Center for Undergraduate Research Opportunities Project 1: Milestone Report

This project depicts an end-to-end Data Science workflow aimed at exploring the student retention rate (whether or not a student graduates from a higher education institution). Attrition results in a severe loss of resources by society, by students, and by colleges that spend to provide programs and services to help retain and graduate students. When a student leaves college prematurely, any debt incurred must be repaid, despite the inability to graduate. In addition, the students may become turned off to the educational system in general, unexpected to benefit from education opportunities that may have aided in job attainment or advancement efforts. Administrators at historically black colleges and universities (HBCUs) report low income as a main reason for student departure among such institutions.

With the escalating cost of higher education tuition and fees, the federal and state government's have increased expenditures to higher education through financial aid programs, retention has become a salient issue. Although large public state universities offer generous aid for students, there are a number of environmental issues that can affect whether or not a university retains their students. My client would be any federal administration that demonstrates an ongoing commitment to advance diveristy and opportunity in higher education, such as the U.S. Department of Education. This analysis would help them look at other factors that may contribute to student retention or departures. Is there another way the education industry can retain students? Are there other factors that play into student retention? To what extent does income really matter when it comes to retaining undergraduates? Do certain institutions offer formal or informal student experiences that contribute to academic integration?

I believe these are important questions to ask since the rapidly growing post-secondary education industry relies on young people essentially acting as patrons of the institutions. It is important to find ways to maintain high student retention and graduation rates.

```
# Native
import datetime
import matplotlib.pyplot as plt
import os
import sys

# Non-Native
import numpy as np
import pandas as pd
import seaborn as sns
sns.set(style='whitegrid')
```

Data Acquisition

The datasets used in this analysis is derived from the National Center for Education Statistics' (NCES). The dataset is available to download from the Integrated Postsecondary Education Data System. Data reporting is mandatory for institutions that participate in or are applicants for participation in any federal student financial aid programs authorized by Title IV of the Higher Education Act of 1965.

I downloaded the data in three separate CSV files due to the nature of the data contained. The first dataframe reports institutional characteristics, the second reports 12 month enrollment counts, and the last contains statistics regarding graduation completion counts.

These 3 datasets contain about()variables that may contribute to student success including under: financial aid status, graduation rates, and demographic breakdowns of the undergraduate cohort, and student-faculty interaction.

This dataset contains approximately 5900 higher education institutions. I read the csv with pd.read_csv() into a pandas data frame.

```
df = pd.read_csv("CSV_142024-95.csv")
```

```
df2 = pd.read_csv("CSV_142024-46.csv")
```

```
df3 = pd.read_csv("CSV_142024-393.csv")
```

Data Cleaning & Pre-Processing

The dataset was already clean with few missing values. Given the dataframes containing the student demographic and achievement data, clean them up by:

- Converting the HBCU Status into a binary variable (0/1 for No and Yes, respectively).
- Convert variable names into shorter, and useable names for each data item.
- Handle any missing/ NaN values
- Ensure reported undergraduate counts and percentages are equivalent to total reported statistics for the 2021-2022 cohort.

```
df.rename(columns= {"institution name": "institution",
            "HD2022.OPE Title IV eligibility indicator code" : "TitleIV_Indicator", ##drop t
            "IC2022.Full time first-time degree/certificate-seeking undergraduate students e
            "SFA2122.Percent of undergraduate students awarded federal, state, local, institu
            "DRVGR2022.Graduation rate, total cohort" : "TotalGradRate",
            "DRVGR2022.Graduation rate, American Indian or Alaska Native" : "AmericanIndianGr
            "DRVGR2022.Graduation rate, Asian" : "AsianGradRate",
            "DRVGR2022.Graduation rate, Native Hawaiian or Other Pacific Islander" : "Pacific
            "DRVGR2022.Graduation rate, Black, non-Hispanic" : "BlackGradRate",
            "DRVGR2022.Graduation rate, Hispanic" : "HispanicGradRate",
```

```
            "DRVGR2022.Graduation rate, White, non-Hispanic" : "WhiteGradRate",
            "DRVGR2022.Graduation rate, two or more races" : "MultiRacialGradRate",
            "DRVGR2022.Graduation rate, Race/ethnicity unknown" : "UnreportedRaceGradRate",
            "DRVGR2022.Transfer-out rate - Bachelor cohort" : "TransferOutRate",
            "HD2022.Historically Black College or University" : "HBCUStatus",
            "HD2022.Postsecondary and Title IV institution indicator" : "PostSecondaryIndic",
            "DRVEF122022.Undergraduate 12-month unduplicated headcount" : "UndergradCount",
            "DRVEF122022.Percent of undergraduate 12-month unduplicated headcount that are An
"DRVEF122022.Percent of undergraduate 12-month unduplicated headcount that are Asian" : "Asi
"DRVEF122022.Percent of undergraduate 12-month unduplicated headcount that are Black or Afri
"DRVEF122022.Percent of undergraduate 12-month unduplicated headcount that are Hispanic/Lati
"DRVEF122022.Percent of undergraduate 12-month unduplicated headcount that are Native Hawaii
"DRVEF122022.Percent of undergraduate 12-month unduplicated headcount that are White" : "Whi
"DRVEF122022.Percent of undergraduate 12-month unduplicated headcount that are two or more r
"DRVEF122022.Percent of undergraduate 12-month unduplicated headcount that are race/ethnicit
```

```
df.drop(columns=["TitleIV_Indicator", "PostSecondaryIndic", "DegreeSeeking"], inplace=True)
```

```
df["TransferOutCount"] = df["UndergradCount"] * df["TransferOutRate"]
df["TotalGrantedAid"] = df["PercentGrantedAid"] * df["UndergradCount"]
```

```
df["unitid"].is_unique
```

```
True
```

```
df["HBCUStatus"].replace(to_replace = dict(Yes = 1, No = 0), inplace=True)
```

```
df.dropna(inplace=True)
```

```
df.isnull().sum()
```

```
unitid                              0
institution                         0
year                                0
PercentGrantedAid                   0
TotalGradRate                       0
AmericanIndianGradRate              0
AsianGradRate                       0
PacificIslanderGradRate             0
BlackGradRate                       0
HispanicGradRate                    0
WhiteGradRate                       0
MultiRacialGradRate                 0
UnreportedRaceGradRate              0
TransferOutRate                     0
HBCUStatus                          0
UndergradCount                      0
AmericanIndianUndergradPercent      0
AsianUndergradPercent               0
BlackUndergradPercent               0
```

```
HispanicUndergradPercent           0
PacificIslanderUndergradPercent    0
WhiteUndergradPercent              0
MultiRacialUndergradPercent        0
UnreportedRaceUndergradPercent     0
TransferOutCount                   0
TotalGrantedAid                    0
dtype: int64
```

```
df2.rename(columns= {"EFFY2022.Level and degree/certificate-seeking status of student" : "St
                     "EFFY2022.Grand total" : "TotalUndergradCount",
                     "EFFY2022.American Indian or Alaska Native total" : "AmericanIndianUnder
                     "EFFY2022.Asian total" : "AsianUndergradCount",
                     "EFFY2022.Black or African American total" : "BlackUndergradCount",
                     "EFFY2022.Hispanic or Latino total" : "HispanicUndergradCount",
                     "EFFY2022.Native Hawaiian or Other Pacific Islander total" : "PacificIsl
                     "EFFY2022.White total" : "WhiteUndergradCount",
                     "EFFY2022.Two or more races total" : "MultiRacialUndergradCount",
                     "EFFY2022.Race/ethnicity unknown total" : "UnreportedRaceUndergradCount"
```

```
df2.drop(columns=["StudentLevel", "IDX_E12"], inplace=True)
```

```
df2.isnull().sum()
```

```
unitid                           0
institution name                 0
year                             0
TotalUndergradCount              0
AmericanIndianUndergradCount     0
AsianUndergradCount              0
BlackUndergradCount              0
HispanicUndergradCount           0
PacificIslanderUndergradCount    0
WhiteUndergradCount              0
MultiRacialUndergradCount        0
UnreportedRaceUndergradCount     0
dtype: int64
```

```
df3.rename(columns= {"C2022_C.Award Level code" : "DegreeType",
                     "C2022_C.Grand total" : "TotalStudentsGraduating",
                     "C2022_C.American Indian or Alaska Native total" : "AmericanIndianGradu
                     "C2022_C.Asian total" : "AsianGraduatingCount",
                     "C2022_C.Black or African American total" : "BlackGraduatingCount",
                     "C2022_C.Hispanic or Latino total" : "HispanicGraduatingCount",
                     "C2022_C.Native Hawaiian or Other Pacific Islander total" : "PacificIsl
                     "C2022_C.White total" : "WhiteGraduatingCount",
                     "C2022_C.Two or more races total" : "MultiRacialGraduatingCount",
                     "C2022_C.Race/ethnicity unknown total" : "UnreportedRaceGraduatingCount
```

```
                              , inplace=True)
df3.drop(columns = ["DegreeType", "IDX_C"], inplace=True)
df3.isnull().sum()
```

```
unitid                            0
institution name                  0
year                              0
TotalStudentsGraduating           0
AmericanIndianGraduatingCount     0
AsianGraduatingCount              0
BlackGraduatingCount              0
HispanicGraduatingCount           0
PacificIslanderGraduatingCount    0
WhiteGraduatingCount              0
MultiRacialGraduatingCount        0
UnreportedRaceGraduatingCount     0
dtype: int64
```

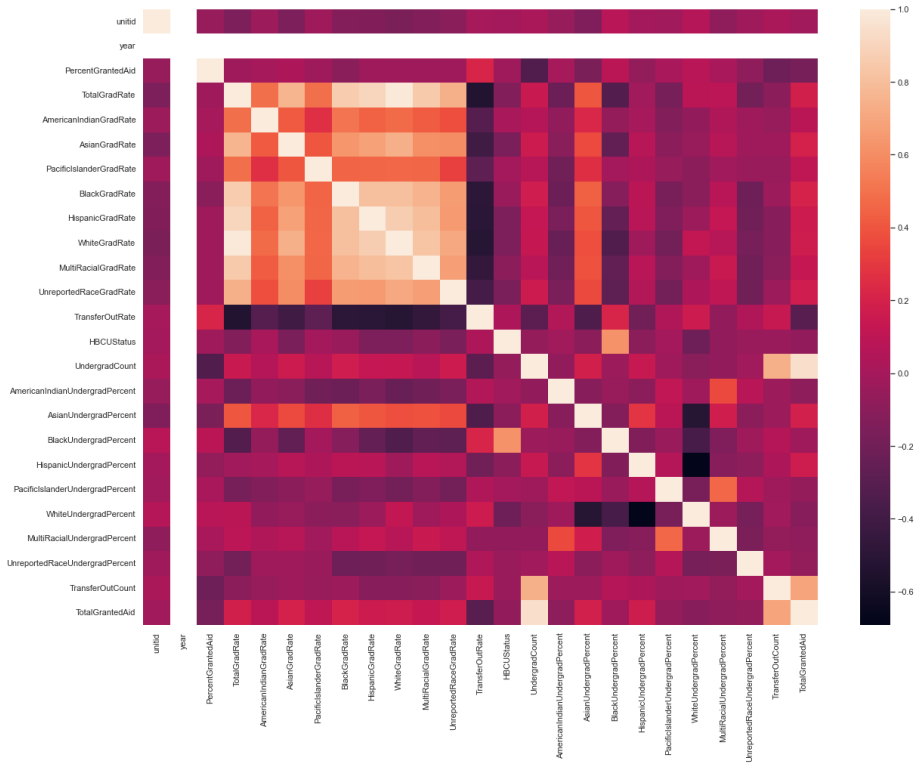
One discrepancy found during the data cleaning process is the reported total count of undergraduate students. When adding up the counts of undergraduate demographics, I found that the reported count did not match the breakdown of the student demographics. - For example, for Alabama A&M, the total undergraduate count was reported as 5663. When adding up the race demographic counts, it totaled as 5610. This is likely due to multiple categorizations of race, and the addition of Unreported or Multi Racial categories for students that identify with two or more, or chose to leave the field unreported. - For this, I created a column in the dataframe to calculate the Actual Undergraduate Count

```
df2['ActualUndergrad_Count'] = df2["AmericanIndianUndergradCount"]+ df2["AsianUndergradCount
```

Exploratory Data Analysis

```
corr1 = df.corr()
plt.subplots(figsize=(20,15))
sns.heatmap(corr1,
            xticklabels=corr1.columns.values,
            yticklabels=corr1.columns.values)
```

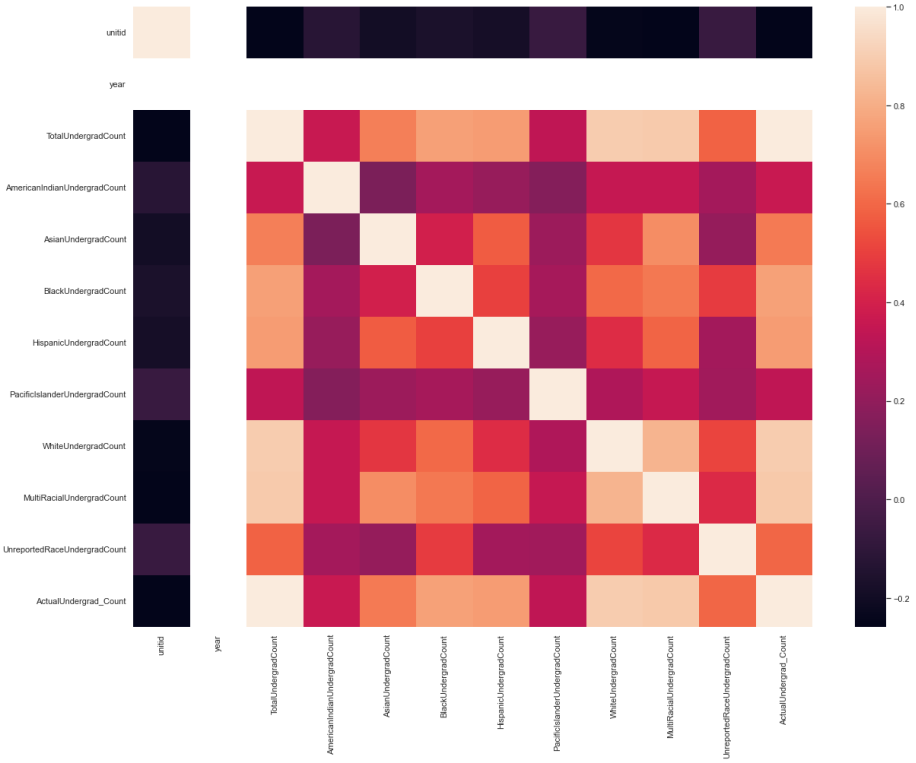
```
<AxesSubplot:>
```



I created a correlation plot to see if any of the institutional characteristic variables were correlated with each other. I found that very few variables were correlated. This correlation plot shows that each demographic's graduation rate and undergraduate percent are all correlated with each other. The total graduation rate is also correlated with the white grad rate, multi-racial grad rate, hispanic grad rate, and black grad rate.

```
corr2 = df2.corr()
plt.subplots(figsize=(20,15))
sns.heatmap(corr2,
            xticklabels=corr2.columns.values,
            yticklabels=corr2.columns.values)
```

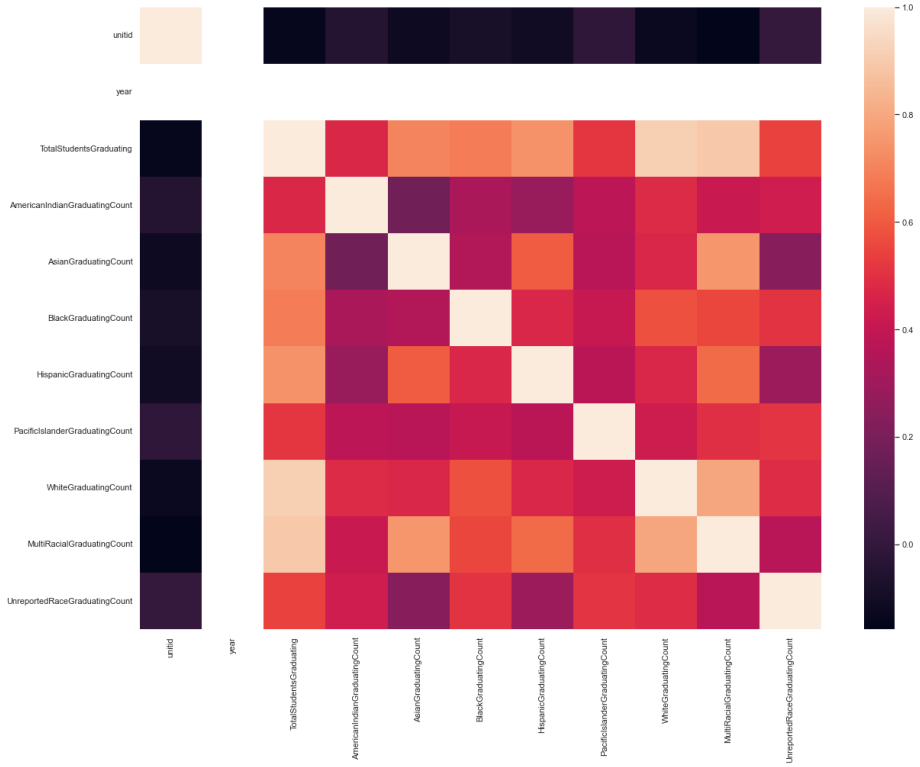
```
<AxesSubplot:>
```



This correlation plot shows if any of the undergraduate 12-month enrollment variables were correlated with each other. I found that very few variables were correlated. This correlation plot shows that the white and multiracial undergraduate count is correlated with the total undergraduate count.

```
corr3 = df3.corr()
plt.subplots(figsize=(20,15))
sns.heatmap(corr3,
            xticklabels=corr3.columns.values,
            yticklabels=corr3.columns.values)
```

```
<AxesSubplot:>
```



This correlation plot to see if any of the variables reporting graduation counts were correlated with each other. I found that very few variables were correlated. This correlation plot shows that each white graduation counts and multiracial graduation counts are correlated. Additionally, the Asian graduation count is correlated with multiracial graduating count.

The analysis of the correlation plots reiterates the importance of drawing insights from subgroups, to gain a picture of what compostitional diversity looks like in American schools.

Correlation Tests

Tests whether two samples have a linear relationship.

```
from scipy.stats import pearsonr
stat, p = pearsonr(df['TotalGradRate'], df['WhiteGradRate'])
print('stat= %.3f, p= %.3f' % (stat, p))
if p > 0.05:
    print('Total Graduation Rate and White Graduation Rate are probably independent')
else:
    print('Total Graduation Rate and White Graduation Rate are probably dependent')
```

```
stat= 0.981, p= 0.000
```

```
Total Graduation Rate and White Graduation Rate are probably dependent
```

```
stat, p = pearsonr(df['TotalGradRate'], df['BlackGradRate'])
print('stat= %.3f, p= %.3f' % (stat, p))
if p > 0.05:
    print('Total Graduation Rate and Black Graduation Rate are probably independent')
else:
    print('Total Graduation Rate and Black Graduation Rate are probably dependent')
```

```
stat= 0.859, p= 0.000
Total Graduation Rate and Black Graduation Rate are probably dependent
```

```
stat, p = pearsonr(df['TotalGradRate'], df['AsianGradRate'])
print('stat= %.3f, p= %.3f' % (stat, p))
if p > 0.05:
    print('Total Graduation Rate and Asian Graduation Rate are probably independent')
else:
    print('Total Graduation Rate and Asian Graduation Rate are probably dependent')
```

```
stat= 0.757, p= 0.000
Total Graduation Rate and Asian Graduation Rate are probably dependent
```

```
stat, p = pearsonr(df['TotalGradRate'], df['HispanicGradRate'])
print('stat= %.3f, p= %.3f' % (stat, p))
if p > 0.05:
    print('Total Graduation Rate and Hispanic Graduation Rate are probably independent')
else:
    print('Total Graduation Rate and Hispanic Graduation Rate are probably dependent')
```

```
stat= 0.902, p= 0.000
Total Graduation Rate and Hispanic Graduation Rate are probably dependent
```

```
stat, p = pearsonr(df['TotalGradRate'], df['MultiRacialGradRate'])
print('stat= %.3f, p= %.3f' % (stat, p))
if p > 0.05:
    print('Total Graduation Rate and Mutli Racial Graduation Rate are probably independent')
else:
    print('Total Graduation Rate and Mutli Racial Graduation Rate are probably dependent')
```

```
stat= 0.848, p= 0.000
Total Graduation Rate and Mutli Racial Graduation Rate are probably dependent
```

```
stat, p = pearsonr(df['TotalGradRate'], df['MultiRacialGradRate'])
print('stat= %.3f, p= %.3f' % (stat, p))
if p > 0.05:
    print('Total Graduation Rate and Mutli Racial Graduation Rate are probably independent')
else:
    print('Total Graduation Rate and Mutli Racial Graduation Rate are probably dependent')
```

```
stat= 0.848, p= 0.000
Total Graduation Rate and Mutli Racial Graduation Rate are probably dependent
```

```
stat, p = pearsonr(df['TotalGradRate'], df['UnreportedRaceGradRate'])
print('stat= %.3f, p= %.3f' % (stat, p))
if p > 0.05:
    print('Total Graduation Rate and Unreported Race Graduation Rate are probably independer
else:
    print('Total Graduation Rate and Unreported Race Graduation Rate are probably dependent'
```

```
stat= 0.738, p= 0.000
Total Graduation Rate and Unreported Race Graduation Rate are probably dependent
```

```
stat, p = pearsonr(df['TransferOutRate'], df['WhiteUndergradPercent'])
print('stat= %.3f, p= %.3f' % (stat, p))
if p > 0.05:
    print('The transfer out rate and White Undergraduate Percent are probably independent')
else:
    print('The transfer out rate and White Undergraduate Percent are probably dependent')
```

```
stat= 0.160, p= 0.003
The transfer out rate and White Undergraduate Percent are probably dependent
```

```
stat, p = pearsonr(df['PercentGrantedAid'], df['TransferOutRate'])
print('stat= %.3f, p= %.3f' % (stat, p))
if p > 0.05:
    print('The transfer out rate and Percent Granted Aid are probably independent')
else:
    print('The transfer out rate and Percent Granted Aid are probably dependent')
```

```
stat= 0.215, p= 0.000
The transfer out rate and Percent Granted Aid are probably dependent
```

```
stat, p = pearsonr(df['TotalGradRate'], df['PercentGrantedAid'])
print('stat= %.3f, p= %.3f' % (stat, p))
if p > 0.05:
    print('Total Graduation Rate and Percent Granted Aid are probably independent')
else:
    print('Total Graduation Rate and Percent Granted Aid are probably dependent')
```

```
stat= -0.020, p= 0.706
Total Graduation Rate and Percent Granted Aid are probably independent
```

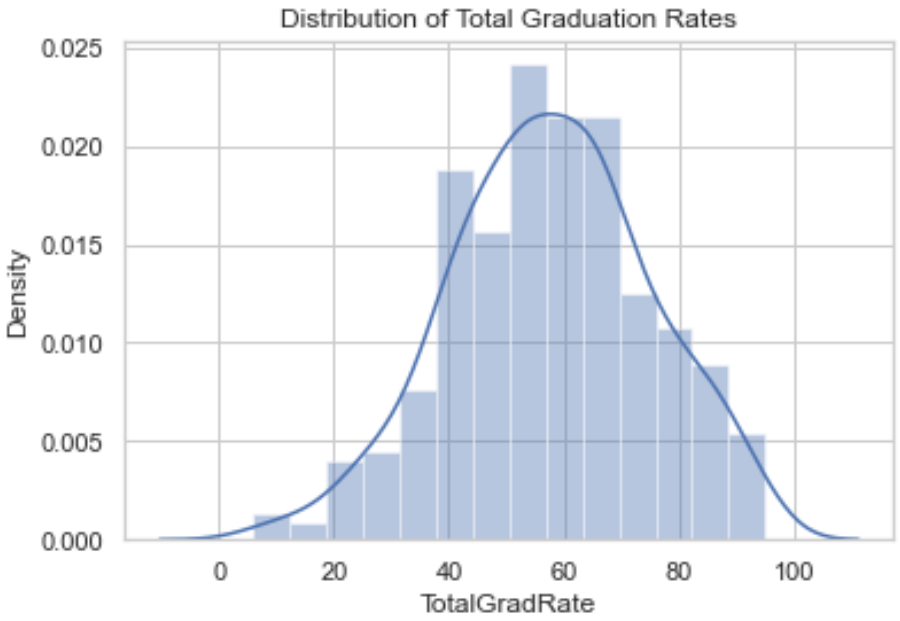
```
stat, p = pearsonr(df2['TotalUndergradCount'], df2['WhiteUndergradCount'])
print('stat= %.3f, p= %.3f' % (stat, p))
if p > 0.05:
    print('Total Undergrad Count and White Undergrad Count are probably independent')
else:
    print('Total Undergrad Count and White Undergrad Count are probably dependent')
```

```
stat= 0.894, p= 0.000
Total Undergrad Count and White Undergrad Count are probably dependent
```

Visualizations

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Total Graduation Rates')
sns.distplot(df.TotalGradRate)
```

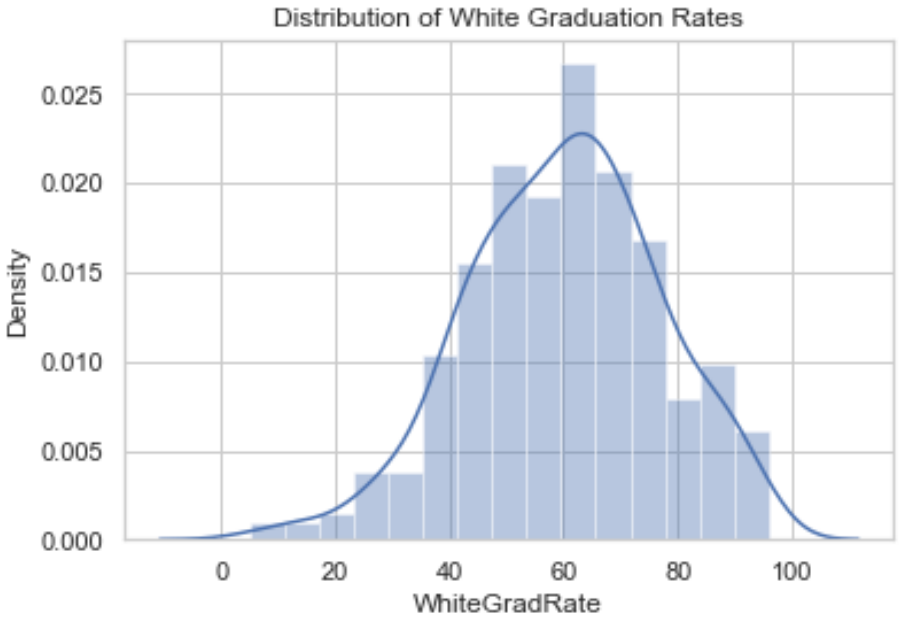
```
<AxesSubplot:title={'center':'Distribution of Total Graduation Rates'}, xlabel='TotalGradRat
```



This graph indicates that the average total graduation rate is approximately 50%.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of White Graduation Rates')
sns.distplot(df.WhiteGradRate)
```

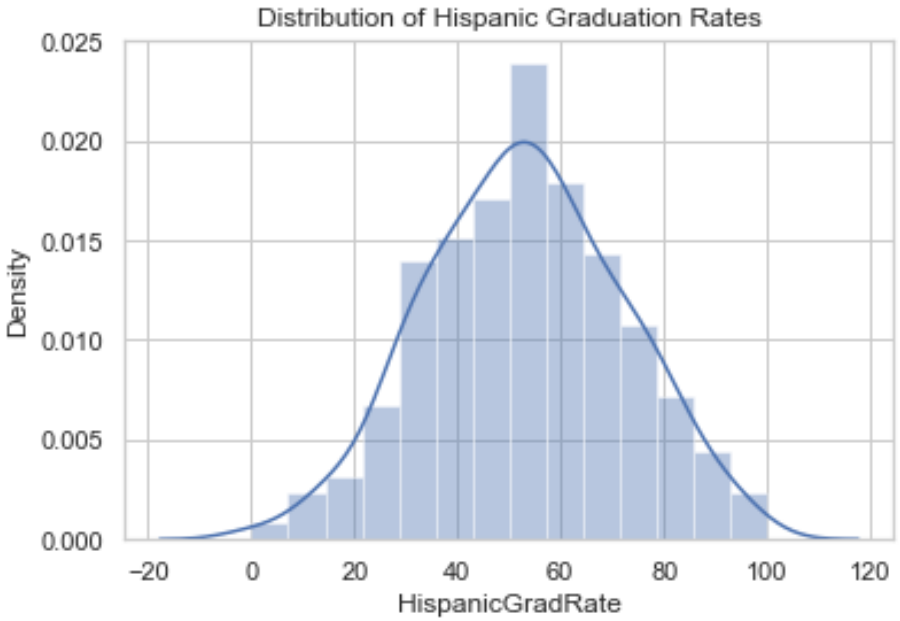
```
<AxesSubplot:title={'center':'Distribution of White Graduation Rates'}, xlabel='WhiteGradRat
```



This graph indicates that the average graduation rate among white unergraduates is approximately 60%.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Hispanic Graduation Rates')
sns.distplot(df.HispanicGradRate)
```

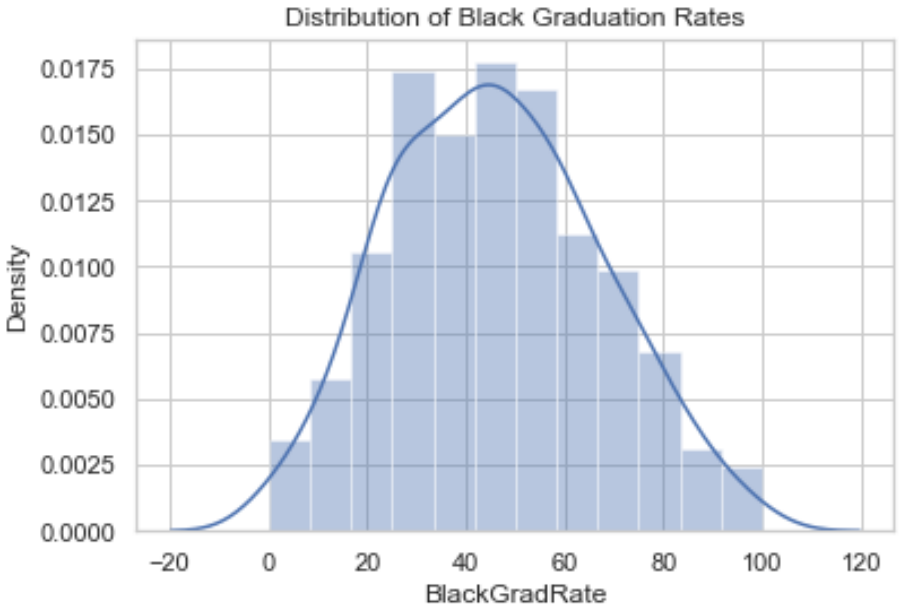
```
<AxesSubplot:title={'center':'Distribution of Hispanic Graduation Rates'}, xlabel='Hispanic(
```



This graph indicates that the average graduation rate among hispanic undergraduates is approximately 50%.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Black Graduation Rates')
sns.distplot(df.BlackGradRate)
```

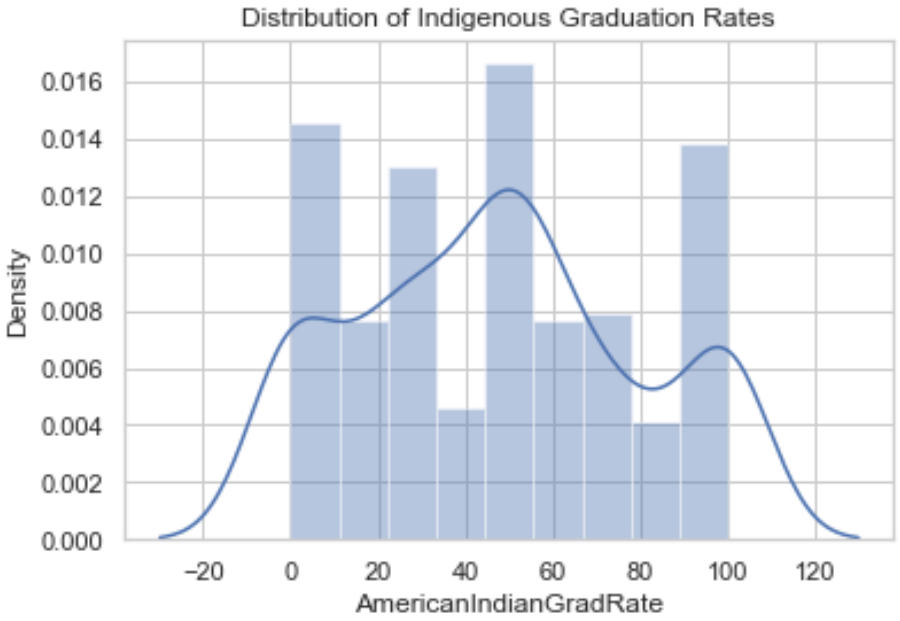
```
<AxesSubplot:title={'center':'Distribution of Black Graduation Rates'}, xlabel='BlackGradRat
```



This graph indicates that the average graduation rate of black undergraduate students is approximately 50%. The graph is slightly right-skewed, indicating that the majority of students have a graduation rate of less than 50%.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Indigenous Graduation Rates')
sns.distplot(df.AmericanIndianGradRate)
```

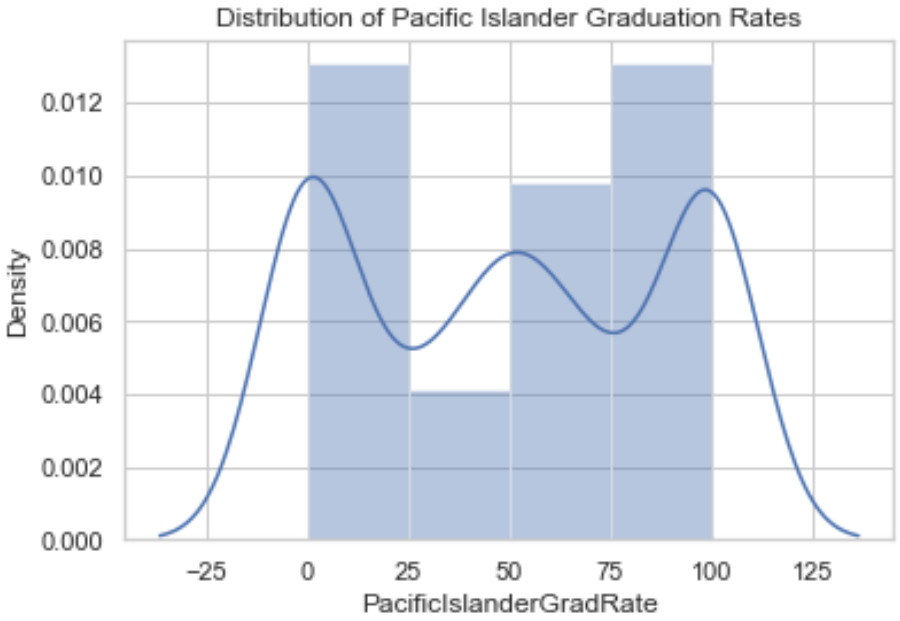
```
<AxesSubplot:title={'center':'Distribution of Indigenous Graduation Rates'}, xlabel='America
```



This graph indicates that the graduation rates of indigenous undergraduate students vary greatly, with an average graduation rate of 50%.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Pacific Islander Graduation Rates')
sns.distplot(df.PacificIslanderGradRate)
```

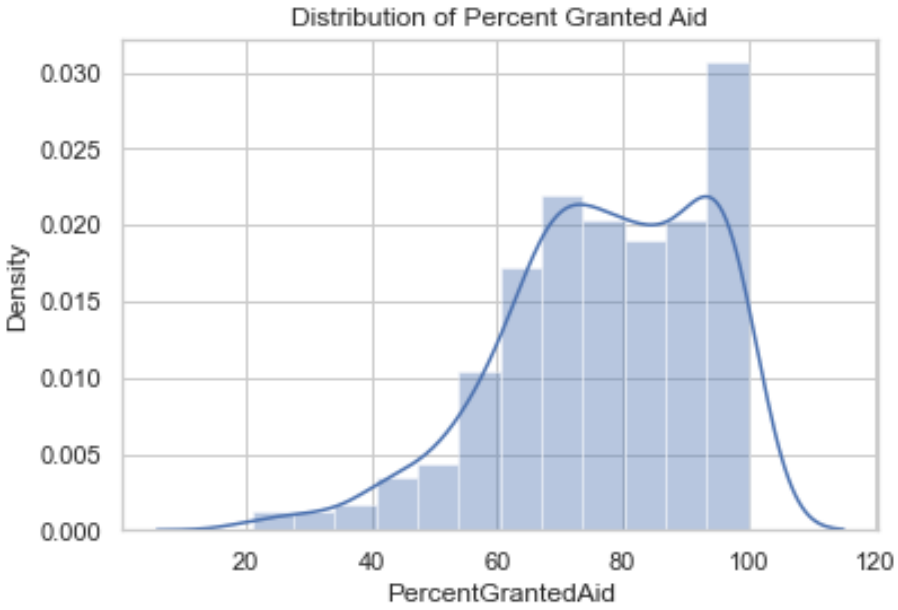
```
<AxesSubplot:title={'center':'Distribution of Pacific Islander Graduation Rates'}, xlabel='P
```



The graph indicates that the average graduation rate of pacific islander students is highly varied.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Percent Granted Aid')
sns.distplot(df.PercentGrantedAid)
```

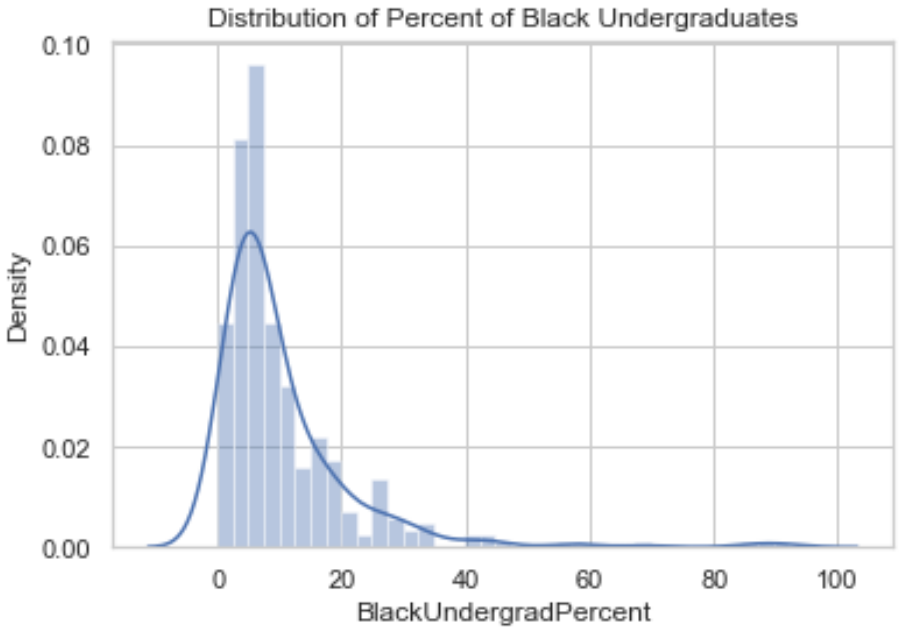
```
<AxesSubplot:title={'center':'Distribution of Percent Granted Aid'}, xlabel='PercentGranted
```



This graph shows that the average percent of students granted financial aid is near 90% and is left-skewed due to the long tail on the left, indicating that the mean is less than the median. The bulk of the observations are toward 80% or greater.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Percent of Black Undergraduates')
sns.distplot(df.BlackUndergradPercent)
```

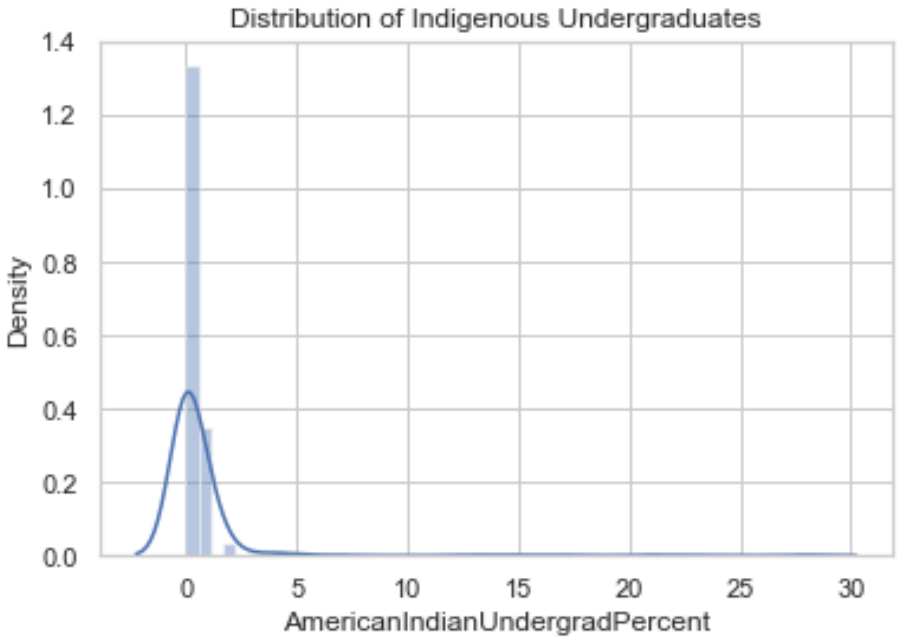
```
<AxesSubplot:title={'center':'Distribution of Percent of Black Undergraduates'}, xlabel='Bla
```



This graph shows that the average percent of black undergraduate students is approximately near 10%, and the graph is right-skewed, indicating that the majority of Title IV schools have less than 15% of black undergraduate students.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Indigenous Undergraduates')
sns.distplot(df.AmericanIndianUndergradPercent)
```

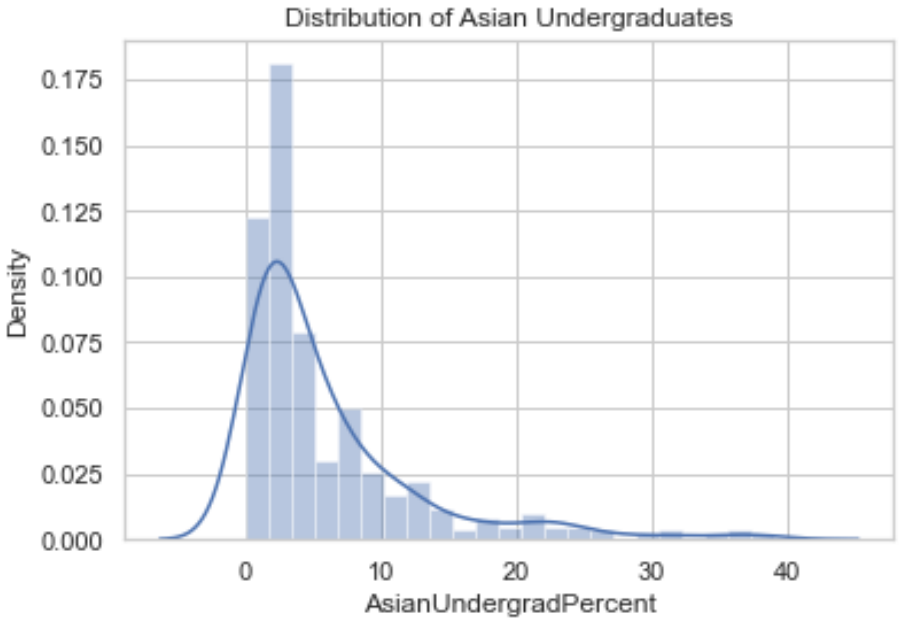
```
<AxesSubplot:title={'center':'Distribution of Indigenous Undergraduates'}, xlabel='American
```



This graph indicates the breakdown of American Indian Undergraduates. The graph is right-skewed, indicating that the majority of undergraduate populations in Title IV schools have less than 5% of Indigenuous students.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Asian Undergraduates')
sns.distplot(df.AsianUndergradPercent)
```

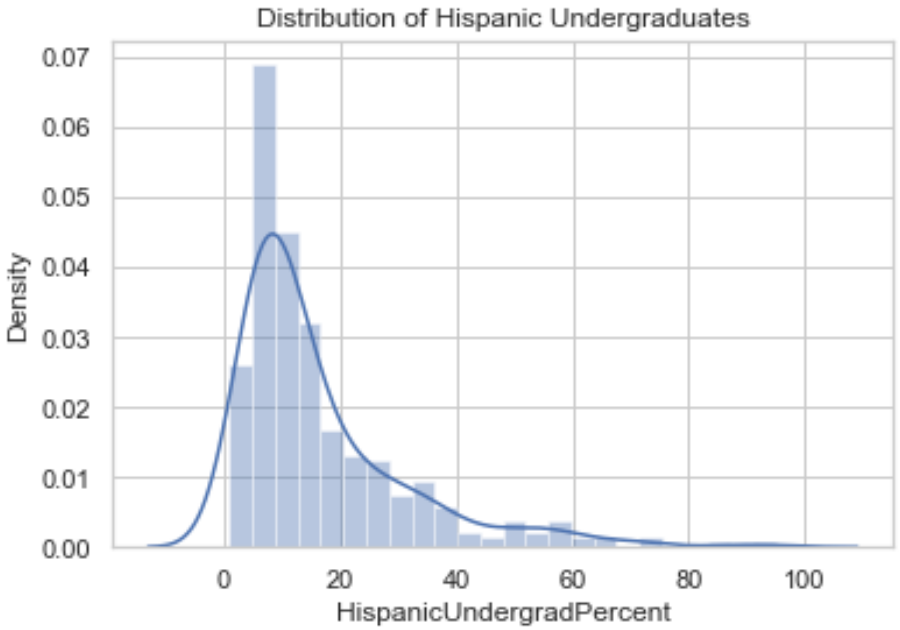
```
<AxesSubplot:title={'center':'Distribution of Asian Undergraduates'}, xlabel='AsianUndergrad
```



This graph indicates the breakdown of Asian Undergraduates. The graph is right-skewed, indicating that the majority of undergraduate populations in Title IV schools have less than 20% of Asian students.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Hispanic Undergraduates')
sns.distplot(df.HispanicUndergradPercent)
```

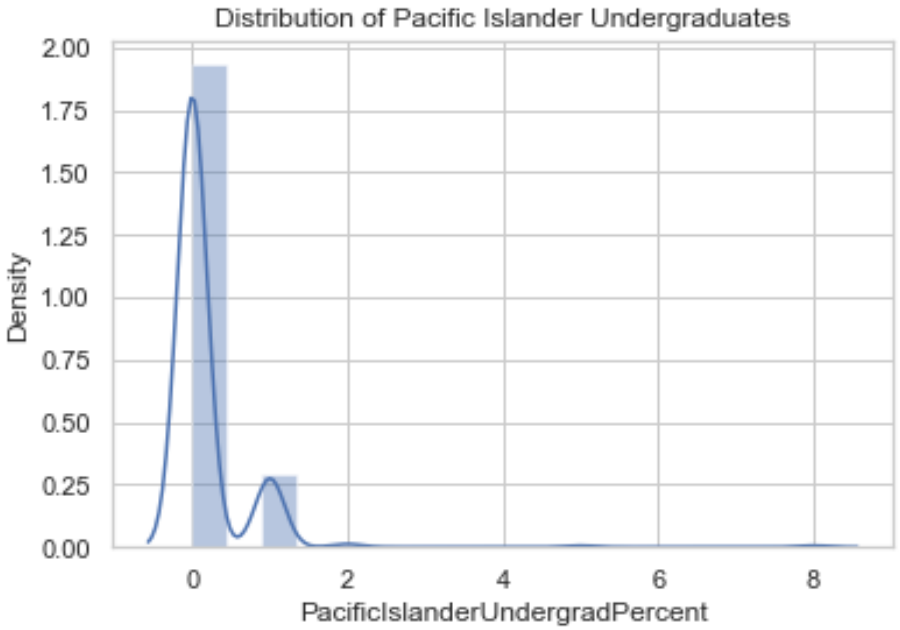
```
<AxesSubplot:title={'center':'Distribution of Hispanic Undergraduates'}, xlabel='HispanicUnd
```



This graph indicates the breakdown of Hispanic Undergraduates. The graph is right-skewed, indicating that the majority of undergraduate populations in Title IV schools have less than 30% of Hispanic students.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Pacific Islander Undergraduates')
sns.distplot(df.PacificIslanderUndergradPercent)
```

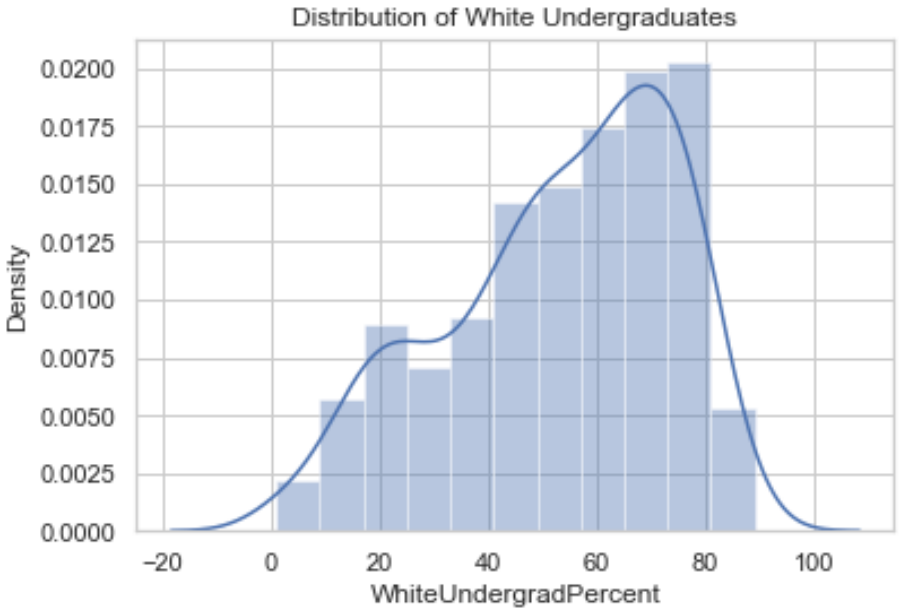
```
<AxesSubplot:title={'center':'Distribution of Pacific Islander Undergraduates'}, xlabel='Pac
```



This graph indicates the breakdown of Pacific Islander Undergraduates. The graph is right-skewed, indicating that the majority of undergraduate populations in Title IV schools have less than 2% of Pacific Islander students.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of White Undergraduates')
sns.distplot(df.WhiteUndergradPercent)
```

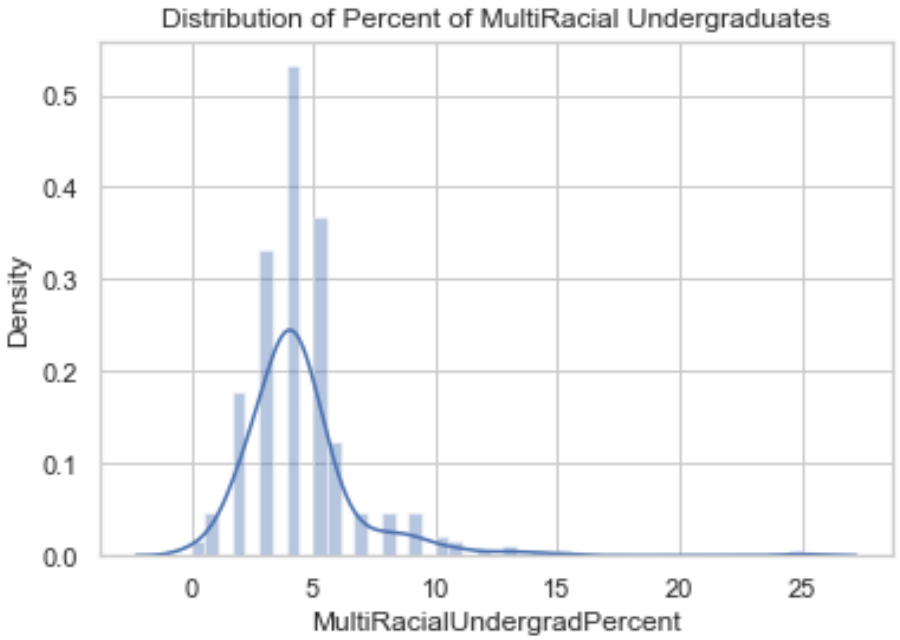
```
<AxesSubplot:title={'center':'Distribution of White Undergraduates'}, xlabel='WhiteUndergrad
```



This graph indicates the breakdown of White Undergraduates. The graph is slightly left-skewed, indicating that the majority of undergraduate populations in Title IV schools have more than 50% of White students.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Percent of MultiRacial Undergraduates')
sns.distplot(df.MultiRacialUndergradPercent)
```

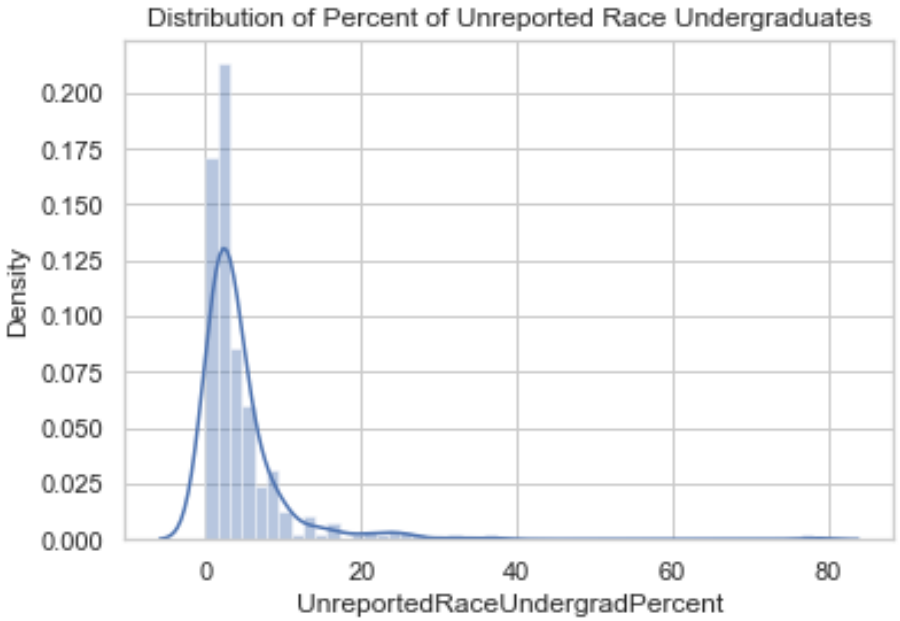
```
<AxesSubplot:title={'center':'Distribution of Percent of MultiRacial Undergraduates'}, xlabe
```



This graph indicates the breakdown of Multi-Racial Undergraduate students. This graph is left-skewed, indicating that the majority of undergraduate populations have between 0-10% of Multi-Racial students.

```
fig, ax = plt.subplots()
ax.set_title('Distribution of Percent of Unreported Race Undergraduates')
sns.distplot(df.UnreportedRaceUndergradPercent)
```

```
<AxesSubplot:title={'center':'Distribution of Percent of Unreported Race Undergraduates'}, x
```



This graph indicates the breakdown of Unreported Race Undergraduate students. This graph is left-skewed, indicating that the majority of undergraduate populations have between 0-20% of Unreported Race students.

```
subset = df.sample(n=200)
```

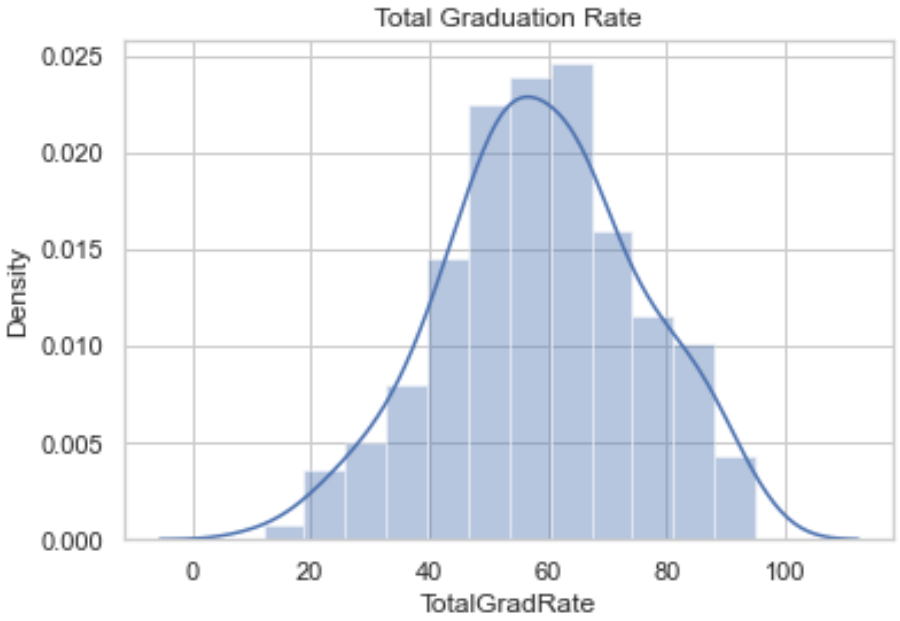
```
subset.shape
```

```
(200, 26)
```

Creating a subset of size (n=200) randomly sampled schools to extract insights from a smaller group of schools. This allows us to perform deeper analysis on variables using a smaller subset of all 5,000 Title IV schools.

```
fig, ax = plt.subplots()
ax.set_title('Total Graduation Rate')
sns.distplot(subset.TotalGradRate)
```

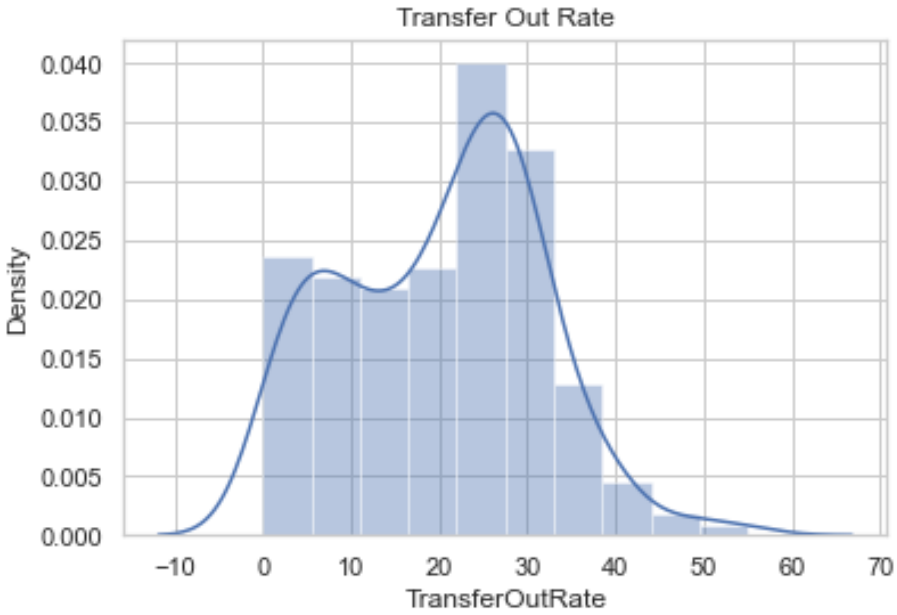
```
<AxesSubplot:title={'center':'Total Graduation Rate'}, xlabel='TotalGradRate', ylabel='Densi
```



This graph indicates the distribution of total graduation rates. This graph is slightly right-skewed, indicating that the majority of institutions have a total graduation rate of greater than 60%.

```
fig, ax = plt.subplots()
ax.set_title('Transfer Out Rate')
sns.distplot(subset.TransferOutRate)
```

```
<AxesSubplot:title={'center':'Transfer Out Rate'}, xlabel='TransferOutRate', ylabel='Density
```

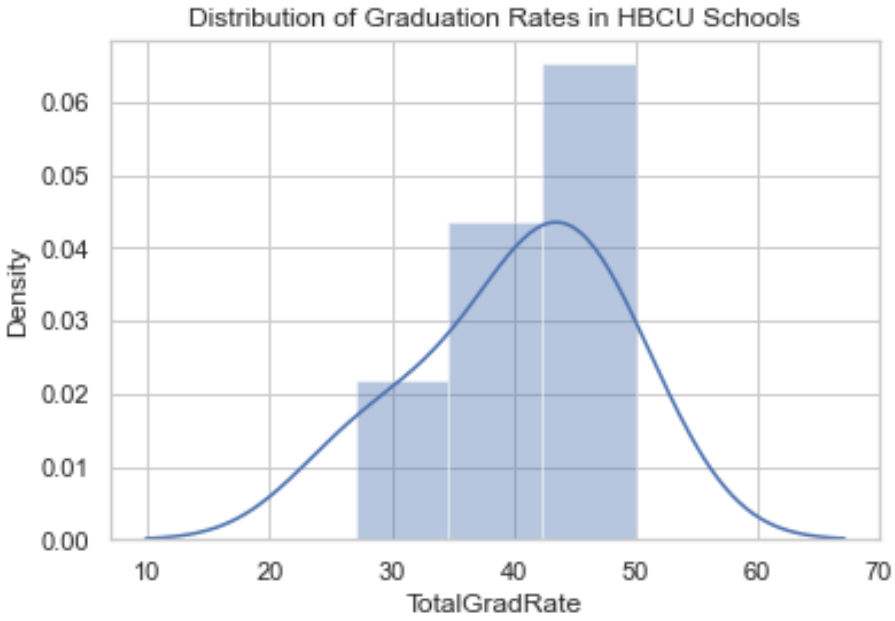


This graph indicates the distribution of transfer rates across Title IV schools. This graph is right-skewed with a few notable outliers. The average transfer out rate is approximately 30%, and the right-skewed nature indicates that the majority of schools have a transfer out rate that is greater than 30%.

```
#Get dataframe where school is HBCU status
HBCU_df = df[df['HBCUStatus'] == 1]
```

```
fig, ax = plt.subplots()
ax.set_title("Distribution of Graduation Rates in HBCU Schools")
sns.distplot(HBCU_df.TotalGradRate)
```

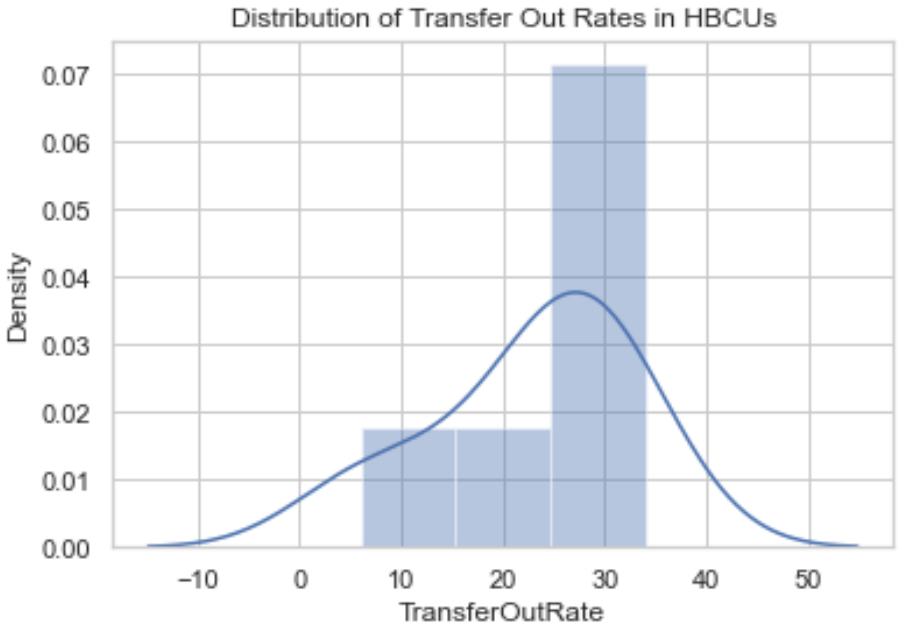
```
<AxesSubplot:title={'center':'Distribution of Graduation Rates in HBCU Schools'}, xlabel='To
```



The graduation rates among HBCU schools are left-skewed, with an average graduation rate of 45%. The left-skewed nature of the graph indicates that the majority of HBCU schools hold graduation rates greater than 45%.

```
fig, ax = plt.subplots()
ax.set_title("Distribution of Transfer Out Rates in HBCUs")
sns.distplot(HBCU_df.TransferOutRate)
```

```
<AxesSubplot:title={'center':'Distribution of Transfer Out Rates in HBCUs'}, xlabel='Transfe
```

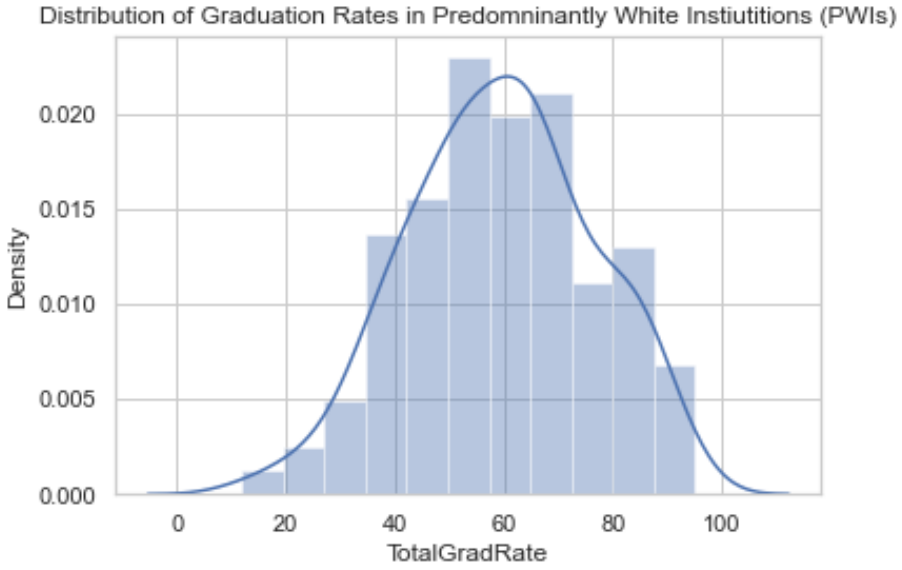


The transfer out rates among HBCU schools are slightly left-skewed, with an average transfer rate of 30%. The left-skewed nature of the graph indicates that the majority of HBCU schools hold transfer rates greater than 30%.

```
PWI_df = df[df['WhiteUndergradPercent'] >= 50]
```

```
fig, ax = plt.subplots()
ax.set_title("Distribution of Graduation Rates in Predomninantly White Instiutitions (PWIs)"
sns.distplot(PWI_df.TotalGradRate)
```

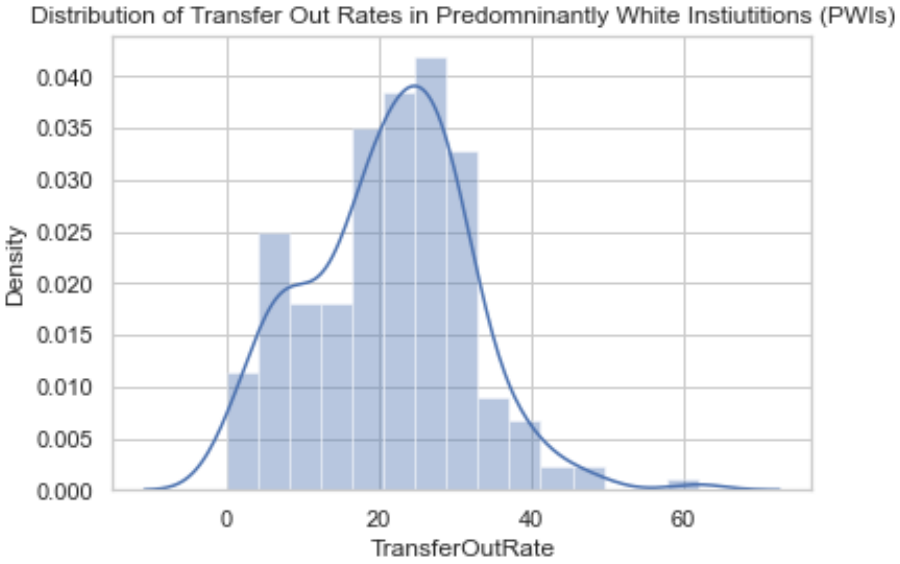
```
<AxesSubplot:title={'center':'Distribution of Graduation Rates in Predomninantly White Insti
```



This graph indicates the distribution of the average graduation rate in institutions where white undergraduates account for greater than 50% of the undergraduate population. The average total graduation rate among PWIs is approximately 60%.

```
fig, ax = plt.subplots()
ax.set_title("Distribution of Transfer Out Rates in Predomninantly White Instiutitions (PWIs
sns.distplot(PWI_df.TransferOutRate)
```

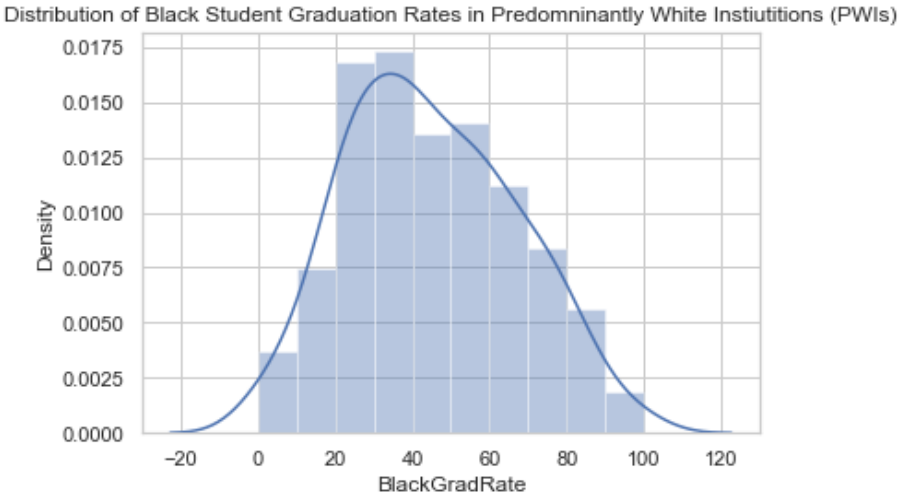
```
<AxesSubplot:title={'center':'Distribution of Transfer Out Rates in Predomninantly White Ins
```



This graph indicates the distribution of the transfer out rate in institutions where white undergraduates account for greater than 1/2 of the undergraduate population. The average transfer rate among PWIs is approximately 30%. The right-skewed nature of this graph indicates that the majority of sampled institutions have a transfer rate less than 30%.

```
fig, ax = plt.subplots()
ax.set_title("Distribution of Black Student Graduation Rates in Predomninantly White Instiut
sns.distplot(PWI_df.BlackGradRate)
```

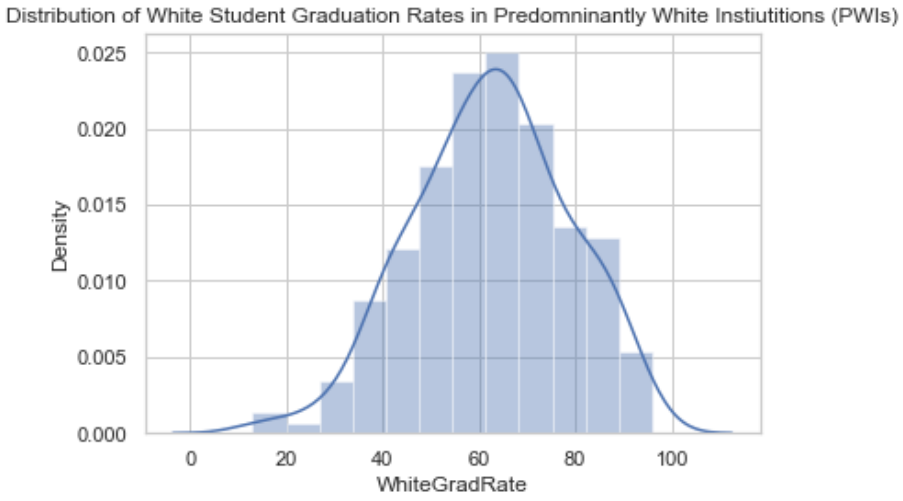
```
<AxesSubplot:title={'center':'Distribution of Black Student Graduation Rates in Predomninant
```



This graph indicates the distribution of black undergraduate graduation rates in institutions where white undergraduates account for greater than 1/2 of the undergraduate population. The average black graduation rate in PWIs is approximately 40%. The left-skewed nature of this graph indicates that the majority of sampled institutions have a black undegraduate graduation rate of greater than 40%.

```
fig, ax = plt.subplots()
ax.set_title("Distribution of White Student Graduation Rates in Predomninantly White Instiut
sns.distplot(PWI_df.WhiteGradRate)
```

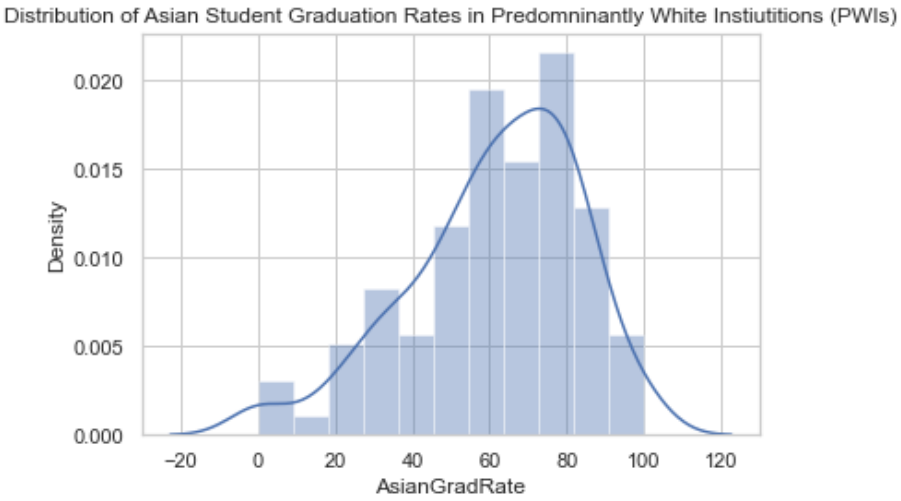
```
<AxesSubplot:title={'center':'Distribution of White Student Graduation Rates in Predomninant
```



This graph indicates the distribution of the white undergraduate graduation rates in institutions where white undergraduates account for greater than 1/2 of the undergraduate population. The average transfer rate among PWIs is approximately 70%. The left-skewed nature of this graph indicates that the majority of sampled institutions have a white undergraduate graduation rate greater than 70%.

```
fig, ax = plt.subplots()
ax.set_title("Distribution of Asian Student Graduation Rates in Predomninantly White Instiut
sns.distplot(PWI_df.AsianGradRate)
```

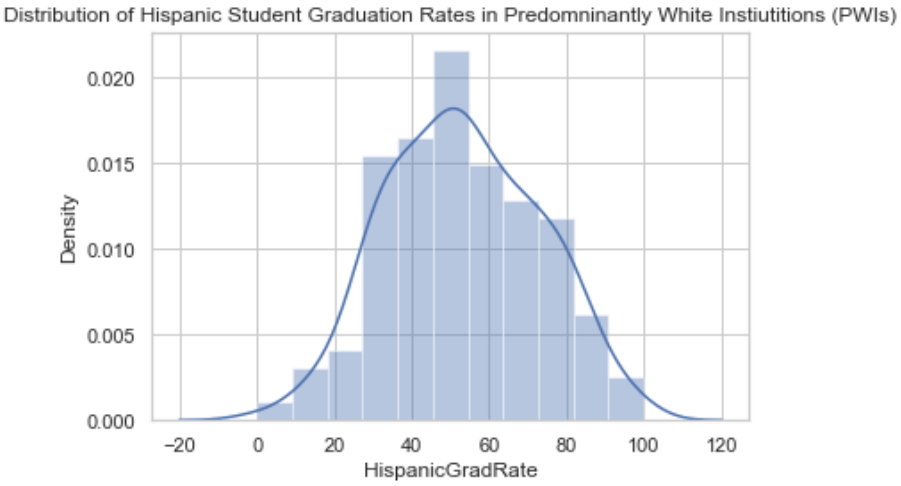
```
<AxesSubplot:title={'center':'Distribution of Asian Student Graduation Rates in Predomninant
```



This graph indicates the distribution of the Asian graduation rates in institutions where white undergraduates account for greater than 1/2 of the undergraduate population. The average Asian graduation rate among PWIs is approximately 75%. The left-skewed nature of this graph indicates that the majority of sampled institutions have an asian undergraduate graduation rate of greater than 75%.

```
fig, ax = plt.subplots()
ax.set_title("Distribution of Hispanic Student Graduation Rates in Predomninantly White Inst
sns.distplot(PWI_df.HispanicGradRate)
```

```
<AxesSubplot:title={'center':'Distribution of Hispanic Student Graduation Rates in Predomnir
```

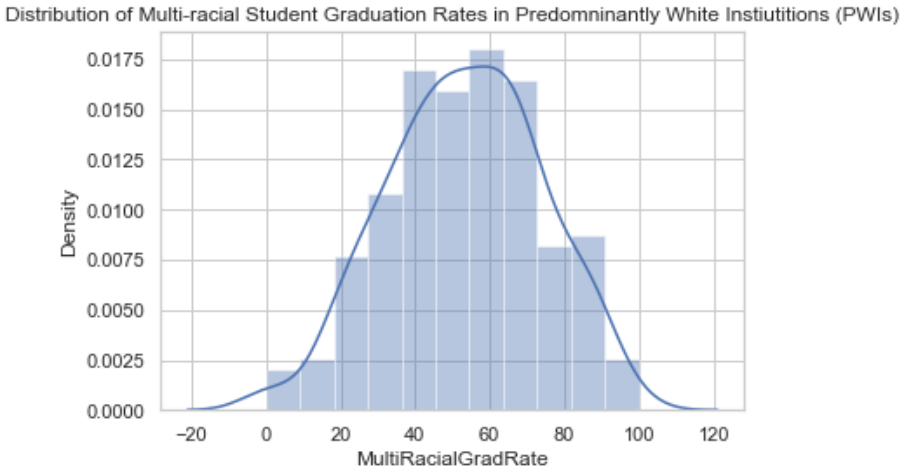


This graph indicates the distribution of the Hispanic undergraduate graduation rates in institutions where white undergraduates account for greater than 1/2

of the undergraduate population. The average Hispanic graduation rate among PWIs is approximately 50%.

```
fig, ax = plt.subplots()
ax.set_title("Distribution of Multi-racial Student Graduation Rates in Predomninantly White
sns.distplot(PWI_df.MultiRacialGradRate)
```

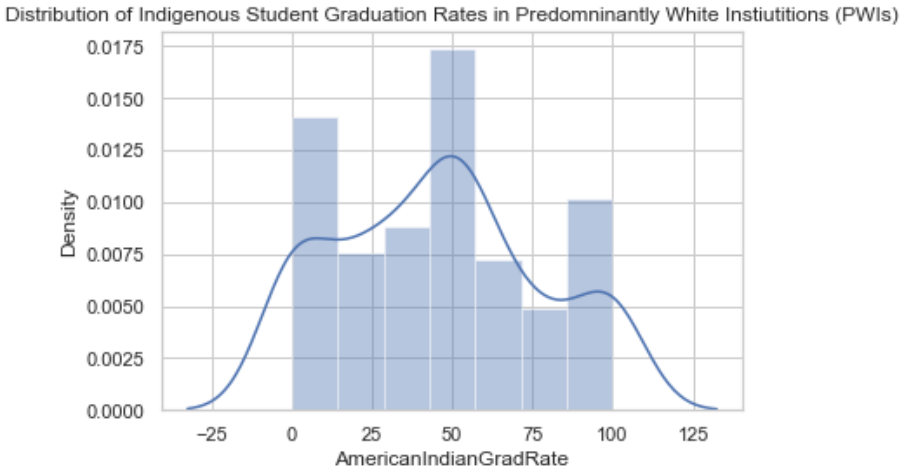
```
<AxesSubplot:title={'center':'Distribution of Multi-racial Student Graduation Rates in Predo
```



This graph indicates the distribution of the multi-racial undergraduate graduation rates in institutions where white undergraduates account for greater than 1/2 of the undergraduate population. The average multi-racial graduation rate among PWIs is approximately 55%.

```
fig, ax = plt.subplots()
ax.set_title("Distribution of Indigenous Student Graduation Rates in Predomninantly White In
sns.distplot(PWI_df.AmericanIndianGradRate)
```

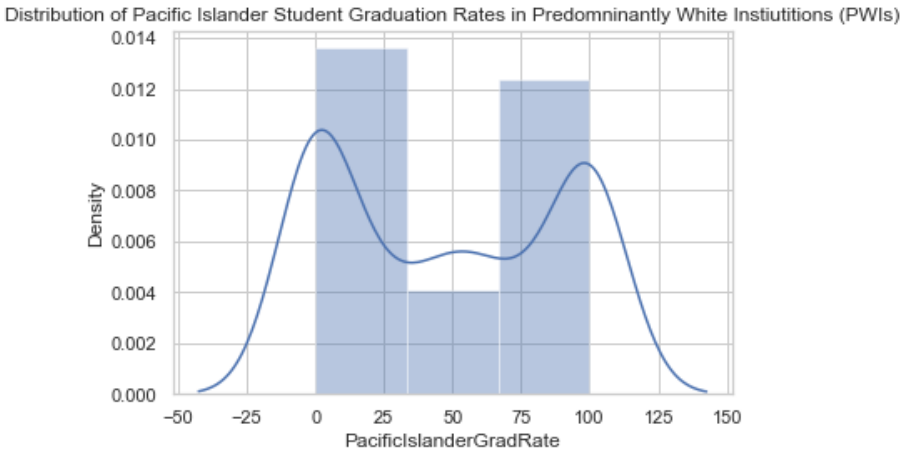
```
<AxesSubplot:title={'center':'Distribution of Indigenous Student Graduation Rates in Predomn
```



This graph indicates the distribution of the American Indian undergraduate graduation rate in institutions where white undergraduate students account for greater than 1/2 of the undergraduate population. The distribution for the American Indian Graduation Rate varies greatly. The average American Indian Graduation Rate among PWIs is approximately 50%.

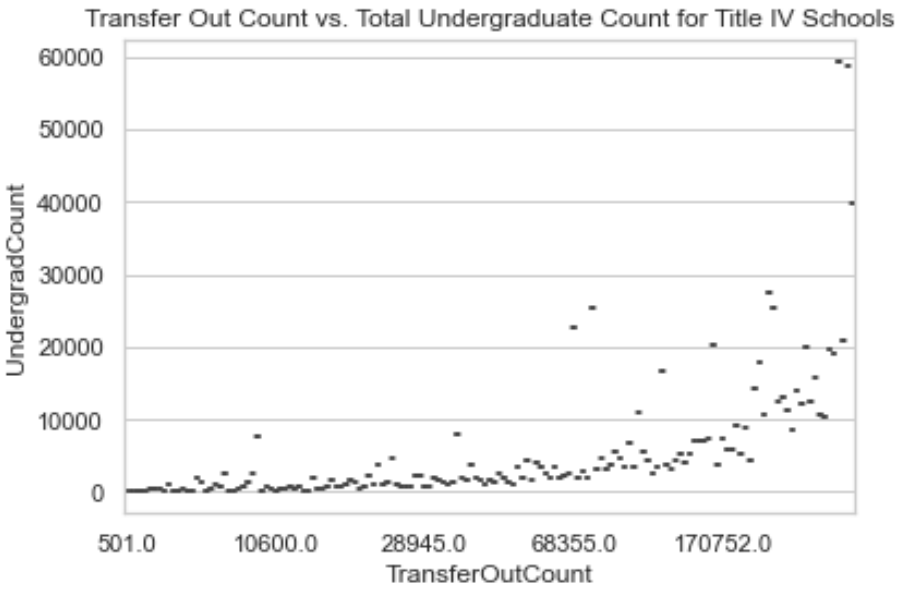
```
fig, ax = plt.subplots()
ax.set_title("Distribution of Pacific Islander Student Graduation Rates in Predomninantly Wh
sns.distplot(PWI_df.PacificIslanderGradRate)
```

```
<AxesSubplot:title={'center':'Distribution of Pacific Islander Student Graduation Rates in P
```



This graph indicates the distribution of the Pacific Islander Graduation rate in institutions where white undergraduates account for greater than 1/2 of the undergraduate population. The average pacific islander graduation rate greatly varies among PWI institutions.

```
fig, ax = plt.subplots()
ax.set_title('Transfer Out Count vs. Total Undergraduate Count for Title IV Schools')
sns.boxplot(x='TransferOutCount', y='UndergradCount', data=subset)
plt.locator_params(axis='x', nbins=5)
```



This plot shows the distribution of students that transferred out versus the students that stayed in their institution.