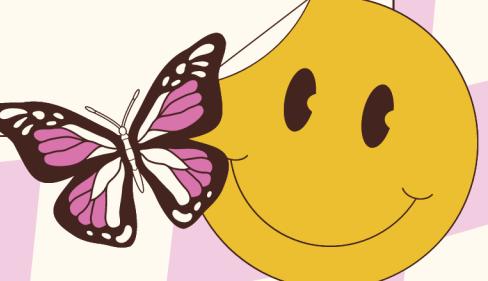
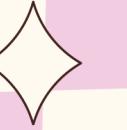
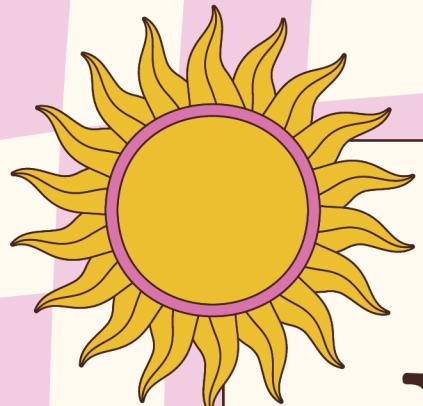


DATA 400 PROJECT

By Margaret
and Ella

PREDICTING COLLEGE DIVERSITY SCORES



Predicting College Diversity Scores

- Our project aims to predict the diversity score of universities based on the composition of students and faculty, college ranking, and endowment.
- The analysis covers **both National Universities and Liberal Arts Colleges**.
- We examine data **from the 2012-2013 to 2021-2022 academic years to determine factors influencing diversity**.
- Main goal: **Predict the student and professor diversity score in 2021-2022 academic year**.





Tractable Data and Data Retrieval



IPEDS: Integrated Postsecondary Education Data System

- System of interrelated surveys conducted annually by the U.S. Department of Education's National Center for Education Statistics
- It gathers data from every college and university participating in federal student financial aid programs, with data collection occurring in the fall, winter, and spring.



U.S. News Ranking

- U.S. News ranks nearly 1,500 U.S. four-year bachelor's degree-granting institutions, organized within 10 distinct categories where colleges and universities are compared with others that share similar academic missions.



Data Overview: National University

- The dataset **df_nu_final** combines **college characteristics and rankings from IPEDS data and US News data**. It contains **126,445 rows and 50 columns**.
- The dataset **df_nu_stu_diversity_training** includes data **from 2012-2013 to 2020-2021**, with **110,928 rows and 50 columns**. It is used to train the student diversity machine learning model.
- The dataset **df_nu_stu_diversity_test** includes data **from 2021-2022**, with **14,028 rows and 50 columns**. It is used to **test the student diversity machine learning model**.
- The dataset **df_nu_prof_diversity_training** includes data **from 2012-2013 to 2020-2021**, with **109,829 rows and 50 columns**. It is used to **train the professor diversity machine learning model**.
- The dataset **df_nu_prof_diversity_test** includes data **from 2021-2022**, with **14,028 rows and 50 columns**. It is used to test the professor diversity machine learning model.

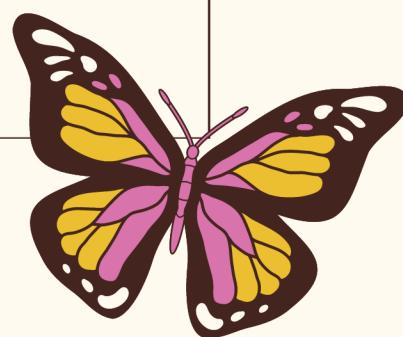
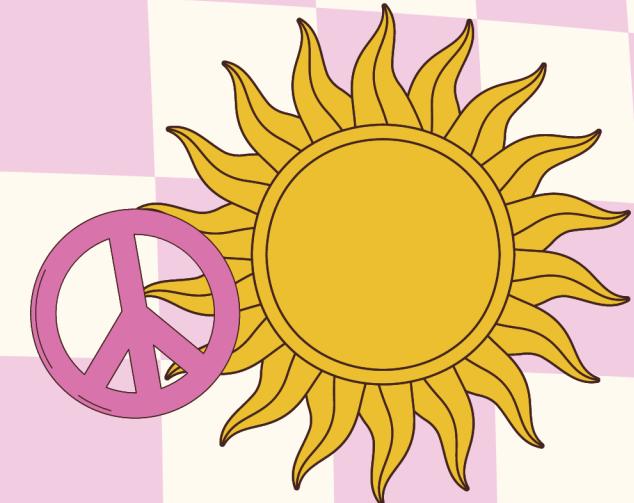
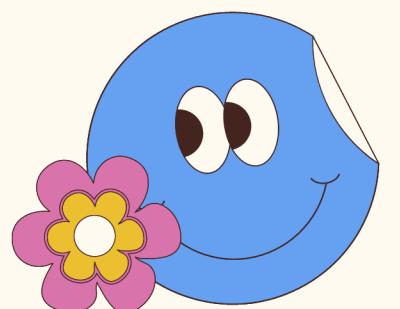


Data Overview: Liberal Art College

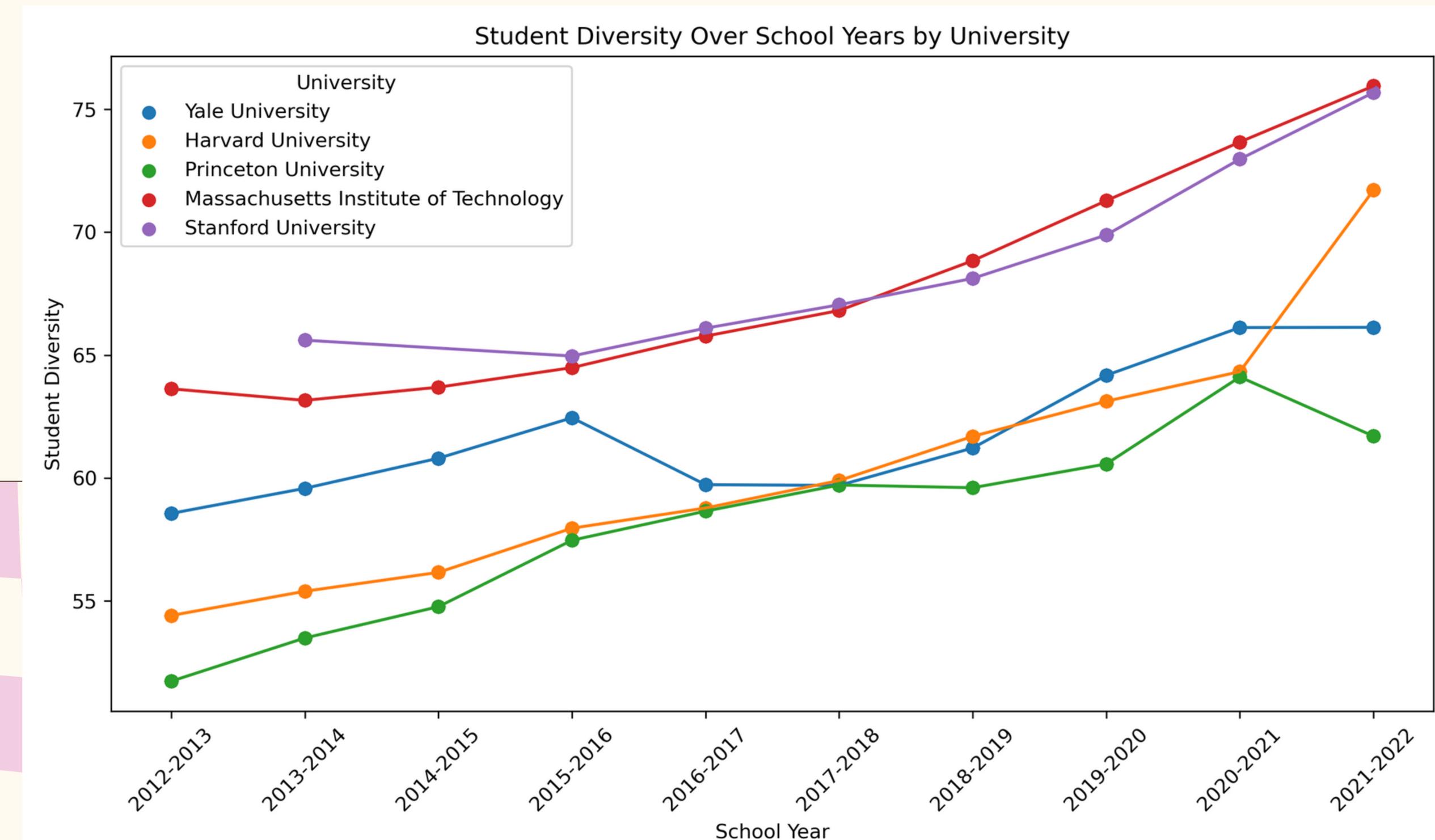
- The dataset **df_lac_final** combines **college characteristics and rankings from IPEDS data and US News data**. It contains **66,684 rows and 50 columns**.
- The dataset **df_lac_stu_diversity_training** includes data **from 2012-2013 to 2020-2021**, with **57,174 rows and 50 columns**. It is used to train the student diversity machine learning model.
- The dataset **df_lac_stu_diversity_test** includes data **from 2021-2022**, with **7,163 rows and 50 columns**. It is used to **test the student diversity machine learning model**.
- The dataset **df_lac_prof_diversity_training** includes data **from 2012-2013 to 2020-2021**, with **54,772 rows and 50 columns**. It is used to **train the professor diversity machine learning model**.
- The dataset **df_lac_prof_diversity_test** includes data **from 2021-2022**, with **7,163 rows and 50 columns**. It is used to **test the professor diversity machine learning model**.



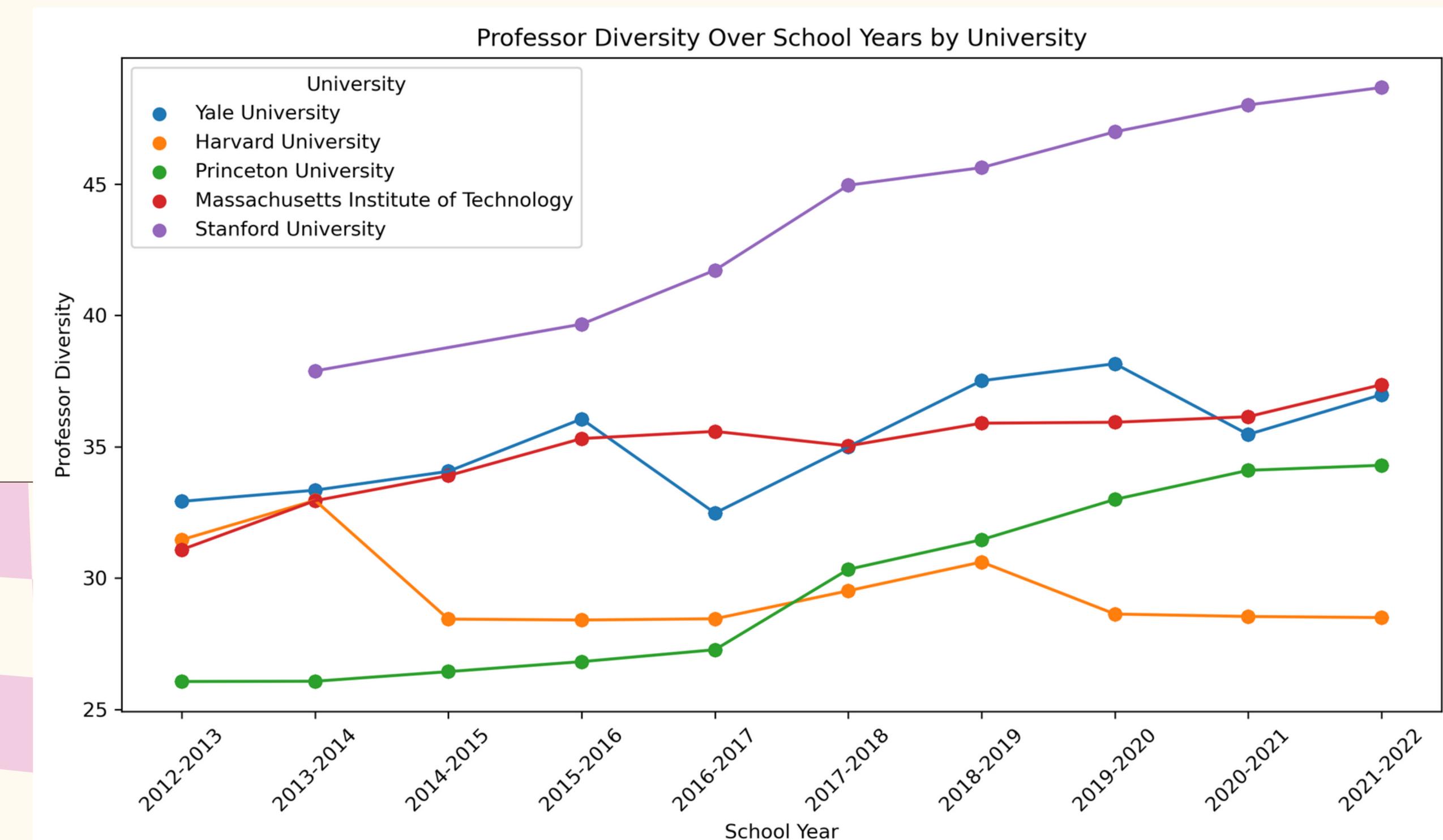
SUMMARY STATISTICS



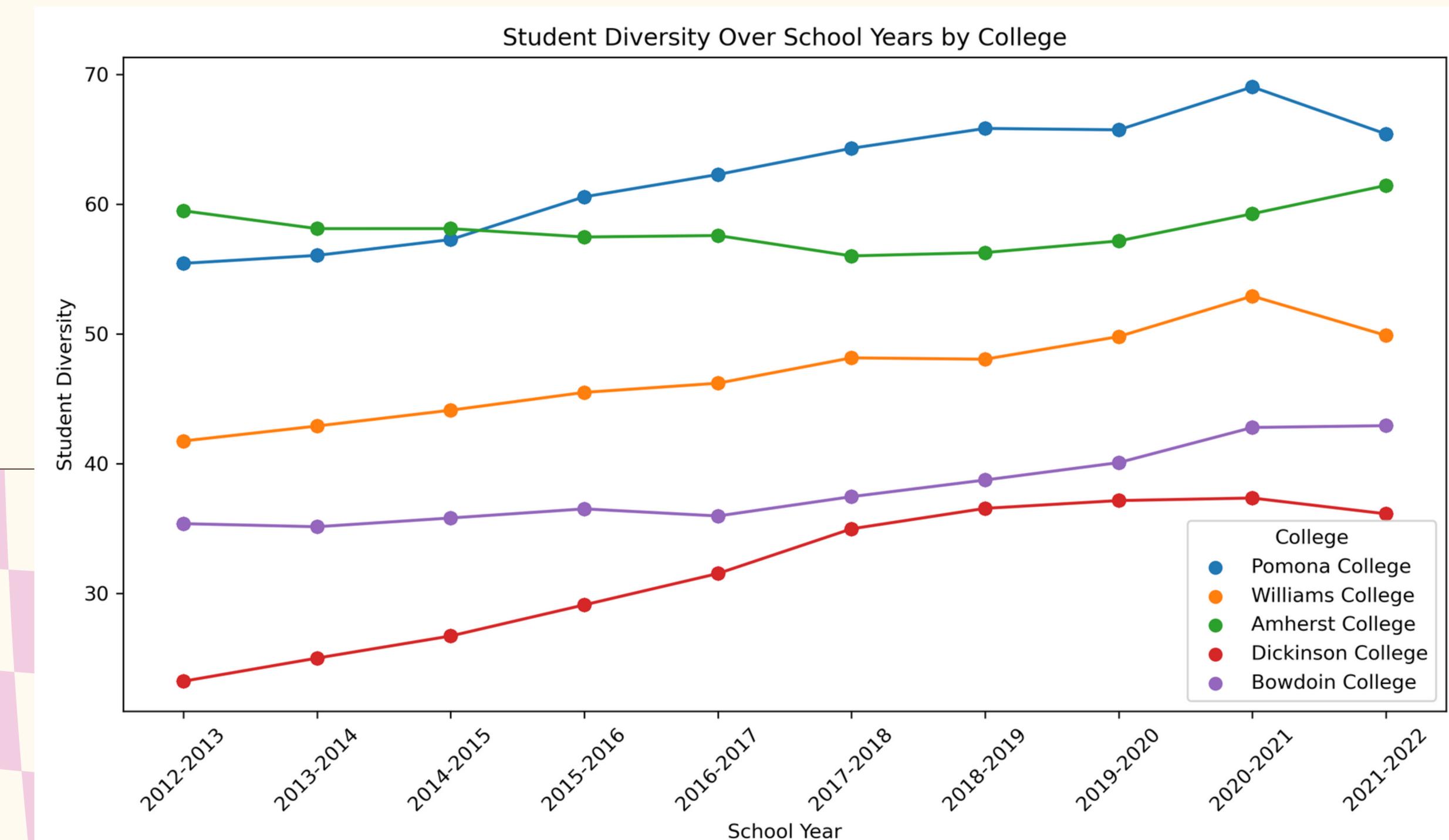
Student Diversity at Top Universities Based on the 2025 Best National University Rankings Over the Years



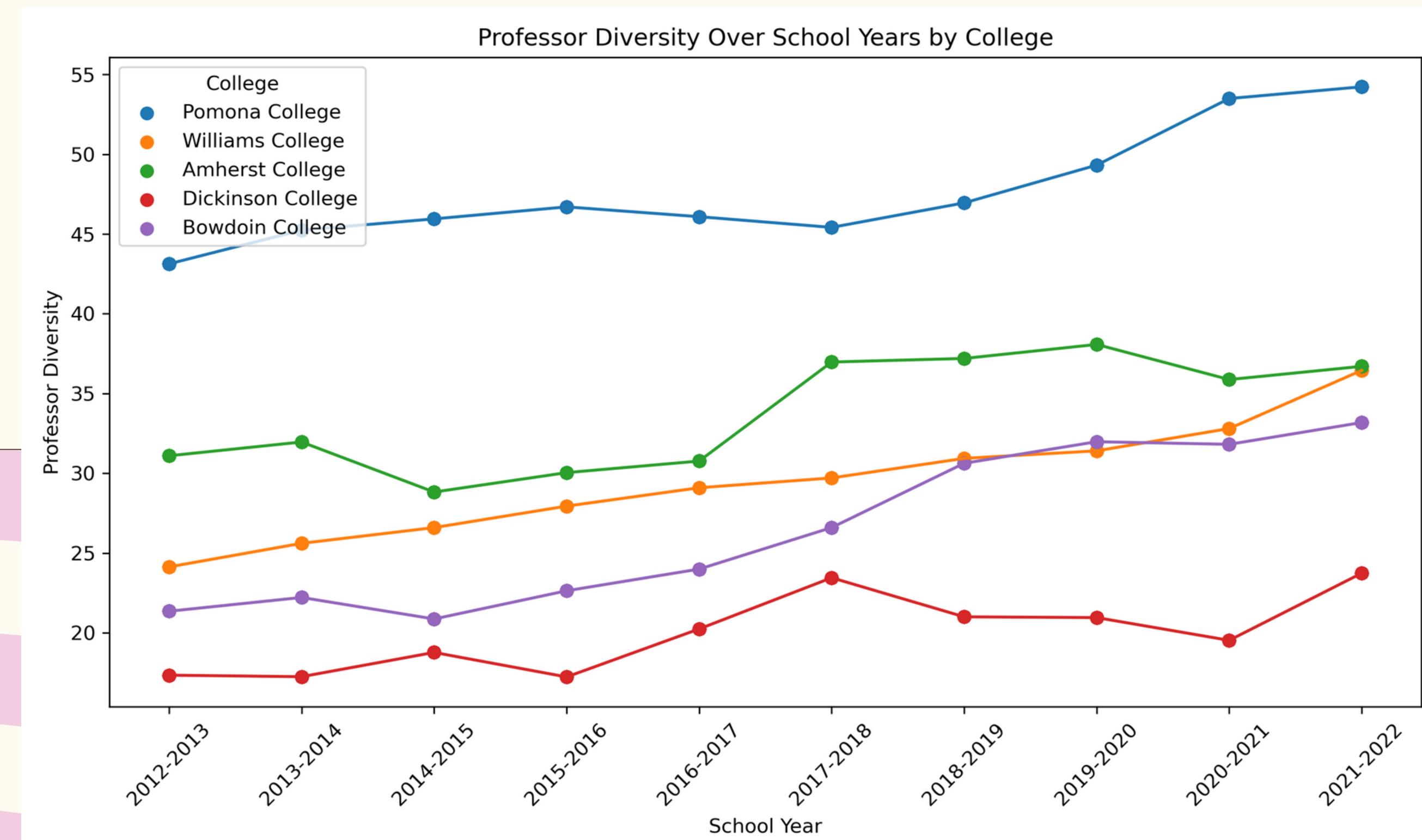
Professor Diversity at Top Universities Based on the 2025 Best National University Rankings Over the Years



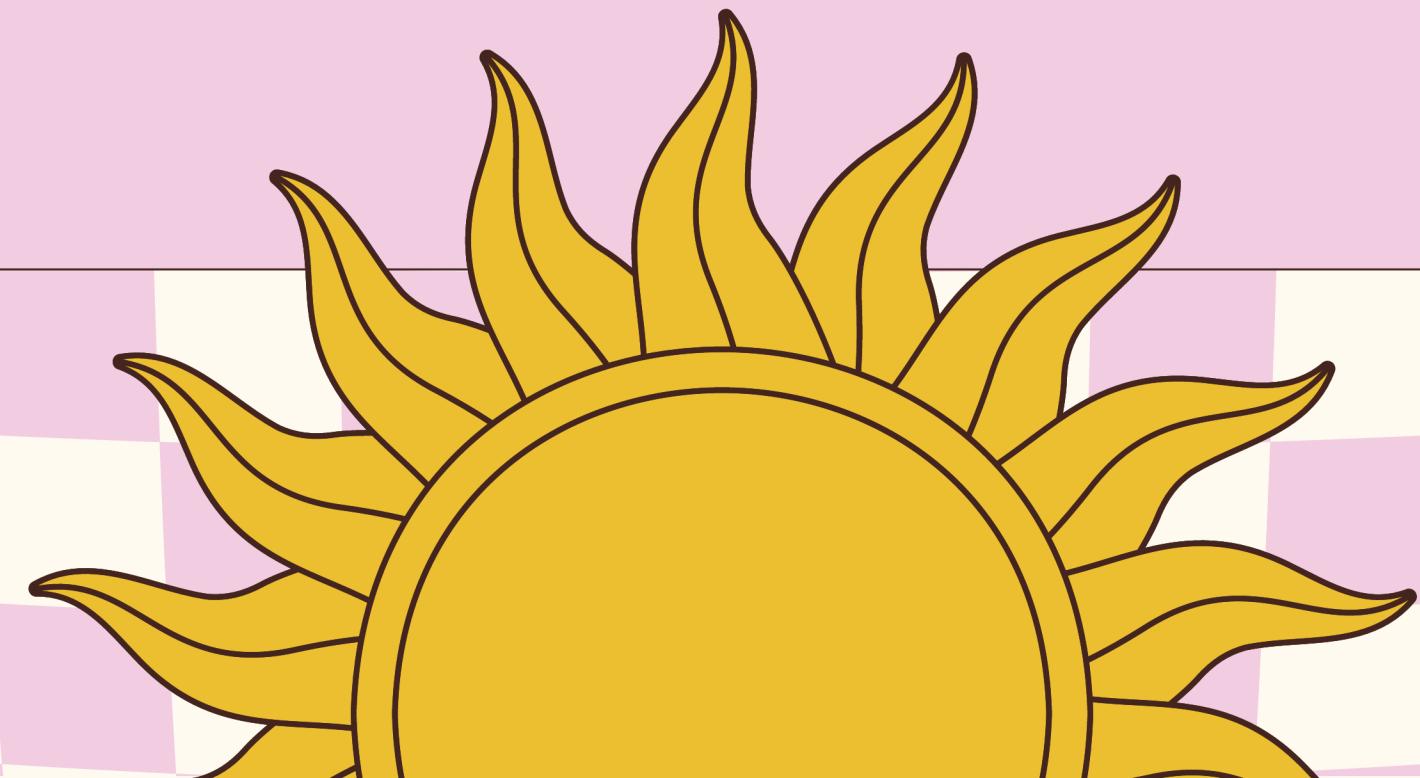
Student Diversity at Top Universities Based on the 2025 Liberal Arts College Rankings Over the Years



Professor Diversity at Top Universities Based on the 2025 Liberal Arts College Rankings Over the Years



METHODOLOGY





Machine Learning Models



- j denotes each university, k denotes each state, and t denotes each school year.
 - The training set covers the years from 2012–2013 to 2020–2021.
 - The test set is the 2021–2022 school year.
- 

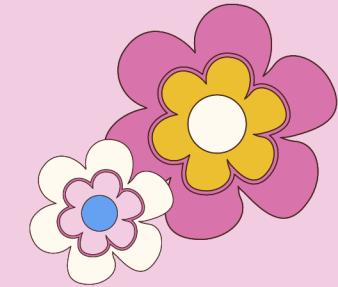
$stu_diversity_{jkt}$

$$\begin{aligned} &= \beta_0 + \beta_1 ranking_{jkt} + \beta_2 hbcu_{jkt} + \beta_3 pct_retenton_{jkt} \\ &+ \beta_4 stu_faculty_ratio_{jkt} + \beta_5 beg_endowment_{jkt} \\ &+ \beta_6 stu_financial_aid_{jkt} + \beta_7 prof_diversity_{jkt} + \beta_8 prop_major_{jkt} \\ &+ \beta_9 pct_female_stu_{jkt} + \beta_{10} pct_male_stu_{jkt} + \beta_{11} graduate_female_{jkt} \\ &+ \beta_{12} graduate_male_{jkt} + \beta_{13} pct_female_prof_salary_{jkt} \\ &+ \beta_{14} pct_male_prof_salary_{jkt} + \alpha_k + \varepsilon \end{aligned}$$

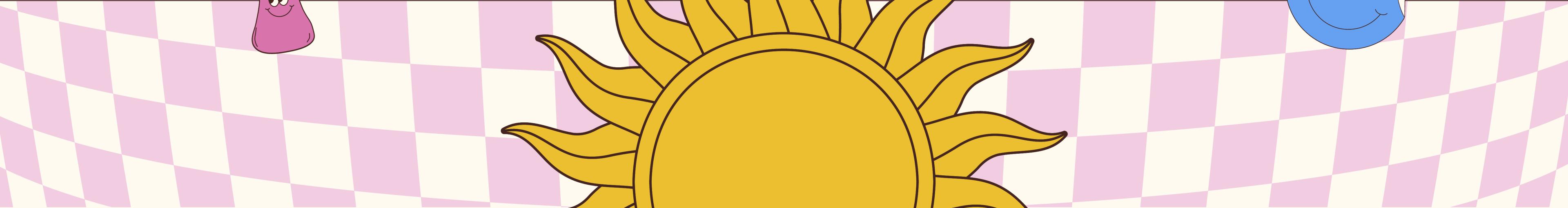
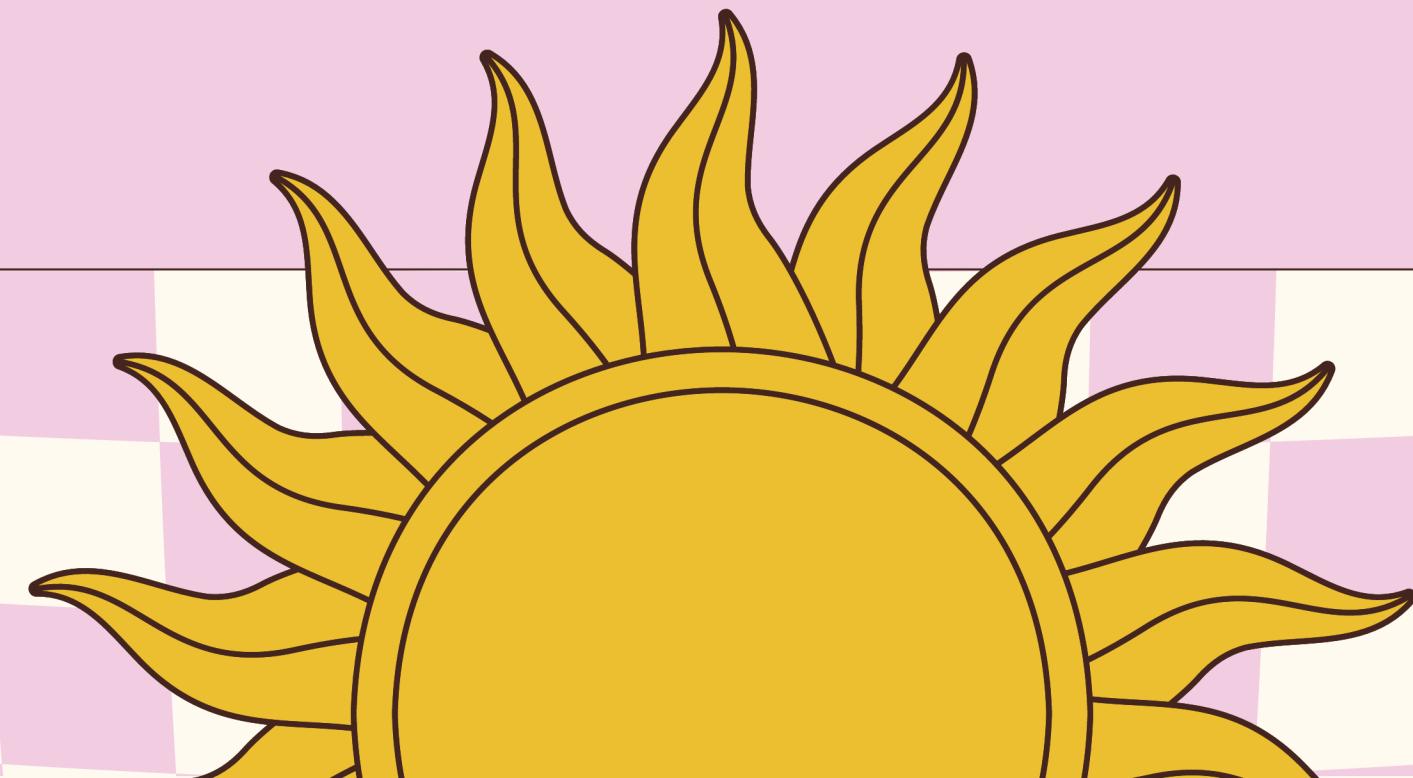
$prof_diversity_{jkt}$

$$\begin{aligned} &= \beta_0 + \beta_1 ranking_{jkt} + \beta_2 hbcu_{jkt} + \beta_3 pct_retenton_{jkt} \\ &+ \beta_4 stu_faculty_ratio_{jkt} + \beta_5 beg_endowment_{jkt} \\ &+ \beta_6 stu_financial_aid_{jkt} + \beta_7 stu_diversity_{jkt} + \beta_8 prop_major_{jkt} \\ &+ \beta_9 pct_female_stu_{jkt} + \beta_{10} pct_male_stu_{jkt} + \beta_{11} graduate_female_{jkt} \\ &+ \beta_{12} graduate_male_{jkt} + \beta_{13} pct_female_prof_salary_{jkt} \\ &+ \beta_{14} pct_male_prof_salary_{jkt} + \alpha_k + \varepsilon \end{aligned}$$

RESULTS



NATIONAL UNIVERSITY



Student Diversity Model

OLS Regression Results									
Dep. Variable:	stu_diversity	R-squared:	0.781						
Model:	OLS	Adj. R-squared:	0.781						
Method:	Least Squares	F-statistic:	7920.						
Date:	Thu, 12 Dec 2024	Prob (F-statistic):	0.00						
Time:	11:16:00	Log-Likelihood:	-3.9287e+05						
No. Observations:	110928	AIC:	7.858e+05						
Df Residuals:	110877	BIC:	7.863e+05						
Df Model:	50								
Covariance Type:	nonrobust								
	coef	std err	t	P> t	[0.025	0.975]			
const	0.0009	0.000	8.320	0.000	0.001	0.001			
AR	-1.2478	0.373	-3.348	0.001	-1.978	-0.517			
AZ	19.8220	0.282	70.177	0.000	19.268	20.376			
CA	31.4038	0.219	143.340	0.000	30.974	31.833			
CO	1.1547	0.294	3.931	0.000	0.579	1.730			
CT	7.7362	0.290	26.673	0.000	7.168	8.305			
DC	6.3193	0.266	23.754	0.000	5.798	6.841			
DE	-0.9522	0.324	-2.937	0.003	-1.588	-0.317			
FL	13.9018	0.245	56.637	0.000	13.421	14.383			
GA	6.7199	0.262	25.647	0.000	6.206	7.233			
IA	-0.7260	0.278	-2.615	0.009	-1.270	-0.182			
IL	18.7834	0.228	82.479	0.000	18.337	19.230			
IN	-1.3364	0.260	-5.135	0.000	-1.846	-0.826			
KS	3.8109	0.288	13.253	0.000	3.247	4.374			
KY	-0.4604	0.339	-1.358	0.174	-1.125	0.204			
LA	-2.8327	0.314	-9.016	0.000	-3.449	-2.217			
MA	12.3140	0.236	52.261	0.000	11.852	12.776			
MD	13.8184	0.279	49.526	0.000	13.272	14.365			
MI	-4.7524	0.262	-18.123	0.000	-5.266	-4.238			
MN	1.2605	0.274	4.597	0.000	0.723	1.798			
MO	0.6479	0.261	2.483	0.013	0.137	1.159			
MS	1.2829	0.413	3.103	0.002	0.472	2.093			
NC	3.0150	0.261	11.551	0.000	2.503	3.527			
NE	0.9080	0.300	3.022	0.003	0.319	1.497			
NH	3.0973	0.299	10.363	0.000	2.511	3.683			
NJ	19.7222	0.262	75.161	0.000	19.208	20.236			
NY	16.8029	0.214	78.397	0.000	16.383	17.223			
OH	-1.2892	0.224	-5.762	0.000	-1.728	-0.851			
OK	13.9338	0.272	51.211	0.000	13.400	14.467			
OR	11.3932	0.289	39.483	0.000	10.828	11.959			
PA	6.9081	0.224	30.791	0.000	6.468	7.348			

- The **R-squared value for the student diversity model is 0.781, indicating that the model's independent variables explain 78.1% of the variation in student diversity.**
- The coefficient for New York is 16.8, indicating that, holding all other factors constant, **being in New York is associated with an increase of 16.8 percentage points in student diversity.**
- The coefficient for Pennsylvania (PA) is 6.9081, suggesting that, all else being equal, **being in Pennsylvania corresponds to an increase of 6.9081 percentage points in student diversity.**

Student Diversity Model

RI	13.9293	0.348	39.977	0.000	13.246	14.612
SC	-3.5072	0.285	-12.287	0.000	-4.067	-2.948
TN	4.8665	0.295	16.472	0.000	4.287	5.446
TX	13.6544	0.225	60.706	0.000	13.214	14.095
UT	13.2045	0.299	44.224	0.000	12.619	13.790
VA	13.2276	0.271	48.893	0.000	12.697	13.758
VT	-4.5214	0.348	-13.000	0.000	-5.203	-3.840
WA	14.2576	0.254	56.054	0.000	13.759	14.756
WI	-0.2819	0.272	-1.038	0.299	-0.814	0.250
ranking	-0.0616	0.001	-41.410	0.000	-0.065	-0.059
hbcu	16.3918	0.467	35.099	0.000	15.476	17.307
pct_retention	0.0863	0.010	8.374	0.000	0.066	0.107
stu_faculty_ratio	-0.2759	0.012	-23.178	0.000	-0.299	-0.253
beg_endowment	1.004e-10	7.25e-12	13.843	0.000	8.62e-11	1.15e-10
stu_financial_aid	2.048e-05	4.23e-06	4.836	0.000	1.22e-05	2.88e-05
prof_diversity	0.7906	0.005	153.048	0.000	0.780	0.801
prop_major	-0.0109	0.005	-2.347	0.019	-0.020	-0.002
pct_female_stu	0.3751	0.007	56.816	0.000	0.362	0.388
pct_male_stu	-0.2552	0.007	-34.908	0.000	-0.270	-0.241
graduate_female	-0.1278	0.006	-21.130	0.000	-0.140	-0.116
graduate_male	0.2477	0.007	35.926	0.000	0.234	0.261
pct_female_prof_salary	-6.2866	0.376	-16.726	0.000	-7.023	-5.550
pct_male_prof_salary	6.2878	0.376	16.730	0.000	5.551	7.024
<hr/>						
Omnibus:	4194.559	Durbin-Watson:		0.007		
Prob(Omnibus):	0.000	Jarque-Bera (JB):		12449.687		
Skew:	-0.090	Prob(JB):		0.00		
Kurtosis:	4.631	Cond. No.		4.84e+19		
<hr/>						

- The coefficient for percentage of female students is 0.3751, meaning that a 1 percentage point increase in the proportion of female students is associated with a 0.3751 percentage point increase in student diversity, holding other factors constant.
- The coefficient for percentage of male students is -0.2552, indicating that a 1 percentage point increase in the proportion of male students is associated with a 0.2552 percentage point decrease in student diversity, controlling for other variables.
- For Harvard: An MSE of 92.03 means that, on average, the squared difference between the predicted values and the actual values is 92.03.

Professor Diversity Model

OLS Regression Results						
Dep. Variable:	prof_diversity	R-squared:	0.737			
Model:	OLS	Adj. R-squared:	0.737			
Method:	Least Squares	F-statistic:	6158.			
Date:	Thu, 12 Dec 2024	Prob (F-statistic):	0.00			
Time:	20:42:32	Log-Likelihood:	-3.1989e+05			
No. Observations:	109829	AIC:	6.399e+05			
Df Residuals:	109778	BIC:	6.404e+05			
Df Model:	50					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	0.0014	5.62e-05	24.978	0.000	0.001	0.002
AR	-0.5096	0.199	-2.564	0.010	-0.899	-0.120
AZ	-2.8917	0.153	-18.857	0.000	-3.192	-2.591
CA	3.5080	0.127	27.584	0.000	3.259	3.757
CO	0.7345	0.159	4.626	0.000	0.423	1.046
CT	4.9133	0.155	31.786	0.000	4.610	5.216
DC	4.5993	0.142	32.297	0.000	4.320	4.878
DE	1.5814	0.173	9.150	0.000	1.243	1.920
FL	2.4756	0.133	18.664	0.000	2.216	2.736
GA	3.8649	0.140	27.684	0.000	3.591	4.138
IA	0.1929	0.148	1.304	0.192	-0.097	0.483
IL	0.2472	0.125	1.977	0.048	0.002	0.492
IN	-0.1260	0.139	-0.910	0.363	-0.398	0.145
KS	-1.7709	0.153	-11.542	0.000	-2.072	-1.470
KY	-1.7004	0.181	-9.409	0.000	-2.055	-1.346
LA	4.8131	0.170	28.271	0.000	4.479	5.147
MA	0.2127	0.128	1.665	0.096	-0.038	0.463
MD	2.8234	0.150	18.813	0.000	2.529	3.118
MI	3.9416	0.140	28.219	0.000	3.668	4.215
MN	-3.1348	0.146	-21.503	0.000	-3.421	-2.849
MO	0.9690	0.140	6.921	0.000	0.695	1.243
MS	-0.7575	0.220	-3.436	0.001	-1.190	-0.325
NC	0.6452	0.141	4.591	0.000	0.370	0.921
NE	-1.4888	0.160	-9.293	0.000	-1.803	-1.175
NH	-1.3555	0.159	-8.502	0.000	-1.668	-1.043
NJ	4.1142	0.144	28.632	0.000	3.833	4.396
NY	-0.9427	0.118	-8.014	0.000	-1.173	-0.712
OH	0.4253	0.119	3.567	0.000	0.192	0.659
OK	-1.6691	0.147	-11.370	0.000	-1.957	-1.381
OR	-0.6335	0.155	-4.088	0.000	-0.937	-0.330

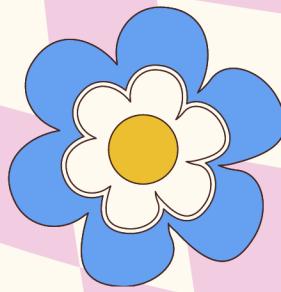
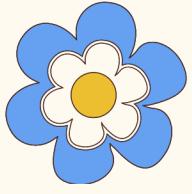
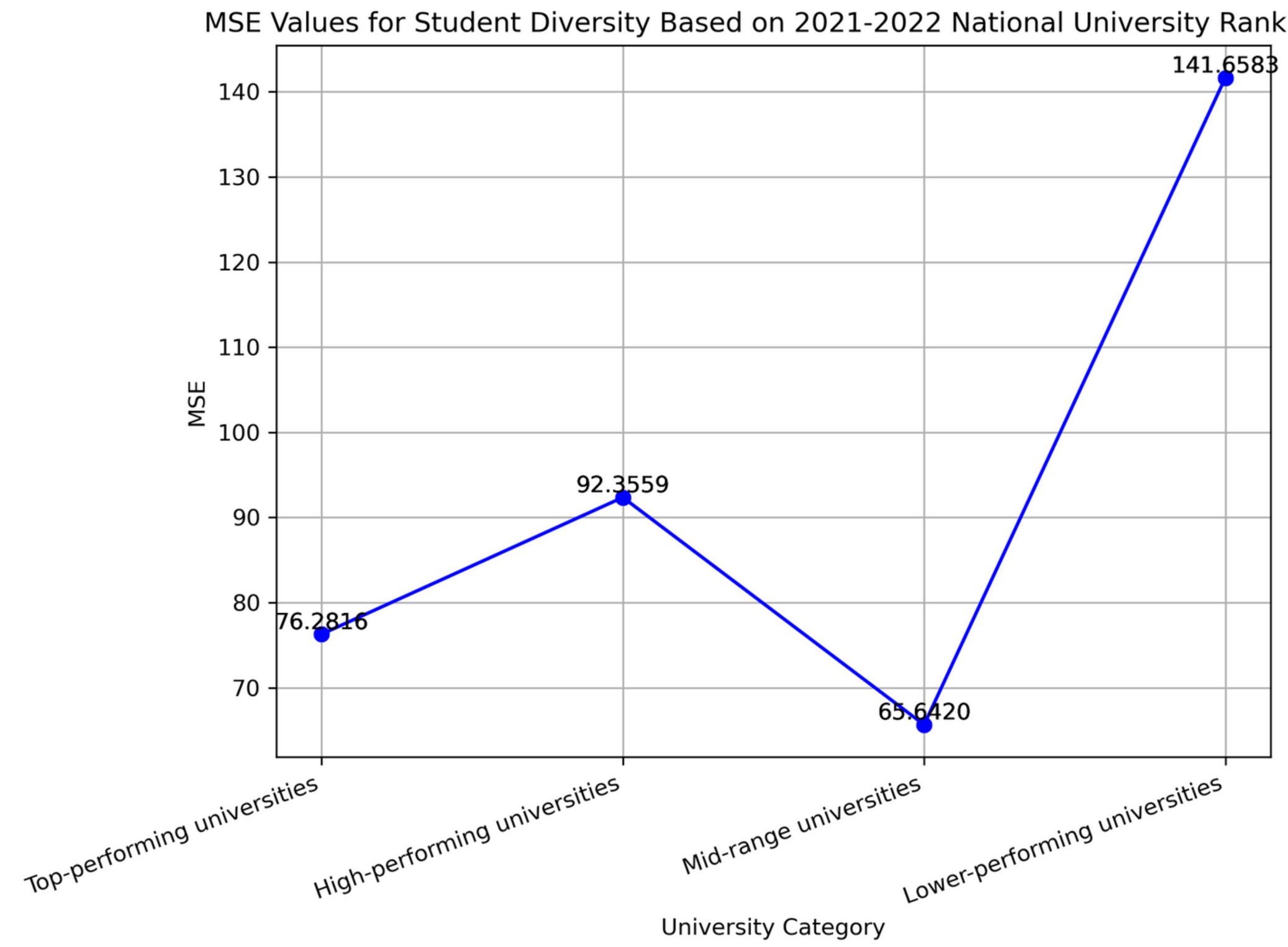
- The **R-squared value for the professor diversity model is 0.737, indicating that 73.7% of the variation in professor diversity is explained by the model's independent variables.**
- The coefficient for New York is **-0.9427**, suggesting that, holding all other factors constant, **being in New York is associated with a 0.9427 percentage point decrease in professor diversity**.
- The coefficient for Pennsylvania (PA) is **1.3161**, indicating that, holding all other factors constant, **being in Pennsylvania corresponds to a 1.3161 percentage point increase in professor diversity**.

Professor Diversity Model

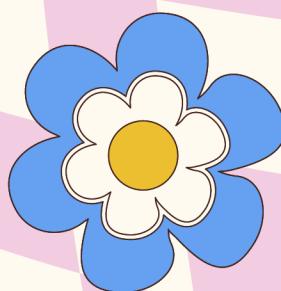
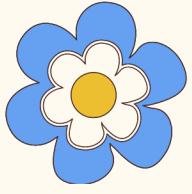
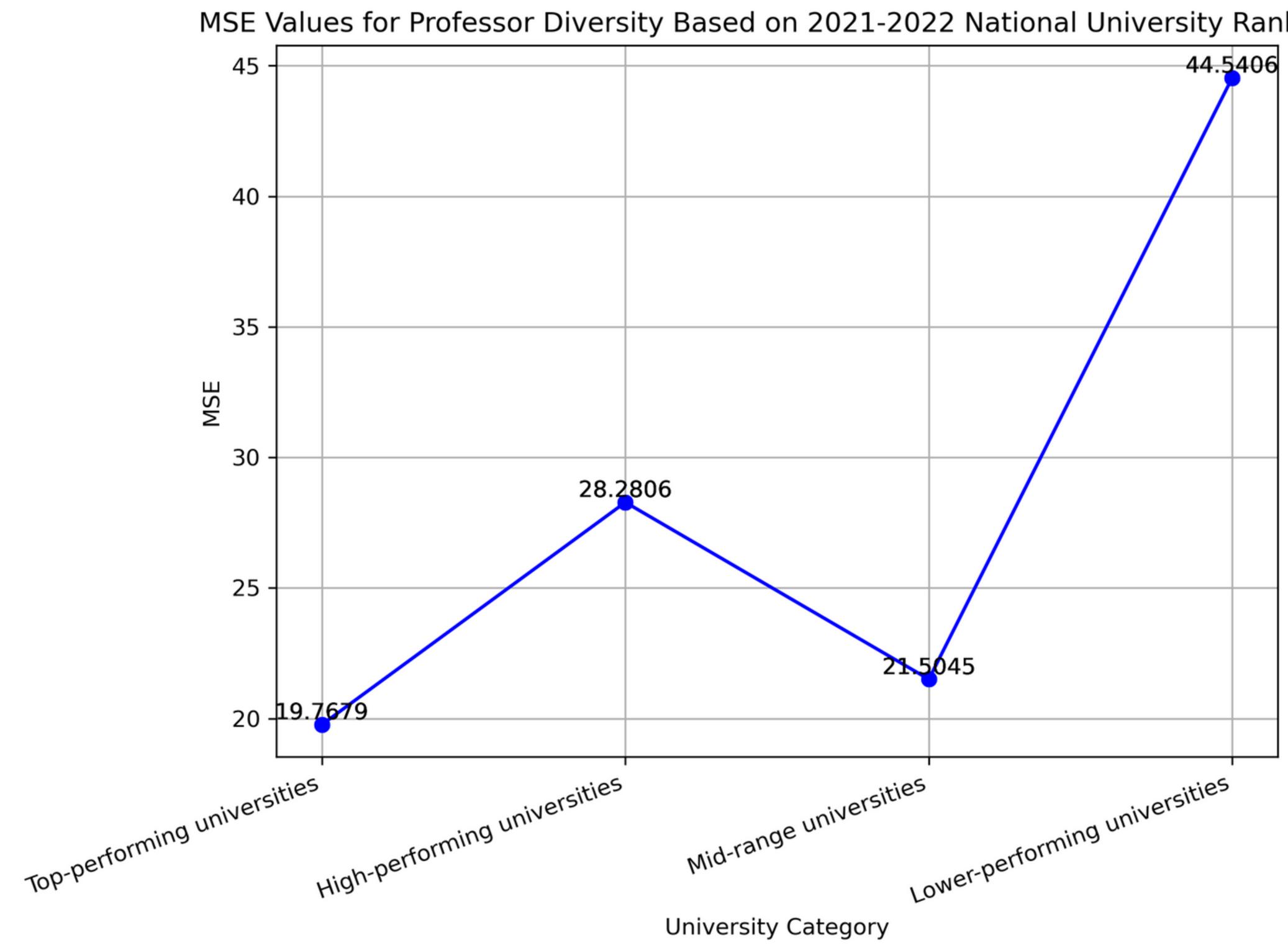
PA	1.3161	0.120	10.963	0.000	1.081	1.551
RI	1.0921	0.187	5.834	0.000	0.725	1.459
SC	0.6560	0.152	4.309	0.000	0.358	0.954
TN	-2.5083	0.157	-15.930	0.000	-2.817	-2.200
TX	-2.3147	0.122	-18.977	0.000	-2.554	-2.076
UT	-4.3803	0.161	-27.271	0.000	-4.695	-4.066
VA	-2.0240	0.146	-13.897	0.000	-2.309	-1.739
VT	-5.2598	0.185	-28.429	0.000	-5.622	-4.897
WA	6.7458	0.136	49.593	0.000	6.479	7.012
WI	-1.0537	0.145	-7.281	0.000	-1.337	-0.770
ranking	0.0576	0.001	73.540	0.000	0.056	0.059
hbcu	42.8559	0.212	202.590	0.000	42.441	43.271
pct_retention	-0.0025	0.006	-0.454	0.650	-0.013	0.008
stu_faculty_ratio	-0.2055	0.006	-32.918	0.000	-0.218	-0.193
beg_endowment	1.116e-10	3.87e-12	28.840	0.000	1.04e-10	1.19e-10
stu_financial_aid	0.0002	2.06e-06	101.042	0.000	0.000	0.000
stu_diversity	0.2321	0.001	157.157	0.000	0.229	0.235
prop_major	0.0018	0.002	0.709	0.478	-0.003	0.007
pct_female_stu	0.0280	0.004	7.811	0.000	0.021	0.035
pct_male_stu	0.1488	0.004	38.049	0.000	0.141	0.156
graduate_female	-0.0497	0.003	-15.329	0.000	-0.056	-0.043
graduate_male	0.2265	0.004	62.440	0.000	0.219	0.234
pct_female_prof_salary	7.9897	0.200	39.981	0.000	7.598	8.381
pct_male_prof_salary	-7.9879	0.200	-39.974	0.000	-8.380	-7.596
<hr/>						
Omnibus:	5404.459	Durbin-Watson:		0.008		
Prob(Omnibus):	0.000	Jarque-Bera (JB):		14401.626		
Skew:	0.261	Prob(JB):		0.00		
Kurtosis:	4.695	Cond. No.		5.10e+19		
<hr/>						

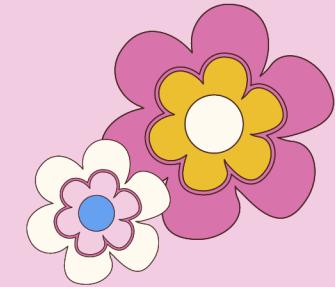
- The coefficient for percentage of female professor salary is 7.9897, meaning a 1 percentage point increase in the proportion of female professor salaries is associated with a 7.9897 percentage point increase in professor diversity, holding other factors constant.
- Conversely, the coefficient for percentage of male professor salary is -7.9879, indicating that a 1 percentage point increase in the proportion of male professor salaries is associated with a 7.9879 percentage point decrease in professor diversity, controlling for other variables.
- For Harvard: An MSE of 93.09 means that, on average, the squared difference between the predicted values and the actual values is 93.09.

MSE Values for Student Diversity Based on 2021-2022 National University Rankings

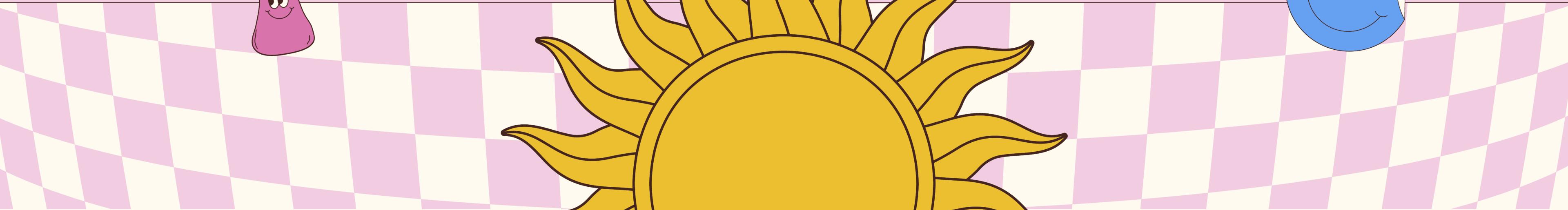
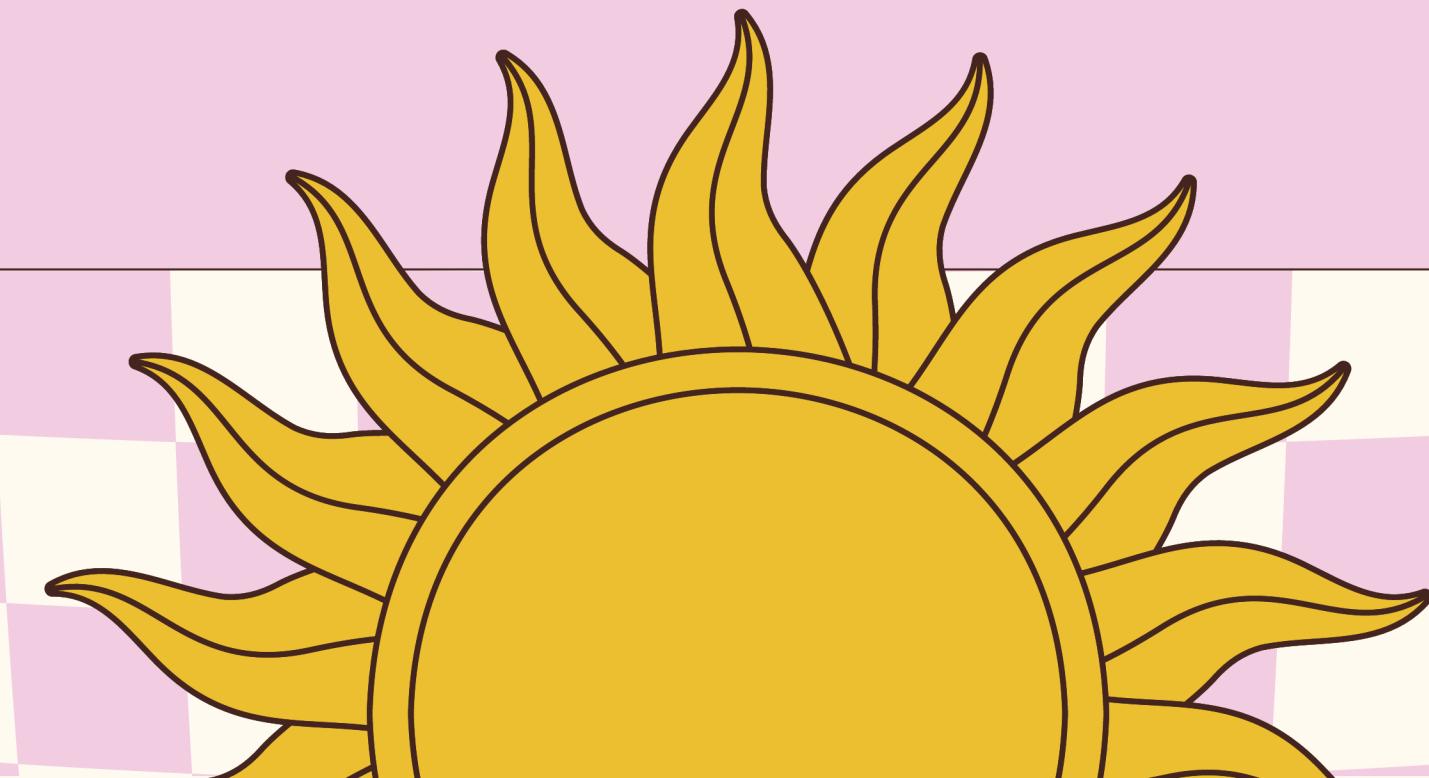
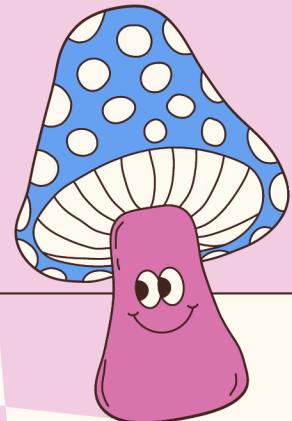


MSE Values for Professor Diversity Based on 2021-2022 National University Rankings





LIBERAL ART COLLEGE



Student Diversity Model

OLS Regression Results						
Dep. Variable:	stu_diversity	R-squared:	0.732			
Model:	OLS	Adj. R-squared:	0.732			
Method:	Least Squares	F-statistic:	3323.			
Date:	Sat, 14 Dec 2024	Prob (F-statistic):	0.00			
Time:	16:55:45	Log-Likelihood:	-1.9830e+05			
No. Observations:	57174	AIC:	3.967e+05			
Df Residuals:	57126	BIC:	3.971e+05			
Df Model:	47					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	2567.9543	192.777	13.321	0.000	2190.110	2945.798
AR	1.4819	0.604	2.453	0.014	0.298	2.666
CA	10.0980	0.525	19.219	0.000	9.068	11.128
CO	-4.0480	0.659	-6.138	0.000	-5.341	-2.755
CT	3.4207	0.560	6.112	0.000	2.324	4.518
FL	-1.9386	0.601	-3.225	0.001	-3.117	-0.760
GA	14.2221	0.558	25.504	0.000	13.129	15.315
IA	-1.7175	0.519	-3.311	0.001	-2.734	-0.701
ID	13.2732	0.675	19.654	0.000	11.950	14.597
IL	5.9383	0.517	11.487	0.000	4.925	6.952
IN	-2.3753	0.528	-4.495	0.000	-3.411	-1.340
KY	-1.6789	0.547	-3.070	0.002	-2.751	-0.607
LA	9.9381	0.714	13.911	0.000	8.538	11.338
MA	1.7798	0.521	3.416	0.001	0.759	2.801
MD	5.6192	0.558	10.067	0.000	4.525	6.713
ME	-2.5766	0.551	-4.676	0.000	-3.657	-1.496
MI	5.4461	0.527	10.336	0.000	4.413	6.479
MN	0.7626	0.519	1.469	0.142	-0.255	1.780
MO	-0.5611	0.576	-0.974	0.330	-1.691	0.568
MS	0.9623	0.658	1.462	0.144	-0.327	2.252
NC	2.8937	0.542	5.337	0.000	1.831	3.956
NE	5.8773	0.669	8.787	0.000	4.566	7.188
NH	-8.6881	0.643	-13.520	0.000	-9.948	-7.429
NJ	14.9514	0.715	20.919	0.000	13.550	16.352
NM	16.3080	2.311	7.057	0.000	11.779	20.837
NY	4.0328	0.517	7.807	0.000	3.020	5.045

- The *R-squared* value for the model is 0.732, indicating that approximately **73.2% of the variation in student diversity is explained by the independent variables in the model.**
- The coefficient for California (CA) is 10.098, meaning that, holding all other factors constant, being in **California is associated with an increase of 10.0980 percentage point in student diversity. This effect is statistically significant with a p-value of 0.00**
- The coefficient for Pennsylvania (PA) is 1.6212, indicating that, **holding all other factors constant, being in Pennsylvania is associated with an increase of 1.6212 percentage point in student diversity. This effect is statistically significant, as the p-value is 0.001 (below the 0.05 threshold).**

Student Diversity Model

- Student-faculty ratio has a coefficient of **-0.3873**, suggesting a **small but significant negative relationship with student diversity** ($p = 0.000$).
- Professorial diversity (*prof_diversity*) has a coefficient of **0.5710**, indicating a **strong positive relationship with student diversity**, holding other variables constant ($p = 0.000$).
- For **Dickinson College**, the Mean Squared Error (MSE) of **0.616** indicates that the model's average squared prediction error for student diversity is moderate, reflecting some variability in accuracy. The Mean Absolute Error (MAE) of **0.783** suggests that, on average, predictions for student diversity at Dickinson College deviate by approximately 0.78 percentage point from the actual values.

NY	4.0328	0.517	7.807	0.000	3.020	5.045
OH	0.7043	0.522	1.349	0.177	-0.319	1.727
OR	7.2043	0.547	13.167	0.000	6.132	8.277
PA	1.6212	0.506	3.205	0.001	0.630	2.612
SC	-1.8943	0.559	-3.390	0.001	-2.990	-0.799
TN	-5.5371	0.571	-9.699	0.000	-6.656	-4.418
TX	12.1434	0.588	20.662	0.000	10.991	13.295
VA	-3.8789	0.517	-7.497	0.000	-4.893	-2.865
VT	2.6045	0.534	4.874	0.000	1.557	3.652
WA	-3.2596	0.571	-5.709	0.000	-4.379	-2.141
WI	-0.2950	0.524	-0.562	0.574	-1.323	0.733
WV	13.1834	1.588	8.304	0.000	10.072	16.295
ranking	-0.0247	0.002	-14.936	0.000	-0.028	-0.021
hbcu	21.6221	0.498	43.434	0.000	20.646	22.598
pct_retention	-0.3133	0.009	-34.933	0.000	-0.331	-0.296
stu_faculty_ratio	-0.3873	0.032	-12.020	0.000	-0.451	-0.324
beg_endowment	9.242e-09	1.08e-10	85.392	0.000	9.03e-09	9.45e-09
stu_financial_aid	-0.0038	6.54e-05	-58.146	0.000	-0.004	-0.004
prof_diversity	0.5710	0.005	126.126	0.000	0.562	0.580
prop_major	0.0081	0.004	1.906	0.057	-0.000	0.016
pct_female_stu	-1.326e+05	9956.640	-13.321	0.000	-1.52e+05	-1.13e+05
pct_male_stu	-1.326e+05	9956.640	-13.321	0.000	-1.52e+05	-1.13e+05
graduate_female	1.326e+05	9953.717	13.321	0.000	1.13e+05	1.52e+05
graduate_male	1.326e+05	9953.717	13.321	0.000	1.13e+05	1.52e+05
pct_female_prof_salary	1333.5392	99.532	13.398	0.000	1138.456	1528.622
pct_male_prof_salary	1318.3002	99.543	13.243	0.000	1123.195	1513.406
<hr/>						
Omnibus:	2299.013	Durbin-Watson:		0.012		
Prob(Omnibus):	0.000	Jarque-Bera (JB):		3068.276		
Skew:	0.420	Prob(JB):		0.00		
Kurtosis:	3.764	Cond. No.		2.29e+17		
<hr/>						

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 4.66e-13. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

Professor Diversity Model

OLS Regression Results						
Dep. Variable:	prof_diversity	R-squared:	0.724			
Model:	OLS	Adj. R-squared:	0.724			
Method:	Least Squares	F-statistic:	3061.			
Date:	Wed, 18 Dec 2024	Prob (F-statistic):	0.00			
Time:	10:46:40	Log-Likelihood:	-1.7963e+05			
No. Observations:	54772	AIC:	3.594e+05			
Df Residuals:	54724	BIC:	3.598e+05			
Df Model:	47					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-1145.1277	160.520	-7.134	0.000	-1459.749	-830.507
AR	-2.7506	0.511	-5.387	0.000	-3.751	-1.750
CA	11.1419	0.435	25.631	0.000	10.290	11.994
CO	12.3783	0.544	22.750	0.000	11.312	13.445
CT	6.4765	0.467	13.864	0.000	5.561	7.392
FL	8.9621	0.496	18.056	0.000	7.989	9.935
GA	-2.2098	0.471	-4.694	0.000	-3.133	-1.287
IA	2.5752	0.430	5.982	0.000	1.731	3.419
ID	-5.0864	0.562	-9.057	0.000	-6.187	-3.986
IL	2.9383	0.429	6.846	0.000	2.097	3.780
IN	5.0852	0.438	11.618	0.000	4.227	5.943
KY	-0.5016	0.456	-1.100	0.271	-1.396	0.392
LA	-0.6269	0.628	-0.999	0.318	-1.857	0.603
MA	5.9758	0.432	13.845	0.000	5.130	6.822
MD	2.6402	0.466	5.667	0.000	1.727	3.553
ME	7.0870	0.456	15.538	0.000	6.193	7.981
MI	9.3880	0.436	21.554	0.000	8.534	10.242
MN	3.4330	0.431	7.970	0.000	2.589	4.277
MO	3.2146	0.489	6.568	0.000	2.255	4.174
MS	3.4352	0.545	6.307	0.000	2.368	4.503
NC	7.0979	0.449	15.793	0.000	6.217	7.979
NE	-4.0156	0.582	-6.902	0.000	-5.156	-2.875
NH	5.8625	0.532	11.011	0.000	4.819	6.906
NJ	10.9795	0.616	17.812	0.000	9.771	12.188
NM	-13.2202	1.914	-6.906	0.000	-16.972	-9.468
NY	3.9842	0.429	9.288	0.000	3.143	4.825
OH	5.6840	0.432	13.154	0.000	4.837	6.531
OR	4.1683	0.457	9.114	0.000	3.272	5.065
PA	4.2720	0.419	10.200	0.000	3.451	5.093
SC	~ 0.000	~ 0.000	~ 0.000	~ 0.000	~ 0.000	~ 0.000

- The R-squared value for the professor diversity model is 0.724, indicating that the independent variables in the **model explain 72.4% of the variation in professor diversity**.
- California (CA) has a positive coefficient (11.1419), meaning universities in California tend to have higher professor diversity scores compared to the reference state.
- The coefficient for Pennsylvania (PA) is 4.272, suggesting that, all else being equal, **being in Pennsylvania corresponds to an increase of 4.27 percentage points in professor diversity**.

Professor Diversity Model

TN	6.6981	0.479	13.980	0.000	5.759	7.637
TX	3.5277	0.492	7.171	0.000	2.564	4.492
VA	1.5253	0.430	3.550	0.000	0.683	2.367
VT	5.5311	0.443	12.472	0.000	4.662	6.400
WA	10.4755	0.475	22.054	0.000	9.544	11.406
WI	4.4015	0.435	10.119	0.000	3.549	5.254
WV	52.9502	1.314	40.310	0.000	50.376	55.525
ranking	-0.0417	0.001	-29.775	0.000	-0.044	-0.039
hbcu	45.7802	0.413	110.856	0.000	44.971	46.590
pct_retention	0.1707	0.008	22.343	0.000	0.156	0.186
stu_faculty_ratio	-0.2373	0.027	-8.650	0.000	-0.291	-0.184
beg_endowment	-1.538e-09	9.64e-11	-15.959	0.000	-1.73e-09	-1.35e-09
stu_financial_aid	0.0007	5.69e-05	12.923	0.000	0.001	0.001
stu_diversity	0.4053	0.003	129.015	0.000	0.399	0.411
prop_major	-0.0085	0.004	-2.347	0.019	-0.016	-0.001
pct_female_stu	5.912e+04	8287.888	7.134	0.000	4.29e+04	7.54e+04
pct_male_stu	5.912e+04	8287.888	7.134	0.000	4.29e+04	7.54e+04
graduate_female	-5.911e+04	8285.454	-7.134	0.000	-7.53e+04	-4.29e+04
graduate_male	-5.911e+04	8285.455	-7.134	0.000	-7.53e+04	-4.29e+04
pct_female_prof_salary	-593.2367	82.851	-7.160	0.000	-755.626	-430.847
pct_male_prof_salary	-588.9073	82.858	-7.107	0.000	-751.310	-426.504
<hr/>						

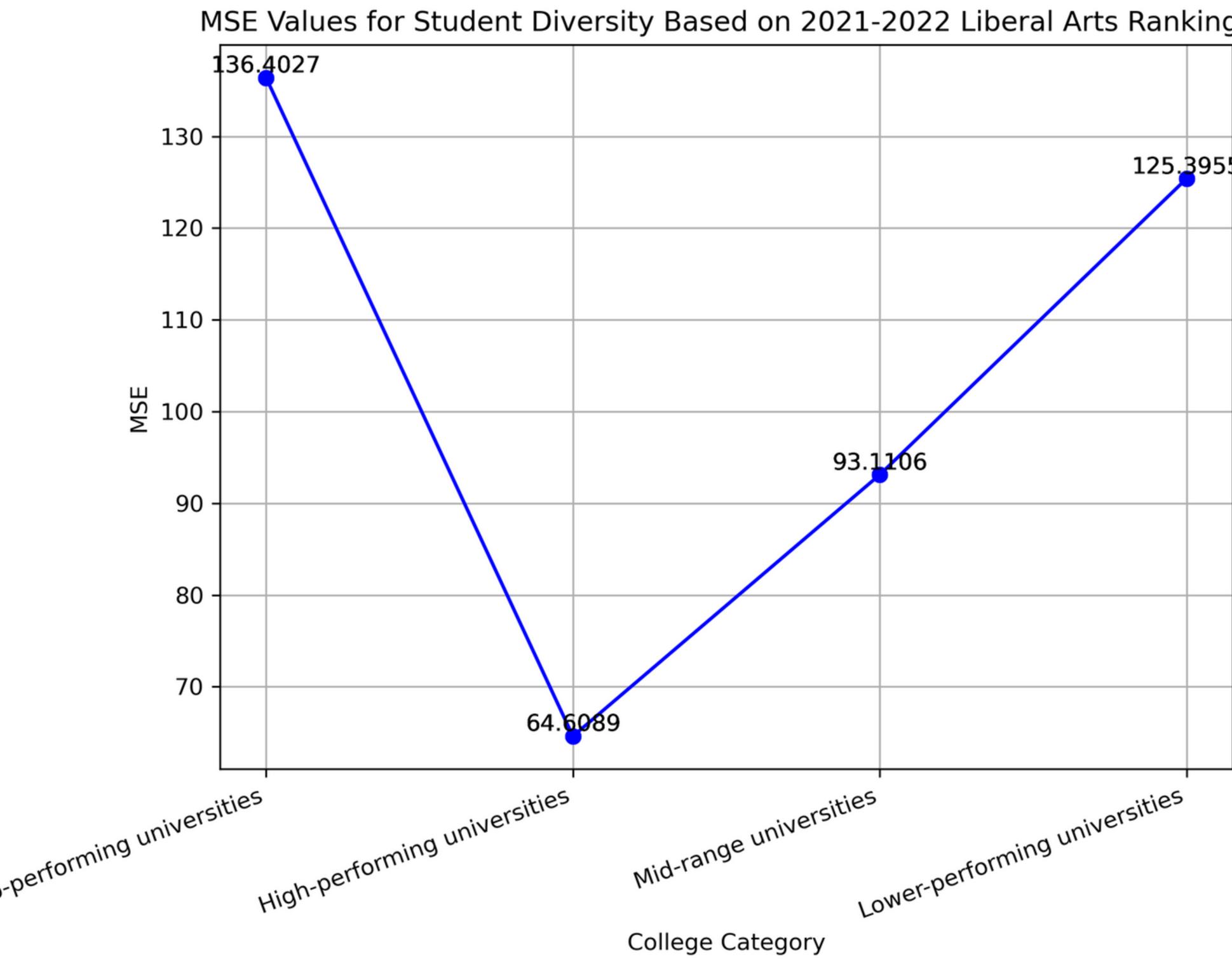
Omnibus:	20687.316	Durbin-Watson:	0.015
Prob(Omnibus):	0.000	Jarque-Bera (JB):	246889.220
Skew:	1.479	Prob(JB):	0.00
Kurtosis:	12.972	Cond. No.	2.29e+17

Notes:

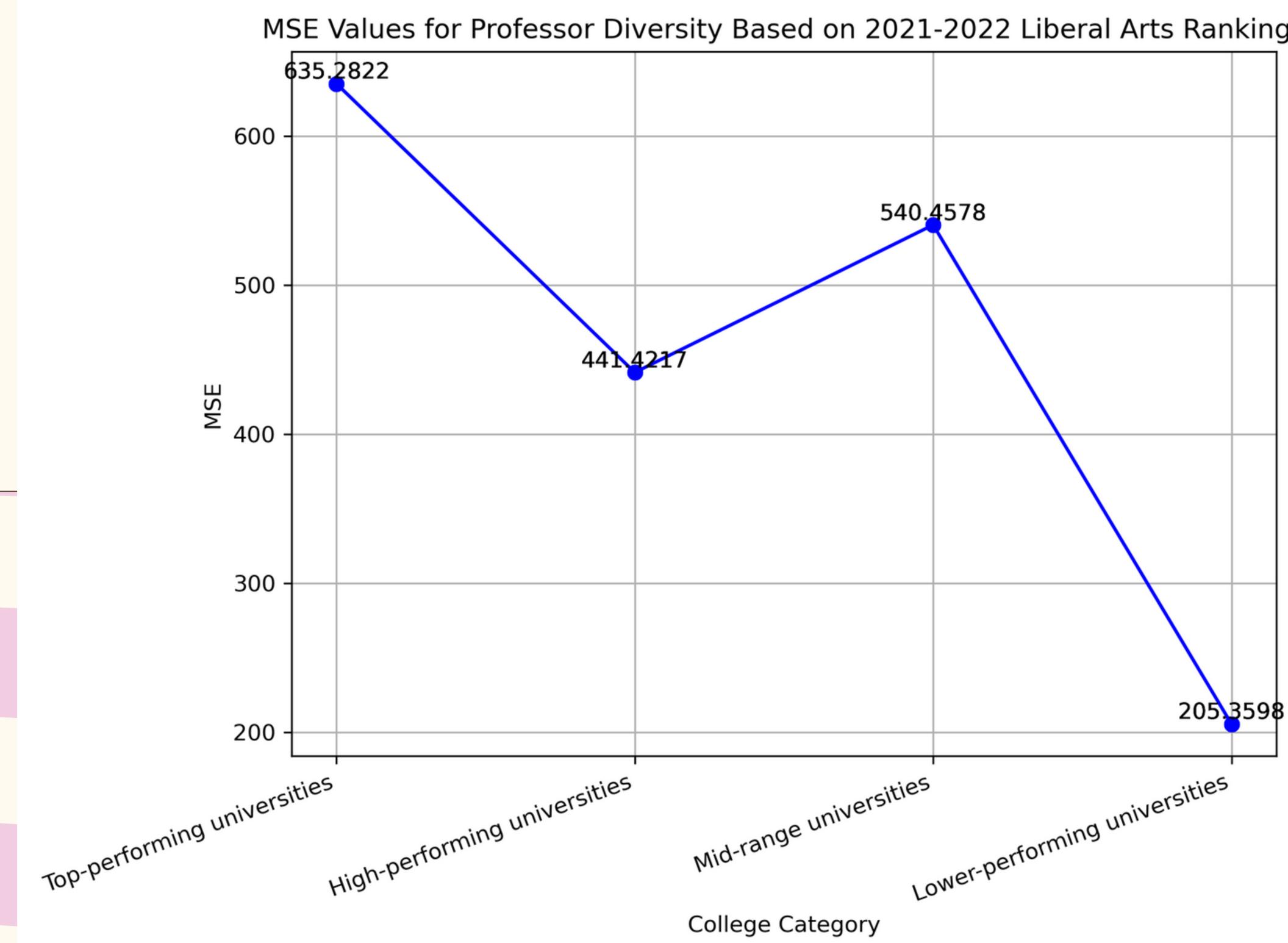
- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The smallest eigenvalue is 4.56e-13. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

- **stu_diversity** (0.4053): Strong positive association with professor diversity, as expected.
- **hbcu** (45.7802): Being a Historically Black College or University is strongly associated with much higher professor diversity scores.
- **stu_financial_aid** (0.0007): Financial aid correlates positively with professor diversity.
- For **Dickinson College**, the Mean Squared Error (MSE) of **0.114** indicates that the model's average squared prediction error is low, highlighting good accuracy.
- The Mean Absolute Error (MAE) of **0.332** suggests that, on average, predictions for professor diversity deviate by about 0.33 percentage point from the actual values

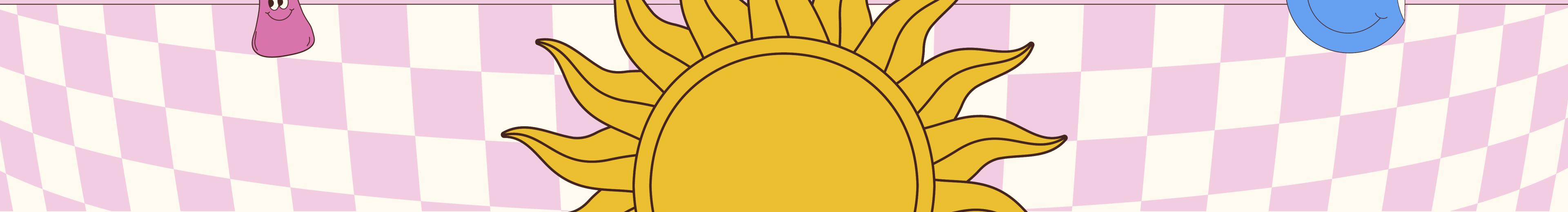
MSE Values for Student Diversity Based on 2021-2022 Liberal Art College Rankings



MSE Values for Professor Diversity Based on 2021-2022 Liberal Art College Rankings



IMPLICATIONS FOR STAKEHOLDERS



Implications for Stakeholders

- **Universities:**
 - **Strategic Planning:** Insight into diversity metrics helps institutions align their strategies to foster inclusive environments and meet diversity targets.
 - **Resource Allocation:** Identifying trends can guide the allocation of scholarships, support programs, and outreach efforts to underrepresented groups.
 - **Reputation Management:** Diversity scores may influence rankings and perceptions of inclusivity, impacting an institution's reputation.



Implications for Stakeholders

- **Students:**
 - **Informed Choices:** Prospective students can use diversity scores to choose institutions aligned with their values and preferences.
 - **Enhanced Environment:** Greater diversity fosters inclusive campuses that improve learning and social experiences.
- **Policy Makers:**
 - **Policy Development:** Helps evaluate the effectiveness of state or federal diversity initiatives and policies.
- **Employers:**
 - **Talent Pipeline:** Diverse campuses indicate universities producing graduates with varied perspectives and experiences, which employers value.





ETHICAL, LEGAL, AND SOCIETAL CONSIDERATIONS

Ethical, Legal, and Societal Considerations

- Ethical Considerations

- **Algorithmic Bias:** The model must ensure fairness, avoiding biases in predictions due to historical data reflecting systemic inequalities.
- **Equity vs. Equality:** Predictions should aim to promote equity rather than reinforce inequalities.

- Societal Considerations:

- Promotes broader access to education for underrepresented groups, addressing systemic barriers.
- Highlights disparities that require targeted interventions at institutional and societal levels.



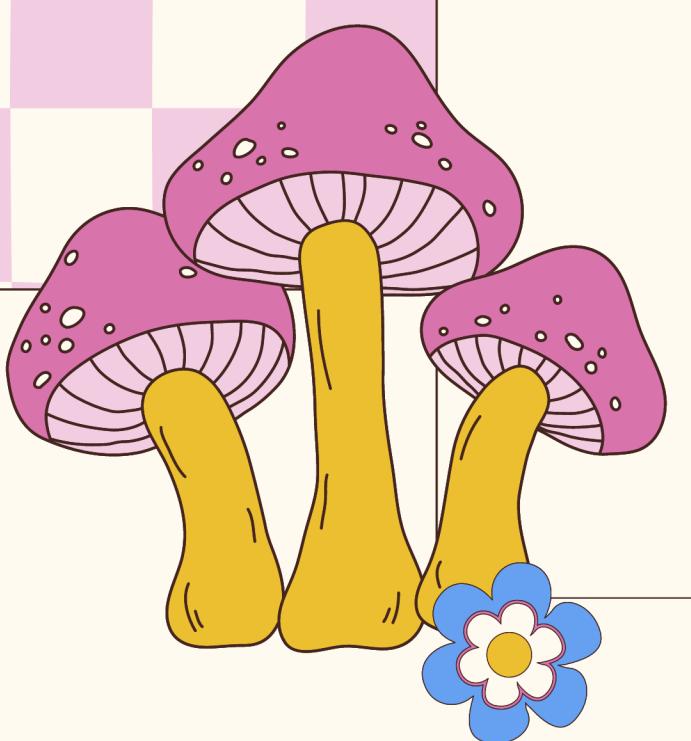
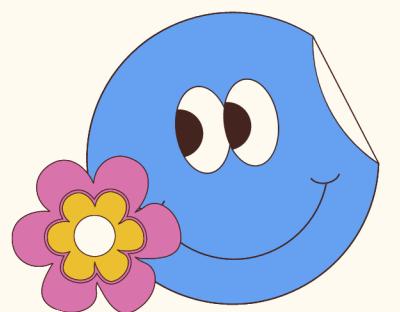
Ethical, Legal, and Societal Considerations

- **Legal Considerations:**
 - Comply with laws such as Title VI of the Civil Rights Act, which prohibits discrimination based on race, color, or national origin in programs receiving federal funding.
 - Be cautious of how diversity metrics are used in admissions decisions, especially in jurisdictions where affirmative action policies are contested.





THANK YOU!



Any questions?

