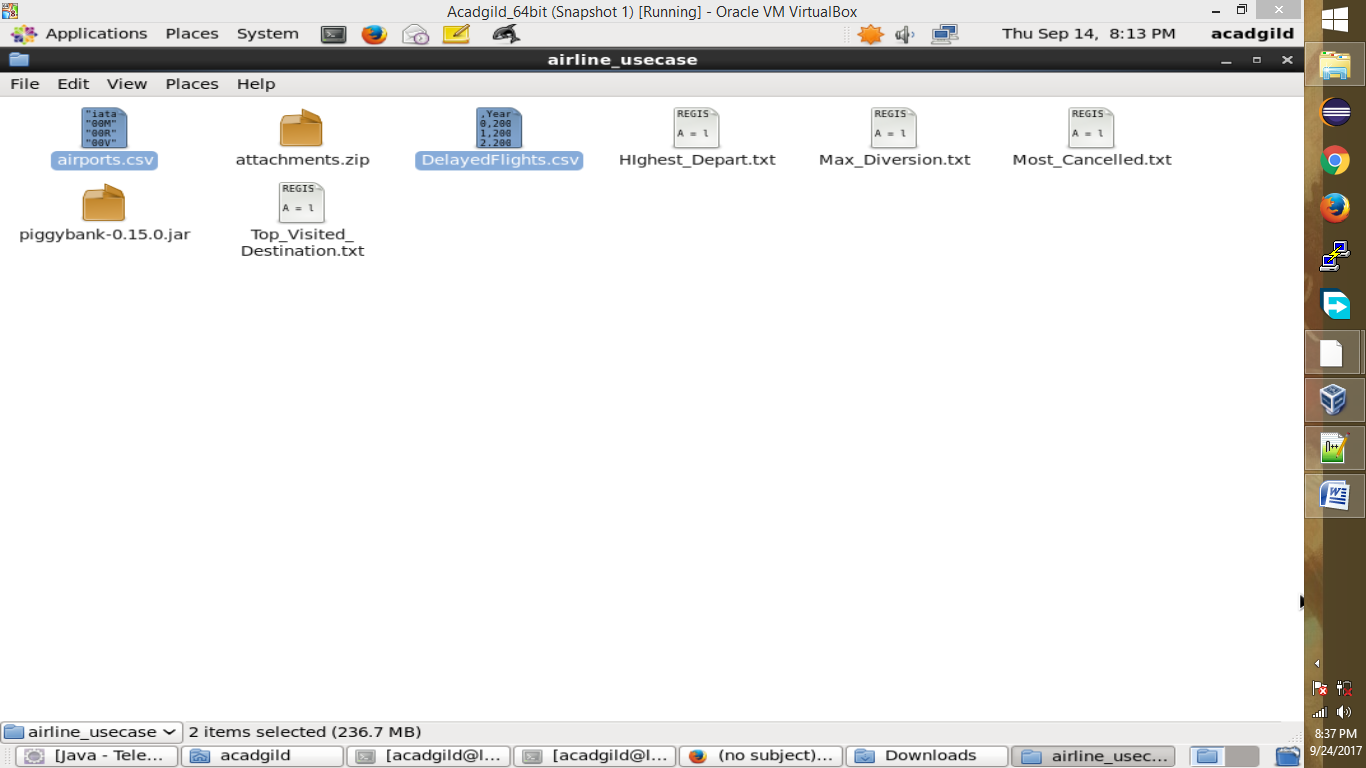
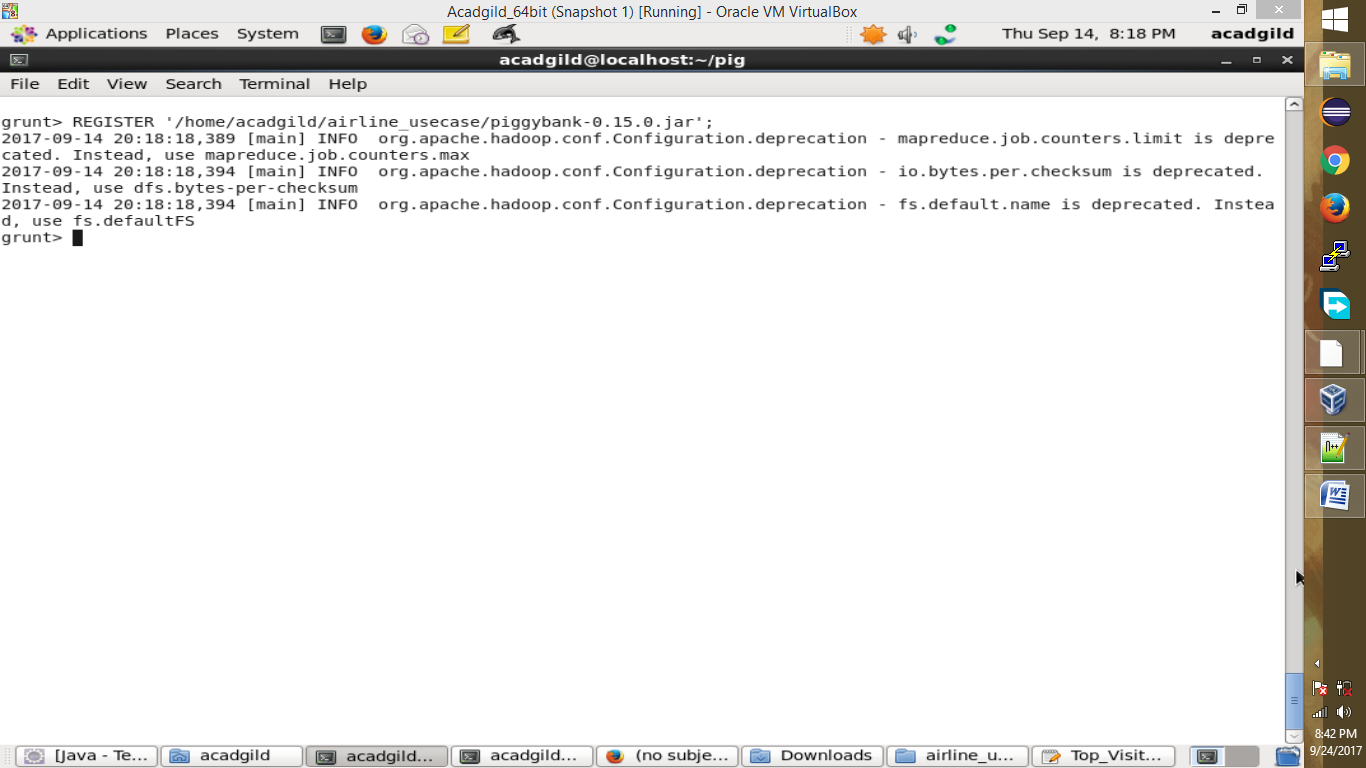
**Problem Statement: Top 5 Most Visited Destination**

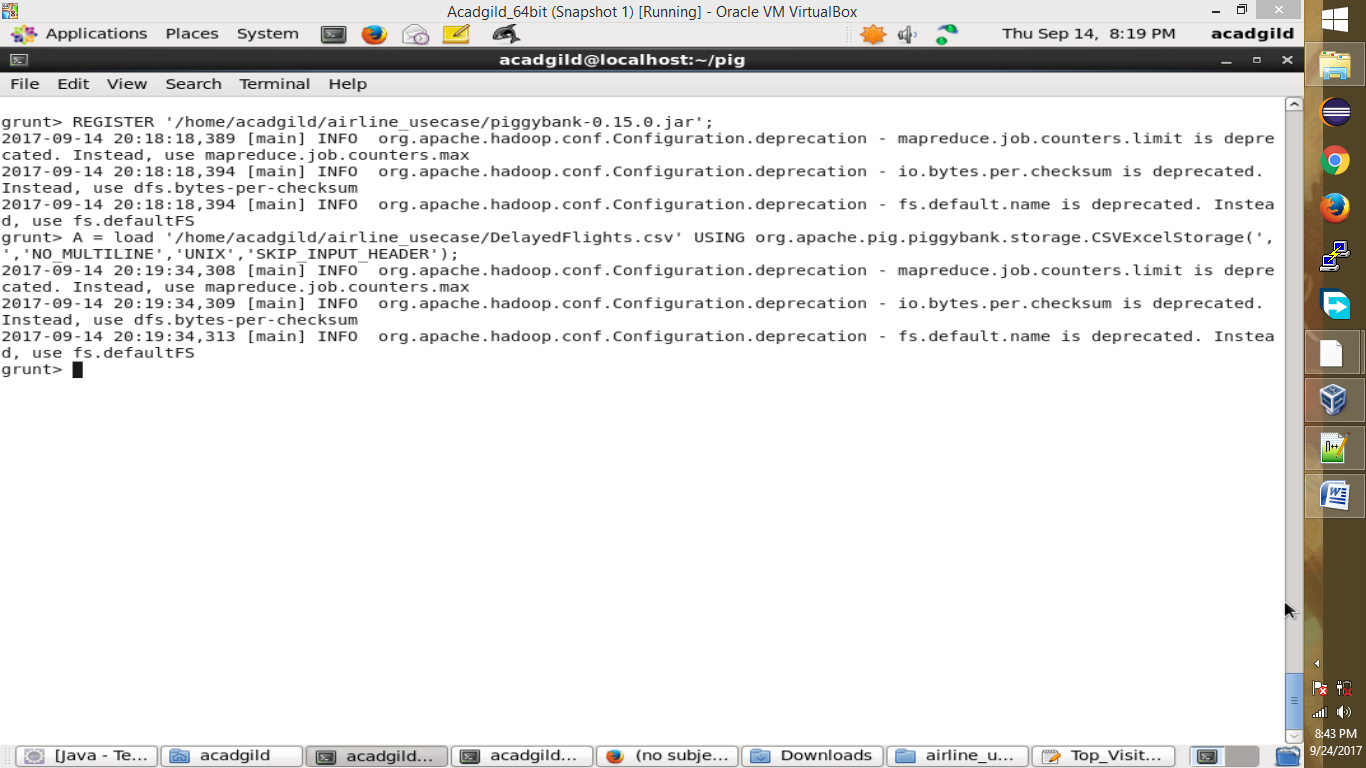
Step 1: Downloaded the file airport.csv and DelayedFlights.csv



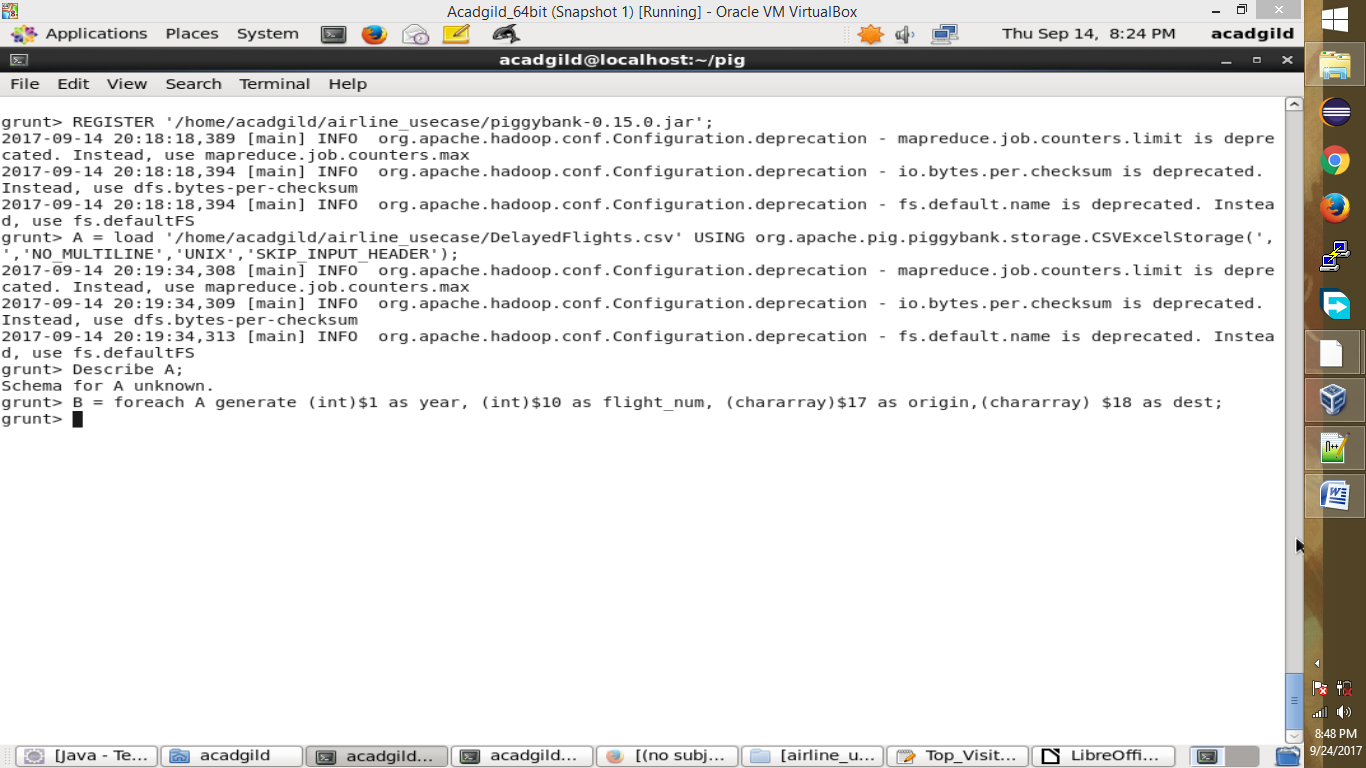
Step 2: We are registering *piggybank* jar in order to use the CSVExcelStorage class.



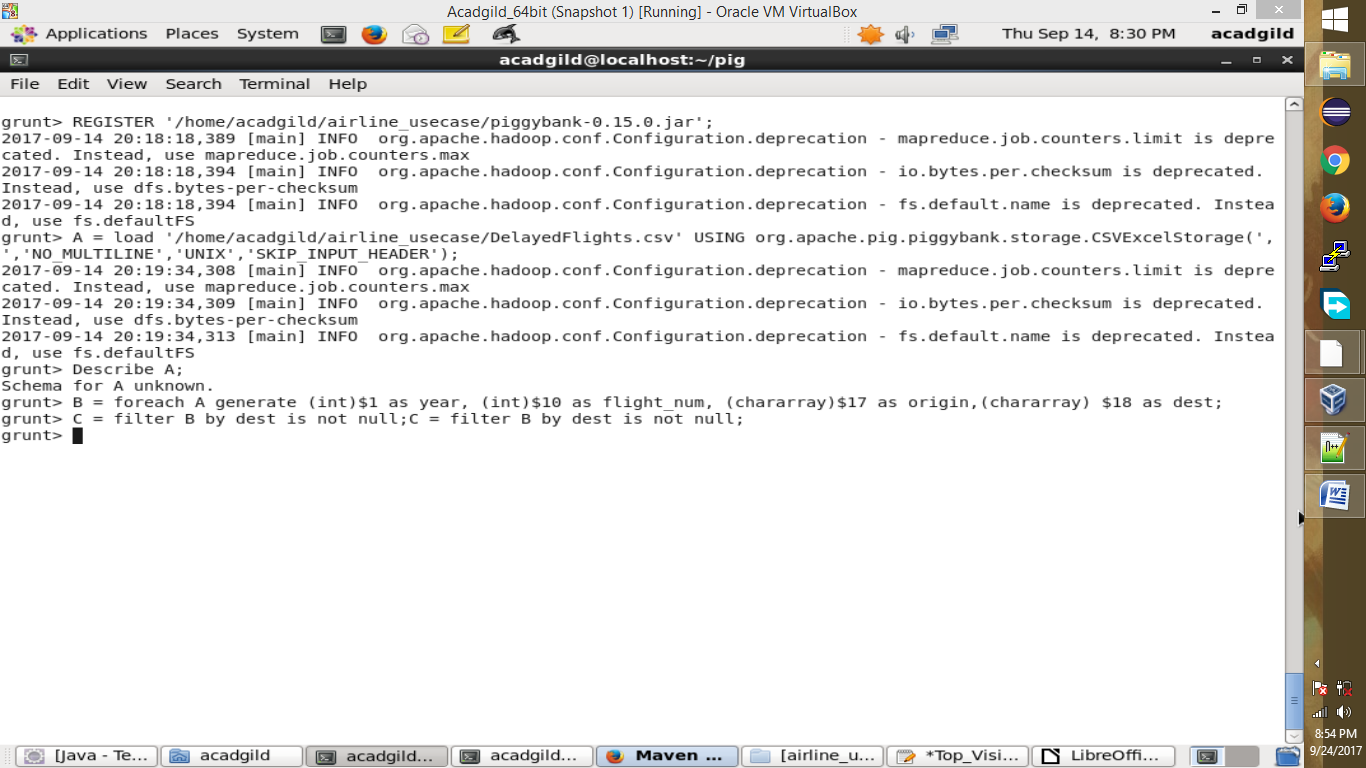
Step 3: In relation **A**, we are loading the dataset using CSVExcelStorage because of its effective technique to handle double quotes and headers.



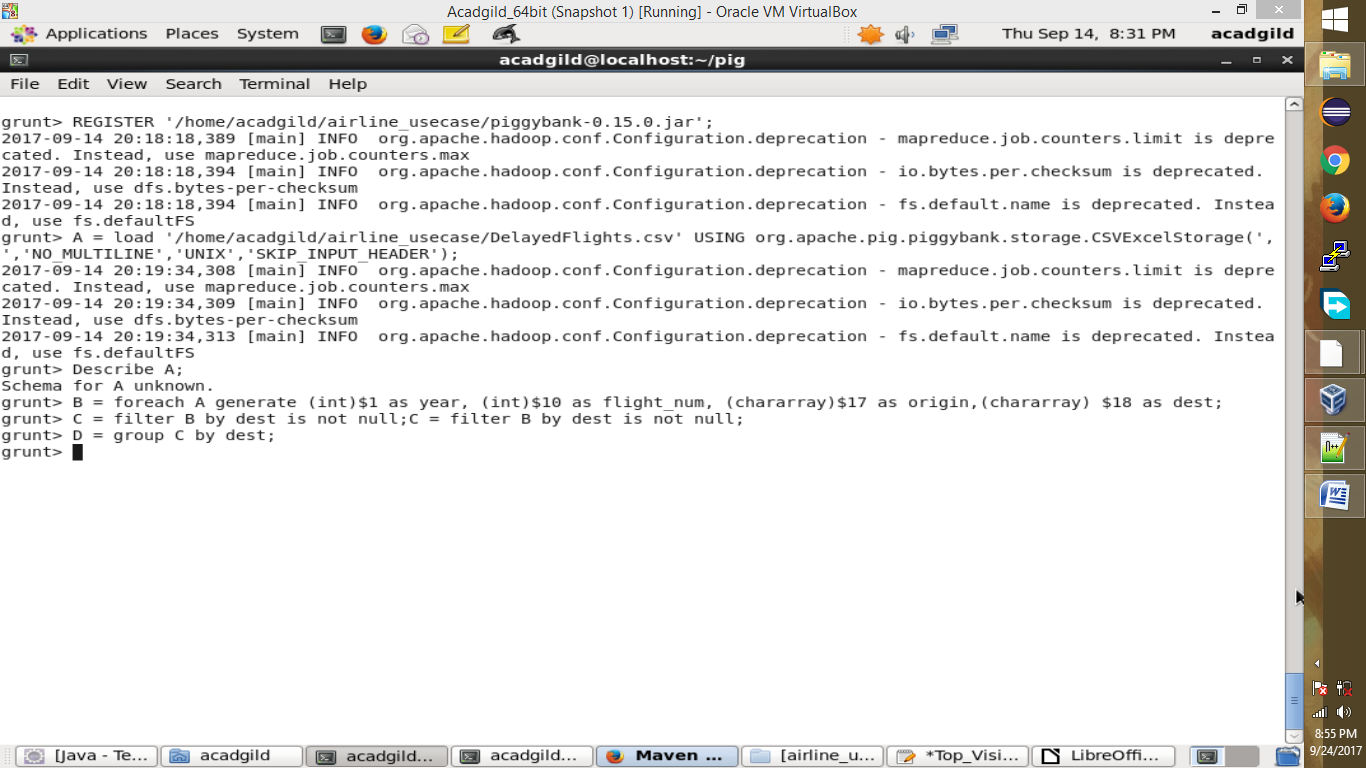
Step 4: In relation **B**, we are generating the columns that are required for processing and explicitly typecasting each of them.



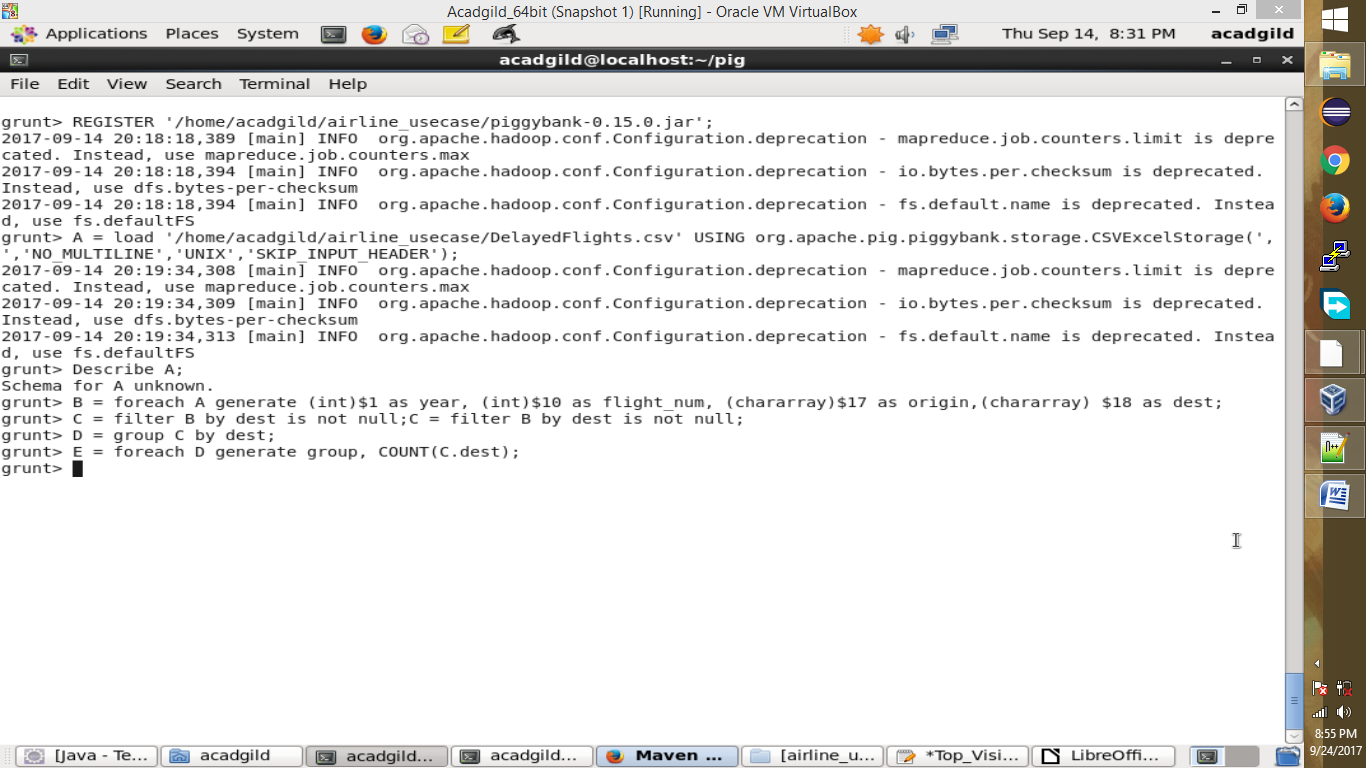
Step 5: In relation **C**, we are filtering the null values from the “dest” column.



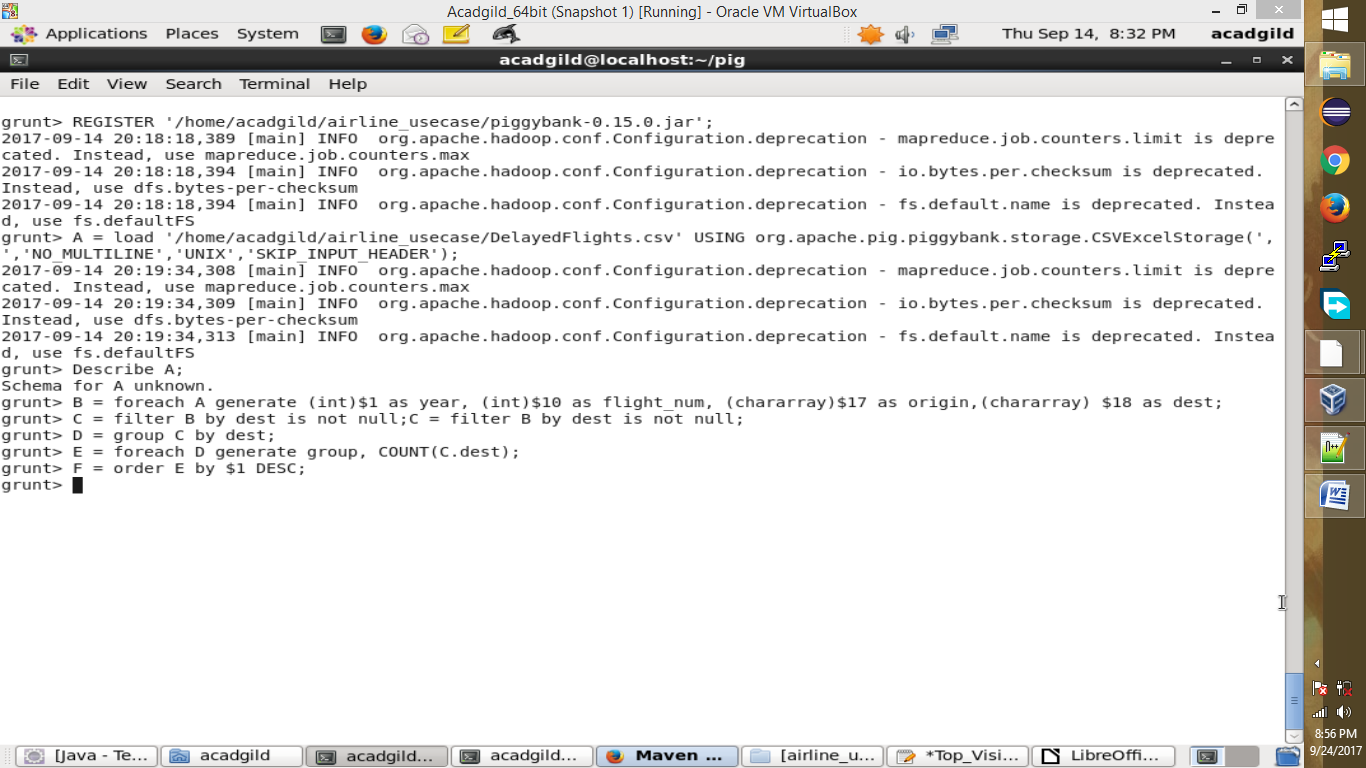
Step6: In relation **D**, we are grouping relation C by “dest.”



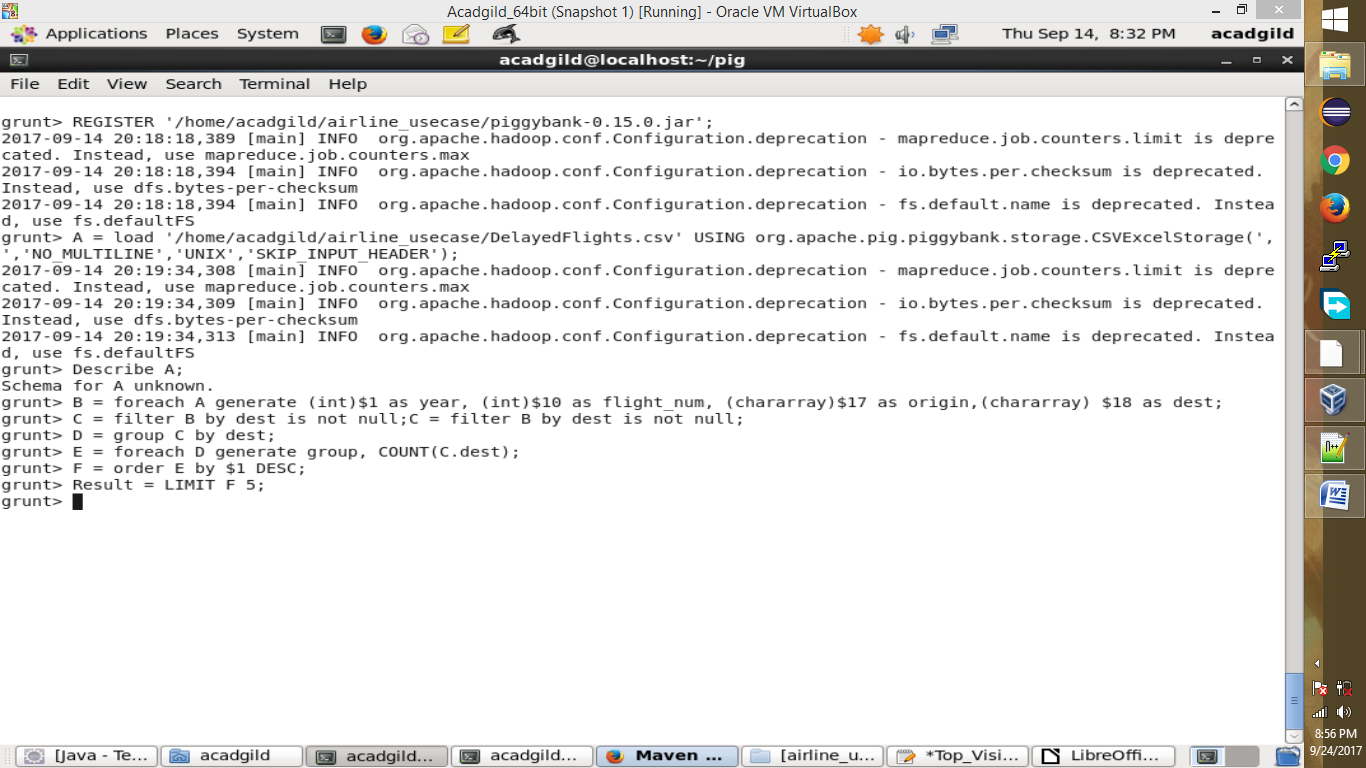
Step7: In relation **E**, we are generating the grouped column and the count of each.



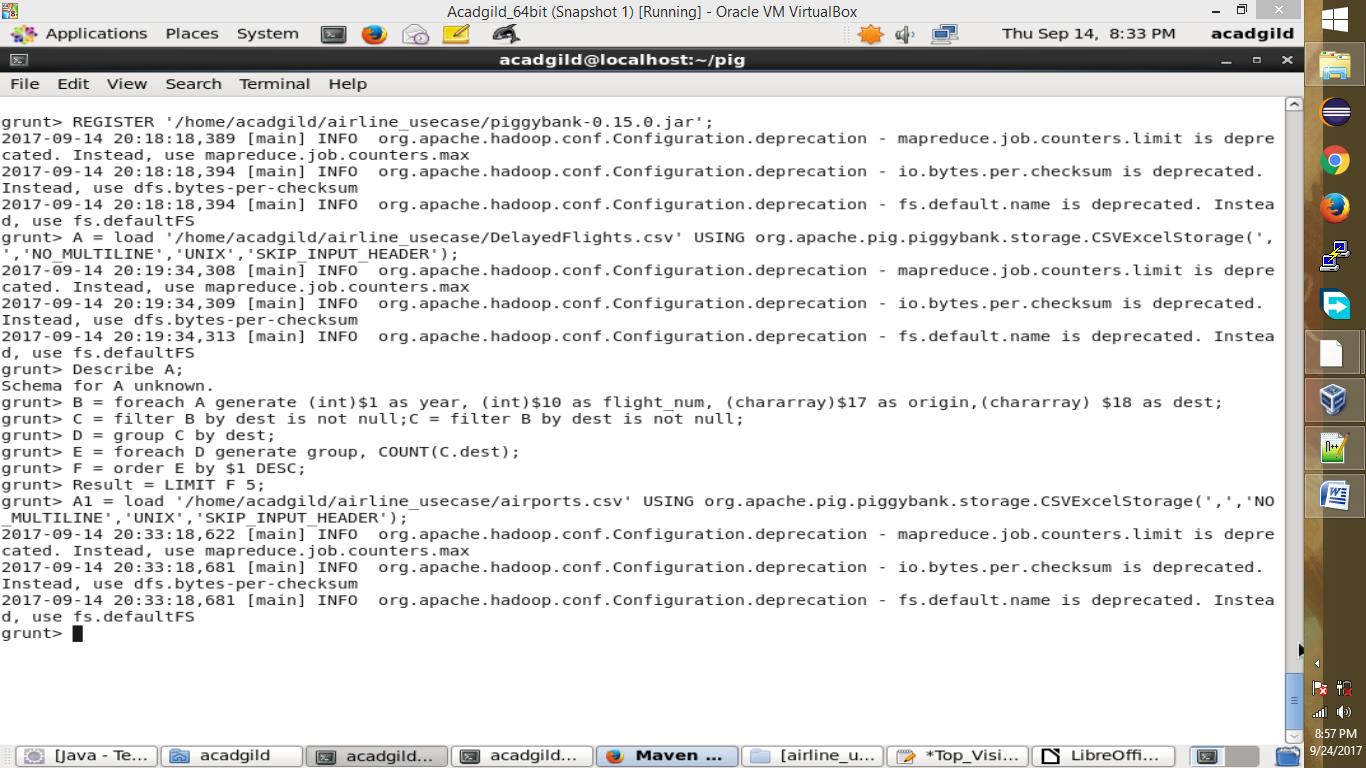
Step8: Relation **F** and **Result** is used to order and limit the result to top 5.



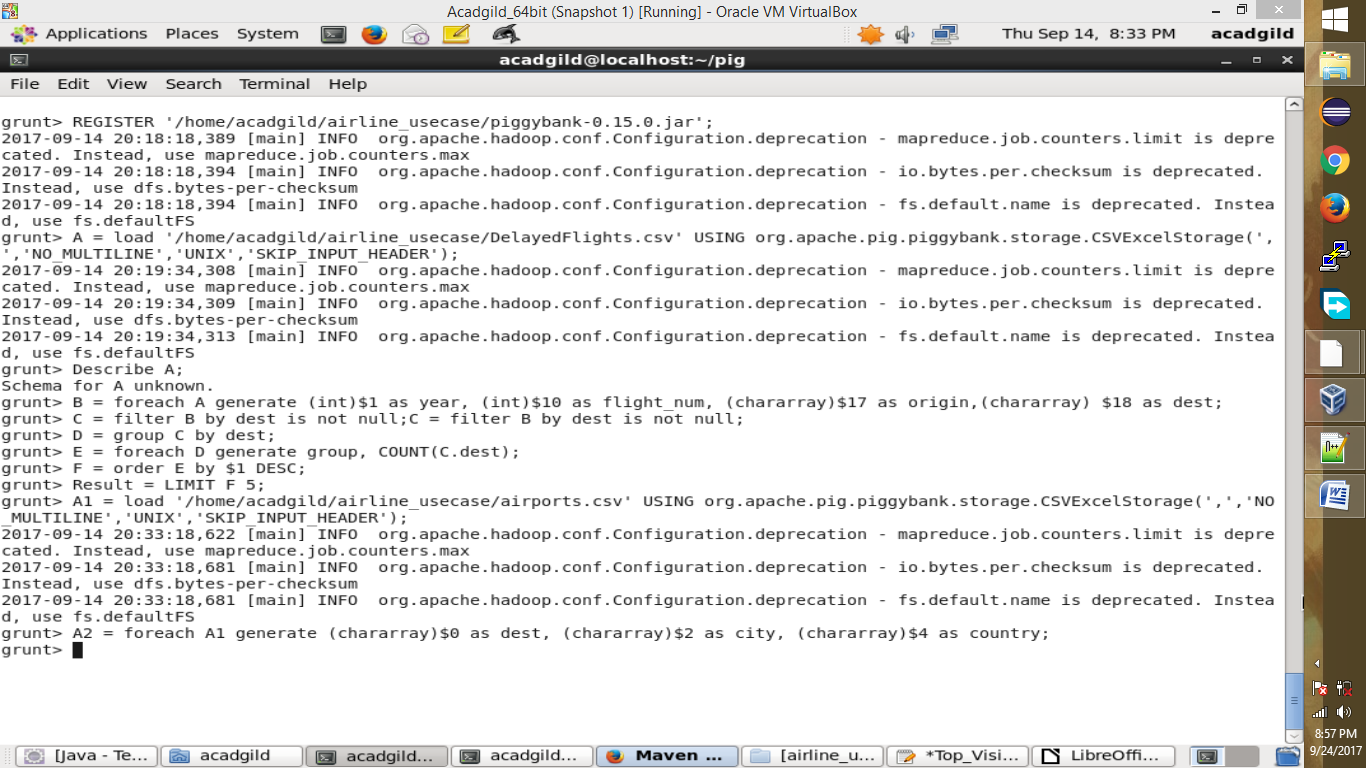
Step 9: These are the steps to find the top 5 most visited destinations. However, adding few more steps in this process, we will be using another table to find the city name and country as well.



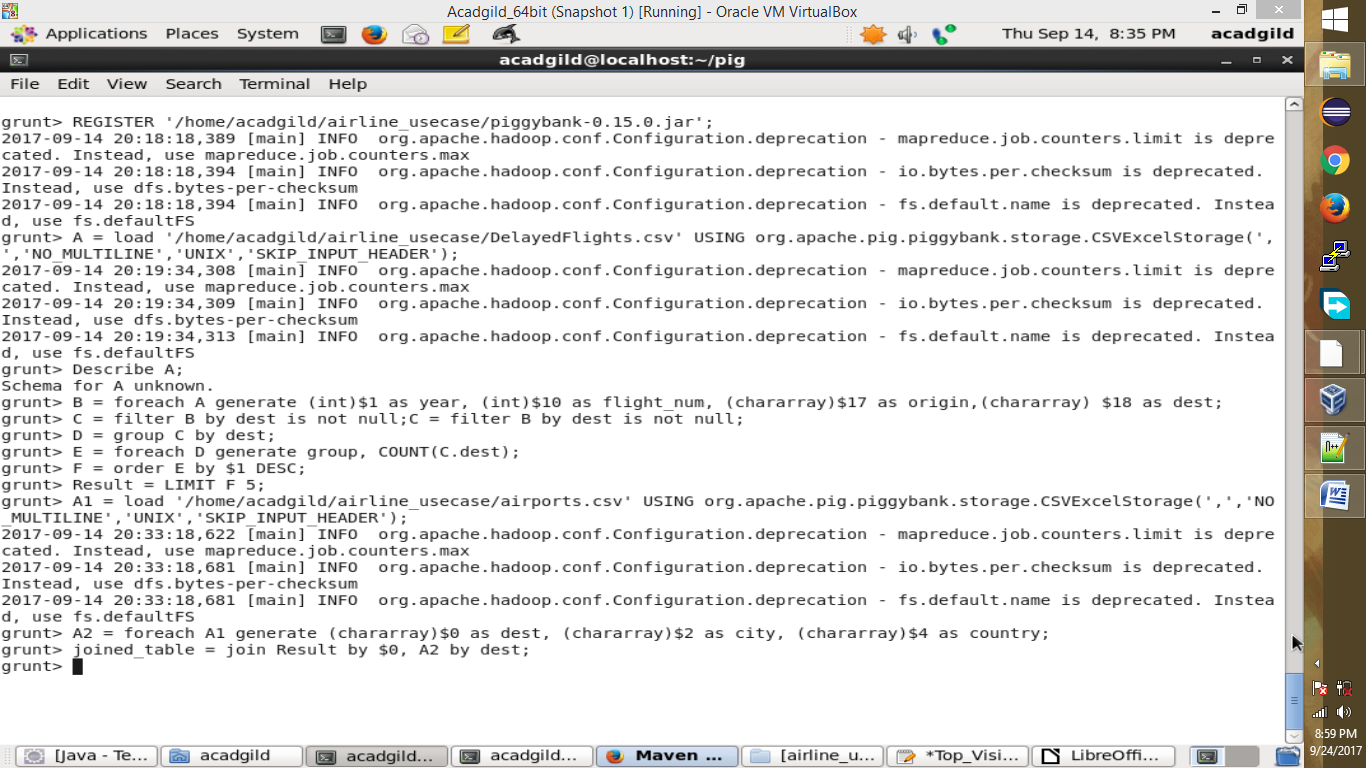
Step 10: In relation **A1**, we are loading another table to which we will look-up and find the city as well as the country.



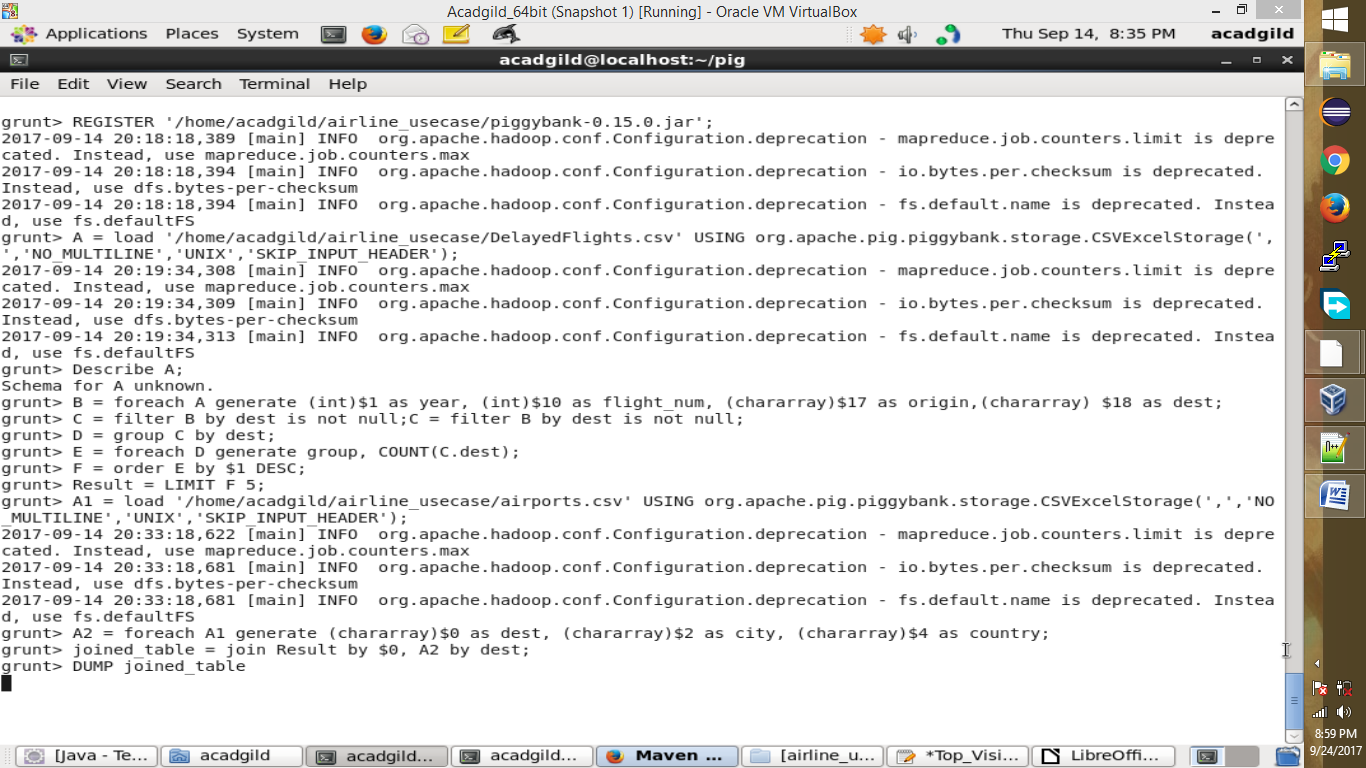
Step 11:In relation **A2**, we are generating dest, city, and country from the previous relation.



Step12: In relation **joined\_table**, we are joining Result and A2 based on a common column, i.e., “dest”



Step 13: Finally, using dump, we are printing the result.



Step 14: Top 5 most visited destinations are printed in the result

