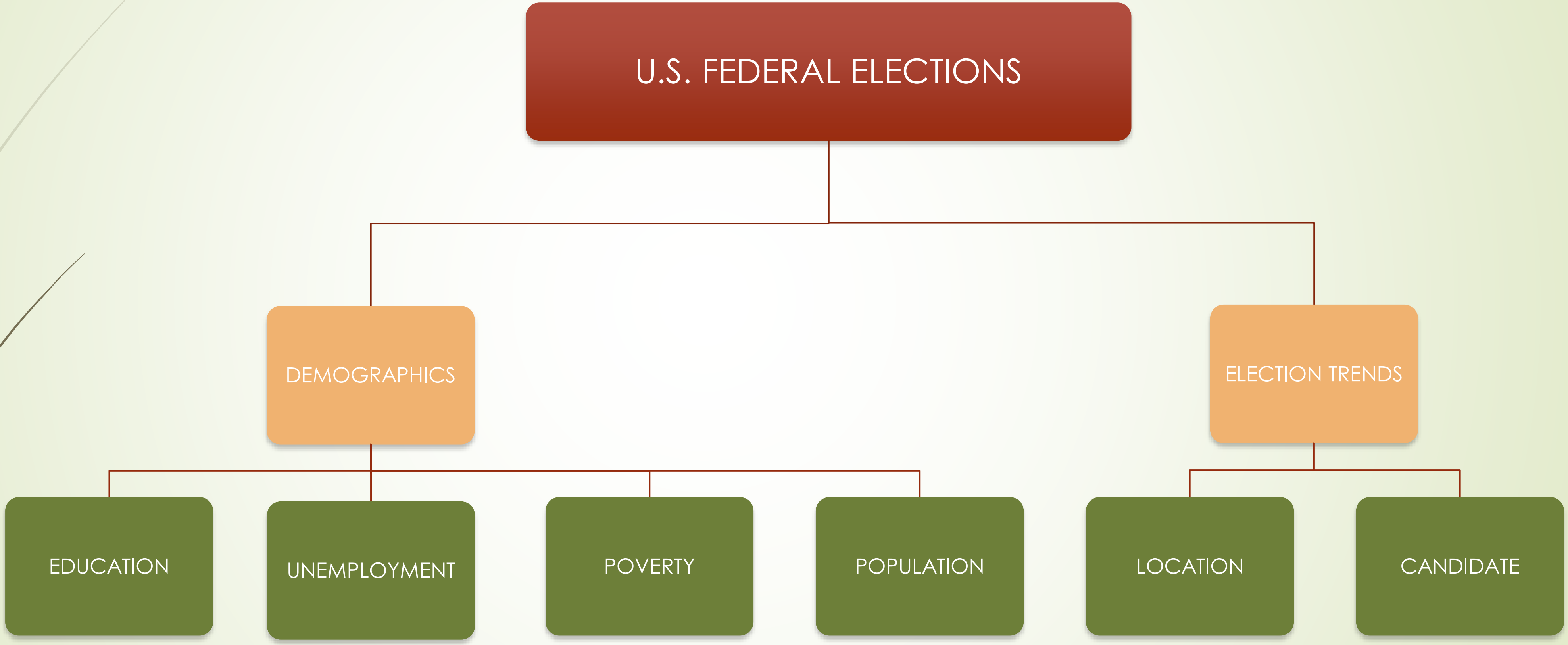
Team Poutine

By: Yusaf Hasan, Muaaz Ahmad

PROBLEM



U.S. FEDERAL ELECTIONS

DEMOGRAPHICS

ELECTION TRENDS

EDUCATION

UNEMPLOYMENT

POVERTY

POPULATION

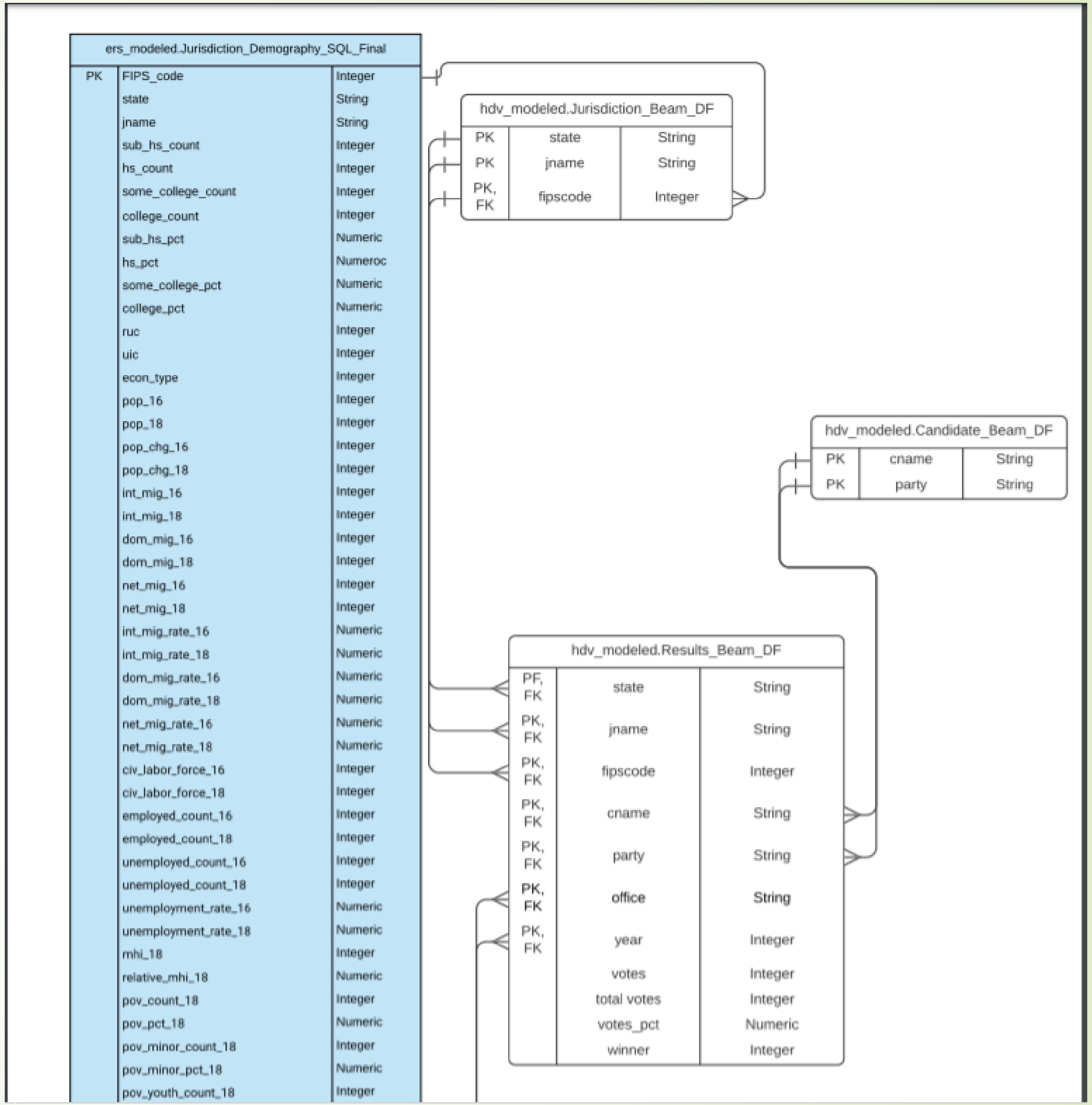
LOCATION

CANDIDATE

DATASETS

- Dataset 1 (hdv_modeled)
  - 2016 and 2018 U.S. Federal Election Data
    - House
    - Senate
    - Presidential
- Dataset 2 (ers_modeled)
  - 2016 and 2018 U.S. Demographic Data
    - Population Estimates (i.e. population change, migration rates)
    - Poverty Estimates (i.e. number of people in poverty, number of minors in poverty)
    - Education (i.e. high school education, college education)
    - Unemployment Estimates (i.e. unemployment rates)

MODELED TABLES



ers_modeled.Jurisdiction_Demography_SQL_Final

PK	FIPS_code	Integer
	state	String
	jname	String
	sub_hs_count	Integer
	hs_count	Integer
	some_college_count	Integer
	college_count	Integer
	sub_hs_pct	Numeric
	hs_pct	Numeroc
	some_college_pct	Numeric
	college_pct	Numeric
	ruc	Integer
	uic	Integer
	econ_type	Integer
	pop_16	Integer
	pop_18	Integer
	pop_chg_16	Integer
	pop_chg_18	Integer
	int_mig_16	Integer
	int_mig_18	Integer
	dom_mig_16	Integer
	dom_mig_18	Integer
	net_mig_16	Integer
	net_mig_18	Integer
	int_mig_rate_16	Numeric
	int_mig_rate_18	Numeric
	dom_mig_rate_16	Numeric
	dom_mig_rate_18	Numeric
	net_mig_rate_16	Numeric
	net_mig_rate_18	Numeric
	civ_labor_force_16	Integer
	civ_labor_force_18	Integer
	employed_count_16	Integer
	employed_count_18	Integer
	unemployed_count_16	Integer
	unemployed_count_18	Integer
	unemployment_rate_16	Numeric
	unemployment_rate_18	Numeric
	mhi_18	Integer
	relative_mhi_18	Numeric
	pov_count_18	Integer
	pov_pct_18	Numeric
	pov_minor_count_18	Integer
	pov_minor_pct_18	Numeric
	pov_youth_count_18	Integer

hdv_modeled.Jurisdiction_Beam_DF

PK	state	String
PK	jname	String
PK, FK	fipscode	Integer

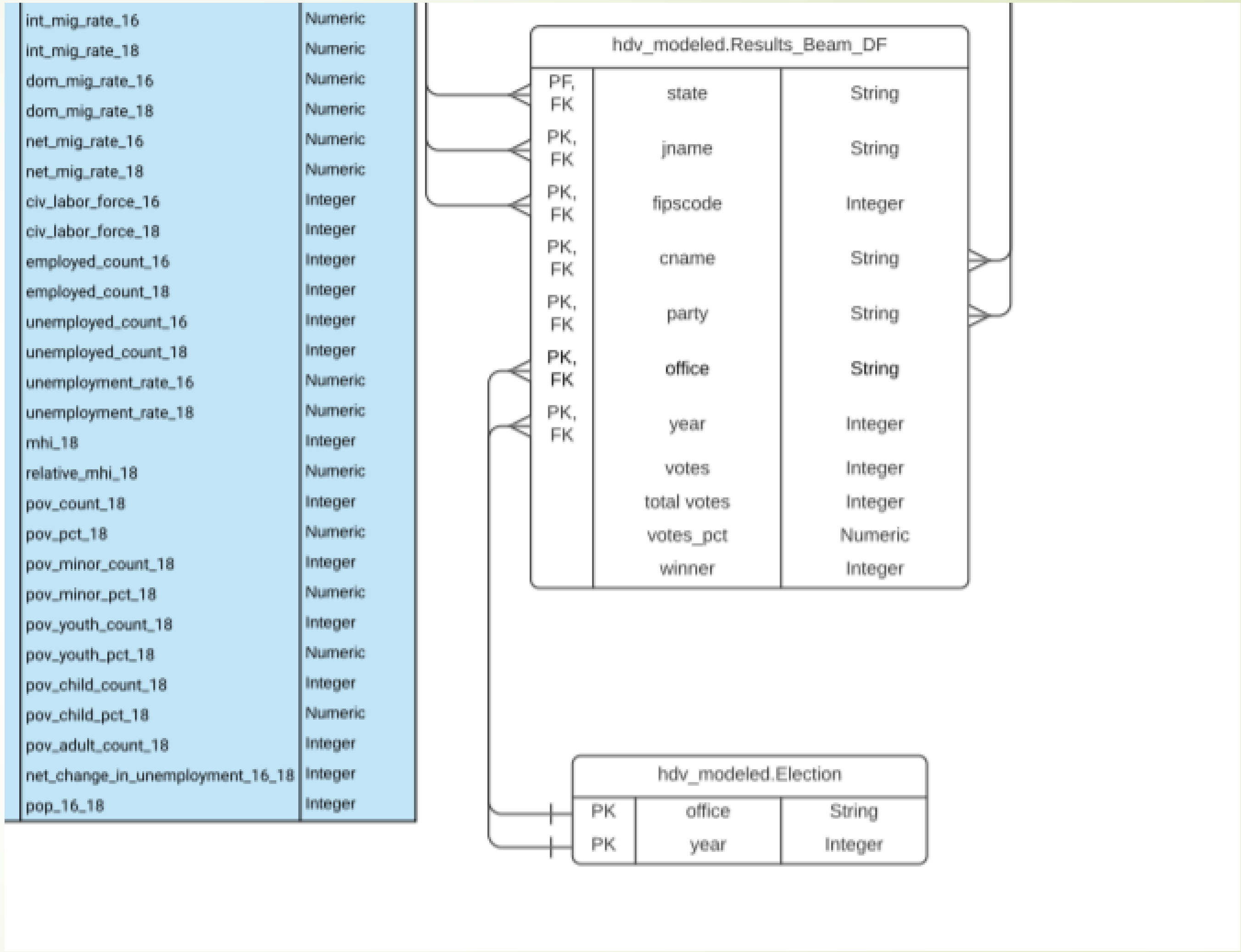
hdv_modeled.Candidate_Beam_DF

PK	cname	String
PK	party	String

hdv_modeled.Results_Beam_DF

PF, FK	state	String
PK, FK	jname	String
PK, FK	fipscode	Integer
PK, FK	cname	String
PK, FK	party	String
PK, FK	office	String
PK, FK	year	Integer
	votes	Integer
	total votes	Integer
	votes_pct	Numeric
	winner	Integer

MODELED TABLES



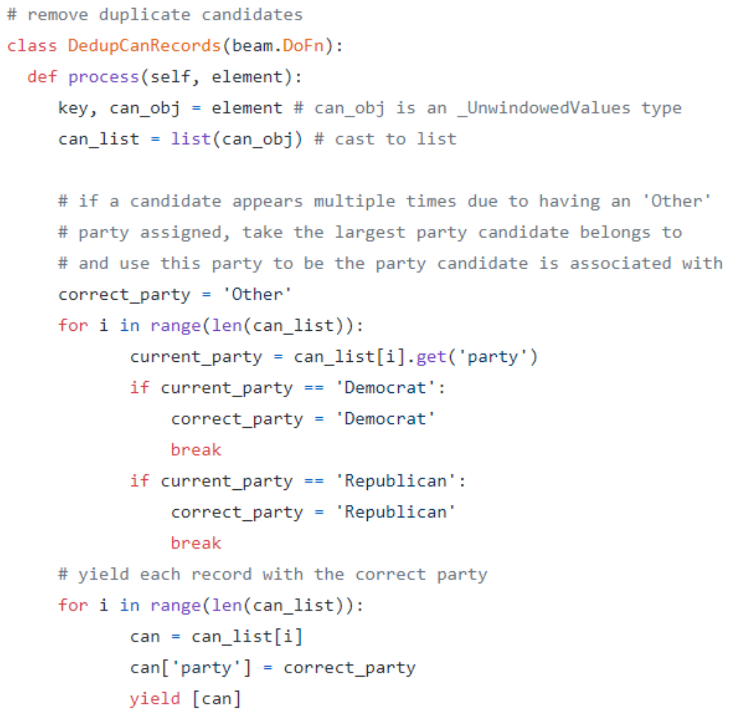
int_mig_rate_16	Numeric
int_mig_rate_18	Numeric
dom_mig_rate_16	Numeric
dom_mig_rate_18	Numeric
net_mig_rate_16	Numeric
net_mig_rate_18	Numeric
civ_labor_force_16	Integer
civ_labor_force_18	Integer
employed_count_16	Integer
employed_count_18	Integer
unemployed_count_16	Integer
unemployed_count_18	Integer
unemployment_rate_16	Numeric
unemployment_rate_18	Numeric
mhi_18	Integer
relative_mhi_18	Numeric
pov_count_18	Integer
pov_pct_18	Numeric
pov_minor_count_18	Integer
pov_minor_pct_18	Numeric
pov_youth_count_18	Integer
pov_youth_pct_18	Numeric
pov_child_count_18	Integer
pov_child_pct_18	Numeric
pov_adult_count_18	Integer
net_change_in_unemployment_16_18	Integer
pop_16_18	Integer

hdv_modeled.Results_Beam_DF

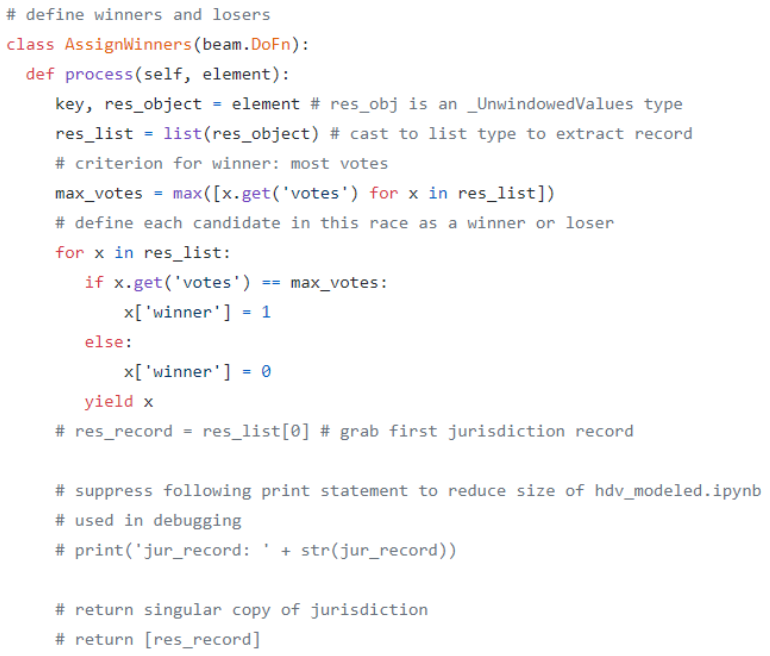
PF, FK	state	String
PK, FK	jname	String
PK, FK	fipscode	Integer
PK, FK	cname	String
PK, FK	party	String
PK, FK	office	String
PK, FK	year	Integer
	votes	Integer
	total votes	Integer
	votes_pct	Numeric
	winner	Integer

hdv_modeled.Election

PK	office	String
PK	year	Integer



```
# remove duplicate candidates
class DedupCanRecords(beam.DoFn):
  def process(self, element):
     key, can_obj = element # can_obj is an _UnwindowedValues type
     can_list = list(can_obj) # cast to list

     # if a candidate appears multiple times due to having an 'Other'
     # party assigned, take the largest party candidate belongs to
     # and use this party to be the party candidate is associated with
     correct_party = 'Other'
     for i in range(len(can_list)):
            current_party = can_list[i].get('party')
            if current_party == 'Democrat':
                correct_party = 'Democrat'
                break
            if current_party == 'Republican':
                correct_party = 'Republican'
                break
     # yield each record with the correct party
     for i in range(len(can_list)):
            can = can_list[i]
            can['party'] = correct_party
            yield [can]
```



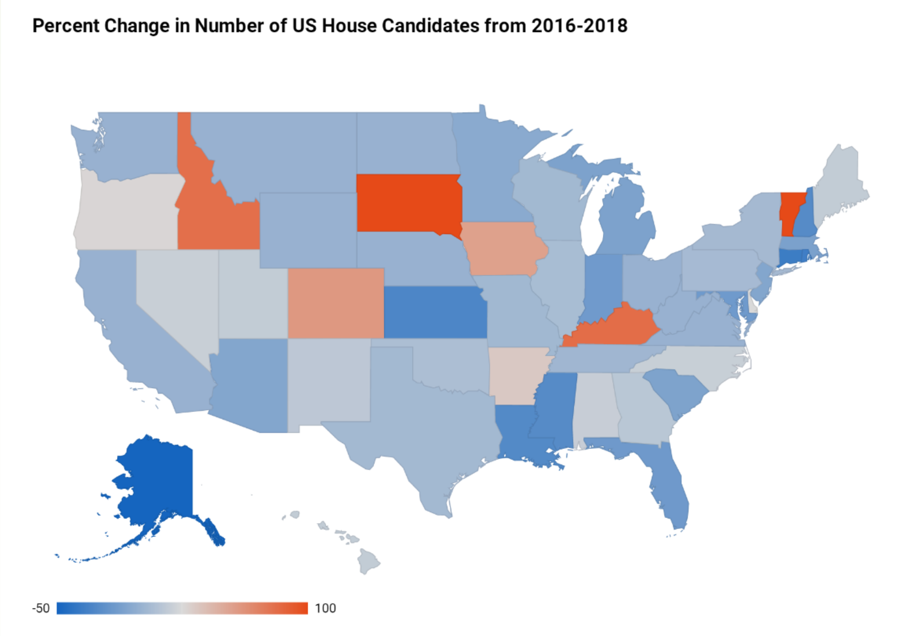
```
# define winners and losers
class AssignWinners(beam.DoFn):
  def process(self, element):
     key, res_object = element # res_obj is an _UnwindowedValues type
     res_list = list(res_object) # cast to list type to extract record
     # criterion for winner: most votes
     max_votes = max([x.get('votes') for x in res_list])
     # define each candidate in this race as a winner or loser
     for x in res_list:
        if x.get('votes') == max_votes:
            x['winner'] = 1
        else:
            x['winner'] = 0
        yield x
     # res_record = res_list[0] # grab first jurisdiction record

     # suppress following print statement to reduce size of hdv_modeled.ipynb
     # used in debugging
     # print('jur_record: ' + str(jur_record))

     # return singular copy of jurisdiction
     # return [res_record]
```

BEAM PIPELINES

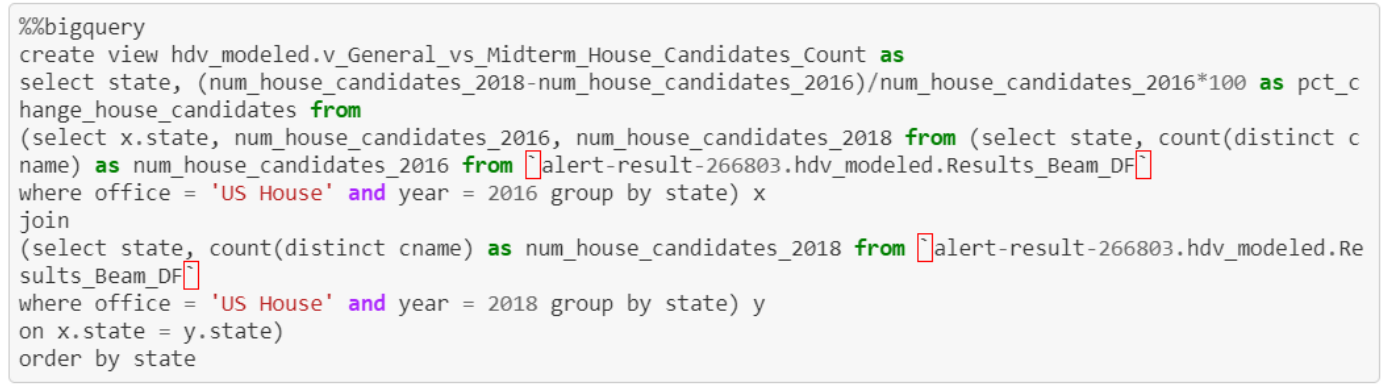
- Executed on Dataset 1
  - Standardizing candidate names, jurisdiction names, party affiliations
  - Removed duplicates
  - For the Results table, a winner column was added

DATASET 1 QUERIES

- Queries investigated simple election trends in both election years
- Example Query
  - Percent change in US House candidates can shed insight on population change



Percent Change in Number of US House Candidates from 2016-2018

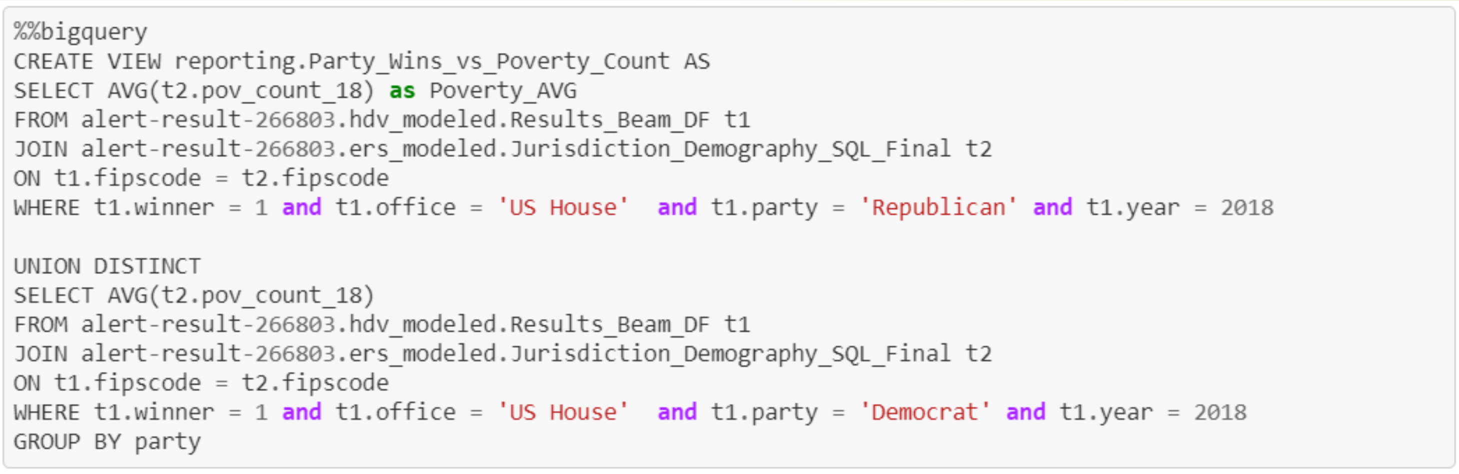


```
%%bigquery
create view hdv_modeled.v_General_vs_Midterm_House_Candidates_Count as
select state, (num_house_candidates_2018-num_house_candidates_2016)/num_house_candidates_2016*100 as pct_c
hange_house_candidates from
(select x.state, num_house_candidates_2016, num_house_candidates_2018 from (select state, count(distinct c
name) as num_house_candidates_2016 from `alert-result-266803.hdv_modeled.Results_Beam_DF`
where office = 'US House' and year = 2016 group by state) x
join
(select state, count(distinct cname) as num_house_candidates_2018 from `alert-result-266803.hdv_modeled.Re
sults_Beam_DF`
where office = 'US House' and year = 2018 group by state) y
on x.state = y.state)
order by state
```

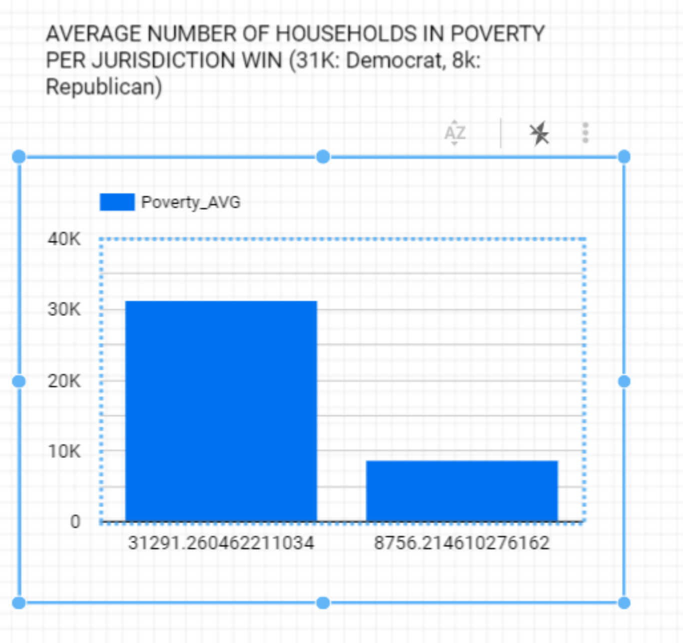
Query View and Code

CROSS DATASET QUERIES

- Query I
  - Compared unemployment rates in each jurisdiction against corresponding 2016 presidential election results
- Query 2
  - Compares House 2016 and House 2018 election results in jurisdiction with at least 5% growth in population
- Query 3
  - Compares average number of household in poverty for Republican and Democratic wins in the 2018 House election



```
%%bigquery
CREATE VIEW reporting.Party_Wins_vs_Poverty_Count AS
SELECT AVG(t2.pov_count_18) as Poverty_AVG
FROM alert-result-266803.hdv_modeled.Results_Beam_DF t1
JOIN alert-result-266803.ers_modeled.Jurisdiction_Demography_SQL_Final t2
ON t1.fipscode = t2.fipscode
WHERE t1.winner = 1 and t1.office = 'US House'  and t1.party = 'Republican' and t1.year = 2018

UNION DISTINCT
SELECT AVG(t2.pov_count_18)
FROM alert-result-266803.hdv_modeled.Results_Beam_DF t1
JOIN alert-result-266803.ers_modeled.Jurisdiction_Demography_SQL_Final t2
ON t1.fipscode = t2.fipscode
WHERE t1.winner = 1 and t1.office = 'US House'  and t1.party = 'Democrat' and t1.year = 2018
GROUP BY party
```



AVERAGE NUMBER OF HOUSEHOLDS IN POVERTY PER JURISDICTION WIN (31K: Democrat, 8k: Republican)

Query 3 View and Code

AIRFLOW DAG ON DATASET 2

CONSOLIDATION

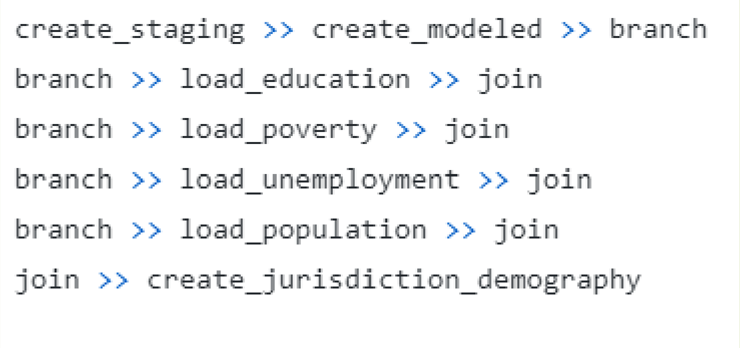
- Jurisdiction_Demography Table

NEW COLUMNS

- pop_16_18
- net_change_in_unemployment_16_18
- pov_adult_count_18

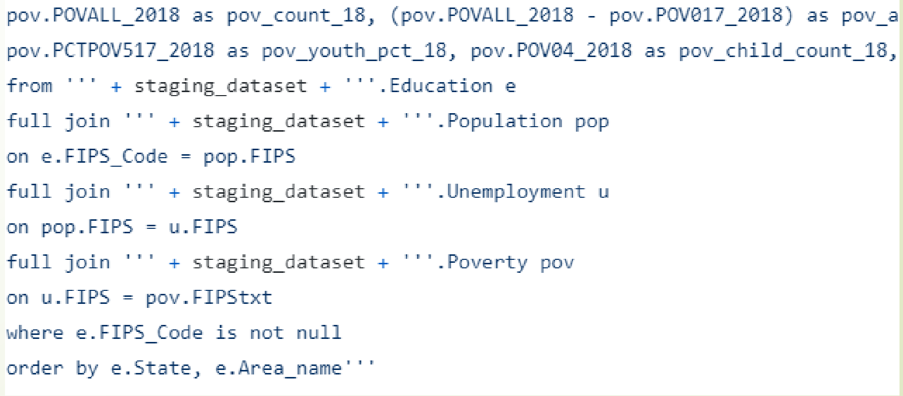
DAG

- Parallel Process: Loading CSV Files
- Dependent Process: Creating demography table



```
create_staging >> create_modeled >> branch
branch >> load_education >> join
branch >> load_poverty >> join
branch >> load_unemployment >> join
branch >> load_population >> join
join >> create_jurisdiction_demography
```

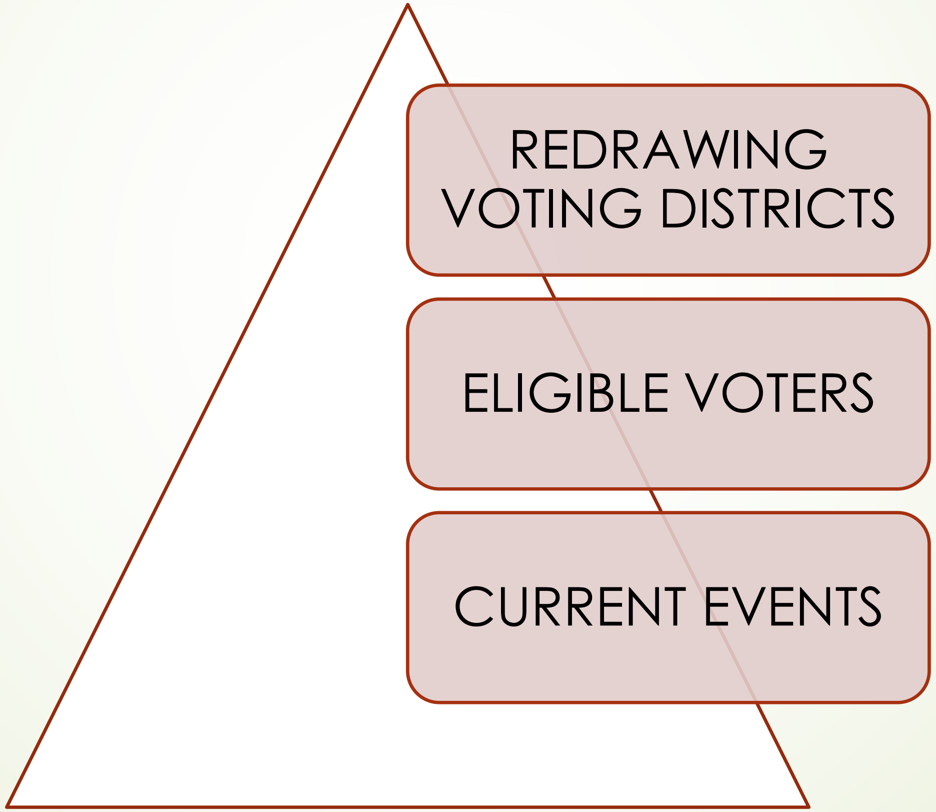
DAG



```
pov.POVALL_2018 as pov_count_18, (pov.POVALL_2018 - pov.POV017_2018) as pov_a
pov.PCTPOV517_2018 as pov_youth_pct_18, pov.POV04_2018 as pov_child_count_18,
from ''' + staging_dataset + '''.Education e
full join ''' + staging_dataset + '''.Population pop
on e.FIPS_Code = pop.FIPS
full join ''' + staging_dataset + '''.Unemployment u
on pop.FIPS = u.FIPS
full join ''' + staging_dataset + '''.Poverty pov
on u.FIPS = pov.FIPStxt
where e.FIPS_Code is not null
order by e.State, e.Area_name'''
```

Modeled Jurisdiction_Demography Table

FUTURE IMPROVEMENTS



REDRAWING VOTING DISTRICTS

ELIGIBLE VOTERS

CURRENT EVENTS