

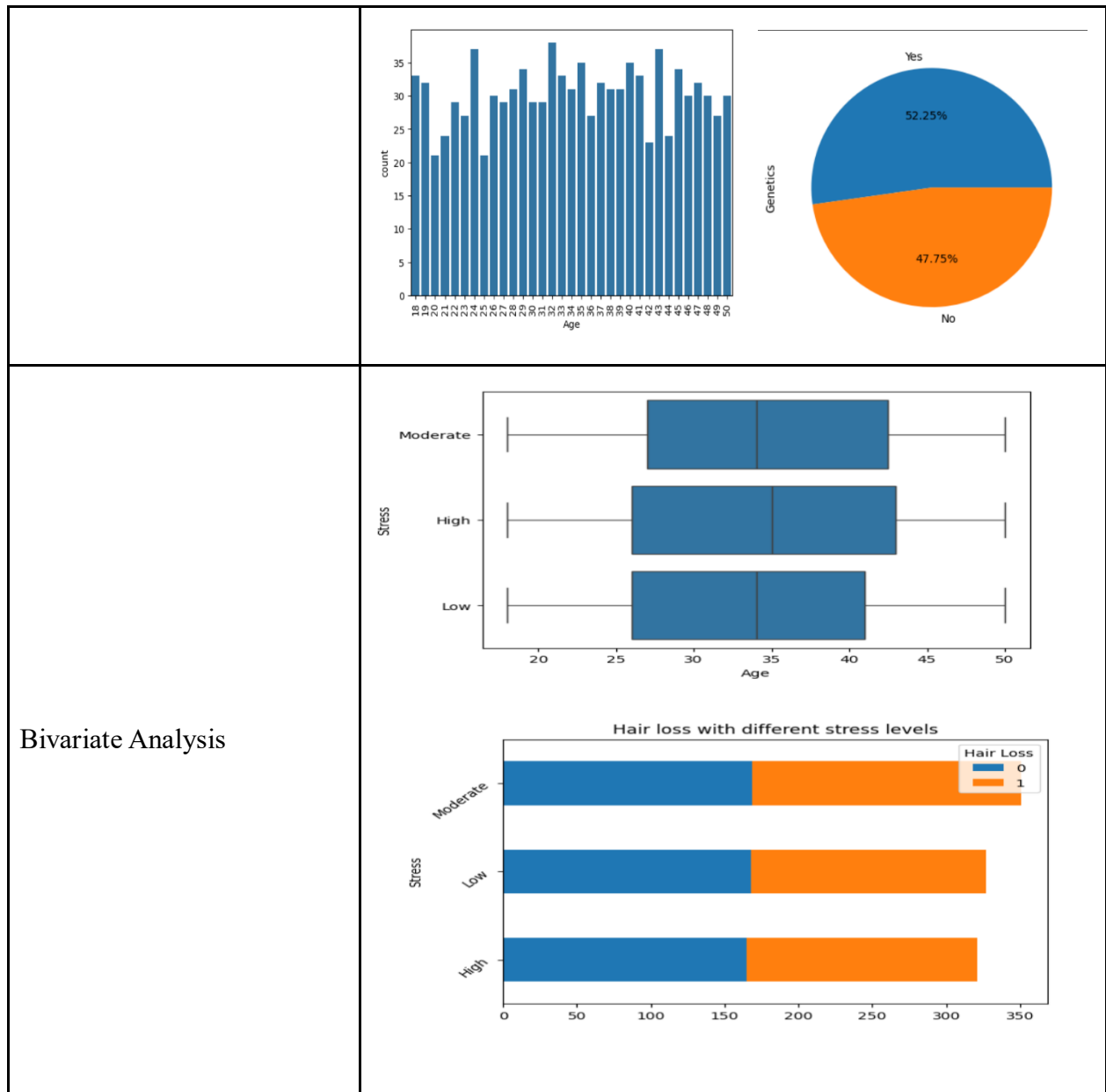
## Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	740017
Project Title	Unveiling Baldness: Genetic And Environmental Dynamics
Maximum Marks	6 Marks

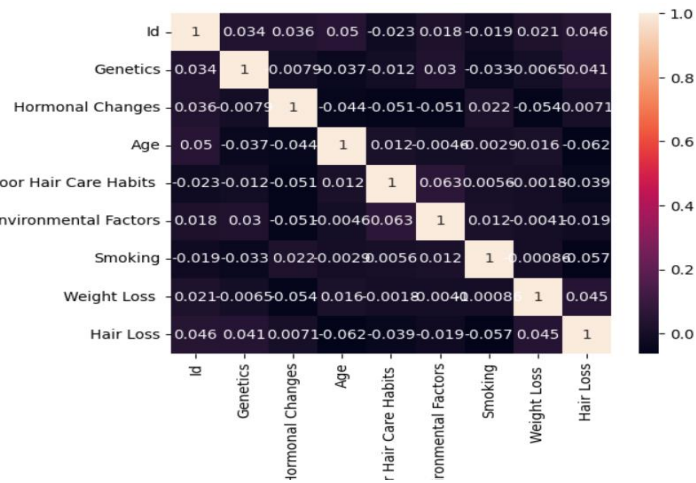
### Data Exploration and Preprocessing Report

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	<div><div>Dimension:</div><div>999 rows × 13 columns</div><div>Descriptive statistics:</div></div> <table><tr><th></th><th>Id</th><th>Genetics</th><th>Hormonal Changes</th><th>Medical Conditions</th><th>Medications &amp; Treatments</th><th>Nutritional Deficiencies</th><th>Stress</th><th>Age</th><th>Poor Hair Care Habits</th><th>Environmental Factors</th><th>Smoking</th><th>Weight Loss</th><th>Hair Loss</th></tr><tr><td>0</td><td>133992</td><td>Yes</td><td>No</td><td>No Data</td><td>No Data</td><td>Magnesium deficiency</td><td>Moderate</td><td>19</td><td>Yes</td><td>Yes</td><td>No</td><td>No</td><td>0</td></tr><tr><td>1</td><td>148393</td><td>No</td><td>No</td><td>Eczema</td><td>Antibiotics</td><td>Magnesium deficiency</td><td>High</td><td>43</td><td>Yes</td><td>Yes</td><td>No</td><td>No</td><td>0</td></tr><tr><td>2</td><td>155074</td><td>No</td><td>No</td><td>Dermatosis</td><td>Antifungal Cream</td><td>Protein deficiency</td><td>Moderate</td><td>26</td><td>Yes</td><td>Yes</td><td>No</td><td>Yes</td><td>0</td></tr><tr><td>3</td><td>118261</td><td>Yes</td><td>Yes</td><td>Ringworm</td><td>Antibiotics</td><td>Biotin Deficiency</td><td>Moderate</td><td>46</td><td>Yes</td><td>Yes</td><td>No</td><td>No</td><td>0</td></tr><tr><td>4</td><td>111915</td><td>No</td><td>No</td><td>Psoriasis</td><td>Accutane</td><td>Iron deficiency</td><td>Moderate</td><td>30</td><td>No</td><td>Yes</td><td>Yes</td><td>No</td><td>1</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>994</td><td>184367</td><td>Yes</td><td>No</td><td>Seborrheic Dermatitis</td><td>Rogaine</td><td>Vitamin A Deficiency</td><td>Low</td><td>33</td><td>Yes</td><td>Yes</td><td>Yes</td><td>Yes</td><td>1</td></tr><tr><td>995</td><td>164777</td><td>Yes</td><td>Yes</td><td>No Data</td><td>Accutane</td><td>Protein deficiency</td><td>Low</td><td>47</td><td>No</td><td>No</td><td>No</td><td>Yes</td><td>0</td></tr><tr><td>996</td><td>143273</td><td>No</td><td>Yes</td><td>Androgenetic Alopecia</td><td>Antidepressants</td><td>Protein deficiency</td><td>Moderate</td><td>20</td><td>Yes</td><td>No</td><td>Yes</td><td>Yes</td><td>1</td></tr><tr><td>997</td><td>169123</td><td>No</td><td>Yes</td><td>Dermatitis</td><td>Immunomodulators</td><td>Biotin Deficiency</td><td>Moderate</td><td>32</td><td>Yes</td><td>Yes</td><td>Yes</td><td>Yes</td><td>1</td></tr><tr><td>998</td><td>127183</td><td>Yes</td><td>Yes</td><td>Psoriasis</td><td>Blood Pressure Medication</td><td>Vitamin D Deficiency</td><td>Low</td><td>34</td><td>No</td><td>Yes</td><td>No</td><td>No</td><td>1</td></tr></table>		Id	Genetics	Hormonal Changes	Medical Conditions	Medications & Treatments	Nutritional Deficiencies	Stress	Age	Poor Hair Care Habits	Environmental Factors	Smoking	Weight Loss	Hair Loss	0	133992	Yes	No	No Data	No Data	Magnesium deficiency	Moderate	19	Yes	Yes	No	No	0	1	148393	No	No	Eczema	Antibiotics	Magnesium deficiency	High	43	Yes	Yes	No	No	0	2	155074	No	No	Dermatosis	Antifungal Cream	Protein deficiency	Moderate	26	Yes	Yes	No	Yes	0	3	118261	Yes	Yes	Ringworm	Antibiotics	Biotin Deficiency	Moderate	46	Yes	Yes	No	No	0	4	111915	No	No	Psoriasis	Accutane	Iron deficiency	Moderate	30	No	Yes	Yes	No	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	994	184367	Yes	No	Seborrheic Dermatitis	Rogaine	Vitamin A Deficiency	Low	33	Yes	Yes	Yes	Yes	1	995	164777	Yes	Yes	No Data	Accutane	Protein deficiency	Low	47	No	No	No	Yes	0	996	143273	No	Yes	Androgenetic Alopecia	Antidepressants	Protein deficiency	Moderate	20	Yes	No	Yes	Yes	1	997	169123	No	Yes	Dermatitis	Immunomodulators	Biotin Deficiency	Moderate	32	Yes	Yes	Yes	Yes	1	998	127183	Yes	Yes	Psoriasis	Blood Pressure Medication	Vitamin D Deficiency	Low	34	No	Yes	No	No	1
		Id	Genetics	Hormonal Changes	Medical Conditions	Medications & Treatments	Nutritional Deficiencies	Stress	Age	Poor Hair Care Habits	Environmental Factors	Smoking	Weight Loss	Hair Loss																																																																																																																																																											
	0	133992	Yes	No	No Data	No Data	Magnesium deficiency	Moderate	19	Yes	Yes	No	No	0																																																																																																																																																											
	1	148393	No	No	Eczema	Antibiotics	Magnesium deficiency	High	43	Yes	Yes	No	No	0																																																																																																																																																											
	2	155074	No	No	Dermatosis	Antifungal Cream	Protein deficiency	Moderate	26	Yes	Yes	No	Yes	0																																																																																																																																																											
	3	118261	Yes	Yes	Ringworm	Antibiotics	Biotin Deficiency	Moderate	46	Yes	Yes	No	No	0																																																																																																																																																											
	4	111915	No	No	Psoriasis	Accutane	Iron deficiency	Moderate	30	No	Yes	Yes	No	1																																																																																																																																																											
	...	...	...	...	...	...	...	...	...	...	...	...	...	...																																																																																																																																																											
	994	184367	Yes	No	Seborrheic Dermatitis	Rogaine	Vitamin A Deficiency	Low	33	Yes	Yes	Yes	Yes	1																																																																																																																																																											
	995	164777	Yes	Yes	No Data	Accutane	Protein deficiency	Low	47	No	No	No	Yes	0																																																																																																																																																											
996	143273	No	Yes	Androgenetic Alopecia	Antidepressants	Protein deficiency	Moderate	20	Yes	No	Yes	Yes	1																																																																																																																																																												
997	169123	No	Yes	Dermatitis	Immunomodulators	Biotin Deficiency	Moderate	32	Yes	Yes	Yes	Yes	1																																																																																																																																																												
998	127183	Yes	Yes	Psoriasis	Blood Pressure Medication	Vitamin D Deficiency	Low	34	No	Yes	No	No	1																																																																																																																																																												
Univariate Analysis																																																																																																																																																																									



## Multivariate Analysis



## Outliers and Anomalies

-

## Data Preprocessing Code Screenshots

## Loading Data

```
# importing the dataset which is in csv file
data= pd.read_csv('/content/Dataset/hairfall_prediction.csv')
data
```

1	Id	Genetics	Hormonal	Medical C	Medicatio	Nutritiona	Stress	Age	Poor Hair	Environm	Smoking	Weight Lo	Hair Loss
2	133992	Yes	No	No Data	No Data	Magnesi	Moderate	19	Yes	Yes	No	No	0
3	148393	No	No	Eczema	Antibiotic	Magnesi	High	43	Yes	Yes	No	No	0
4	155074	No	No	Dermatos	Antifungal	Protein de	Moderate	26	Yes	Yes	No	Yes	0
5	118261	Yes	Yes	Ringworm	Antibiotic	Biotin Def	Moderate	46	Yes	Yes	No	No	0
6	111915	No	No	Psoriasis	Accutane	Iron defici	Moderate	30	No	Yes	Yes	No	1
7	139661	Yes	No	Psoriasis	Antibiotic	Magnesi	Low	37	No	Yes	No	Yes	1

Handling Missing Data	<pre>df.info()</pre> <pre>&lt;class 'pandas.core.frame.DataFrame'&gt; RangeIndex: 999 entries, 0 to 998 Data columns (total 13 columns): #   Column                                Non-Null Count  Dtype ---  ---                                --- 0   Id                                    999 non-null    int64 1   Genetics                             999 non-null    object 2   Hormonal Changes                     999 non-null    object 3   Medical Conditions                   999 non-null    object 4   Medications &amp; Treatments             999 non-null    object 5   Nutritional Deficiencies             999 non-null    object 6   Stress                               999 non-null    object 7   Age                                  999 non-null    int64 8   Poor Hair Care Habits                999 non-null    object 9   Environmental Factors                999 non-null    object 10  Smoking                              999 non-null    object 11  Weight Loss                          999 non-null    object 12  Hair Loss                            999 non-null    int64 dtypes: int64(3), object(10) memory usage: 101.6+ KB</pre> <pre>df.isnull().sum()</pre> <pre>Id                                0 Genetics                          0 Hormonal Changes                  0 Medical Conditions                0 Medications &amp; Treatments          0 Nutritional Deficiencies          0 Stress                           0 Age                               0 Poor Hair Care Habits             0 Environmental Factors             0 Smoking                          0 Weight Loss                       0 Hair Loss                         0 dtype: int64</pre>
Feature Engineering	Attached the codes in final submission.
Save Processed Data	-