

Analysis of CA State Demography

Background:

Through this project, I would like to analyze the demography of California state to find the following:

1. Predicting migration patterns of various ethnical groups. For an example, increase/decrease in populations of the Hispanic and Latino Americans in a county.
2. The county with fastest growing aging population.
3. Median age of the county for different ethnicity.

The required data is obtained from Open Data portal of California. ^[1] The data contains population projections by county, age, gender, and ethnicity. It takes 2016th population as a baseline for the prediction. The population projection data is available from the year 2010 to 2060. The data contains 2,099,650 records.

Motivation:

Understanding demography provides better insights for organizations. Business owners often require demographic information to use in the business plans for raising the funds. For businesses, it is easier and effective to target marketing to groups of people with similar characteristics over targeting to every individual. To give an example, health care products can target the county with aging population and cosmetics companies can look for the age group of 15 - 40. Another example can be giving insight regarding the expected future appreciation/depreciation of immovable assets such as house, property by knowing the rate of migration.

Proposed methods:

- Storing the data in MongoDB.
- Processing the information using Apache Hadoop MapReduce in the local system.
- Creating HBase clusters using Amazon EMR services for remote access.
- Providing a simple UI to query the information.

Preliminary results/progress:

Created a database and stored the data as we need only a partial data for certain queries and having a database will help for fast processing.

```
{ "id": ObjectId("58c8d24dfceeb63a0fcd6c48"), "className": "com.se.data.DemographyEntry", "year": 2053, "ethnicity": "\"Multiracial (two or more of above races), Non-Hispanic\"", "ethnicityCode": 6, "gender": "Female", "age": 99, "county": "El Dorado", "countyCode": 6017, "population": NumberLong(2) }
{ "id": ObjectId("58c8d24dfceeb63a0fcd6c49"), "className": "com.se.data.DemographyEntry", "year": 2053, "ethnicity": "\"Multiracial (two or more of above races), Non-Hispanic\"", "ethnicityCode": 6, "gender": "Female", "age": 101, "county": "El Dorado", "countyCode": 6017, "population": NumberLong(2) }
{ "id": ObjectId("58c8d24dfceeb63a0fcd6c4a"), "className": "com.se.data.DemographyEntry", "year": 2053, "ethnicity": "\"Multiracial (two or more of above races), Non-Hispanic\"", "ethnicityCode": 6, "gender": "Female", "age": 103, "county": "El Dorado", "countyCode": 6017, "population": NumberLong(1) }
{ "id": ObjectId("58c8d24dfceeb63a0fcd6c4b"), "className": "com.se.data.DemographyEntry", "year": 2053, "ethnicity": "Hispanic (any race)", "ethnicityCode": 7, "gender": "Female", "age": 0, "county": "El Dorado", "countyCode": 6017, "population": NumberLong(293) }
{ "id": ObjectId("58c8d24dfceeb63a0fcd6c4c"), "className": "com.se.data.DemographyEntry", "year": 2053, "ethnicity": "Hispanic (any race)", "ethnicityCode": 7, "gender": "Female", "age": 1, "county": "El Dorado", "countyCode": 6017, "population": NumberLong(260) }
{ "id": ObjectId("58c8d24dfceeb63a0fcd6c4d"), "className": "com.se.data.DemographyEntry", "year": 2053, "ethnicity": "Hispanic (any race)", "ethnicityCode": 7, "gender": "Female", "age": 2, "county": "El Dorado", "countyCode": 6017, "population": NumberLong(271) }
{ "id": ObjectId("58c8d24dfceeb63a0fcd6c4e"), "className": "com.se.data.DemographyEntry", "year": 2053, "ethnicity": "Hispanic (any race)", "ethnicityCode": 7, "gender": "Female", "age": 3, "county": "El Dorado", "countyCode": 6017, "population": NumberLong(272) }
Type "it" for more
>
```

Figure 1 Content in the collection

References:

[1] <https://data.ca.gov/dataset/california-population-projection-county-age-gender-and-ethnicity>