| **DATA VISUALIZATION** **LAB** | |
| --- | --- |
| **Course Code: ISL56** | **Credits: 0:0:1** |
| **Pre – requisites: Python** | **Contact Hours: 14** |
| **Coordinator: Dr. Lincy Meera Mathews** | |
| **Course Contents** | |
| **PART B** | |
| **Experiment 1:** Using Python, create your own having columns plant name, sunlight exposure, plant height and answer the following questions:   1. Is there a relationship between the number of hours of sunlight exposure and the height of the plants? 2. Visualize the relationship between sunlight exposure and plant height using a scatterplot. 3. Calculate the correlation coefficient between sunlight exposure and plant height. Is the correlation positive or negative? Is it strong or weak? 4. Based on the correlation coefficient, can we conclude that there is a significant association between sunlight exposure and plant growth rate?   **Experiment 2:** In a solar panel efficiency study, researchers want to investigate the relationship between the temperature and the efficiency of solar panels. They collected data on the temperature (in Celsius) and the corresponding efficiency (in percentage) of solar panels over a period of time. The dataset contains measurements from 50 different days.   1. Using Simple Linear Regression, can you develop a model to predict the efficiency of solar panels based on the temperature? 2. Perform an F-test to determine whether temperature significantly predicts the efficiency of solar panels. 3. Conduct a t-test to assess the significance of the regression coefficient for temperature.   **Experiment 3:** Given the dataset of 30 students' study hours and exam scores, how would you build a linear regression model to predict exam scores? Describe the steps you would take to diagnose the regression model, including checking assumptions, identifying outliers, and handling influential points. Finally, evaluate the model's performance and discuss any insights gained.  **Experiment 4:** In a retail experiment, we want to understand how advertising expenditure, store location, and competition affect sales revenue. Using synthetic data, implement multiple linear regression in Python to analyse these factors. Interpret the coefficients, perform an F-test to assess overall model significance, and conduct t-tests to evaluate the significance of individual coefficients.  **Experiment 5:** Given a dataset that contains information about different types of flowers (e.g., Iris dataset), perform classification using the **k-Nearest Neighbors (kNN)** algorithm. Evaluate the performance of the model by calculating its **accuracy** and visualize the results using appropriate techniques.  **Experiment 6:** Given a dataset that contains customer information (such as Age, Income, and Spending Score), perform K-means clustering to group customers into clusters. Use visualization chart, plot the data before and after grouping. Also, use the Elbow Method to determine the optimal number of clusters.  **Experiment 7:** Compare the effectiveness of two teaching methods, A and B, in helping students pass a test. Analyse the proportions of passing students, calculate confidence intervals for the difference in proportions, conduct significance tests, and evaluate the area under the ROC curve for predictive accuracy. | |
| **PART A** | |
| **Experiment 1:** Exploring and Visualizing Sales Data  Import a sample sales dataset into Tableau. Create a dashboard that includes the following visualizations:   * A bar chart showing total sales by product category. * A line chart showing monthly sales trends over the past year. * A geographic map showing sales distribution by region.   **Experiment 2:** Advanced Visualization Techniques  Using a dataset on customer purchases, create a dashboard with the following advanced visualizations:   * A scatter plot showing the relationship between customer age and total spending. * A heat map illustrating product sales intensity by store location. * A dual-axis chart comparing sales and profit margins over time.   **Experiment 3**. Analysing Sales Performance: Create various visualizations to analyze the sales performance.   * Create a line chart to show sales trends over time. * Create a bar chart to compare sales across different regions. * Use a pie chart to display the distribution of sales among different product categories. * Create a dashboard combining these visualizations to provide an overall view of sales performance.   **Experiment 4**. Financial Performance Dashboard: Create a comprehensive dashboard to monitor the financial performance of a company.   * Create a combination of line and bar charts to compare actual revenue and expenses against budgeted figures. * Use a gauge chart to display key financial metrics such as net profit margin. * Create a waterfall chart to show the contribution of different factors to the overall profit. * Design an interactive dashboard that allows users to filter data by different time periods (e.g., monthly, quarterly, yearly). | |

**Course outcomes (COs):**

At the end of the course, the student will be able to:

1. Interpret various data visualization techniques and principles along with pre-processing methods. (PO- 2, 3, 4)
2. Gain hands-on experience and proficiency in using data visualization tools. (PO – 2, 3, 4)
3. Apply data visualization techniques to solve real-world problems, making data-driven decisions in various domains such as business, healthcare, finance etc. (PO – 1, 2, 3, 4, 5)

Reference books:

1. <https://www.nrigroupindia.com/e-book/Introduction%20to%20Machine%20Learning%20with%20Python%20(%20PDFDrive.com%20)-min.pdf>

Web links:

[1. https://realpython.com/python3-object-oriented-programming/](about:blank)

[2. https://python.swaroopch.com/oop.html](about:blank)

[3. https://python-textbok.readthedocs.io/en/1.0/Object\_Oriented\_Programming.html](about:blank)

[4. https://www.programiz.com/python-programming/](about:blank)

[5. https://www.geeksforgeeks.org/python-programming-language/](about:blank)