COMP - 8157 Advanced Database Topics – PART 2

MASTERS IN APPLIED COMPUTING

ASSIGNMENT – 1

Analysis of Windsor Municipal Election Results 2014

Submitted to

Dr. Stephanos Mavromoustakos

Submitted by

Shiva Sai Mashetty
(1051555423)

ANALYSIS OF ELECTION RESULTS IN THE YEAR 2014

BY

SHIVA SAI MASHETTY

INTRODUCTION:

We are living in a world where population is increasing day to day, so its becoming very tough for the government to maintain the democracy.so it came to a conclusion that they have to elect the leaders based on the people's opinion by their votes percentage. This is the best possible way to elect the leaders based on the analysis of the votes casted by every individual.

DATA SET:

The dataset which I used in analyzing data is extracted from the City of Windsor official website .

The following are the requirements related to this particular dataset

Dataset Description:

This dataset contains the information about the results of the Municipal Election held in Windsor which was held in October 27, 2014.

Dataset Attributes:

The attributes include the Ward, Poll, Poll Name, Contest Title, Candidate, and Total.

- **Ward:** This field contains the information about the different wards numbers available in the current city.
- **Poll:** This field contains the information about the different poll numbers which are associated with the particular area in the city.
- **Poll name:** This field contains information about the different poll names for a particular poll.

- **Contest Title**: This field contains information about the different roles contesting in the elections such as Mayor, counsellor, Trustee.

- **Total:** This field contains the information about the total number of votes.

The dataset is available in the following website:
(https://opendata.citywindsor.ca/)

LINK: https://opendata.citywindsor.ca/opendata/details/198

FILE: Election2014.xlsx

USAGE OF RAPID MINER STUDIO:

The following are the steps that I followed for analyzing the data using the rapid miner studio tool:

Step1:

In the first step download the dataset from the given website and see which type of the dataset it is, based on its type import it in the rapid miner studio workspace.

Step2:

After importing the dataset in to the rapid miner tool observe all the attributes in the dataset . If there are any null values in it create a filter and put 'is not missing' for that particular attribute. Otherwise leave the dataset as it is.

Step3:

Next, Go to the Turbo prep in the tool and analyze the data according to the requirements.

Step4:

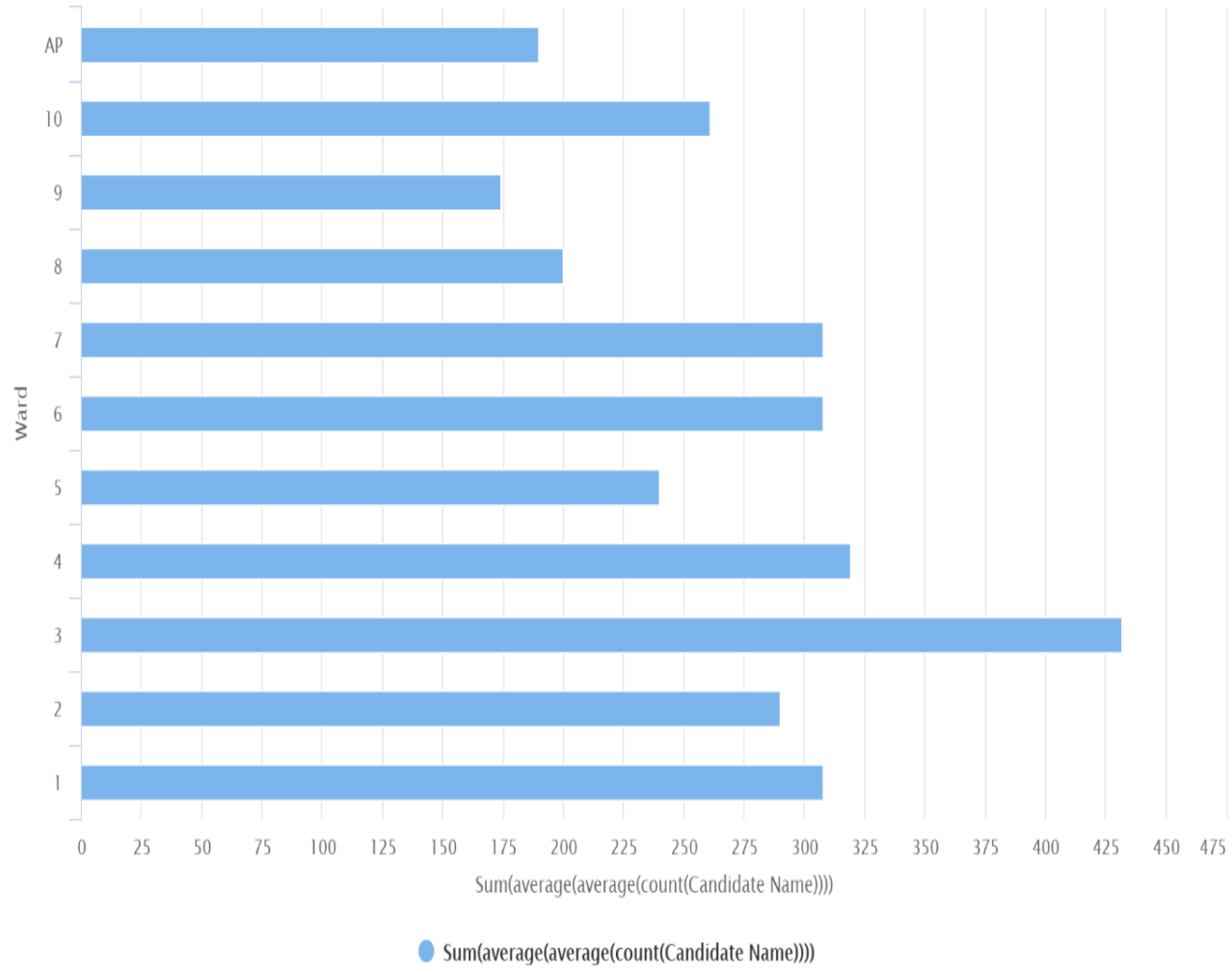
Finally, design the graphs based on the requirements and display the results on the screen.

SUMMARY OF RESEARCH QUESTIONS:

QUESTION 1 :

To analyze the total number of candidates per each ward?

Answer:



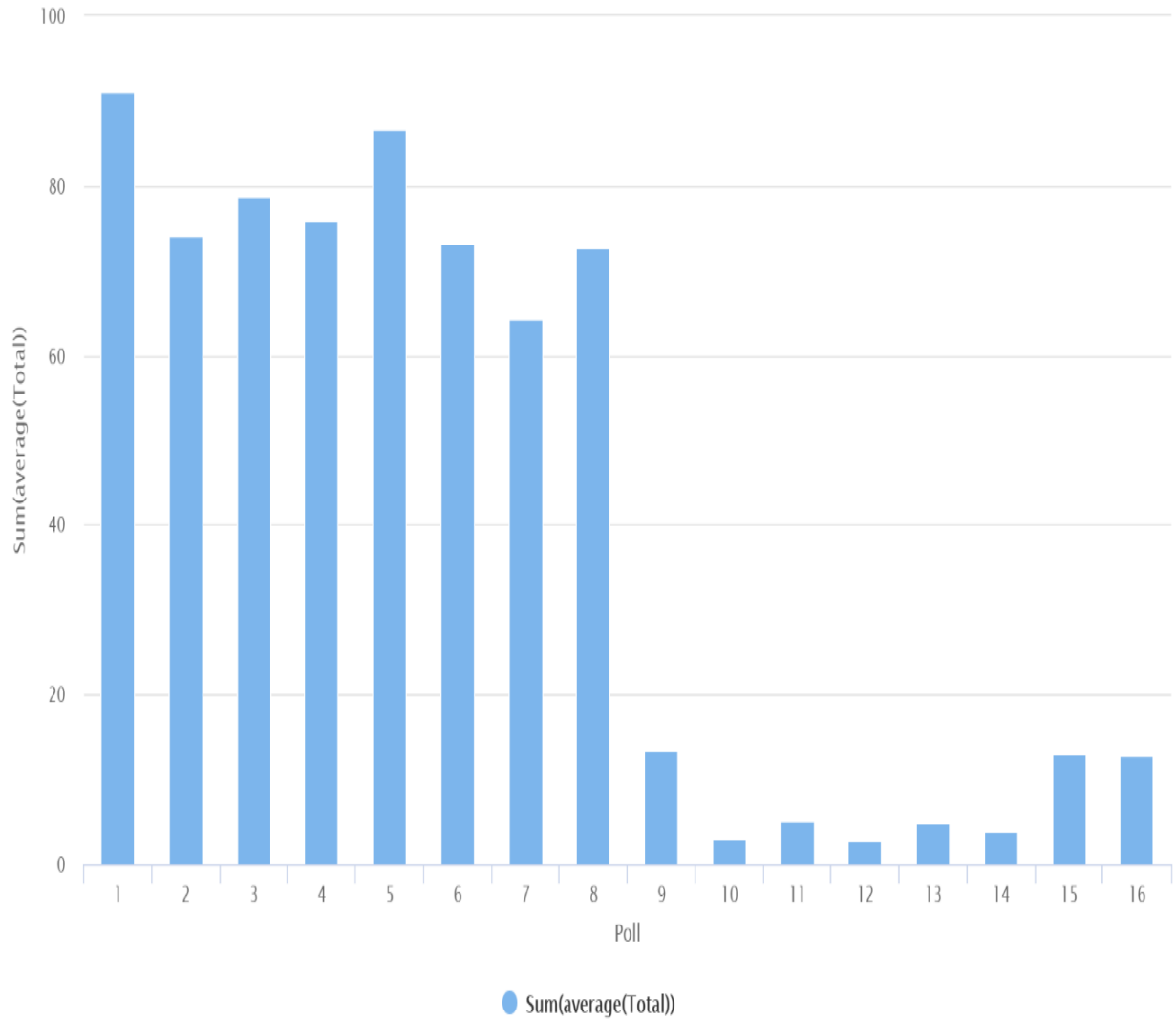
For the above analysis, I have created a pivot table between the ward and count of candidate name (using the aggregate function as count).In this analysis I have Observed that ward 3 has the maximum number of candidates where as ward 9 has least number of candidates. Both ward 6 and ward 7 has same number of candidates. On plotting the values from the pivot table I have obtained the

following graph which depicts the information about the total number of candidates per each ward. The X-axis shows the information about the different wards and Y-axis shows the information about the total number of candidates.

QUESTION 2 :

To analyze the total number of Votes per each Poll?

Answer:



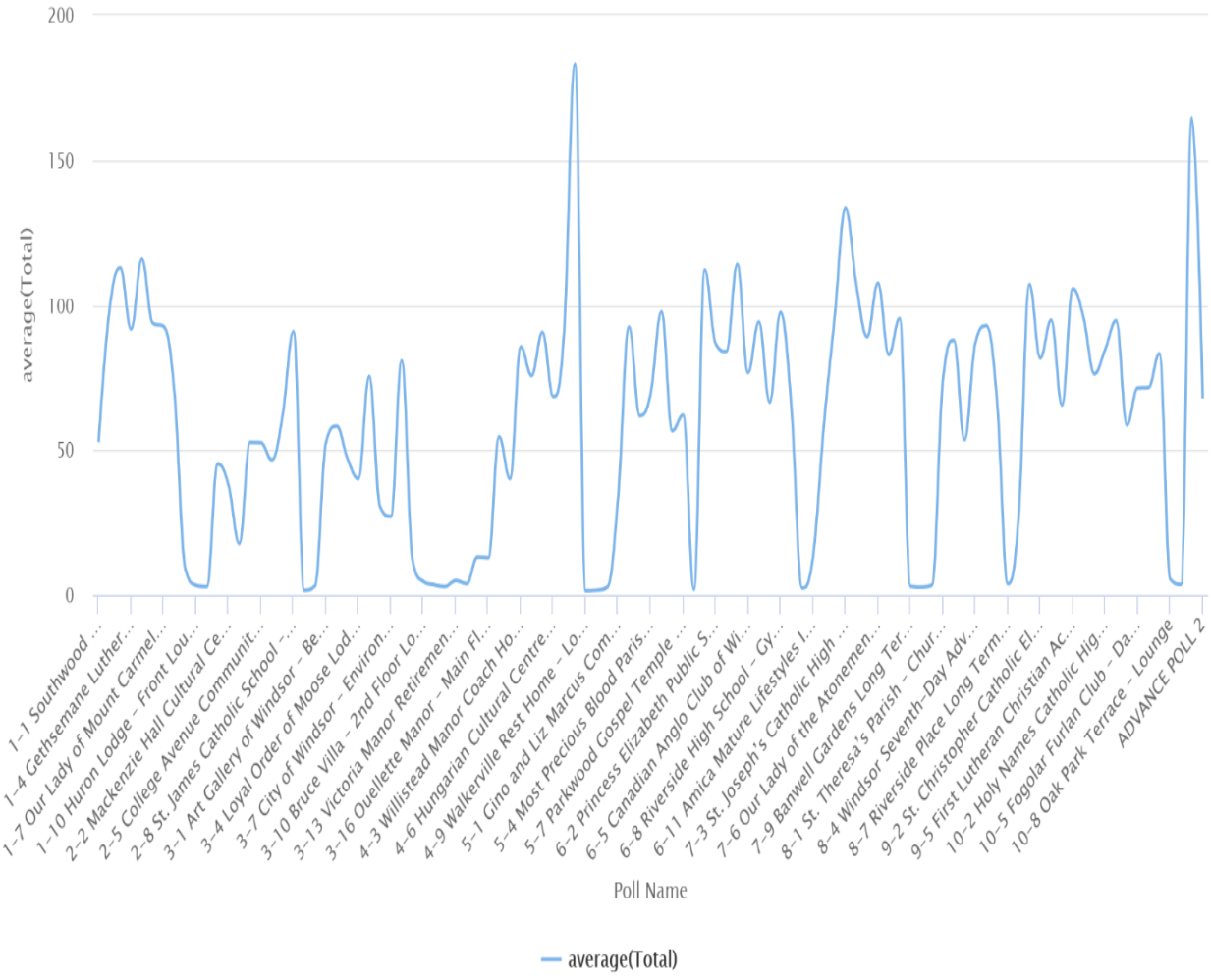
For the above analysis, I have created a pivot table between the Poll and Total number of votes(using the aggregate function as Sum).In this analysis I have Observed that poll 1 has maximum number of votes where as poll 10 and poll 12 has least number of votes and poll 11 and poll 13 has equal number of votes. On plotting the values from the pivot table I have obtained the following graph which

depicts the information about total number of votes per each poll. The X-axis shows the information about the poll and Y-axis shows the information about the sum of total number of votes.

QUESTION 3 :

To analyze the total number of poll names per each ward and poll?

Answer:



For the above analysis, I have created a pivot table between the ward, poll and average of total number of poll names (using the aggregate function as average).In this analysis I have Observed that ward 4, poll9 has the highest number of poll names where as ward 5 and poll 7 has the least number of poll names. On plotting the values from the pivot table I have obtained the following

graph which depicts the information about total number of poll names per each ward and poll. The X-axis shows the information about the poll, ward and poll name where as Y-axis shows the information about the total number of poll names.

Conclusion:

Based on the above analysis we can study the patterns of current elections and also predict the results of future elections. This data is divided in to both training and testing data which helps in prediction of future election results.