



INSTITUTE FOR ADVANCED COMPUTING
AND SOFTWARE DEVELOPMENT AKURDI,
PUNE

Documentation On

“Hotel Recommendation System using Python”

PG-DBDA SEPT 2023

Submitted By:
Group No: 01

Roll No.
239508
239511

Name:
Anuj Thawre
Ashlesh Chinchkhede

Mrs. Priyanka Bohr
Project Guide

Mr. Rohit Puranik
Centre Coordinator

ABSTRACT

The objective of this project is to develop a machine learning-based hotel recommendation system that assists users in finding suitable accommodations based on their preferences and requirements. By employing various machine learning algorithms, the system aims to enhance user experience and satisfaction in the hotel booking process. Therefore the aim of this project is to predict the quality of the question using ML models so that the user can make a calculated guess before posting the question.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to everyone who has contributed to the completion of our project.

First and foremost, we would like to thank our project guide **Mrs. Priyanka Bohr** mam for their constant guidance and support throughout the project. We extend our sincere thanks to our respected centre coordinator, **Mr Rohit Puranik** for allowing us to use the facilities available.

We would also like to express our appreciation to the faculty members of our department for their constructive feedback and encouragement. Their insights and suggestions have helped us to refine our ideas and announce the quality of our work.

Further more, we would like to thank our families and friends for their unwavering support and encouragement throughout our academic journey, their love and support have been a constant source of motivation and inspiration for us.

Thank you for all your valuable contributions to our project.

Anuj Thawre (239508)
Ashlesh Chinchkhede (239511)

Table of contents

1. Introduction	5
1.1 PROBLEM STATEMENT.....	5
1.2 Abstract	5
1.3 Product Scope	5
1.4 Aims & Objectives.....	6
2. Overall Description.....	7
2.1 Workflow of Project:	7
2.2 Data Preprocessing and Cleaning	7
2.2.1 Data Cleaning.....	7
2.2.2 Label encoding	8
2.3 Exploratory Data Analysis	8
3. Requirements Specification.....	17
3.1 Hardware Requirement.....	17
3.2 Software Requirement.....	17
4. Conclusion:	18
5. Future Scope	19
6. References	20

Figure A Workflow Diagram.....	7
Figure 1 Correlation heatmap between variables.....	9
Figure 2 Below graph shows the Correlation between Price of Hotel vs Start Rating...	10
Figure 3 Showing count of hotels having ratings 2,3,4.....	11
Figure 4 Top 10 most booked Hotels.....	12
Figure 5 Top 10 cities where hotels are most booked.....	13
Figure 6 Top 10 countries where hotels are most booked.	14
Figure 7 Graph showing count of hotels in each country.	15
Figure 8 Top property types preferred by users while hotel booking.	16

1. Introduction

1.1 PROBLEM STATEMENT

Hotel Recommendation System using Python

Motivation

In this project we are building a hotel recommender system for the customer who needs an insight about how their hotel is ranked compared with other hotels from scratch.

1.2 Abstract

The objective of this project is to develop a machine learning-based hotel recommendation system that assists users in finding suitable accommodations based on their preferences and requirements. By employing various machine learning algorithms, the system aims to enhance user experience and satisfaction in the hotel booking process. Therefore the aim of this project is to predict the quality of the question using ML models so that the user can make a calculated guess before posting the question.

1.3 Product Scope

The main use of this classification models is to check the quality of the question posted by the user in our web interface. The user will input the question's title and body and press the Predict Question Button after this the model will predict the quality of the question and send that quality back to the user.

1.4 Aims & Objectives

The aim of a hotel recommendation system is to provide users with relevant and useful recommendations that meet their specific needs and preferences, while also benefiting hotels by increasing bookings and revenue

2 Overall Description

2.1 Workflow of Project:

The diagram below shows the workflow of this project.

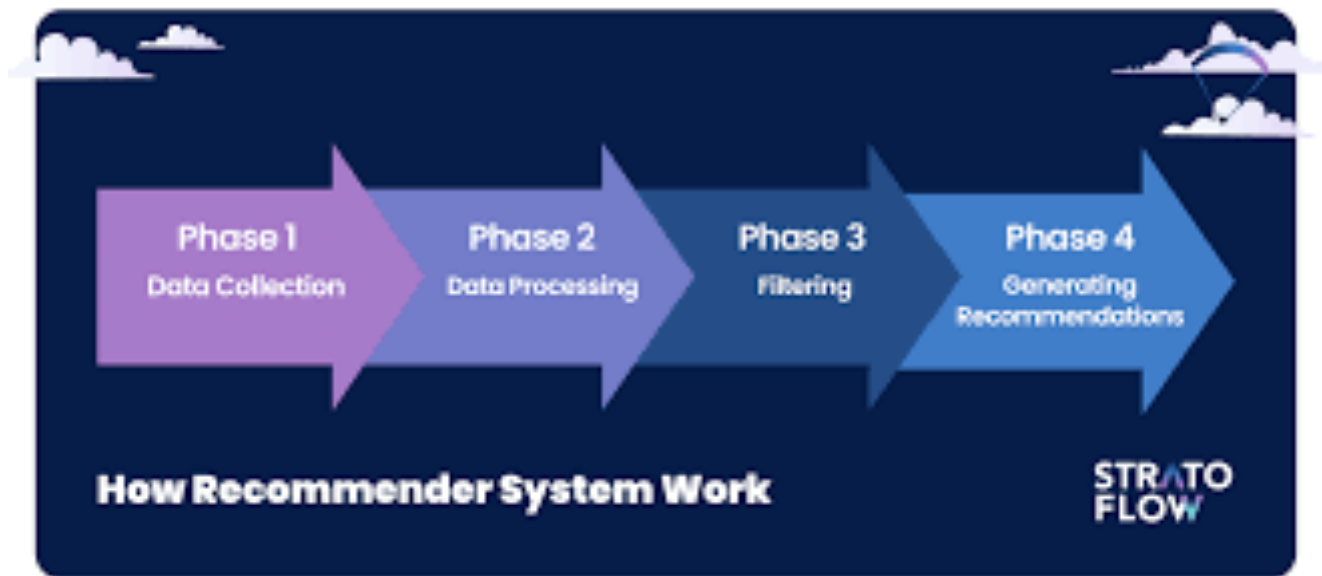


Figure Workflow Diagram

2.2 Data Preprocessing and Cleaning:

2.2.1 Data Cleaning:

The data can have many irrelevant, missing parts, HTML tags, links. To handle this part, data cleaning is done.

1. Remove HTML tags

The attributes present in our dataset like title and body have HTML tags like[p, a, strong etc.]

2. Removing StopWords

The stop words present in our dataset don't have any significance when we train our model. We don't lose any information while training the data. But removing stopwords will make the size of our dataset small and this reduces training time.

3. Lemmatizing our dataset

The data contains many words that can be reduced down to its basic form.

This step will reduce training time but the dataset will still retain its original meaning.

2.2.2 Label encoding:

To make the data understandable or in human readable form, the training data is often labeled in words. Label Encoding refers to converting the labels into numeric form so as to convert it into the machine-readable form. Machine learning algorithms can then decide in a better way on how those labels must be operated.

2.3 Exploratory Data Analysis:

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

Following are some plots we used to extract some useful information

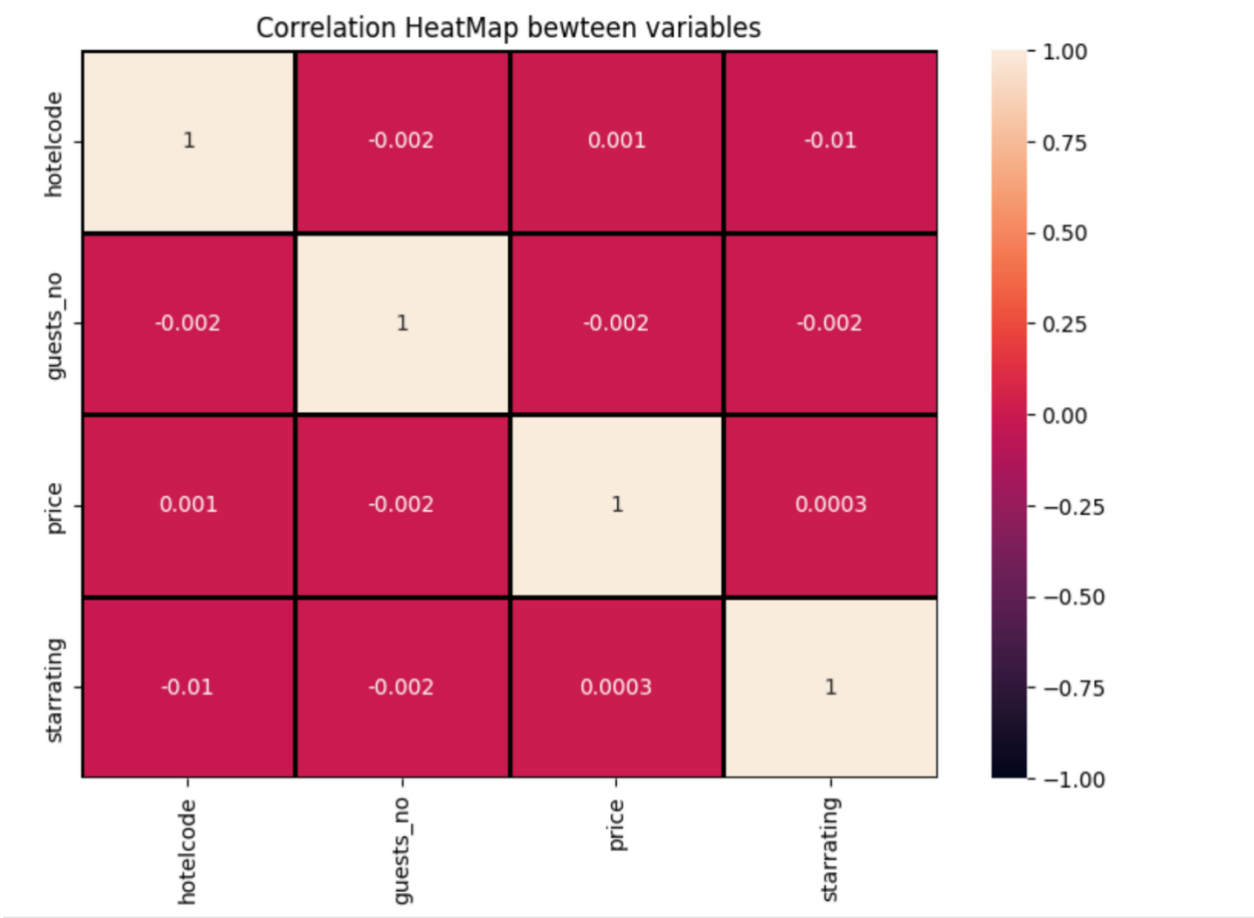


Figure 1 Correlation heatmap between variables

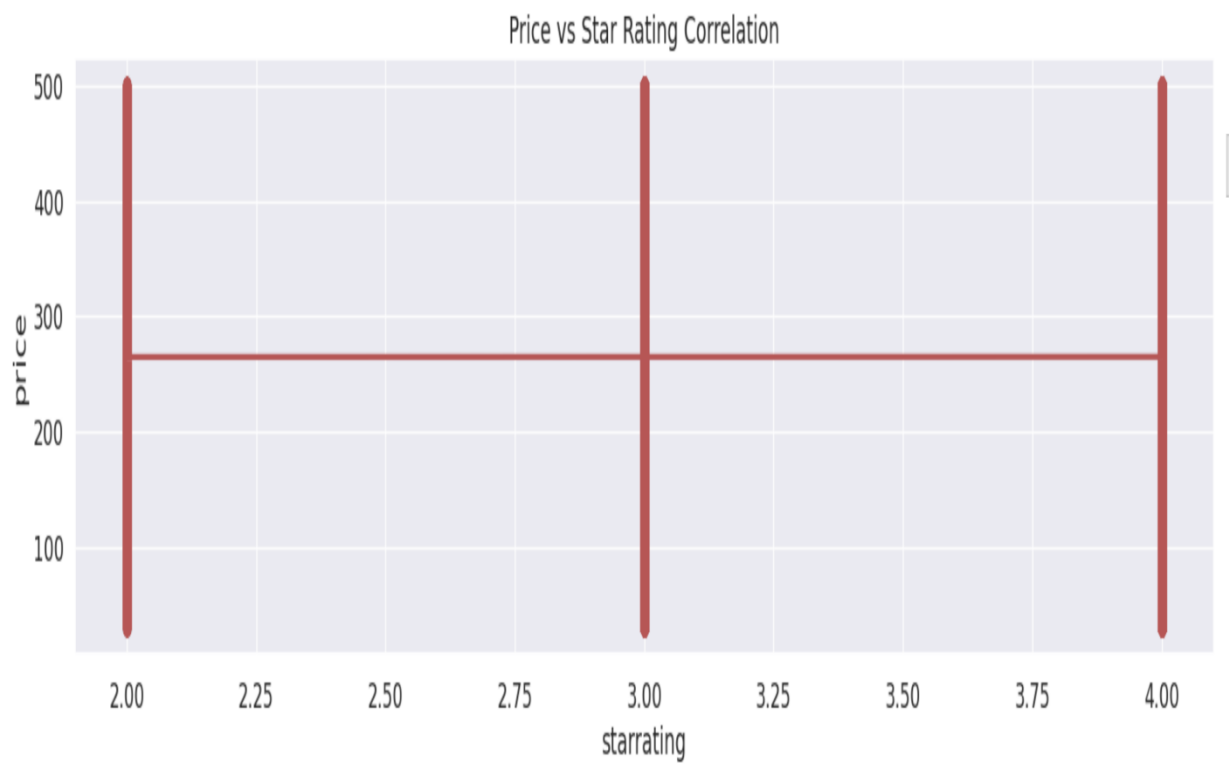


Figure 2 Below graph shows the Correlation between Price of Hotel vs Start Rating.

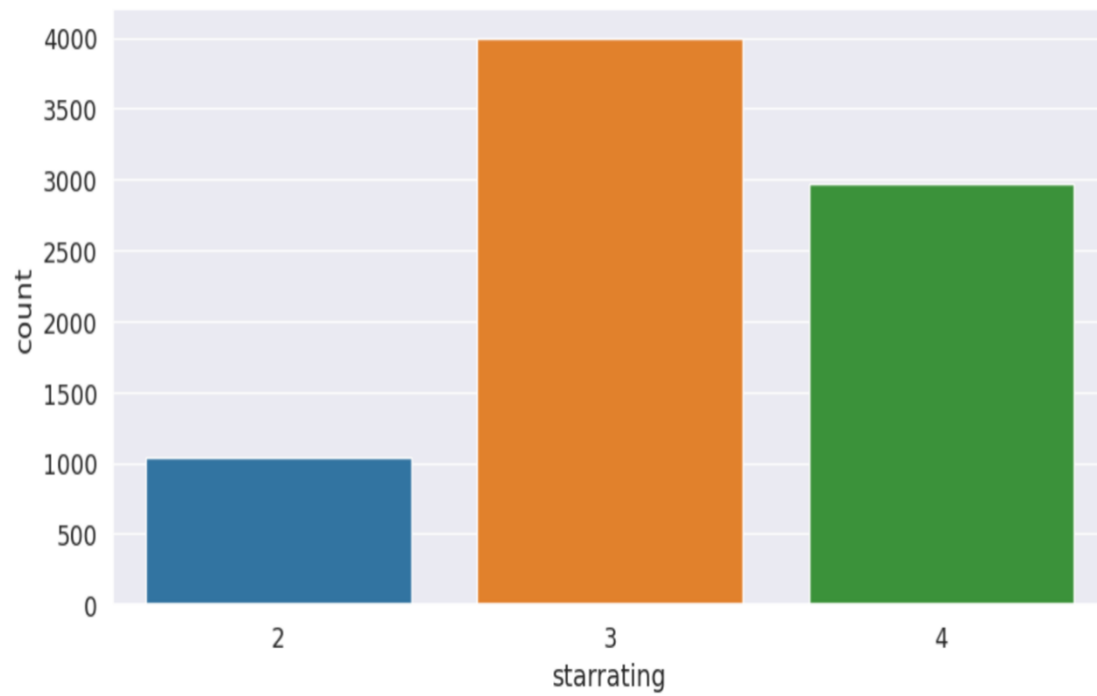


Figure 3 Showing count of hotels having ratings 2,3,4

Exp: Here the rating 3 was voted by most of the hotel visitors therefore star rating 3 is highest as shown in the graph above

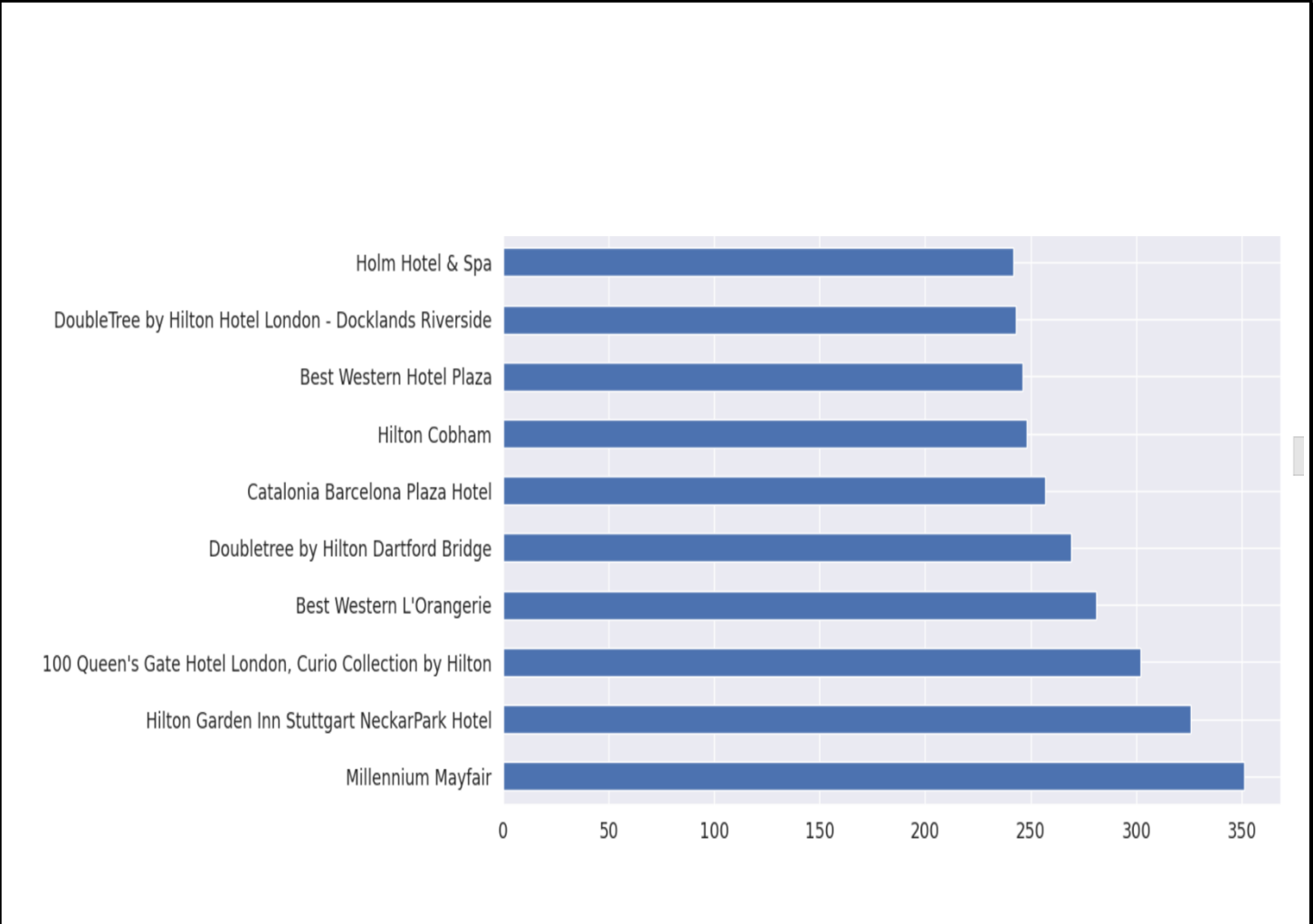


Figure 4 Top 10 most booked Hotels..

Exp: Here the top 10 hotels are shown which were most booked and thus the best hotel was Millennium Mayfair

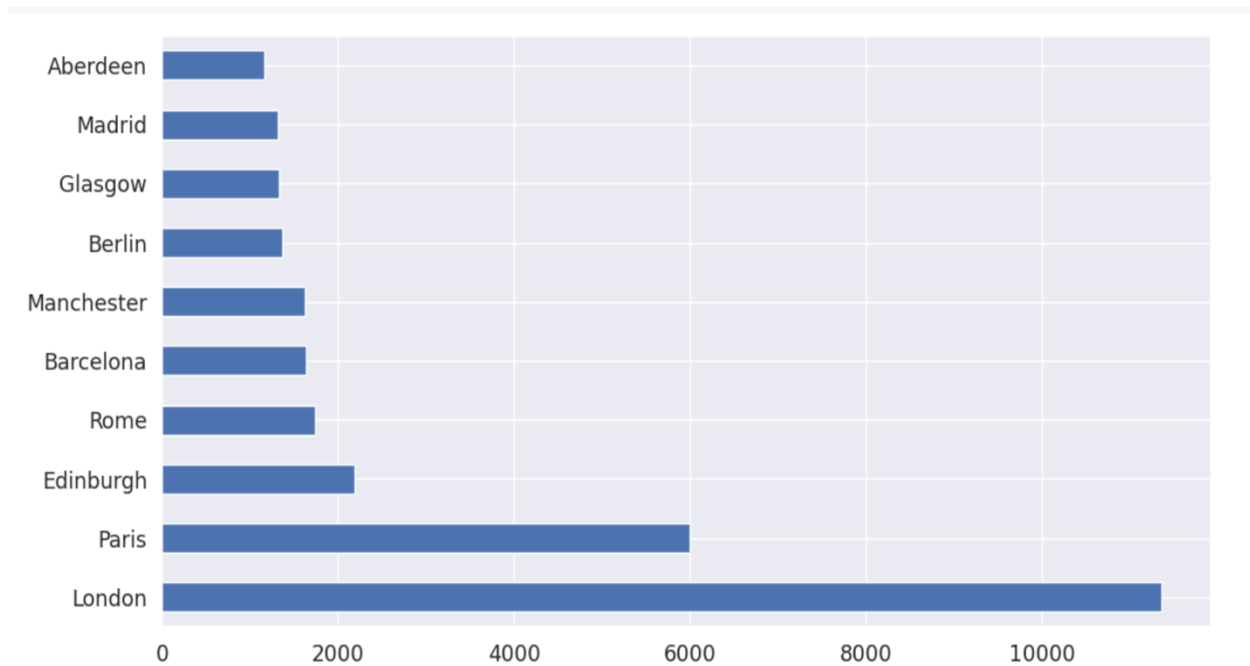


Figure 5 Top 10 cities where hotels are most booked

Exp: Here we can observe that most of the visitors visited city London Thus London is the most visited city where hotels were most booked followed by cities Paris, Edinburgh, Rome, Barcelona.

10k+ hotels were booked in city London

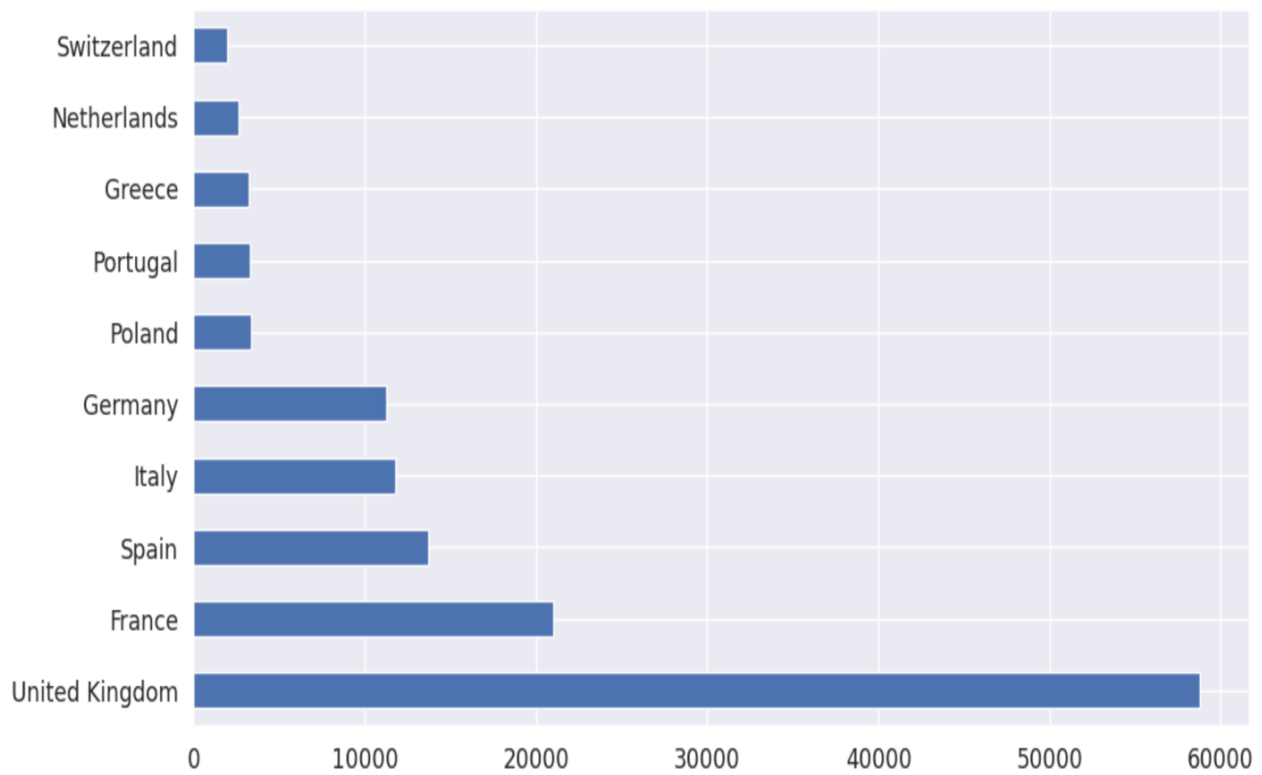


Figure 6 Top 10 countries where hotels are most booked.

Exp: Here we can observe that most of the visitors visited country UK. Thus UK is the most visited country where hotels were most booked followed by countries France, Spain, Italy, Germany.

50k+ hotels were booked in country UK.

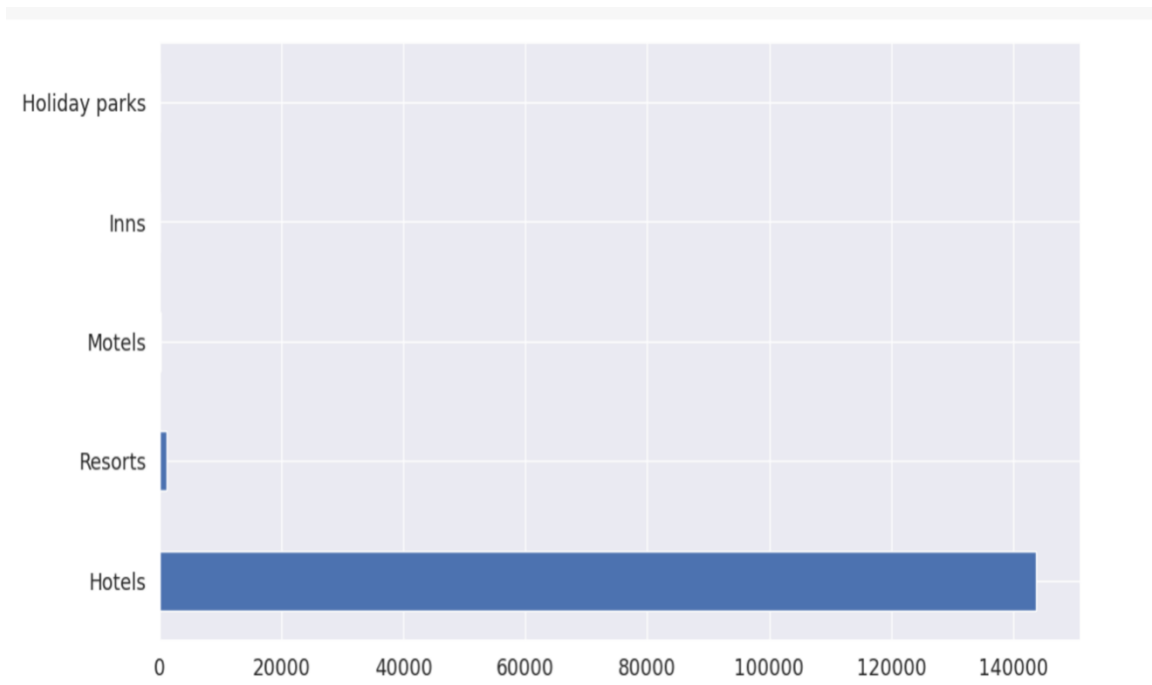


Figure 7 Graph showing count of hotels in each country.

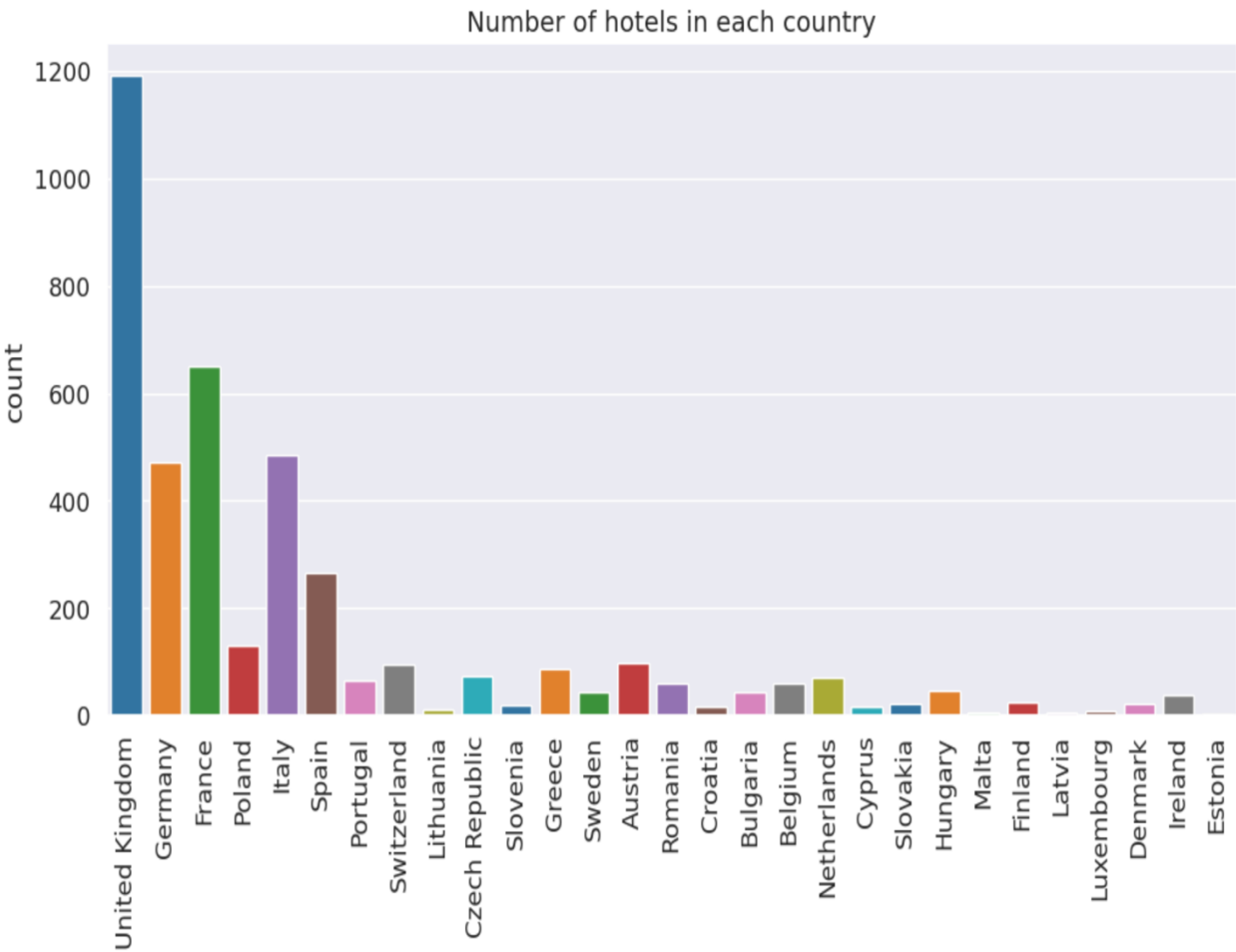


Figure 8 Top property types preferred by users while hotel booking.

Exp: From the graph above we can observe that the country United Kingdom has most number of hotels. 1k+ hotels are present.

3 Requirements Specification

4.1 Hardware Requirement:

- 500 GB hard drive (Minimum requirement)
- 8 GB RAM (Minimum requirement)
- PC x64-bit CPU

4.2 Software Requirement:

- Windows/Mac/Linux
- Python-3.9.1
- VS Code/Anaconda/Spyder
- Python Extension for VS Code
- Libraries:
 - Numpy 1.18.2
 - Pandas 1.2.1
 - Matplotlib 3.3.3
 - Scikit-learn 0.24.1
 - Seaborn

4. Conclusion:

- In this project we have built a system for Low Quality Question.
- For Model building we gathered the data from Kaggle.com
- Built a different models on the cleaned dataset.
- The app gives a comparison of the different trained models as well.
- Using the feedback obtained for the question classification the user can improve the question to get a High Quality rating.

5 Future Scope

The hotel recommendation system project demonstrates the effectiveness of machine learning techniques in providing personalized recommendations to users in the hospitality industry.

By leveraging collaborative and content-based filtering approaches, the system delivers accurate and relevant hotel suggestions, thereby enhancing user satisfaction and overall experience in the hotel booking process.

Continued exploration and refinement of machine learning models will contribute to further advancements in recommendation systems, benefiting both users and stakeholders in the hospitality industry.

6 References

Dataset –

[https://www.kaggle.com/datasets/keshavramaiah/hotel-recommendation.](https://www.kaggle.com/datasets/keshavramaiah/hotel-recommendation)

<https://towardsdatascience.com/a-machine-learning-approach-building-a-hotel-recommendation-engine-6812bfd53f50>

https://www.researchgate.net/publication/336577136_An_Intelligent_Data_Analysis_for_Hotel_Recommendation_Systems_using_Machine_Learning

https://www.researchgate.net/publication/335259229_Hotel_Recommendation_System