

Bishop's University
Department of Computer Science

CS503 - Project

Presentation on Data Visualization Tools with Python Customization

Instructor: R. Hedjam

Objective:

For your upcoming project, the focus will be on **Data Visualization** using an **open-source tool**, similar to **Tableau**¹. The objective is not only to explore how these powerful tools can be employed to visualize real-world problems and complex datasets, but also to integrate **Python scripts** to **customize** and **enhance** the visualizations. By incorporating Python, you will be able to take full advantage of the flexibility offered by open-source platforms, allowing you to manipulate data, apply advanced analytics, and create more interactive and insightful visual representations that go beyond the default capabilities of the tool.

Task

Prepare an informative and engaging presentation

Key Steps for the Presentation:

1. Introduction to the Open-Source Tool:

- Choose an open-source tool like Apache Superset, Metabase, or Redash (you can also explore Plotly Dash if you'd prefer more Python-driven customization).
- Provide an overview of the tool, its features, and how it compares to commercial tools like **Tableau** in terms of accessibility, functionality, and integration capabilities.

2. Selecting a Real-Life Problem and Dataset:

- Pick a real-life problem that can be visualized using data. Some examples could include:
 - Sales data from a retail company
 - Climate change trends (temperature, CO2 levels over time)
 - Health data analysis (tracking diseases, healthcare trends)
 - Customer behavior data (website or app user analytics)
- Define what you want to visualize, such as trends, distributions, comparisons, or geographic data.

3. Using Python to Enhance Visualizations:

- Demonstrate how to integrate Python scripts with the open-source tool to customize the visualizations.
- For example, you can use Python libraries such as Pandas for data manipulation, and Plotly, Matplotlib, or Seaborn for creating or enhancing visualizations. Discuss how Python helps refine and improve your output.
- Customizations might include:

¹ <https://www.tableau.com/>

- i. Changing chart types or creating advanced visualizations (e.g., interactive graphs, heatmaps, or 3D plots).
- ii. Adding custom labels, annotations, or interactive features (such as hover effects).
- iii. Performing data preprocessing (e.g., filtering or cleaning data) using Python before feeding it into the visualization tool.

4. Case Study and Example:

- Walk through an example of how you can visualize a dataset using the chosen tool. For instance, if you're working with sales data from a retail company, you might use Apache Superset to create a dashboard that visualizes monthly sales across different regions.
- Then, use Python (with libraries like Plotly or Pandas) to enhance that visualization, such as adding custom interactive features (e.g., a filter to display data for specific months or regions).

5. Visualization Process:

- Discuss the entire process you followed, including:
 - i. Data exploration and preparation (using Python for data cleaning, handling missing values, etc.).
 - ii. How you used the open-source tool to create the base visualization (e.g., a bar chart or line graph).
 - iii. Customizations you applied using Python (e.g., creating a custom tooltip or adding more advanced visual features).

6. Key Takeaways and Insights:

- Highlight the key insights that can be drawn from your visualizations. For example, if using sales data, what trends or patterns can you identify? How does your visualization help uncover valuable information?
- Discuss the benefits of using both open-source tools and Python in data visualization, especially when working with complex datasets.

Deliverables:

- A PowerPoint/Google Slides presentation that covers the above steps in detail.
- Include screenshots or live demos of the visualizations created with the open-source tool and Python scripts.
- Be prepared to explain the code used for customization, and how it enhances the final visualization.

Resources:

- Apache Superset: <https://superset.apache.org/>
- Metabase: <https://www.metabase.com/>
- Redash: <https://redash.io/>
- Plotly Dash: <https://dash.plotly.com/>

Note: The project will be presented in two stages:

Step 1: Around weeks 7-8, the team will present their progress.

Step 2: Around weeks 11-12, the team will deliver/present the final presentation.