# Data Collection and Preprocessing Phase

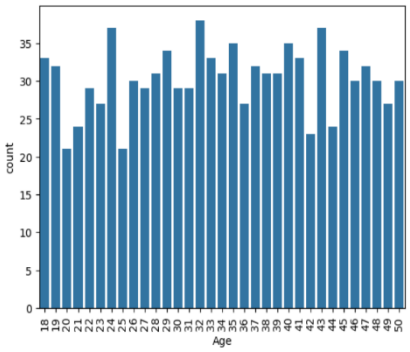
| | |
|---|---|
| Date | 10 July 2024 |
| Team ID | 740017 |
| Project Title | Unveiling Baldness: Genetic And Environmental Dynamics |
| Maximum Marks | 6 Marks |

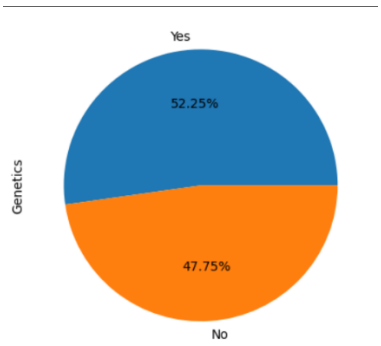
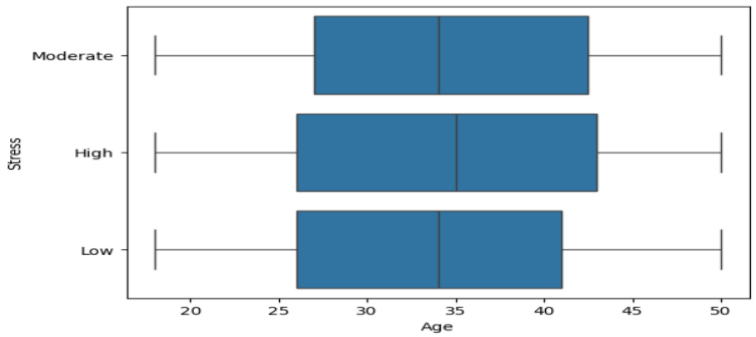
**Data Exploration and Preprocessing Report**

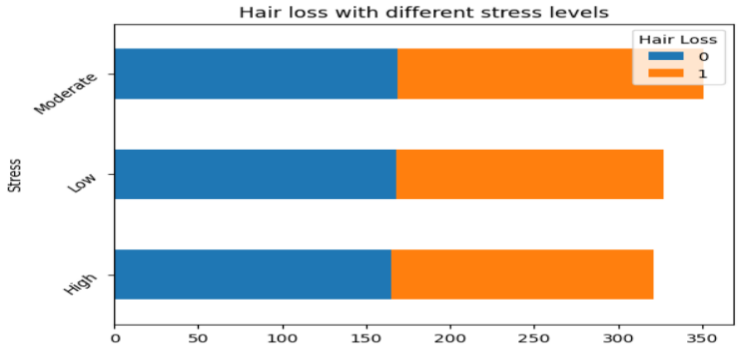
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong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

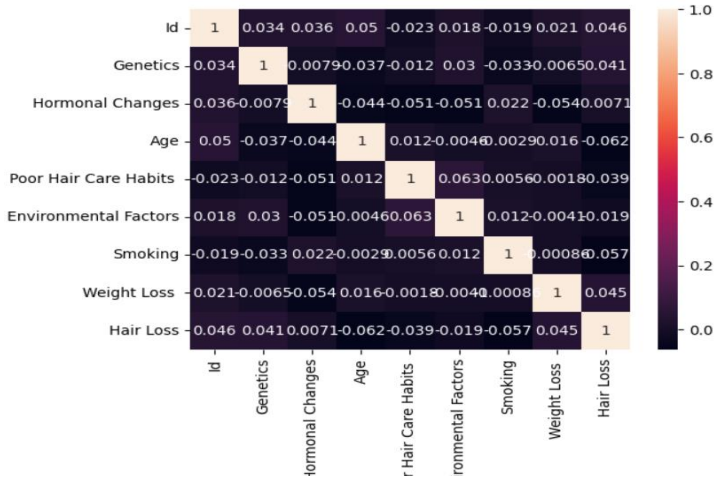
| Section | Description |
|---|---|
| Data Overview | Dimension:<br>999 rows × 13 columns<br>Descriptive statistics:<br> |
| Univariate Analysis | |

| | 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 |

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


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


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

| | |
|---|---|
| Multivariate Analysis |  |



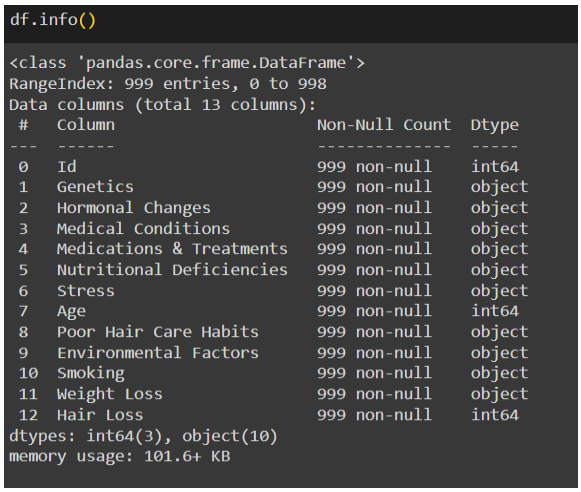
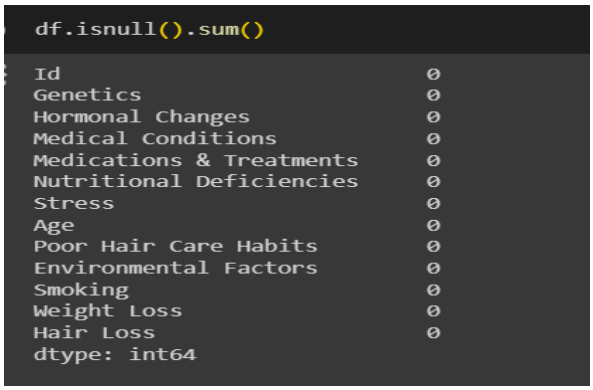
| | |
|---|---|
| Outliers and Anomalies | - |

**Data Preprocessing Code Screenshots**

Loading Data

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

| | 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 | Magnesiu | Moderate | 19 | Yes | Yes | No | No | 0 |
| 3 | 148393 | No | No | Eczema | Antibiotic | Magnesiu | 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 | Magnesiu | Low | 37 | No | Yes | No | Yes | 1 |

| | |
|---|---|
| Handling Missing Data |  <br>  |
| Feature Engineering | Attached the codes in final submission. |
| Save Processed Data | - |