# Aviation Data Analysis

**Problem Statement 1**

**Find out the top 5 most visited destinations.**

First egister the piggybank jar.

1. Load the data using CSVExcelStorage( ), which can handle double quotes and headers.



1. Fetch only columns which are required accoding to problem statement.



1. Filter out the null data.



1. Since we need to find out top 5 destination, do group by destination

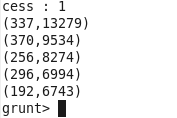


1. Count:



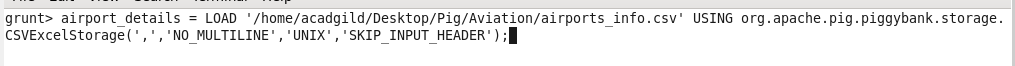
1. Top 5 visited destinations in descending order:





Now join this result data with the Airport details to obtain final complete results.

1. Load the Airport details.



1. Defining the columns of Airport\_details data:



1. Now join both the data.



OR



1. Fetch the final results.

## Problem Statement 2

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

1. Load data:



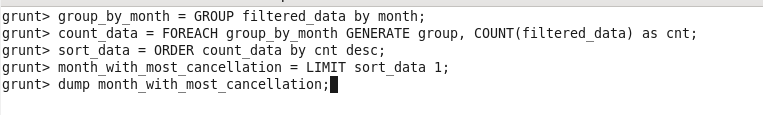
1. Fetch only required data.



1. Filter the data which contains flights cancelled due to bad weather.



1. Group the results by month and arrange the grouped data in descending and top most data is the one which has seen the most number of cancellations due to bad weather.

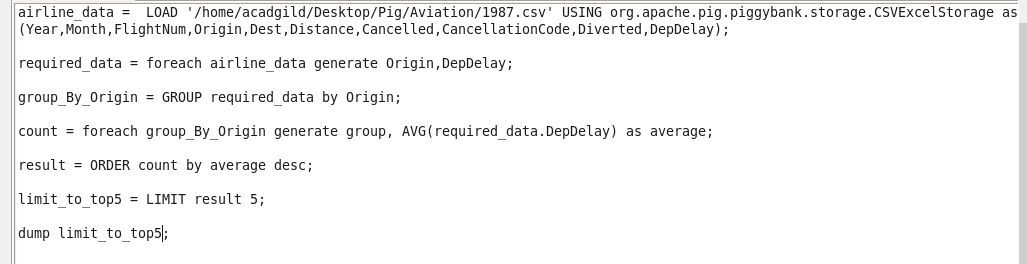




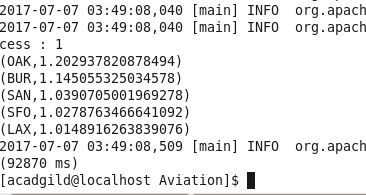
## Problem Statement 3

Top ten origins with the highest AVG departure delay

Below screenshot is the script to achieve this task. Executing this is local mode.



Output:



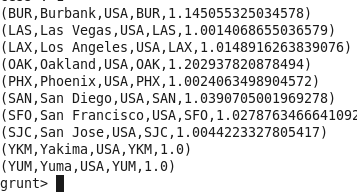
Now join the result with airport info.







Result:



## Problem Statement 4

Which route (origin & destination) has seen the maximum diversion?

1. Load data and fetch only required data.



1. Filter only data which seen Diveersion.



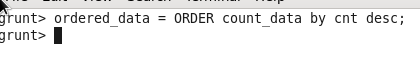
1. Group the data with Origin and dest.



1. Count the grouped data



1. Now sort by descending order



1. Top ten route (origin & destination) which has seen the maximum diversion

