

**PES University, Bangalore**

(Established under Karnataka Act No. 16 of 2013)

**B.Tech, Sem III**  
**Session : Aug-Dec, 2018**

**UE17CS203 – INTRODUCTION TO DATA SCIENCE****REPORT****EXPLORATORY ANALYSIS ON  
US Presidential Election 2016**

<b>DATA SET LINK :</b>	<a href="https://www.kaggle.com/benhamner/2016-us-election">https://www.kaggle.com/benhamner/2016-us-election</a>
<b>TEAM MEMBERS</b>	<b>NAME : Vihan Srinath Murthy</b> <b>SRN : PES1201701603</b> <b>EMAIL ID : vihanmurthy@gmail.com</b> <b>CONTACT NO. : 7022346024</b>
	<b>NAME : Shreyas Mavanoor</b> <b>SRN : PES1201700837</b> <b>EMAIL ID : shreyasmavanoor@gmail.com</b> <b>CONTACT NO. : 9901917506</b>

# DATASET

Exploratory Data is the process of analysing data sets to summarise their main characteristics. This is often accompanied by with various visual methods of representing data. It involves analysis and chalking out reasons or results.

Bar Graphs, Pie charts are among the multiple resources used to provide easily understandable insights into the data.

# ABSTRACT

2016 US Election is a small dataset available on [kaggle.com](https://www.kaggle.com) which contains information pertaining to the last United States elections which resulted in a Republican victory after two continuous Democrat terms.

The dataset contains eight columns and spans across twenty-four thousand rows. The premier column describes the concerned state's name. The United States has a total of fifty states. The following column mentions the state's abbreviation, that is used to denote the state's name within two alphabets.

The third column tells us the name of the county in which the electoral poll took place. There exist 3142 counties in the United States Of America. The fourth column represents Federal Information Processing Standards which uniquely identifies counties equivalent in the United States. The FIPS number is a 5 digit number

The fifth column provides information with respect to the two major political parties in the United States, the Democrats and the Republicans, while the following column, names the candidate standing for elections in the region and hence their votes contributed from said region. The last column mentions the fraction vote description of each county

# INTRODUCTION

The US presidential election of 2016 was an event that had left the entire planet in wonder and question. It was spearheaded by Donald Trump Jr. alongside the Republicans and Hillary Clinton backed by the Democrats.

While the latter candidate was all for women's rights, climate change issues and bringing in fresh foreign talent to the country, the former was pushing capitalist economic policies, reduction in imports and lax gun owning rules.

The elections play a huge role in the standing of the international market. The stock market is seeing its longest decline in years in the run-up to the 2016 General Election. Analysts were blaming concerns about the election for the short-term economic woes. As the election nears, investors worldwide begin to take less risk, and the Volatility Index or 'the fear gauge' tends to jump up, because people feel the market might suddenly change and become unstable after new policies launched by the brand-new president.

Also, the new commander in chief of the United States Armed Forces has the power to decide which countries to ally with and whether or not to pursue the raging war in the middle east.

# PROCESSING

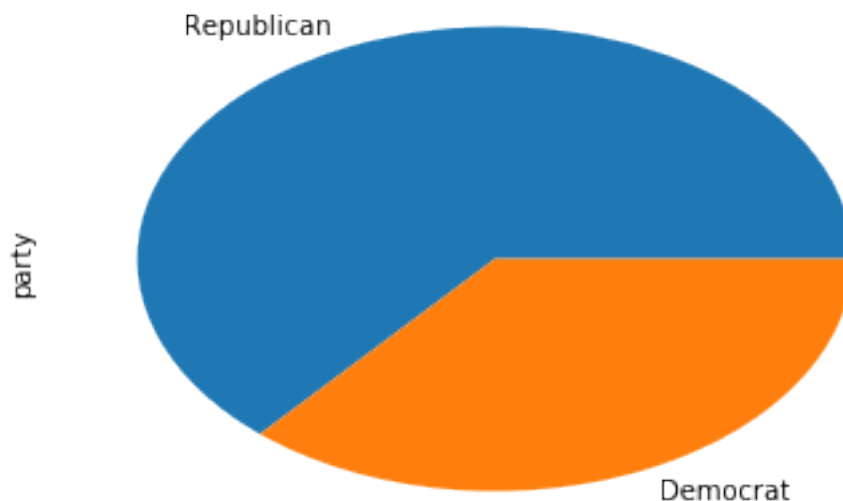
Steps involved in the cleaning of data:

- Checking the state abbreviations column to ensure all the rows were of type string and had length of not more than two
- Ensuring the Federal Information Processing Standards (FIPS) number was only a numerical value and did not contain any other data type
- Making sure the vote fractions of various candidates in a specific county was lesser than or equal to one
- Checking the names of political parties

# Exploratory Analysis

The United States Elections has 2 major political parties that compete for the presidential seat and senate, the Republicans and the Democrats.

Their respective representation among the polling crowd for the 2016 General Election in terms of their preferred party was.



The total number of votes per candidate amounted to

Bernie Sanders: 11959102

Donald Trump: 13302541

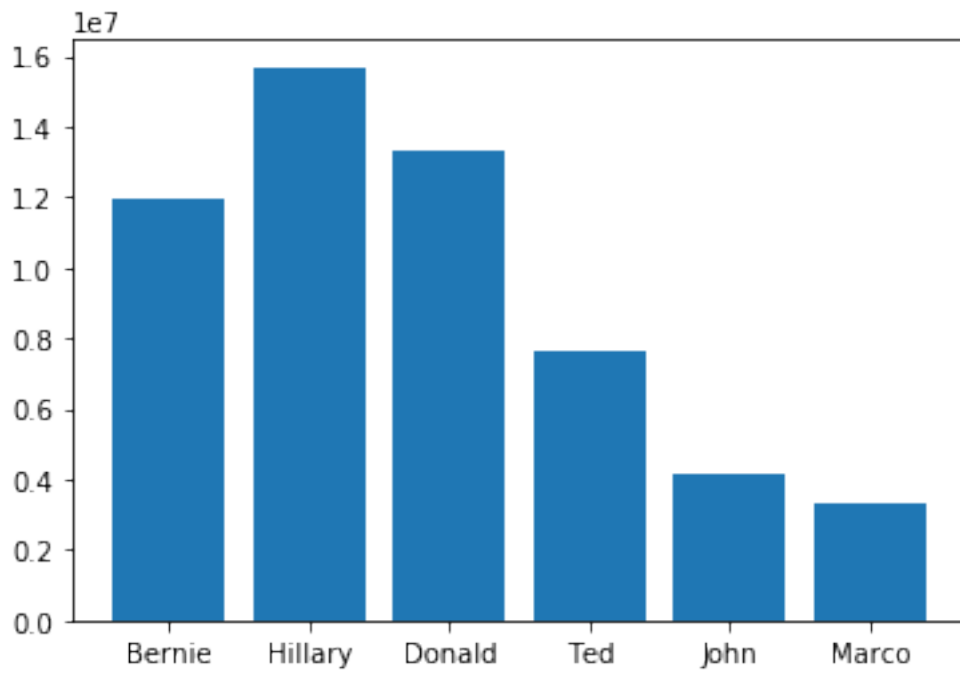
Hillary Clinton: 15692452

John Kasich: 4159949

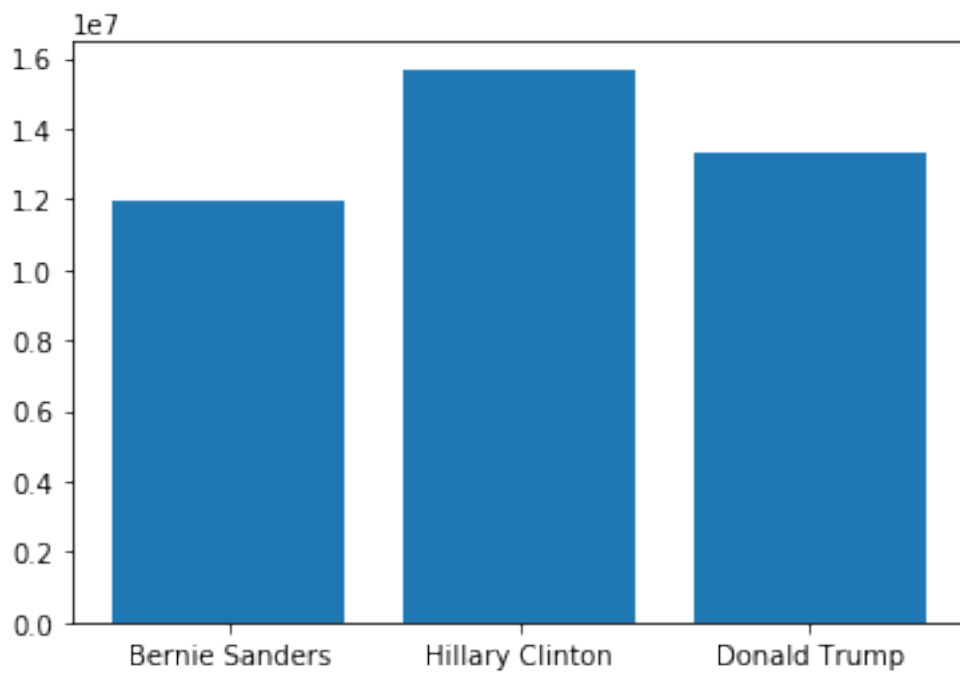
Marco Rubio: 3321076

Ted Cruz: 7603006

Other smaller candidates included Mike Huckabee, Jeb Bush, Carly Fiorina and Chris Christie.

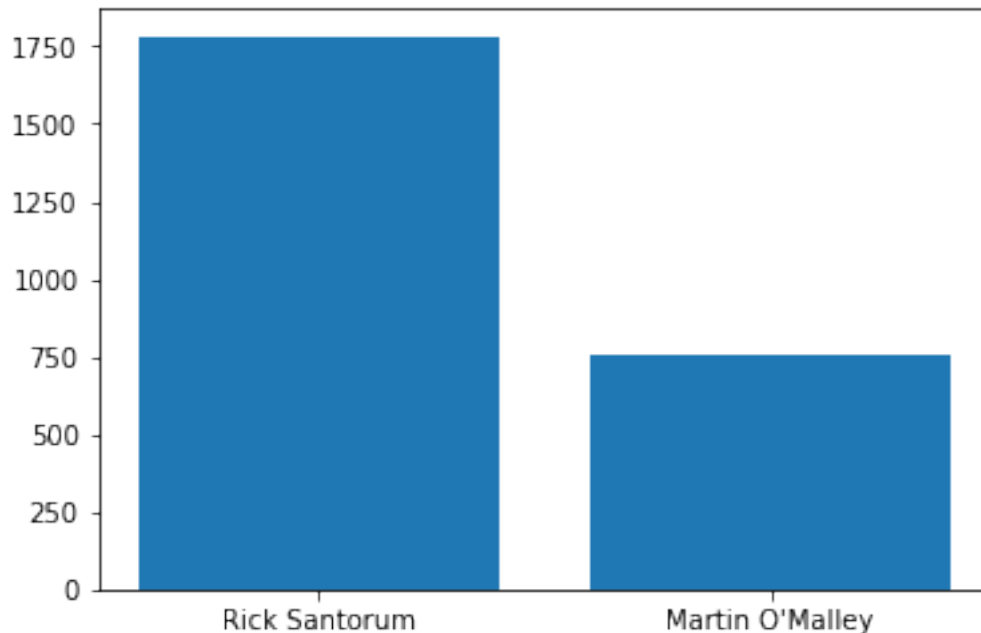


The top 3 candidates turned out to be Bernie Sanders, Hillary Clinton and Donald Trump.



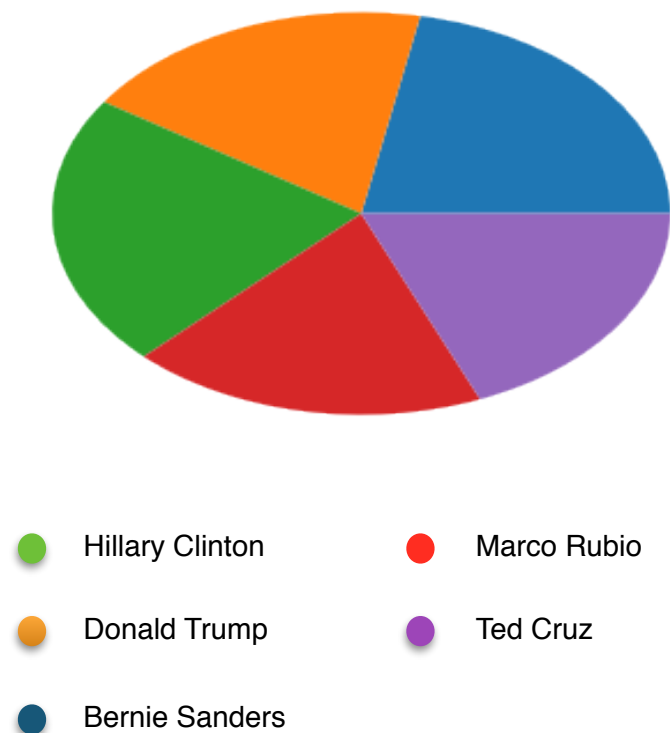
Hillary Clinton was leading with Bernie Sanders and Donald Trump quite close to each other

Whereas on the other hand there existed, Martin O'Malley and Rick Santorum at the bottom of the ladder, with a total of around 2500 votes.



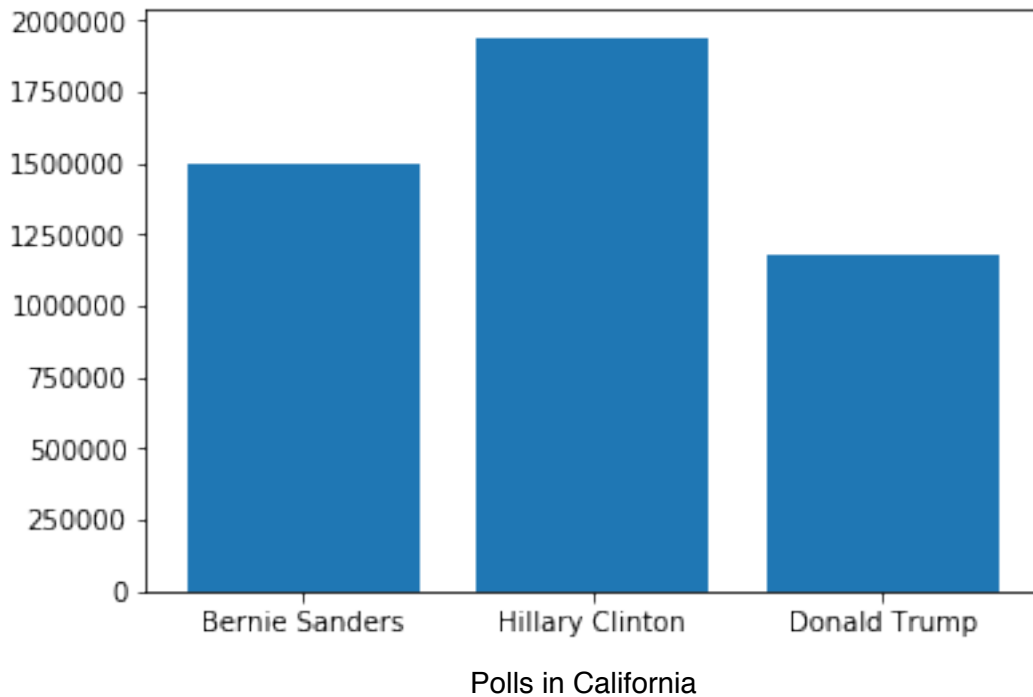
The distribution of counties with respect to their preferred candidate

Bernie Sanders: 4205  
Carly Fiorina: 109  
Chris Christie: 109  
Donald Trump: 3586  
Hillary Clinton: 4205  
Jeb Bush: 155  
John Kasich: 3586  
Marco Rubio: 2555  
Martin O'Malley: 99  
Mike Huckabee: 99  
Rand Paul: 99  
Rick Santorum: 99  
Ted Cruz: 3586

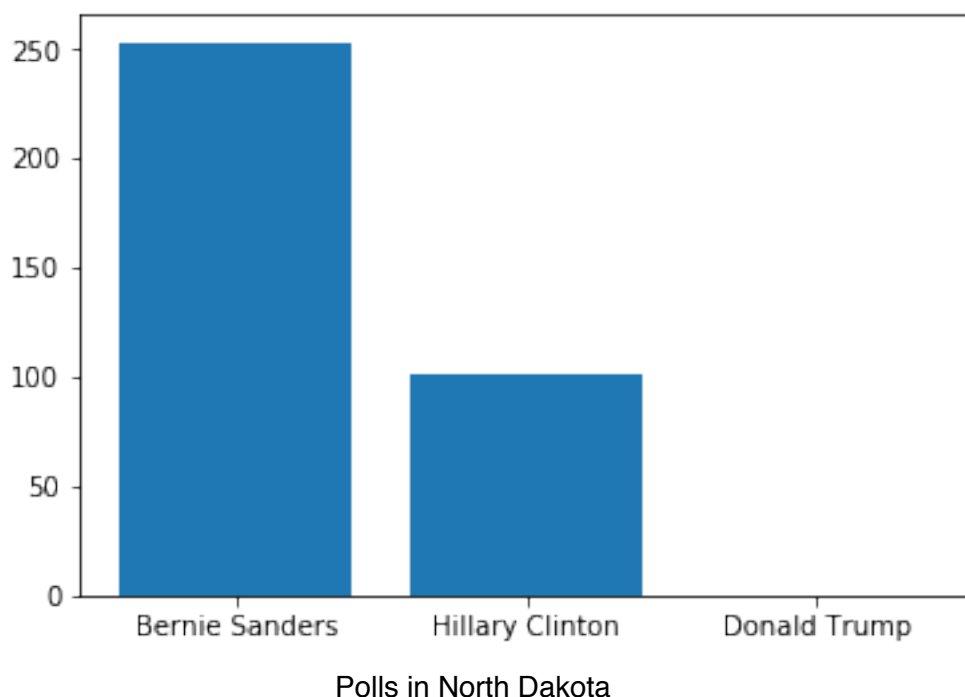




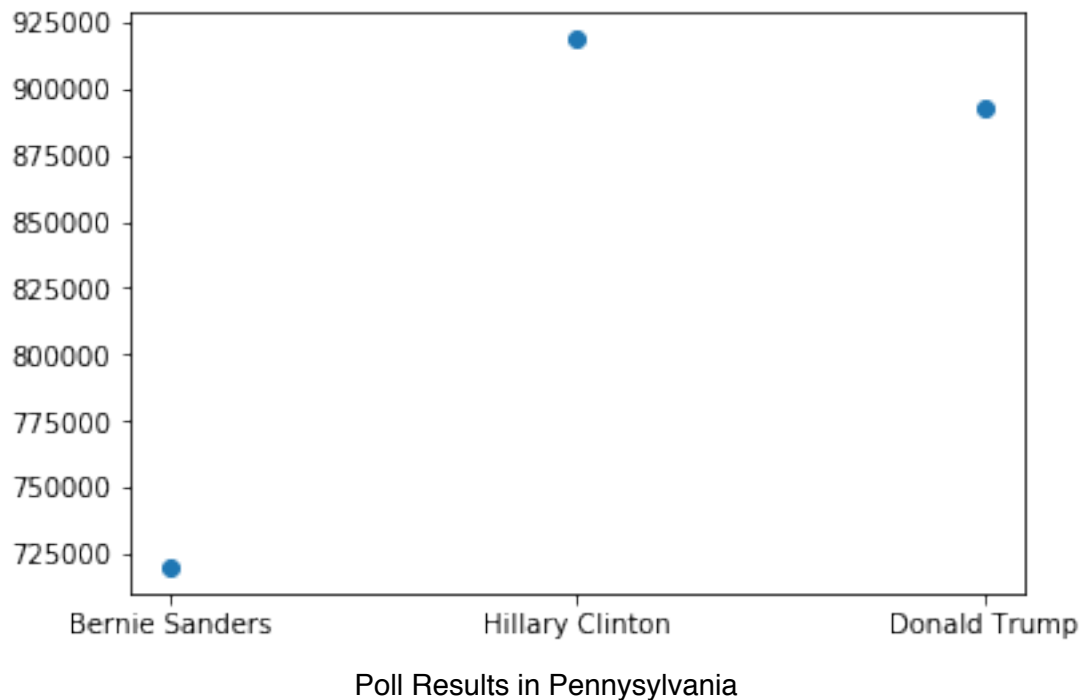
California is found to be the state with the greatest number of voters. With a total of 4.9 million voters, California has a greater number of Democrat supporters siding with Hillary Clinton, than Republican voters



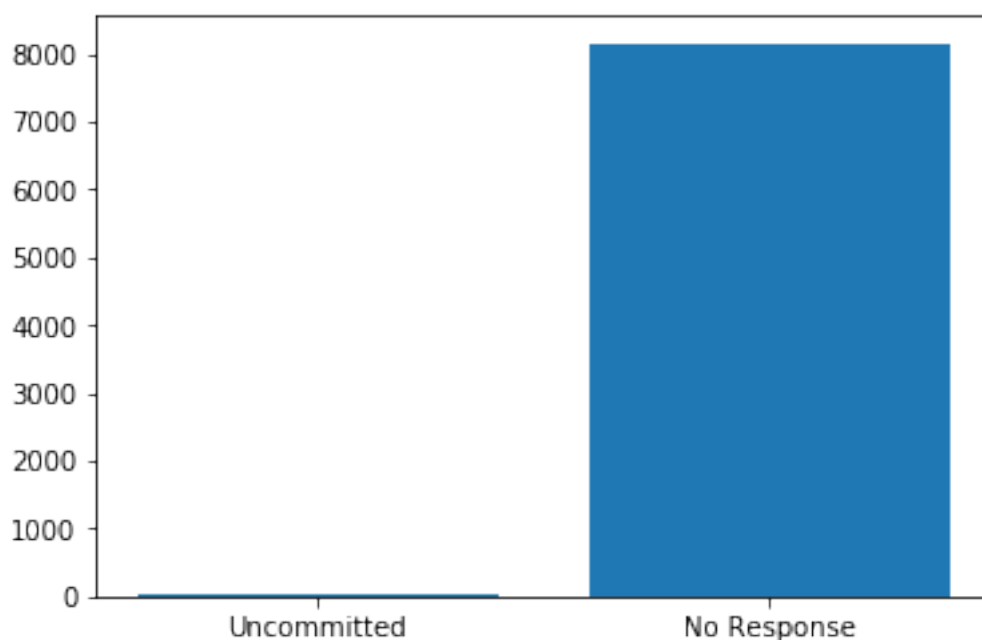
On the contrary, North Dakota has the least amount of voters numbering to around three hundred and fifty. North Dakota has majoritarian Bernie Sanders support and half that amount for Hillary Clinton. To much amazement, it has no voters for Donald Trump at all.



The most Pro-Republican state is found to be Pennysylvania. Most of the voters in Pennysylvania tend to favour Donald Trump, who has almost 900,000 votes in the region.



Overall American Population that was not actively engaged in the election process amounted to around eight thousand individuals who either did not provide any response regarding their political views or were not committed to any specific party or candidate.



# CONCLUSION

The overall american population voted in favour of the republican party. Conservative states such as Idaho and Alabama leaned more towards the Republican side whereas liberal states like Washington and Illinois supported the Democrats .

The US General Election of 2016 was won by the Republican party after lead candidate Donald Trump Jr. won 304 electoral college seats thus clearing the minimum 270 required to win.

Our take away from this exercise is that data driven statistical insights with the help of teamwork and thorough checking of any given subject provides meaningful numbers such that a few facts can be drawn from them. Although it is true that the results derived from data can vary and stray far from the actual truth, it doesn't also mean that valid statements which contain rational logic cannot be used to further the argument. Hence, we can say that one can use statistical data to form a valid statement.