Standardized Test Analysis

- Maryam Amjad

Problem Statement :

To compare and study the participation rate and scores of ACT and SAT examinations and to give a better idea to the participants about the ACT and SAT results and competition in recent years across all the states in USA.

Datasets used

- act_2018.csv - Gives ACT scores for the year 2018 across all the states in USA.
- act_2019.csv - Gives ACT scores for the year 2019 across all the states in USA.
- sat_2018.csv - Gives SAT scores for the year 2018 across all the states in USA.
- sat_2019.csv - Gives SAT scores for the year 2019 across all the states in USA.

Final dataframe:

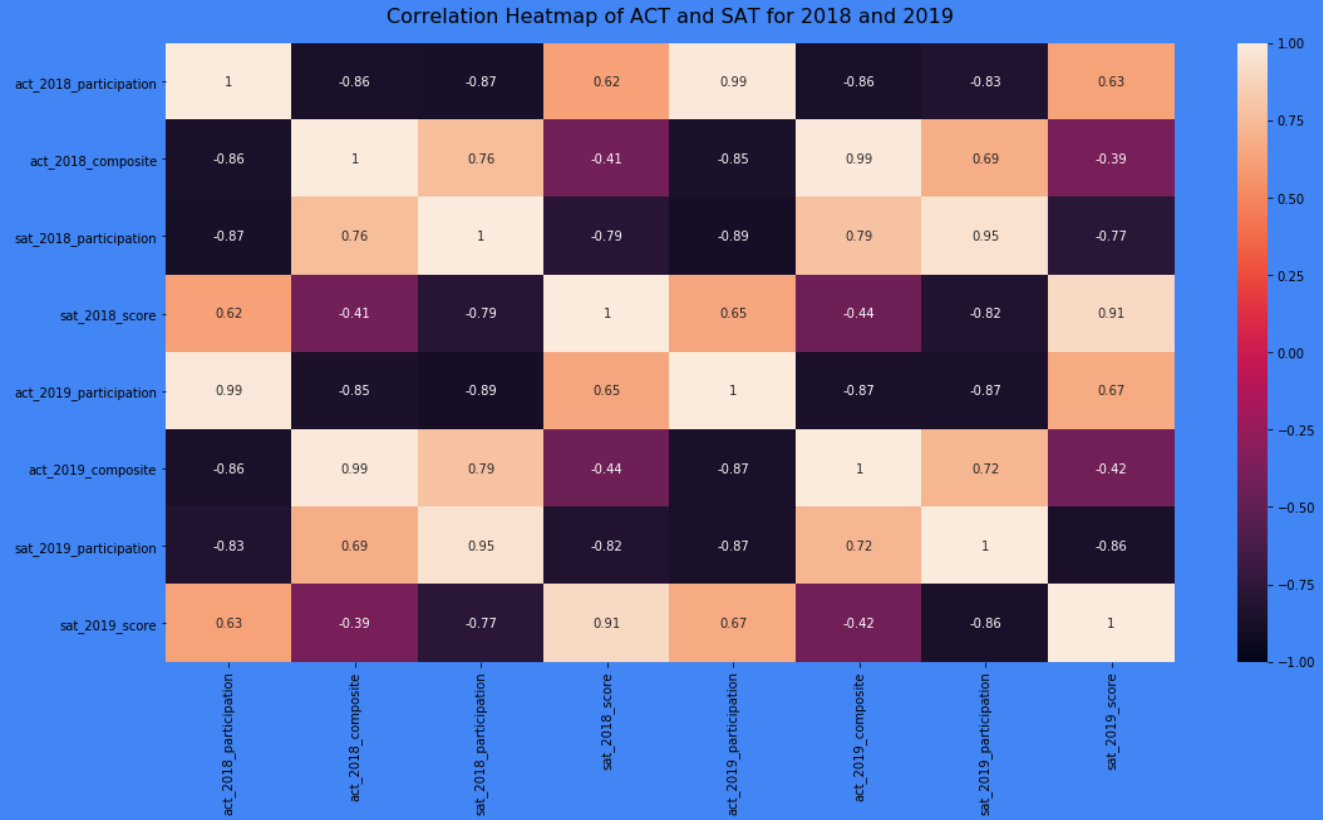
Feature	Type	Dataset	Description
state	object	final_data	name of the state in USA
act_2018_participation	float	final_data	ACT-2018 particiption rate (0.5 represents 50%)
act_2019_participation	float	final_data	ACT-2019 particiption rate (0.5 represents 50%)
sat_2018_participation	float	final_data	SAT-2018 particiption rate (0.5 represents 50%)
act_2019_participation	float	final_data	SAT-2019 particiption rate (0.5 represents 50%)
act_2018_composite	float	final_data	ACT-2018 composite score
act_2019_composite	float	final_data	ACT-2019 composite score
sat_2018_composite	float	final_data	SAT-2018 score
sat_2019_composite	float	final_data	SAT-2019 score

Data Cleaning

- Replacing NULL with appropriate values
- Removing redundancy
- Dropping unwanted columns
- Fixing data
  - Correcting inconsistent data like symbols
  - Checking Data type
- Renaming columns
- Merging dataframes
- Exporting final dataframe

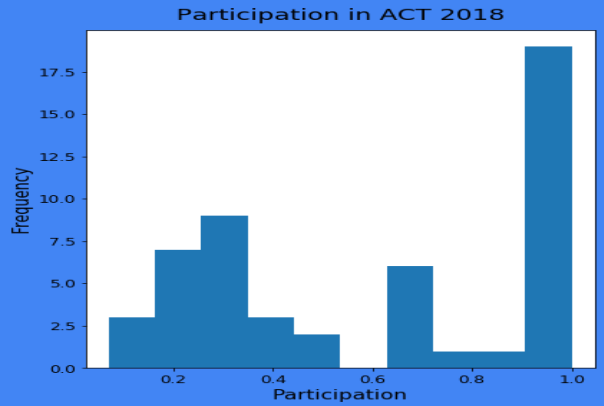
Then Dictionary of standard deviation(SD) is created and required observations are done with the help of the final dataframe.

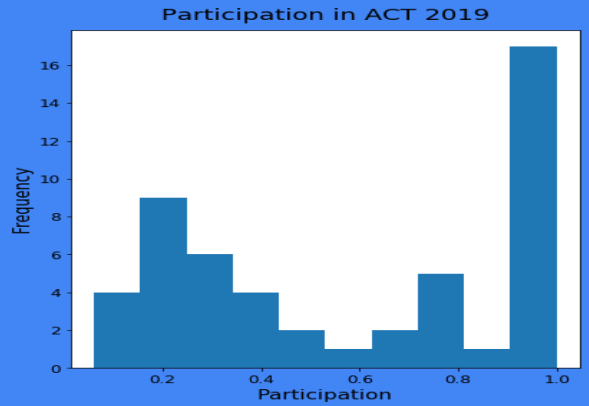
Heat Map

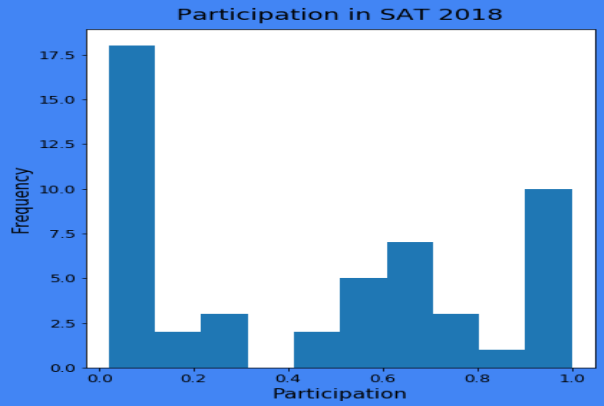


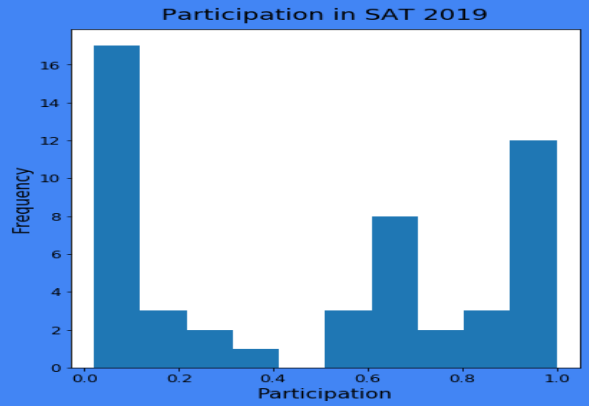
It is observed from the heatmap that most of the features shows a strong relationship with all other features.

Histogram:



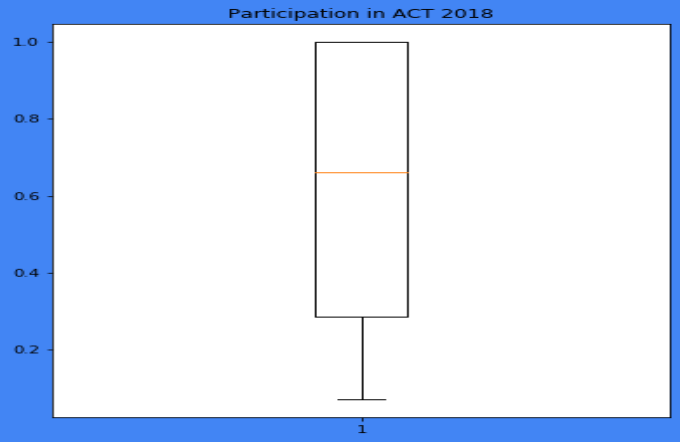


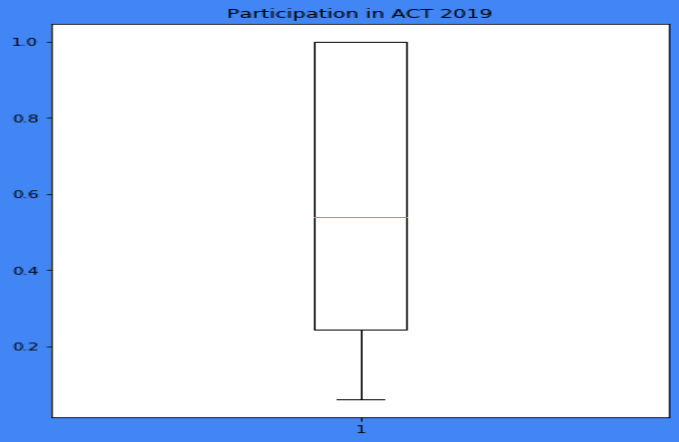


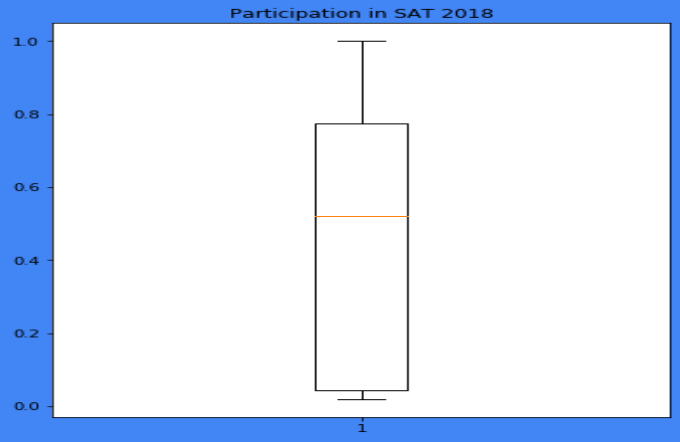


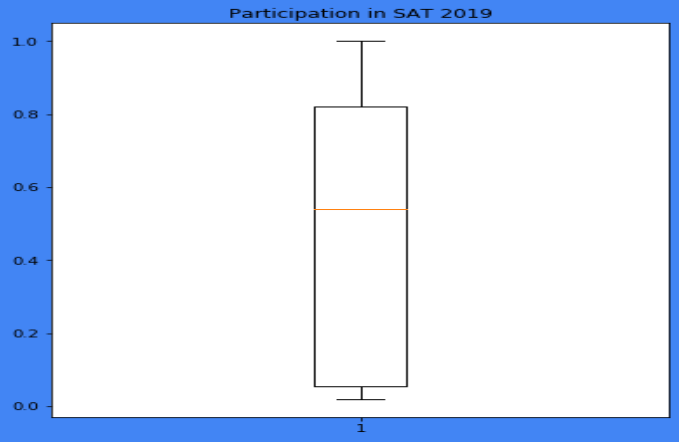
It is observed from the histogram that most of the states shows higher rate of participation in ACT and less participation rate in SAT. This pattern is observed in 2018 as well as in 2019.

BoxPlot:



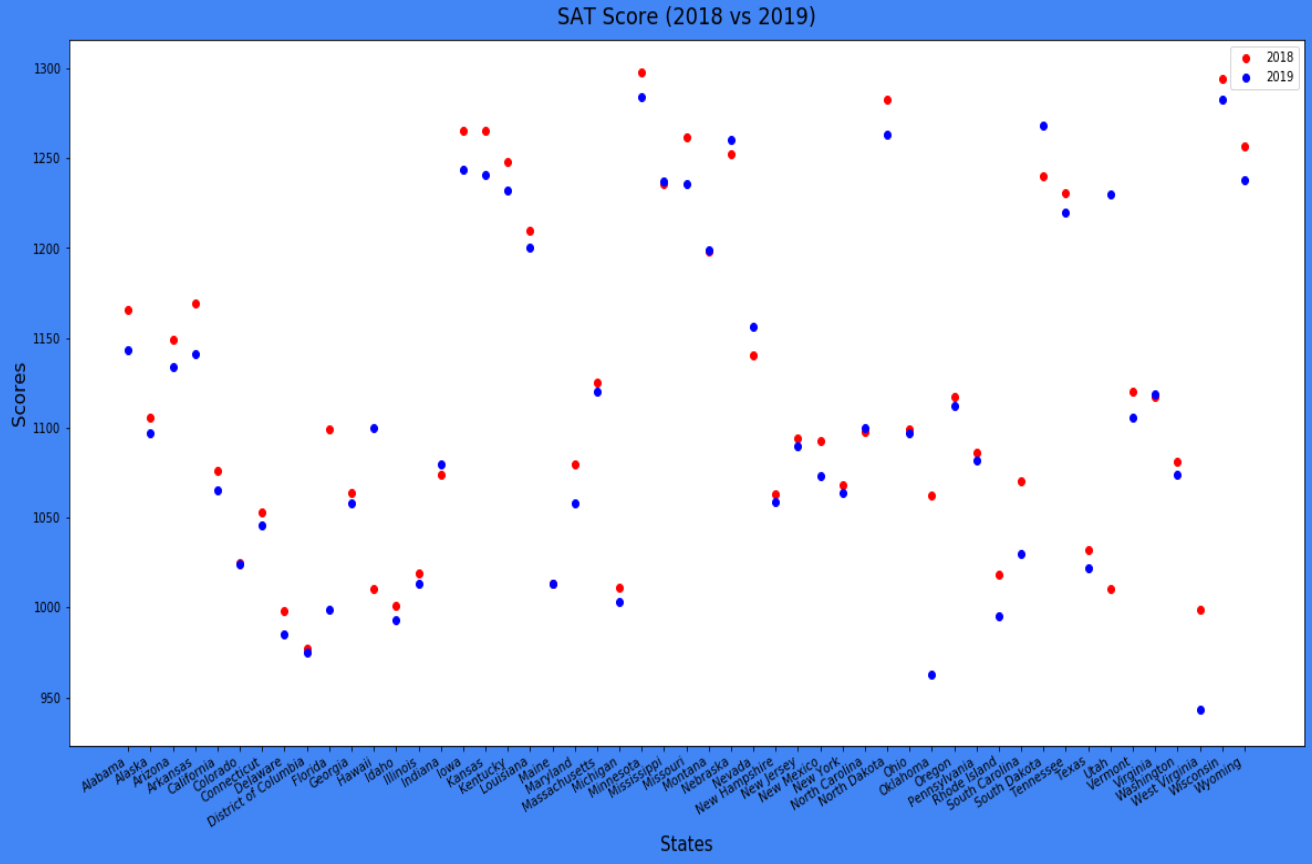






From this boxplots we can derive the respective median, 25th percentile and 75th percentile for SAT and ACT participation rate of 2018 and 2019.

Scattered Plot:



We observed from the scatter plot that most of the states shows almost similar results in ACT and SAT each year. only few states shows significant difference in results.

Conclusion

From the visualization of the data, we can observe the trend in participation rate and scores across the states of USA remains almost the same for the years 2018 and 2019. The standard deviations found previously supports the above. This gives better understanding of the recent trends in ACT and SAT in USA. The final dataset can be used to build a predictive machine learning models.

Thank you