

Farhan Alavi
April 12th, 2024

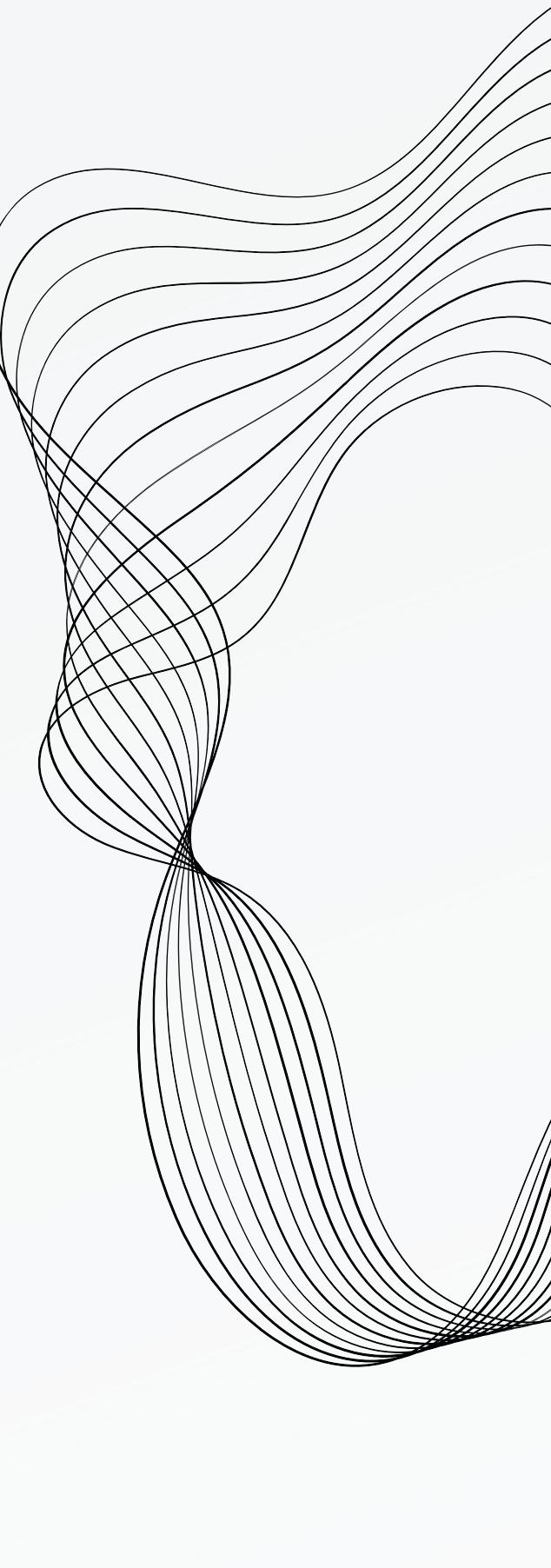
Infectious Disease Cases in California from 2001 to 2022

REINTRODUCTION

There were **511,874** cases reported in California in the last **21 years**.

Does the ethnic makeup of a county impact disease cases?

What can we do with this information?



THE ROAD SO FAR

Sprint 0

Finding an interesting topic + dataset

Data from <https://data.gov/>

Sprint 1:

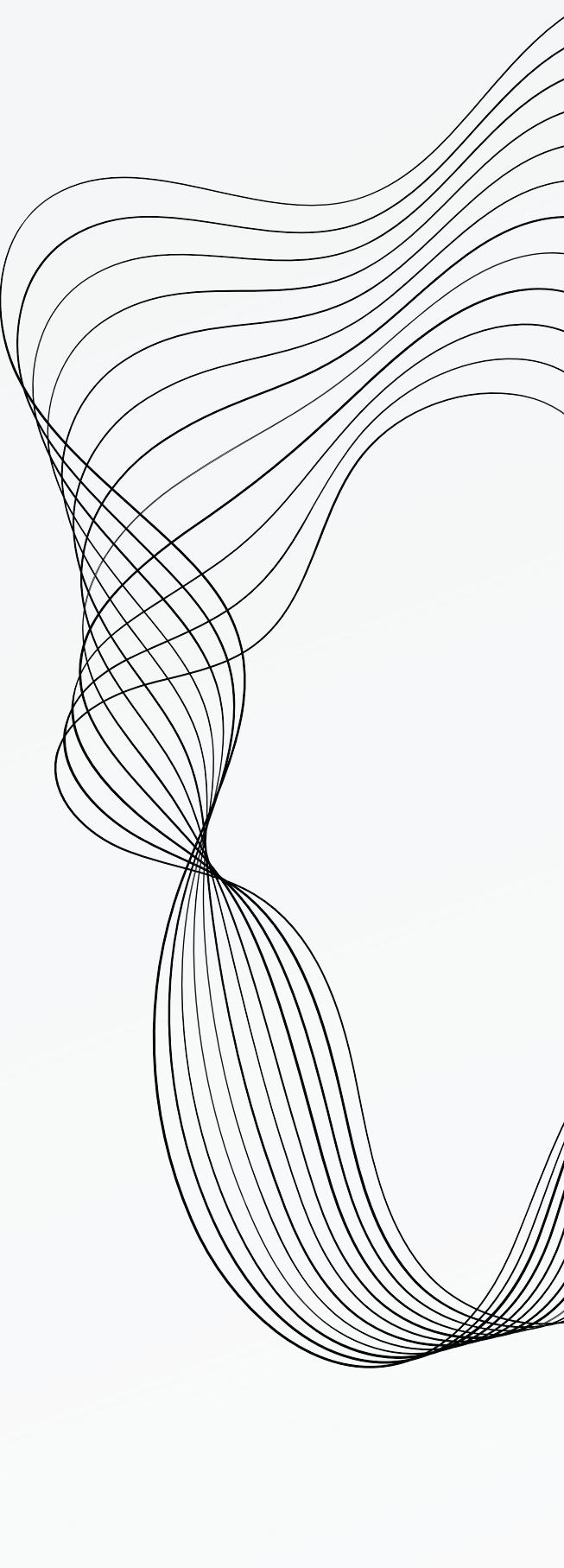
EDA- trimming from 180,000+ rows to
63,000

Sprint 2:

Preprocessing and base modelling- LinReg

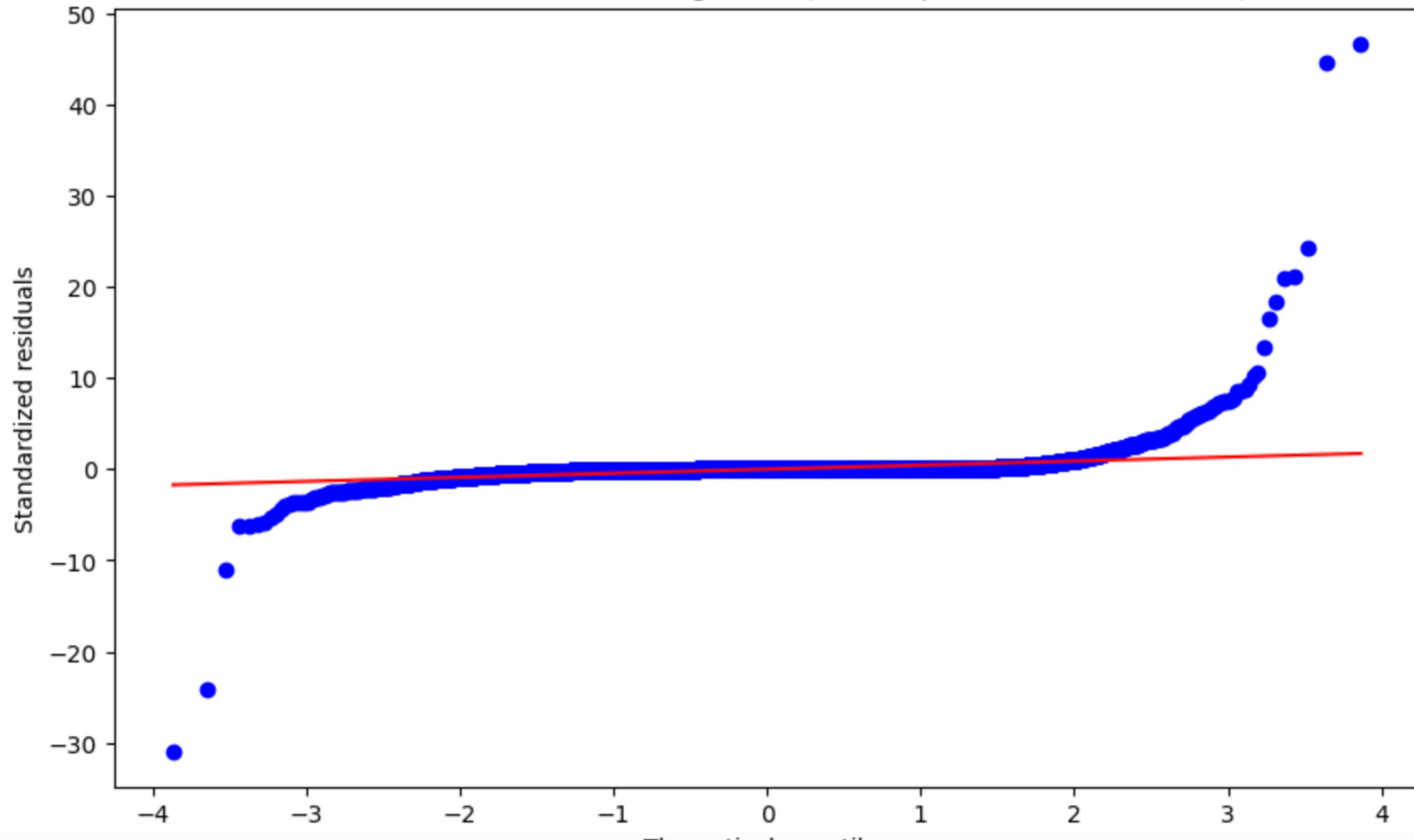
Sprint 3:

Compared LinReg to KNN and DT
Optimizing and evaluating.



DECISION TREE PERFORMANCE

Q-Q Plot for Decision Tree Regressor (Max Depth: 19, Test Residuals)



COMPARING THE MODELS

KNN/Linear Regression

Training MAE: 13.498010701950157

Training MSE: 5170.465471838032

Training R-squared: 0.040140512846411625

Test MAE: 11.840710270807483

Test MSE: 2437.960452981889

Test R-squared: 0.03968677842278867

Decision Tree

Training MAE: 3.903553341367285

Training MSE: 385.679405169057

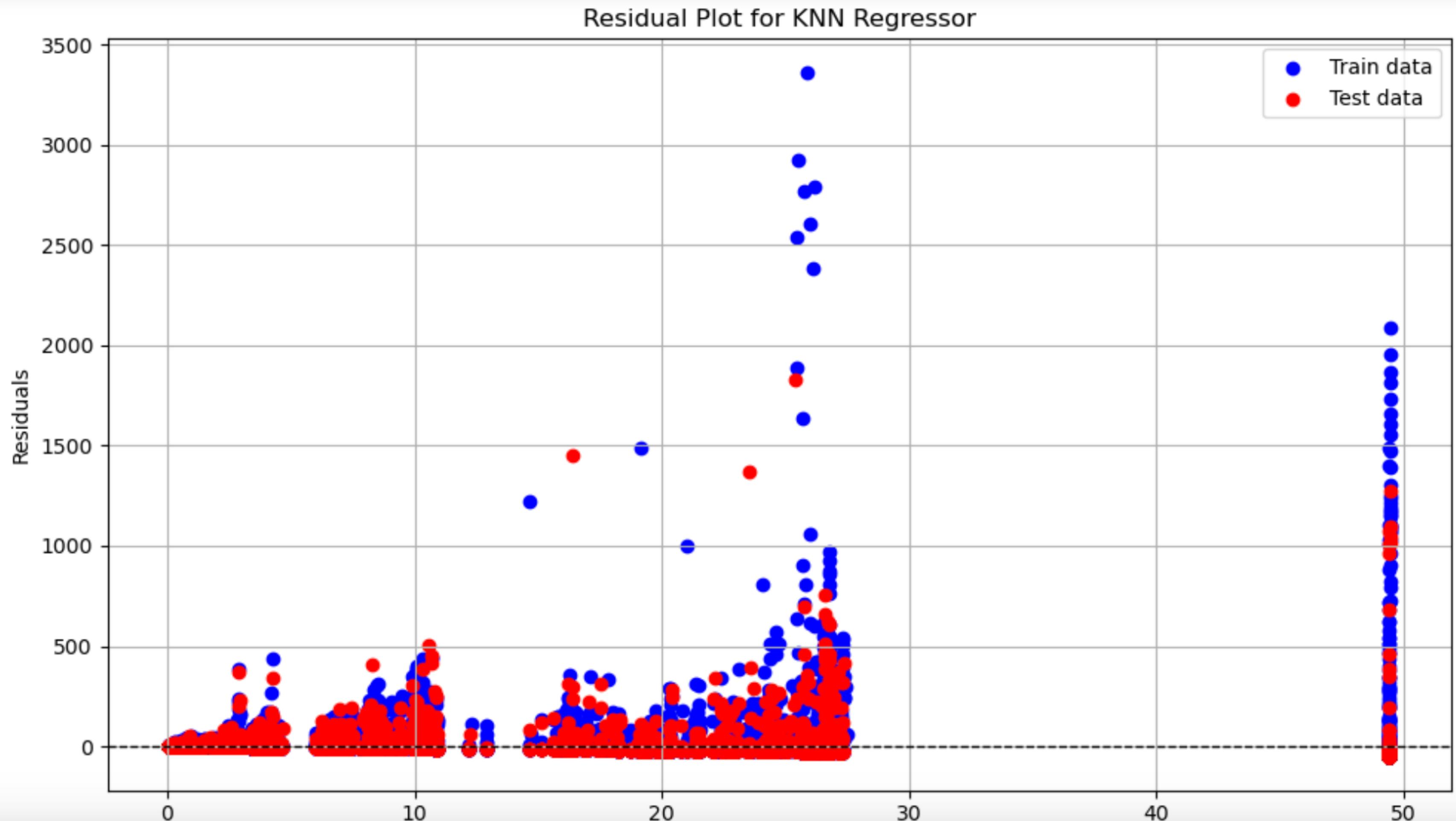
Training R-squared: 0.92840141026613

Test MAE: 4.715665543632342

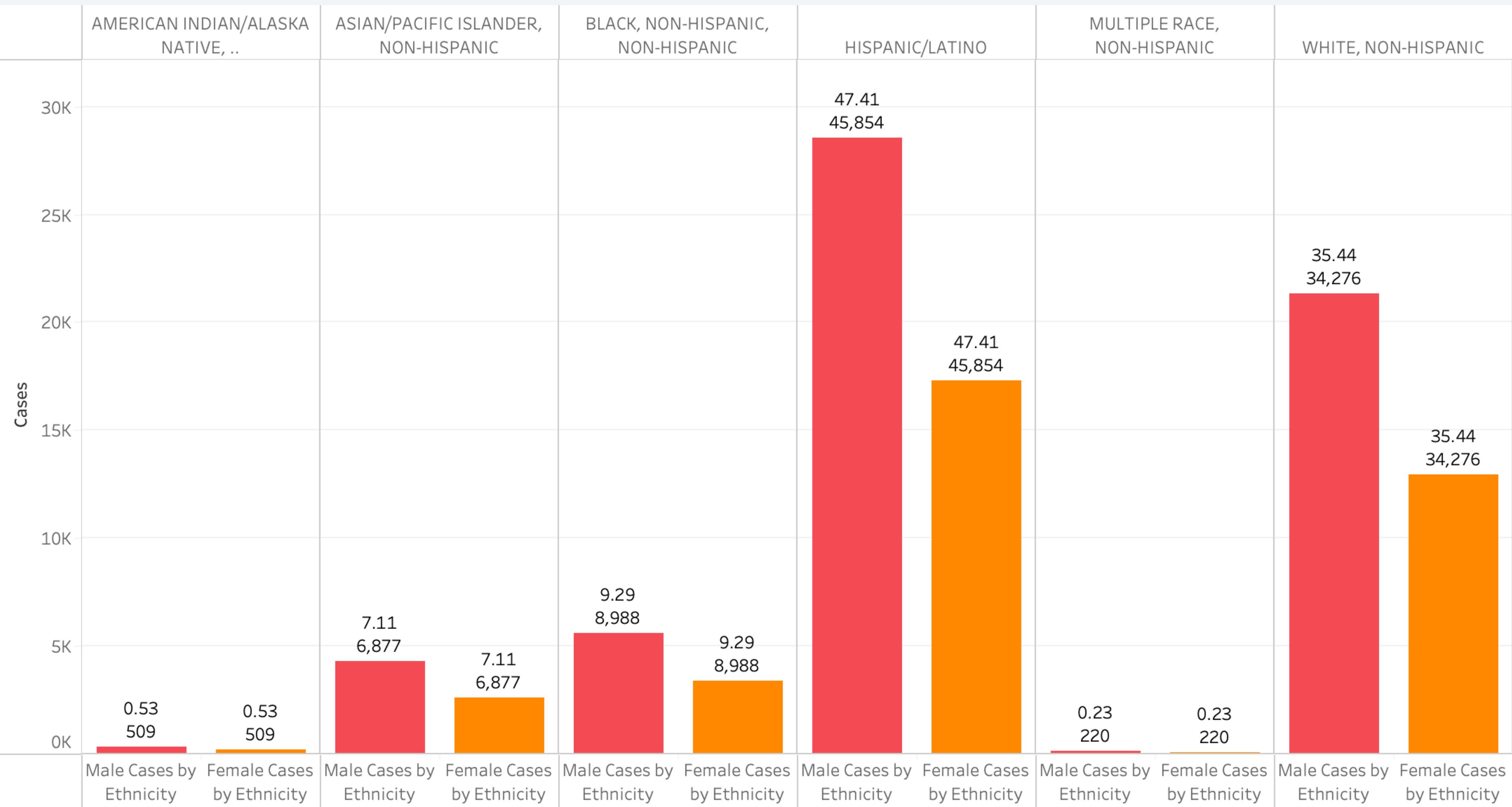
Test MSE: 494.04534152182237

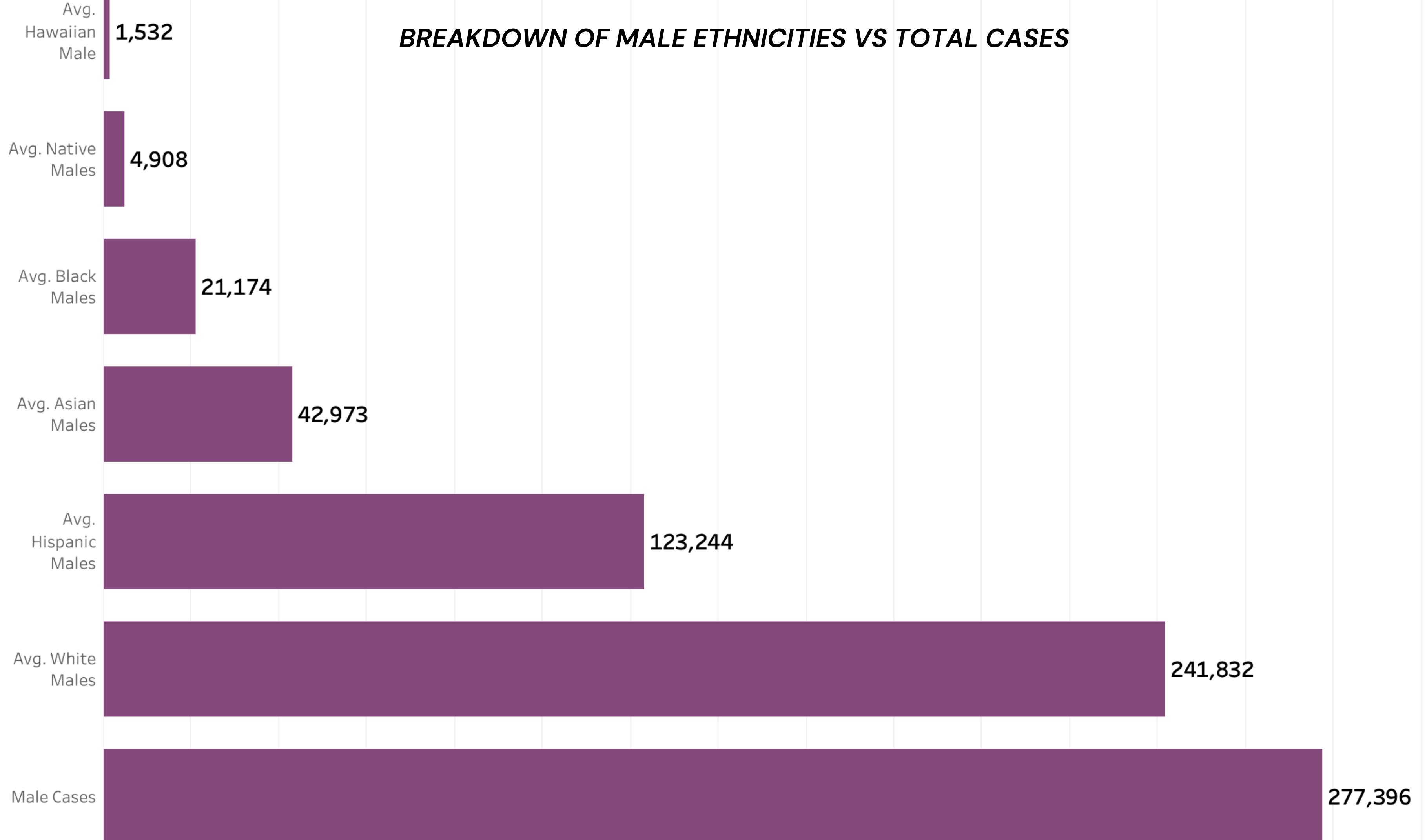
Test R-squared: 0.8053954185590888

KNN PERFORMANCE WITH 2700 NEIGHBORS



VALLEY FEVER ETHNIC BREAKDOWN





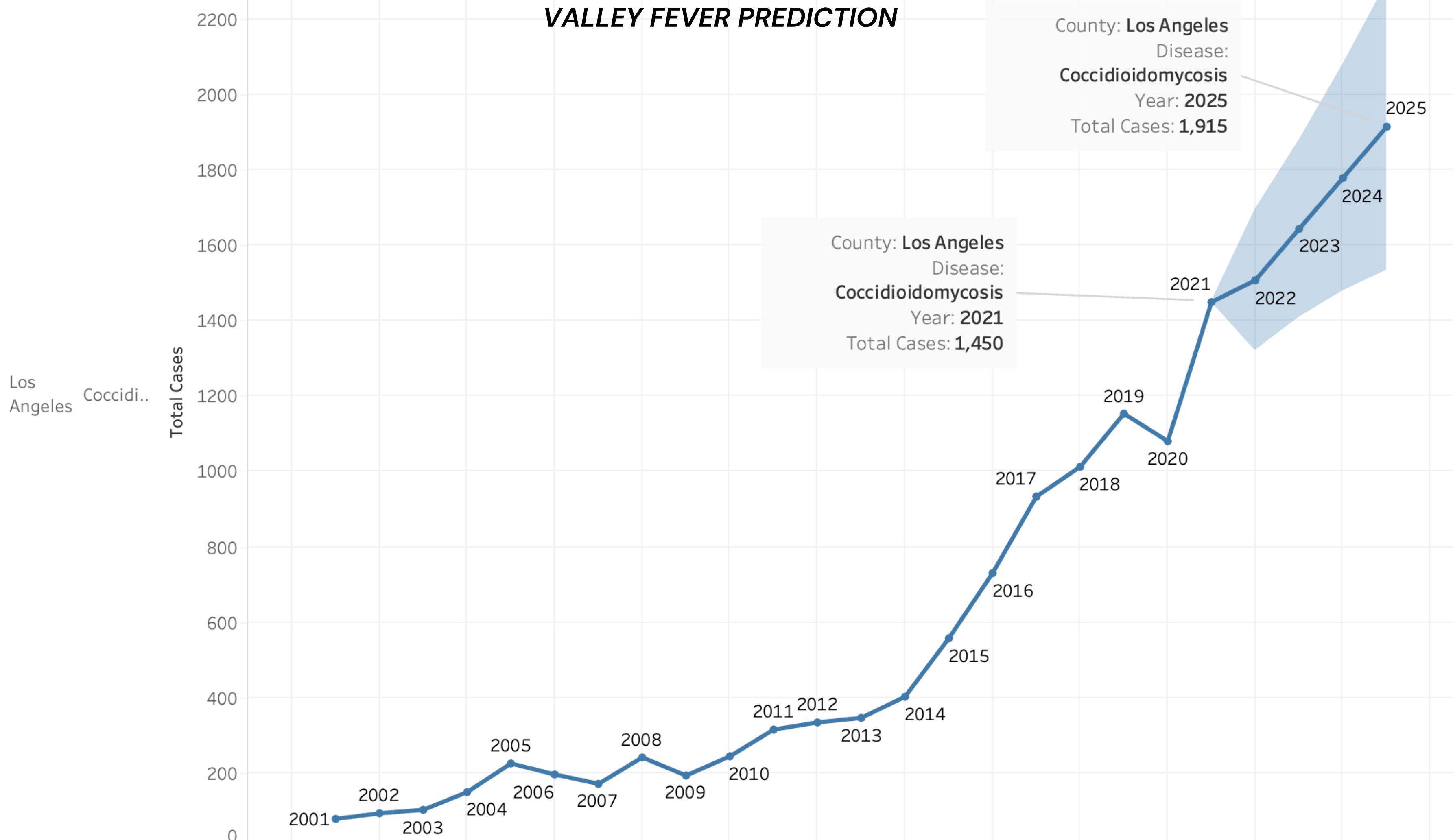
NEXT STEP: STREAMLIT APP

Dropdown menu with diseases,
counties.

Path of disease cases and future
prediction.

Ethnic breakdown?

VALLEY FEVER PREDICTION



**Thanks
for watching!
Questions?**

