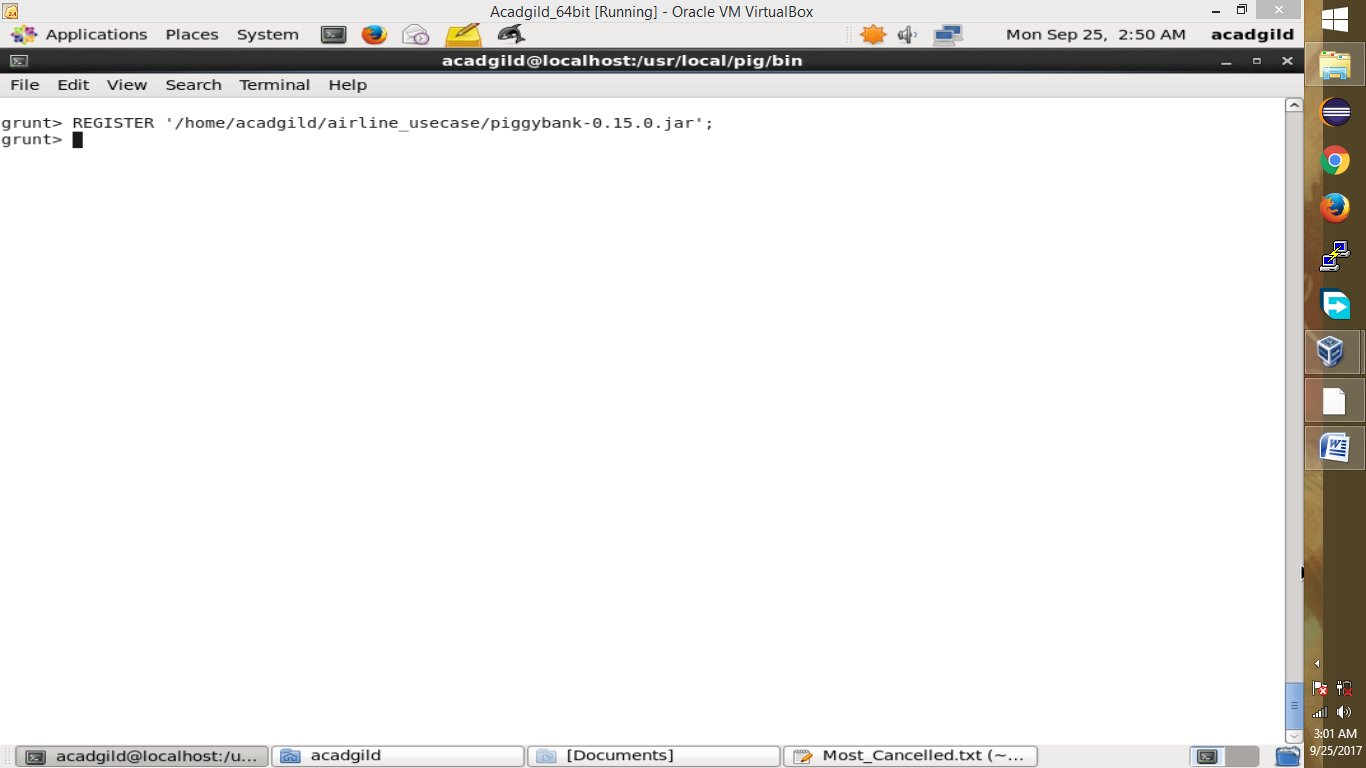
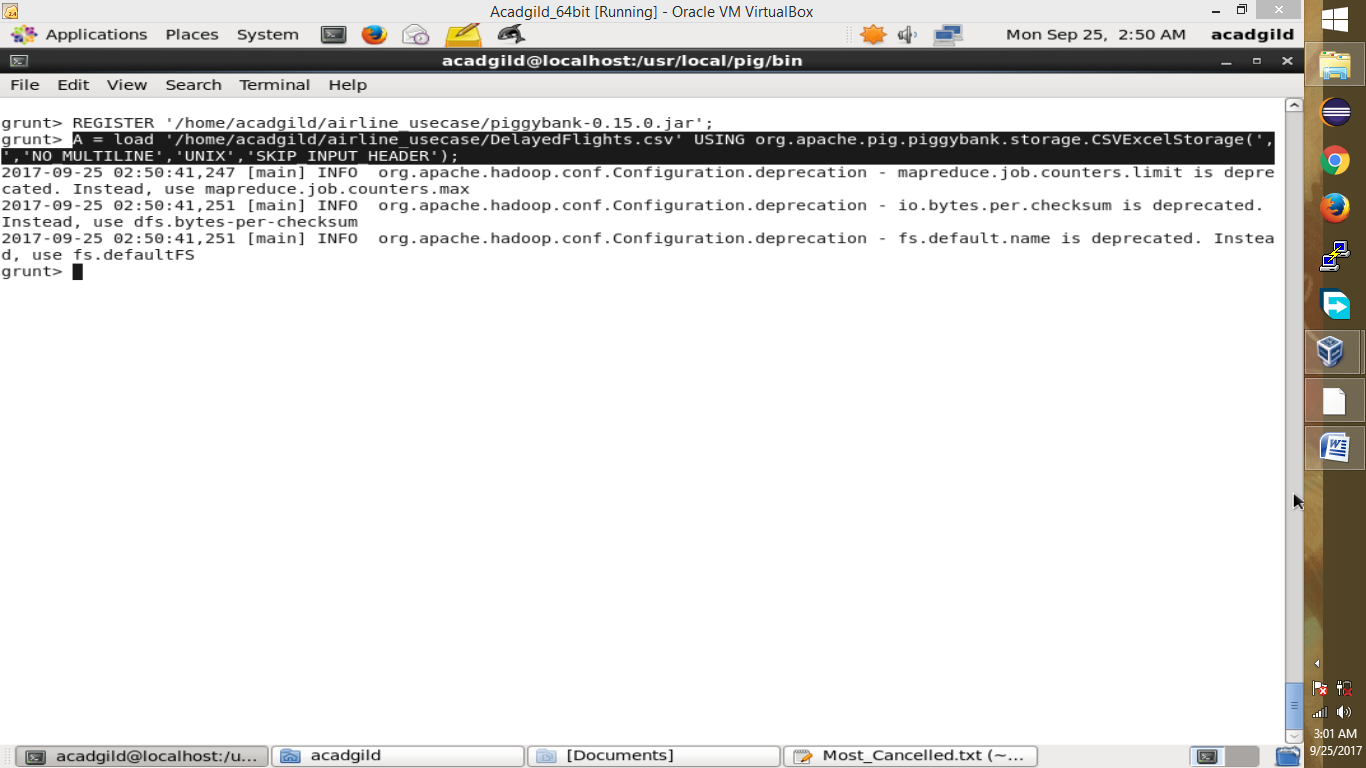
**Problem Statement 2**

**Which month has seen the most number of cancellations due to bad weather?**

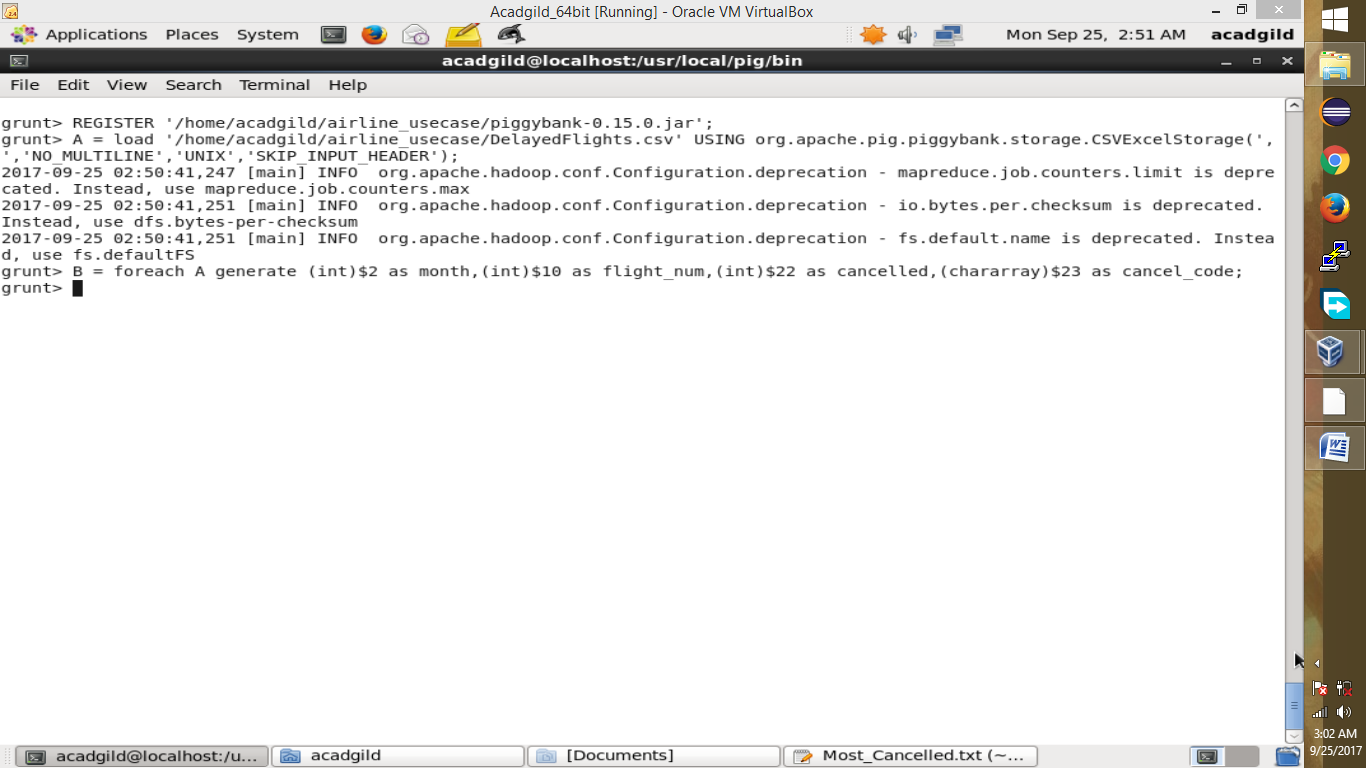
**Step 1** : We are registering *piggybank* jar in order to use the CSVExcelStorage class.



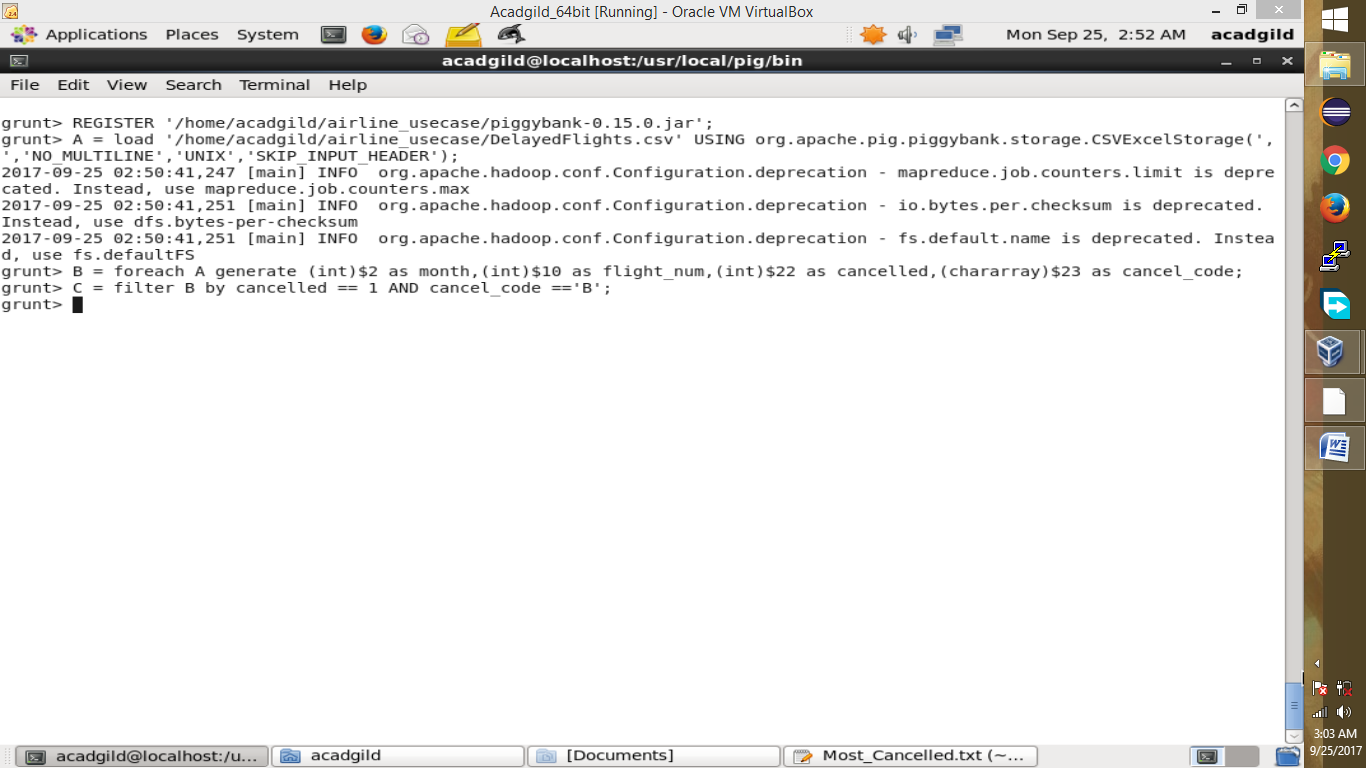
**Step 2:** In relation **A**, we are loading the dataset using CSVExcelStorage because of its effective technique to handle double quotes and header.



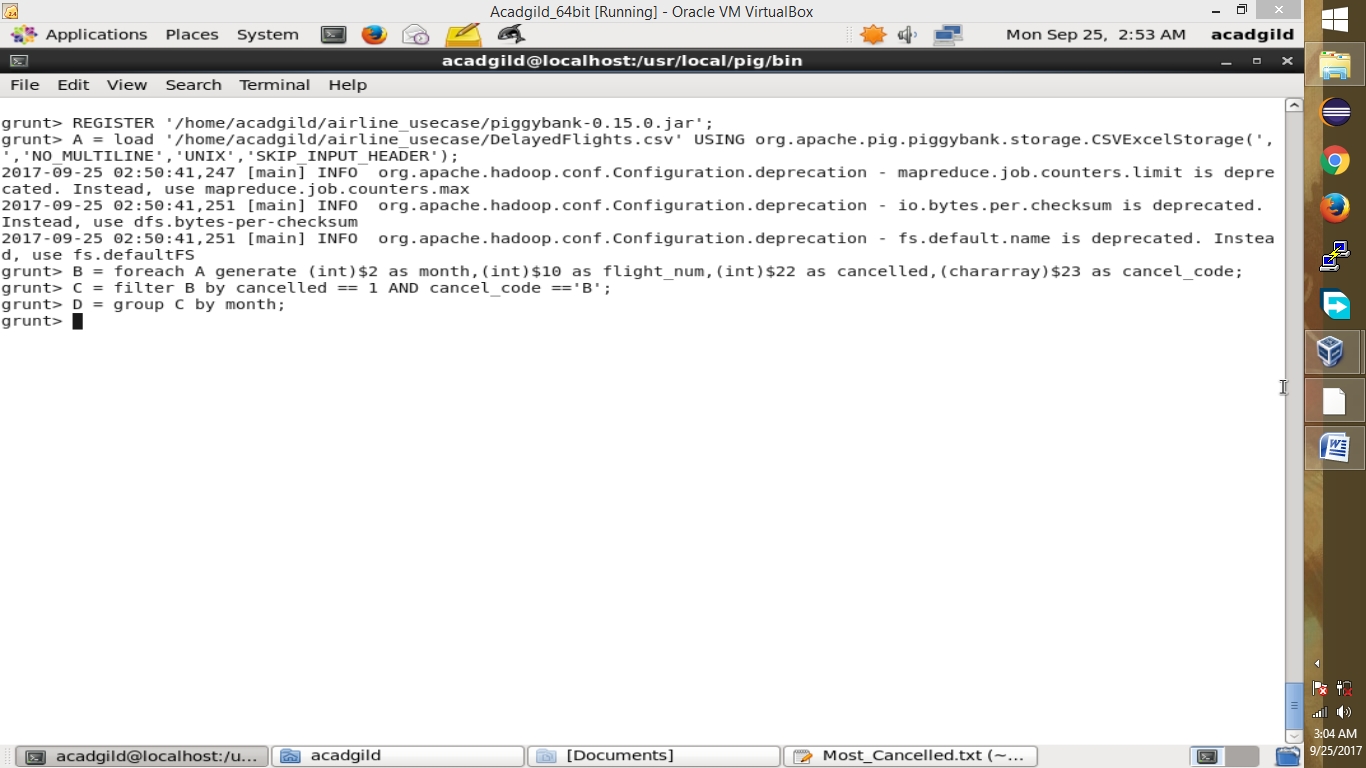
**Step 3:** In relation **B**, we are generating the columns which are required for processing and explicitly typecasting each of them.

****

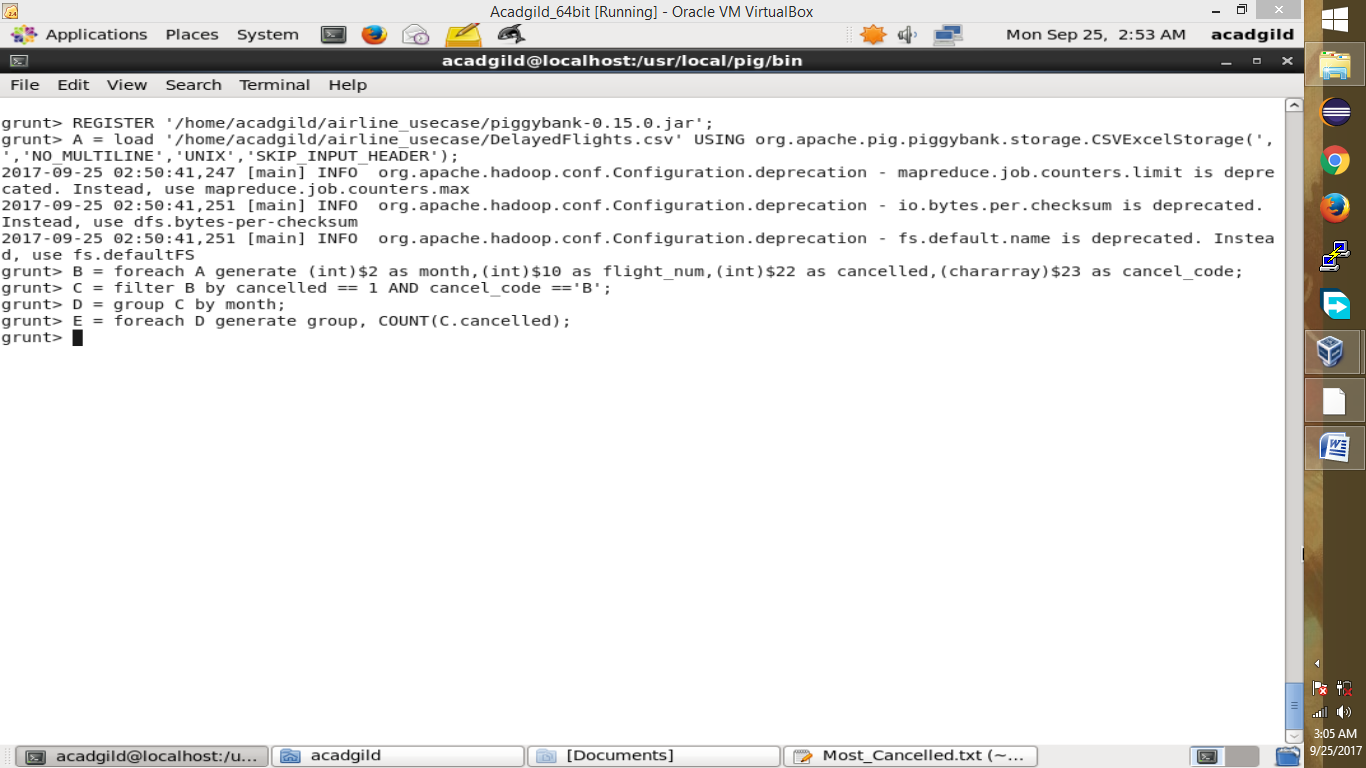
**Step 4:** In relation **C**, we are filtering the data based on cancellation and cancellation code, i.e.,  canceled = 1 means flight have been canceled and cancel\_code = ‘B’ means the reason for cancellation is “weather.” So relation C will point to the data which consists of canceled flights due to bad weather.

****

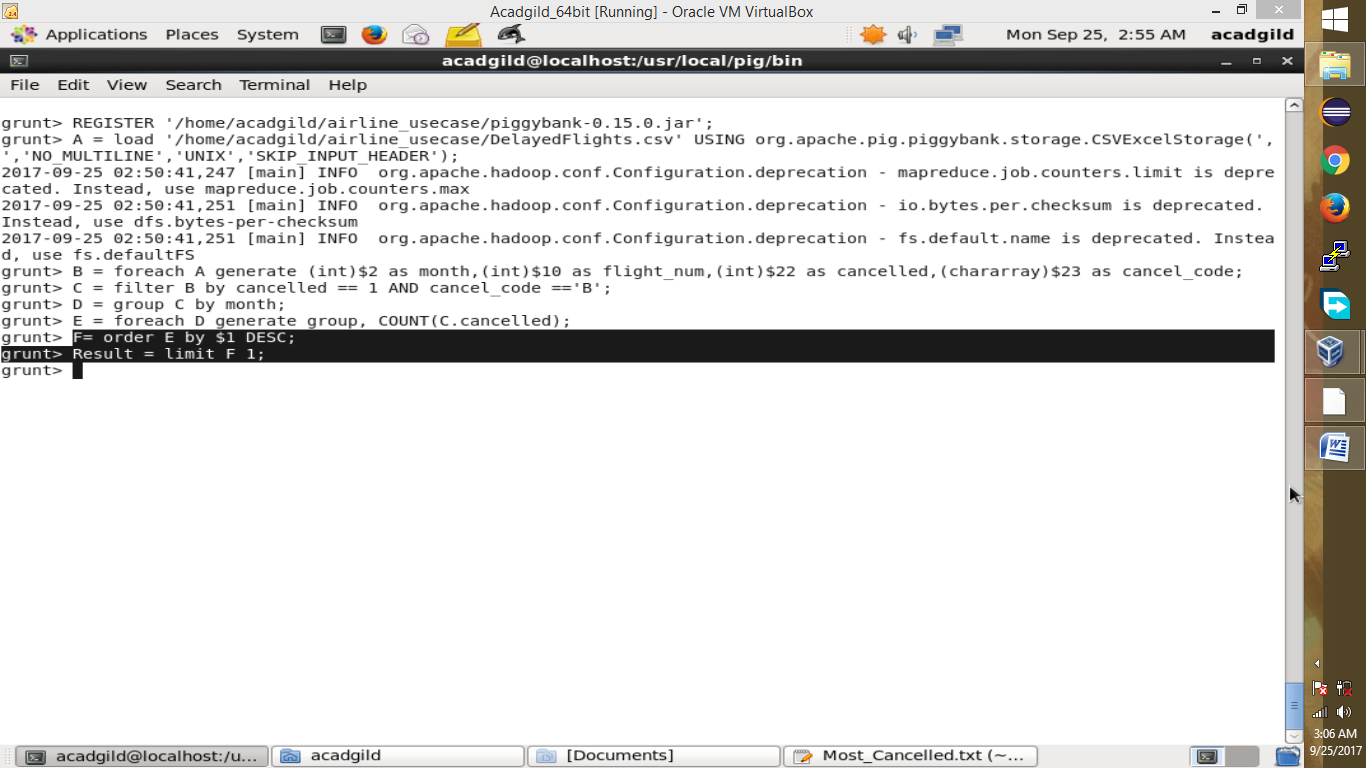
**Step 5:** In relation**D**, we are grouping the relation C based on every month.

****

**Step 6:** In relation **E**, we are finding the count of canceled flights every month.

****

**Step 7:** Relation **F** and **Result** is for ordering and finding the top month based on cancellation.



**Step 8:** Most number of cancellation s due to bad whether are printed below:

