‘Q’ THE PEOPLE

Insight into voting trends in Costa Mesa

ADRIAN LLAMAS

ISAAC CARRANZA

HELMA ZECENA

JUNGJE SON

REGGIE PELZ

COSTA MESA
CITY OF THE ARTS
CALIFORNIA

OBJECTIVE

- Use Data Analytics to inform candidate for Costa Mesa Mayor, Quinten Pullen, on the geographical areas where his campaign's efforts may be most effective.

MEET THE CANDIDATE: QUENTIN PULLEN

- Former Marine & Gulf War vet
- Certified personal trainer
- Business owner – “Self Made Training Facility”
- Change candidate

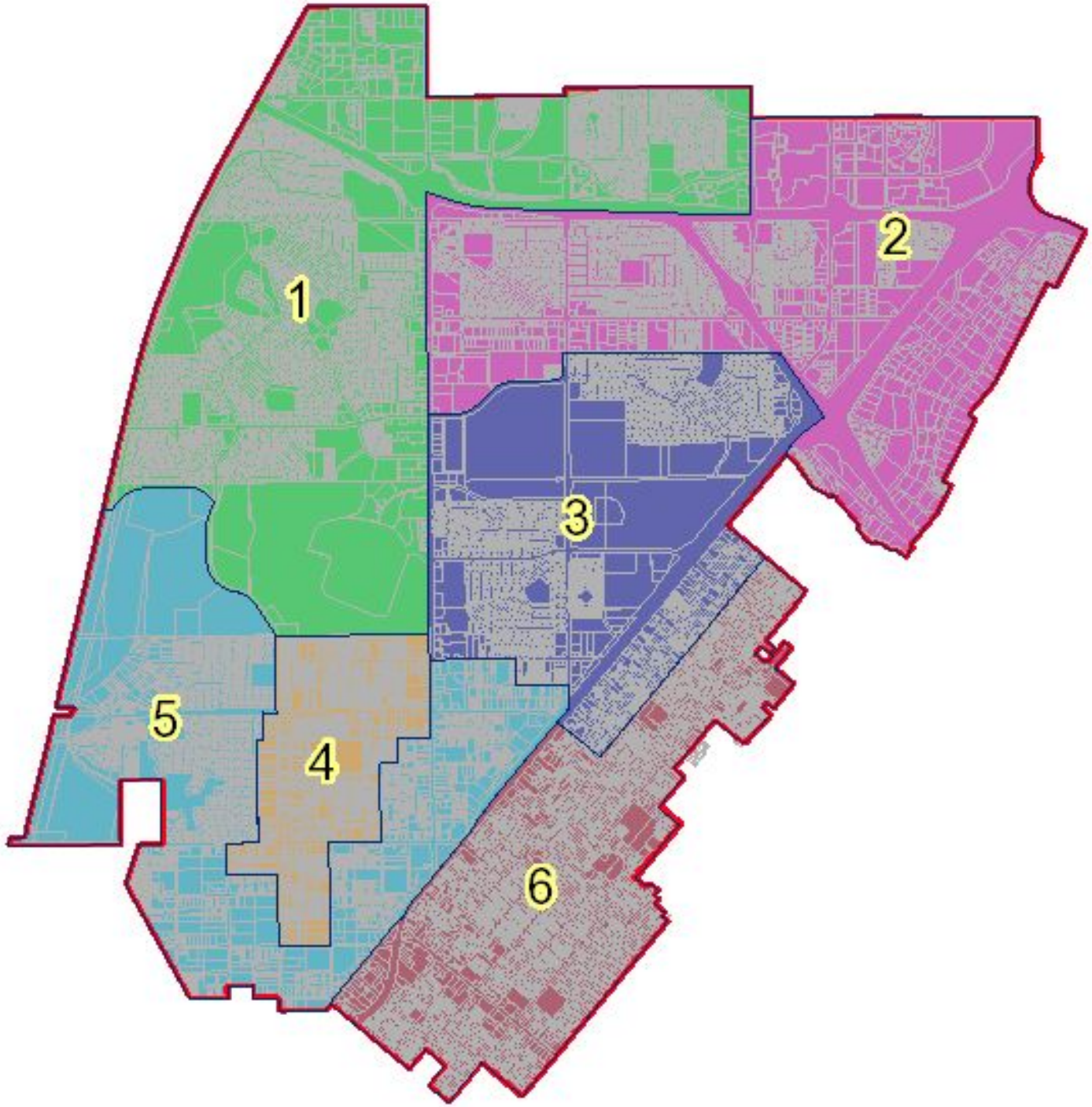
OVERVIEW

- Understanding Costa Mesa's Demography and Voter Participation
- Data Mining
  - Data Sources
  - ETL
  - Machine Learning
- Analysis
- Additional Considerations

UNDERSTANDING COSTA MESA'S DEMOGRAPHY AND VOTER PARTICIPATION

Districts / Wards

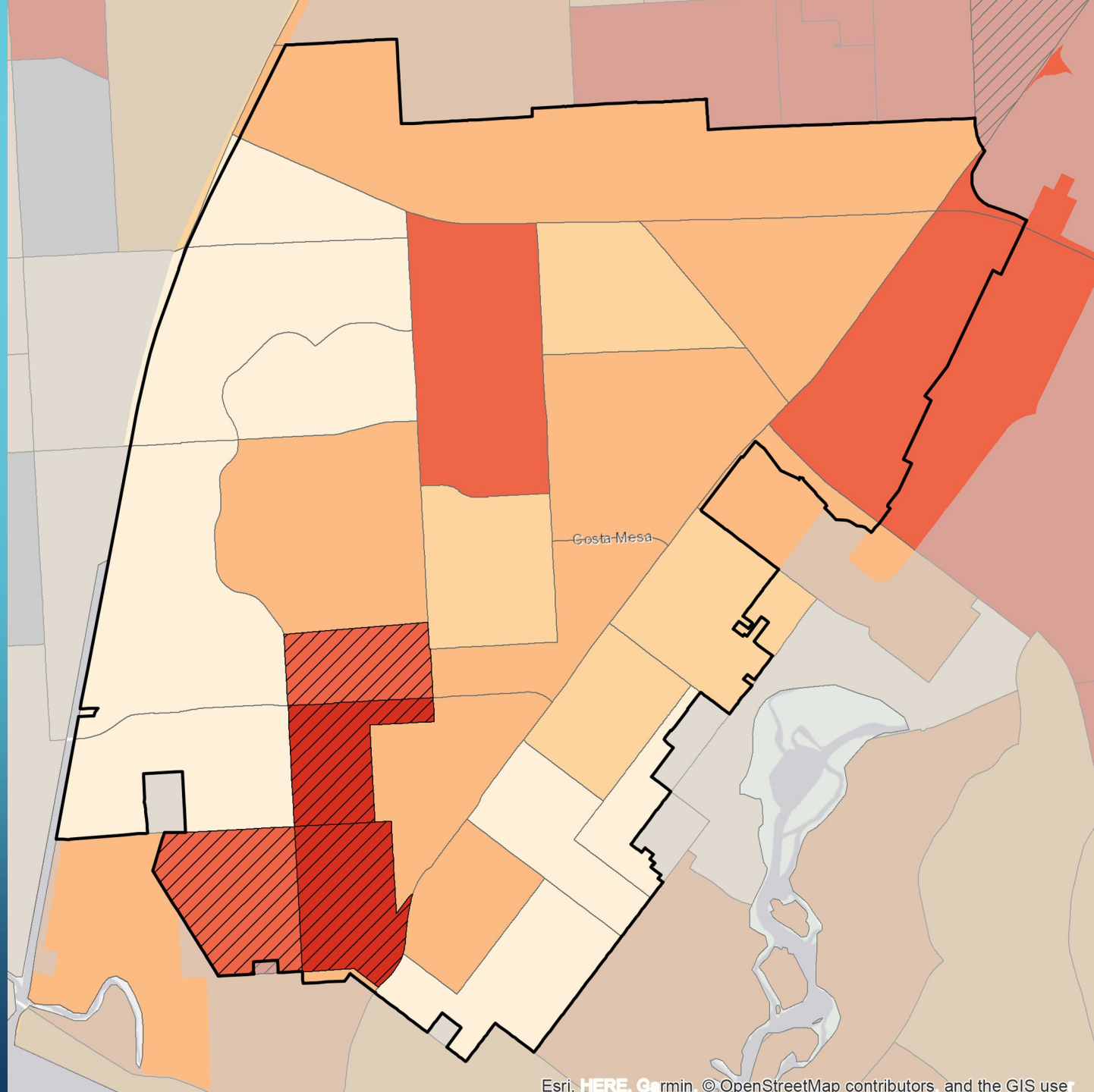
At a high-level, the city is divided into districts, also known as wards. These geographic regions determine where Costa Mesa residents may vote in elections.



UNDERSTANDING COSTA MESA'S DEMOGRAPHY AND VOTER PARTICIPATION

Census Tracts

They are the most common used unit of measurement for the presentation of statistical data. Census tracts generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people.



UNDERSTANDING COSTA MESA'S DEMOGRAPHY AND VOTER PARTICIPATION

Census Blocks / Precincts

They are the area of focus for this project and are the smallest subdivision for which relevant data is available.

DATA MINING

- **Data Sources**
  - Census
  - American Community Survey (ACS) 2018
  - OC Public Works API
  - Orange County Registrar of Voters
  - Esri ArcGIS

DATA MINING

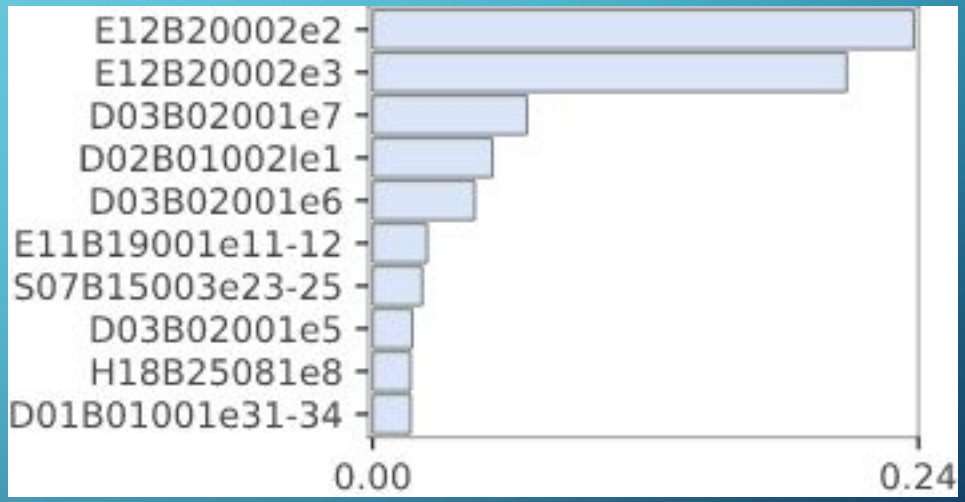
- **ETL**
  - Customized data from OCPW as JSON (Demographic, Economic, Housing, Social) and extracted Costa Mesa data as CSV
  - Found over 400 potential features from four different measures at the block level
  - Performed initial Feature Reduction based on preliminary research on what factors are most important to elections (around 130 features)
  - Google Colab

DATA MINING

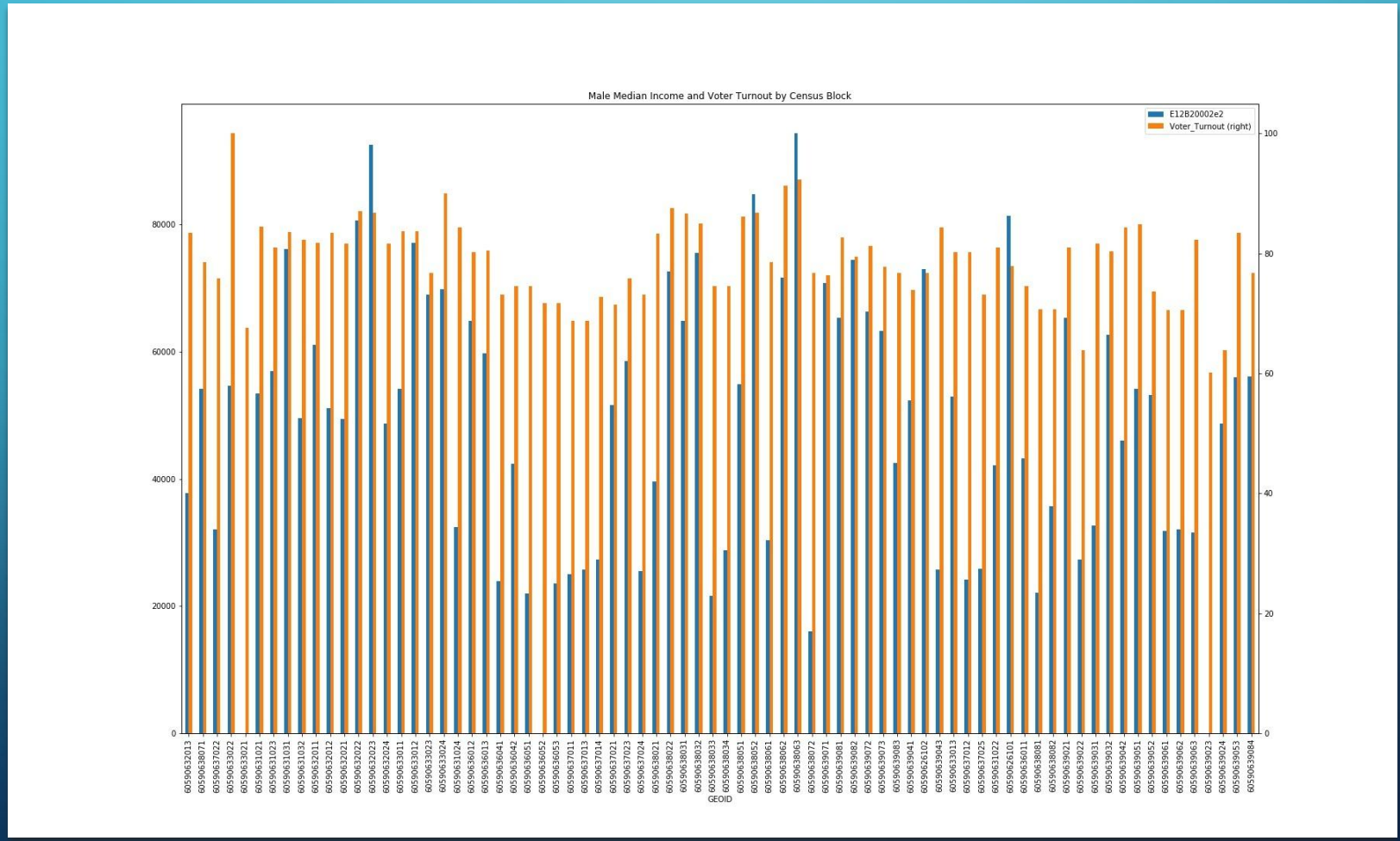
- **Machine Learning**
  - Calculated level of importance (correlation with voter turnout) on our selected features.
  - The scikit-learn Random Forest feature importance strategy was unreliable. To get reliable results, we used permutation importance, provided in the **rfpimp package** in the **src dir**.
  - rfpimp is an increasingly-ill-suited name, but we still like it.
  - Jupyter Notebook

ANALYSIS:
TOP SELECTED FEATURES BY ALGORITHM

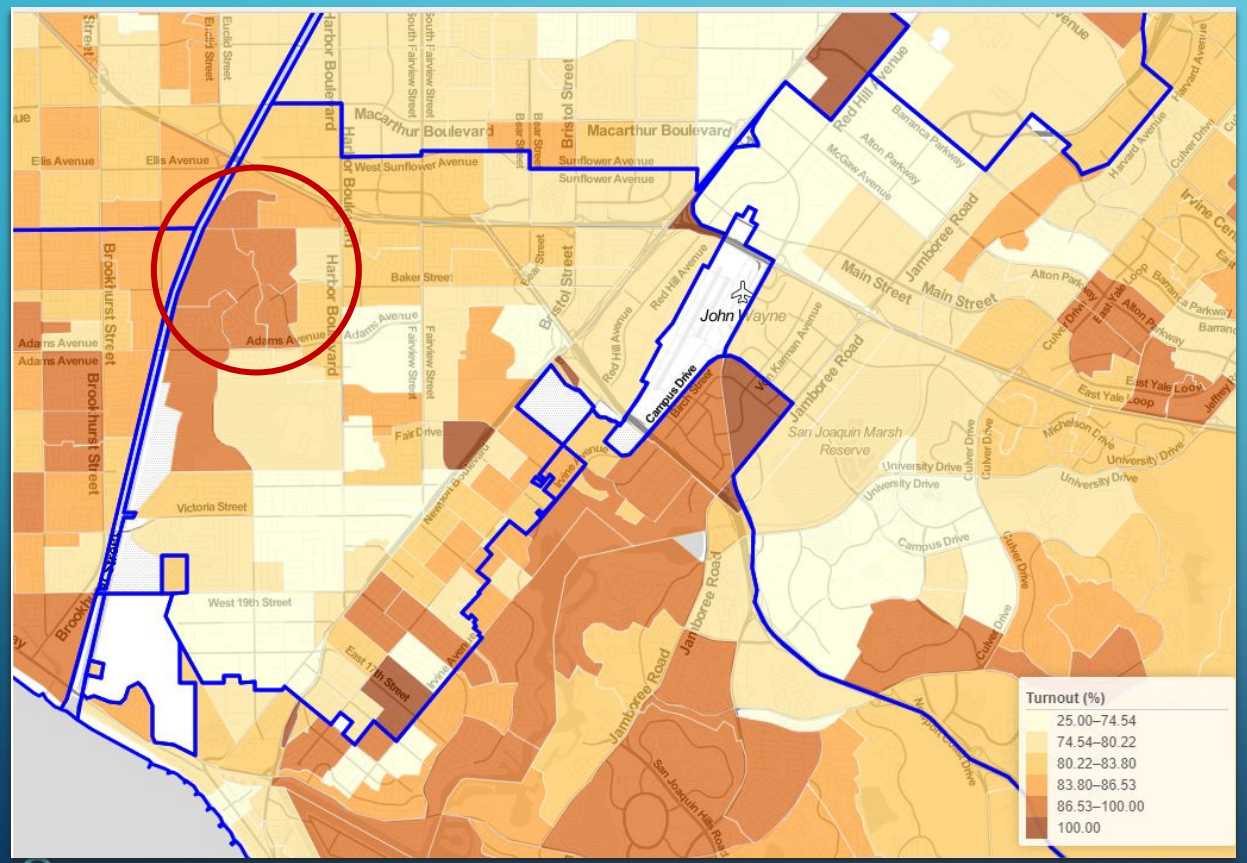
1. Median earnings, male population
2. Median earnings, female population
3. Some other race alone
4. Median age of Hispanic/Latino
5. Native Hawaiian and other PI alone
6. Households, $50,000 to $74,999
7. Masters+
8. Asian alone
9. Housing units without a mortgage



ANALYSIS:
MEDIAN EARNINGS MALE POPULATION (BLUE) VS VOTER TURNOUT (ORANGE)



ANALYSIS:
MEDIAN EARNINGS MALE POPULATION VS VOTER TURNOUT



ADDITIONAL CONSIDERATIONS

- This project serves as a proof of concept for predicting voter participation.
- Due to population and geographic size of Costa Mesa, more data is required to draw stronger conclusions for wider applications.
- Information obtained will help Quentin Pullen target his campaign efforts.