

DIAGNOSTIC ANALYTICS
FOR COLLEGE
COMPLETION RATES

BACKGROUND



A College Degree Is The New High
School Diploma

PREDICTIVE MODEL DESIGN / RESULTS



DATA AND METHODOLOGY

Data

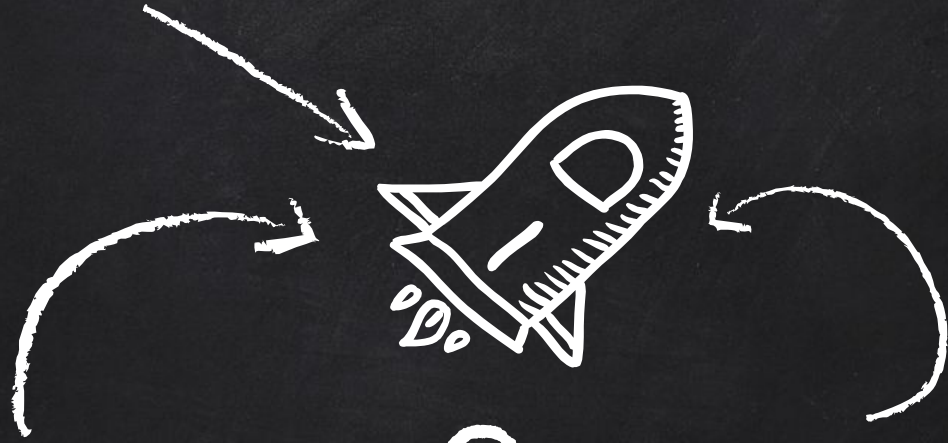
US Department of Education College
Scorecard Data

Approx. 5,000 records from 2009 ,2010,
2011, 2012 and 2013 were sampled

Methods

Supervised Classification Machine
Learning Algorithms –

Decision Tree, Random Forest, Logistic
Regression, Support Vector Machines,
Gradient Boosting Machines, K Nearest
Neighbors



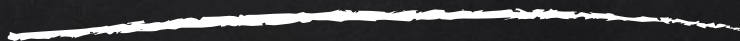
COLLEGE COMPLETION RATES

5113

5113



51.17%

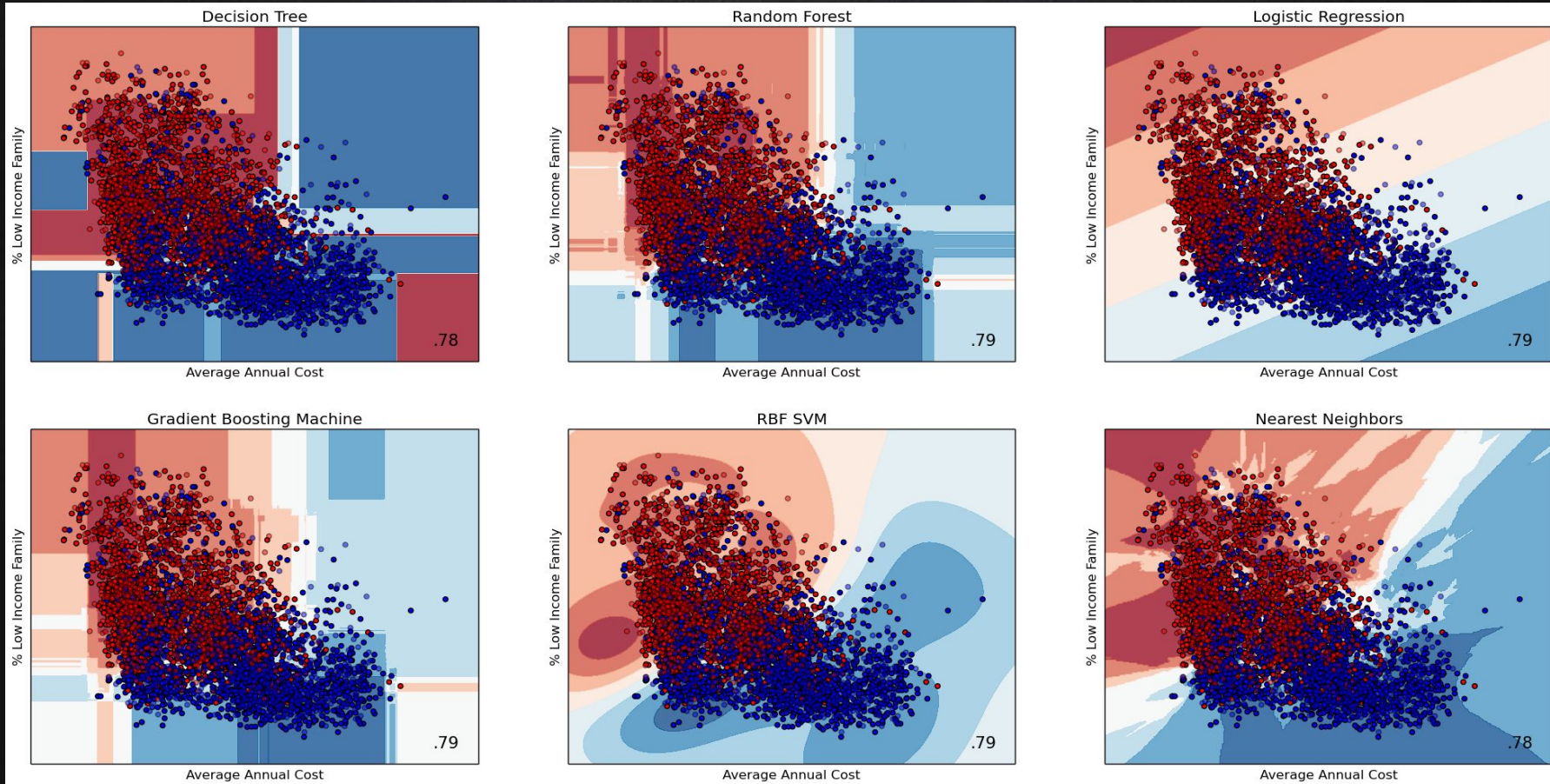




MODEL FEATURES

- X Admission rate
- X Average annual cost
- X % of Student Body – black
- X % of Student Body – low Income family
- X % of Student Body – 1st generation student
- X % of Student Body – received federal loan

MODEL DECISION SURFACES



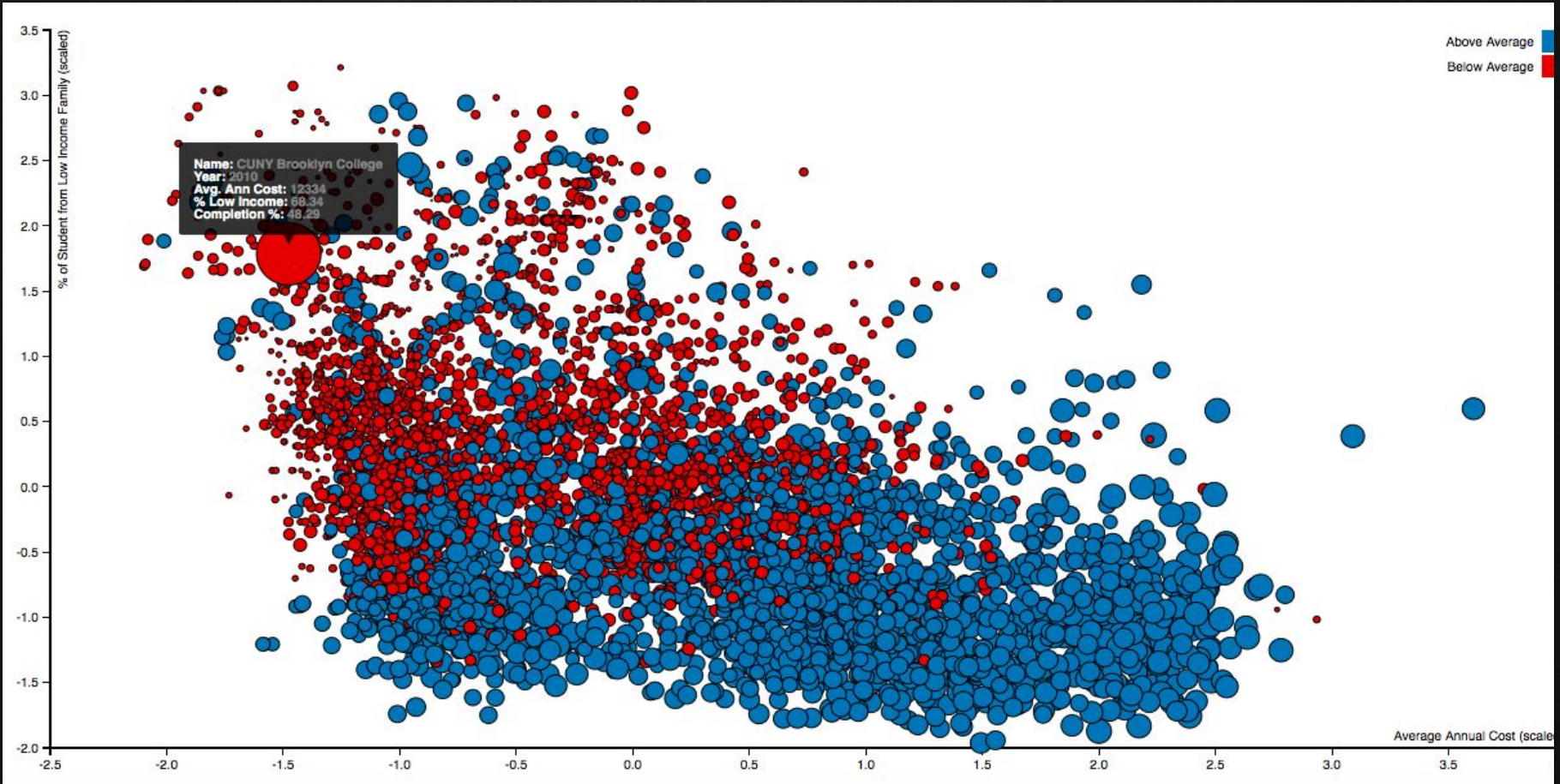
DIAGNOSTIC ANALYTICS



EXTRACTING KEY FEATURES

Random Forest Model indicates that % of students from low income families, % of 1st generation students and average annual cost are most influential features.

DASHBOARD TOOLS



KEY TAKEAWAYS / NEXT STEPS

1.

QUANTITATIVE SUPPORT FOR QUALITATIVE ARGUMENTS

2.

DRILL DOWN TO FIND OUT
WHAT WORKS

3.

DRIVE POLICY AND SUPPORT
PROGRAM DESIGN



THANKS!

Any questions?