

EDUCATION INEQUALITY

Bayan Farag, Anabel Pinto, Maxine Lanier

DO SOCIOECONOMIC FACTORS PLAY A ROLE IN STUDENTS ACT SCORES?

Background

- Given socioeconomic factors can we predict an ACT score?
- Address educational inequality



SOCIAL FACTORS

Added Question

- Does the number of people living in one household play a role in impacting ACT scores?

PART 1 OF THE CLEAN UP

STEPS FOR CLEAN UP

- Started with two datasets
- Subset of the data = school ID, school type, year, state, and zip codes.
- Merged on school ID
- Invalid ACT scores and percentages were dropped from the combined dataframe.

BEFORE

```
J>
      id rate_unemployment percent_college percent_married median_income average_act percent_lunch
0 100001600143          0.117962         0.445283         0.346495         42820.0      20.433455      0.066901
1 100008000024          0.063984         0.662765         0.767619         89320.0      19.498168      0.112412
2 100008000225          0.056460         0.701864         0.713090         84140.0      19.554335      0.096816
3 100017000029          0.044739         0.692062         0.641283         56500.0      17.737485      0.296960
4 100018000040          0.077014         0.640060         0.834402         54015.0      18.245421      0.262641

J> school_info.head()
J>
      year      id state zip_code school_type school_level
0 2016-2017 10000200277  AL    35220  Alternative School      High
1 2016-2017 10000201667  AL    36057  Alternative School      High
2 2016-2017 10000201670  AL    36057  Alternative School      High
3 2016-2017 10000201705  AL    36057  Alternative School      High
```

AFTER

rate_unemployment	percent_college	percent_married	median_income	average_act	percent_lunch	year	state	zip_code	school_type
0.117962	0.445283	0.346495	42820.0	20.433455	0.066901	2016-2017	DE	19804	Regular School
0.063984	0.662765	0.767619	89320.0	19.498168	0.112412	2016-2017	DE	19709	Regular School
0.056460	0.701864	0.713090	84140.0	19.554335	0.096816	2016-2017	DE	19709	Regular School
0.044739	0.692062	0.641283	56500.0	17.737485	0.296960	2016-2017	DE	19958	Regular School
0.077014	0.640060	0.834402	54015.0	18.245421	0.262641	2016-2017	DE	19934	Regular School

PART 2 OF THE CLEAN UP

BEFORE

STEPS FOR CLEAN UP

- merged on zip codes
- Renamed columns so that they were more informative
- Split the data into training and testing data
- Iterative imputer

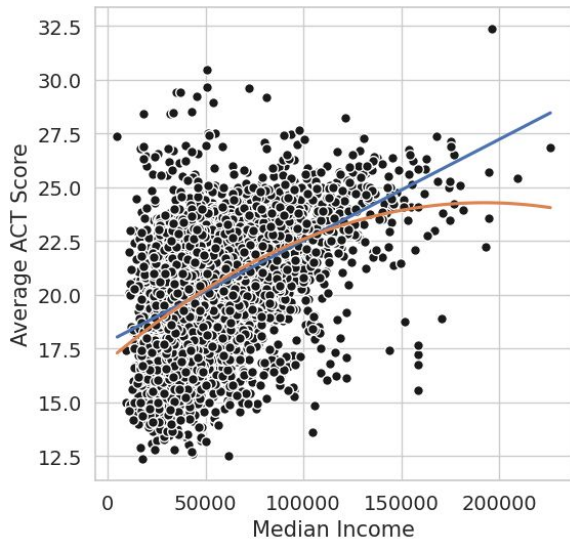
0	8600000US00601	ZCTAS00601	5818	NaN	249	NaN	1416
1	8600000US00602	ZCTAS00602	12719	NaN	374	NaN	2938
2	8600000US00603	ZCTAS00603	19009	NaN	503	NaN	5037
3	8600000US00606	ZCTAS00606	1959	NaN	154	NaN	425
4	8600000US00610	ZCTAS00610	9120	NaN	375	NaN	2208

AFTER

	median_income	one_person	percent_college	percent_lunch	percent_married	rate_unemployment	state	three_people	two_people
0	41793.0	2554.41932	0.602419	0.542056	0.574034	0.111111	NJ	1400.381764	2852.091229
1	38173.0	699.00000	0.469225	0.339655	0.711429	0.135246	IN	417.000000	866.000000
2	39635.0	1021.00000	0.567361	0.270175	0.694514	0.083419	PA	441.000000	991.000000
3	40978.0	276.00000	0.467614	0.315556	0.766901	0.062531	MO	178.000000	467.000000
4	36875.0	1247.00000	0.604470	0.548410	0.803435	0.071429	FL	704.000000	2055.000000

ANALYSIS

ANALYSIS OVERVIEW



Regression Analysis

- Created scatter plots with regression lines for each variable of interest to determine which are relevant

Completed Model

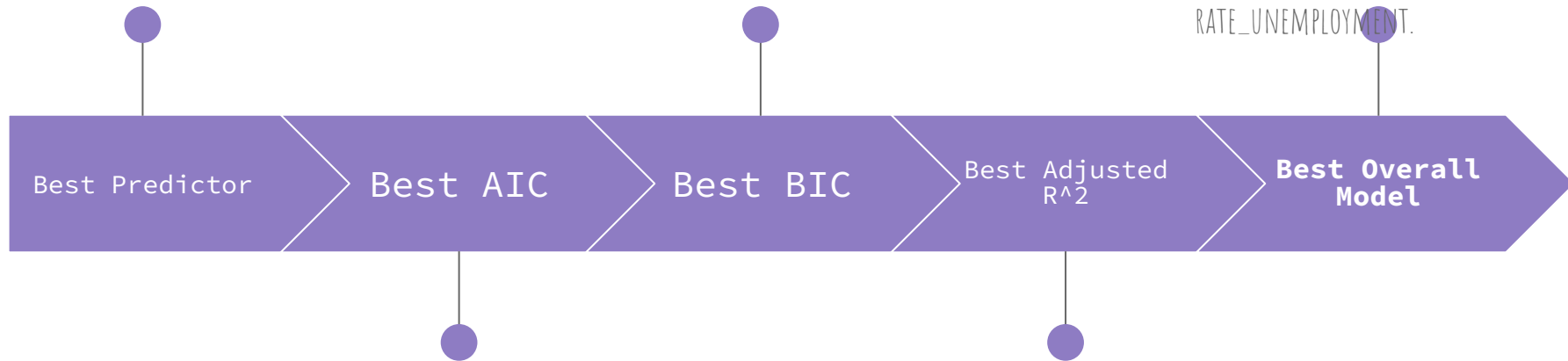
- Used a best subset selector to compare relevant models to find which produces the most accurate prediction of ACT scores

FINDINGS

BOTH WITH AND WITHOUT THE
HOUSEHOLD SIZE DATA, THE BEST
PREDICTOR WAS
`PERCENT_LUNCH`.

6 PREDICTORS: 21077.07

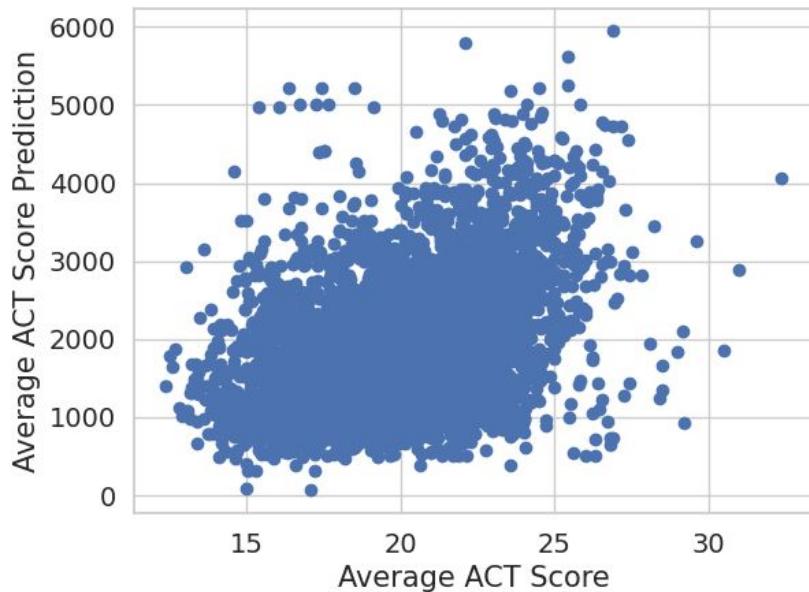
8 PREDICTORS: ONE_PERSON,
TWO_PEOPLE, THREE_PEOPLE,
FOUR_PEOPLE, PERCENT_COLLEGE,
PERCENT_LUNCH, PERCENT_MARRIED,
RATE_UNEMPLOYMENT.



7 PREDICTORS: 21028.24

8 PREDICTORS: 0.65

A DECENT PREDICTION OF ACT SCORES

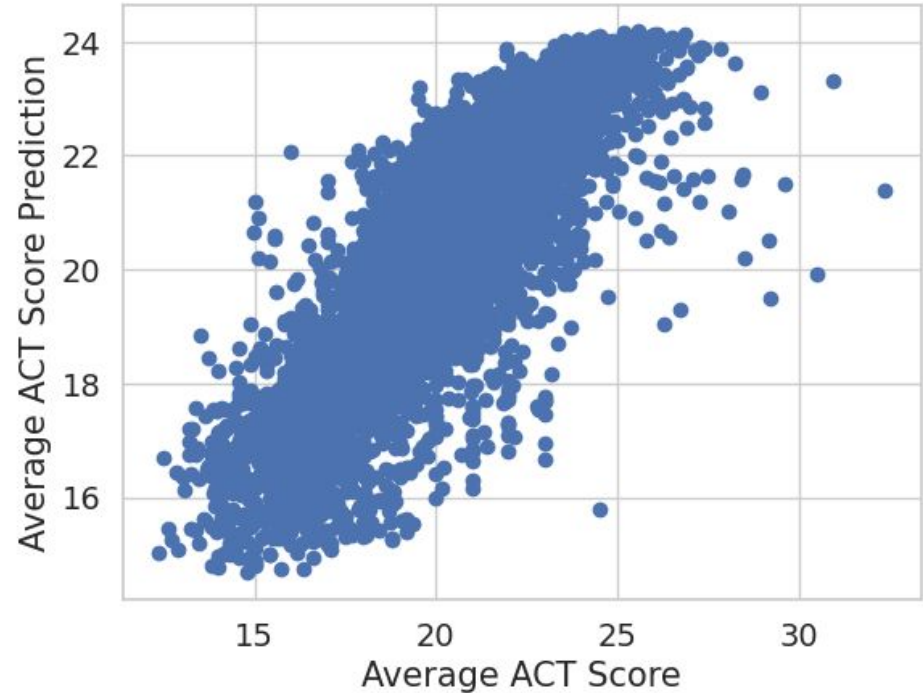


The full model (9 predictors) is a good prediction of ACT scores, but not great – the relationship is **roughly linear**.

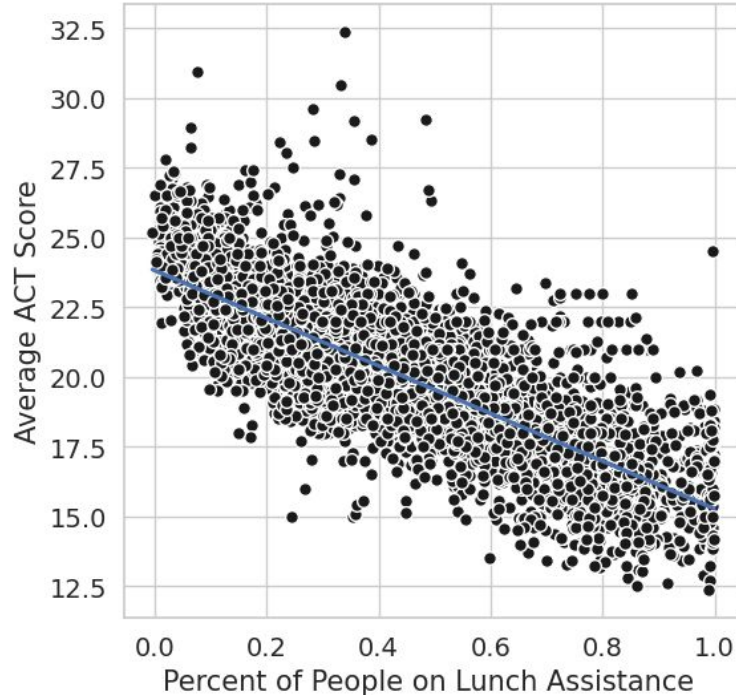
THE BEST MODEL INCLUDES 8 PREDICTORS

one_person, two_people,
three_people, four_people,
percent_college,
percent_lunch,
percent_married,
rate_unemployment.

This model is **more positively linear** than the full model.



PERCENT_LUNCH AS THE BEST PREDICTOR



Very **negatively linear relationship** between percent of people on lunch assistance and the average ACT scores.

CONCLUSION

CONCLUSION

- Median income was not a relevant factor in predicting ACT scores
- The best model included 8 predictors
- Percent_lunch was the best individual predictor of ACT scores

