**Spark Use Case – Travel Data Analysis**

**Travel Sector Dataset Description**

Column 1: City pair (Combination of *from* and *to*): String

Column 2: From location: String

Column 3: To Location: String

Column 4: Product type: Integer (1=Air, 2=Car, 3 =Air+Car, 4 =Hotel, 5=Air+Hotel, 6=Hotel +Car, 7 =Air+Hotel+Car)

Column 5: Adults traveling: Integer

Column 6: Seniors traveling: Integer

Column 7: Children traveling: Integer

Column 8: Youth traveling: Integer

Column 9: Infant traveling: Integer

Column 10: Date of travel: String

Column 11: Time of travel: String

Column 12: Date of Return: String

Column 13: Time of Return: String

Column 14: Price of booking: Float

Column 15: Hotel name: String

**Problem Statement 1**

Top 20 destination people travel the most: Based on the given data, we can find the most popular destination that people travel frequently. There are many destinations out of which we will find only first 20, based on trips booked for particular destinations.

**PROBLEM2: Top** 20 locations from where people travel the most: We can find the places from where most of the trips are undertaken, based on the booked trip count.

**Problem Statement 3**

Top 20 cities that generate high airline revenues for travel, so that the site can concentrate on offering discount on booking, to those cities to attract more bookings.