

# Analysis of Travel data

Nandini Goswami  
Indiana University  
Bloomington  
F16-IG-3007  
goswamin@uemail.iu.edu

Sarita Bhateja  
Indiana University  
Bloomington  
F16-IG-3003  
sbhateja@uemail.iu.edu

Kavya Guruprasad  
Indiana University  
Bloomington  
F16-IG-3008  
prasadk@uemail.iu.edu

## ABSTRACT

The analysis of travel data will be done in order to gain insights to data and help the travel and hospitality industry to offer various benefits to the customers, increase marketing, strategic decisions etc. It will help the industry to better understand their target audience.

## 1. INTRODUCTION

The project is to analyse the travel data which includes data such as destination city, source city, mode of commute, the hotel they will be staying at, type of audience such as youth, children, adults etc. This project will prove helpful for both the hospitality industry and customers.

## 2. PROBLEM STATEMENT

Based on the data, we will find the top 10 destinations people travel to, location where they travel from and duration of the year when people travel to such destinations the most.

## 3. ARTIFACTS TO BE CREATED

Project phase and Tasks to be done are as follows:-

-> Initiating- In the initiating phase we will be storing the data in database and preprocessing data. We will be cleaning and formatting data.

->Analysis and Development- In this phase we will be analyzing the dataset and extracting and scaling the important features. After data pruning and feature selection, we will start extracting the necessary information. Data processing will be done using Spark and programming language will be python.

->Testing and enhancements- After the completion of implementation of algorithm, we will be testing with different datasets. Depending on the test results, if required we will do bug fixing and enhancements.

->Closing- A successfully running program with the desired results.

## 4. TEAM

We will be working in group of three namely Nandini Goswami (F16-IG-3007), Sarita Bhateja (F16-IG-3003) and Kavya Guruprasad (F16-IG-3008).

## 5. TIMELINE

-> Initiating:- We will be initially doing data processing till November 22.

->Analysis and Development:- We will be doing analysis and development in the month of October and finish it till November 27.

->Testing and enhancements:- We will start our testing and do enhancements if required.

->Closing :- Our entire project will be completed by December 1.

[1]

## 6. REFERENCES

[1] Use cases of big data. web page.