Data Collection and Preprocessing Phase
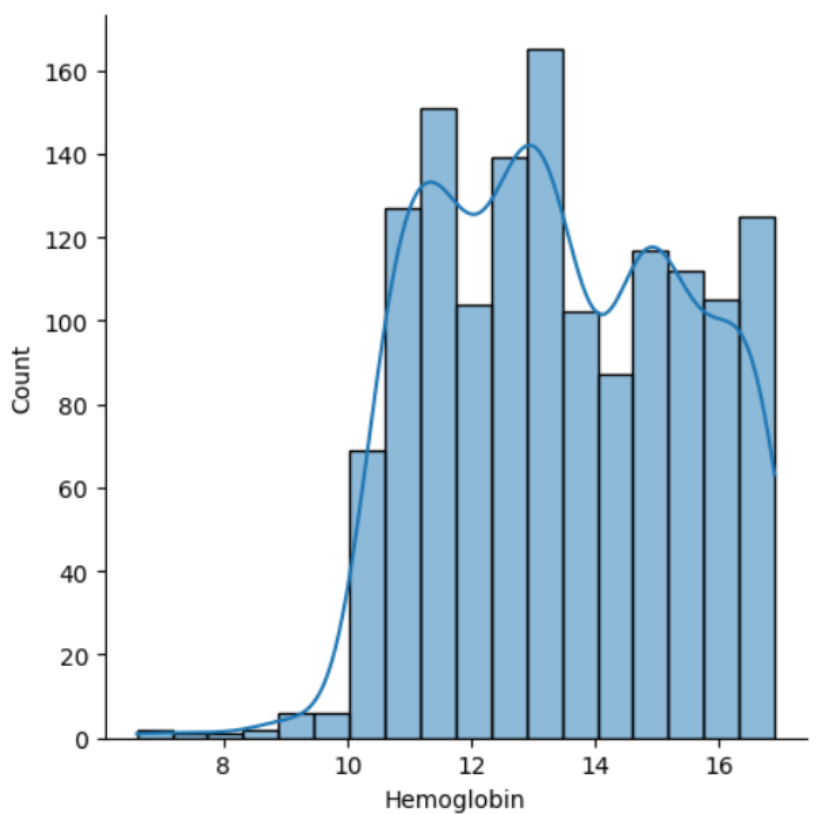
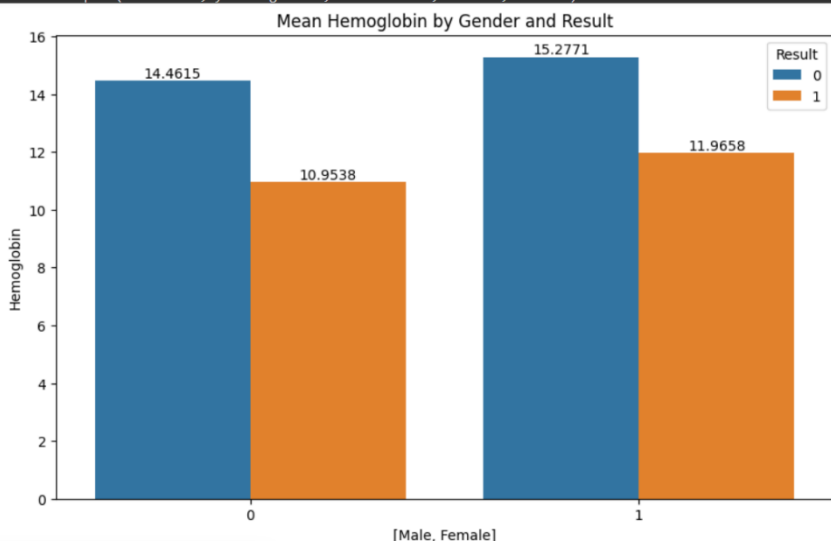|	
Date	8 August 2025
Skillwallet ID	SWUID20250188620
Project Title	Anemia Sense: Leveraging Machine Learning for Precise Anemia Recognition
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

Dataset variables were statistically analyzed to identify patterns, distributions, and possible anomalies. Python was employed for preprocessing tasks, including handling missing values, normalizing features, and encoding categorical variables. This ensured a clean dataset for model training and provided a strong foundation for accurate anemia prediction.

Section	Description
Data Overview	Dimension: 1,421 rows x 6 columns Descriptive statistics: |	Gender	Hemoglobin	MCH	MCHC	MCV	Result
---	---	---	---	---	---	---
count	1421.000000	1421.000000	1421.000000	1421.000000	1421.000000	1421.000000
mean	0.520760	13.412738	22.905630	30.251232	85.523786	0.436312
std	0.499745	1.974546	3.969375	1.400898	9.636701	0.496102
min	0.000000	6.600000	16.000000	27.800000	69.400000	0.000000
25%	0.000000	11.700000	19.400000	29.000000	77.300000	0.000000
50%	1.000000	13.200000	22.700000	30.400000	85.300000	0.000000
75%	1.000000	15.000000	26.200000	31.400000	94.200000	1.000000
max	1.000000	16.900000	30.000000	32.500000	101.600000	1.000000	
Univariate Analysis	

`<seaborn.axisgrid.FacetGrid at 0x79501b49ab10>`

| Bivariate Analysis | |

Mean Hemoglobin by Gender and Result

| Multivariate Analysis | |

| Outliers and Anomalies | - |

Data Preprocessing Code Screenshots

| Loading Data | |

Handling Missing Data	
| | ```
df.shape

(1421, 6)

df.isnull().sum()

                    0
        Gender      0
    Hemoglobin      0
        MCH         0
        MCHC        0
        MCV         0
        Result      0

dtype: int64
``` |
| Feature Engineering | Attached the codes in final submission. |
| Save Processed Data | - |