# PDP Assignment 2

Samuel Kuik,

580221,

21-06-2022

<https://github.com/Zalmwel/PDP>

## Explanation

First off, we need to download the orders.csv dataset from Moodle. From the Ambari dashboard, navigate to the Files View -> user/maria\_dev, create a folder called ‘diplomacy’ and move the file into this folder.

We load the CSV file ‘orders.csv’, set this to ‘orders’ and set the column names as variables. We use CSVExcelStorage, as PigStorage was unfortunately not working (for me) with this csv-file.

Text

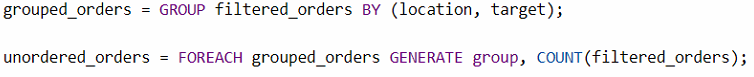
Description automatically generated with low confidence

Next, we filter for target ‘Holland’, to speed up the group query. Otherwise, the group query will group *all* the data in the csv, which we do not want.



We then group the data by location and target, using our filter for target; ‘Holland’.

We iterate over each group and count the amount records in the group, meaning we take the sum of each unique, grouped location.



Finally, we order the groups alphabetically, naturally starting from the first position (indicated by $0). We want to list the groups from A to Z, which is done with the ‘ASC’ keyword.

To finish up, we print the results using ‘DUMP’.

A picture containing text

Description automatically generated

## Results

Text

Description automatically generated