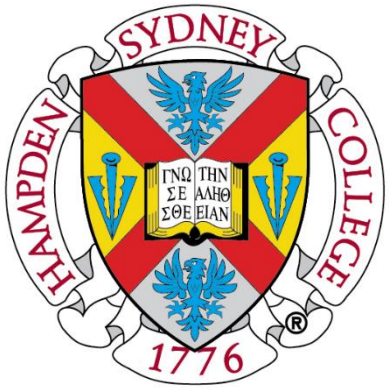


Hampden Sydney College

Computer Science Department 2021

COMS 385 Special Topics in Computer Science

Report on:

"Impact on Enrollment due to Tuition"

Prepared by:

Michael Davis

Table of Contents

Introduction:

For our final exam in COMS 385 Special Topics in Computer Science, we were tasked to investigate a problem or question of our choosing that could be solved or answered with data. For my topic I chose to discuss my findings *on Changes in Enrollment in Colleges around the US due to Tuition changes.* To discuss this further I will be explaining my data collection and research process with the certain data that I chose for my project in the Methodology section below.

Methodology:

During the research process I had many different ideas of how I would visualize this data from different data sources and data types. I first wanted to look back over the last five years of data which included changes in enrollment, Tuition from private, in state and out of state public colleges, admission, admission rate, and admission to enrollment ratio to show the changes. I found most of my data on Wikipedia within the College Admission in the US section. The websites table went back 20 years, so I decided to base my data off a 20-year scale just to help visualize the data more than only 5 years. Using Beautiful Soup, I was able to parser the html data and then finding the certain table within the html code that I wanted to pull the data. After that I converted the tables into an appended data frame that was converted to a CSV file making it easier to plot using MatPlot lib. The collection process was interesting as I found data on tuition changes comparing the different school types on US News but couldn't scrap the data off the html code. I was able to download the data into a CSV file and then put the data into my code. With that I was able to open the csv file and easily plot using Matplot lib to show the data.

In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt

cost = pd.read_csv("schooldata.csv")
costchange = pd.DataFrame(cost)
costchange
```

In [ ]:
```
fig, ax = plt.subplots(dpi=150)
costchange.plot(kind='line', x ='Year', marker = 'o',mfc ='white' ,ax=ax)
ax.set_xlabel("Year Change")
ax.set_ylabel("Price Change")
plt.title("Changes in Tuition over the last 20 years")

plt.show()
```

In [ ]:
```
import requests
import pandas as pd
from bs4 import BeautifulSoup
```

In [ ]:
```
# Downloading contents of the web page
url = "https://en.wikipedia.org/wiki/College_admissions_in_the_United_States"
data = requests.get(url).text
```

In [ ]:
```
soup = BeautifulSoup(data, 'html.parser')
```

In [ ]:
```
print('Classes of each table: ')
for table in soup.find_all('table'):
    print(table.get('class'))
```

In [ ]:
```
tables = soup.find_all('table')

table = soup.find('table', class_='wikitable')
```

In [ ]:
```
df = pd.DataFrame(columns=['Admit Year', 'Apps', 'Admits', 'Enroll' 'Admit Rate', 'Admit: Enroll'])

for row in table.tbody.find_all('tr'):
    columns = row.find_all('td')

    if(columns != []):
        admityear = columns[0].text.strip()
        Apps = columns[1].text.strip()
        Admits = columns[2].text.strip()
        Enroll = columns[3].text.strip()
        Admit_Rate = columns[4].text.strip()
        Admit_Enroll = columns[5].text.strip()

        df = df.append({'Admit Year': admityear, 'Apps': Apps, 'Admits': Admits, 'Enroll': Enroll, 'Admit Rate': Admit_Rate, 'Adr
        df.to_csv("schoolwikidata.csv", sep=',')
```

In [ ]:
```
admission = pd.read_csv("wikidata.csv")
admissionchanges = pd.DataFrame(admission)
admissionchanges
```

In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import csv

filename = 'wikidata.csv'

with open(filename) as f:
    reader = csv.reader(f)
    header_row = next(reader)
    apps = []
    for row in reader:
        if row[1]=='':
            continue
        app = int(row[1])
        apps.append(app)


#reading the database

data = pd.read_csv("wikidata.csv")

fig, ax = plt.subplots(dpi=150)
plt.plot(data['Admit Year'], apps, c = 'red' , marker = 'o', mfc = 'white')


# Adding Title to the Plot
plt.title("Application Numbers over the Last 20 Years")

# Setting the X and Y Labels
plt.xlabel('Admit Year')
plt.ylabel('Applications in Millions')
```

In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import csv

filename = 'wikidata.csv'

with open(filename) as f:
    reader = csv.reader(f)
    header_row = next(reader)
    rates = []
    for row in reader:
        rate = float(row[2])
        rates.append(rate)


#reading the database

data = pd.read_csv("wikidata.csv")

fig, ax = plt.subplots(dpi=150)
plt.plot(data['Admit Year'], rates, c = 'blue' , marker = 'o', mfc = 'white')


# Adding Title to the Plot
plt.title("Students Admitted over the Last 20 Years")

# Setting the X and Y Labels
plt.xlabel('Admission Year')
plt.ylabel('Student Admitted')


plt.show()
```

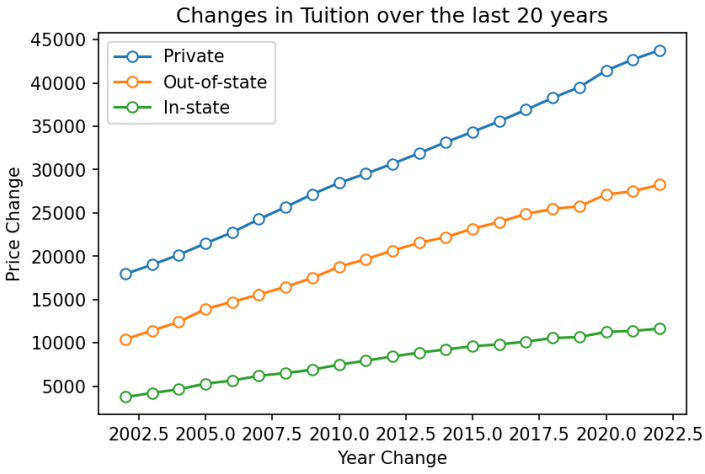
In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import csv

filename = 'wikidata.csv'

with open(filename) as f:
    reader = csv.reader(f)
    header_row = next(reader)
    rates = []
    for row in reader:
        rate = float(row[3])
        rates.append(rate)


#reading the database

data = pd.read_csv("wikidata.csv")

fig, ax = plt.subplots(dpi=150)
plt.plot(data['Admit Year'], rates, c = 'green' , marker = 'o', mfc = 'white')


# Adding Title to the Plot
plt.title("Students Enrolled over the Last 20 Years")

# Setting the X and Y Labels
plt.xlabel('Admission Year')
plt.ylabel('Students Enrolled')


plt.show()
```

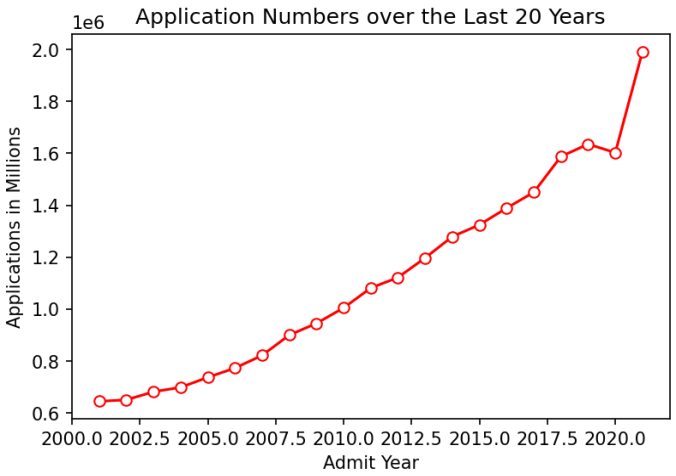
In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import csv

filename = 'wikidata.csv'

with open(filename) as f:
    reader = csv.reader(f)
    header_row = next(reader)
    rates = []
    for row in reader:
        rate = float(row[4])
        rates.append(rate)


#reading the database

data = pd.read_csv("wikidata.csv")

fig, ax = plt.subplots(dpi=150)
plt.plot(data['Admit Year'], rates, c = 'black' , marker = 'o', mfc = 'white')


# Adding Title to the Plot
plt.title("Admission Rate over the Last 20 Years")

# Setting the X and Y Labels
plt.xlabel('Admission Year')
plt.ylabel('Admission Rate')


plt.show()
```

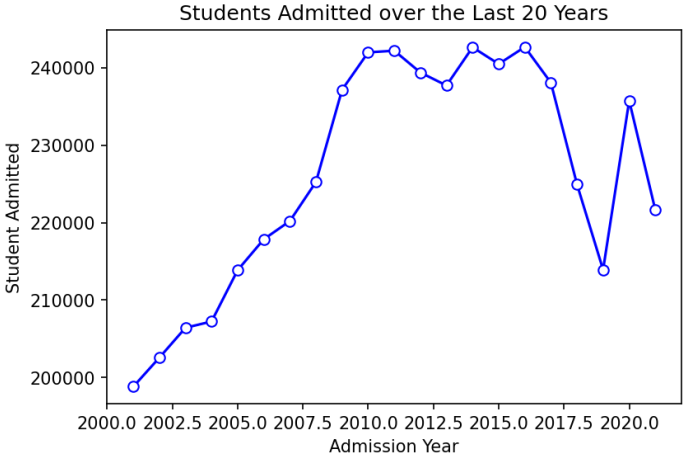
In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import csv

filename = 'wikidata.csv'

with open(filename) as f:
    reader = csv.reader(f)
    header_row = next(reader)
    rates = []
    for row in reader:
        rate = float(row[5])
        rates.append(rate)


#reading the database

data = pd.read_csv("wikidata.csv")

fig, ax = plt.subplots(dpi=150)
plt.plot(data['Admit Year'], rates, c = 'purple' , marker = 'o', mfc = 'white')


# Adding Title to the Plot
plt.title("Admission Enrollment Ratio the Last 20 Years")

# Setting the X and Y Labels
plt.xlabel('Admission Year')
plt.ylabel('Admit: Enroll Ratio')


plt.show()
```

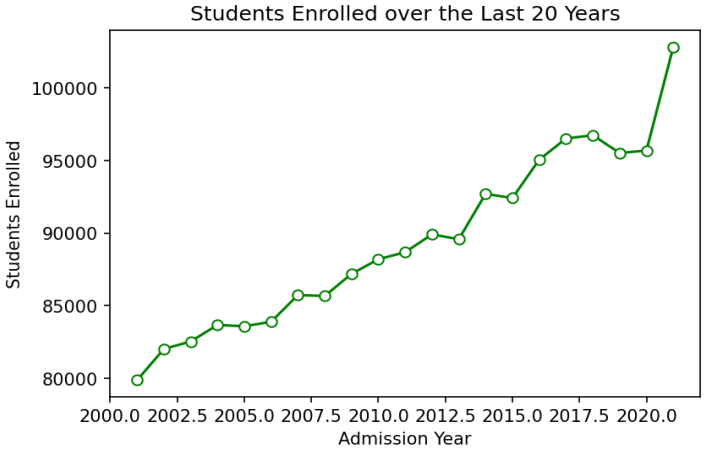
Results:

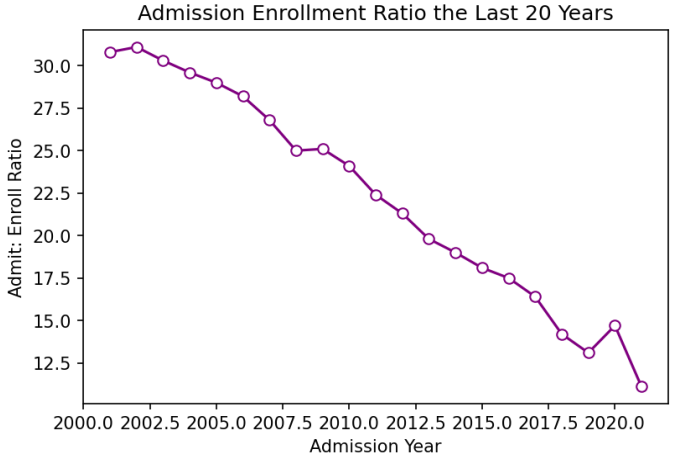
After showing some screenshots of my code to help the understanding of the process of plotting and collecting my data we can see the visualizations of my graphs. For my graphs I decided to plot them in a basic line graph style to easily demonstrate the changes over the last 20 years correlating to each of the data types that I chose. Using the code segments above I was able to show the overall changes of each variable in the y axis with correlation to the time difference in years as the x axis.

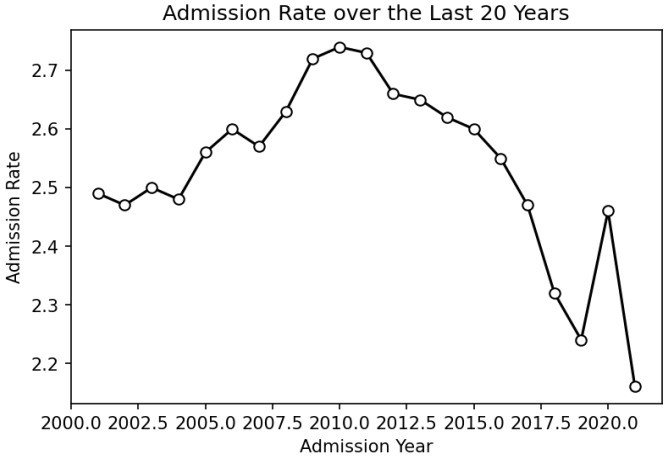

Changes in Tuition over the last 20 years
Private
Out-of-state
In-state
Price Change
Year Change


Application Numbers over the Last 20 Years
1e6
Applications in Millions
Admit Year


Students Admitted over the Last 20 Years
Student Admitted
Admission Year


Students Enrolled over the Last 20 Years
Students Enrolled
Admission Year


Admission Enrollment Ratio the Last 20 Years
Admit: Enroll Ratio
Admission Year


Admission Rate over the Last 20 Years
Admission Rate
Admission Year

Discussion:

As we can see from the previous results, the constant marginal increase in tuition prices has been going up for the last twenty years for all three types of colleges including Private, Public Out of State and In State. Though with the two public variables we can see the same amount of incline of price for their margin but comparing the price to private we can see a massive difference of an average of 45,000 dollars a year for private colleges nationwide compared to in state and out of state averaging around 15,000 dollars. This is quite interesting to see how these prices vary so much for tuition but as we know from colleges that public colleges get state funding making it much easier to maintain the college without raising tuition to cost justify expenses unlike private colleges. Looking over the next variable we can see a huge increase of applications numbers ranging from 600,000 around the beginning of 2000 to now in 2020 around 2 million. That is a huge exponential increase especially in 2020 with the graph having a high upward shift from around 1.6 million to 2 million in only a year. Going off of that we shift to the next data graph displaying the changes in students admitted as we see a exponential increase in numbers till 2010 ranging from an increase of 30,000 from 20,000 to 230,000 then hitting a plateau until taking a huge dip in numbers around 2016 until 2018. Average dropped back down around to 210,000 then shifting back up around 2019 but 2020 then dropping again. This is interesting to see such a dip in numbers, but we can see a direct connection with the admission rate graph, but it is starting slightly higher but still has the same decrease correlations with the admitted as expected. As stated earlier we see around 2 million applications with only around 220,000 averages of students being admitted but from 2000 we only saw 80,000 students being enrolled that makes a steady exponential climb up to around 100,000 in 2020. Only 100,000 students are enrolling into college with over 2,000,000

applications it's so interesting to see such a huge marginal difference because looking at the graph lines alone it seems to stay at a constant increase between Applications, admitted, and enrolled numbers but the actual number values of each is the value that most people look over as stated. This leads to the final graph showing the admission enrollment ratio percentage over the last 20 years as it has been steady decreasing from 2000 at a 30% mark to a 12.5% in 2020. With this we can see a linear comparison between the difference of tuition changes and the admission enrollment ratio. With the other graphs its difficult to see any real changes to make any conclusions on as many are close in correlation. We can see that this is the complete opposite as the constant change of exponential increase in tuition prices compared to the admission enrollment ratio taking a hard fall over the last 20 years.

Conclusion:

After reviewing and discussing the data findings and the graphs generated from the data to make an easier idea of correlations between data and linear regressions, I can conclude my hypothesis. My original hypothesis or question that I wanted to solve as stated earlier was that we would see a direct correlation that enrollment in colleges were decreasing due to a rise in tuition. As we can see within the results and graphs, that this has been proven true as tuition prices have been rising at a constant increasing rate over the last 20 years the admission enrollment ratio has been steadily decreasing from a 30% margin to a 12.5% in only 20 years. Amazing to show these findings

References:

Boyington, Briana Sarah Wood, et al. “See 20 Years of Tuition Growth at National Universities” *U.S News and World Report,* 17 Sept. 2021, https://www.usnews.com/education/best-colleges/paying-for-college/articles/2017-09-20/see-20-years-of-tuition-growth-at-national-universities.

Wikipedia. “College Admissions in the United States.” *Wikipedia*, Wikimedia Foundation, 7 Nov. 2021, https://en.wikipedia.org/wiki/College_admissions_in_the_United_States.

Acknowledgements

I would like to express my sincere gratitude to and thanks to Dr. Dara Jaiyeola for allowing me to conduct this research into a topic of my choosing. Allowing me to gain a broader understanding of my project topic. It allowed me to get an understanding of how data reports and findings will help me in my future endeavors.