Data Collection and Preprocessing Phase

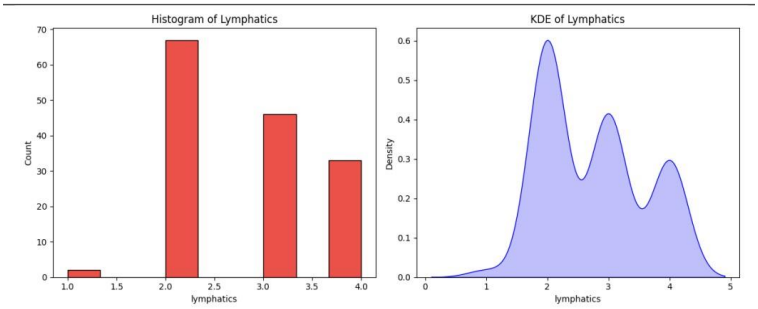
Date	09 July 2024
Team ID	739734
Project Title	Evolving efficient classification patterns in Lymphography
Maximum Marks	6 Marks

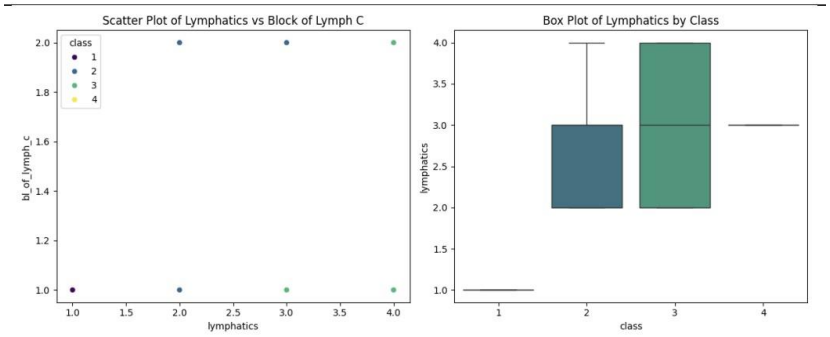
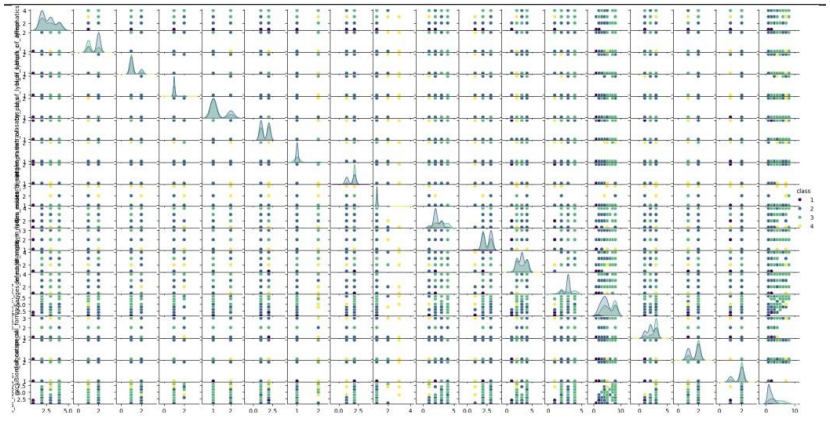
Data Exploration and Preprocessing Template

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

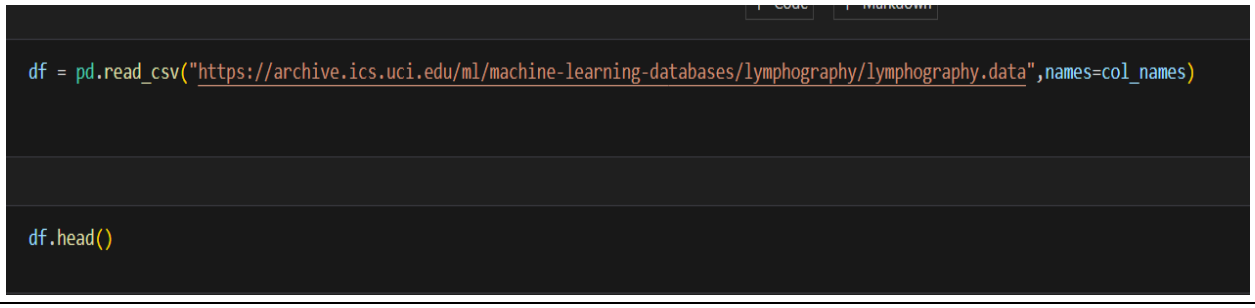
Section	Description

Data Overview	Attribute information: --- NOTE: All attribute values in the database have been entered as numeric values corresponding to their index in the list of attribute values for that attribute domain as given below. 1. class: normal find, metastases, malign lymph, fibrosis 2. lymphatics: normal, arched, deformed, displaced 3. block of affere: no, yes 4. bl. of lymph. c: no, yes 5. bl. of lymph. s: no, yes 6. by pass: no, yes 7. extravasates: no, yes 8. regeneration of: no, yes 9. early uptake in: no, yes 10. lym.nodes dimin: 0-3 11. lym.nodes enlar: 1-4 12. changes in lym.: bean, oval, round 13. defect in node: no, lacunar, lac. marginal, lac. central 14. changes in node: no, lacunar, lac. margin, lac. central 15. changes in stru: no, grainy, drop-like, coarse, diluted, reticular, stripped, faint, 16. special forms: no, chalices, vesicles 17. dislocation of: no, yes 18. exclusion of no: no, yes 19. no. of nodes in: 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, >=70 Missing Attribute Values: None Class Distribution: Class: Number of Instances: normal find: 2 metastases: 81 malign lymph: 61 fibrosis: 4 Dimension: 614 rows × 13 columns Descriptive statistics

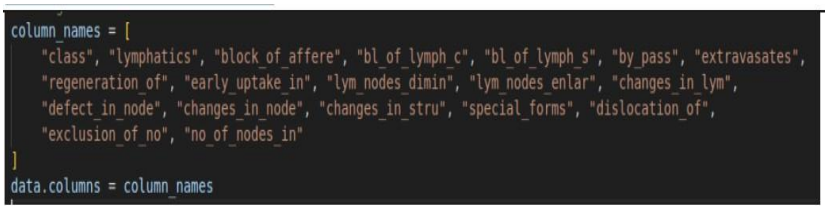
Univariate Analysis	

Bivariate Analysis	
Multivariate Analysis	
Outliers and Anomalies	Outliers in lymphography data can be data points that fall outside the expected range for lymph node size, flow patterns, or tracer distribution. These outliers may indicate errors during imaging, unusual anatomical features, or potential diseases requiring further investigation. Anomalies in lymphography could be specific patterns, like unexpected blockages or leakage, that deviate from the norm and warrant specialist attention.

Data Preprocessing Code Screenshots

Loading Data	

Handling Missing Data	No missing attributes

Data Transformation	
Feature Engineering	Attached the codes in final submission.
Save Processed Data	Done