MIS781 Business Intelligence and Database

Example - Election Database Design and Report

Group Assignment

This report has been prepared by:

GROUP Number
X, Y, Z NAME OF ALL STUDENTS

TABLE OF CONTENTS

INTRODUCTION

1. Introduction (2 marks)

-Overall introduction of your report (Database for simulated phishing campaigns dashboard)

-Mention your three main Target Audience

Example:

1. Phishing Campaign Manager
2. Individual User or Department Manager
3. HR Manager or CEO

Example:

General Introduction

The database designs presented in this report made use of presidential election data in the United States of America across different years. With the country's electoral college process being unique, the election system can be deemed complicated and perplexing for most people. Hence, to assist different stakeholders in understanding the process, evaluating results, and making decisions on various aspects of the election setting, these database designs have been created.

Target Audiences

- **Database 1:** Civil society
- **Database 2:** Political parties
- **Database 3:** Election campaign teams

Database Design Objectives

- **Database 1:** In consideration of the civil society's role in voter education and election monitoring (International Foundation for Electoral Systems et. al, 2003), the *Voter Turnouts and Voting Results* database design aims to provide easily accessible, perceptive, and reliable information on both voter participation of various segments of the electorate and official popular vote results in order to assist with the beneficiary's voter engagement initiatives.
- **Database 2:** ….
- **Database 3:** …

Database Design Benefits

- **Database 1:** With the *Voter Turnouts and Voting Results* database design, the civil society would have a clear picture of the turnout results of their segments of interest by finding all information needed in one place. This would accelerate preparations for voter education training sessions for recipients that are at risk of not voting due to lack of information, apathy, or sense of alienation, or those susceptible to manipulation (International Foundation for Electoral Systems et. al, 2003). Having the popular vote result together in the same database would also allow them to easily interpret the impact of voter turnout to actual voting results. Also, the presence of the details in each state's voting-eligible population and registered voters would assist them in strategy and decision formulations. In addition, information on the political parties and their respective ideologies would provide a helpful tool in voter education. In this case, target setting, and information dissemination would be accurate, therefore making initiatives on increasing voting turnouts efficient and effective.
- **Database 2:** …
- **Database 3:** …

Assumptions

DATABASE 1

- *Voting-eligible population* is composed of people who are above 18 years old and have the right and capacity to vote regardless of voter registration status.
- *Registered voters* are people who have officially registered themselves to vote.
- Data on *voting turnouts* and *popular votes* are official presidential election results data collected from valid resources.

- Data for *poll turnout* are collected from an official surveying body which conducts survey manually and electronically.
- *New voters* are people voting for the first time regardless of their registration date.
- *Ethnic minorities* are ethnic groups not belonging to the following groups: White/European American, Black/African American, Alaska Native/American Indian, Hispanic/Latino, Asian American, Hawaiian Native
- *Remote communities* are towns which populations are less than 15,000.
- *Candidates* are official political party presidential representatives that have won the preliminary elections.
- *Ideology* refers to the political party's current ideology and can change over time.
- ……

DATABASE 2

- The data represented in the database is for a particular day in between the voting days for the different regions of states in US.
- The *candidate* must be a natural-born citizen of the United States, be at least 35 years old and above.
- …

DATABASE 3

- …

...

Attributes and Business Rules

Attribute	Data Type	Description	Business Rule
Turnout ID	Character	Characters that represent the turnout results for a specified year and state	• Composed of a unique combination of 4 characters: 2 numbers (last 2 digits of year yyyy) followed by 2 letters (State ID) • Cannot be NULL
State ID	Character	Characters that represent the state name	• Composed of a unique combination of 2 letters • Cannot be NULL
State Name	Varchar	Name of states in the United States	• Can contain up to 30 characters
Voting-eligible Population	Integer	By state population of eligible voters	• Must be a whole number • Cannot be less than *Registered Voters*
Registered Voters	Integer	By state number of registered voters	• Must be a whole number • Cannot be greater than *Voting-eligible Population*
Poll Turnout	Floating Point Number	By state percentage of poll survey turnout results for presidential election in the United States conducted before the official voting	• Floating Point Number
Actual Turnout	Floating Point Number	By state percentage of official overall turnout results for the presidential election in the United States	• Composed of 2 whole numbers and 2 decimals
New Voter Turnout	Floating Point Number	By state percentage of official first-time voter turnout results for the presidential election in the United States	• Composed of 2 whole numbers and 2 decimals
Female Turnout	Floating Point Number	By state percentage of official female voter turnout results for the presidential election in the United States	• Composed of 2 whole numbers and 2 decimals
Ethnic Minorities Turnout	Floating Point Number	By state percentage of official ethnic minority voters turnout results for the presidential election in the United States	• Composed of 2 whole numbers and 2 decimals
Remote Communities Turnout	Floating Point Number	By state percentage of official remote community voters turnout results for the presidential election in the United States	• Composed of 2 whole numbers and 2 decimals
Candidate ID	Integer	Numbers that represent the candidate details	• Composed of a unique combination of 3 numbers (first 2: last 2 digits of year yyyy; last 1: numbers from 0-9) • Cannot be NULL
Candidate Name	Varchar	Name of official presidential candidates in the United States	• Composed of the last and first names of the candidate in the format <last name, first name> • Can contain up to 30 characters

2-DATABASE DESIGNS

(4 marks for each database and commentary: 1 mark for 1NF, 1m for 2NF, 1m for 3NF, 1 m for comments)

Example:

Database 1: Voter Turnouts and Voting Results

Example:

The Voter Turnouts and Voting Results Database Design includes by state historical data of general presidential election popular vote results and voter turnouts of various categories, state information, and details of each political party's presidential candidate.

1NF

TURNOUT ID	STATE ID	STATE NAME	VOTING-ELIGIBLE POPULATION	REGISTERED VOTERS	POLL TURNOUT	ACTUAL TURNOUT	NEW VOTER TURNOUT	FEMALE TURNOUT	ETHNIC MINORITIES TURNOUT	REMOTE COMMUNITIES TURNOUT	CANDIDATE ID	CANDIDATE NAME	PARTY NAME	IDEOLOGY	POPULAR VOTE
20OK	OK	Oklahoma	2,845,835	1,861,367	79.14	54.84	59.19	68.57	55.07	68.79	161	Trump, Donald	Republican	American conservatism	65.37
20AR	AR	Arkansas	2,182,375	1,456,364	83.56	55.86	51.86	71.68	55.41	56.77	161	Trump, Donald	Republican	American conservatism	62.40
20HI	HI	Hawaii	1,007,920	530,237	81.47	57.00	65.57	57.08	58.21	56.71	201	Biden, Joe	Democratic	Modern Liberalism	63.73
16HI	HI	Hawaii	1,007,920	530,237	76.38	53.61	59.36	52.61	48.23	52.38	162	Hillary Clinton	Democratic	Liberalism	56.34
20WV	WV	West Virginia	1,394,028	913,897	83.79	57.00	50.87	54.02	58.97	59.12	161	Trump, Donald	Republican	American conservatism	68.63
20TN	TN	Tennessee	5,124,867	3,251,024	78.49	59.59	51.71	50.54	56.51	58.70	161	Trump, Donald	Republican	American conservatism	60.66
20MS	MS	Mississippi	2,201,950	1,725,486	84.61	59.67	56.86	58.53	56.99	60.38	161	Trump, Donald	Republican	American conservatism	57.60
20TX	TX	Texas	18,784,280	11,724,032	73.13	60.24	51.97	59.47	45.31	65.37	161	Trump, Donald	Republican	American conservatism	52.06
20IN	IN	Indiana	5,000,007	3,298,895	81.95	60.66	55.59	61.63	58.51	67.91	161	Trump, Donald	Republican	American conservatism	57.02
16IN	IN	Indiana	5,000,007	3,298,895	75.61	56.79	52.96	58.74	55.63	60.47	161	Trump, Donald	Republican	American conservatism	49.53
20MN	MN	New Mexico	1,515,355	916,146	79.59	60.97	62.29	57.28	45.22	59.00	201	Biden, Joe	Democratic	Modern Liberalism	54.29
20LA	LA	Louisiana	3,373,932	2,446,231	77.36	61.40	53.82	55.15	55.93	63.03	161	Trump, Donald	Republican	American conservatism	58.46
20NY	NY	New York	13,670,596	9,142,785	82.46	62.87	58.83	64.40	50.26	63.28	201	Biden, Joe	Democratic	Modern Liberalism	60.87
20AL	AL	Alabama	3,683,055	2,526,364	79.98	63.08	63.10	63.99	45.26	57.15	161	Trump, Donald	Republican	American conservatism	62.03
20DC	DC	District of Columbia	540,685	420,964	81.00	63.69	58.21	59.08	40.97	69.50	201	Biden, Joe	Democratic	Modern Liberalism	92.15
20KS	KS	Kansas	2,087,946	1,438,012	84.51	63.87	58.58	63.98	50.32	60.99	161	Trump, Donald	Republican	American conservatism	56.21
20SC	SC	South Carolina	3,926,305	2,575,945	77.63	64.01	59.02	64.10	48.04	58.94	161	Trump, Donald	Republican	American conservatism	55.11

Table 1.2. Voter turnouts and popular vote results per state

Table 1.2 has been created in order to meet the requirements of the first normal form, where all cells are single valued, entries in a column are of the same type, and rows are uniquely identified:

* *Turnout ID, State ID, and Candidate ID* columns were created to act as primary keys in order to distinctly determine the records.

Dependencies and Key Attributes

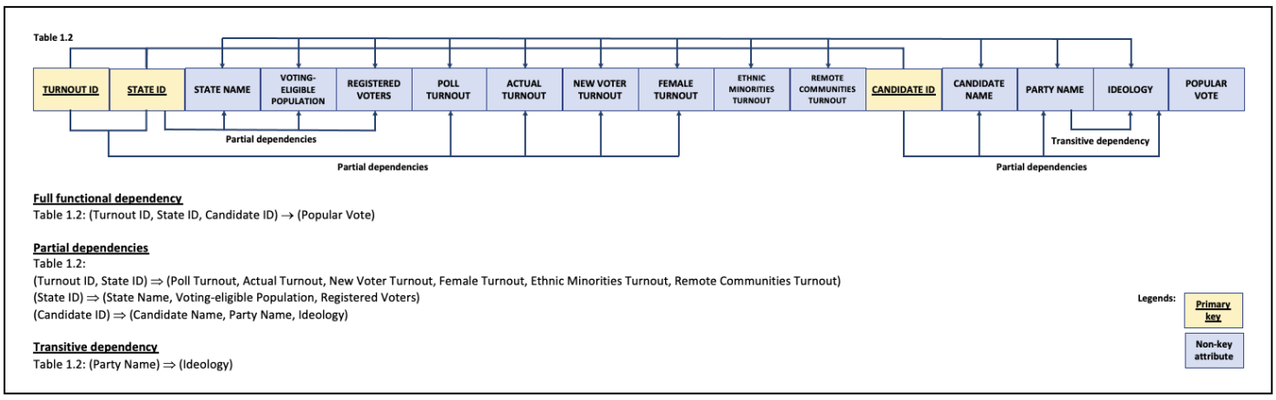

Figure 1.1. 1NF table dependency diagram

Even in its first normal form, the following anomalies still exist in *Table 1.2*:

* Insertion: Candidates who have not yet won the popular votes cannot be inserted
* Insertion: Turnouts for states that does not have results for the popular votes cannot be entered
* Deletion: If a candidate is deleted from the table, the corresponding records for the *Popular vote results, Voter turnouts, Voting-eligible population, and Registered voters* tables will not be lost
* Deletion: Given the relationship of party name and ideology, changing one item would remove the other information from the database
* Update: If a presidential candidate was replaced before the general election, update will be done for all tuples with the candidate's information

To get rid of these anomalies, we further normalize the table into the second normal form.

2NF

TURNOUT ID	STATE ID	POLL TURNOUT	ACTUAL TURNOUT	NEW VOTER TURNOUT	FEMALE TURNOUT	ETHNIC MINORITIES TURNOUT	REMOTE COMMUNITIES TURNOUT
20OK	OK	79.14	54.84	59.19	68.57	55.07	68.79
20AR	AR	83.56	55.86	51.86	71.68	55.41	56.77
20HI	HI	81.47	57.00	65.57	57.08	58.21	56.71
16HI	HI	76.38	53.61	59.36	52.61	48.23	52.38
20WV	WV	83.79	57.00	50.87	54.02	58.97	59.12
20TN	TN	78.49	59.59	51.71	50.54	56.51	58.70
20MS	MS	84.61	59.67	56.86	58.53	56.99	60.38
20TX	TX	73.13	60.24	51.97	59.47	45.31	65.37
20IN	IN	81.95	60.66	55.59	61.63	58.51	67.91
16IN	IN	75.61	56.79	52.96	58.74	55.63	60.47
20MN	MN	79.59	60.97	62.29	57.28	45.22	59.00
20LA	LA	77.36	61.40	53.82	55.15	55.93	63.03
20NY	NY	82.46	62.87	58.83	64.40	50.26	63.28
20AL	AL	79.98	63.08	63.10	63.99	45.26	57.15
20DC	DC	81.00	63.69	58.21	59.08	40.97	69.50
20KS	KS	84.51	63.87	58.58	63.98	50.32	60.99
20SC	SC	77.63	64.01	59.02	64.10	48.04	58.94

Table 1.3. Voter turnouts by state broken down into categories

TURNOUT ID	STATE ID	CANDIDATE ID	POPULAR VOTE
20OK	OK	161	65.37
20AR	AR	161	62.40
20HI	HI	201	63.73
16HI	HI	162	56.34
20WV	WV	161	68.63
20TN	TN	161	60.66
20MS	MS	161	57.60
20TX	TX	161	52.06
20IN	IN	161	57.02
16IN	IN	161	49.53
20MN	MN	201	54.29
20LA	LA	161	58.46
20NY	NY	201	60.87
20AL	AL	161	62.03
20DC	DC	201	92.15
20KS	KS	161	56.21
20SC	SC	161	55.11

Table 1.4. Popular vote results by state

STATE ID	STATE NAME	VOTING-ELIGIBLE POPULATION	REGISTERED VOTERS
OK	Oklahoma	2,845,835	1,861,367
AR	Arkansas	2,182,375	1,456,364
HI	Hawaii	1,007,920	530,237
WV	West Virginia	1,394,028	913,897
TN	Tennessee	5,124,867	3,251,024
MS	Mississippi	2,201,950	1,725,486
TX	Texas	18,784,280	11,724,032
IN	Indiana	5,000,007	3,298,895
MN	New Mexico	1,515,355	916,146
LA	Louisiana	3,373,932	2,446,231
NY	New York	13,670,596	9,142,785
AL	Alabama	3,683,055	2,526,364
DC	District of Columbia	540,685	420,964
KS	Kansas	2,087,946	1,438,012
SC	South Carolina	3,926,305	2,575,945

Table 1.5. State details

CANDIDATE ID	CANDIDATE NAME	PARTY NAME	IDEOLOGY
201	Biden, Joe	Democratic	Modern Liberalism
161	Trump, Donald	Republican	American conservatism
202	Jorgensen, Jo	Libertarian	Libertarianism
203	Hawkins, Howie	Green	Ecopolitics
000	Other	Other	Other
162	Clinton, Hillary	Democratic	Modern Liberalism
163	Johnson, Gary	Libertarian	Libertarianism
164	Stein, Jill	Green	Ecopolitics

Table 1.6. Candidate details

By creating *Tables 1.3, 1.4, 1.5, and 1.6*, partial dependencies have been removed and the tables were reduced to the second normal form, where all non-key attributes in the table are fully functionally dependent on the primary key. With this, the following anomalies were eliminated:

* Insertion: Can now enter candidates who have not yet won the popular votes
* Insertion: Can now enter turnouts for states that does not have results for the popular votes
* Deletion: If a candidate is deleted from the table, the corresponding records for the *Popular vote results*, *Vote turnouts*, *Voting-eligible population*, *and Registered voters* tables will not be lost
* Update: If a presidential candidate was replaced before the general election, update will only be done on the *Candidate details* table

Dependencies and Key Attributes

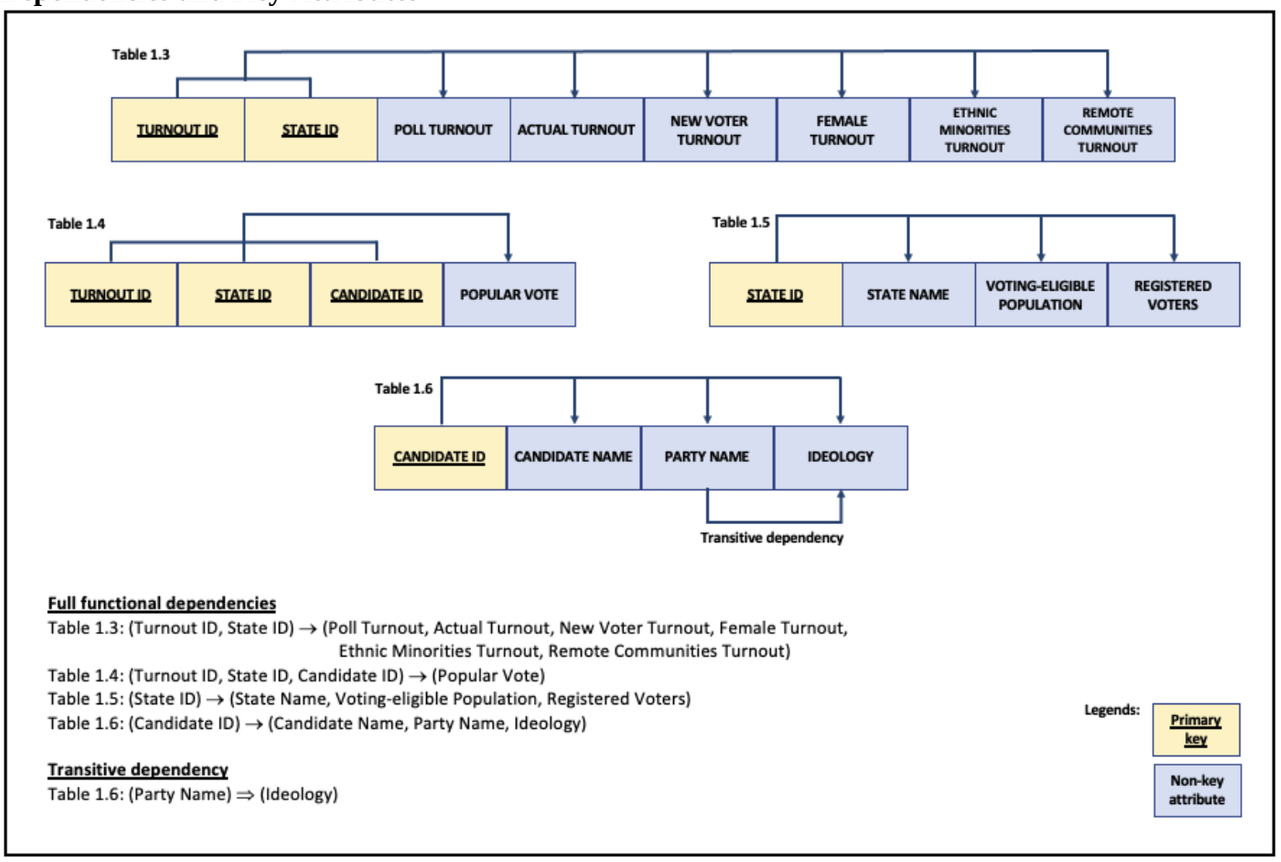

Figure 1.2. 2NF tables dependency diagram

Even in its second normal form, the following anomalies still exist in *Table 1.6*:

* Insertion: Cannot enter the fact that a particular political party has an ideology unless *Party Name* is recorded
* Deletion: If a particular candidate switches party before the general election, the information on the previous party's ideology is lost
* Update: If the ideology of the political party changes, all candidate records need to be updated

To get rid of these anomalies, we further normalise the table into the third normal form.

3NF

TURNOUT ID	STATE ID	POLL TURNOUT	ACTUAL TURNOUT	NEW VOTER TURNOUT	FEMALE TURNOUT	ETHNIC MINORITIES TURNOUT	REMOTE COMMUNITIES TURNOUT
20OK	OK	79.14	54.84	59.19	68.57	55.07	68.79
20AR	AR	83.56	55.86	51.86	71.68	55.41	56.77
20HI	HI	81.47	57.00	65.57	57.08	58.21	56.71
16HI	HI	76.38	53.61	59.36	52.61	48.23	52.38
20WV	WV	83.79	57.00	50.87	54.02	58.97	59.12
20TN	TN	78.49	59.59	51.71	50.54	56.51	58.70
20MS	MS	84.61	59.67	56.86	58.53	56.99	60.38
20TX	TX	73.13	60.24	51.97	59.47	45.31	65.37
20IN	IN	81.95	60.66	55.59	61.63	58.51	67.91
16IN	IN	75.61	56.79	52.96	58.74	55.63	60.47
20MN	MN	79.59	60.97	62.29	57.28	45.22	59.00
20LA	LA	77.36	61.40	53.82	55.15	55.93	63.03
20NY	NY	82.46	62.87	58.83	64.40	50.26	63.28
20AL	AL	79.98	63.08	63.10	63.99	45.26	57.15
20DC	DC	81.00	63.69	58.21	59.08	40.97	69.50
20KS	KS	84.51	63.87	58.58	63.98	50.32	60.99
20SC	SC	77.63	64.01	59.02	64.10	48.04	58.94

Table 1.3. Voter turnouts by state broken down into categories

TURNOUT ID	STATE ID	CANDIDATE ID	POPULAR VOTE
20OK	OK	161	65.37
20AR	AR	161	62.40
20HI	HI	201	63.73
16HI	HI	162	56.34
20WV	WV	161	68.63
20TN	TN	161	60.66
20MS	MS	161	57.60
20TX	TX	161	52.06
20IN	IN	161	57.02
16IN	IN	161	49.53
20MN	MN	201	54.29
20LA	LA	161	58.46
20NY	NY	201	60.87
20AL	AL	161	62.03
20DC	DC	201	92.15
20KS	KS	161	56.21
20SC	SC	161	55.11

Table 1.4. Popular vote results by state

STATE ID	STATE NAME	VOTING-ELIGIBLE POPULATION	REGISTERED VOTERS
OK	Oklahoma	2,845,835	1,861,367
AR	Arkansas	2,182,375	1,456,364
HI	Hawaii	1,007,920	530,237
WV	West Virginia	1,394,028	913,897
TN	Tennessee	5,124,867	3,251,024
MS	Mississippi	2,201,950	1,725,486
TX	Texas	18,784,280	11,724,032
IN	Indiana	5,000,007	3,298,895
MN	New Mexico	1,515,355	916,146
LA	Louisiana	3,373,932	2,446,231
NY	New York	13,670,596	9,142,785
AL	Alabama	3,683,055	2,526,364
DC	District of Columbia	540,685	420,964
KS	Kansas	2,087,946	1,438,012
SC	South Carolina	3,926,305	2,575,945

Table 1.5. State details

CANDIDATE ID	CANDIDATE NAME	PARTY ID
201	Biden, Joe	P01
161	Trump, Donald	P02
202	Jorgensen, Jo	P03
203	Hawkins, Howie	P04
000	Other	P00
162	Clinton, Hillary	P01
163	Johnson, Gary	P03
164	Stein, Jill	P04

Table 1.7. Candidate details (version 2)

PARTY ID	PARTY NAME	IDEOLOGY
P01	Democratic	Modern Liberalism
P02	Republican	American conservatism
P03	Libertarian	Libertarianism
P04	Green	Ecopolitics
P05	Alliance	Anti-corruption
P06	Constitution	Christian right
P07	Reform	Populism
P00	Other	Other

Table 1.8. Political parties and ideologies

By creating the *Political parties and ideologies* table, the transitive dependency between *Party Name* and *Ideology* has been removed and the *Candidate details* table was reduced to the third normal form, where all non-key fields in a table are determined solely by its primary key. With this, the following anomalies were eliminated:

* Insertion: Can now enter ideologies for new/other parties that take part in the election
* Deletion: If a particular candidate switches party before the general election, the information on the previous party's ideology is maintained
* Update: The ideology of each political party appears only once

Dependencies and Key Attributes

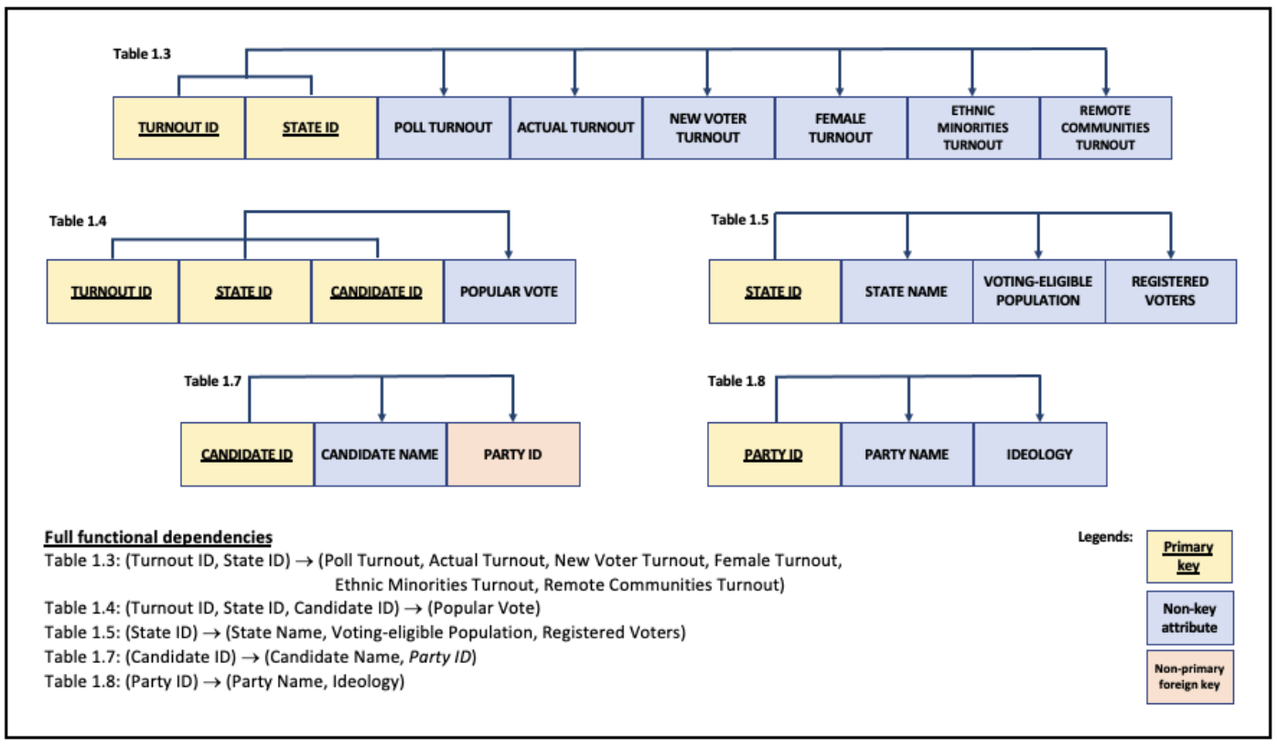

Figure 1.3. 3NF tables dependency diagram

With all the anomalies removed, we are now confident that the relations in *Database 1* are satisfactory and are more effective in inserting, updating, and deleting data.

Database 2:…

Database 3:…

REFERENCES

(use APA ref 1 mark) and report presentation (1 mark)

See: https://www.deakin.edu.au/students/studying/study-support/referencing/apa

Brancati, D. (n.d.). *Global Elections Database.* The Project. Global Elections Database. http://www.globalelectionsdatabase.com/

Harewood, W. G. (2016). *Comparing the advantages & disadvantages of first past the post, alternative voting [AV] and proportional voting representation [PR] electoral systems*. [Excerpt]. Parliament of Canada. https://www.ourcommons.ca/content/Committee/421/ERRE/Brief/BR8555618/br-external/HarewoodWGlenn-e.pdf

International Foundation for Electoral Systems, International IDEA, UN Department of Economic and Social Affairs (2003). *Voter Education*. ACE Project. *https://aceproject.org/main/english/ve/index.htm*

McDonald, M. P. (2018, September 5). *2016 November General Election Turnout Rates.* United States Elections Project. http://www.electproject.org/2016g

McDonald, M. P. (2020, December 7). *2020 November General Election Turnout Rates.* United States Elections Project. http://www.electproject.org/2020g

Open Knowledge Foundation and Open Data for Development Network. (2016). *Election Results*. Global Open Data Index. https://index.okfn.org/dataset/elections/

Pew Research Center (2008). *Inside Obama's Sweeping Victory*. Pew Research. https://www.pewresearch.org/2008/11/05/inside-obamas-sweeping-victory/

Sharda, R. et. al. (2016). *Business Intelligence, Analytics, and Data Science: A Managerial Perspective (4th edition).* Pearson.

MIS781 – Business Intelligence and Database

Trimester 1, 2024

Group Assignment
GROUP MEMBER CONTRIBUTION FORM

Group Number:

Name (Print)	**Student ID**	**%Contribution**	**Signature***
1.		**33.3%**	
2.		**33.3%**	
3.		**33.3%**	

Note: * By signing here, I hereby declare that this is my original work.

If every member of the group contributes equally, the figure entered in the '%Contribution' column should be 33.3% for a 3-member group. This page should be signed by each member of the group and attached to the end of the report.

Note: Where there is significant variation in group contributions, marks may be adjusted individually among group members.