Project(2): Hotel Reviews Analysis

Can you make your trip more cozy by using data science?. Can you predict what score a reviewer will give a hotel using features about the hotel in combination with the reviewer history and each review's language?.

Dataset:
This dataset contains 515,000 customer reviews and scoring of 1493 luxury hotels across Europe.

Dataset Snapshots:

	A	B	C	D	E	F	G	H
1	Hotel_Address	Additional_Number_of_Scori	Review_Date	Average_S	Hotel_Name	Reviewer_Nation	Negative_Review	Review_Total_Negative_Word_Counts
2	s Gravesandestraat 5	194	8/3/2017	7.7	Hotel Arena	Russia	I am so angry that i ma	397
3	s Gravesandestraat 5	194	8/3/2017	7.7	Hotel Arena	Ireland	No Negative	0
4	s Gravesandestraat 5	194	7/31/2017	7.7	Hotel Arena	Australia	Rooms are nice but fo	42
5	s Gravesandestraat 5	194	7/31/2017	7.7	Hotel Arena	United Kingdom	My room was dirty an	210
6	s Gravesandestraat 5	194	7/24/2017	7.7	Hotel Arena	New Zealand	You When I booked w	140
7	s Gravesandestraat 5	194	7/24/2017	7.7	Hotel Arena	Poland	Backyard of the hotel	17
8	s Gravesandestraat 5	194	7/17/2017	7.7	Hotel Arena	United Kingdom	Cleaner did not chang	33
9	s Gravesandestraat 5	194	7/17/2017	7.7	Hotel Arena	United Kingdom	Apart from the price f	11
10	s Gravesandestraat 5	194	7/9/2017	7.7	Hotel Arena	Belgium	Even though the pictu	34
11	s Gravesandestraat 5	194	7/8/2017	7.7	Hotel Arena	Norway	The aircondition make	15
12	s Gravesandestraat 5	194	7/7/2017	7.7	Hotel Arena	United Kingdom	Nothing all great	5
13	s Gravesandestraat 5	194	7/6/2017	7.7	Hotel Arena	France	6 30 AM started big n	75
14	s Gravesandestraat 5	194	7/6/2017	7.7	Hotel Arena	United Kingdom	The floor in my room	28
15	s Gravesandestraat 5	194	7/4/2017	7.7	Hotel Arena	Italy	No Negative	0

~Dataset header Continued:

	I	J	K	L	M	N	O	P	Q
1	Total_Number_of_R	Positive_Review	Review_Total_Positive_Word	Total_Number_of_Revie	Reviewer_Score	Tags	days_since	lat	lng
2	1403	Only the park outside	11	7	2.9	[' Leisure trip ', '	0 days	52.36058	4.915968
3	1403	No real complaints th	105	7	7.5	[' Leisure trip ', '	0 days	52.36058	4.915968
4	1403	Location was good an	21	9	7.1	[' Leisure trip ', '	3 days	52.36058	4.915968
5	1403	Great location in nice	26	1	3.8	[' Leisure trip ', '	3 days	52.36058	4.915968
6	1403	Amazing location and	8	3	6.7	[' Leisure trip ', '	10 days	52.36058	4.915968
7	1403	Good restaurant with	20	1	6.7	[' Leisure trip ', '	10 days	52.36058	4.915968
8	1403	The room is spacious	18	6	4.6	[' Leisure trip ', '	17 days	52.36058	4.915968
9	1403	Good location Set in a	19	1	10	[' Leisure trip ', '	17 days	52.36058	4.915968
10	1403	No Positive	0	3	6.5	[' Leisure trip ', '	25 days	52.36058	4.915968
11	1403	The room was big end	50	1	7.9	[' Leisure trip ', '	26 days	52.36058	4.915968
12	1403	Rooms were stunning	101	2	10	[' Leisure trip ', '	27 days	52.36058	4.915968
13	1403	Style location rooms	4	12	5.8	[' Business trip ',	28 days	52.36058	4.915968
14	1403	Comfy bed good loca	6	7	4.6	[' Leisure trip ', '	28 days	52.36058	4.915968
15	1403	This hotel is being rer	59	6	9.2	[' Business trip ',	30 days	52.36058	4.915968

Dataset Description:

Feature	Description
Hotel Address	
Review Date	
Average Score	Average Score of the hotel, calculated based on the latest comment in the last year.
Hotel Name	
Reviewer Nationality	

Negative Review	Negative Review the reviewer gave to the hotel. If the reviewer does not give the negative review, then it should be: 'No Negative'
Review Total Negative Word Counts	Total number of words in the negative review.
Positive Review	Positive Review the reviewer gave to the hotel. If the reviewer does not give the negative review, then it should be: 'No Positive'
Review Total Positive Word Counts	Total number of words in the positive review.
Reviewer Score	Score the reviewer has given to the hotel, based on his/her experience.
Total Number of Reviews Reviewer Has Given	Number of Reviews the reviewers has given in the past.
Total Number of Reviews	Total number of valid reviews the hotel has.
Tags	Tags reviewer gave the hotel.
Days since review	Duration between the review date and scrape date.
Additional Number of Scoring	There are also some guests who made a scoring on the service rather than a review. This number indicates how many valid scores without review are in there.

Milestone 1 tasks:

1. Apply pre-processing on the provided dataset.
2. Experiment with regression techniques to reduce the error on prediction of reviewer score of a hotel (Deliver at least two techniques).