G Summit

Technical Report

Pat Badou Yosief Marshal Matthew Ray

PROJECT #2 VISUALIZE ME!

INTRODUCTION

G Summit has expanded its ETL best practices and now provides Data Analytics and Visualizations. G Summit is utilizing the latest research techniques and algorithms in Data Science, with interactive presentations of data insights for prescient visuals that will aide in research and decision making, telling a story straight from the data.

SCOPE

G Summit worked with the Dr-A focus group to identify themes for today's story telling. Extensive research was executed for each theme that came up, and it yielded potential data sources that were evaluated by the G Summit data experts. The chosen theme was churned through the G Summit Analytics and Visualization Insights (SAVI) to present its story, interactively.

RESEARCH

LIFE EXPECTANCY was at the top of the focus list. **Roadblock:** Research revealed data damaged, with similar feedback from earlier data diggers.

OPIOID was an exciting theme. From Kaggle, it brought us to Washington Post for several articles and documents. It provided a robust and heavy unraveling of how Opioid impacted each state or local town. Washington Post secured access to the DEA database tracking the drugs, from manufacturer, to distribution, to pharmacies, with an available API to DEA. **Roadblock:** data volume was in terabytes. This theme was deferred for the upcoming release (class) on Big Data and partnering with Google to leverage their claim on using qubits in quantum computing.

GRADUATE SCHOOL was evaluated next. A web scraper code was originally used to capture relevant data into the dataset used for the visualizations. This presented a useful story to tell for a special group of students at the UNCC Data Analytics Bootcamp. Initial focus was for the top 3 graduate programs, and corresponding related programs, with an interactive visualization to tell its story. It included currency calculations to present corresponding average tuition costs, and tag countries where programs are offered.

EXTRACT, TRANSFORM and LOAD (ETL)

G Summit applied its expertise with the ETL process that is efficiently carried out by cleaning, filtering, and aggregating data found in the csv files obtained from Kaggle. A relational database was used (PostgreSQL) to load the data for the next exciting stage

VISUALIZATION

Top 3 graduate programs were identified using SQL, followed by an extract of related graduate programs. These were both presented in a pie chart for visualization, side by side for comparison.

Average Tuition Cost for each of the Top 3 programs were calculated. Native currencies were kept as is to preserve the global characteristic of the data. In addition, base currency buttons were presented to convert each to the chosen base currency. As a result, it presented the graph in one comparable (base) currency for comparative analysis. Base Currency allowed is either US Dollars or Euros.

Top 7 countries were identified and presented on a stacked bar chart for the Top 3 programs. An additional feature allows each Top 3 program to be marked on a map for visuals.

SUMMARY and FINDINGS

The chosen theme on Graduate Schools shows that the Top 3 graduate disciplines share an approximate equal footing, with Computer Science getting a larger share.

However, once the related programs from each are included, both Business Administration and Economics graduate programs took a larger share of the pie.

Average Tuition Cost is larger for an MS in Economics except in Australia where MBA is more expensive.