**HOTEL REVIEW SENTIMENT ANALYSIS**

****

**Submitted By:**

Khushi bathla

102053024

Bhavya gupta

102003409

**Submitted To:**

Ms. Kudratdeep Aulakh

July 2022 – December 2022

**HOTEL REVIEW SENTIMENT ANALYSIS**

**Python Libraries**

* Import pandas:

Pandas is a python library for data analysis and input. shape attribute in pandas help us to obtain shape of the data frame.

* Import missingno:

Missingno is an excellent and simple to use Python library that provides a series of visualisations to understand the presence and distribution of missing data within a pandas data frame.

* Import nltk:

Natural language toolkit is python package that you can use for NLP. A lot of data that you could be analyzing is unstructured data and contains human readable text.

* Import string:

It’s a built-in module and we have to import it before using any of its constants and classes. Python String module contains some constants, utility function, and classes for string manipulation.

* Import matplotlib:

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in python.

* Import seaborn:

It provides a high-level interface for drawing attractive and informative statistical graphics.

* Import wordcloud:

A word cloud (also known as a tag cloud or text cloud) is a visual representation of a text, in which the words appear bigger the more often they are mentioned. Word clouds are great for visualizing unstructured text data and getting insights on trends and patterns.

**Overview of the Project**

Sentiment analysis is part of the Natural Language Processing (NLP) techniques that consists in extracting emotions related to some raw texts. The goal of this study is to show how sentiment analysis can be performed using python. We will use here some hotel reviews data. Each observation consists in one customer review for one hotel. Each customer review is composed of a textual feedback of the customer's experience at the hotel and an overall rating. For each textual review, we want to predict if it corresponds to a good review (the customer is happy) or to a bad one (the customer is not satisfied). The reviews overall ratings can range from 2.5/10 to 10/10.

In order to simplify the problem we will split those into two categories:

* bad reviews have overall ratings < 5
* good reviews have overall ratings >= 5

**Code and Output**

































