Submission sho Assignment II	eet
	ntains questions to help you reflect about the operations applied to s assignment. You have to fill it in and submit it on Studium.
TASK 1: warm u	p
Number of records	9816 (after removing duplicates)
Number of attribut	tes (excluding User and Query) 500
TASK 2: frequen	t itemsets
What is the most f	requent item (i.e., single keyword)? "of"
In how many quer	ies does it occur: 955
Which support val	ue corresponds to a support count of 100 records? 100/9816 = 0.0101874491
How many frequer	nt itemsets have you found with this min-support? 28
What is the maxim	num size of your frequent itemsets? 2
	of the support parameter
-	nt itemsets for some selected values of support:
Mininum support	Number of Frequent Itemsets
0.001	855
0.004	140
0.007	55
0.01	27
TASK 4: rule gen	in-support leads to the discovery of about 150 itemsets? 0.004 -> leads to 140 itemset eration ave you identified with min-confidence .8?
Indicate a high-cor	nfidence rule $X \rightarrow Y$ where $Y \rightarrow X$ has lower confidence.
	York

Why is the confidence of Y -> X lower than the confidence of X -> Y?

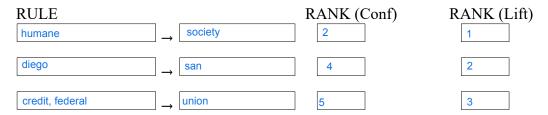
TASK 5: impact of confidence

Number of rules for some selected values of minimum confidence:

Mininum confidence	Number of Rules
0.1	18
0.3	14
0.5	7
0.7	5
0.9	1

TASK 6: rule interpretation

What are the top-3 interesting rules, and what is their rank w.r.t. confidence and lift?



What are the top-3 unexpected rules, and what is their rank w.r.t. confidence and lift?

