



VIT[®]
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

CSE5007 - Project Component

1. Choose the domain for a dataset belonging to any one of the following:
 - Healthcare/ Vaccination
 - Governance/Census
 - Transportation/Fuel
 - Election Board/Polling
 - Electricity/Water/Gas/Household Utilities Consumption
 - University/Examination
 - Meteorological/Climate Change
 - Exports/Imports/Trade
2. Students can choose their team on their own with minimum of 2 and maximum of 4 members. Provide the details such as the domain, the team members, the link for accessing the relevant dataset in the google drive link on or before 2-2-24.

<https://forms.gle/3b7TY5JgPAFbgLi18>

3. Students are instructed to choose a suitable dataset that is unclean and proceed using the techniques to
 - Clean the dataset
 - Find outliers
 - Dimensionality reduction
 - Feature Selection

This task is weighted for challenging Digital Assignment component also.

4. Prepare a detailed report and submit the pdf for review 3 upload in VTOP
 - The collage pic of all the team members.
 - Name and registration number details.
 - The link for the dataset.
 - You can derive or transform the dataset based on your needs.
 - The Jupyter notebook-based manipulations you have used to derive the data for all of the EDA made.
 - Mention the quantitative and qualitative attributes available in your dataset.
 - The figures representing the results used to analyze the dataset
 - Provide the link to your dash in a google drive link. Remember to provide the access to me padmapriya.r@vit.ac.in
5. Present the necessary information being inferred from the respective exploratory data analysis.
6. Students are hereby informed to incorporate any four of the following as part of the project preparation. Atleast any 2 must be interactive.
 - Scatter plot
 - Bar graph
 - Stacked Bar graph

- Heatmap
- Pie chart
- Network/Tree
- Histogram
- Boxplot
- Line chart