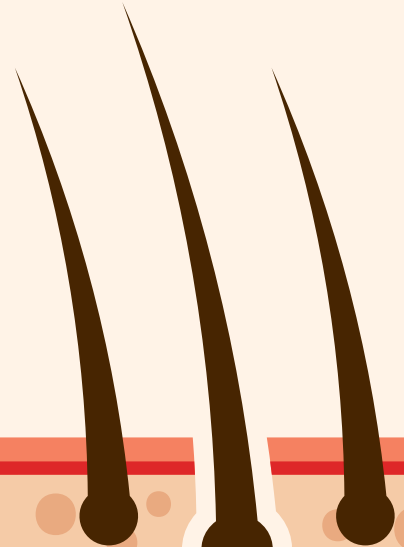


Hair Loss Prediction

Presented By: Maryam Montilla



Introduction/Background of Data

Hair loss can be something that most people aim to prevent, but taking a deep dive into the contributing and non-contributing factors that affect hair health in individuals can be interesting to analyze for those currently experiencing hair loss or wanting to prevent it. Using the Hair Loss Kaggle Dataset, we can examine specific conditions and features that determine whether a person experiences hair loss or not and create models that can make useful predictions for hair loss as well. These are some key features in our dataset:

Data Overview:

- **Genetics:** Indicates whether the individual has a family history of baldness (Yes/No).
- **Hormonal Changes:** Indicates whether the individual has experienced hormonal changes (Yes/No).
- **Medical Conditions:** Lists specific medical conditions that may contribute to baldness, such as Alopecia Areata, Thyroid Problems, Scalp Infection, Psoriasis, Dermatitis, etc.
- **Medications & Treatments:** Lists medications and treatments that may lead to hair loss, such as Chemotherapy, Heart Medication, Antidepressants, Steroids, etc.
- **Nutritional Deficiencies:** Lists nutritional deficiencies that may contribute to hair loss, such as Iron deficiency, Vitamin D deficiency, Biotin deficiency, Omega-3 fatty acid deficiency, etc.
- **Stress:** Indicates the stress level of the individual (Low/Moderate/High).
- **Age:** Represents the age of the individual.
- **Poor Hair Care Habits:** Indicates whether the individual practices poor hair care habits (Yes/No).
- **Environmental Factors:** Indicates whether the individual is exposed to environmental factors that may contribute to hair loss (Yes/No).
- **Smoking:** Indicates whether the individual smokes (Yes/No).
- **Weight Loss:** Indicates whether the individual has experienced significant weight loss (Yes/No).
- **Baldness (Target):** Binary variable indicating the presence (1) or absence (0) of baldness in the individual.

Our dataset contains 999 rows and 13 columns.

Purpose of Analysis

01

Data Cleaning

Cleaning up the data by making 'Yes'/'No' values into binary 1/0 columns for better analysis and predictions.

02

Visualizations

Creating visualizations between different features so individuals can evaluate which factors they experience actually correlate to hair loss.

03

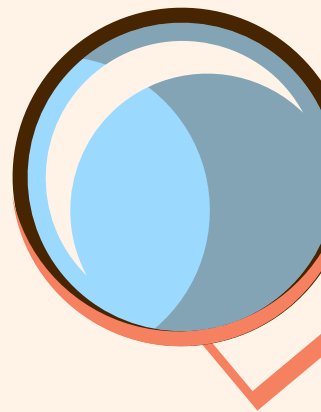
Classification Model

Using KNN, RandomForest, and Logistic Regression to find the best model that beats a baseline accuracy score of 50.25% to predict hair loss.

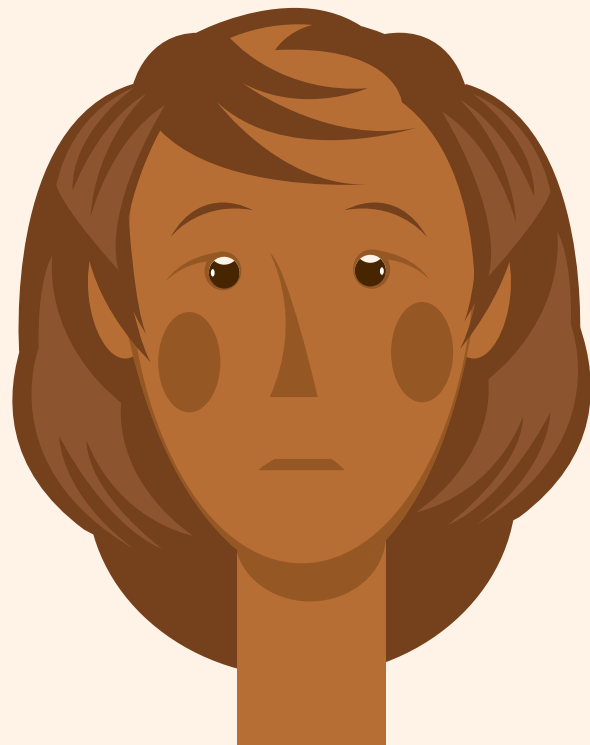
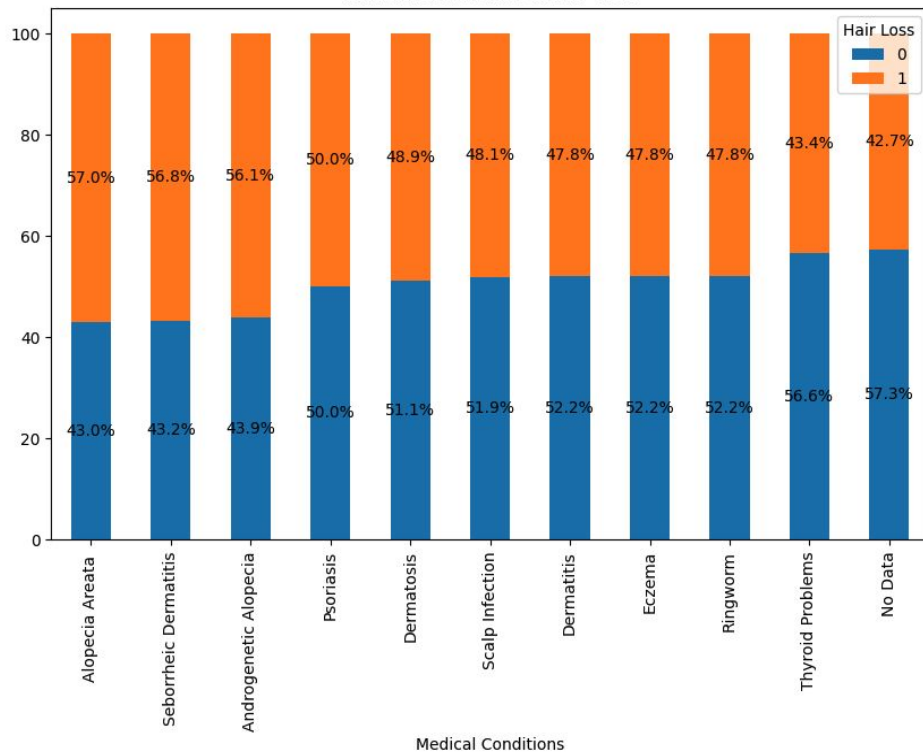
04

Conclusions

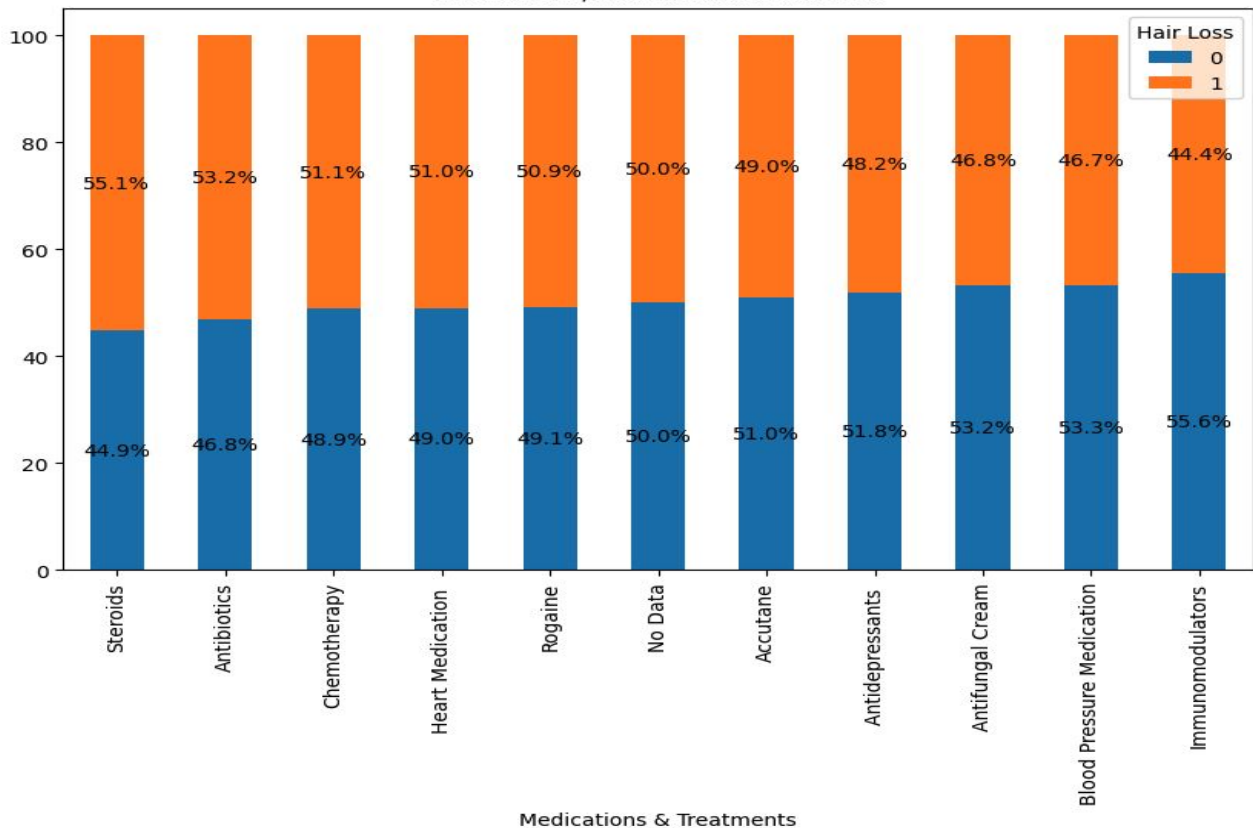
Collecting analysis to make accurate predictions on hair loss using the features in dataset. Those who are interested in predicting whether or not they may experience hair loss and what factors contribute to hair loss as well.

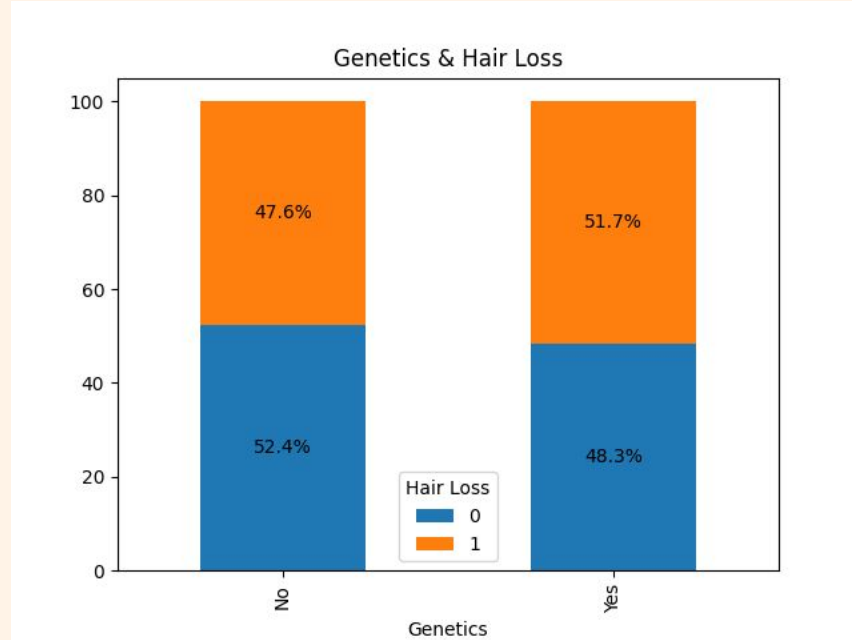
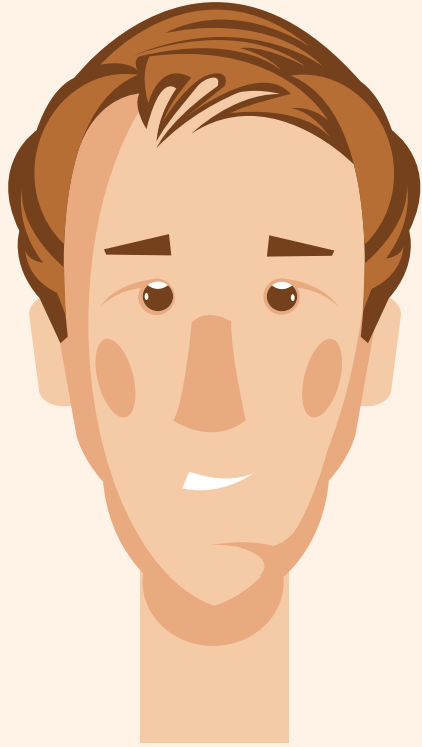


Medical Conditions & Hair Loss

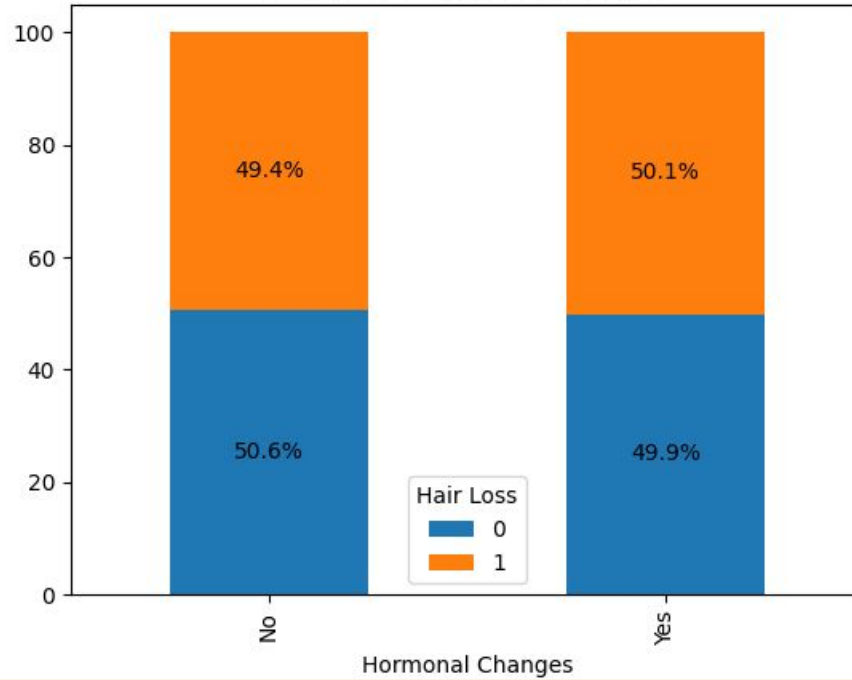


Medications/Treatments & Hair Loss

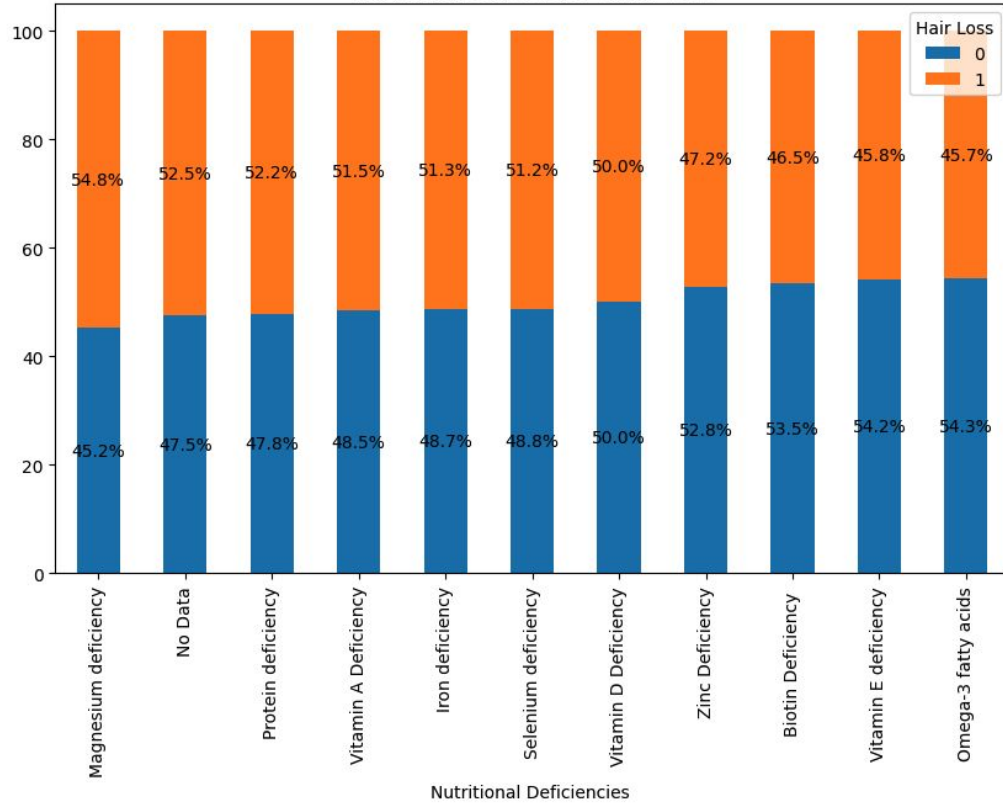




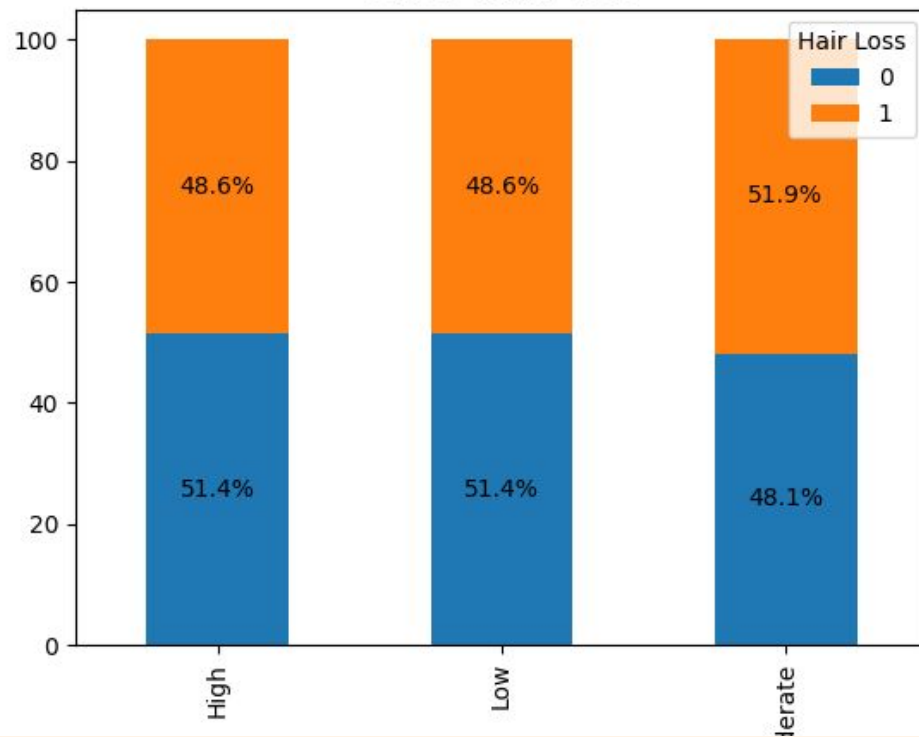
Hormonal Changes & Hair Loss

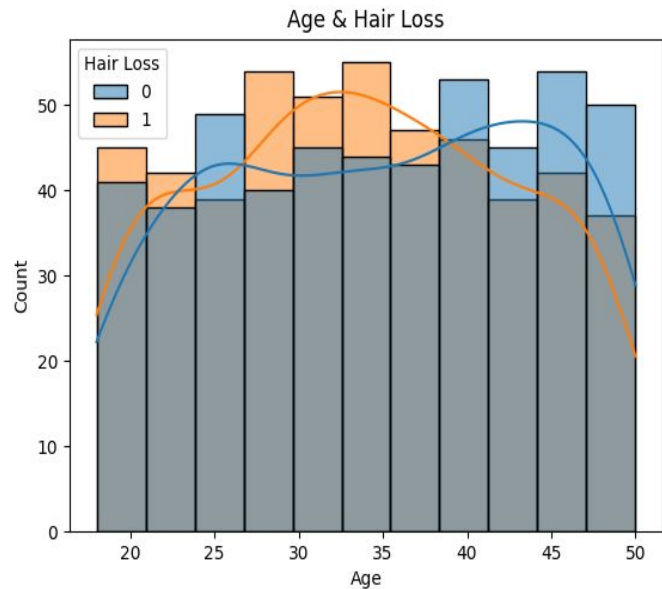
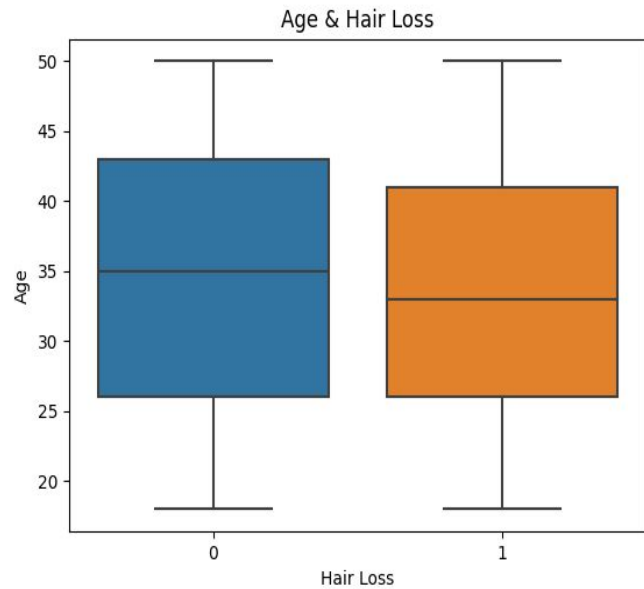


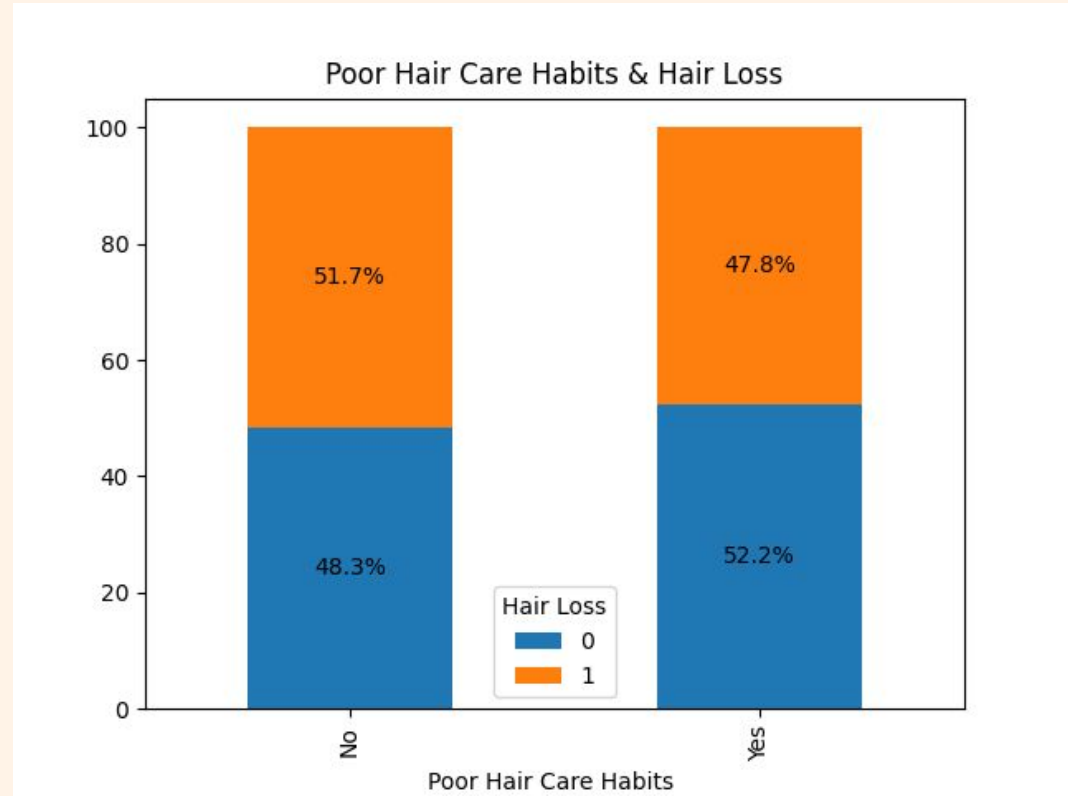
Nutritional Deficiencies & Hair Loss



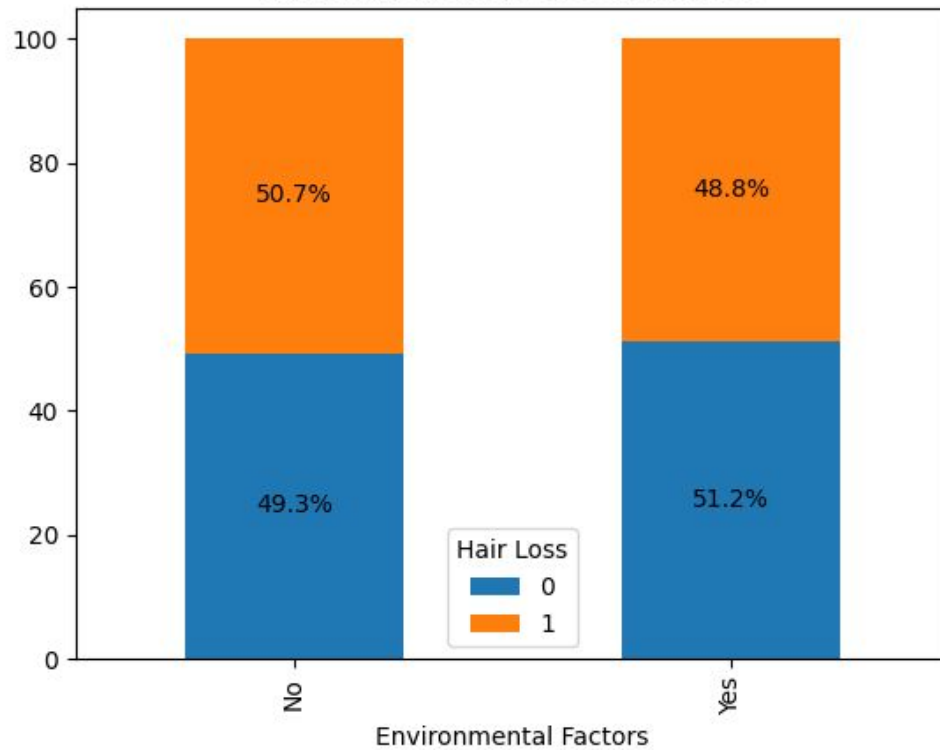
Stress & Hair Loss



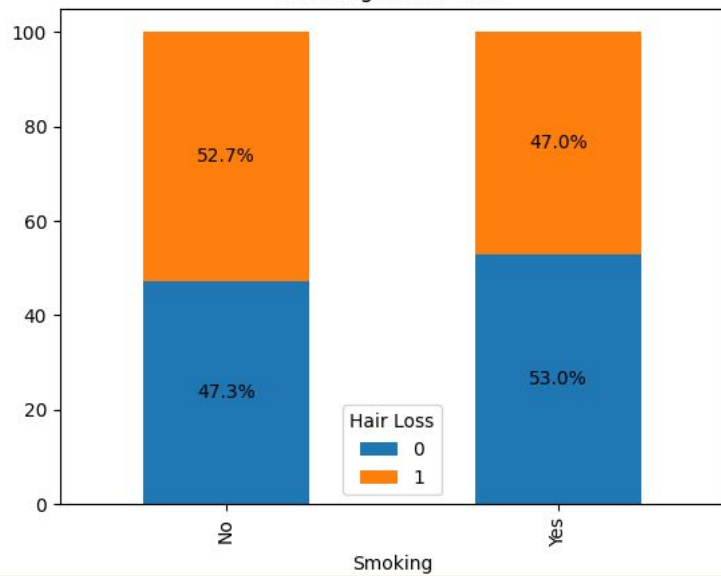




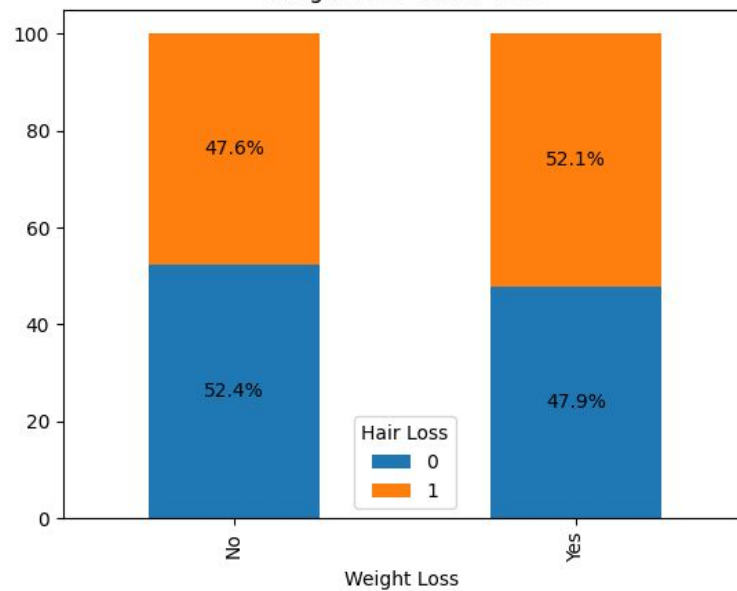
Environmental Factors & Hair Loss



Smoking & Hair Loss



Weight Loss & Hair Loss





Hair Loss



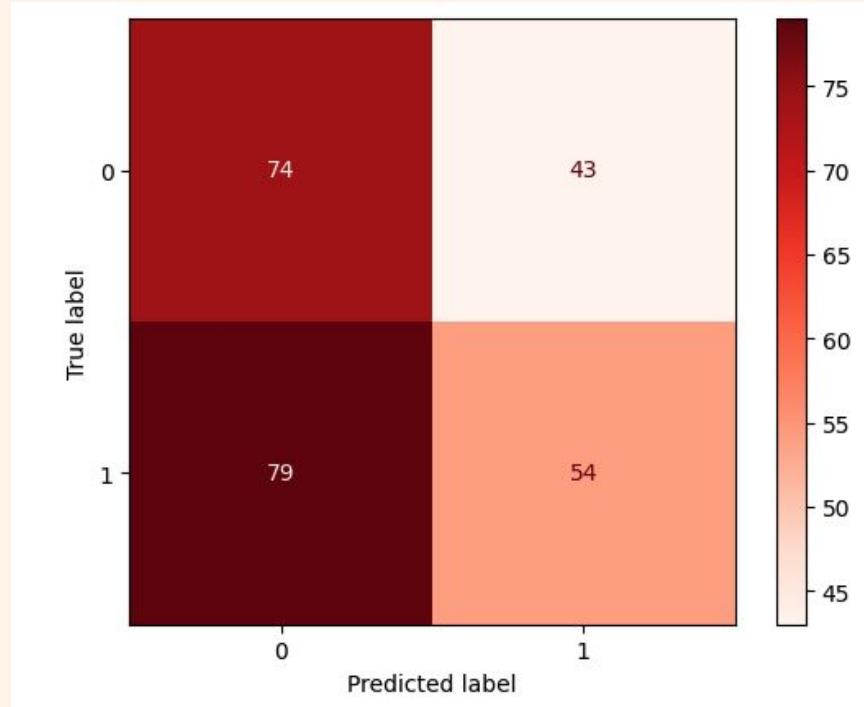
Using Classification Model to Make Predictions!

Baseline Accuracy Score for predicting hair loss: 50.25%. Our model must beat this baseline score!



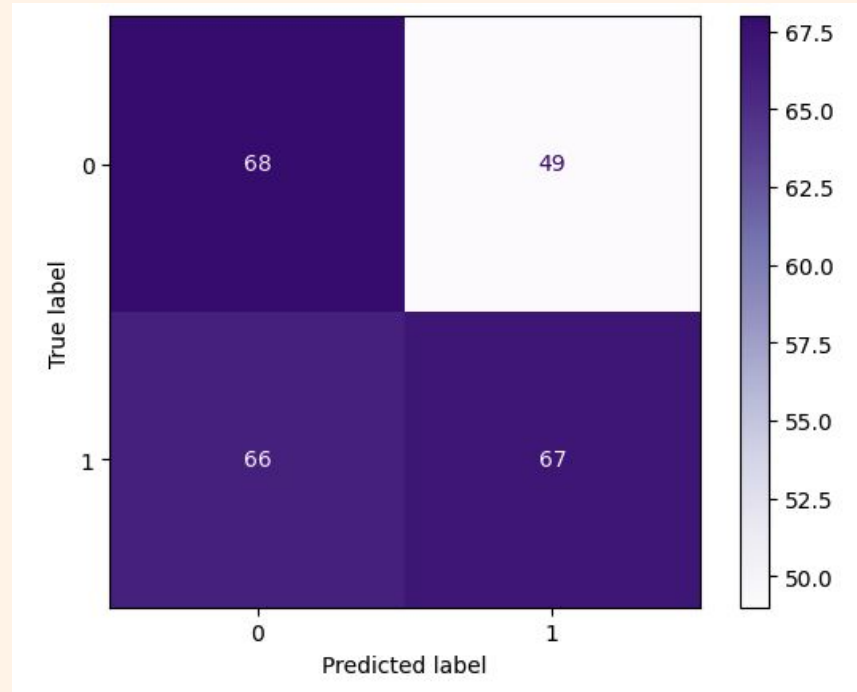
Making Predictions using Logistic Regression

Using Logistic Regression, our model was able to beat the baseline accuracy score, with a score of 51.2%!



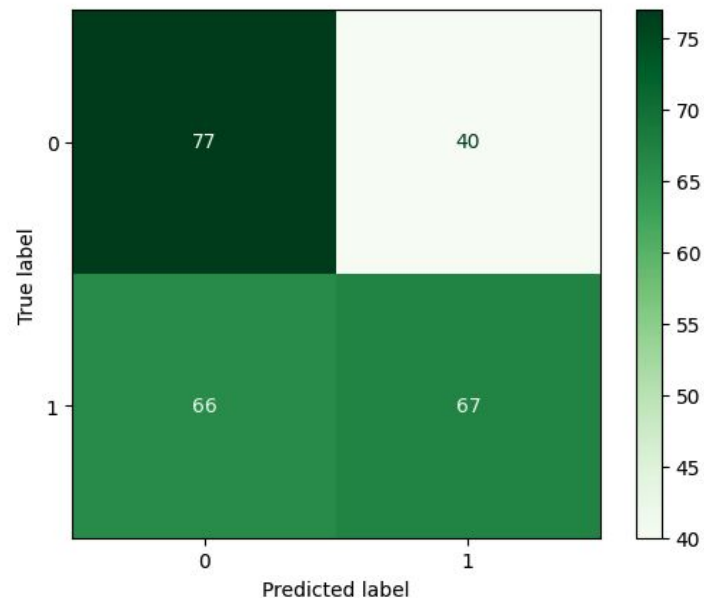
Making Predictions using Random Forest

Using Random Forest, our model was able to beat the baseline accuracy score, with a score of 54%



Making Predictions using KNN

Using KNN, our model was able to best beat the baseline accuracy score, with a KNN score of 57.6% when $k = 9$!





Conclusion

From the data collected, we were able to visualize and compare multiple features of hair health to get a better understanding of they can or cannot affect the result of hair loss. Using the data in our dataframe to create different models, we are able to use models that perform better than our baseline. Our original baseline accuracy score was calculated to be 50.25% accurate at predicting the presence of hair loss in individuals.

Using the models we created we were able to beat that baseline score with the highest score being 57.6 using the KNN Model. This proves that our KNN model is better at predicting hair loss and will be useful to those individuals who are interested in understanding what contributes to experiencing hair loss and may be useful to those who want to prevent hair loss when visualizing the medical conditions, medications/treatments, nutritional deficiencies as well as the habits that impact hair loss.

Thanks!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

