

**Introduction to Data Management Project Report**

(Project Semester August-December 2021)

**PROJECT REPORT ON Lok Sabha General Election** **2019**

Submitted by

**Niraj Kumar** 11915346

Programme: Bachelor of Technology

**Section: KM009**

**Course Code: INT217**

Under the Guidance of

**Sandeep Kaur: 23614**

**Assistant Professor**

**Discipline of CSE/IT**

**Lovely School of Computer Science & Engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Niraj Kumar bearing Registration no. 11915346 has completed INT217 project titled, **“Introduction to Data Management”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 12/18/2021

**DECLARATION**

I,Niraj Kumar, student of Computer Science & Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12/18/2021

Niraj Kumar

Registration No: 11915346

ACKNOWLEDGEMENT

Primarily I'd thank god for being able to complete my project with success. Then I'd like to thank my mentor Ms. Sandeep Kaur, whose valuable guidance has been the ones that helped me patch this project and make it full proof success in contribution towards the completion of this project. Last but not least I'd rather thanks to Lovely Professional University, and my parent’s inspiration, who gave me this golden opportunity to learn many new things, to learn other aspects of life.

Niraj Kumar

11915346

**CONTENTS:**

|  |  |  |
| --- | --- | --- |
| So.N | Title | Page. no |
| 1. | Introduction | 5 |
| 2. | Objectives/Scope of the Analysis | 6 |
| 3. | Source of dataset | 7 |
| 4. | ETL Process | 8 |
| 5. | Analysis of dataset | 16 |
| 6. | List of Analysis with results | 28 |
| 7. | Final Dashboard | 29 |
| 8. | Bibliography | 31 |

**INTRODUCTION**

* Data management is important because the data your organization creates is a very valuable resource.
* The last thing you want to do is spend time and resources collecting data and business intelligence, only to lose or misplace that information.
* In that case, you would then have to spend time and resources again to get that same business intelligence you already had.
* And on that data analysis is carried out which show visualization of our problems in efficient way.
* Data Analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision- making.
* This project is based on such data analysis on General Election 2019.

**OBJECTIVES/SCOPE OF ANALYSIS**

After analysis of the dataset, the aim of this project is to give answer of given objectives in easy way:

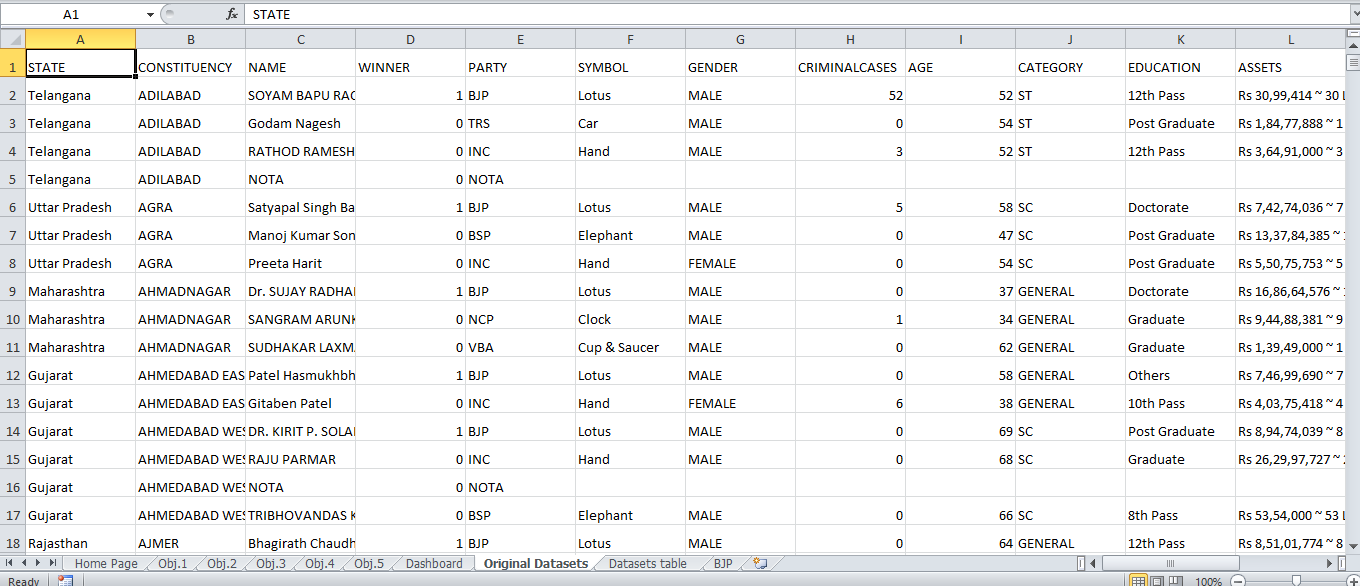
1. **Total number of seats won by party in the state.**
2. **Data analysis of total percentage of votes gain by party Constituency wise.**
3. **Total number of male and female Candidate State wise.**
4. **Party wise Criminal Cases of candidate.**
5. **Qualification of candidate.**

**SOURCE OF DATASET:**

Source of dataset: Kaggle The dataset is based on lok sabha general election 2019. The columns included in the dataset are given below:

* State
* Name of Sate
* Constituency List of Constituency
* Name
* Name of Candidate
* Winner
* Winner of candidate
* Party
* Name of parties
* Symbol
* Symbol of Parties
* Gender
* Gender of Candidate
* Criminal Case
* Criminal Case over Party of Candidate
* Age
* Age of Candidate
* Categories
* Caste Categories
* Education
* Qualification of Candidate
* Assets
* Assets of Candidate
* Liabilities
* Liabilities by Candidate
* General Votes
* Votes by normal people
* Postal Votes
* Votes by government workers
* Total Votes
* Total votes gain in constituency
* Over total elector of constituency
* Percentage of votes over total elector of Constituency
* Over Total Votes polled in Constituency
* Percentage of total votes polled in constituency
* Total Electors
* Total Number of elector in each Consituency

Sample of dataset with data fields is given below:

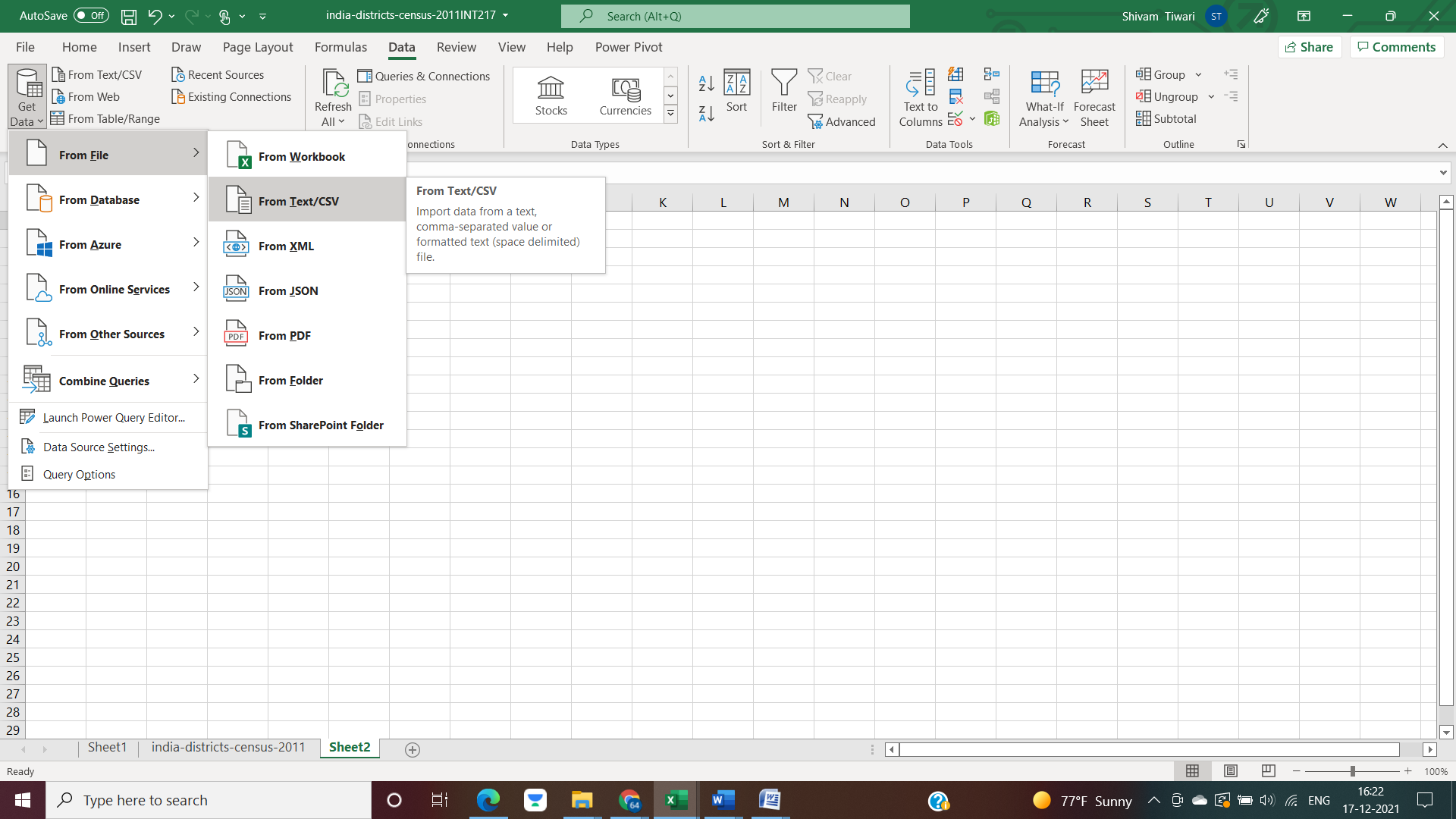


**ETL PROCESS:**

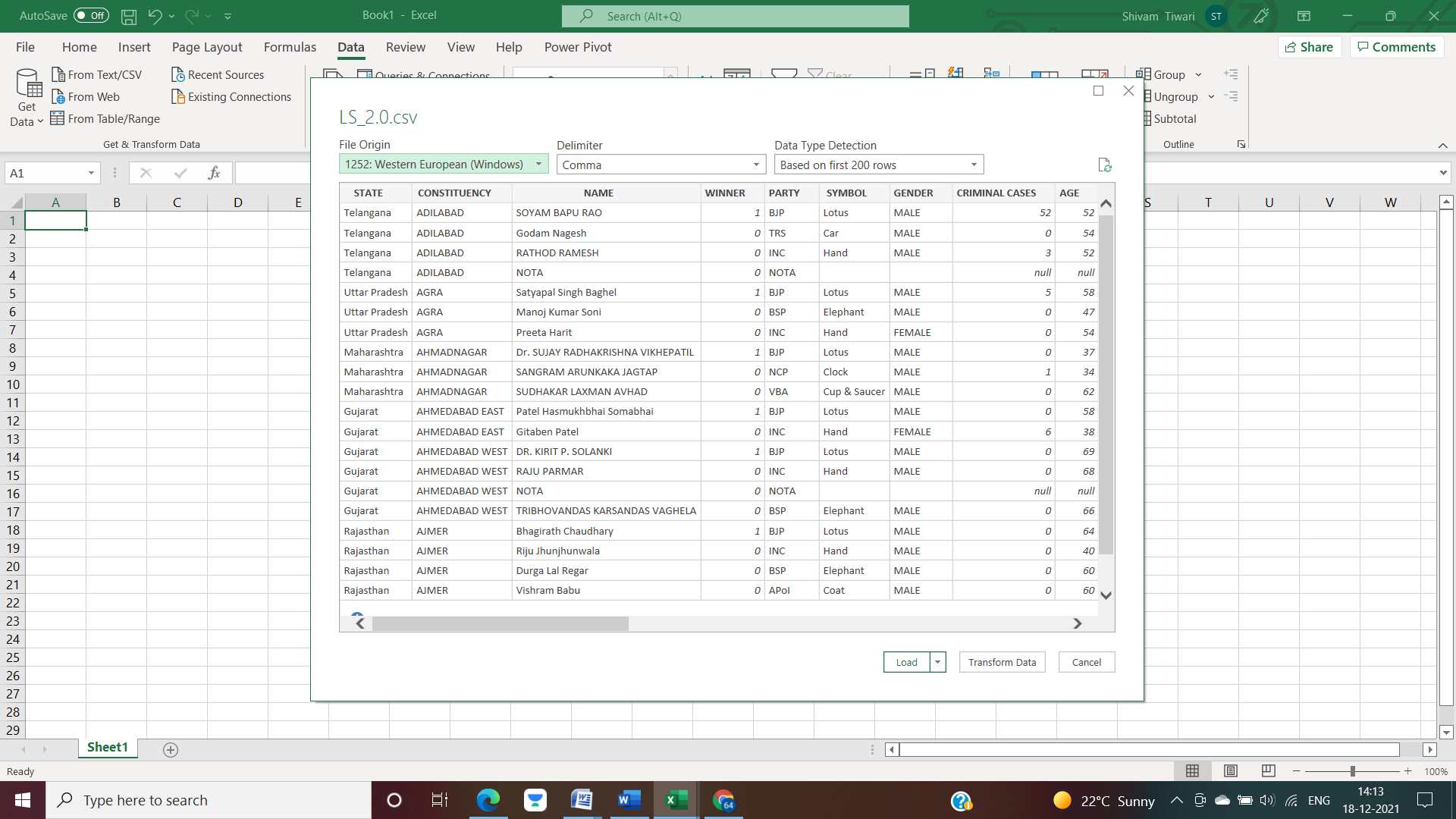
* + ETL is a process that extracts the data from different source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system.
  + Full form of ETL is Extract, Transform and Load.
  + The triple combination of ETL provides crucial functions that are many times combined into a single application or suite of tools that help in the following areas:
* Enhances Business Intelligence solutions for decision making.
* Allows verification of data transformation, aggregation and calculations rules.
* Allows sample data comparison between source and target system.
* Helps to improve productivity as it codifies and reuses without additional technical skills.

The ETL process of the Dataset is shown as below:

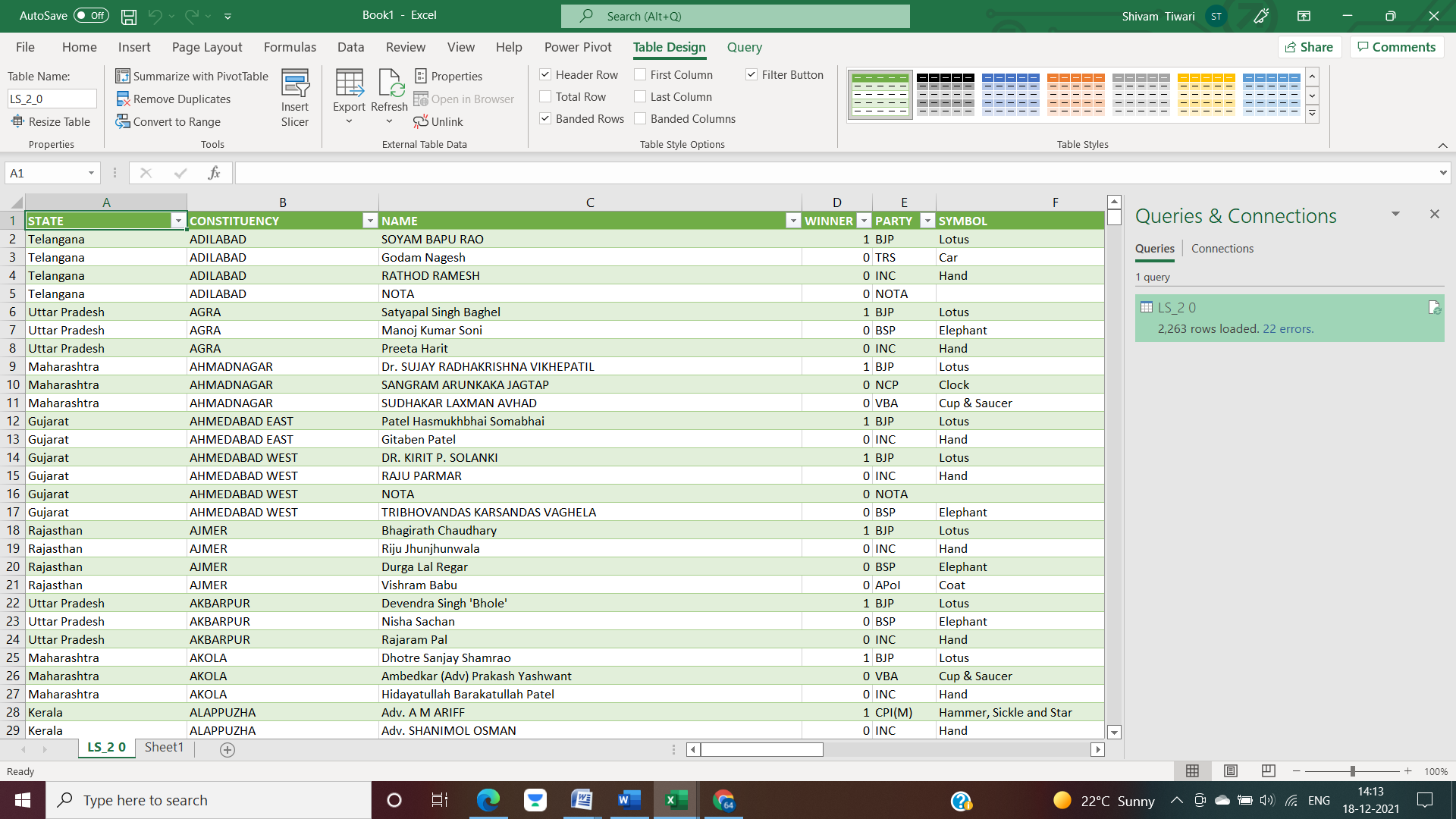
* Go to data and choose the CSV file as new query



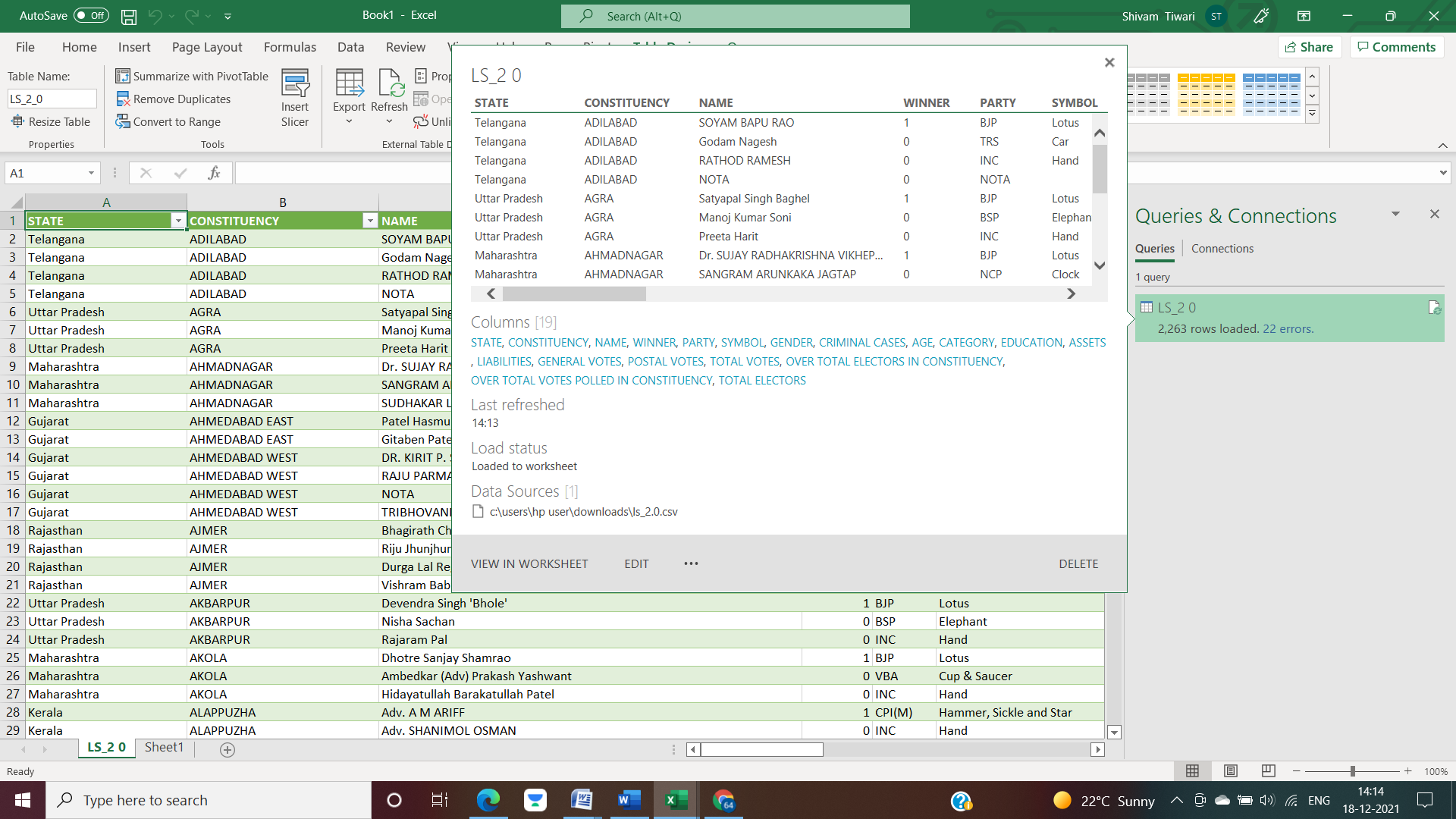
* Select the CSV file and load the data



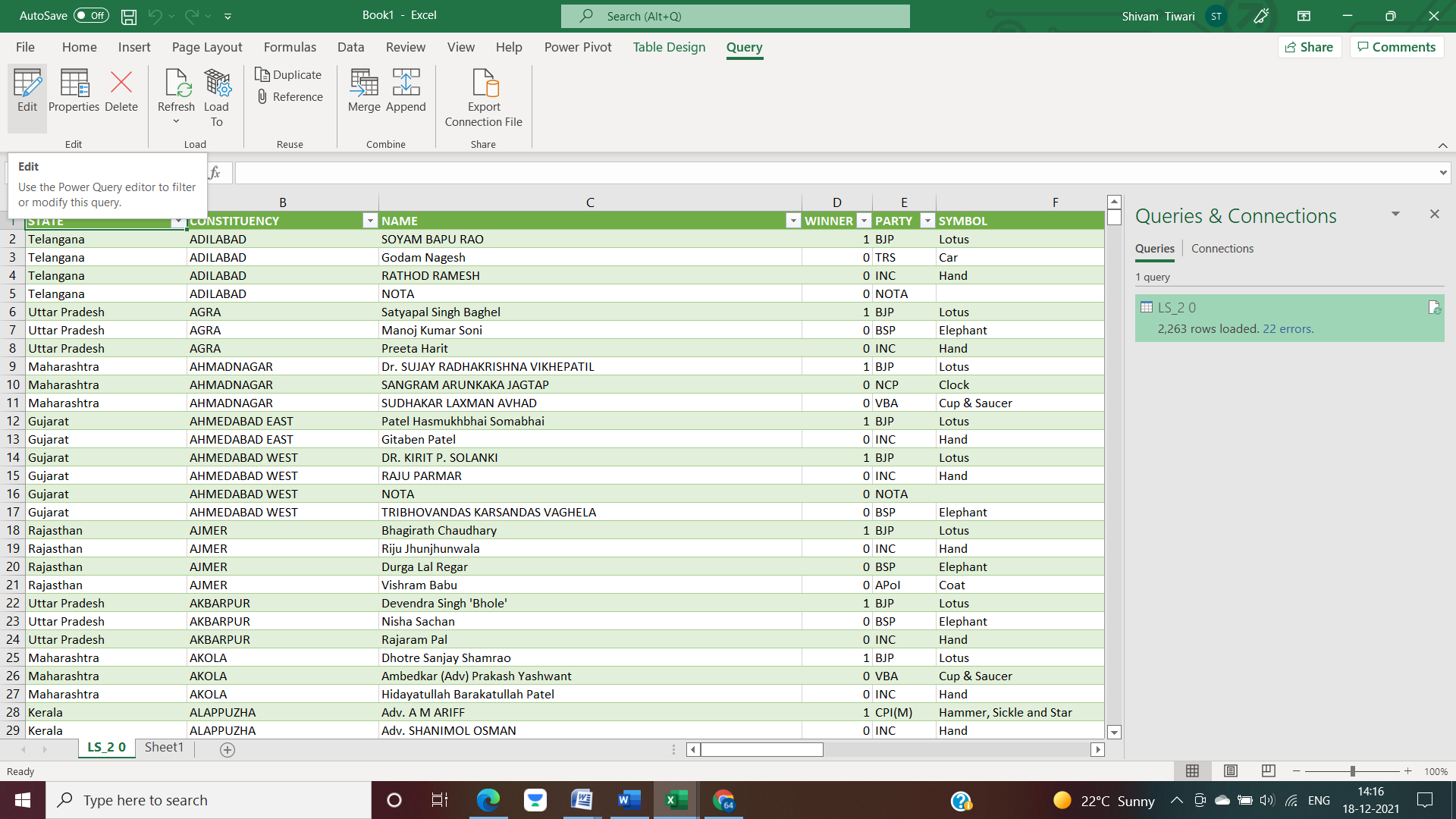
* The raw dataset will look like as shown:

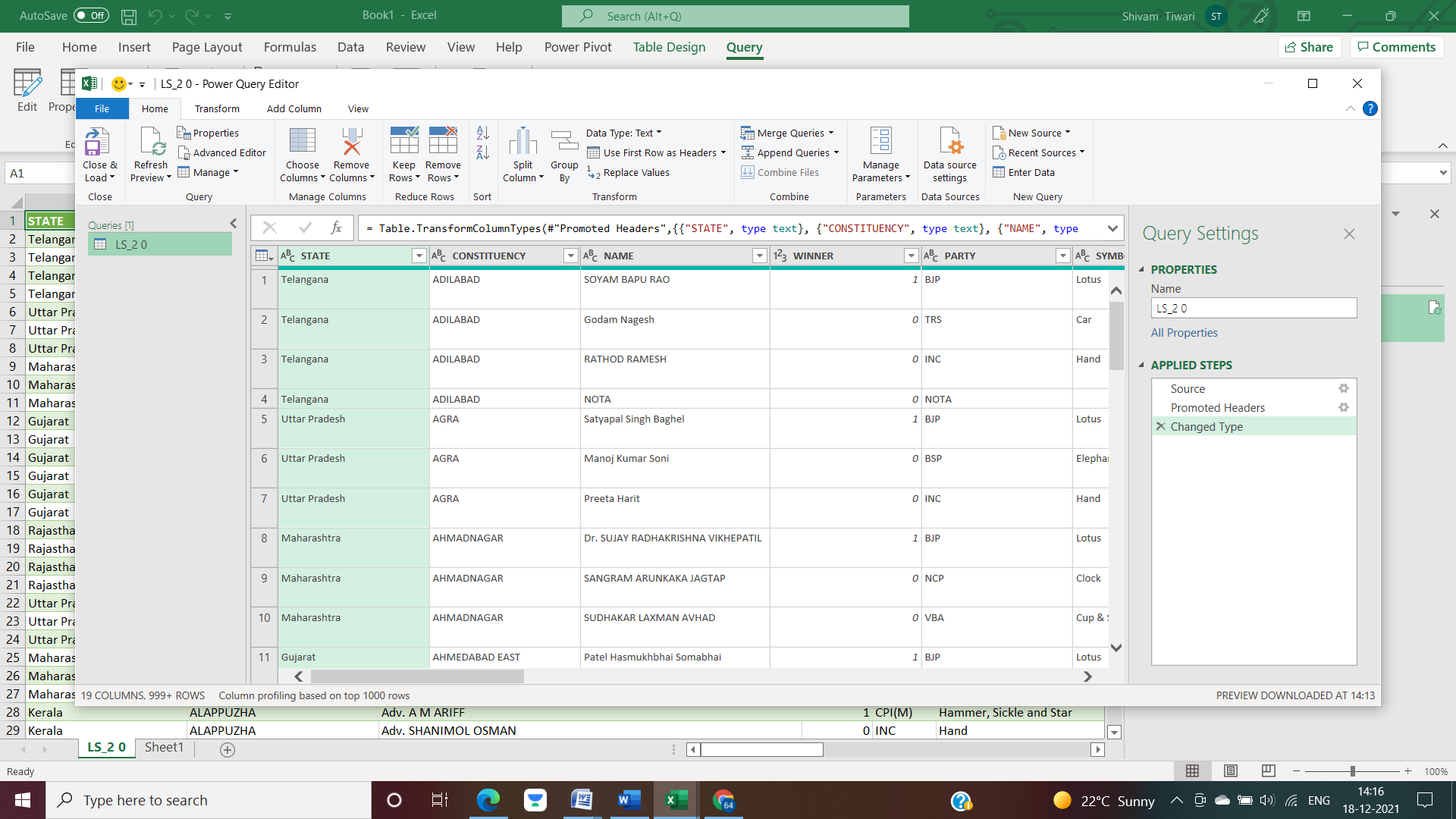


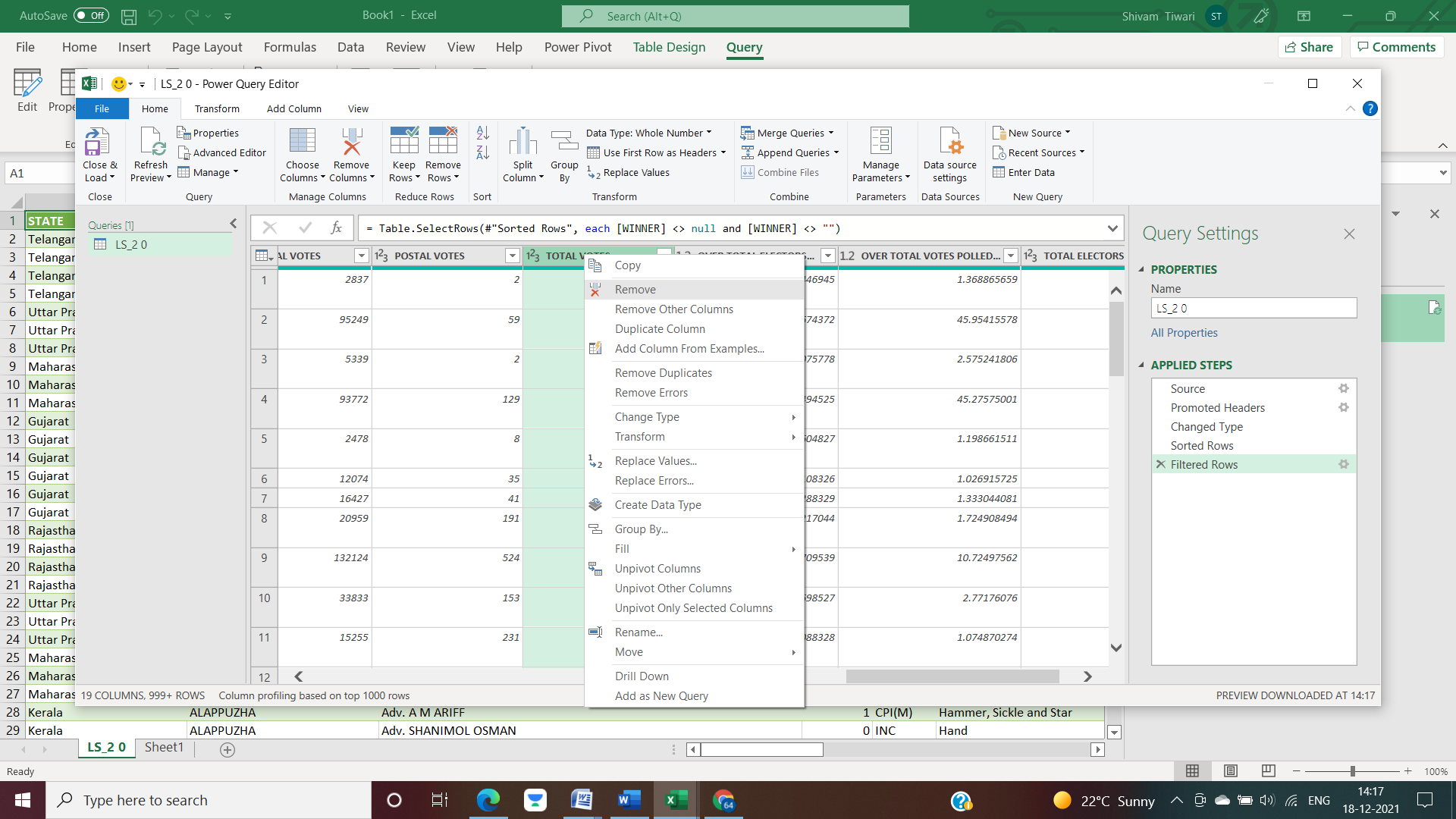
* The details of the dataset can be seen systematically.

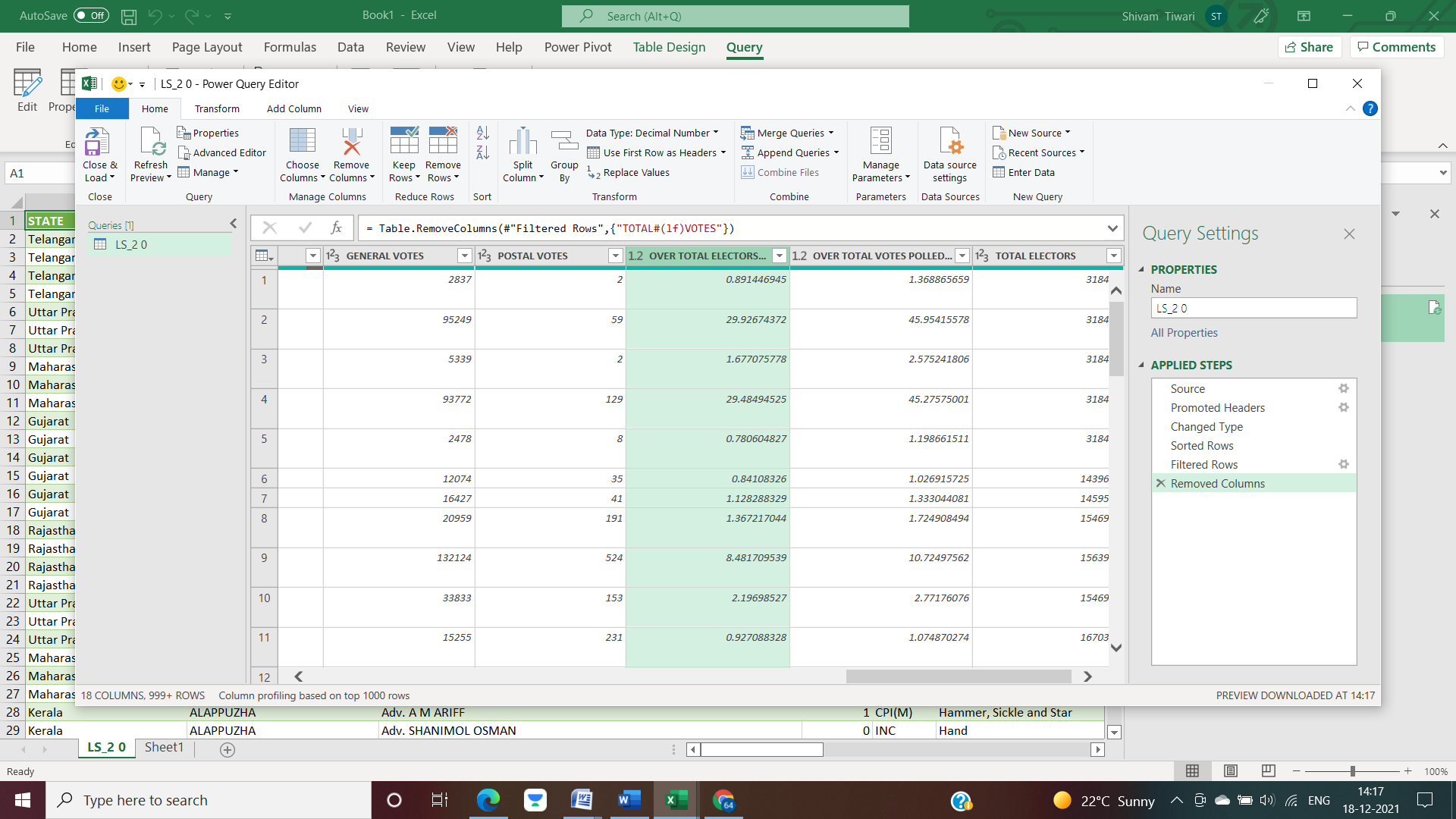


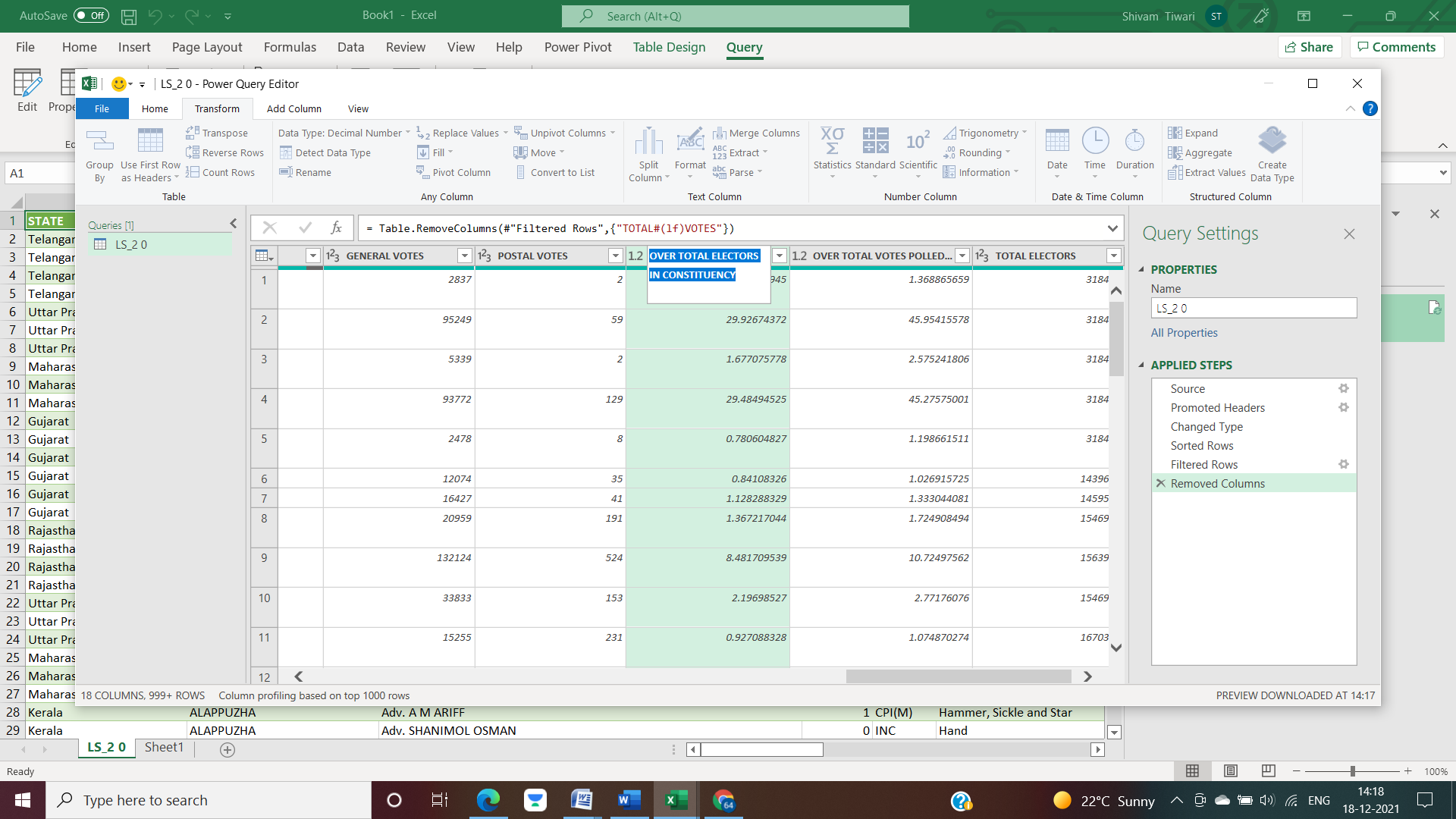
* We can change, modify and delete dataset according to the need of our analysis as well.

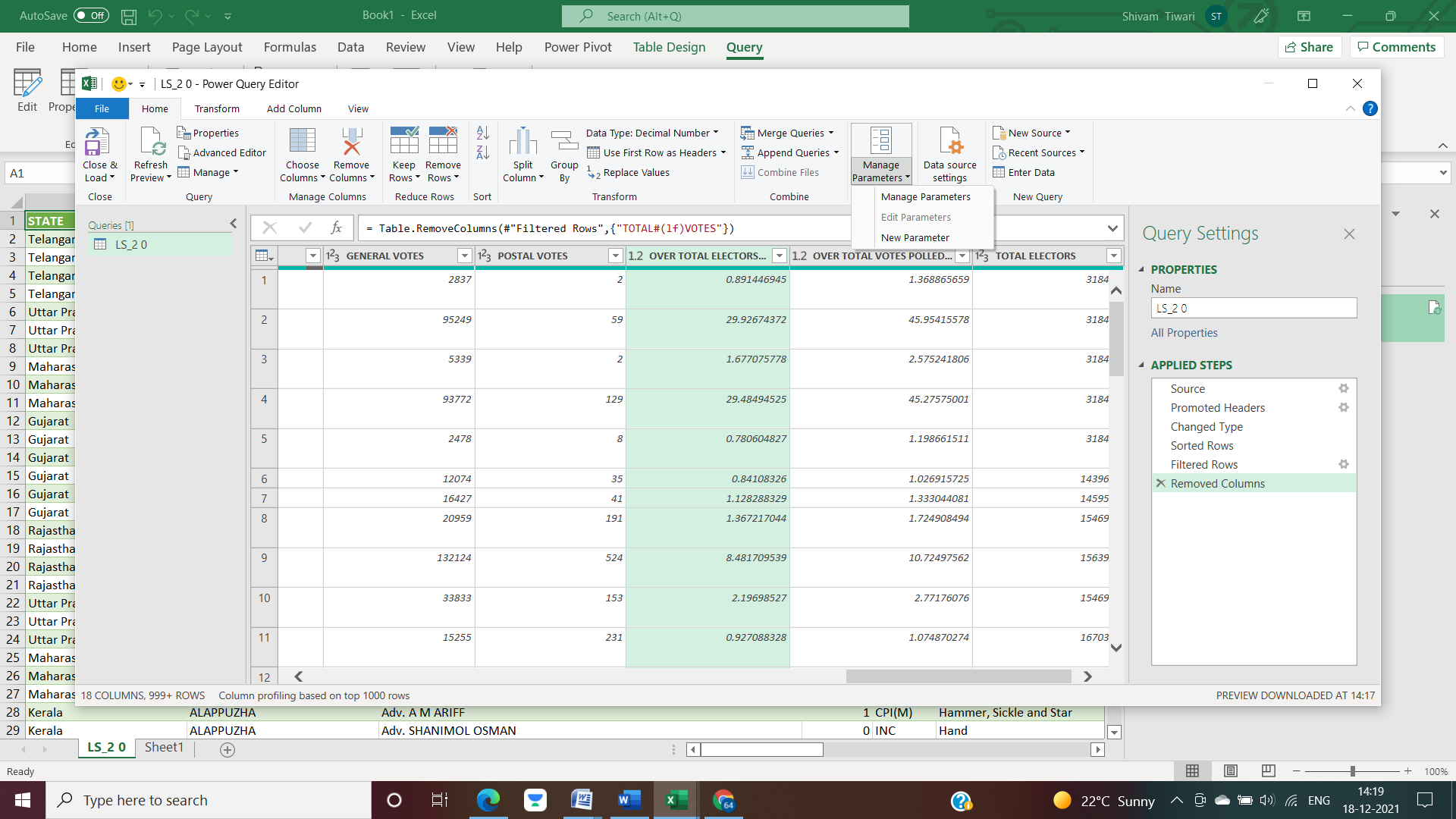




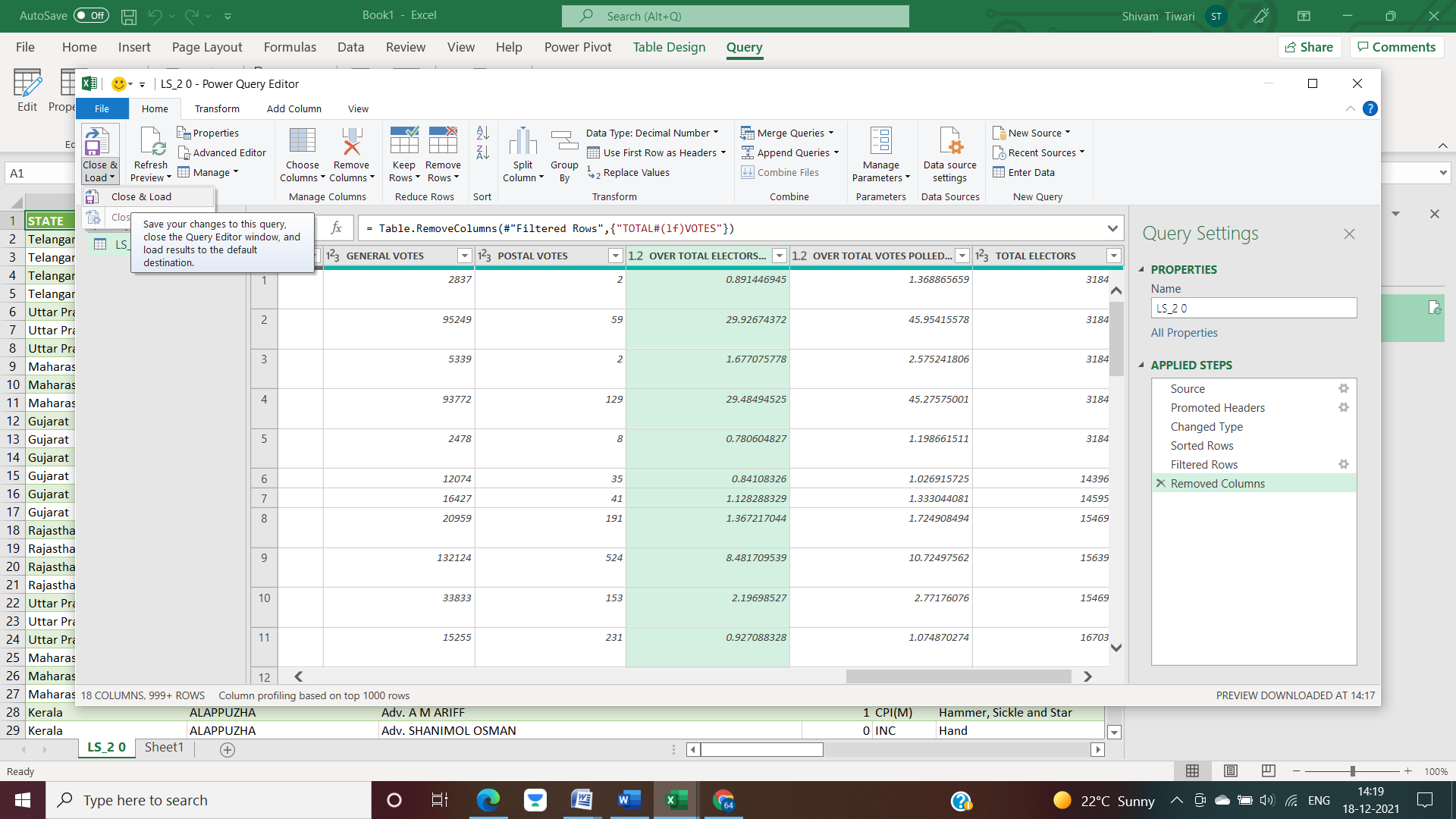


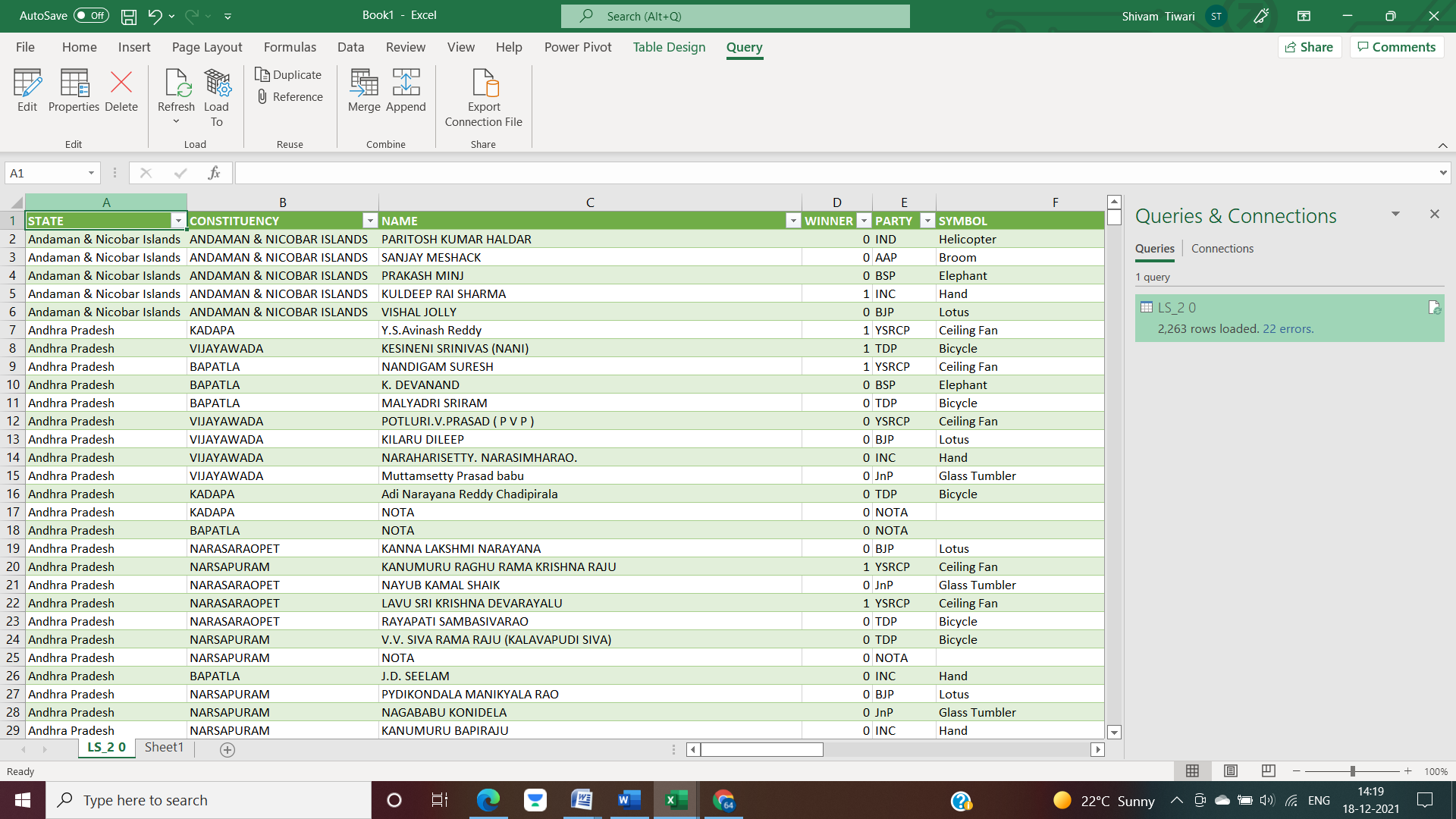






* After performing all these transformations, we can simply load it to worksheet.





**Analysis on dataset:**

1. **Total number of seats won by party in the state.**

**Introduction**

* By performing this analysis, We will get the result of seats won by party in different states.

**Description**

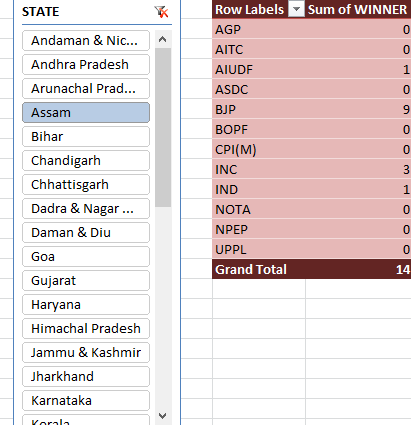
* The analysis is based on State, Party and winner candidates.

**Specific requirements, functions and formulas**

* To get Result, used Power Pivot, Slicer, Pivot table and Bar graph for visualize the data.

**Analysis results**

* When States are selected in slicer box and party add in row label.



**Visualization**

* By using Bar graph , Visualize maximum number of seat won by any party in Selected state.

1. **Data analysis of total percentage of votes gain by party Constituency wise.**

**Introduction**

* By Performing this Objectives , We will get the Vote percentage gain by parties.

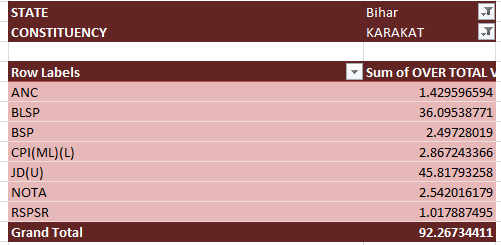
**Description**

* The analysis is based on States, Constituency , Parties and Vote percentage.

**Specific requirements, functions and formulas**

* To get Result, used Power Pivot, Pivot table and Pie Chart for visualize the data.
* Using formula, VLOOKUP(=VLOOKUP(J23,A8:E43,5,0), DATA VALIDATION

**Analysis results**



**Visualization**

1. **Total number of male and female candidate Statewise.**

**Introduction**

* By performing this Objectives , We will get the gender of candidates in election 2019.

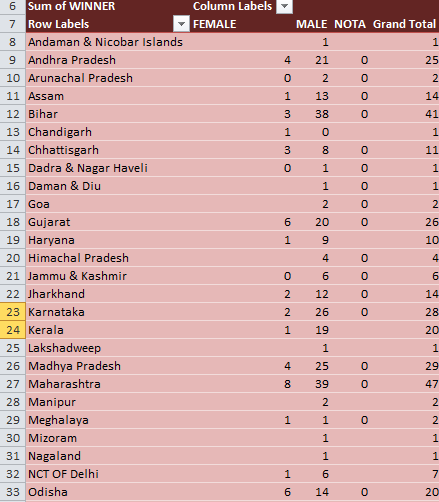
**Description**

* The analysis is based on States, Gender , Winner male/female Candidates.

**Specific requirements, functions and formulas**

* To get Result, used Power Pivot, Pivot table and Bar graph for visualize the data.
* Using formula, VLOOKUP(=VLOOKUP(J23,A8:E43,5,0), DATA VALIDATION

**Analysis results**



**Visualization**

1. Party wise Criminal Cases of candidate.

**Introduction**

* By performing this Objectives , We will get the party wise Criminal Case of Candidates in election 2019.

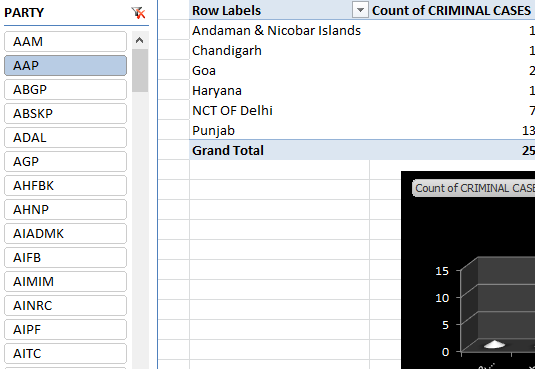
**Description**

* The analysis is based on States, Parties , Criminal case against Candidates.

**Specific requirements, functions and formulas**

* To get Result, used Power Pivot, Pivot table and Bar graph for visualize the data.

**Analysis results**



Visualization

1. Qualification of candidate

**Introduction**

* By performing this Objectives , We will get the qualification of candidates in election 2019.

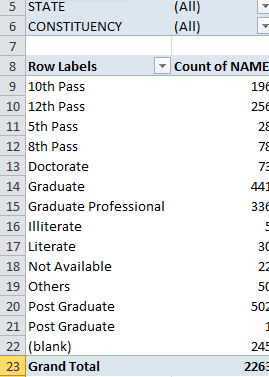
**Description**

* The analysis is based on States, Parties, Constituency , Qualification of Candidates.

**Specific requirements, functions and formulas**

* To get Result, used Power Pivot, Pivot table and Bar graph for visualize the data.

**Analysis results**

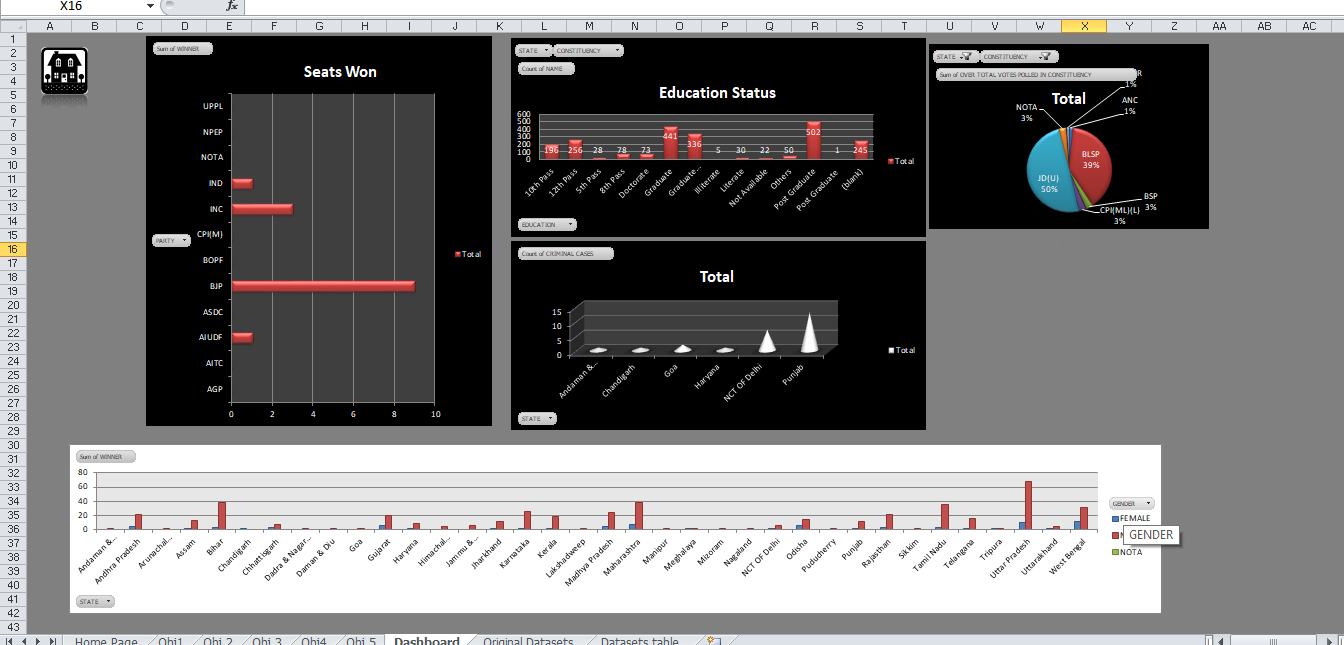


Visualization

**List of Analysis with results**

1. **Total number of seats won by party in the state.**
2. **Data analysis of total percentage of votes gain by party Constituancy wise**
3. **Total number of male and female candidate Statewise**
4. **Criminal case of Candidate party wise**
5. **Qualification of candidate**

**FINAL DASHBOARD:**

****

**BIBLIOGRAPHY:**

Dataset source: https://www.kaggle.com/prakrutchauhan/indian-candidates-for-general-election-2019

Information about Data Management: https://www.blue-pencil.ca/what-is-data-management-and-why-it-is-important/