Horizon Data Engineering Coding Assessment

There are four questions in this assessment. You are encouraged to import/install any packages you would like. Use of Python is required, pandas is highly encouraged.

For our dataset we have prepared this excel sheet https://storage.googleapis.com/hznp_public/Horizon_Data_Engineering.xlsx. In this excel sheet there are the following sheets.

- hcp : Healthcare provider table
- hco : Healthcare organization table
- ziptoter : maps zip codes to territories

For each question, please provide a data prep section and an answer section. Do all prep for the question in the data prep section and print your final outputs in the answer section. Finally save the final output for each question to an excel sheet.

The four questions are below.

1. What is the number of distinct specialties per territory?
2. What is the straight-line distance (km) for each hcp with a RHEUMATOLOGY or NEPHROLOGY specialty to the closest IC (infusion center) in their territory?
3. For each IC, what are the closest 3 hcps with a RHEUMATOLOGY or NEPHROLOGY specialty in the ICs territory?
4. All hcps within 100 miles for each HOSP (hospital) only if they are not closer to another HOSP.