A Sample *ACM* SIG Proceedings Paper in LaTeX

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

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

Kavya Guruprasad
F16-IG-3008
Indiana University
Bloomington
Address Three
prasadk@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. DATASET DESCRIPTION
The dataset was taken from:-[1] Following are the attributes of the dataset:-

- **City Pair** : The origin-destination pair of travel (STRING)
- **From Location** : The origin of travel (STRING)
- **To Location** : The destination of travel (STRING)
- **Product Type** : This attribute tells us about the mode of travel and the
- **Combinations** ->Here 1 represents Air, ->2 represents car, ->3 represents Air+Car, ->4 represents Hotel, ->5 represents Air+Hotel, ->6 represents Hotel+Car, ->7 represents Air+Hotel+Car) (INTEGER)
- **Adults Traveling** : Number of adults travelling (INTEGER)
- **Seniors Traveling** : Number of seniors travelling (INTEGER)
- **Children Traveling** : Number of children travelling (INTEGER)
- **Youth Traveling** : Number of youth travelling (INTEGER)
- **Infant Traveling** : Number of infants travelling (INTEGER)
- **Date of Travel** : The date of travel(STRING)
- **Time of Travel** : The time of travel(STRING)
- **Date of Return** : The date of return (STRING)
- **Time of Return** : The time of return (STRING)
- **Price of booking** : The price of booking hotel (FLOAT)
- **Airline code** : The code with which each airlines is denoted (STRING)
- **Airline** : The airline name (STRING)
- **Hotel** :The hotel name (STRING)

3. DATA ANALYTICS
We have done analysis on this dataset by writing some queries.The inspiration of this project was taken from:- [2]. We have implemented this using Apache Spark and Map reduce and used python as the language for writing the code. We have tried to find the top 20 destinations that people travel to, from and the most frequently visited city pairs where people prefer travelling. We have printed the top 20 places in descending order. We have found the top 10 cities that generate high airline revenues for travel. From this analysis airline companies can give better deals and make better profit. Top source-destination sites that people prefer travelling is also found. This can help airlines give better deals for in these routes. We also found The most often visited hotels by travellers, what combination of traveling do people prefer(Air+Hotel or Hotel+Car or Air+Hotel+Car), Top places from where people generally travel by car, Which airlines do people prefer the most. This kind of analysis is especially helpful to the travel industry to make wise decisions in making their costumer's stay a better one.

4. DATABRICKS
Instructions on how to import Databricks notebook and run the code:-

1) Go to https://community.cloud.databricks.com and Sign Up.

2) Select the Community Edition.

3) Sign Up for Databricks Community Edition by entering personal details.

4) Select Workspaces from the left-hand column, and under workspaces, choose shared and select import under the dropdown. 5) Create Clusters by choosing the clusters in the left-hand column, click on the create cluster tab and enter the cluster name and Apache Spark Version Spark 2.0(Auto-Updating, Scala 2.0)

5. REFERENCES

[1] Google Drive.

[2] K. K. Innamuri. Spark use case âĂŞ travel data analysis. Webpage, May 2016.