**DATA WRANGLING AND VISUALIZATION – PYTHON/**

**R PROGRAMMING/POWER BI**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Course** | **B.Tech.-III-Sem.** | **L** | **T** | **P** | **C** |
| **Subject Code** | **22CAPC36** | - | - | **2** | **1** |

**Course Outcomes (COs) & CO-PO Mapping (3-Strong; 2-Medium; 1-Weak Correlation)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **COs** | **Upon completion of course the students will be able to** | **PO3** | **PO4** | **PO5** | **PO9** | **PO12** | **PSO2** |
| **CO1** | create python shell script for data validation | 3 | 3 | 3 | 3 | 3 | 3 |
| **CO2** | demonstrate how to import data into tableau | 3 | 3 | 3 | 3 | 3 | 3 |
| **CO3** | apply the tableau concepts of dimensions and measures | 3 | 3 | 3 | 3 | 3 | 3 |
| **CO4** | develop programs, map visual layouts and graphical properties | 3 | 3 | 3 | 3 | 3 | 3 |
| **CO5** | create a dashboard that links multiple visualizations | 3 | 3 | 3 | 3 | 3 | 3 |

**List of Experiments**

|  |  |
| --- | --- |
| **Week** | **Title/Experiment** |
| **Data Wrangling** | |
| 1 | Understanding Data, what is data, where to find data, data wrangling, data clean up basics - formatting, outliers, duplicates, normalizing and standardizing data. |
| 2 | Develop the python script to parse the pdf files using pdfminer. |
| 3 | Develop the python Shell Script to do the basic data cleanup on child labour and child marriage data.xlsx a) check duplicates and missing data b) eliminate mismatches c) cleans line breaks, spaces, and special characters. |
| 4 | Draw the chart between perceived corruption scores compared to the child labour percentages using matplotlib. |
| 5 | Write a python program to download & display content of robot.txt for en.wikipedia.org. |
| **Data Visualization** | |
| 6 | Foundations for building data visualizations, Creating first visualization. |
| 7 | Getting started with tableau software using data file formats, connecting data to tableau, creating basic charts (line, bar charts, tree maps) using the show me panel. |
| 8 | Tableau calculations, overview of SUM, AVG and aggregate features, creating custom calculations and fields. |
| 9 | Applying new data calculations to in visualizations, formatting visualizations, formatting tools and menus, formatting specific parts of the view. |
| 10 | Editing and formatting axes, manipulating data in tableau data, pivoting tableau data. |
| 11 | Structuring the data, sorting and filtering tableau data, pivoting tableau data. |
| 12 | Advanced visualization tools: using filters, using the detail panel, using the size panels, customizing filters, using and customizing tooltips, formatting data with colors. |
| 13 | Creating dashboards and storytelling, design for different displays, adding interactivity in dashboard, distributing, publishing data visualization. |
| 14 | Creating custom charts, cyclical data and circular area charts, dual axis charts. |
| **References** | |
| 1. Data Wrangling & Visualization - Python/R Programming/Power BI Manual, Dept. of CSE(AI&ML), CMRIT. | |
| **Micro-Projects:** Student should submit a report on one of the following/any other micro-project(s) approved by the lab faculty before commencement of lab internal examination. | |
| 1. Apply the raw data set, and implement the different data wrangling functionalities. 2. Perform Exploratory Data Analysis (EDA) and Data Wrangling in Pandas. 3. Perform Feature Engineering, one-hot encoding and deal with missing data.  Import Datasets and Perform Basic Statistical Data Analysis.  1. Develop a Scatter Plot with Matplotlib. 2. Basic Interactive Binned Scatter Plot with Altair. 3. Histogram with Plotnine (ggplot). 4. Create a Viz on Cricket Stadium. 5. Creating common visualizations on various charts and assembling a dashboard layout. 6. Develop data visualization on interactive plot with Plotly (using Cufflinks). | |