Qatar Airways EDA / ML Recommendation on passenger reviews

Steps Performed

1. Imported dataset

2. Removed `Unnamed: 0` Column

3. We have 2 columns on date hence verified that both have same data, and hence removed one of the column to reduce redundancy

4. Started checking for null values

   a. Had 1 null values in `Rating` and `Max Rating`, hence removed as it wont affect our analysis.

   b. Imputed missing values in `Type Of Traveller` with `Unknown`, Since we didn't have any dependent column/data to find and fill missing values with some sensible/relatable data, we will need to impute that data with `Unknown` value.

- [x] ~~Tried Using `from` and `to` extraction of route destinations from `review body` however it didn't worked.~~

   1. Further tried to create two columns named `From` and `To` from route to get appropriate destinations for segregation, but as we have some missing columns, we let those as it is as of now.

   2. Creating `VIA` Column from To column.

   3. Extracted `to` from `via` column, and deleted old `to` column.

      a. Faced new challenge in to column, as we've extracted from via column, some route column didn't had via hence they're mentioned as NaN.

   4. Rearranged the columns.

5. Imputed `Date Flown` column from `Date Published` changing date format and imputing it.

6. Worked on `Verified` column, where 1 is verified and 0 is Not verified, here we have Verified Review as text in Review Body, hence filtering it out and using those to impute 1.0 in NaN values.

7. Filled Nan values in `Recommended` on logic based on `Rating` column, where any rating below 8 will be no and above will be yes as recommendation, and imputed all the blank values using this logic.

8. Manually entered `From` and `To` locations for 10 remaining records

9. Changed names of Airport to `DOHA` from `doh` `DOH` and `doha, SIN, and others in` FROM, TO, & VIA.