Visual analytical trends using Power BI

Graduate Directed Study

Arun Aryal

California State University-Los Angeles

College of Business and Economics

aaryal@calstatela.edu

Era Singh Kajal

California State University-Los Angeles

College of Business and Economics

ekajal@calstatela.edu

Tanmai Aurangabadkar

California State University-Los Angeles

College of Business and Economics

taurang@calstatela.edu

Shwetha Seetharam

California State University-Los Angeles

College of Business and Economics

slnu2@calstatela.edu

Abstract **- As data becomes more reachable, using huge amounts of available data to drive 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 US universities dataset based on different 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 on 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 a challenge such as which type of students they will make an offer for admission to maintain respectable performance in terms of number of students finishing their courses in time. They also need to forecast how the recent batch of intake will perform. Universities cannot relax the entry criteria too much as that way the quality of education that they provide gets diluted. In case of others, they must make offers to the candidates who are not only having appropriate profiles but are also who are most likely to accept the offers. On the other side, students have a task in deciding which colleges they should apply i.e. the universities that provide best excellence in education at a minimal cost given their own profiles. Student profile ca be determined by not only the performance in examination such as SAT and ACT, but also other data points such as their ethnicity, immigration status, gender etc.

IPEDS is the prime foundation of information on US Colleges, Universities, technical and vocational institutions. This dataset has facts about the universities and colleges in United states for year 2013. It includes degrees offered, enrollment,

graduation and fees and so on. This dataset is explored using Power BI to get valuable insights about the universities in United States, though the data is old.
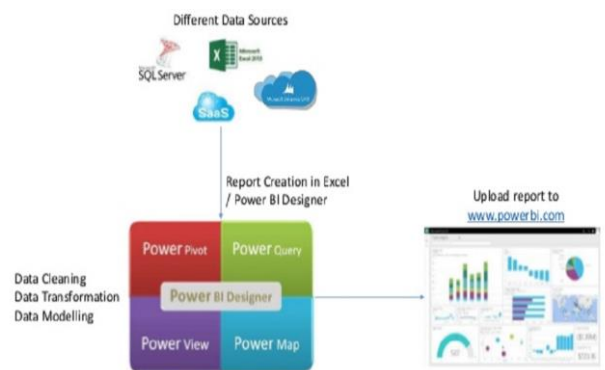

Figure (a)

Power BI can be easily integrated with various Big Data sources to collect and analyze huge amounts of data. Power Query helps in enhancing the BI experience through its ability to extract the data in Microsoft Excel and further utilize it and break down as needed. Data transformations are decided by the users and it is used to remove errors and redundant data, correct formatting, and prepare data for further analysis by organizing them into suitable normalized forms, and so on. Based on the report being developed, filtering the data to only include the relevant bits that enables one to focus on only the data that matters. Upon further study on the dataset we were able to discover some interesting facts like religious affiliation to the university, types of degrees offered, minority and majority community students studying, applicant information etc.

II. RESEARCH BACKGROUND

The need for Big data in Academics:

Big data analytics is actively being used in various business organizations in the field of business intelligence, financial

forecasting, and banking. However, in the recent years, big data has been widely used in academia. Many academic institutions are moving to cloud architectures and with help of these ecosystems are leading to a situation where more data is being gathered in these institutions than ever before, creating significant opportunities for using Big Data to analyze and correlate information that enhances decision-making. Educators are using big data results to find flaws in the education system to improve the teaching methods and standards and to gain a better understanding of student life. The application of Big data in education has a paved path for technological advancement like Hadoop, cloud services, Data Warehouse.

Big Data challenges in Educational area:

Analytics is a growing interest in the success of higher education. Most of the educational Institutions are believing in its importance for their success. To improve the quality of learning outcomes, it is necessary that a large volume of data generated by educational systems should be analyzed cautiously and effectively walk through its appropriate responses to its new challenges. Roadblocks to Big Data in Education generally include lack of executive vision, Inability to associate important business problems with big data solutions, Users or executives rooted in an old technology, cost, lack of data warehouses and analytical tools, data quality issues and simply leaving most facts to be uncollected leading to no analyses. Observed that widespread analytics use is limited mainly to the areas of enrollment management, student progress, and resource optimization. This limited scope of analytics is mainly due to barriers of affordability, data, culture, ability, and communication.

There are other challenges of big data in education.

  i. **Transparency**: Learners have the right to know where the gathered data are used.
 ii. **Privacy**: Who gets to see all the data or parts of the data. Levels of privacy must be assigned.
iii. **Presenting data**: There is a need to adopt a strategic approach to present the data. How do we display and describe the data that makes learners understand?
 iv. **Infrastructure**: Education institutions must upgrade to the latest technological developments to use big data techniques

III. RESEARCH METHODOLOGY

There are two purposes for this research. The first one is to understand what the framework of Power BI is and how it works. And the second purpose is to figure out how to create the analyzes by using Power BI using the dataset. In this paper, various research questions will be answered such as: (a) Which county have the highest number of enrollments? (b) Which is the highest state having total price for out-of-state students living on campus in year 2013-14? (b) What will be a likely graduation rate? (c) Find the top 10 states having highest number of applicants with total enrollment based on degree of urbanization.
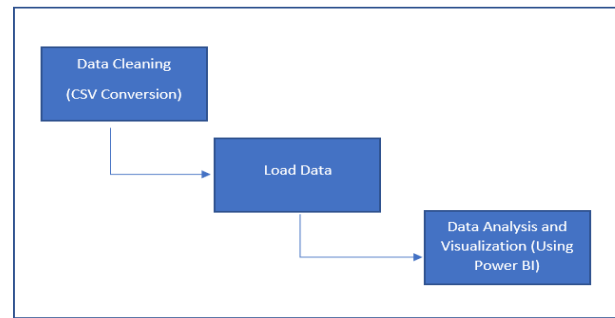
Figure(b)

BI architecture is the framework used to organize different components together and set up the standards and policies for a business intelligence solution. Architecture plays a vital role in business intelligence operation which is shown in the above figure. Dataset used in this research is published by IPEDS which contains details such as graduation rate, graduate & undergraduate enrollment, degrees offered, financial aid, tuition fees, etc. As per the architectural diagram, data cleaning is the first process to detect and remove inaccurate records from the dataset which is then organized in a CSV format. Once, data cleansing is done next step is to load data into the business intelligence tool which can be used further for visualization. Power BI is a of business analytical tool allowing users to extract valuable facts from data to solve business problems. Final step is to perform data analysis and make visualizations with the help of Power BI as it is easy platform to deploy and manage.

IV. RESULTS

Data Model representation:

Below figure is the data model representation which shows relationships between two tables: Students and Institutions. Every table has their own attributes and their relationships are shown through common attributes.

As per this data model, 3 relationships can be derived.

  a) Firstly, one student living on campus (for year 2013-14) can pay the total price for only one institution at a time. Single institution can have total price (for year 2013-14) for one student living on campus.
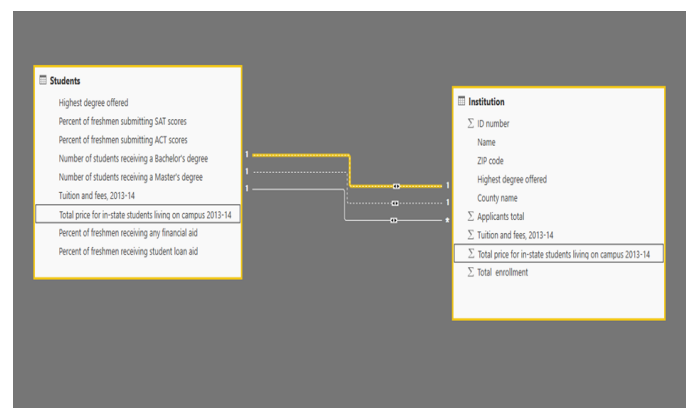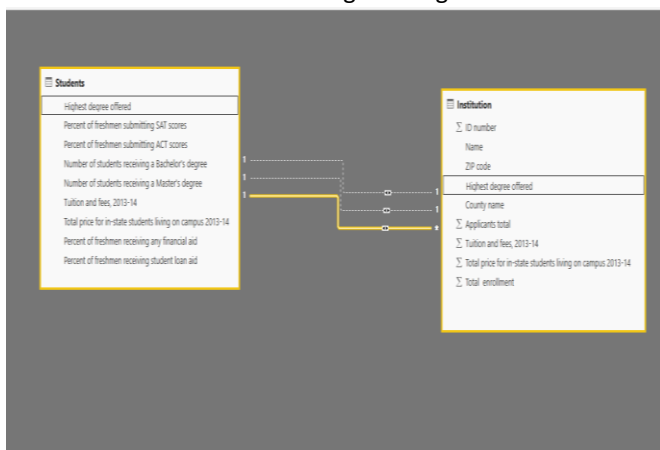

Figure(c)

b) Secondly, one student (for year 2013-14) can pay tuition fees for only one institution at a time. Sole institution can have tuition fees (for year 2013-14) for one student.
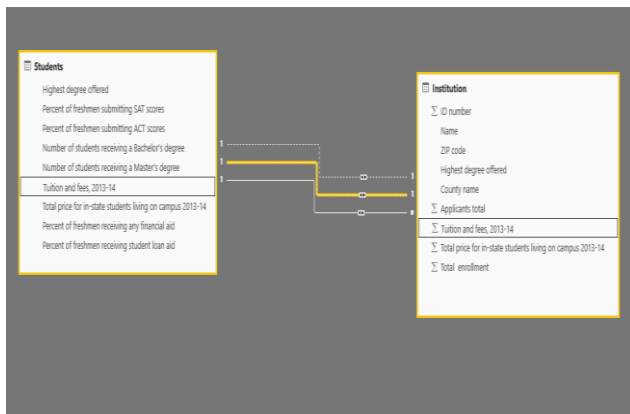
Figure(d)

c) Thirdly, one student can take up many highest offered degrees from any institution. And single institution can offer highest degrees to one student.

Figure(e)

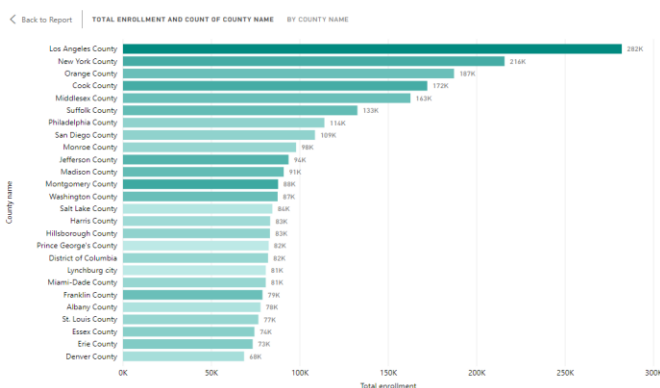**Which county have the highest number of enrollments?**

Figure (f)

In the above clustered bar chart, total enrollment can be found using calculated field based on county. I have created a column called total enrollment and added total undergraduate enrollment and total graduate enrollment. This chart shows us that which county has the highest number of enrollments for the year 2013 which happens to be the default year in our dataset. From this analysis we can find out that the highest number of enrollment is 282000 in the county of Los Angeles the least number of enrollment is 680000 in the county of Denver. We can arrive to a conclusion that Los Angeles county has highest number of enrollments is due to various factors such as job opportunities, weather, recreation etc.

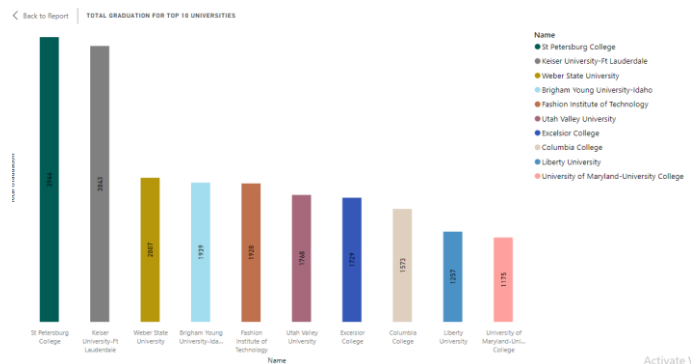**Total Graduation among Top 10 universities**

Figure (g)

In this we have used the Calculated field method to calculate the Total graduation for Post Baccalaureate and post master's degree with associate degree. We created a new column called Total graduation "Total Graduation = Data [Number of students receiving a Postbaccalaureate or Post-master's certificate] + Data [Number of students receiving an Associate degree]" hence we got the total number of graduation for each university and then we filtered the University by the 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**.
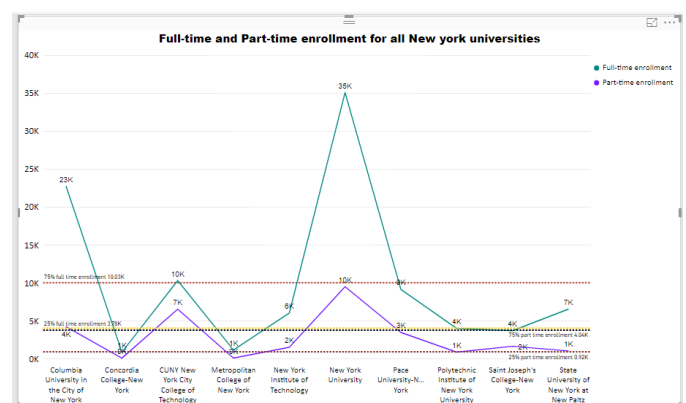
Figure(h)

From the above line chart, legends have navy blue for full time enrollment and purple for part time enrollment of students in New York universities. As per this line chart, brown Percentile

line depicts 25% part time enrollment which is about 0.92K and 75% part time enrollment is around 4.04K which is represented by 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 red percentile line. For New York University, full time enrollment is approximately 27% more than part time enrollment. For Columbia University in the City of New York, full time enrollment is approximately 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 highest state having total price for out-of-state students living on campus in year 2013-14?
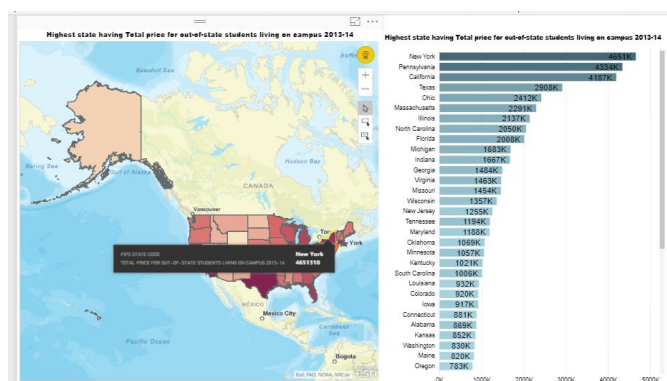
Figure(i)

Above report has combination of ArcGIS maps and Bar chart which shows listing from highest to lowest total price for out-of-state students living on campus in 2013-14. As per bar chart, 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 lowest total price for out-of-state students living on campus in 2013-14 is seen in Wyoming which is approximately 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 where dark pink represents highest total price for out-of-state students living on campus in 2013-14 and light pink shows lowest total price for out-of-state students living on campus in 2013-14. Using Analytics feature in ArcGIS maps, 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 red pointer which is seen for New York.

Determine the analysis for top 10 states having highest number of applicants along with total enrollment based on degree of urbanization.
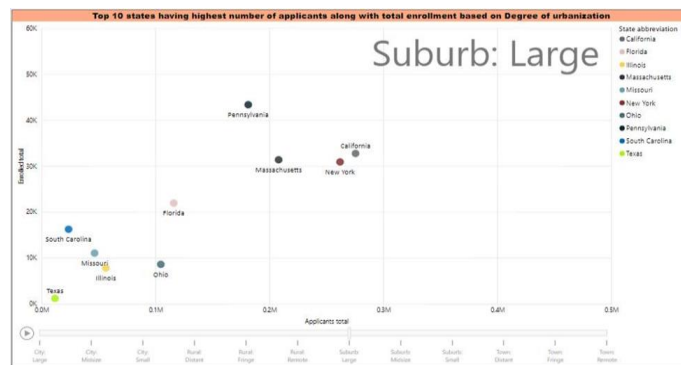
Figure(j)

From the above scatter plot, legends have top 10 states representing 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 scatter plot, California is the city with the maximum number of applicants with the enrollment of 53848 whereas Florida is the city with 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 enrollment of 1366. Also, in large suburb areas Texas ranks the least with total applicants of 11577 and enrollment rate of 1264. Moving towards the comparison of distant town, New York is the state with highest number of applicants of 91712 with enrollment of 9315 and Illinois has the lowest number of applicants 44517 with 5797 enrolled applicants.

Plot correlation between total enrollment and graduation rate.
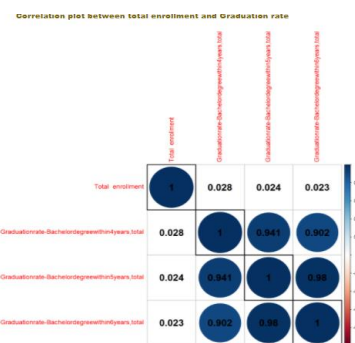
Figure(k)

Correlation Analysis is used for statistics which quantifies the relation between two continuous variables. The correlation coefficient is an indicator which indicates the correlation 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 shows 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 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 way learning and 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. Based on our research, Los Angeles county has the highest number of enrollments. Total graduation is seen highest for St. Petersburg College whereas the least is seen for University of Maryland-university College. For New York University, full time enrollment is approximately 27% more than part time enrollment. New York is the state having highest total price for out-of-state students living on campus in year 2013-14. Our study also indicates that New York is the state with highest number of applicants enrolled and Illinois has the lowest number of applicants enrolled.

REFERENCES

[1] http://www.ijcsi.org/papers/IJCSI-11-5-1-58-63.pdf
[2] https://files.eric.ed.gov/fulltext/EJ982669.pdf
[3] http://www.theijes.com/papers/vol6-issue6/Version-2/C0606021421.pdf
[4] http://ir.inflibnet.ac.in:8080/ir/bitstream/1944/2116/1/24.pdf
[5] https://www.irjet.net/archives/V4/i4/IRJET-V4I4721.pdf
[6] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.922.3825&rep=rep1&type=pdf