TEAM NAME- **Akez**
TEAM ID- **IVS3891**

APPROACH

1)The input dataset was converted to dataframe & basic analysis was performed.

2)The NaN values of price were handled & converted to 0 meaning operator did not post price for it.

3)The Seat Fare Type 1 values were converted to a list row wise. Similar was done for Seat fare Type2

4)New column ‘Seat Fare Type1 Mean’ was made which was mean of values for Seat Fare Type 1 values row wise which were a list. Similar was performed for Seat Fare Type2

5)New column ‘Total_seats Type1’ was created which was length of list of Seat Fare Type1 which meant the no. of seats posted by operator on a day. Similar was done for Seat Fare Type2

6)The Buses where sum of Total seats Type1 & Type2 were 0, were placed in a list which indicated these buses were unnecessary data as none of the operators of the bus posted a non-null price for any date.

7)For Buses except these buses, for every Service date this bus posted a price for, standard deviation was calculated from the Seat Fare Type1 Mean column. The Standard deviation indicated how much the bus operator changed the price for a bus on particular Service date. With this we can infer that the bus operators that changed values in a more range(higher standard deviation) were following bus operators that changed values in a smaller range(lower standard deviation). Then these buses were scored accordingly on their standard deviation(Lower standard deviation means more score & vice-versa). Similar was done for buses that only posted prices for Seat Fare Type 2 & both scores were appended in a single data frame.

8)But these scores were biased to some bus operators which only posted for 1 service date & was against those bus operators which had slightly higher std dev but posted for multiple service dates. This was then handled by balancing the score appropriately according to the no. of unique Service Dates the bus operator posted prices.

9)The scores were then sorted in order & the output file was generated according to immediate neighbours for a bus service above & below in terms of follows, followed by & the scores.