Group Name	Mining and Analysis of Information related to California Hotels	Date	09/26/2023
Member (Coordinator) Name	Xiangyi ZHU	Signature	Xiangyi.Z
Member Name	Priyansh Sanjaybhai Zinzuvadia	Signature	Priyansh.S
Member Name	Aditya Rakesh Desale	Signature	Aditya_Desale
Member Name	Jiachen Wang	Signature	Jiachen Wang
Project Statement (200 words)	First, we will obtain the hotel information dataset from a hotel reservation website and clean it (Booking.com). The collected fields, including hotel name, price, score, review count, coordinates, and guest reviews. We will collect the average hotel ratings in each region of California and present them in the form of a pie chart. For hotel prices, we will draw a line chart of price changes for some hotels in the month before and after the festival. Also, we will count the average number of reviews received by hotels in each region and present them in the form of a histogram (if possible, a map related to hotel reviews will be produced, with colors used to distinguish the number of reviews or the level of ratings). If possible, we will try to apply NLP to our project, training the model to rate each review as 'good' or 'bad', and extracting the keywords in comments. For keywords, we may make a word cloud to describe what people like or dislike about this hotel. Then compare the review information output by the machine with the actual rating of the hotel. And we will try to look for any discrepancies between machine ratings and actual ratings. We can consider those hotels that have actual ratings higher than those given by the machine to be overrated, and vice versa. There could be many reasons behind this, it could be a bot rating.		

Objectives and/or steps toward project completion)	- Data Collection (Scraping the data from website)
	- Data Transformation & Data Cleaning
	- Exploratory Data Analysis:
	    (1) Feature Engineering
	    (2) Correlation Analysis
	    (3) Pattern Recognition
	- Data visualization (charts on hotel rating, price and reviews)
	- If possible:NLP:
	    (1)Extract keywords of comments
	    (2)Train models to rate comments
	- Compare machines' rates and hotels' real rates		

| Weekly Schedule and individual assignments | We plan to meet 2 or 3 times per week to discuss and work collaboratively on any problem encountered or to brainstorm features for the project.……
The individual assignments are
Xiangyi: Collect data, Clean data, Create Number of Reviews chart, learn NLP
Jiachen: Collect data, Clean data, Create Number of Reviews chart, learn NLP
Priyansh: Collect data,Create Hotel Rating chart & Price Change chart, learn NLP
Aditya: Collect data, Create Hotel Rating chart & Price Change chart, learn NLP

Weekly Schedule:
● 9/27-10/3: Find dataset, meet as team and discuss where we want to go with the project.
● 10/4-10/10: Clean dataset, Feature Engineering,
● 10/11/-10/17: Correlation Analysis
● 10/18-10/24: Pattern Recognition
● 10/25-10/31: Data visualization(1)
● 11/1-11/7: Data visualization(2)
● 11/8-11/14: NLP learning and application(1)
● 11/15-11/21: NLP learning and application(2)
● 11/22-11/28: NLP learning and application(3)
● 11/29-12/5: Creating website and writing project report
● 12/6-12/11: Presentation and Video creation |