Inclusion Dependency Discovery - Wild Catz

Step 1

Take the distinct elements of a column and store them in a new RDD -> Col A

Step 2

Compare the RDD to all other columns in all csv-files -> Col B

- Is A a subset of B; Yes, then A<B
- Is B a subset of A; Yes, then B<A

Step 3

Save the result into a map with dependent columns as keys and a list of referenced columns as values

--> Repeat from Step 1 with a new column

Finally, sort the map lexicographically by key and print each Key-Value pair to the console :)