1

**Department of Information Technology**

**IT 7113 Data Visualization Development Project**

**Project Details:**

Finding meaningful Insights from Georgia Schools Data for Enrollment, School Climate, Diversity, and Milestones Test



**Executive Summary:**

Before the 17th century, data visualization was mostly limited to maps, which showcased land markers, cities, roads, and resources. As the demand grew for more accurate mapping and physical measurement, better visualizations were needed. Theme visualization first appeared in the 18th century. Near the end of the 21st century, attempts were made at the thematic mapping of geologic, economic, and medical data. At this time, abstract graphs of functions, measurement error, and empirical data collection were introduced.

The latter half of the 20th century is what friendly calls the ‘rebirth of data visualization’ with the emergence of computer processing. Computers made it all possible to process large amounts of data at lightning-fast speeds. Over the last three decades, the field of data visualization has exploded into dozens, if not hundreds, of focus areas. Dashboards, data discovery tools, and a variety of other software tools allow businesses, researchers, and individuals to explore their data in new and increasingly creative ways.

In this project, I will showcase a variety of data visualizations in multiple dashboards to get meaningful insights from Georgia Schools’ data.

**Background:**

Data visualization is the graphical representation of information and data. Data visualization tools, which use visual elements such as charts, graphs, and maps, provide an easy way to see and understand trends, outliers, and patterns in data. Data visualization tools and technologies are critical in the Big Data world for analyzing massive amounts of data and making data-driven decisions.

Our culture is visual, containing everything from art and advertising to television and movies. Data visualization is another form of visual art that captures our attention and keeps it there. When we look at a chart, we can quickly identify trends and outliers. If we can see something, we quickly identify it. The main goal of data visualization is to communicate information clearly and effectively through graphical means. So, if this is successfully achieved, the user can find out the major insights from the individual visualizations. In the business context, it will help the stakeholders and business owners to make proper decisions that profit the business. In the same way, my intention or my goal is to showcase the meaningful insights from Georgia school’s Data which is spread across different counties. The objective is to showcase the school’s data with various factors like Enrollment/Admissions, grades, percentages, and diversity/race with different visualizations and Dashboards. The other objective is to give a holistic picture of Georgia school data to the common audience so that they can get meaningful information out of it.

**Data Description:**

For this development, the data source I have chosen is related to Georgia schools’ data. The link for the data source is as follows: <https://schoolgrades.georgia.gov/dataset/school-level-data>

The data sources presented in the above links are in CSV format. There are a total of 6 six data sets present in the provided link. It consists of Georgia state schools’ information, and it is available from 2014 to 2019. Each data set holds various columns and huge numbers of records. The data set contains different columns such as School Id, School name, County name, Cluster, Grade, Total enrollment, Single score, School climate Percentage of students who are Asian, American Indian, Black, Hispanic, White, etc. Other important information like percentages of different subjects which are as part of milestones test.

The most important thing which I did as part of this project is data cleansing. All the unwanted rows and columns are removed. Only the columns that are required for the dashboard purpose are considered. Some of the news columns are also created as per the dashboard requirement. One more important column which has been created is the Region. All the counties which are present as part of Georgia school data are segregated concerning the region. This will help to filter out the counties concerning the region. Keeping the data consistency in mind the datasheets from 2016 to 2019 are considered as part of this project. Based on the requirement separate data sheets are prepared and used for building an individual dashboard. Excel is the input data format that is fed to the Tableau tool for visualization purposes.

**Dashboard Design:**

A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at one stretch.

Especially for this project, I have created four dashboards. All the dashboards give meaningful insights. Each Dashboard contains different visualizations and from each visualization, the end-user or audience can find useful and important information. The main objective of dashboard design is to showcase the data in a meaningful way to provide more clarity and reduce complexity.

Below are the steps to follow in designing a Dashboard:

* The first and starting step that needs to be followed in the case of dashboard creation is requirements gathering. We need to make sure that all the requirements are gathered correctly without any issues.
* After gathering requirements and analyzing them, the problem statement must be stated clearly. Without any proper clarity on this, it is very difficult to proceed further in dashboard creation.
* Once both above steps are completed the main and important steps are Data analysis followed by Data cleansing. This is all to make sure that data is cleaned by removing or eliminating unwanted rows and columns.
* Align and segregate the data that is only required for the dashboard design.
* Make proper visualizations or choose charts that are required for the dashboard.
* Finally create a dashboard by following the best practices.
* Proper filters, colors, and labeling are the minimum things that need to be followed and taken while creating a dashboard.

**Best Practices:**

A dashboard is successful when people can easily use it to derive answers. Even a beautiful dashboard with an interesting data source could be rendered useless if your audience can’t use it to discover insights. This can be improved by properly following the best practices. I always follow the below practices in my dashboard design.

* The main and starting step is anticipating the user’s needs. To be more precise what are his goals.
* The next step is to choose proper visualizations that make a difference if they are present in the dashboard. Displaying too much data with more visualizations becomes quite confusing to derive insights or data.
* We need to choose proper metrics that are required and only those need to be displayed in the dashboard or the individual visualizations. So, the lesser would be better as far as it serves user needs.
* The main interactivity for the user with the dashboard would happen if proper interactivity features were used while designing a dashboard. So, we need to make sure that most of the interactivity features are used while designing a dashboard.
* Finally, a complete picture of the dashboard needs to be shown on a single screen with a proper and meaningful dashboard name.

By implementing effective dashboard design practices, I can ensure that users can quickly identify critical information.

**Dashboard-1:**

Graphical user interface, chart, application

Description automatically generated **Georgia Schools Enrollment**

**Purpose of the Dashboard-1:**

The main purpose of the above dashboard is to give information about Georgia school enrollment. It consists of different visualizations in it. Each visualization will give a unique insight. It has info related to Total Enrollments, Enrollment count by region and counties, Enrollment by Region, Enrollment by cluster, and Enrollment for each grade from 2016 to 2019. It gives information about how the variation in Enrollment count for different years and different counties. With this, they can come to conclusions like which county or which region has low or high enrollments. So, the schools in that county can make certain decisions like how they can make an increasing enrollment. The above dashboard consists of a bar chart, bubble chart, pie chart, map, and line chart.

**Contents of Dashboard-1:**

Graphical user interface, text, application

Description automatically generated **Bar chart: Schools Enrollment by cluster**

* The above is the bar chart that gives the school’s enrollment by cluster.
* Here the highest enrollment is for Cluster E and the lowest enrollment is for EMH.
* On the right-hand side there is a filter for the cluster. We can also select only the required clusters by checking the checkboxes and by removing the other ones.

Graphical user interface, map

Description automatically generated **Map: Schools Enrollment count by Region and Counties**

* The above is the Map that gives the school’s Enrollment count by Region and Counties.
* In the above map we can see the highest enrollments for the Metro-Atlanta region whereas the lowest enrollments are for the South region.
* If we see concerning counties the highest enrollment is for Gwinnett County from the Metro-Atlanta region and the lowest enrollment is for clay county from the west region.
* On the right-hand side we can see the filter for the region with different colors.
* The other filter which we can see is the size of the enrollment.

Graphical user interface, chart, application, pie chart

Description automatically generated

**Pie chart: Schools Enrollment count by Grade**

* The above is the pie chart that gives the school’s Enrollment count by Grade.
* Here is the Grade C with the highest enrollments and Grade F has the minimum enrollments.
* On the right-hand side we can see the filter with different colors for Grade.

Chart, bubble chart

Description automatically generated

**Bubble chart: Schools Enrollment count by Region**

* The above is the bubble chart that gives the information about Enrollment concerning the region.
* The highest enrollment here is shown in the big bubble size and the lowest is in the small bubble size.
* The highest enrollment is for the Metro-Atlanta region and the lowest enrollment is for the South region.
* On the right-hand side we can see the filter for the size of the enrollment.

Graphical user interface, text, application, email

Description automatically generated **Total Enrollment Count**

* The above is the small KPI that gives the Total Enrollment count value.
* This is very helpful in the Dashboard which gives the enrollment count for an individual year.

Chart, line chart

Description automatically generated **Line chart: Schools Enrollment for each grade from 2016 -to 2019**

* The above is the line chart which gives the information about the Enrollment count for each grade for 4 different years i.e; 2016,17,18,19.
* Overall, the highest enrollment is in 2019 for grade C.
* The lowest is for grade F in the year 2019.
* The trends in enrollment for Grade A and Grade B are quite similar.
* Only grade C has a consistent increase in enrollment.
* Only for grade D and grade F there is a decrease in enrollment count from 2018 to 2019.

**Usage of Dashboard-1:**

The above dashboard gives the information about Georgia school’s enrollment. I have used the interactivity features for the above dashboard. One can get the stats of the respective item by clicking on the required visualization. The other option is the usage of a region filter. By using this one can get the stats for the individual region for different visualizations present in the dashboard.

**Dashboard-2:**

Graphical user interface, chart, application

Description automatically generated **Georgia Schools Climate**

**Purpose of Dashboard-2:**

The main purpose of the above dashboard is to give information about Georgia school’s climate. It consists of different visualizations in it. Each visualization will give a unique insight. It has info related to Avg school climate by county, Avg school climate by region, Avg school climate score and enrollment count for each grade, Avg school climate, and Avg school enrollment by region. It gives information about Avg school climate for individual counties and different regions. It explains how the climate score varies for different grades. Understanding all these visualizations provides information and aids decision-making to improve the climate score for the respective county or region. The above dashboard consists of a map, dual-axis chart, treemap, and butterfly chart.

**Contents of Dashboard-2:**

Graphical user interface, application

Description automatically generated **Map: Avg school climate by County**

* The above is the map that gives the information about Avg school climate by county.
* It shows the Avg school climate for different counties.
* Out of which the counties with Avg school climate of (5.0) are Monroe County, Glasscock County, and Pierce County.
* The county with Avg school climate of (1.667) in Macon County. These are the counties with the Maximum and Minimum ranges of Avg school climate.
* On the right-hand side, there is a filter for counties and another filter for the size of Avg school climate.

Chart, treemap chart

Description automatically generated **Treemap: Avg School climate by region**

* The above is the treemap which gives the information about Avg school climate by region.
* The highest avg school climate is for the Northeast region.
* The lowest avg school climate is for the west region.
* On the right-hand side there are two filters. One filter is for the Enrollment year.
* The Avg school climate can be seen for 2 different years 2018 and 2019.
* The other filter is for the size of Avg school climate.

Chart, bar chart

Description automatically generated **Dual-axis chart: Avg school climate score and enrollment count for each grade**

* The above is the dual axis chart which gives the information about Avg school climate score and enrollment count for each grade.
* On the right-hand axis it is showing is Avg school climate and on the left-hand side, the axis is the enrollment count.
* The x-axis it is shows the value of the grade.
* Grade F has the lowest enrollment count and Grade C has the highest enrollment count.
* When it comes to Avg school climate Grade A is the highest one and Grade F is the lowest one.
* The bars here represent the enrollment count, and the line represents the Avg school climate.
* On the right-hand side there is a filter with different colors which tells about Avg school climate and enrollment count.

Chart, bar chart

Description automatically generated **Butterfly chart: Avg school climate and Avg Enrollment count by region**

* The above is the butterfly chart which gives the information about Avg school climate and Avg Enrollment count by region.
* On the left-hand side it shows the Avg school climate score and on the right-hand side, it shows the Avg enrollment count.
* So, for both sides, the common attribute here is the region. So, we can check Avg school climate and Avg enrollment count concerning the region.
* The Highest avg school climate is for the Northeast region and the lowest is for the west region.
* The Highest avg enrollment count is for the Metro-Atlanta region and the lowest is for the west region.
* One more insight is that the west region has both Avg school climate and Avg enrollment as the lowest.

**Usage of Dashboard-2:**

The above dashboard gives information about Georgia’s school’s climate. I have used the interactivity features for the above dashboard. One can get the stats of the respective item by clicking on the required visualization. The other option is the usage of filters. By using a region filter one can get the stats for the individual region for different visualizations present in the dashboard. With an enrollment year filter, one can get the details for only the selected year.

**Dashboard-3:**

Graphical user interface, table

Description automatically generated **Georgia Schools student Diversity**

**Purpose of the Dashboard-3:**

The main purpose of the above dashboard is to give information about Georgia school’s student diversity. It consists of different visualizations in it. Each visualization will give a unique insight. It has info related to student count of different diversities. It also gives information on the top 5 schools with the count of Asian, black, and white students. The other insight which it gives is the count of different diversity of students concerning the region. With this, they can come to conclusions like which region and which school are having a specific count of students white, black, Asian. Hispanic, native, and multi. The above dashboard consists of a bar chart, text table, and KPIs.

**Contents of Dashboard-3:**

Graphical user interface, text, Word

Description automatically generated with medium confidence

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated with medium confidence

Graphical user interface, text, Word

Description automatically generated with medium confidence

Graphical user interface, text, application, Word

Description automatically generated with medium confidence

* The above all are the KPIs that gives the count of different diversity of students present in Georgia schools.
* With this the user can get a clear count of different diversity of students.
* The different diversities are White, Black, Hispanic, Asian, Native, and multi.
* These KPIs have great importance once they come and sit in the dashboard.

Text

Description automatically generated

Graphical user interface, text

Description automatically generated

Graphical user interface, text

Description automatically generated **Bar chart: Top 5 schools with students Count for Asian Black, and White**

* The above are the bar charts that give the information for the top 5 schools with students of Asian, Black, and White.
* When comes to the first visualization it gives the information on the top 5 schools with the highest count of Asians. Northview High school is the top school having students of Asian.
* The second visualization gives the information of the top 5 schools with the highest count of Black. Peachtree Ridge High school is the top school having students of black.
* The last visualization gives the information of the top 5 schools with the highest count of White. South Forsyth High school is the top school having students of White.

Graphical user interface

Description automatically generated **Text table: Students counted by Diversity and Region**

* The above is the text table that gives the information of students counted by diversity and Region.
* There are a total of 6 different diversity groups and 12 regions.
* The text table gives the full picture of the different diversity of students present across the 12 regions of Georgia state.
* As you can see from the text table white students are larger in number and they have spread over all regions. When comes to the lowest it is the native type.
* Likewise, we can find see which region has which type of students as higher in number.

**Usage of Dashboard-3:**

The above dashboard gives information about Georgia School’s Student Diversity. I have used the interactivity features for the above dashboard. One can get the stats of the respective item by clicking on the required visualization. The other option is the usage of filters. By using the Enrollment filter one can get the stats for the selected year for different visualizations present in the dashboard. Here the KPIs give the more important insights into where the user can get the student’s count of respective diversity.

**Dashboard-4:**

Graphical user interface, chart

Description automatically generated

**Georgia student’s Milestone test**

**Purpose of the Dashboard-4:**

The main purpose of the above dashboard is to give information about Georgia school milestone tests. It consists of different visualizations in it. Each visualization will give a unique insight. It has information related to Georgia school students’ milestone tests. This milestone test comprises subjects like English, Maths, Science, and Social. Besides this will also give the student’s Avg test percentage by county and students’ Avg test percentage concerning different categories. Understanding all these visualizations provides information and aids decision-making to improve the student’s percentage in milestones tests. They can also concentrate on individual counties who are having low avg test percentages. The above dashboard consists of a map, bubble chart, and bar chart.

**Contents of Dashboard-4:**

Chart, bubble chart

Description automatically generated **Bubble chart: Student count by School level**

* The above is the bubble chart which gives the information about student count for different school levels.
* It has a total of three different school levels i.e., Elementary, High, and Middle.
* The highest student count is for the Elementary level whereas the lowest is in the middle level.
* The count with the highest has the bigger bubble size.
* On the right-hand side, it has the student count filter where we can see the student count for 2 different years i.e., 2018 and 2019.

Graphical user interface

Description automatically generated **Map: Student’s Avg test pct by county**

* The above map gives information about the student’s average test percentages by county.
* The map consists of Avg test percentage for different counties.
* All the county’s Avg test percentages range between 24.99375 to 25.00741.
* On the right-hand side there is a filter with a list of county names and the other one is the range for Avg test percentage.

Graphical user interface, chart, bar chart

Description automatically generated **Bar chart (with categories): Student’s Avg test pct by diff categories**

* The above is the bar chart which gives the information about students’ average test percentages concerning different categories.
* The different categories are School level, Test name, and Proficiency.
* Here from the above bar chart, we can see a complete picture of Avg test percentages.
* The highest Avg test percentage is at the elementary school level for test name social with proficiency level as developing.
* The lowest Avg test percentage is at the high school level for test name math social with proficiency level as distinguished.

**Usage of Dashboard-4:**

The above dashboard gives information about Georgia’s student’s milestone tests. I have used the interactivity features for the above dashboard. One can get the stats of the respective item by clicking on the required visualization. The other option is the usage of filters. By using the Test name filter one can get the stats for the individual subjects from the different visualizations present in the dashboard. One of the visualizations i.e.; a Bubble chart will also give the count of students at different school levels for the selected test names. With an enrollment year filter, one can get the details for only the selected year.

**Conclusion:**

From all the above dashboards a user can get meaningful insights from it. Each dashboard gives important and useful information. In a similar way, we can create a multiple number of dashboards in tableau. The only thing is we need to follow proper design steps and the best practices.

**Learnings from this project:**

I learned so many things while completing this course. To start with I got a clear picture of how data is being handled while designing dashboards and getting meaningful insights from it. The biggest and most important thing which I learned was understanding the data and using only the necessary parameters that are required for the project. Once the data is ready then analyze the best visualizations for the available data. This I learned with enough practice. Finally following best practices while designing the dashboard. These are all things I learned in a day-by-day and weekly process by completing lab modules on my own.

**Acknowledgment:**

Today I can design a Tableau dashboard by following all the best practices. I believe this learning will be helpful for my future course i.e; Power BI. I want to thank Professor Dr. Jack Zheng for teaching this course to all students by providing useful references, and important documents. This all made to improve my knowledge of Data visualization.

**References:**

* Data and Information visualization: <https://en.wikipedia.org/wiki/Data_and_information_visualization>
* History of Data visualization:
* <https://www.dundas.com/resources/blogs/introduction-to-business-intelligence/brief-history-data-visualization>
* <https://www.sas.com/en_us/insights/big-data/data-visualization.html>
* What is Data Visualization: <https://www.tableau.com/learn/articles/data-visualization>
* Dashboard Design best practices:

<https://www.toptal.com/designers/data-visualization/dashboard-design-best-practices>