Report:

Problem statement states that to find out which bus services is/are pricing independently (the price leaders) and which services are the followers (and following whom?), given the history of prices set by the operators.

Data contains five fields named Seat Fare Type 1, Seat Fare Type 2, Bus, Service Date, Recorded At.

Our Approach is as follows. As the data is mostly clean, we started with deleting the duplicate rows in data and also gave unknown values a separate category. Followed by we changed the data types to their respective categories (since all are of data type string initially). Followed by we created 3 dictionaries to accommodate the values of average of each bus total Seat Fare Type 1, Seat Fare Type 2 and also the occurrences of each bus in Data. Followed by we encoded each bus id with a distinct integer. Followed by the final pre-processed data is fitted into K-mean Clustering Algorithm with 117 centres, and also defined a method to result the predictions by model along with the confidence metrics.