

Everything Data

[Aaronea Wiggins, Lojain Idris](#)

CTEC 298-101

Hello all,

My name is Aaronea Wiggins, and my name is Lojain Idris, and we are currently participants within the course CTEC 298, Spring 2025-week seven semester. Today, this group will happily share summaries and descriptions of CTEC 128 and CTEC 298 tasks. This paper includes our individual submissions, as well as our collaborative efforts to complete the final group project.

Summaries of CTEC 128 Papers

Aaronea Wiggins: “The overall viewpoint of my CTEC 128 report was how Black Americans, specifically men, are racially profiled, considered the most dangerous race, and falsely connected to the most crimes. With the raw dataset provided via FBI website, my group at the time and I, had the ability to visually prove that Black Americans are not the most dangerous race. Though, in fact, White people have committed the most crimes.”

Lojain Idris: “In summary of my CTEC 128 paper report, I explore the data about the cyber-attacks on global businesses, using reliable source from Kaggle “Top Data breaches (2004-2021)” dataset. With this data I focused on understanding how often cyber-attacks occur each year from 2004 to 2021, what type of attacks are most common. And how different organization types are affected. Also focuses on visualizing the data with pivot table to compare attack methods across organization type over time (2004-2021). This visualization helped analyzed that most global businesses have poor security and that common methods are through sharing too much information or social engineering.”

Description of CTEC 298 Material Submitted

Aaronea Wiggins: “During my time within this course, I have been tasked to connect to my previous drive of a different course and retrieve datasets from that course to successfully create my work for the class. The course in question--CTEC 128. Within CTEC 298, I have been introduced to many data related tutorials and tasks. A brief review of the tutorials, visualizations, and repositories are Matplotlib, Jupyter Notebook, Tableau, and GitHub.

- Jupyter Notebook is an interactive notebook which integrates code, visualizations, equations, and more. File extension. ipynb
- Matplotlib: A comprehensive library for creating static, animated, and interactive visualizations in Python
- Tableau: software for data input, visual analytical tool to help you understand data.
- GitHub: considerable repository to manage and share code.

With that said, I have imported and uploaded several files within my repositories and data software to create plots and python code utilizing CTEC 128's data."

Lojain Idris: "I completed LearnPython tutorials part1, LearnPython Tutorials part2, Anaconda Installation completion, GitHub Account completion, Matplotlib tutorials, matplotlib python plotting, Panda's tutorials, NumPy assignment, Tableau tutorials, Jupyter python code prove of completion. LearnPython Tutorials 1 focused on basic learning such as printing hello World, variables and types, lists, basic operators, string formatting, basic string operations. LearnPython Tutorial Part Two focuses on basic learning too such as conditions, loops, functions, classes and objects, dictionaries, modules, and packages. Anaconda installation completion focused on completing anaconda 3(python) on the computer and launching Jupyter Notebook. GitHub account set up focused on setting up a GitHub account and creating repositories which also include all CTEC 298 martials. Matplotlib Tutorials focused on installing matplotlib python libraries in Jupyter Notebook and creating one assigned plot which pie chart. Matplotlib Python plotting focused on creating six plots that were taught in class and creating based on my dataset I used in CTEC 128. Pandas Tutorials focused on using panda's python libraries in Jupyter Notebook for creating DataFrames and reading Excel file and wrangling my CTEC 128 dataset to clean and organized dataset. NumPy assignment focuses on using NumPy python libraries to numeric data and dealing with arrays operation. Tableau Tutorials focused on installing tableau for students and creating two plots of my choice which are packed bubble plot, horizontal bar using my CTEC 128 dataset file. Jupyter Notebook Python Code which focuses in writing all python code in Jupyter Notebook and uploading the completion."

Description of CTEC 298 Plot Deliverables

Aaronea Wiggins: "Of the six plots, I have been assigned the 'Histogram' plot. To convert statistics from number to plot to python code, a reference back to my CTEC 128 information was required. Once retrieved, I utilized the statistics and imported the information within a new excel workbook. Following the information import, I was able to create a new graph by selecting **insert** within the ribbon --> statistical graph **Histogram**. Automatically the information was graphed but not exactly how I would have

liked. With arrangements by adding a title, switching the race to x-axis, and switching the number of crimes to the y-axis—my plot was now complete. Next, I have been tasked with the creation of my stats within the newly downloaded software, Tableau. The class has been tasked to select two additional plots to visualize the CTEC 128 data.”

Lojain Idris: “I completed Six Matplotlib plots which are pie chart, histogram, Scatter Plot, Bar Graph, Area Plot, Hexagonal Bin Plot, and Two tableau plots which are Packed Bubble, Horizontal Bar. Pie chart was created to visualize the top five breaches by organization type. The plot uses the organization type columns from CTEC 128 dataset. By counting the five largest breaches by the organization type into each different labels each category represents a color and percentage of breaches by organization type. The plot was made using matplotlib and pandas. Histograms plot was created to visualize the distribution of records count. The plot uses records columns to count the number of records most breached and bars are represented in orange color and black edge color. Histograms plot was created but python matplotlib and pandas. Scatter Plot was created to visualize the breaches over time 2004-2021. The plot uses records and year columns. This plot shows the number of breaches changed over time of 2004-2021 and the dots are represented in yellow color. This plot was created using python matplotlib and pandas. Bar Graph was created to visualize the entities by records. The plot uses entity and records columns. This plot shows the top ten entities and how many records were breached, each entity or bar represented in color red. Bar Graph plot was created using python matplotlib and pandas. Area Plot was created to visualize the total of records breached per year of 2004-2021. This plot uses records and year columns. The plot calculates the breaches per year by grouping the records each year. Area Plot was created using python matplotlib and pandas. Hexagonal Bin Plot was created to visualize the count of effected record over time. This plot uses year and records columns to show the effected records through 2004-2021, also in this plot I fix my data because hexbin plot could not read the string in year which was 2018-2019 to fix that I convert the year values all to numeric values. Hexagonal Bin Plot was created using python matplotlib and pandas. Packed bubbles were created to visualize most methods to breach the data in plot. This plot uses records and methods columns; each bubble represents a method and size of the bubble represents the number of records that were affected using that method. Packed Bubble was created using tableau. The Horizontal Bar plot was created to visualize total records breached by method. This plot uses records and method columns; it displays the sum number of records compromised by each method. The length of the bar represents the number records was affected. Horizontal Bar Plot was created using tableau.”

Summary/Conclusion

In conclusion, this class has taught me punctuality, how to data—visually, via code, data handling, and teamwork. We have learned to work with data visually through coding. Although we work individually throughout assignments, we still make efforts to help each other to complete our assignments and the final project. This project was created through a lot of lessons and assignments to teach us teamwork. Throughout the project we gain knowledge and experience with python libraries, and learned coding, creativity, and data visualization. Completing this project has improved our understanding of python and data visualization.

References

Top Data Breaches (2004-2021)

<https://www.kaggle.com/datasets/hishaamarmghan/list-of-top-data-breaches-2004-2021>