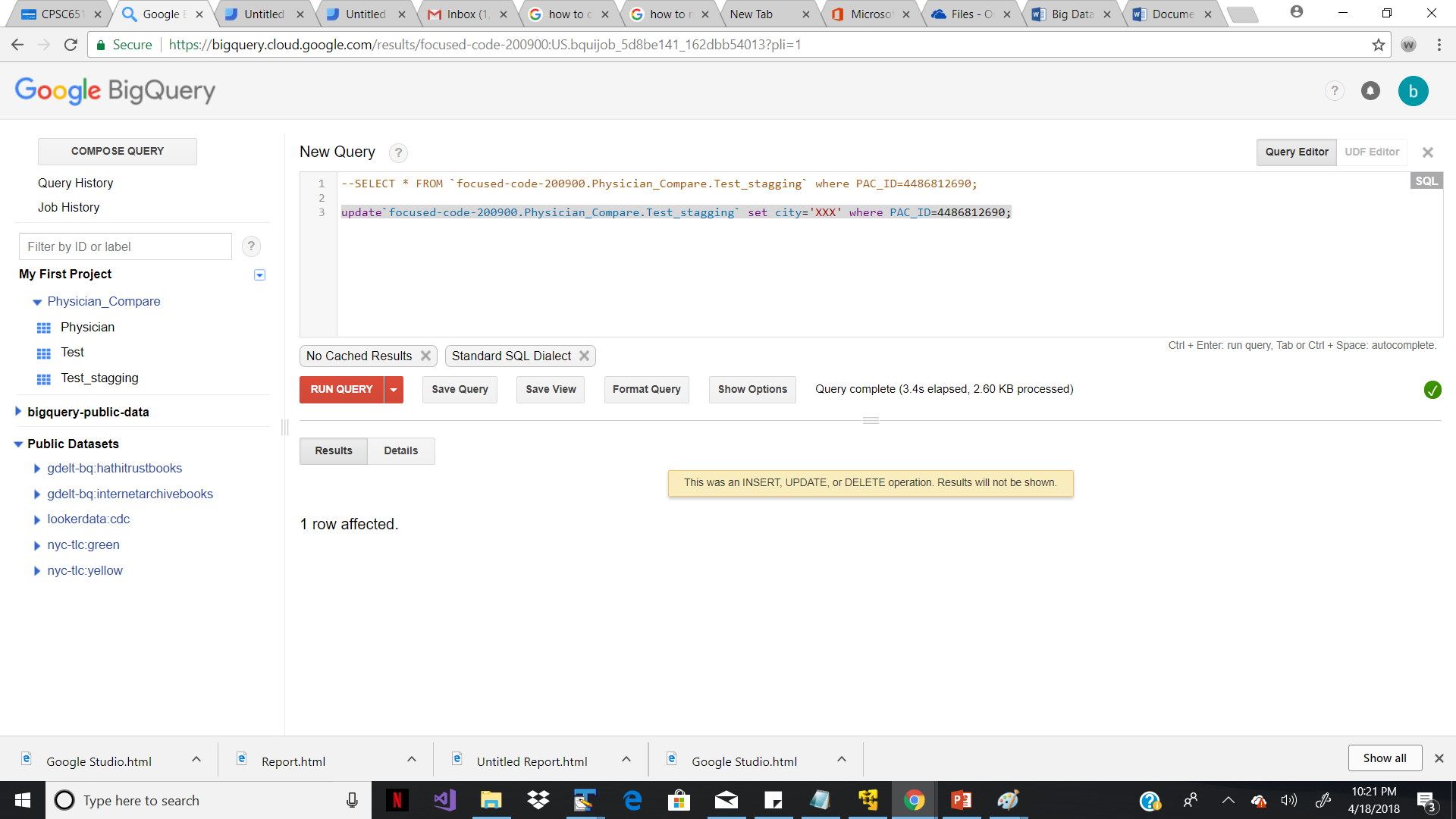
cs.Committed\_to\_heart\_health\_through\_the\_Million\_Hearts\_\_\_initiative\_ , 'N');

A) Modified Existing record:-

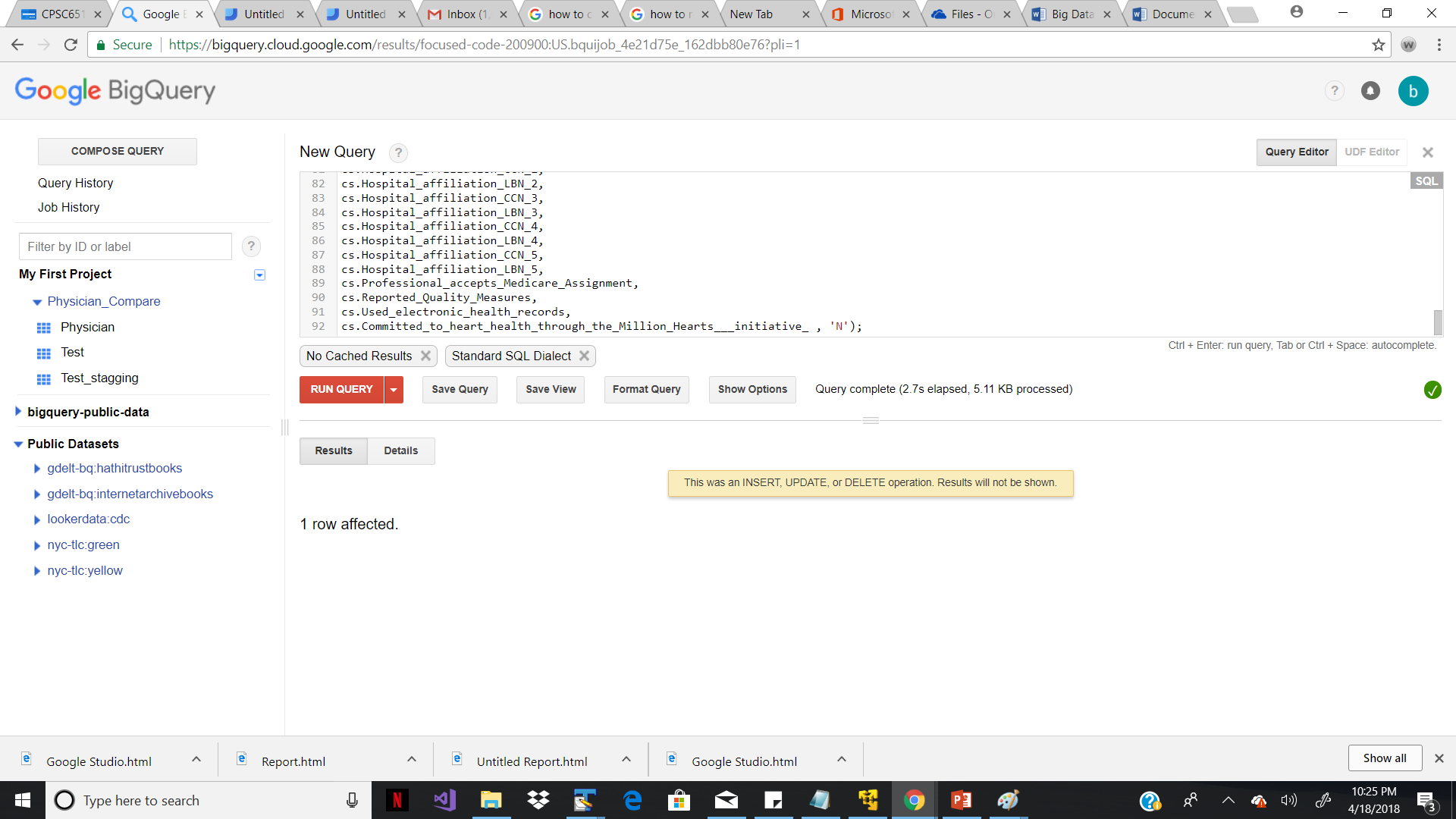
Before modification:-



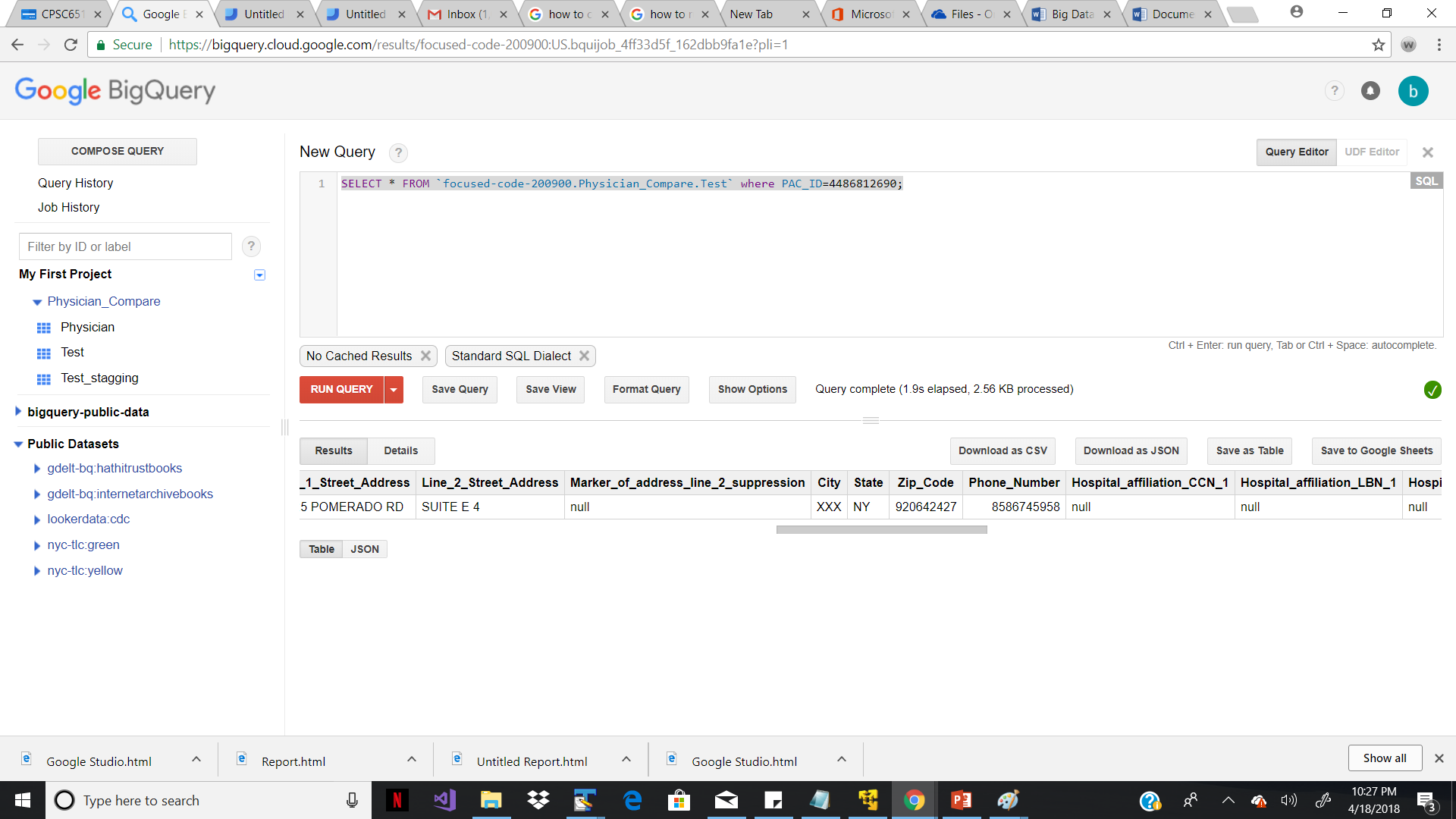




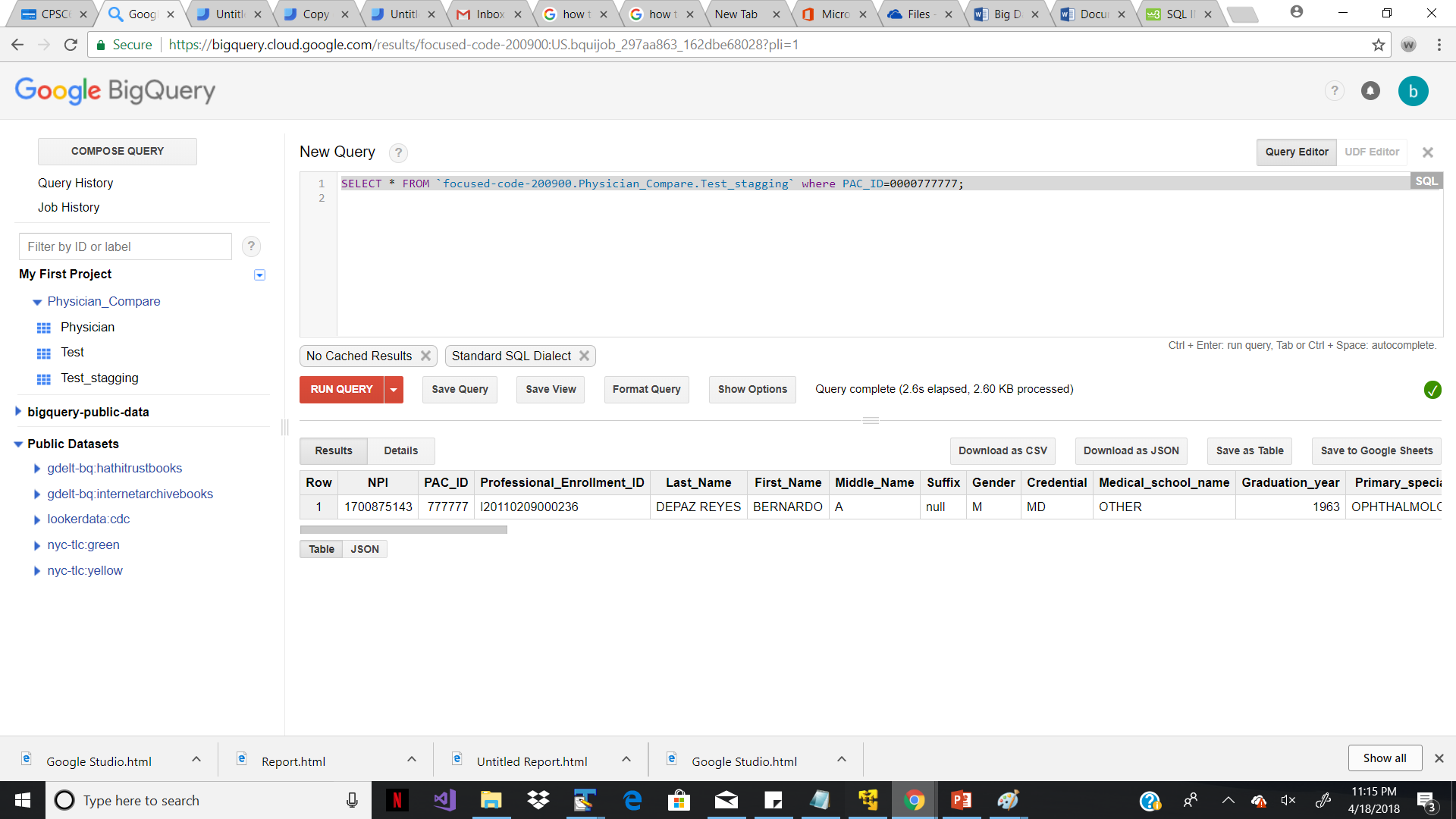
After running query :-

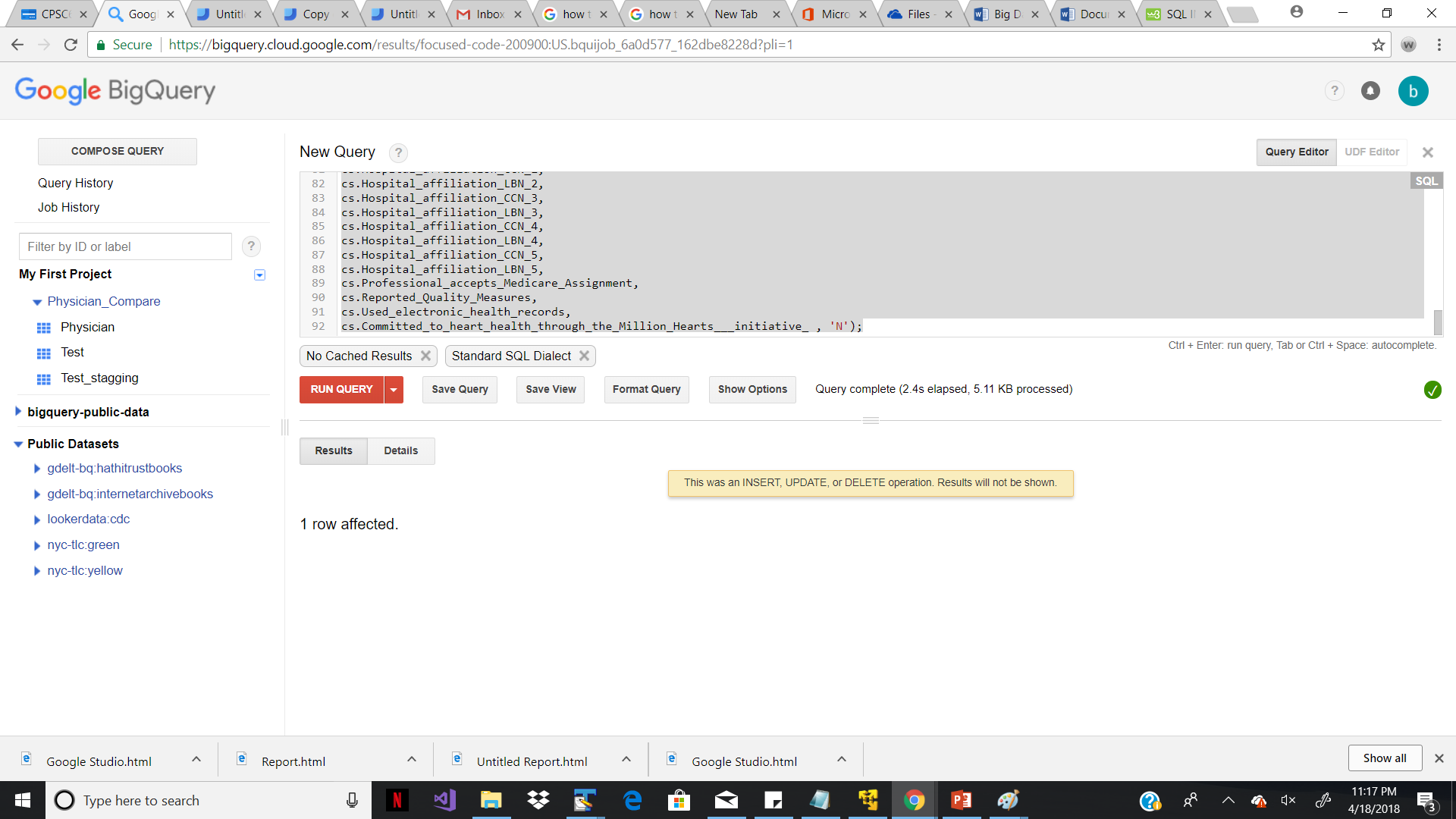


Below are the updated values in target table :-

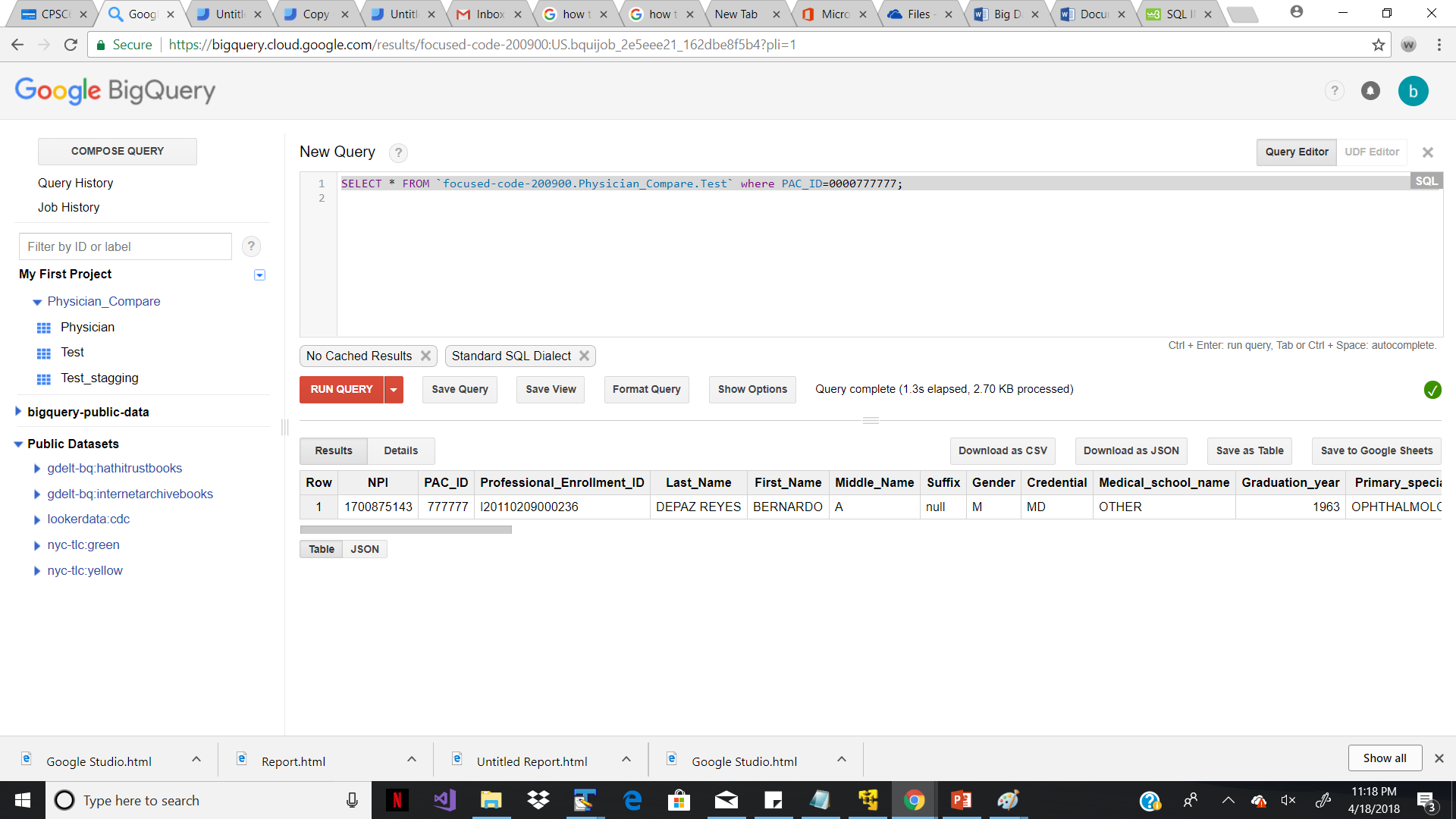


B) Inserting completely new value from staging to target table: -





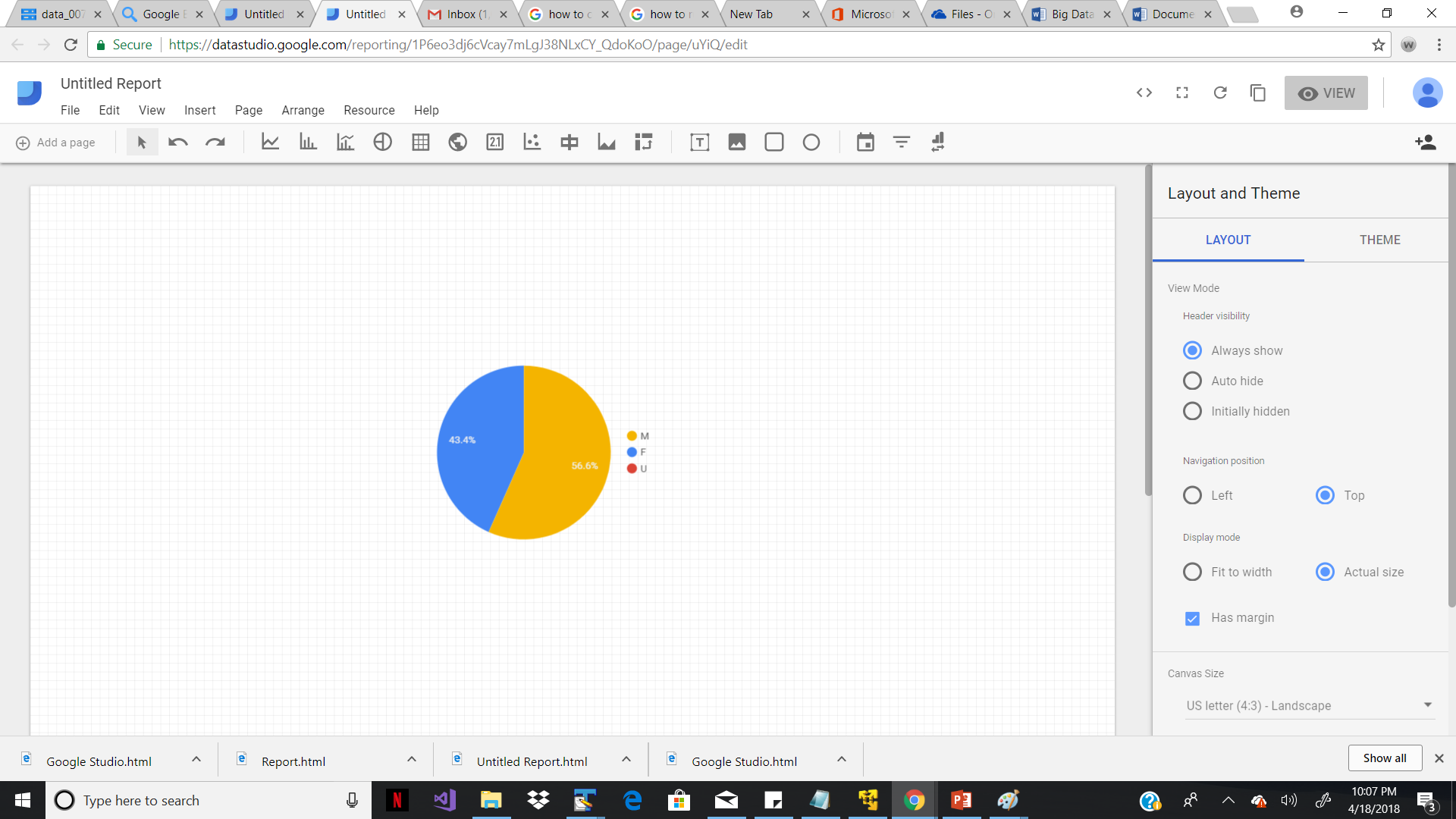
New record inserted in target:-



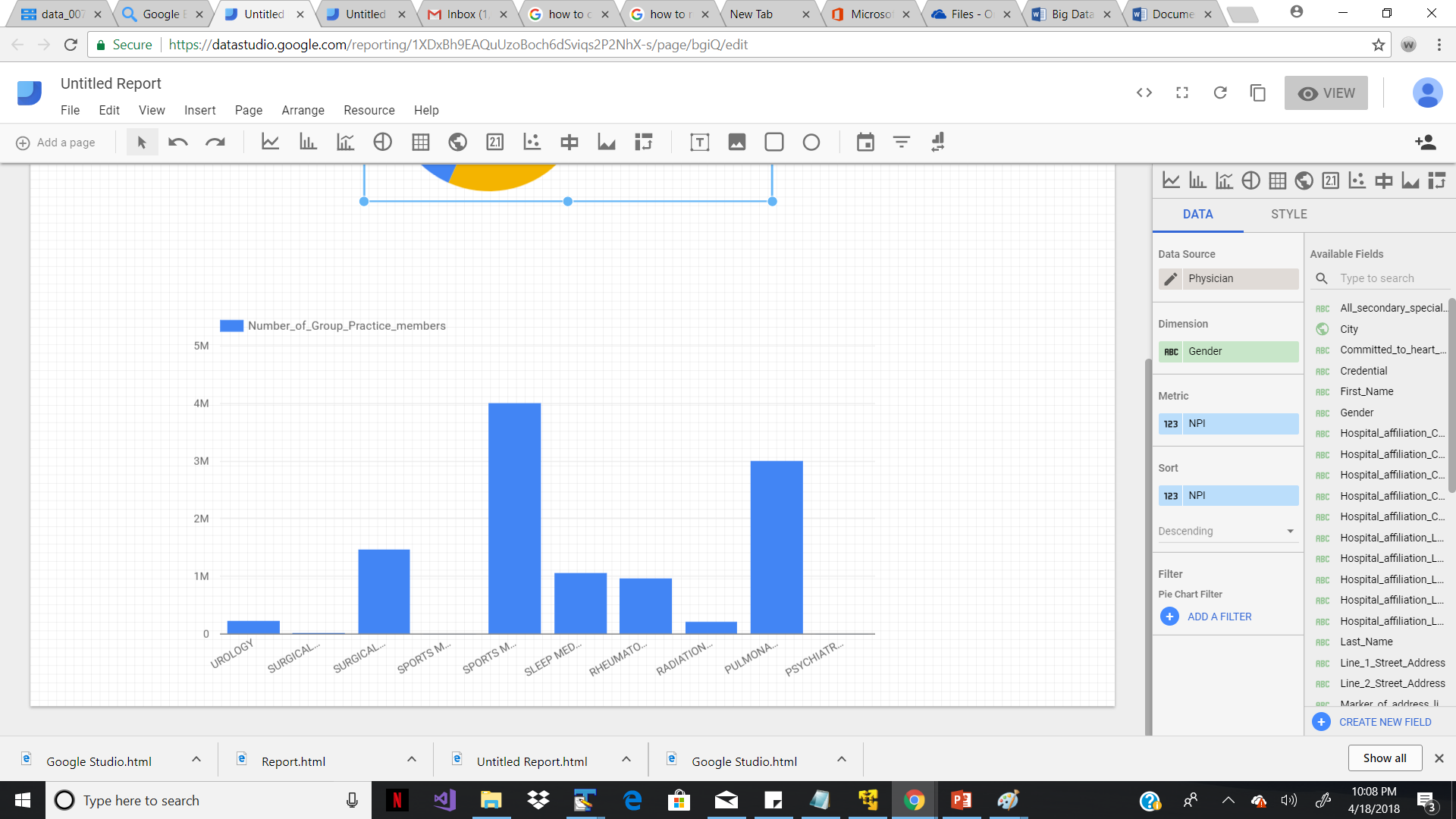
**Google Studio: -**

A) Its used to generate the report for male and female percentage among the entire population.

Sample.csv file is imported on Google Cloud Storage.



B) Depend on number of number of practice groups as well as on specialization below graph is generated.



After applying filter:-

