1. CustomerID

Unique customer ID

1. ProdTaken

ProdTaken is our variable which we want to predict. It can be 0 (Wellness Tourism Package not taken) or 1 (Wellness Tourism Package taken)

1. Age  
   Age of customer
2. TypeofContact  
   How customer was contacted (Company Invited or Self Inquiry)
3. CityTier

City tier depends on the development of a city, population, facilities, and living standards. The categories are ordered i.e. Tier 1 > Tier 2 > Tier 3

1. DurationOfPitch

Duration of the pitch by a salesperson to the customer

1. Occupation

Occupation of customer

1. Gender

Gender of customer

1. NumberOfPersonVisiting

Total number of persons planning to take the trip with the customer

1. NumberOfFollowups

Total number of follow-ups has been done by the salesperson after the sales pitch

1. ProductPitched

Product pitched by the salesperson

1. PreferredPropertyStar

Preferred hotel property rating by customer

1. MaritalStatus

Marital status of customer

1. NumberOfTrips

Average number of trips in a year by customer

1. Passport

The customer has a passport or not (0: No, 1: Yes)

1. PitchSatisfactionScore

Sales pitch satisfaction score

1. OwnCar

Whether the customers own a car or not (0: No, 1: Yes)

1. NumberOfChildrenVisiting

Total number of children with age less than 5 planning to take the trip with the customer

1. Designation

Designation of the customer in the current organization

1. MonthlyIncome

Gross monthly income of the customer

# Korelasi Pearson

pearson\_corr = df\_raw[['Age', 'MonthlyIncome']].corr(method='pearson')

print("Korelasi Pearson:")

print(pearson\_corr)

# Korelasi Spearman

spearman\_corr = df\_raw[['Age', 'MonthlyIncome']].corr(method='spearman')

print("Korelasi Spearman:")

print(spearman\_corr)

from scipy.stats import mannwhitneyu

# Misalnya, 'Age' sebagai fitur numerik

result = mannwhitneyu(df\_raw[df\_raw['ProdTaken'] == 'Taken']['Age'],

df\_raw[df\_raw['ProdTaken'] == 'Not Taken']['Age'])

print(f"U Statistic: {result.statistic}, P-Value: {result.pvalue}")

from scipy.stats import f\_oneway

# Misalnya, 'Age' sebagai fitur numerik

f\_statistic, p\_value = f\_oneway(df\_raw[df\_raw['ProdTaken'] == 'Taken']['Age'],

df\_raw[df\_raw['ProdTaken'] == 'Not Taken']['Age'])

print(f"F-Statistic: {f\_statistic}, P-Value: {p\_value}")

from scipy.stats import chi2\_contingency

contingency\_table = pd.crosstab(df\_raw['TypeofContact'], df\_raw['ProdTaken'])

chi2, p, dof, expected = chi2\_contingency(contingency\_table)

print(f"Chi-Square: {chi2}, P-Value: {p}")