Visual Analytics for Graduation trends using Power BI

Graduate Directed Study

Arun Aryal
California State University-Los Angeles
College of Business and Economics
aaryal@calstatela.edu

Tanmai Aurangabadkar
California State University-Los Angeles
College of Business and Economics
taurang@calstatela.edu

Era Singh Kajal
California State University-Los Angeles
College of Business and Economics
ekajal@calstatela.edu

Shwetha Seetharam
California State University-Los Angeles
College of Business and Economics
slnu2@calstatela.edu

***Abstract* - As data become more reachable, using huge amounts of available data to derive insights and make business results can be a task. Many big data technologies are used in organizations to perform analytics and acquire powerful acumens. Microsoft Power BI carries advanced analytics to the regular business decision methods, allowing users to extract useful information from data to unravel business problems. This paper shows analysis & does descriptive analytics on all universities in the United States based on diverse factors such as graduation rate, enrollment of graduates & undergraduates, degrees offered, financial aid, tuition fees, etc. This data is indicated in the form of visualizations to get valuable insights and develop future predictions about it.**

Keywords: Big data, analytics, Microsoft Power BI, dashboard, predictive analysis

I. INTRODUCTION

Most of the American Institutions that provide undergraduate or graduate education face the challenge for the type of applicants, they must make an offer for the admission. To maintain respectable performance concerning the number of students finishing their courses on time. Likewise, they need to forecast on how the recent intake of students might perform. Universities cannot lower the entry criteria too much, else the quality of education that they provide might get diluted. [5] In case of others, they must make offers to the candidates who are not only having appropriate profiles, but also who are the ones to accept the offers. On the other side, students have a task in deciding which schools they should apply, i.e. the universities that provide excellent education at a minimal cost for their profiles. Student profile can be determined by not only the performance in examination such as SAT and ACT but also on the data points based on the ethnicity, immigration status, gender etc.

IPEDS is the prime basis of information on Schools Universities, technical and vocational institutions in the United States. The dataset used from the Kaggle presents facts on the universities in the United States for the year 2013. It includes degrees offered, enrollment, graduation rate, fees and so on. Though the data being old, it can be explored using Business Intelligence tools called the Power BI to get significant insights about the universities in the United States.

Figure (a)

Visual analytics is beneficial in allowing universities to use Big data technologies to derive value from the various sources of data. Analytics has the potential to absolutely impact all the major areas that are of importance for an institution of learning; in areas such as student enrollment, information management, and reporting, and research. Through analytics, learning colleges can perform thorough analyses of student and learning data to make informed decisions on future course offerings to cater to the requirements of probable and existing students.

The notion of Business Intelligence has been around in innumerable forms for more than a century and a half. At its core, BI is about understanding the facts - and the relationship between facts - in a way that guides decision-making and action. Power BI can be easily integrated with numerous Big

Data sources to assemble and analyze massive amounts of data. Data transformations are decided by the users and it is used to remove faults and redundant data, correct formatting, and prepare data for further analysis by making them into suitable normalized forms, and so on. Power Query supports in enhancing the BI experience through its ability to extract the data in Microsoft Excel and further utilize it and break down as needed. Using the concept of this Power Query, filtering the data includes the relevant bits that enable one to focus only on the data that matters. The main purpose of this research is to figure out how to develop analysis using Power BI for the above-mentioned dataset. It does help in answering different research questions such as (a) Which county has the highest number of enrollments? (b) Which can be the highest state having total price for out-of-state students living on campus in the year 2013-14? (c) Find the top 10 states having the highest number of applicants with total enrollment based on the degree of urbanization.

II. RESEARCH BACKGROUND

The use of Big data and Visual Analytics to improve graduation rates in all the universities of United States:

Whenever we browse the Web, some sophisticated algorithms are tracking where we go, comparing us with the other millions of people. They are trying to predict what are you likely to do. At least, 45 % of the universities report that they are trying to improve their graduation rate by using the new technology based on the number of enrollments. This highlights on the predictive analytic concept. According to the survey, the latest figures for the enrollment in the year of 2017 is shown as 20.4 million and it increases by 5.1 million from the year 2000. But data remains to show that a lot of students attending four-year institutions will not graduate within the time frame. [5] It can be also seen that only 38 percent of students at a four-year flagship institution complete a bachelor's degree on time. On an average, it takes a student six years to complete, and academic institutions often quote their graduation rate using a six-year figure.

Institutions are increasingly using data to identify key areas where students are succeeding and struggling with the goal of completing a four-year degree. To identify the potential flaws, visual analytics uses the Power BI tool that provides several visualization features and predicts the problem and the time frame required to improve the student's performance.

Digital transformation is a multistep process. [6] Initial phases involve modernizing outdated practices like paper applications and automating and connecting manual-heavy processes across departments and systems. The advanced phase involves using data collected from the students to enhance their academic experience. Power BI enables data analytics to track a student's entire academic career from course completion to student activities. "A lot of the data is there, but not correlated, such as, how students are performing in extracurricular activities, how well they make progress in their program, what sort of upper division classes they take."

There are other challenges of Big data in education. There were many research papers conducted based on enhancing the graduation rates. The success was achieved by Georgia State University, Atlanta, where they built a predictive model, a system that uses Big data and Visual Analytics, they have named it has **GPS - Graduation and Progression Success.**

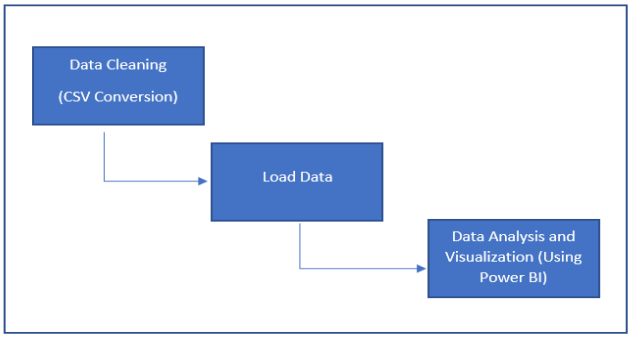
[7] Georgia State University is a large public university in Atlanta with more than 24,000 undergraduate students. Of those students, 60 percent are nonwhite, as with many public universities, resources for student counseling are limited. Large institutions incline to have staggering caseloads, a few years ago, GSU's ratio was 700 students per adviser. Researchers at GSU spent years examining student's grades, test scores and other information to identify those in potential distress and promptly assist them. The study shows that the number of students graduating has jumped by 30%.

Visual analytics are used in anticipating the retention and graduation rates. Many universities are adopting new data mining technologies with the use of visualization for recognizing the errors and connect with the students to rectify their mistakes.

[8] Wayne State University, which is in Detroit, serves more than 27,000 students. With the help of Visual Analytics to dig into massive amounts of data stored in Hadoop, the university scrutinizes details on enrollment, demographics, classes, students and more. Among the dashboards, the university makes the retention and graduation rates by different cohorts, admissions, peer comparisons and overall university enrollment. Administrators can see all these views on one screen and easily investigate variables. In just five months by means of Visual Analytics, Wayne State was able to reconstruct and modernize 80 percent of its old, non-penetrable reports (that previously had taken five years to create) into new interactive reports and dashboards.

III. RESEARCH METHODOLOGY

There is a rising inquisitiveness across higher education in the power of actionable insight resulting from data analytics. Academics and higher education administrators need access to data visions to drive computable performance improvements across a broad choice of tasks from basic reporting and accountability requirements for enhancing the student experience at all phases of the engagement lifecycle. [1]



Figure(b)

Visual analytics of unstructured data in Power BI deliver users to access their interest in the analysis; processing of the data into tables required by the visuals; bind a suitable table column

to visual fields; unites visuals into appropriate dashboard blends and filtering relationships for the analytic questions. [4] Power BI offers a simple, intuitive experience for interacting with data. From creating and sharing dashboards to exploring and enhancing reports which make it easy to engage with data from heterogeneous sources, fueling faster, more insightful business conclusions. With Power BI, one can get a rich, combined view of key information, no matter where all the underlying data is stored. [3] Architecture plays a vital role in the business intelligence operation which is represented in the above diagram. It acts as a framework that can be used to shape different sections together and set up standards and policies for a business intelligence solution. Dataset adopted from the Kaggle website (Source link) which is published by IPEDS comprises details on graduation rate, graduate & undergraduate enrollment, degrees offered, financial aid, tuition fees, etc.

Phase 1: Initially, unstructured data are observed from the dataset acquired from the Kaggle website. To convert from unstructured to structured data, the first stage is to perform data cleaning. Data cleaning is a process to detect and remove inaccurate records from the dataset which is then organized in a CSV format.

Phase 2: Once, data cleansing is done where structured data is achieved, the next stage is to load the data into business intelligence tool which can be further used for visualization. Microsoft Power BI acts as a business analysis tool which allows users to extract valuable facts from data to solve business problems.

Phase 3: The final stage is to complete data analysis and make visualizations with the help of the Power BI tool as it is an easy platform to deploy and manage. Data can be analyzed using Power BI features such as ArcGIS maps, Bar charts, Scatter plots, Line charts, Correlations, etc.

IV. RESULTS

Which county has the highest number of enrollments?

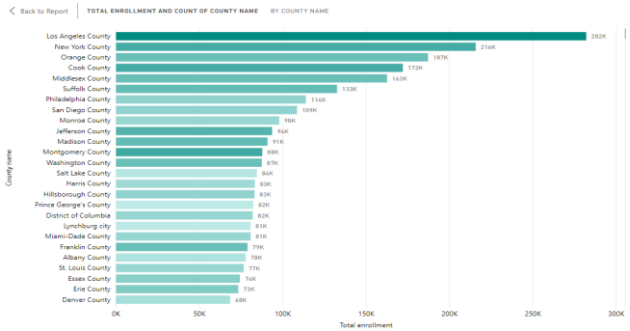


Figure (c)

In the above-clustered bar chart, total enrollment can be found using a calculated field based on county. A column called as the total enrollment is created and added for total undergraduate and graduate enrollment. This chart shows that which county has the highest number of enrollments for the year 2013. Analysis helps in finding out that the highest number of enrollments is 282000 for the county of Los Angeles and the least number of enrollments is 680000 for the county of Denver. Thus, one can arrive at a conclusion that Los Angeles County has the highest number of enrollments because of several factors such as job opportunities, weather, recreation etc.

Total Graduation among Top 10 universities

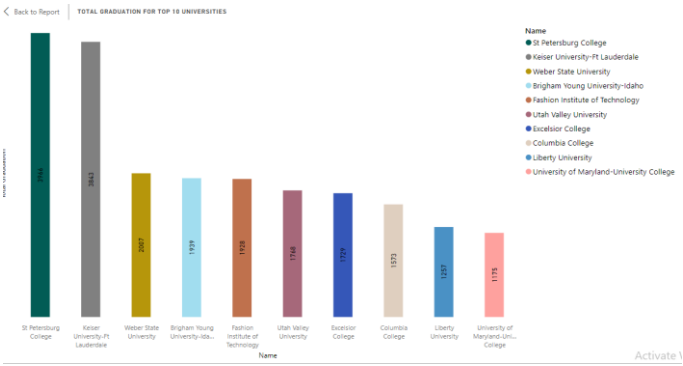
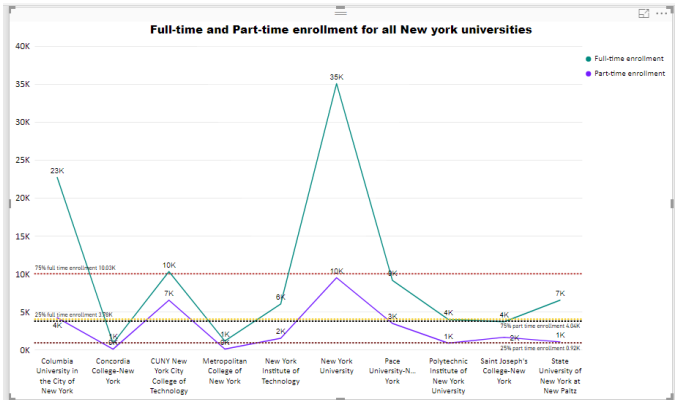


Figure (d)

Above Bar chart uses a calculated field method to compute the Total graduation for Post Baccalaureate and Post Master's degree with an associate degree. A single new column called as Total graduation can be calculated as "Total Graduation = Data [Number of students getting a Post-baccalaureate or Post-master's certificate] + Data [Number of students receiving an Associate degree]". Hence, one can get the total number of graduations for each university which can then be filtered out based on the University having criteria as top 10. The above visualization is the result of the top 10 universities. The highest Total graduation is 3966 for St. Petersburg College and the least is 1175 for University of Maryland-university College.

Determine analysis for Full-time and Part-time enrollment for all New York universities.

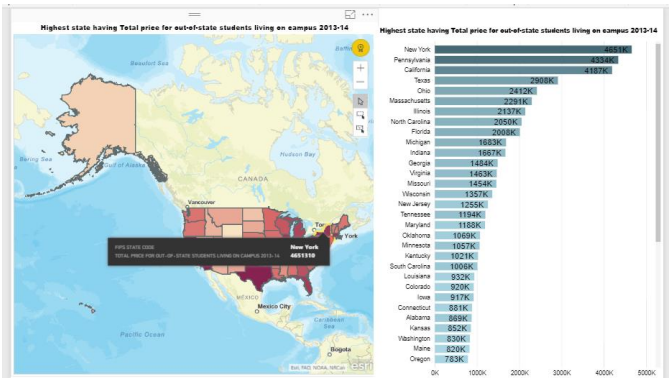


Figure(e)

From the above line chart, the legends have navy blue for full-time enrollment and purple for part-time enrollment of students in New York universities. As per the analysis, brown Percentile line depicts 25% part-time enrollment, which is about 0.92K and 75% part-time enrollment is around 4.04K which is denoted by the yellow percentile line. It is also seen that black percentile line depicts 25% full-time enrollment, which is about 3.78K and 75% full-time enrollment is around 10.03K which is represented by the red percentile line. For New York

University, full-time enrollment is 27% more than part-time enrollment. For Columbia University in the City of New York, full-time enrollment is 19 times more than part-time enrollment. It can be observed that for all New York universities 25% full-time enrollment closely matches with the 75% part-time enrollment.

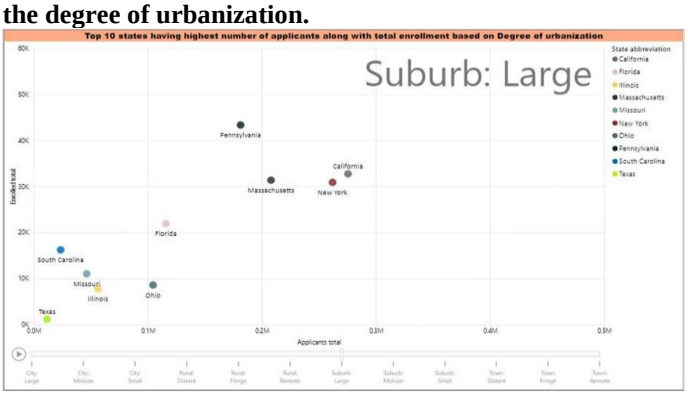
Which is the state having highest total price for out-of-state students living on campus in the year 2013-14?



Figure(f)

Above report has a combination of ArcGIS maps and bar chart which shows a list from highest to lowest total price for out-of-state students living on campus in 2013-14. As per bar chart, the highest total price for out-of-state students living on campus in 2013-14 is seen in New York which is about 4651K followed by Pennsylvania and California. And the lowest total price for out-of-state students living on campus in 2013-14 is seen in Wyoming which is 39K. Georgia, Virginia, and Missouri have closer total prices for out-of-state students living on campus in 2013-14. From ArcGIS maps, every state is highlighted based on different shades of pink were dark pink that represents the highest total price for out-of-state students living on campus in 2013-14 and light pink shows the lowest total price for out-of-state students living on campus in 2013-14. Using the Analytics feature in ArcGIS maps, the highest state having total price for out-of-state students living on campus in 2013-14 is pinned by locating on the map, which is shown by a red pointer which is seen for New York.

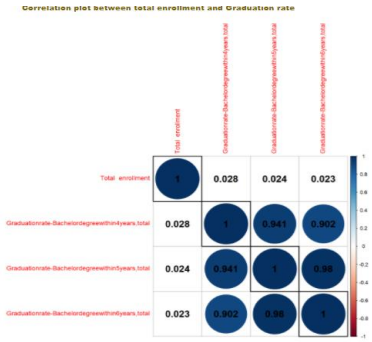
Perform analysis for the top 10 states having the highest number of applicants along with total enrollment based on the degree of urbanization.



Figure(g)

From the above scatter plot, legends have the top 10 states representing the highest number of applicants for every degree of urbanization. The visualization focuses on four degrees of urbanization such as city, rural, suburb and town. As per the plot, California is the city with the maximum number of applicants with the enrollment of 53848 whereas Florida is the city with a minimum number of applicants with 2503 enrollments in the large cities. Amongst the distant rural areas, New York ranks the top with 13511 number of total applicants with an enrollment rate of 1366. Also, in large suburb areas Texas is ranked the least with total applicants of 11577 and an enrollment rate of 1264. Moving towards the comparison of a distant town, New York is the state with the highest number of applicants of 91712 with an enrollment of 9315 and Illinois has the lowest number of applicants 44517 with 5797 enrolled applicants.

Plot correlation between total enrollment and graduation rate.



Figure(h)

Correlation Analysis is used for statistics, which quantifies the relation between two continuous variables. The correlation coefficient is an indicator which indicates the correlation between strength and direction. They have a probability, which shows the probability that the relationship between the two variables is equal to zero. The correlation coefficient can lie in between -1 and +1. The dark blue circles in a diagonal line from top left to bottom right show a correlation of an attribute with itself, which is always the strongest or 1. So this should not be read as correlation, but just as a separator line. The more the circle has a dark blue color, it signifies a stronger positive correlation. The darker the red color, it signifies a negative correlation. Lighter or white colors signifies weak or no correlation. The scale can be used to estimate the correlation coefficient value. The highest correlation is 0.98 which is between graduation rate bachelor's degree within 5 years and graduation rate bachelor's degree within 6 years. The lowest correlation is 0.023 which is in between total enrollment and graduation rate bachelor's degree within 6 years. In addition, there are four clusters in this plot which partitions the data points into groups based on their similarity.

CONCLUSION

Technology is revolutionizing the system of learning and how development practitioners do their work. Leveraging big data is the next logical step in this evolution. We now have access to volumes of data, but we must understand what it can tell us,

what it does tell us, and as importantly, what it can't capture. Colleges and universities need to meet a lot of challenges that have already started to impact the elite position that American higher education has enjoyed for decades. Big data has evolved to change the way of industries including the educational field. But the way of this new era has just begun and there are many difficulties such as the lack of experienced personnel on the science of big data and data analytics. Furthermore, the tutors and academics must be trained and involved in them and finally the students must accept and use these new tools. [1]

In the new era of big data, the traditional difficulties will no longer exist, keeping the best methods. The education system will be enriched with new learning ways, making more effective and targeted. To overcome these difficulties monitoring individual student performance could the best possible solution. Disaggregating student performance by selecting characteristics such as major, year of study, ethnicity, etc. and identifying outliers for early intervention could be other solutions. Predictive analytics can be performed by estimating the potential so that all students achieve optimal results, preventing attrition from a course or program and testing and evaluation of curriculum. Higher education administrators will do well by evaluating whether they can be used in their institutions and determining the role they can play. [2]

REFERENCES

[1] Drigas, Athanasios & Leliopoulos, Panagiotis. (2014). The Use of Big Data in Education. IJCSI International Journal of Computer Science Issues.11.58-63.

[2] Picciano, Anthony. (2014). Big Data and Learning Analytics in Blended Learning Environments: Benefits and Concerns. International Journal of Artificial Intelligence and Interactive Multimedia.2.35-43. 10.9781/ijimai.2014.275.

[3] https://powerbi.microsoft.com/en-us/documentation/powerbi-desktopgetting-started/

[4] Edge, Darren, Jonathan Larson, and Christopher White. "Bringing AI to BI: Enabling Visual Analytics of Unstructured Data in a Modern Business Intelligence Platform." *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018.

[5] Murumba, Julius, and Elyjoy Micheni. "Big data analytics in higher education: A Review." *International Journal of Engineering Science* 6.6 (2017): 14-21.

[6] Krishnan, Vijay. "Research Data Analysis with Power BI." (2017).

[7] Kamenetz, A. "How one university used big data to boost graduation rates. NPR." (2018).

[8] "Wayne State Univ. Enrolls SAS Analytics to Improve Retention, Graduation Rates." *Analytics Magazine*, 18 July 2016, analytics-magazine.org/wayne-state-univ-enrolls-sas-analytics-improve-retention-graduation-rates/.