Marketing Proposal for the WTA

At the WTA (Women's Tennis Association), the marketing team would like to display the 2016 tournament winners and their total play time in minutes for their winning match. Given the WTA data, sort the players alphabetically (to avoid favoritism), as well as the country they represent, with their total final match play time for all tournaments won.

The data includes information about the winner and the loser, as well as the tournament information and final match information for the 2016 tournament year. The data was downloaded from https://www.kaggle.com/gmadevs/wta-matches.

To solve this problem, the python MRJob code included a single mapper and reducer. The mapper took the 49 columns, and yielded the winners' names and representing countries, as well as the total minutes of final match play time. The reducer took this information and yielded the player's information, along with the total final match play time for all the tournaments they won. The output was a text file, *winner.txt*.

A snapshot from Excel of the data is given below:

["Caroline Wozniacki", "DEN"]	208
["Christina Mchale", "USA"]	66
["Darya Kasatkina", "RUS"]	135
["Dominika Cibulkova", "SVK"]	131
["Ekaterina Makarova", "RUS"]	135
["Eugenie Bouchard", "CAN"]	183
["Irina Camelia Begu", "ROU"]	86
["Irina Falconi", "USA"]	154
["Jelena Ostapenko", "LAT"]	72
["Julia Goerges", "GER"]	356
["Katerina Siniakova", "CZE"]	139
["Kirsten Flipkens", "BEL"]	234
["Madison Brengle", "USA"]	105
["Marina Erakovic", "NZL"]	119
["Monica Niculescu", "ROU"]	106
["Nao Hibino", "JPN"]	247
["Naomi Broady", "GBR"]	231
["Nicole Gibbs", "USA"]	50
["Qiang Wang", "CHN"]	214
["Roberta Vinci", "ITA"]	206
["Saisai Zheng", "CHN"]	41
["Samantha Crawford", "USA"]	225
["Samantha Stosur", "AUS"]	133
["Shuai Zhang", "CHN"]	67
["Sloane Stephens", "USA"]	441
["Su Wei Hsieh", "TPE"]	85
["Svetlana Kuznetsova", "RUS"]	49
["Tamira Paszek", "AUT"]	434
["Timea Babos", "HUN"]	252