

Name:- Vaishnav Ramgiri
PRN :- 202401100078
Roll no. 16

20 problem statements with
Numpy and pandas of OpinRank
Review dataset.

1. Count the total number of reviews available for each hotel.
2. Find the average rating given to each hotel.
3. Identify the hotel with the highest average rating.
4. Find the most frequently used words in reviews for a specific hotel.
5. Determine the average length (Number of words) of reviews per hotel.
6. Count the Number of reviews per year for trend analysis.
7. Find the percentage of positive or negative reviews using keywords.

8. Filter out hotels with less than 50 reviews.

9. Find the variance and standard deviation in rating across all hotels.

10. Identify the top 5 hotels with the most consistent ratings (lowest std. dev)

11. Check for missing or null values in the dataset and handle them.

12. Determine the correlation between review length and rating.

13. Find out which hotels have received 5-star ratings most frequently.

14. Group reviews by city and compute average rating per city.

15. Find out which hotels have received 5-star ratings most frequently.

16. detect duplicate reviews and remove them.
17. plot the distribution of review scores using NumPy histogram data.
18. Analyze the sentiment of reviews using a basic word count approach.
19. List top 10 adjectives used in high-rated vs low-rated reviews.
20. Merge this dataset with a Fictional hotel pricing dataset and analyze price vs rating.