

A REPORT OF HOTEL RECOMMENDATION ENGINE
At

CHANDIGARH UNIVERSITY

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD
OF THE DEGREE OF

BACHELOR OF ENGINEERING

(Computer Science & Engineering)



AUG-DEC, 2018

SUBMITTED BY:

ABHISHEK JOSHI

UNIVERSITY UID:-

16BCS3171

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
CHANDIGARH UNIVERSITY GHARUAN, MOHALI

CONTENTS

<i>Candidate's Declaration</i>	4
<i>Abstract</i>	5
<i>Acknowledgement</i>	6
<i>About the Company/ Industry / Institute</i>	7
<i>List of Figures</i>	8

CHAPTER 1 INTRODUCTION

1.1	SW/HW REQUIRED.....	9
1.2	BACKGROUND DETAILS.....	10

CHAPTER 2 WORK UNDERTAKEN

2.2	Coding.....	12
-----	-------------	----

CHAPTER 3 RESULTS AND DISCUSSION

3.1	Result.....	17
-----	-------------	----

CHAPTER 4 CONCLUSION AND FUTURE SCOPE

4.1	Conclusion.....	18
4.2	Future Scope.....	19

REFERENCES.....	20
-----------------	----

CANDIDATE’S DECLARATION

I “ABHISHEK JOSHI” uid 16bcs3171 hereby declare that during a period from AUG –DEC 2018 I sincerely worked on the project of the hotel recommendation engine in partial fulfillment of requirements for the award of degree of B.E (COMPUTER SCIENCE & ENGINEERING) at CHANDIGARH UNIVERSITY GHARUAN, MOHALI. The work which is being presented in the report submitted to Department of Computer Science & Engineering at CHANDIGARH UNIVERSITY GHARUAN, MOHALI is an authentic record of my work.

Signature of the Student

ABSTRACT

My Project is based on the Recommendation engine .The idea was to build a smart engine that best finds out the desired place for customer as a hotel and provide the best possible recommendation for the same.

Recommendation engine works on the principle of User rating along with the various special features provided i.e the various facilities provided by the Hotel

The User rating is the prior step for the hotel recommendation .

A recommendation engine, also known as a recommender system, is software that analyzes available data to make suggestions for something that a website user might be interested in, such as a book, a video or a job, among other possibilities.

Recommendation engines use a variety of technologies and techniques that enable them to filter large amounts of data and provide a smaller, focused body of suggestions for the user.

Recommendation engines use a variety of technologies and techniques that enable them to filter large amounts of data and provide a smaller, focused body of suggestions for the user. Netflix, for example, uses metadata tagging on videos in conjunction with data about user behavior to come up with recommended movies and TV shows for specific members.

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my teacher (Mr. Sunil Kumar) as well as my friend(Daksh Agarwal) who gave me the golden opportunity to do this wonderful project on the topic (Web Scrapping), which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to them. Secondly I would also like to thank my parents and friends who helped me a lot in finalizing this project within the limited time frame.

1. INTRODUCTION

1.1 SOFTWARE/HARDWARE DETAIL

1.1.1 HARDWARE DETAILS

1. 4 GB RAM
2. Laptop / Desktop with supported internet

1.1.2 SOFTWARE REQUIRED

1. Python 3.5.1
2. Sqlite
3. Anaconda 4.2
4. Apache Server

1.2 BACKGROUND OF THE PROJECT

(Recommendation Engine)

My Project is based on the Recommendation engine .The idea was to build a smart engine that best finds out the desired place for customer as a hotel and provide the best possible recommendation for the same.

Recommendation engine works on the principle of User rating along with the various special features provided i.e the various facilities provided by the Hotel

The User rating is the prior step for the hotel recommendation.

A recommendation engine, also known as a recommender system, is software that analyzes available data to make suggestions for something that a website user might be interested in, such as a book, a video or a job, among other possibilities.

Recommendation engines use a variety of technologies and techniques that enable them to filter large amounts of data and provide a smaller, focused body of suggestions for the user.

Recommendation engines use a variety of technologies and techniques that enable them to filter large amounts of data and provide a smaller, focused body of suggestions for the user. Netflix, for example, uses metadata tagging on videos in conjunction with data about user behavior to come up with recommended movies and TV shows for specific members.

LANGUAGE USED

1 .PYTHON:-

Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License (GPL). This tutorial gives enough understanding on Python programming language

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, and Unix shell and other scripting languages.

Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).

Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

2. Libraries used:-

1. **Numpy :- NumPy** is the fundamental package for scientific computing with **Python**. It contains among other things: a powerful N-dimensional array object. sophisticated (broadcasting) functions.
2. **Python Pandas** - DataFrame. A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns.
3. **Matplotlib :-** Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API **forembedding** plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+.

2.2 CODING (USING PYTHON)

Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License (GPL). This tutorial gives enough understanding on Python programming language

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, and Unix shell and other scripting languages.

The concepts and rules used in python programming provide these important benefits:

- Interactive
- Interpreted
- Modular
- Dynamic
- Object-oriented
- Portable
- High level
- Extensible in C++ & C

Analytics Report

1. Overview Of the Dataset :- The dataset used in the analytics for the Recommendation is the scrap data of the website make-mytrip. It had the 19 attributes with the data of 2500 users.

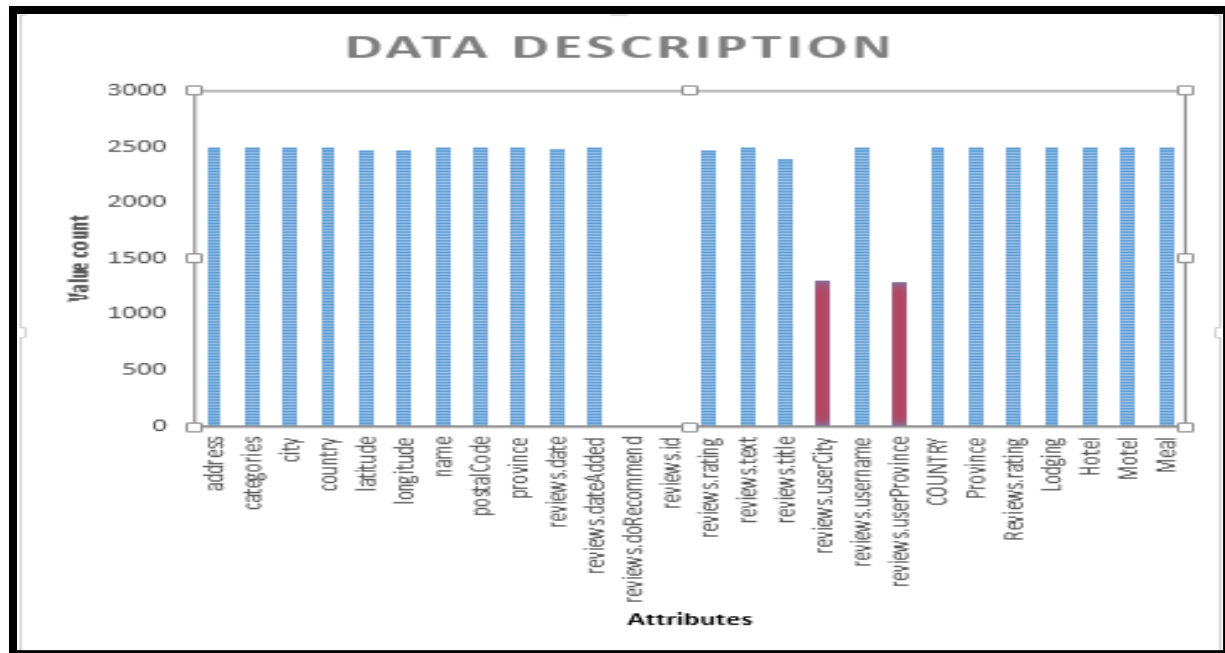
Main Attributes of the Dataset :-

1. **Address :-** defines the address of the each hotel
2. **Longitude and Latitude :-** defining the location
3. **Categories :-** defining various attires of the dataset
4. **Name :-** defining the name of the Hotel
5. **Postal code :-** defining the postal code to define the reign
6. **Province :-** defining the area
7. **review.rating :-** defining the user based rating

```
#data informmation
data.info()

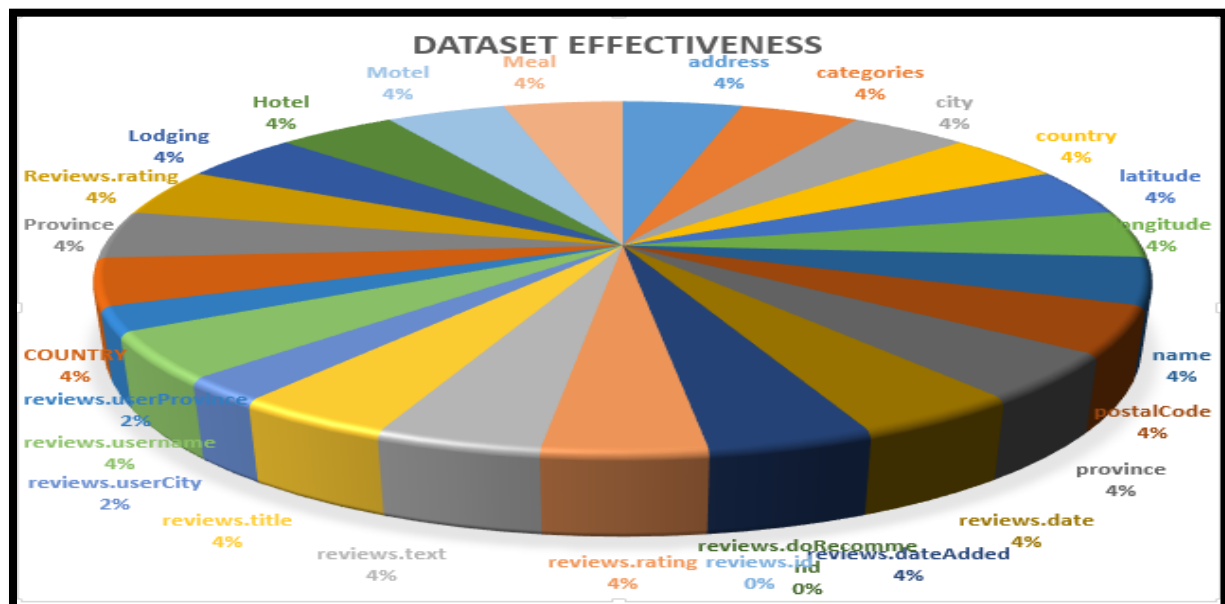
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2500 entries, 0 to 2499
Data columns (total 19 columns):
address                2500 non-null object
categories             2500 non-null object
city                   2500 non-null object
country                2500 non-null object
latitude               2474 non-null float64
longitude              2474 non-null float64
name                   2500 non-null object
postalCode             2500 non-null object
province               2500 non-null object
reviews.date           2483 non-null object
reviews.dateAdded      2500 non-null object
reviews.doRecommend    0 non-null float64
reviews.id             0 non-null float64
reviews.rating         2468 non-null float64
reviews.text           2500 non-null object
reviews.title          2385 non-null object
reviews.userCity       1307 non-null object
reviews.username       2499 non-null object
reviews.userProvince   1293 non-null object
```

DATA DESCRIPTION:- Below is the data description chart that is contrasting the availability of the data and also clarifying the missing data and the Null values.



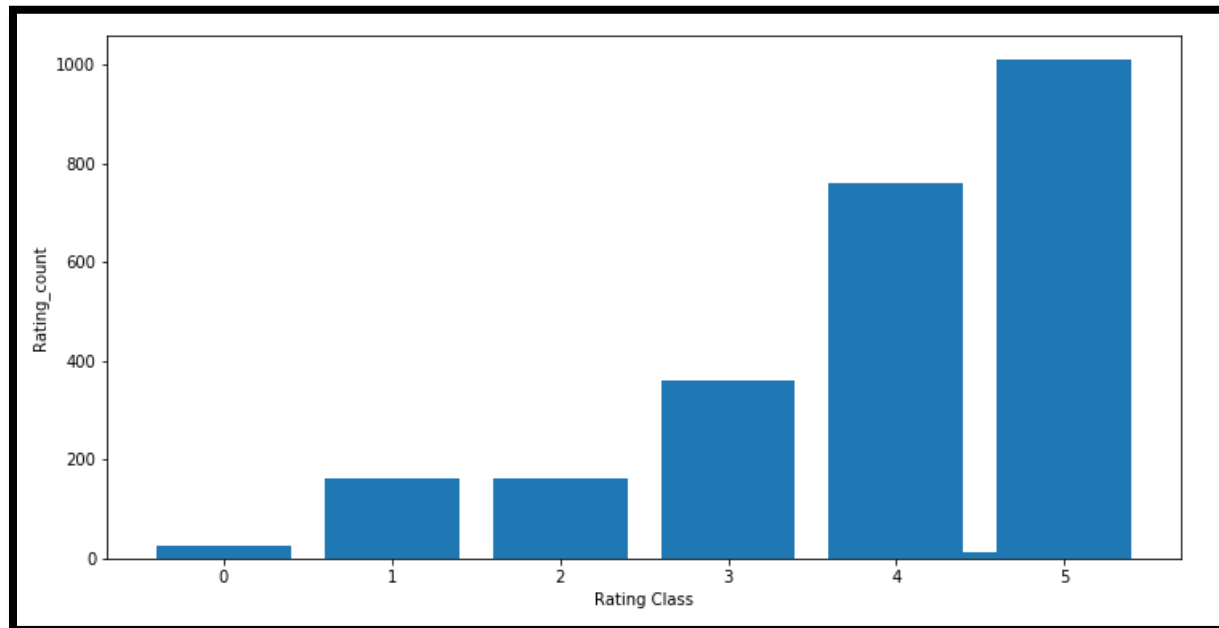
1. The empty graph shows the complete null values i.e reviews.dorecommend and reviews.id
2. The Red line in graph shows the values that have 50% missing value
3. Rest other either have complete values or same had to be filled with new values.

DATA EFFECTIVENESS:- Below is the pie chart that best defines the effectiveness of the data. The effectiveness can be calculated as more % of data is available for filtering and manipulation.



1. As we can clearly see in pie chart reviews.id and reviews.dorecommend contribute 0% so they must be deleted.
2. Rest of the attributes share a good extent to contribute in the analysis.

Closure look at Rating:-As rating is an important attribute because it is going to build a recommendation based on the user satisfaction so it must be analysed properly.Below is the rating class graph .There are 6 unique values in the rating:-0,1,2,3,4,5 .

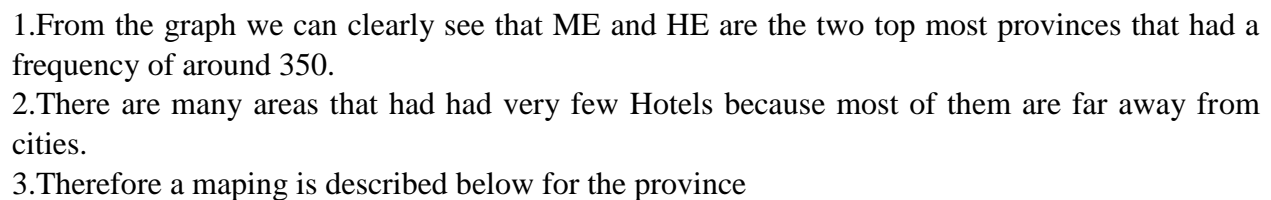


- 1.We can clearly see that some ratings are even going 0 that means there are certain Hotels that are untouched by the customers.
2. As for the 25% data is pointing to the rating of 2
3. For the 50% data it points to the rating of the 4
4. For the 70% data it is pointing to the rating of 5 as well.
- 5.Below are the numerical values stating prove too the above conclusion:-

count	2468.000000
mean	3.901499
std	1.248004
min	0.000000
25%	3.000000
50%	4.000000
75%	5.000000
max	5.000000

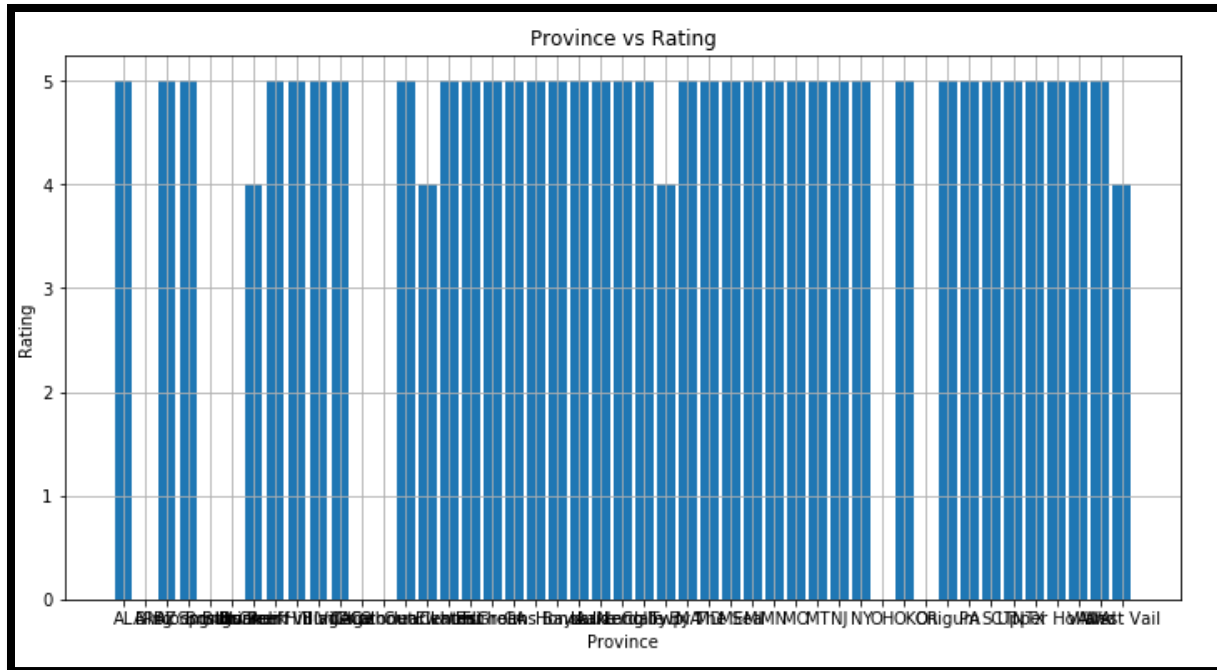
Therefore from above conclusion we can replace the two missing value by 4.0

Closure look at Province :- Province is the another important factor that as the whole filtering is done on the basis of **province** . Province has a unique 47 areas that are used for the recommend generation. Below is the bar chart describing the the count value of the unique provinces.



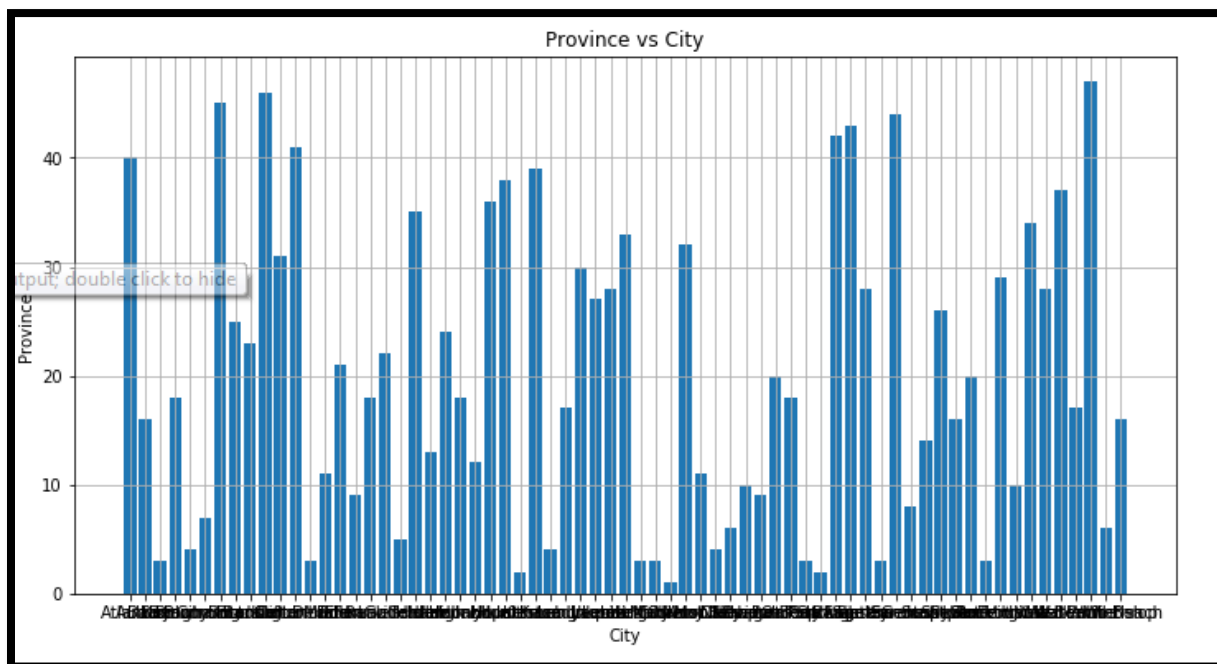
Binary conversion has been used instead of dummy variable so that to ensure the strength Reduction and even convert the oobject into the binary form.

Page 14 of 20

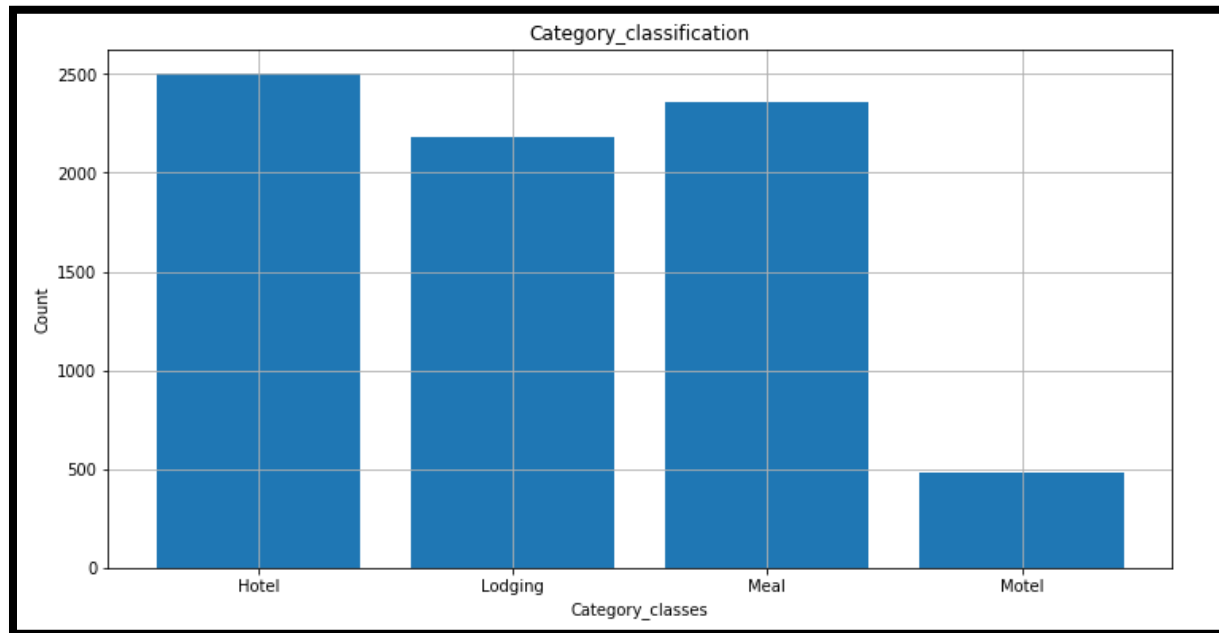


1. From the chart we can see that atleast each province is having a rating of the 4
2. It is for sure that if most of value correspond to the same value then generating recommendation is very difficult .
3. As learning from this we have to shift our filtering to little bit of content based.

Looking at City:-City is the another factor that can be used to find the exact location fo the recommendation .There are 67 unique cities combined with the 47 province .Below is the Description of province with the cities

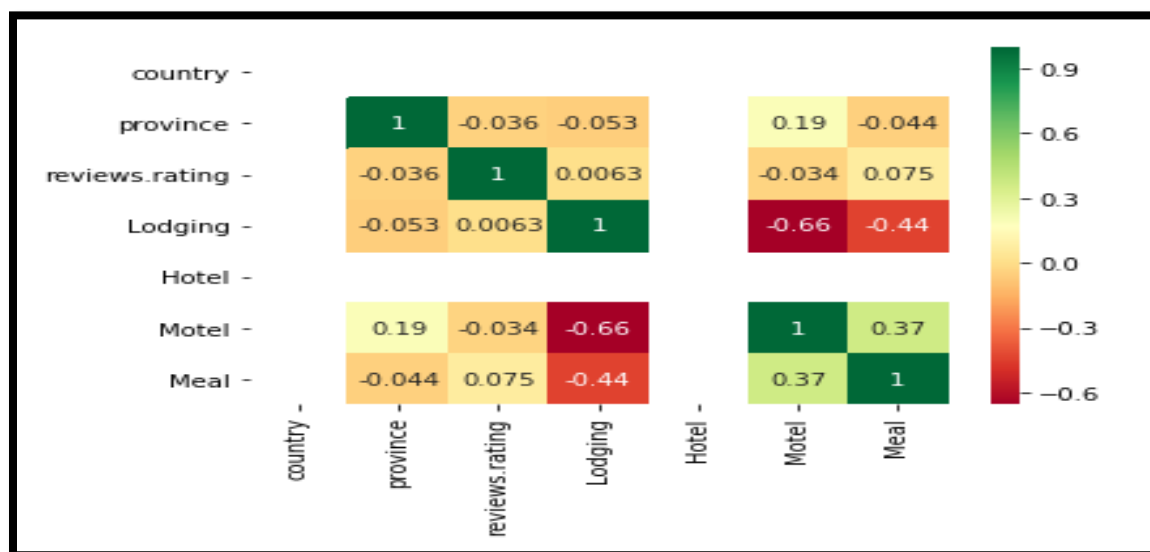


Introducing the categories :- Category attribute is the messy data that comprise of the various features of the Hotel . The topmost categories of the attribute are:- Hotels,Lodging,Motel,Meal .Below is the chart describing the classes of categories.



1. We can see that Hotel is a attribute that is present in 100% data
2. Lodging facility is provided by the 80% of the Hotels.
3. Meal provider Hotel are around 85%
4. Motel service is provided by the least number of the hotels i.e 20% .

Finding Co-relation :- As our object has been converted to the numerical oor integer data so it is best suited to find the corelation among the attributes.Below is the Heatmap describing the Co-relation among attributes.



1. There is hardly any co relation between attributes
2. Even some of attribute had a negative co relation as well.
3. We had a least co relation therefore it proves that data is completely unbiased
4. Now we have least amount of the Underfit and Overfit

Modelling:-The main part of the Engine i.e Modelling. I have used many different model like :-

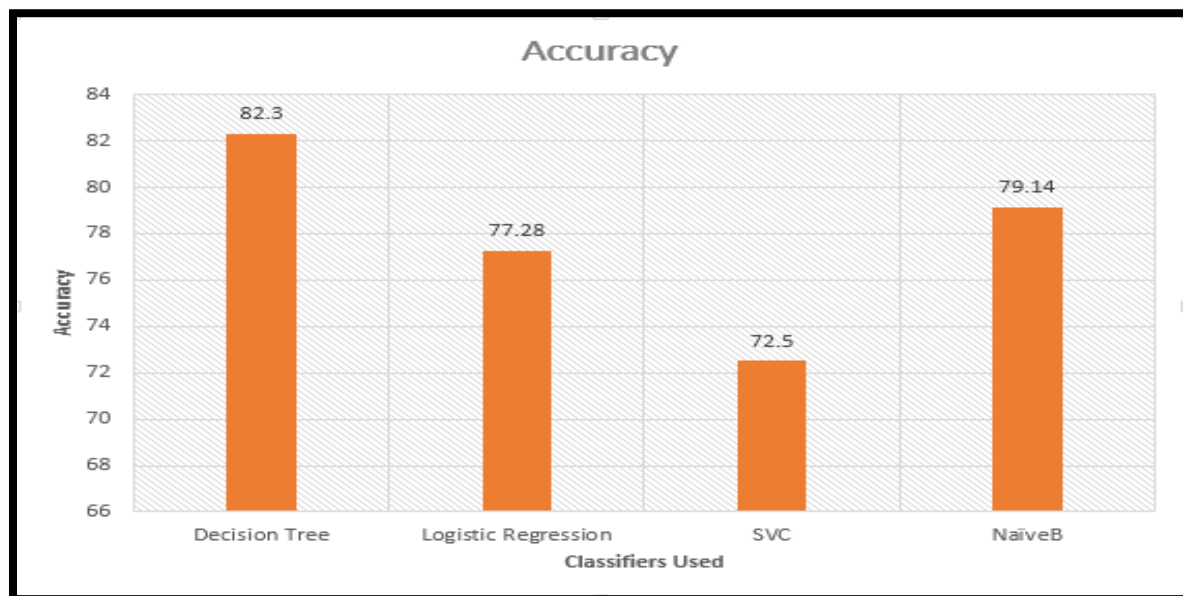
Decision Tree :-To generate one recommendation strictly acc. to user desire

Naïve bayes :-To best find out the probability based decision

SVM:-To better classify the objects in 3D Hyperspace

Logistic Regression:-The main advantage of using this Model is that it would work as a regression but then with the help of sigmoid function works as a classifier.

Below is the graph showing the accuracy of each classifier:-



1. Decision Tree has the max. accuracy of 82.3%
2. Logistic Regression has the accuracy rate of the 77.28%
3. SVC has the least efficiency of the 72.5%
4. Naïve Bayes has a good efficiency of the 79.14%

3. RESULTS AND DISCUSSIONS

3.1 RESULTS

3.1.1 BENEFIT OF PROJECT:-

This project results in enhancement of the DATA ANALYSIS in the can be viewed as a vast and way better application for users to save time in order to find the best suitable choice for them as a recommendation.

3.1.2 RESULT ON INDIVIDUAL DEVELOPMENT:-

On an individual basis this project helped me a lot in understanding concepts of **Python**. By this I was able to explore the use of Anaconda, Apache Server and in enhancement the concepts of hybrid programing

At the due of all these things, I am able to create the best suitable engine to predict the amazing recommendations/

4. FUTURE SCOPE

4.2 FUTURE SCOPE

1. Broadening of the recommendations offered to the client.
2. Utilization of social information accessible in the present for better visualization.
3. Change in improved recommendations through utilizing the additional capacities of the data.
4. Improving the recommendation standard through advanced machine learning.
5. Exploring the human nature states

REFERENCES

1. P. Resnick, N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl, “GroupLens: an open architecture [View at Publisher](#) · [View at Google Scholar](#)
2. U. Shardanand and P. Maes, “Social information filtering: algorithms for automating ‘Word of Mouth’,” in Proceedings of the Human Factors in Computin. [View at Scopus](#)
3. R. Ghani and A. Fano, “Building recommender systems using a knowledge base of product semantics,” in Proceedings
4. K.-W. Cheung, J. T. Kwok, M. H. Law, and K.-C. Tsui, “Mining customer product ratings for personalized marketing,” Decision Support Systems, vol. 35, no. 2, pp. 231–243, 2003. [View at Publisher](#) · [View at Google Scholar](#) · [View at Scopus](#)