# **Tuition Assistance Program - Funding Analysis Across NYC Institutions**

Rahul Ganesan raga7935@colorado.edu University of Colorado Boulder Boulder, Colorado, USA Atharv Kadam atka9416@colorado.edu University of Colorado Boulder Boulder, Colorado, USA Dnyaneshwari Rakshe dnra6230@colorado.edu University of Colorado Boulder Boulder, Colorado, USA

#### **Abstract**

In this project, we have examined the patterns and trends within the New York State Tuition Assistance Program (TAP) dataset. This analysis will help us understand the relationships between different features like financial aid, demographic factors, and educational institutions. We conducted a thorough analysis across several attributes like academic age, income ranges, recipient age groups, and sector types (public vs. private, community vs. non-community colleges). We specifically focused on areas like analyzing trends over academic years, evaluating the impact of income range on financial aid distribution between CUNY and SUNY institutions, and exploring the relationship between TAP dollars awarded and full-time equivalent (FTE) enrollments. We implemented both basic and advanced statistical methodologies to provide actionable insights by comprehensive analysis into TAP funding patterns and suggest the patterns or trends to the institutions and policymakers.

#### **Keywords**

Statistical Methods, Two-sample test, Hypothesis Testing, ANOVA, Linear Regression, Chi-square test, Tuition Assistance Program, Funding Analysis, NYC, CUNY, SUNY, TAP, TAP Receipient Dollars, Age Group Disparities, Income Range Impact, Sector Type Analysis, Educational Financial Aid, TAP Financial Status, Statistical Significance, Higher Education Funding, Demographic Analysis, Welch's ANOVA, Dependent and Independent Variables, Statistical Modeling

#### 1 Introduction

The New York State Tuition Assistance Program (TAP) is one of the largest need-based grant programs in the United States. It provides a critical assistance to the students by providing financial aid for pursuing higher education across various academic institutions. It was established with a purpose to reduce financial barriers for the students. This program offers grants based on different eligibility factors like family income, dependency status, and tuition costs. TAP grants awards of up to \$ 5,665 per year.

This program is different from granting loans because these grants do not require repayment. Hence, this makes TAP a crucial resource for low- and middle-income families. TAP funding supports both public institutions like SUNY and CUNY and private colleges across the state so that students have equitable access to education.

Over the last 23 years, TAP policies have evolved significantly by taking the changing economic and educational landscapes into consideration. Based on the information from TAP official website, it has been observed that this program has expanded eligibility to part-time students, adjusted income thresholds, and increased

funding allocations over the time. These changes have helped in enhancing the program's reach, and allowing it to support a broader range of students.

Year	Milestone
2000	TAP adjusted income eligibility thresholds to include more middle- income families.
2006	Expanded eligibility to part-time students under specific conditions.
2010	Integrated it with the FAFSA to streamline the TAP application process.
2015	Increased funding allocations to meet rising tuition costs at SUNY, CUNY, and other private colleges.
2020	Introduced changes to improve equity in grant distribution across institutional sectors.
2023	Enhanced TAP grants for non-degree and workforce training programs to address skill gaps.

Figure 1: Key Changes in the TAP over the years

# 2 Dataset Description:

The dataset that we considered while working on this project included key information such as income ranges, age groups, financial dependency status, and institutional types such as CUNY, SUNY, and independent colleges. This dataset consisted of records for over 23 academic years. We majorly focused on identifying meaningful insights from the dataset so that we can help decision makers to understand into how financial aid is distributed and utilized, based on different factors like demographic and institutional disparities.

The dataset consists a total of 259,983 rows and 16 columns. It includes both numerical and categorical variables. It consists of following key variables:

Attribute Name	Туре	Description					
Academic Year	Categorical	The academic year during which the					
		funding was received					
TAP Financial Status	Categorical	Indicates whether the recipient is					
		financially independent or dependent.					
Income Range	Categorical	The income range of the recipient					
		(e.g., Low, Medium, High)					
TAP Level of Study	Categorical	Level of study of the recipient (e.g.,					
		Undergraduate, Graduate)					
Sector Type	Categorical	The type of educational institution					
		(e.g., Public, Private)					
TAP Recipient Headcount	Numerical	The number of students who received					
		TAP assistance.					
TAP Recipient FTEs	Numerical	Full-Time Equivalents for TAP					
		recipients.					
TAP Recipient Dollars	Numerical	The total amount in dollars of TAP					
		funding received by the recipients					
TAP Degree/NonDegree	Categorical	Indicates if the recipient is pursuing a					
		degree or non-degree program.					
Recipient Age Group	Categorical	Age group of the recipients (e.g., 18-					
		24, 25-34)					

Figure 2: Attribute Description

# 3 Statistical Methodologies:

In this project, we conducted a comprehensive analysis by starting with exploratory data analysis (EDA) and then implementing advanced statistical methodologies. Statistical tests played an important role in uncovering significant relationships between different attributes of the data. We implemented following statistical tests to achieve different objectives:

- A two-sample t-test
- Chi-square test
- ANOVA (Analysis Of Variance)
- Linear Regression Analysis

# 3.1 Exploratory Data Analysis (EDA):

#### (1) Data Collection:

- Objective: Loading and inspecting the dataset so that we can understand its basic structure and properties.
- Steps Implemented: We started by loading the dataset into a Pandas DataFrame for comprehensive analysis. We assessed the dataset information like the types of the columns, checking for null/missing values. Basically, we tried to get an overview of the dataset.
- Observations: We observed that the dataset consists of 259,983 rows and 16 columns. The columns include both numerical attributes (e.g., TAP Recipient Dollars, TAP Recipient Headcount) and categorical attributes (e.g., TAP Financial Status).

#### (2) Data Cleaning:

- **Objective**: We cleaned the dataset by handling missing values and correcting data inconsistencies.
- Steps Implemented: Checked for missing values in the dataset using .isna().sum() function. Removed rows with missing values using .dropna() function. Renamed the inconsistent column names e.g., TAP Recipient Headcount to TAP Recipient Headcount).
- Observations: We ensured that data integrity is maintained by handling the missing values, and renaming the inconsistent column names.

#### (3) Descriptive Statistical Analysis:

- Objective: We obtained summary statistics for numerical attributes and understood the data distribution.
- Steps Implemented: We used .describe() function to compute descriptive statistics, like calculating the mean, median, standard deviation, minimum, and maximum values for each numerical attribute.
- Observations: With the help of .describe() function, the data was displayed as a range of values for key numerical variables. This helped us in detecting potential outliers or extreme values in the dataset.

#### (4) Data Normalization:

- Objective: We implemented data normalization for scaling the numerical variables in order to ensure more accurate comparisons and modeling.
- Steps Implemented: We applied the StandardScaler function from sklearn.preprocessing to standardize numerical columns such as TAP Recipient Headcount, TAP Recipient FTEs, and TAP Recipient Dollars.

• **Observations**: The dataset attributes were normalized. This allowed for more comparable scaled values across numerical variables. We implemented this step because it was especially important for implementing further statistical modeling techniques.

#### 3.2 Data Distributions:

We created various visualizations to gain meaningful insights from the data, to explore trends, distributions, and relationships within the dataset. These visualizations helped us in understanding the key aspects of TAP funding, such as recipient demographics, funding trends, and correlations between different attributes like headcount, FTEs, and recipient dollars. We specifically implemented following visualizations: trends in recipient headcounts by dollar range, funding distributions for dependent and independent students, age group trends across academic years, trends across degree versus non-degree programs, and student group sizes. These visualizations helped us by providing a comprehensive view of the dataset in order to derive valuable insights and implement statistical methodologies on the data.

#### 3.2.1 Trend of Recipient Headcounts by Dollar Range.

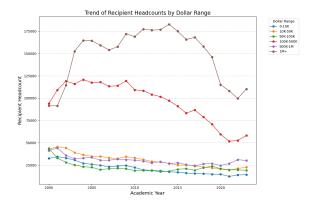


Figure 3: Trend of Recipient Headcounts by Dollar Range

• Insights: We observed notable patterns while analyzing the trend of recipient headcounts w.r.t dollar range. The "1M+" range showed a significant rise in headcounts during the early 2000s, it then peaked and sharply declined after 2015. The "100K-500K" range also shows a steady decline after peaking around 2009–2010. On the other hand, lower dollar ranges like "0-10K" and "10K-50K" remained relatively stable, and showed minor fluctuations over time. The dominance of higher dollar ranges ("1M+" and "100K-500K") in headcounts helped us in highlighting disparities in the funding distribution, as recent declines were observed in these ranges which potentially reflected the implementation of policy changes, economic shifts, or revised allocation strategies.

### 3.2.2 Trend of Dependent and Independent Funding.

 Insights: The trend of dependent and independent funding helped us in highlighting a fact that financially dependent

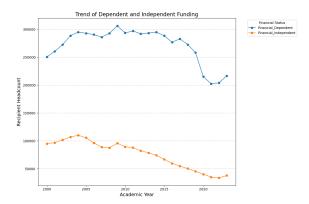


Figure 4: Trend of Dependent and Independent Funding

recipients(blue line) consistently outnumbered financially independent recipients(orange line) throughout the observed period of time. We can also observe that the dependent head-counts peaked around 2010 but experienced a steep decline after 2015, and then stabilizing slightly after 2020. In contrast, the independent headcounts exhibited a steady decline over the years, with a sharper drop after 2015. While both the groups experienced declines, the decline was proportionally sharper for independent recipients. These trends might correspond to policy, economic, or funding shifts, such as changes in eligibility criteria, program budgets, or external factors like tuition trends or student demographics.

#### 3.2.3 Trend of Recipient Age Groups Across Academic Years.

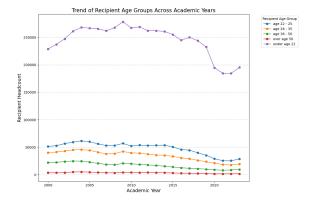


Figure 5: Trend of Recipient Age Groups Across Academic Years

Insights: The trend of recipient age groups across academic
years helped us understand key patterns in TAP funding
distribution. We can observe that the students "under age 22"
group consistently have the highest recipient headcounts,
which peaked around 2010 before it experienced a sharp
decline after 2015 and was stabilized slightly after 2020. For
other age groups, such as "age 22-25" and "age 26-35," we
can observe relatively steady trends with minor fluctuations

over time but it also exhibits declines starting around 2015. The "age 36-50" and "over age 50" groups remain the smallest categories throughout the period, with minimal changes. Basically, the overall decline across all age groups after 2015 may be attributed to changes in funding policies, eligibility criteria, or demographic shifts affecting the TAP program's reach.

#### 3.2.4 Trend of Degree or Non Degree Across Academic Years.

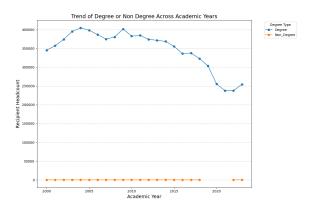


Figure 6: Trend of Degree or Non Degree Across Academic Years

• Insights: The trend of degree versus non-degree recipients across academic years highlights a disparity in headcounts. It can be observed that the degree recipients consistently formed the overwhelming majority, which peaked around 2010 and subsequently showed a steady decline after 2015, with a slight stabilization post-2020. However, on the other hand, the headcount for non-degree recipients remains negligible throughout the period, and this indicates that TAP funding primarily supports degree-seeking students. In this way, the overall decline in degree recipients after 2015 could be considered as a result of the policy changes, funding constraints, or evolving student enrollment patterns.

### 3.2.5 Number of Students in Each Group.

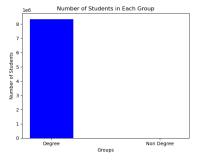


Figure 7: Number of Students in Each Group

• Insights: The visualization displaying the number of students in each group underscores the clear dominance of degree-seeking students in the TAP program. The count of degree recipients is significantly higher and has crossed 8 million, whereas the number of non-degree recipients is negligible as compared to the degree-seeking students. This stark difference highlights that TAP funding highly supports students pursuing degrees, but have minimal allocation to non-degree programs. This distribution highlights us that the program majorly focuses on aiding higher education through structured degree pathways.

# 3.2.6 Correlation Heatmap of Headcount, FTEs and recipient dollars.

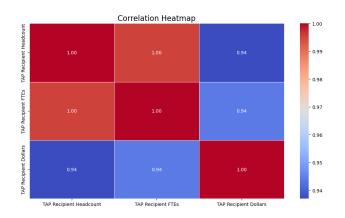


Figure 8: Correlation Heatmap of Headcount, FTEs and recipient dollars

• Insights: The correlation heatmap helps in revealing strong relationships among three different attributes including TAP recipient headcounts, full-time equivalents (FTEs), and recipient dollars. Based on the visualization, we can observe that there is a perfect correlation (1.00) between headcount and FTEs, which indicates that these metrics are directly proportional and closely linked. Additionally, the recipient dollars showcase a very high correlation (0.94) with both headcount and FTEs, and suggests that funding allocation is strongly tied to the number of recipients and their enrollment status. These insights strongly indicate that the distribution of TAP funds align proportionally with student participation, and reflects a systematic funding approach based on enrollment metrics.

## 4 Hypothesis Testing:

**Overview:** In a two-sample t-test we compared the mean TAP recipient dollars between Financial Independent and Financial Dependent groups; in Chi-Square test we examined the relationship between TAP Financial Status and TAP Degree or NonDegree classifications; we implemented ANOVA tests to evaluate differences in TAP recipient dollars across different TAP Sector Groups and Recipient Age Group categories; and linear regression analyses to investigate relationships between TAP Recipient Headcount and

TAP Recipient Dollars, as well as the impact of TAP Level of Study, Sector Type, and Income Range on TAP recipient dollars.

#### 4.1 2 sample T-Test

- **Objective**: We conducted a two-sample one-sided t-test to evaluate whether there is a significant difference in the mean TAP Recipient Dollars between financially independent and financially dependent groups.
- Output:



Figure 9: 2 sample T-Test

• Interpretation: The results of the t-test received during this hypothesis testing, gave us a T-statistic value of 14.1451 and a p-value of 0.0000, indicating strong statistical significance (p < 0.05). As a result, we reject the null hypothesis (H<sub>0</sub>) and concluded that there is a significant difference in TAP Recipient Dollars between the two groups. Specifically it was observed that the mean TAP Recipient Dollars for the financially independent group is significantly greater as compared to the financially dependent group. This observation highlighted that the financial independence might be associated with higher TAP Recipient Dollars, indicating further exploration of the underlying factors that contribute to this disparity.

# 4.2 ANOVA - Testing Differences in TAP Recipient Dollars Across Recipient Age Group Categories

- Objective: We implemented a Welch's ANOVA test to investigate whether TAP Recipient Dollars significantly differ across various age groups, where we considered unequal variances between the groups.
- Output:

<b>+</b>	+	 +		 	 	+	
Intercept	Ī			1.81544e-53			-0.0527087
							8.00958159
			0.00596034		0.0149055		
			0.00788953				0.270863

Figure 10: ANOVA analysis across Recipient Age Groups

• Interpretation: After Levene's test was implemented, it was identified that the variances among age groups were significantly different (p < 0.05), which justifies the application of Welch's ANOVA. The ANOVA results highlighted that TAP Recipient Dollars varied significantly across different age groups (F = 489.27, p < 0.001). After additional analysis through regression coefficients we received following insights:

- (1) Age 26-35: For age 26-35, the p-value (0.841) is large, so we fail to reject the null hypothesis, meaning there is no significant difference in TAP recipient dollars for this group as compared to the reference group.
- (2) For the other age groups (age 36-50, over age 50, and under age 22), the p-values are all less than 0.05, meaning we reject the null hypothesis in each case and conclude that the differences in TAP recipient dollars are statistically significant for these age groups.

#### 4.3 Chi-Square Test

- **Objective**: We implemented a Chi-Square Test for Independence to examine and understand the association between TAP Financial Status and TAP Degree or NonDegree.
- Output:

```
Chi-square statistic: 44.8589
P-value: 0.0000
Begresc of Freedom: 1
Espected Freedom: 1
[59441.6130000 192.36599131]
[[50401.31000013] 309.41300009]
[Conclusion: New righet the null hypothesis. There is an association between TAP Financial Status and TAP Degree or NonDegree
```

Figure 11: Chi-Square Test

• Interpretation: The results received from the Chi-Square test i.e., chi-square statistic = 44.8589, and p-value = 0.0000, indicate a significant association between the two attributes - TAP Financial Status and TAP Degree or NonDegree because the p-value is less than the significance level of 0.05. This indicates that the two attributes are related to each other which means that TAP Financial Status is strongly associated with whether individuals hold a degree or not.

# 4.4 ANOVA - Testing Differences in TAP Recipient Dollars Across multiple Sector Groups

- Objective: We performed ANOVA test on the transformed dataset to determine whether there are significant differences in the mean TAP Recipient Dollars across different TAP Sector Groups.
- Output:

Variable						
Intercept	0.0403981	0.00510272	7.91697	2.44312e-15	0.0303959	0.0503993
	-8.8356845	0.00878054		5.01636e-05		
C(TAP_Sector_Group)[T.3-SUNY SO]	-8.8463814	0.00560796	-7.01902	2.23978e-12		
C(TAP_Sector_Group)[T.5-IMDEPENDENT]		0.00660502		0.014959	-0.0290182	
C(TAP_Sector_Group)[T.6-BUS. DESREE]	-0.0885183			2.68964e-29	-0.0945516	-8.0664691
C(TAP_Sector_Group)[T.9-CHAPTER XXII]	-8.034675	0.0158855		0.0298512	-0.8658101	-0.00353978

Figure 12: ANOVA analysis Across multiple Sector Groups

• Interpretation: We noted that the ANOVA test results showed a significant difference in TAP Recipient Dollars across the various TAP Sector Groups with scores like F-statistic = 35.97, p-value < 0.05. Hence, we reject the null hypothesis. This helped us detect that the mean TAP Recipient Dollars varies significantly across different sectors.

Specifically, groups such as "BUS. NON-DEG" and "OTHER" showcased the largest reductions in TAP dollars as compared to the baseline group. Additionally, the p-values for most groups were observed to be well below the significance threshold and this helped us ensure that these differences are statistically significant.

#### 4.5 Linear Regression analysis

- Objective: We implemented the linear regression analysis to assess the impact of TAP level of study, sector type, and income range on TAP Recipient Dollars.
- Output:

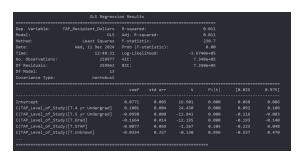


Figure 13: Linear Regression analysis

• Interpretation: Upon observing the results, we found that the overall regression model is statistically significant (F-statistic = 230.7, p-value < 0.05), and indicates that at least one of the independent variables have an effect on TAP Recipient Dollars. Among the significant predictors, the "4-year Undergrad" group receives more TAP dollars, whereas the "5-year Undergrad" and "Graduate" groups receive less TAP dollars as compared to the baseline group (2-year Undergrad). The remaining categories for TAP level of study were not found to significantly affect TAP dollars. Additionally, it was found that the higher income ranges were associated with a decrease in TAP Recipient Dollars.

#### 4.6 Linear regression

- Objective: We conducted a Pearson correlation and linear regression analysis in order to investigate the relationship between TAP Recipient Headcount and TAP Recipient Dollars.
- Output:

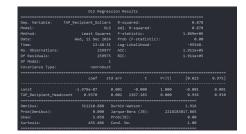


Figure 14: Linear regression

• Interpretation: Based on the results we received, we found that the Pearson correlation coefficient of 0.9370 indicates a very strong positive linear relationship between the two attributes - TAP Recipient Headcount and TAP Recipient Dollars, and a p-value of 0.0000 confirms the statistical significance of this correlation. Even the linear regression model further supports this, with a very high R-squared value of 0.878, which suggests that TAP Recipient Headcount explains approximately 87.8% of the variation in TAP Recipient Dollars. In this way, it implies that as the headcount increases, TAP recipient dollars also tend to increase in a strongly predictable manner.

#### 5 Conclusion

In this way, the project provides a comprehensive analysis of different factors affecting the TAP (Tuition Assistance Program) recipient dynamics. With the help of statistical tests and modeling techniques we have derived significant insights into the patterns of TAP disbursement.

The results highlighted that the **financial independence** is associated with higher TAP Recipient Dollars, which showcased how the economic factors play a major role in shaping TAP distributions. Similarly, the age group differences highlighted through ANOVA tests demonstrated that TAP amounts vary significantly for different stages of life, with younger and older groups receiving different level of support as compared to the middle-aged groups. An analysis of TAP sector groups helped us understand the substantial variability in TAP Recipient Dollars across different sectors, for e.g., certain groups like "BUS. NON-DEG" and "OTHER" receive significantly reduced funding as compared to the other sectors. Additionally, the association between TAP Financial Status and Degree/NonDegree attributes, provided us an evidence for a strong interdependence between educational status and the financial attributes. Furthermore, linear regression analyses provided us a clarity by quantifying how study levels, sector types, and income ranges impact the TAP disbursements. It was observed that the lower levels of study and lower income ranges were associated with higher TAP dollars. Moreover, the strong positive correlation between TAP Recipient Headcount and the TAP Recipient Dollars highlighted the scalability of funding relative to the number of beneficiaries.

In summary, this study underscores the multidimensional nature of TAP distribution across the United States, which is based on demographic, financial, and educational factors. We believe that these findings can guide policymakers and administrators in understanding and reforming the TAP funding allocation strategies to ensure equitable and efficient support for students, and addressing disparities for enhancing the program's overall effectiveness.

#### References

- [1] A. Agresti. 2018. An introduction to categorical data analysis. Wiley.
- [2] M. Dynarski and J. Scott-Clayton. 2013. The impact of need-based financial aid on college enrollment, major choice, and college completion. *Journal of Human Resources* 48, 3 (2013), 599–662.
- Higher Education Services Corporation (HESC). 2024. Tuition Assistance Program (TAP). https://www.hesc.ny.gov/pay-for-college/apply-for-financial-aid/nys-tap.html Accessed: 2024-12-12.
- [4] B. C. Long and A. J. Weiss. 2004. The impact of state student aid on college enrollment and completion. *Journal of Human Resources* 39, 2 (2004), 391–428.

- [5] J. S. Long and J. Freese. 2014. Regression models for categorical dependent variables using Stata. Stata Press.
- [6] City University of New York (CUNY). 2023. Impact of TAP on Enrollment and Equity at CUNY. https://www.cuny.edu/impact-of-tap-program Accessed: 2024-12-12.
- [7] University of Rochester Financial Aid Office. 2024. New York State Tuition Assistance Program Fact Sheet. https://www.rochester.edu/financial-aid/wp-content/uploads/2024/11/TAP-Fact-Sheet.pdf Accessed: 2024-12-12.
- [8] The Atlantic 1945. As We May Think. The Atlantic Monthly 176, 1 (1945), 101– 108 pages.
- [9] Education Week. 2024. Analysis of Financial Aid Programs and Their Impact on College Enrollment. (2024). https://www.edweek.org/education-analysis/ financial-aid-tap-impact Accessed: 2024-12-12.