Statistical & Sentiment Analysis of Factors Affecting State-wise Tourism in India

(Data Mining and Warehousing Project)

Made By:

Akshay Nagpal Akhil Atreya

12-CSU-022 12-CSU-018

Supervised By:

Dr. Latika Singh

CSE & IT Department

THE NORTHCAP UNIVERSITY (Formerly ITM University, Gurgaon)

Contents

[Introduction 3](#_Toc435285214)

[Data Mining Objectives 4](#_Toc435285215)

[Project Planning and Execution 5](#_Toc435285216)

[Data Understanding 6](#_Toc435285217)

[Data Exploration 7](#_Toc435285218)

[Descriptive Data Analysis 8](#_Toc435285219)

[Sentiment Analysis of Twitter Data 9](#_Toc435285220)

# Introduction

# Data Mining Objectives

4 main objectives are:

1. To plot the graph of each state and union territory of India showing tourism, crime rate and number of hotel rooms between 2003-2013 and find the visual relationship between them.
2. To plot the above parameters for each state for every year from 2003-2013 and analyze how the states compare against each other for every year.
3. To perform regression analysis to quantify the dependence of crime and infrastructure on tourism.
4. To perform sentiment analysis of tweets from 2003-2013 for the state showing most positive trends.

# Project Planning and Execution

What we planned for and how we executed it with all the problems faced etc.

# Data Understanding

Data was collected from 3 locations:

1. [Open Government Data(OGD) Platform India](https://data.gov.in/catalogs/ministry_department/national-crime-records-bureau-ncrb)

From the above site we collected the crime statistics from 2003-2013

1. [Ministry of Tourism](http://tourism.gov.in/market-research-and-statistics)

From the above website we collected publications in PDF format titled ‘Indian Tourism Statistics at a Glance’ for the year 2003-2013

1. [Kimono API](https://www.kimonolabs.com)

We used this API to collect tweets from 2003-2013 with hashtags relevant to our requirement.

# Data Exploration

Regarding the attributes discovered and decisions made such as considering foreign and domestic tourism together and other things with attributes.

# Descriptive Data Analysis

All the graphs and the inferences and regression etc.

# Sentiment Analysis of Twitter Data

The whole procedure and calculations etc.