

```
In [8]: import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
import os

In [10]: hair = pd.read_csv('PredictHairFall.csv')
hair.head()

Out[10]:
```

	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

```
In [5]: hair = pd.read_csv('PredictHairFall.csv')
hair.head()
#hair.query('Age == 30')
#hair.query(['Hormonal Changes'] == "Yes")
#plt.hist(hair['Environmental Factors'])
hair = hair.rename(columns={'Hormonal Changes': 'Hormonal_Changes'})
hair = hair.rename(columns={'Hair Loss': 'Hair_Loss'})

#print(hair.Hormonal_Changes)

#print (hair)
#hair.head()

#hair.query('Hormonal_Changes == "No"')

-----
NameError                                Traceback (most recent call last)
Cell In[5], line 1
----> 1 hair = pd.read_csv('PredictHairFall.csv')
      2 hair.head()
      3 #hair.query('Age == 30')
      4 #hair.query(['Hormonal Changes'] == "Yes")
      5 #plt.hist(hair['Environmental Factors'])

NameError: name 'pd' is not defined

In [302]: hair = hair.rename(columns={'Medical Conditions': 'Medical_Conditions'})
hair.query('Medical_Conditions == "Eczema" and Stress == "High"')

EHquery = hair.query('Medical_Conditions == "Eczema" and Stress == "High"')
EHquery['Hair_Loss'] = EHquery['Hair_Loss'].astype(str)

EHquery['Hair_Loss'].replace('1', 'Yes', inplace=True)
EHquery['Hair_Loss'].replace('0', 'No', inplace=True)
print (EHquery)
```

	Id	Genetics	Hormonal_Changes	Medical_Conditions	\
1	148393	No	No	Eczema	
10	118958	Yes	Yes	Eczema	
14	150154	Yes	No	Eczema	
256	133865	No	No	Eczema	
276	130779	No	Yes	Eczema	
290	119252	No	Yes	Eczema	
416	136076	No	Yes	Eczema	
421	149565	No	No	Eczema	
447	192088	No	Yes	Eczema	
543	121286	Yes	Yes	Eczema	
579	184114	Yes	No	Eczema	
758	131979	No	No	Eczema	
830	158275	Yes	Yes	Eczema	
875	132040	No	No	Eczema	
896	160757	Yes	Yes	Eczema	
907	181021	Yes	Yes	Eczema	
918	116831	Yes	Yes	Eczema	
932	148041	No	No	Eczema	
958	149642	Yes	No	Eczema	

	Medications & Treatments	Nutritional Deficiencies	Stress	Age	\
1	Antibiotics	Magnesium deficiency	High	43	
10	Blood Pressure Medication	Biotin Deficiency	High	26	
14	Antibiotics	Biotin Deficiency	High	34	
256	Antidepressants	Vitamin D Deficiency	High	48	
276	Antibiotics	Zinc Deficiency	High	37	
290	Rogaine	Vitamin E deficiency	High	24	
416	Blood Pressure Medication	Vitamin A Deficiency	High	45	
421	Antibiotics	Protein deficiency	High	36	
447	Rogaine	Magnesium deficiency	High	48	
543	Antidepressants	Iron deficiency	High	36	
579	Antibiotics	Biotin Deficiency	High	26	
758	Heart Medication	Vitamin E deficiency	High	31	
830	Blood Pressure Medication	Zinc Deficiency	High	34	
875	Antidepressants	Protein deficiency	High	33	
896	Antidepressants	Protein deficiency	High	43	
907	Accutane	Protein deficiency	High	45	
918	Accutane	Biotin Deficiency	High	48	
932	Heart Medication	Protein deficiency	High	31	
958	Heart Medication	Biotin Deficiency	High	35	

	Poor Hair Care Habits	Environmental Factors	Smoking	Weight Loss	\
1	Yes	Yes	No	No	
10	Yes	Yes	Yes	No	
14	Yes	Yes	No	Yes	
256	No	Yes	Yes	Yes	
276	No	Yes	No	No	
290	No	Yes	Yes	No	
416	No	No	No	No	
421	Yes	No	No	Yes	
447	Yes	Yes	Yes	Yes	
543	No	Yes	Yes	Yes	
579	Yes	Yes	No	No	
758	Yes	No	No	Yes	
830	Yes	Yes	No	Yes	
875	Yes	No	Yes	No	
896	Yes	No	Yes	No	
907	Yes	No	No	No	
918	No	Yes	Yes	Yes	
932	Yes	No	No	Yes	
958	No	Yes	No	No	

	Hair_Loss
1	No
10	No
14	No
256	No
276	No
290	No
416	No
421	Yes
447	Yes
543	No
579	Yes
758	Yes
830	No
875	Yes
896	Yes
907	No
918	Yes
932	Yes
958	Yes

C:\Users\preet\AppData\Local\Temp\ipykernel_4824\3479039015.py:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
EHquery['Hair_Loss'] = EHquery['Hair_Loss'].astype(str)
C:\Users\preet\AppData\Local\Temp\ipykernel_4824\3479039015.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

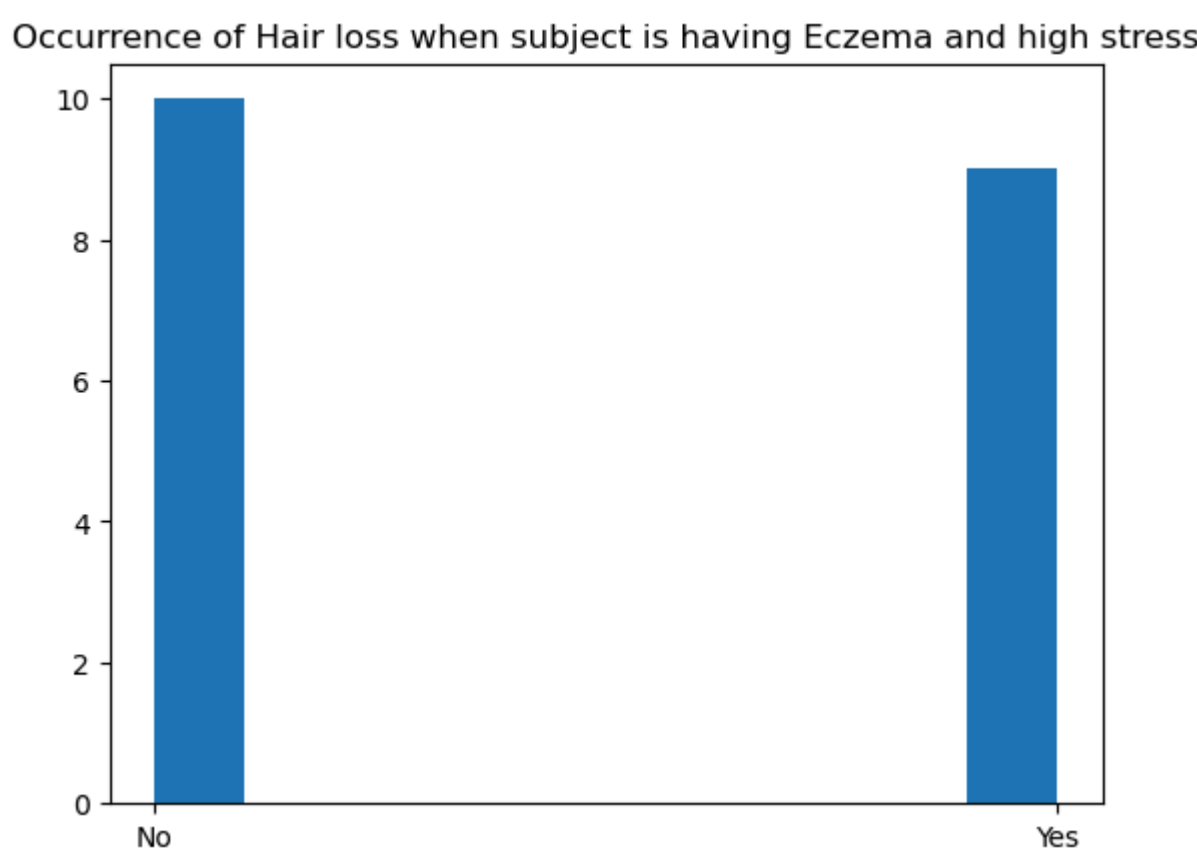
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
EHquery['Hair_Loss'].replace('1', 'Yes', inplace=True)
C:\Users\preet\AppData\Local\Temp\ipykernel_4824\3479039015.py:8: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
EHquery['Hair_Loss'].replace('0', 'No', inplace=True)

```
In [324]: plt.hist(EHquery.Hair_Loss)
plt.title('Occurrence of Hair loss when subject is having Eczema and high stress')
plt.figure(figsize=(55, 5))

#It seems having eczema and a high stress life does not affect hair loss that much

Out[324]: <Figure size 5500x500 with 0 Axes>
```



```
<Figure size 5500x500 with 0 Axes>

In [30]: #Creating a pie chart to show the stress levels of those with Hair Loss

hair = pd.read_csv('PredictHairFall.csv')
hair = hair.rename(columns={'Hair Loss': 'Hair_Loss'})
hair['Hair_Loss'] = hair['Hair_Loss'].astype(str)

hair['Hair_Loss'].replace('1', 'Yes', inplace=True)
hair['Hair_Loss'].replace('0', 'No', inplace=True)
#print (hair)
YesHigh = hair.query('Hair_Loss == "Yes" and Stress == "High").count()[0] #156
YesModerate = hair.query('Hair_Loss == "Yes" and Stress == "Moderate").count()[0] #182
YesLow = hair.query('Hair_Loss == "Yes" and Stress == "Low").count()[0] #159
#print (YesHigh)
#print (YesModerate)
#print (YesLow)
NoHigh = hair.query('Hair_Loss == "No" and Stress == "High").count()[0] #165
NoModerate = hair.query('Hair_Loss == "No" and Stress == "Moderate").count()[0] #169
NoLow = hair.query('Hair_Loss == "No" and Stress == "Low").count()[0] #168
#print (NoHigh)
#print (NoModerate)
print (NoLow)
#EHquery = hair.query('Medical_Conditions == "Eczema" and Stress == "High"')

#Psoriasis = hair.loc[hair['Medical_Conditions'] == 'Psoriasis'].count()[0]
#Eczema = hair.loc[hair['Medical_Conditions'] == 'Eczema'].count()[0]

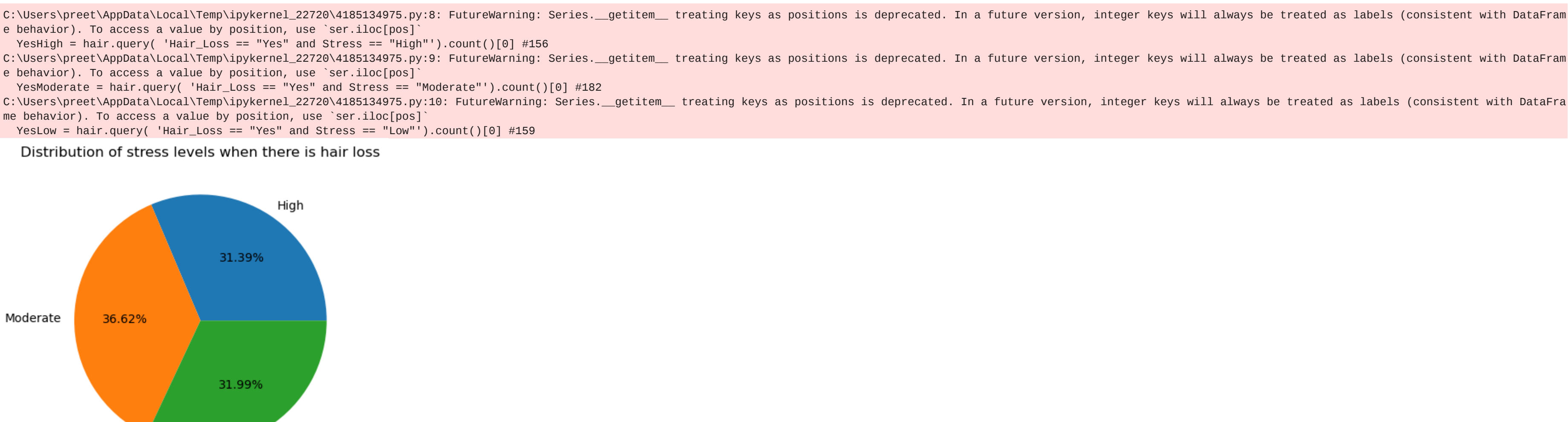
#print(Psoriasis)
#print(Eczema)

168

C:\Users\preet\AppData\Local\Temp\ipykernel_22720\878492342.py:10: FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a future version, integer keys will always be treated as labels (consistent with DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
YesHigh = hair.query('Hair_Loss == "Yes" and Stress == "High").count()[0] #156
C:\Users\preet\AppData\Local\Temp\ipykernel_22720\878492342.py:11: FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a future version, integer keys will always be treated as labels (consistent with DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
YesModerate = hair.query('Hair_Loss == "Yes" and Stress == "Moderate").count()[0] #182
C:\Users\preet\AppData\Local\Temp\ipykernel_22720\878492342.py:12: FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a future version, integer keys will always be treated as labels (consistent with DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
YesLow = hair.query('Hair_Loss == "Yes" and Stress == "Low").count()[0] #159
C:\Users\preet\AppData\Local\Temp\ipykernel_22720\878492342.py:16: FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a future version, integer keys will always be treated as labels (consistent with DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
NoHigh = hair.query('Hair_Loss == "No" and Stress == "High").count()[0] #165
C:\Users\preet\AppData\Local\Temp\ipykernel_22720\878492342.py:17: FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a future version, integer keys will always be treated as labels (consistent with DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
NoModerate = hair.query('Hair_Loss == "No" and Stress == "Moderate").count()[0] #169
C:\Users\preet\AppData\Local\Temp\ipykernel_22720\878492342.py:18: FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a future version, integer keys will always be treated as labels (consistent with DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
NoLow = hair.query('Hair_Loss == "No" and Stress == "Low").count()[0] #169

In [45]: hair = pd.read_csv('PredictHairFall.csv')
hair = hair.rename(columns={'Hair Loss': 'Hair_Loss'})
hair['Hair_Loss'] = hair['Hair_Loss'].astype(str)

hair['Hair_Loss'].replace('1', 'Yes', inplace=True)
hair['Hair_Loss'].replace('0', 'No', inplace=True)
#print (hair)
YesHigh = hair.query('Hair_Loss == "Yes" and Stress == "High").count()[0] #156
YesModerate = hair.query('Hair_Loss == "Yes" and Stress == "Moderate").count()[0] #182
YesLow = hair.query('Hair_Loss == "Yes" and Stress == "Low").count()[0] #159
#print (YesHigh)
#print (YesModerate)
#print (YesLow)
plt.title('Distribution of stress levels when there is hair loss')
labels = ['High', 'Moderate', 'Low']
colors = ['Red', 'Yellow', 'Green']
plt.pie([YesHigh, YesModerate, YesLow], labels = labels, autopct = '%.2f%%')
plt.show()
```



```
In [51]: hair = pd.read_csv('PredictHairFall.csv')
hair = hair.rename(columns={'Hair Loss': 'Hair_Loss'})
hair['Hair_Loss'] = hair['Hair_Loss'].astype(str)

hair['Hair_Loss'].replace('1', 'Yes', inplace=True)
hair['Hair_Loss'].replace('0', 'No', inplace=True)
#print (hair)
NoHigh = hair.query('Hair_Loss == "No" and Stress == "High").count()[0] #165
NoModerate = hair.query('Hair_Loss == "No" and Stress == "Moderate").count()[0] #169
NoLow = hair.query('Hair_Loss == "No" and Stress == "Low").count()[0] #168

plt.title('Distribution of stress levels when there is no hair loss')
labels = ['High', 'Moderate', 'Low']
colors = ['Red', 'Yellow', 'Green']
plt.pie([NoHigh, NoModerate, NoLow], labels = labels, autopct = '%.2f%%')
plt.show()
```

C:\Users\preet\AppData\Local\Temp\ipykernel_22720\1913885946.py:8: FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a future version, integer keys will always be treated as labels (consistent with DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
NoHigh = hair.query('Hair_Loss == "No" and Stress == "High").count()[0] #165
C:\Users\preet\AppData\Local\Temp\ipykernel_22720\1813885946.py:9: FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a future version, integer keys will always be treated as labels (consistent with DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
NoModerate = hair.query('Hair_Loss == "No" and Stress == "Moderate").count()[0] #182
C:\Users\preet\AppData\Local\Temp\ipykernel_22720\1813885946.py:10: FutureWarning: Series.__getitem__ treating keys as positions is deprecated. In a future version, integer keys will always be treated as labels (consistent with DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
NoLow = hair.query('Hair_Loss == "No" and Stress == "Low").count()[0] #168

Distribution of stress levels when there is no hair loss

