Data Collection and Preprocessing Phase

Date	8th July 2024
Team ID	SWTID1720080033
Project Title	Anemia Sense: Leveraging Machine Learning For Precise Anemia Recognitions
Maximum Marks	6 Marks

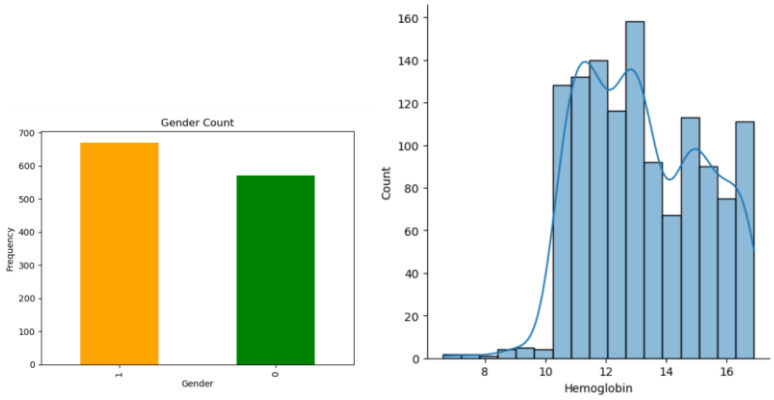
Data Exploration and Preprocessing Template

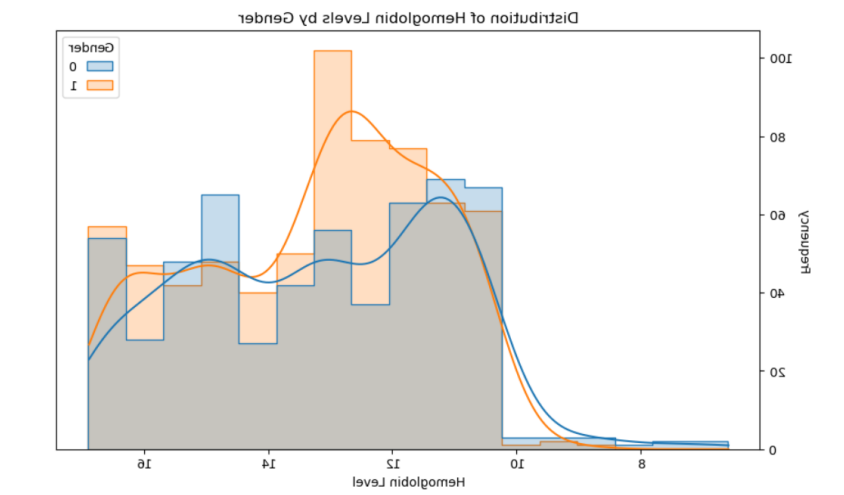
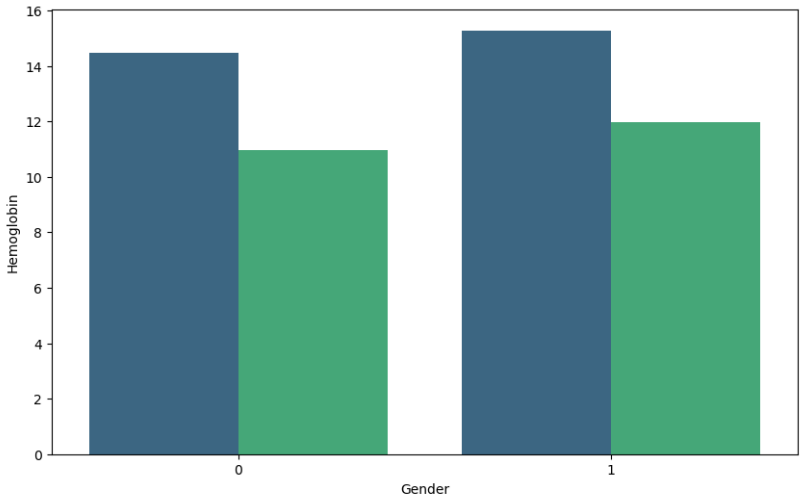
Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

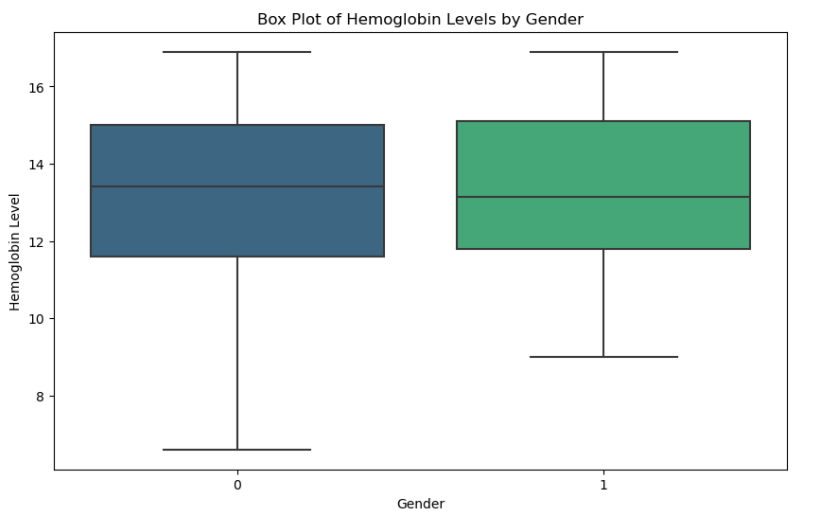
Section	Description
Data Overview	

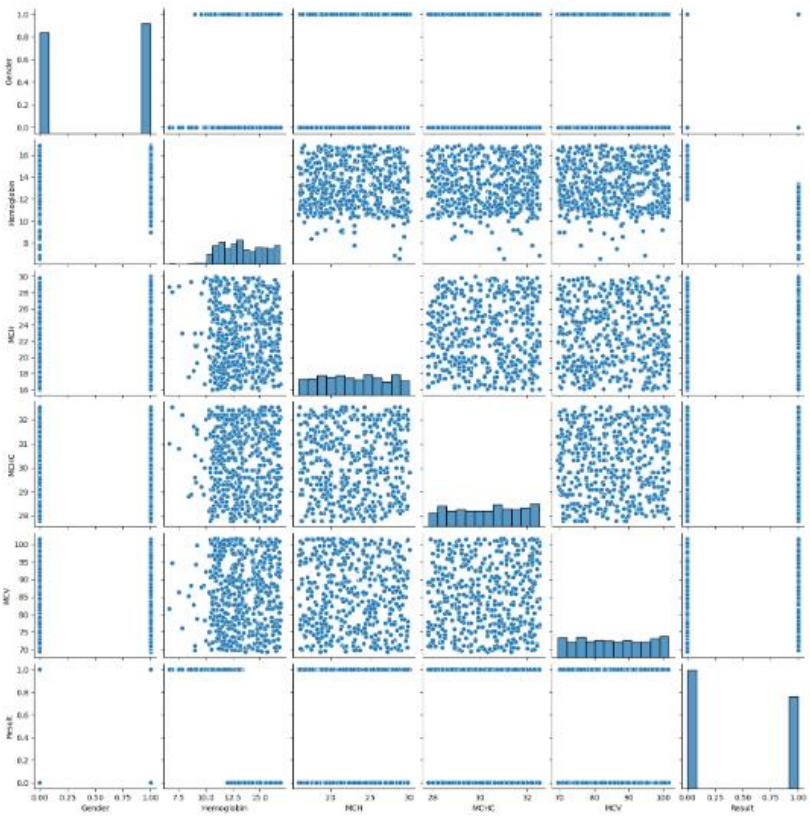
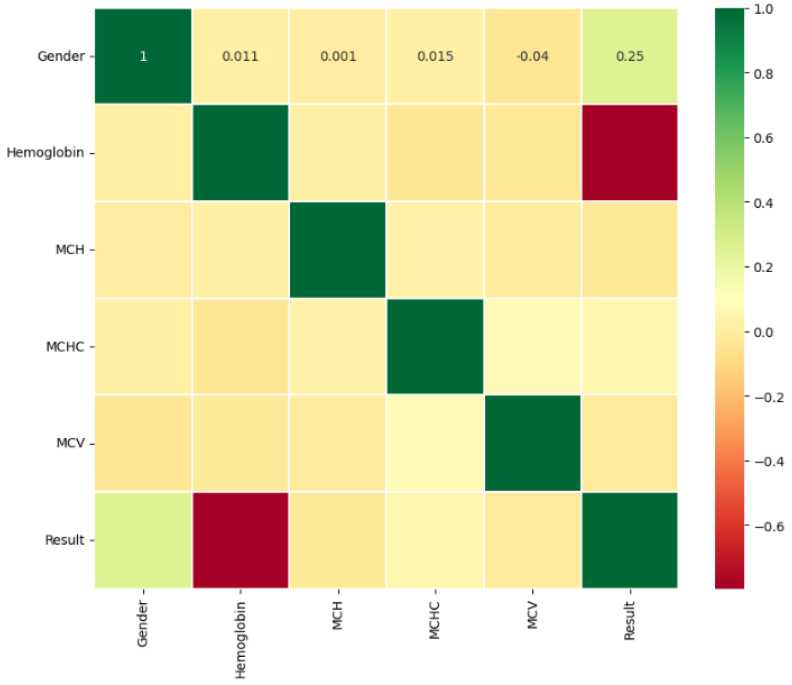
	Gender	Hemoglobin	MCH	MCHC	MCV	Result
count	1240.000000	1240.000000	1240.000000	1240.000000	1240.000000	1240.000000
mean	0.540323	13.218145	22.903952	30.277984	85.620968	0.500000
std	0.498573	1.976190	3.993624	1.394515	9.673794	0.500202
min	0.000000	6.600000	16.000000	27.800000	69.400000	0.000000
25%	0.000000	11.500000	19.400000	29.100000	77.300000	0.000000
50%	1.000000	13.000000	22.700000	30.400000	85.300000	0.500000
75%	1.000000	14.900000	26.200000	31.500000	94.225000	1.000000
max	1.000000	16.900000	30.000000	32.500000	101.600000	1.000000

Univariate Analysis



<table>
<tr>
<td></td>
<td> 
</td>
</tr>
<tr>
<td>Bivariate Analysis</td>
<td> 


</td>
</tr>
</table>

Multivariate Analysis	 
Outliers and Anomalies	NIL

Data Preprocessing Code Screenshots	
Loading Data	df=pd.read_csv('anemia.csv') [2]: `df=pd.read_csv('anemia.csv')`
Handling Missing Data	df.isnull().sum() [8]: `df.isnull().sum()` [8]: Gender 0 Hemoglobin 0 MCH 0 MCHC 0 MCV 0 Result 0 dtype: int64
Data Transformation	NIL
Feature Engineering	NIL
Save Processed Data	NIL