Execute Summary :

The purpose of the project is to analyze factors that affect sleep health and predict which sleep disorder a person can have based on the other factors such as Sleep Quality, Stress Level, BMI Index, Blood Pressure, and etc. I have used the dataset from kaggle. I have performed Data preprocessing by filling out null values in the dataset.

Performed Exploratory Data Analysis (EDA) on the dataset and made following observations :

- Individuals with Age greater than 43 years are more likely to experience sleep disorders.
- Females tend to have greater sleep quality on average compared to Males.
- Individuals with Engineer occupation have more average sleep quality and Individuals with Sales Representative occupation have less average sleep quality.
- Individuals with higher stress levels tend to have more chances of having a sleep disorder.
- Individuals with Sleep disorder have less quality of sleep rating compared to individuals with No disorder
- Individuals in Obese and Overweight BMI category tend to experience more sleep disorders.
- Individuals have sleep duration over 7 hours have very less chance to get have a sleep disorder

After performing EDA, I have splitted my dataset in 3 : 1 ratio for training set and testing set.
I have used three models: Logistic Classification, KNN Classifier, Random Forest. I have used the same training set and testing set for these three models. After testing the predictions of these three models, Random forest classifier was a bit better compared to K-Nearest Neighbors classifiers. These models outperformed logistic classifiers which were only able to give 86% accuracy. Therefore, Random Forest classifier is the best model for this chosen dataset with 89% accuracy. Random Forest was able to achieve following scores :

Accuracy: 89%, Recall: 89%, Precision: 90% and F1-Score: 89%

After the evaluation, based on Random Forest Classifier I have identified important features to detect Sleep disorders of individuals. The Top 3 important features according to Random Forest are **Blood Pressure, BMI Category and Age.**

In conclusion, I was able to find that Random Forest worked the best to predict the target variable i.e., what sleep order does an individual has and found the Top three features contributing to prediction those are Blood Pressure, BMI Category and Age