PROG8430 – Data Analysis, Modeling and Algorithms

Assignment 1

Descriptive Analysis with Excel

DUE BEFORE MIDNIGHT JANUARY 29, 2019

1. Submission Guidelines

All assignments must be submitted via the econestoga course website before the due date in to the assignment folder.

You may make multiple submissions, but only the most current submission will be graded.

Assignments submitted after the due date will receive a mark of 0.

Submissions should be accessible in Excel

Written answers should be presented clearly on an additional tab in the spreadsheet workbook.

2. Grading

This assignment will be worth 5% of your total grade in the course. Late assignments will receive a mark of 0.

3. Data

Each student will be using a dataset with an identical format but with unique data. Datasets are located on the econestoga website under 'Assignment' in the zipped file:
https://conestoga.desire2learn.com/d2l/common/viewFile.d2lfile/Database/NDQxMDI2MQ/pollresults_resultatsbureauCanada.zip?ou=244190

Each student is assigned a separate dataset per the file:

https://conestoga.desire2learn.com/d2l/common/viewFile.d2lfile/Database/NDQxMDQ1NQ/Assignment%201%20-%20Student%20Assignments.pdf?ou=244190

The data dictionary from Elections Canada is in the Appendix.

4. Background

The Canadian Federal Election of 2015 was overseen and administered by Elections Canada. There are 338 electoral districts in Canada, each with several candidates (one for each party and independents) running to win the right to represent that electoral district.

Each Electoral district is divided in to a number of polling stations (locations at which voters place their votes). After the election, each candidate is interested in the poll by poll results to determine

which areas of their district gave them strong support and which gave them weak support. They can then use this information to help determine advertising expenditures in subsequent elections.

You have been engaged to assist each of the candidates in a post-election analysis of the results.

5. Assignment Tasks

1. Identify the feature type of each column in the dataset.
2. Summarize the results of the vote (number of votes, percentage of vote, percentage of voters casting a ballot (i.e. "turn out"), and percentage of possible electors) for each candidate and each polling station using the main numeric measures discussed in class and at least one graphical method you deem appropriate (please justify your decision). **Please note:** To do this, you may have to reformat the dataset using an appropriate technique.
3. Identify the outliers for each key measure using a technique (or techniques) discussed in class. Discuss what to do with the outlier(s) in your analysis.
4. Is there a correlation between the performance of each candidate and "voter turner out" in that poll?
5. Assuming the patterns for the Election you are studying hold for the next election, at dw2vwhich five polling stations should each candidate spend their advertising money next Election, and which five should they not spend any money on? Justify your answer with appropriate evidence and reasoning.

APPENDIX: DATA DICTIONARY

Data Dictionary	
Field Name	**Description**
Electoral District Number	The electoral district number.
Electoral District Name_English	The English name of the electoral district.
Electoral District Name_French	The French name of the electoral district.
Polling Station Number	The number assigned to the polling station,
	for example: 3, 45A, 48-3, 601.
Polling Station Name	A name that generally represents the locality of the polling division boundary.
Void Poll Indicator	Indicates that a poll exists but has no electors.
No Poll Held Indicator	Indicates that the returning officer intended to hold this poll, but unforeseen circumstances prevented it.
Merge With	Indicates the number of the polling station with which the results of this poll were merged.
Rejected Ballots for Polling Station	The number of rejected ballots at this polling station.
Electors for Polling Station	The number of electors on the list of electors for this polling station.
Candidate's Family Name	The family name of the candidate.
Candidate's Middle Name	The middle name of the candidate.
Candidate's First Name	The first name of the candidate.
Political Affiliation Name_English	The short-form English name of the candidate's political affiliation.
Political Affiliation Name_French	The short-form French name of the candidate's political affiliation.
Incumbent Indicator	"Y" if candidate was the incumbent, "N" otherwise.
Elected Candidate Indicator	"Y" if candidate was elected, "N" otherwise.
Candidate Poll Votes Count	The number of valid votes the candidate obtained at this polling station.