**UTSA AND ASPIRATIONAL GRADUATION RATES DASHBOARD DOCUMENTATION**

**I – Data Source**

IPEDS Data center

**II – Get The Dataset**

Hyperlink: <https://nces.ed.gov/ipeds/use-the-data>

- Click on Compare Institutions

- Select Use final release data

- For Comparison Institution, type in The University of Texas at San Antonio and **Select**

- For Comparison Group, select By Names or UnitIDs, type in the Institution Name, and Select. Repeat the process until the end of list of comparison group.

- Click Continue

- In **Graduation Rates** category, Select ‘Graduation rate data within 200 percent of normal time – all institutions’

- Select ‘Four-year institutions – Bachelor’s degree-seeking cohort’

- Select years wanted for the report

- Select ‘4-year Graduation rate – bachelor’s degree within 100% of normal time’

- Select ‘6-year Graduation rate – bachelor’s degree within 100% of normal time’  
- Select **Continue** twice

- Select ‘View on screen’ or ‘Download in comma separated format’ from the third row of radio buttons

- Select Continue

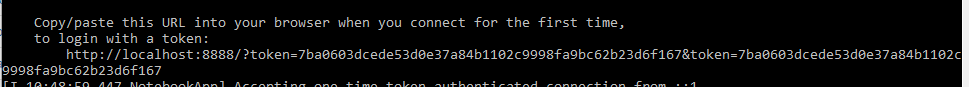
- Rename the downloaded csv file to PeerData and move it to the same directory where the dashboard is

**III – Quick Guide to Customize Dataset**

- Move the dataset into the same folder with Seperator.ipynb  
- From Windows symbol, select **Anaconda3 (64-bit) -> Jupyter Notebook** (if you have not had Anaconda installed, here’s the link: <https://www.anaconda.com/download/> (Select the version for Python 3)

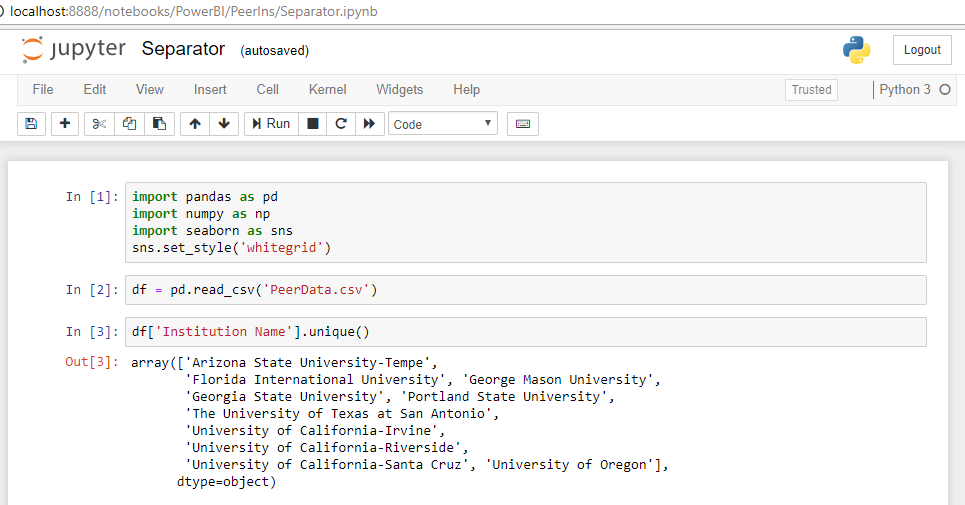
- A picture like this should pop up as an Internet Explorer or Google Chrome window

- If the window does not appear, click on the Jupyter Notebook symbol from Windows tool bar and **select** the similar hyperlink like below, and **open it from a web browser**.



- Navigate to where Separator.ipynb is. If you’re having trouble finding this file from Jupyter Notebook, it should be similar to how you get to it from regular Windows Navigator.

- Open Separator.ipynb. A window like this should appear:



- Select **Kernel -> Restart & Run All**

- Scroll down to the last 2 cells, select the 40th or 41st if you want to get the data in excel or csv format

- Select all of the content within the selected cell and Ctrl + /

- Shift + Enter

- The new excel or csv files should now be in the same directory with Separator.ipynb

- Remember to delete the first columns (index columns) on the two sheets for Power Bi to pick up the first row as parameters row

**IV. Step-by-Step Jupyter Notebook Tutorial (Cell by cell)**

**A. Introduction**

- About Jupyter Notebook: <http://jupyter.org/about>

- About Python libraries for data science and tutorials:

a. Pandas

Tutorial: <https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/>

Documentation: <https://pandas.pydata.org/pandas-docs/stable/>

b. Numpy

Documentation: <https://docs.scipy.org/doc/>

c. Seaborn

Tutorial: <https://www.youtube.com/playlist?list=PLE50-dh6JzC7X8VFX40yoIXnhctF2bR8F>

Documentation: <https://seaborn.pydata.org/>

**B. Tutorial (Cell by cell)**

1. Importing necessary libraries, I used pd, np and sns because they are universal aliases for these libraries.

2. Using method ‘read\_csv’ from library pandas to read a file call ‘PeerData.csv’ as a Pandas Dataframe object called ‘df’.

Your csv should be in the same directory with your Jupyter Notebook because it would be easier to read the csv file.

3. Check how many universities are there in this dataset.

4. Check the head of the dataset.

5. Check the information of the dataset.

6. Drop irrelevant columns (ID and Unnamed: 12)

The column Unnamed: number of parameters will be attached to the csv file every time a csv file is downloaded from IPEDS and read by pd.read\_csv.

7. Create a list of new names for columns of the dataset.

8. Rename the dataset’s columns based on the list created previously.

9. Check the dataset’s info again.

10. Check the dataset as a whole.

11. Create a new dataframe object from df that does not contain UTSA and drop ‘Institution Name’ column for data manipulation. This object is called ‘peers’.

12. Create a Python dictionary object named ‘medians’ with key as ‘Institution Name’ and value as ‘medians’.

13. Extract the median from each column of ‘peers’ and assign to ‘medians’.

14. Display the result.

15. Transform ‘medians’ into a Python dataframe object and called it ‘median\_df’.

16. Create a new dataframe object from df that contains UTSA’s data called ‘UTSA’.

17. Display the object.

18. Append UTSA dataframe object to median\_df.

19. Display median\_df.

20. Transpose median\_df by using built-in ‘melt’ method from Pandas library and call the new object ‘test\_df’.

21. Display test\_df.

22. Sort values of test\_df based on ‘Institution Name’ column.

23. Display test\_df

24. Create new column ‘Median’ for test\_df dataframe that only get values that are ‘Median’ from ‘Institution Name’ column.

25. Create new column ‘The University of Texas at San Antonio’ for test\_df dataframe that only get values that are ‘The University of Texas at San Antonio’ from ‘Institution Name’ column.

26. Display test\_df.

27. Create new dataframe object call ‘median’ that only takes ‘Year – Type’ and ‘Median’ column

28. From ‘median’, drop rows contain null values.

29. Create new dataframe object call ‘UTSA’ that only takes ‘Year – Type’ and ‘The University of Texas at San Antonio’ column.

30. From ‘UTSA’, drop rows contain null values.

31. Reassign median\_df to a new dataframe by merging ‘median’ and ‘UTSA’ (same mechanism in SQL).

32. Sort values of ‘median\_df’ by ‘Year – Type’ column.

33. Display ‘median\_df’

34. Create a new dataframe object, called ‘four\_year’ that only contains rows that have data about graduation in 4 years.

35. Create a new dataframe object, called ‘six\_year’ that only contains rows that have data about graduation in 6 years.

36. Create a list of new names for ‘four\_year’ dataframe object.

37. Rename columns of ‘four\_year’

38. Create a list of new names for ‘six \_year’ dataframe object.

39. Rename columns of ‘six \_year’

40. To uncomment this cell, select all of the content of this cell, select Ctrl + / and select Shift + Enter to run the cell. This cell will create an Excel file called ‘UTSA and Peers.xlsx’

41. To uncomment this cell, select all of the content of this cell, select Ctrl + / and select Shift + Enter to run the cell. This cell will create two csv files named ‘4-year.csv’ and ‘6-year.csv’

If the cell can’t be run, delete ‘UTSA and Peers.xlsx’ in the directory and run the cell again