```
import pandas as pd
import numpy as np
import scipy.stats as stats
import statsmodels.api as sm
import seaborn as sns
import matplotlib.pyplot as plt

data = pd.read_csv('./Q 7- Copy.csv')

# View the first few rows of the dataset
# print(data.head())
data.head()

# Select numerical columns for analysis
# numerical_data = data.select_dtypes(include=np.number)
# numerical_data.head()
```

	Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	NaN
1	2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	NaN
2	3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	NaN
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea

```
# Importing necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.decomposition import PCA

# Load the dataset (assuming the dataset is in CSV format)
# Replace 'dataset.csv' with the actual file path
df = pd.read_csv('./Q 7- Copy.csv')

# Display the first few rows of the dataset
print(df.head())

# Data Preprocessing
# Handle missing values by filling with median or mean (for simplicity, using mean here)
# df.fillna(df.mean(), inplace=True)

# Encode categorical variables
label_encoders = {}
for column in ['Gender', 'Sleep Duration']:
    le = LabelEncoder()
    df[column] = le.fit_transform(df[column])
    label_encoders[column] = le

# Standardize numerical columns
numerical_features = ['Age', 'Sleep Duration', 'Physical Activity Level', 'Stress Level', 'Quality of Sleep']
scaler = StandardScaler()
df[numerical_features] = scaler.fit_transform(df[numerical_features])

# Apply PCA
pca = PCA(n_components=2)  # We reduce to 2 components for visualization purposes
```

```
pca_components = pca.fit_transform(df[numerical_features])

# Create a DataFrame with PCA components
pca_df = pd.DataFrame(data=pca_components, columns=['PC1', 'PC2'])

# Visualize the explained variance ratio
plt.figure(figsize=(8, 5))
plt.plot(range(1, len(pca.explained_variance_ratio_) + 1), pca.explained_variance_ratio_, marker='o', linestyle='--')
plt.title('Explained Variance Ratio')
plt.xlabel('Principal Components')
plt.ylabel('Variance Explained')
plt.show()

# Scatter plot of the PCA results
plt.figure(figsize=(10, 7))
sns.scatterplot(x='PC1', y='PC2', data=pca_df, hue=df['Gender'], palette='viridis', alpha=0.6)
plt.title('PCA of Sleep Health Factors')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.show()

# Display PCA components (loading scores)
components_df = pd.DataFrame(data=pca.components_, columns=numerical_features, index=['PC1', 'PC2'])
print("\nPCA Components:")
print(components_df)

# Interpretation of findings
# The PCA components can be interpreted by examining the loading scores to see which features contribute most to each comp
# For example, higher values in 'PC1' might indicate the importance of sleep duration and stress level.
# Higher values in 'PC2' could highlight physical activity and caffeine intake.

# Additional: Saving the PCA results for further analysis
pca_df['Gender'] = df['Gender']
# pca_df.to_csv('pca_results.csv', index=False)
```

```
   Person ID Gender  Age           Occupation  Sleep Duration  \
0          1   Male   27    Software Engineer             6.1
1          2   Male   28               Doctor             6.2
2          3   Male   28               Doctor             6.2
3          4   Male   28 Sales Representative             5.9
4          5   Male   28 Sales Representative             5.9

   Quality of Sleep  Physical Activity Level  Stress Level BMI Category  \
0                 6                       42             6   Overweight
1                 6                       60             8       Normal
2                 6                       60             8       Normal
3                 4                       30             8        Obese
4                 4                       30             8        Obese

  Blood Pressure  Heart Rate  Daily Steps Sleep Disorder
0         126/83          77         4200            NaN
1         125/80          75        10000            NaN
2         125/80          75        10000            NaN
3         140/90          85         3000    Sleep Apnea
4         140/90          85         3000    Sleep Apnea
```