PROJECT - NOAH CHOE

PART 1

QUESTION 1

- PRECINCT -> GEO, STATE, LOCALITY

- GEO -> PRECINCT, STATE, LOCALITY

- PRECINCT, TIMESTAMP -> GEO, STATE, LOCALITY, TOTALVOTES, BIDEN, TRUMP, FILESTAMP, ID

- GEO, TIMESTAMP -> PRECINCT, STATE, LOCALITY, TOTALVOTES, BIDEN, TRUMP, FILESTAMP, ID

QUESTION 2

Penna is not in BCNF.

PRECINCT -> GEO, LOCALITY, STATE

PRECINCT, TIMESTAMP -> TOTALVOTES, BIDEN, TRUMP, FILESTAMP, ID

PART 2

QUESTION 1 - The Precinct

a) winner(precinct)



|  |  |  |
| --- | --- | --- |
| @winner | @percentage | @totalvotes |
| Trump | 52.02 | 694 |

b) rankall(precinct)



|  |
| --- |
| ranks |
| 1056 |

c) rankcounty(precinct)



|  |
| --- |
| @ranking |
| 194 |

d) plotprecinct(precinct)



e) earliestprecinct(vote\_count)



|  |
| --- |
| Precinct |
| HARRIS EAST |

QUESTION 2 – The Candidates

a) precinctswon(candidate)



Results:

<https://rutgers.box.com/s/bhiof7q9raxymdy1hoichw9mo0xadv3f>

b) precinctswoncount(candidate)



|  |
| --- |
| Count(trump) |
| 691 |

c) precinctsfulllead(candidate)



Results:

<https://rutgers.box.com/s/f0rkn2bobk3hfhvegcomxyr2fdrnwxwj>

d) plotcandidate(candidate)

DELIMITER //

CREATE PROCEDURE plotcandidate(IN input\_candidate VARCHAR(255))

BEGIN

DECLARE candidate VARCHAR(255);

SET candidate = input\_candidate;

IF candidate = 'biden' THEN

CREATE TABLE temp(

id INT NOT NULL AUTO\_INCREMENT,

timestamp DATETIME,

votes INT,

PRIMARY KEY (id)

);

INSERT INTO temp (timestamp, votes)

SELECT timestamp, sum(biden) as votes

FROM votes

GROUP BY timestamp

ORDER BY timestamp ASC;

SELECT x.timestamp, IFNULL(y.votes-x.votes, 0) as votes

FROM temp x INNER JOIN temp y ON y.id = x.id + 1;

DROP TABLE temp;

END IF;

IF candidate = 'trump' THEN

CREATE TABLE temp(

id INT NOT NULL AUTO\_INCREMENT,

timestamp DATETIME,

votes INT,

PRIMARY KEY (id)

);

INSERT INTO temp (timestamp, votes)

SELECT timestamp, sum(biden) as votes

FROM votes

GROUP BY timestamp

ORDER BY timestamp ASC;

SELECT x.timestamp, IFNULL(y.votes-x.votes, 0) as votes

FROM temp x INNER JOIN temp y ON y.id = x.id + 1;

DROP TABLE temp;

END IF;

END //

DELIMITER ;

CALL plotcandidate('biden');

END IF;

IF candidate = 'trump' THEN

CREATE TABLE temp(

id INT NOT NULL AUTO\_INCREMENT,

timestamp DATETIME,

votes INT,

PRIMARY KEY (id)

);

INSERT INTO temp (timestamp, votes)

SELECT timestamp, sum(biden) as votes

FROM votes

GROUP BY timestamp

ORDER BY timestamp ASC;

SELECT x.timestamp, IFNULL(y.votes-x.votes, 0) as votes

FROM temp x INNER JOIN temp y ON y.id = x.id + 1;

DROP TABLE temp;

END IF;

END //

DELIMITER ;

CALL plotcandidate('biden');

e) precinctswoncategory()

precinctswontownships()



|  |  |  |  |
| --- | --- | --- | --- |
| Won Township Precincts | Vote Difference | Votes for Biden | Votes for Trump |
| Trump Won | 18821 | 28655 | 47476 |

precinctswonboroughs()



|  |  |  |  |
| --- | --- | --- | --- |
| Won Borough Precincts | Vote Difference | Votes for Biden | Votes for Trump |
| Trump Won | 7597 | 15977 | 23574 |

precinctswonwards()



|  |  |  |  |
| --- | --- | --- | --- |
| Won Ward Precincts | Vote Difference | Votes for Biden | Votes for Trump |
| Biden won | 33211 | 73348 | 40137 |

QUESTION 3 – The Timestamp

a) totalvotes(timestamp, category)

Results:

<https://rutgers.box.com/s/drc99lr62ckoyhu3djsgwyxuee5bvftf>

DELIMITER //

CREATE PROCEDURE totalvotes(IN input\_timestamp DATETIME, IN input\_category varchar(255))

BEGIN

DECLARE category varchar(255);

SET category = input\_category;

IF category = 'all' THEN

SELECT precinct, totalvotes

FROM votes

WHERE timestamp = input\_timestamp

ORDER BY totalvotes DESC;

END IF;

IF category = 'biden' THEN

SELECT precinct, biden

FROM votes

WHERE timestamp = input\_timestamp

ORDER BY biden DESC;

END IF;

IF category = 'trump' THEN

SELECT precinct, trump

FROM votes

WHERE timestamp = input\_timestamp

ORDER BY trump DESC;

END IF;

END //

DELIMITER ;

CALL totalvotes('2020-11-04 03:58:36', 'biden');

b) gaindelta(timestamp)



|  |  |  |
| --- | --- | --- |
| DELTA | GAIN | GAIN/DELTA |
| 00:06:41 | 8927 | 13.9267 |

c) ranktimestamp()



Results: <https://rutgers.box.com/s/f79qx8ed0std3tcg34s6j8j94pvno4md>

d) votesperday(day)



|  |  |  |
| --- | --- | --- |
| Bidenvotes | Trumpvotes | Allvotes |
| 111237 | 26239 | 139328 |

QUESTION 4 – Suspicious or Interesting Data

SELECT precinct

FROM (SELECT precinct, totalvotes

FROM votes x, (SELECT max(timestamp) as maxtime FROM votes)x

WHERE timestamp = maxtime)b

WHERE b.precinct not in (SELECT precinct FROM votes x, (SELECT min(timestamp) as mintime FROM votes)x WHERE timestamp = mintime)

Results:

<https://rutgers.box.com/s/b613b0o27yxjath1rm4oyjyqim39vghd>

What I found was that many precincts did not have data at the minimum timestamp. This is suspicious because the data is missing while other precincts have values at the minimum timestamp even if the value is zero.

PART 3

QUESTION A



QUESTION B



QUESTION C



PART 4

ALTER TABLE place ADD PRIMARY KEY (precinct)

ALTER TABLE votes ADD PRIMARY KEY (precinct, timestamp)

ALTER TABLE votes

ADD FOREIGN KEY (precinct) REFERENCES place(precinct)

CONSTRAINTS FROM PART 3 AS PROCEDURES

A



B



C





INSERT PROCEDURE





UPDATE PROCEDURE









DELETE PROCEDURE







4.1 – Triggers and Update driven Stored Procedures

TABLE: place





TABLE: votes





4.2 Stored Procedures







VIDEO RECORDING

<https://rutgers.box.com/s/ns5e55xc1ik2yxqxrsr41dcykxkr7wz2>