

Q1)The missing country is most likely to be Canada. My hypothesis is based on the data analysis in "dataexploration.ipynb".

Q2) The recommendation in the question is on the assumption that the user is a new user (his previous search record is unknown). The implemented function is non-personalized recommendation as shown in "non_personalized_recommend" module, it recommend cities searched most or cities that related to searched cities in the session.

Q3) The feature of users: user_id, joining_date, country is useful to predict most likely search cities. Once these information is available, it means that the website knows the user is a registered user and has his record history. It becomes a personalized recommendation problem and implemented in "city_recommend" module. The intuition is item-item CF recommendation.

Q4) For implicit recommendations, 20% of the user/item interaction in original data are masked as "0" (no interaction), it is taken as the training data. For the test data, each user contains at least one masked item. Calculate ROC, and AUC, get the average AUC.