

### **Case Study - Data Engineering Intern**

The data below is a stringified table where the delimiter is ';' and line\_terminator is '\n'.

Please create a new table with the following transformations on the above command separated data.

1. FlightCodes column: Some values are null. Flight Codes are supposed to increase by 10 with each row so 1010 and 1030 will have 1020 in the middle. Fill in these missing numbers and make the column an integer column (instead of a float column).
2. To\_From column: Should be split into two separate columns for better analysis! Split on '\_' to create two new columns respectively. Also, the case of the column is not very readable, convert the column into capital case.
3. Airline Code column: Clean the Airline Codes to have no punctuation except spaces in the middle. E.g. '(Porter Airways. )' should become 'Porter Airways'.

```
data = 'Airline Code;DelayTimes;FlightCodes;To_From\nAir Canada (!);[21, 40];20015.0;WATERLOO_NEWYORK\n<Air France> (12);[ ];Montreal_TORONTO\n(Porter Airways. );[60, 22, 87];20035.0;CALgary_Ottawa\n12. Air France;[78, 66];Ottawa_VANcouver\n\".\\Lufthansa.\\\";[12, 33];20055.0;london_MONTreal\n'
```

4. Write a mock SQL query on the above table to find all flights leaving from Waterloo