Scott Berry

Text

Description automatically generated

For ease, I moved “people.json” and “people.txt” to my working directory. I used the initially recommended methods of native python csv and json libraries to import the data, but Spark would throw many WARNs about how the methods had become deprecated, recommending the new Spark-native methods to Spark 2.0+. My final code uses these Spark-native libraries. The four outputs below are (in order): 1) the schema of the DF created from “people.json”, 2) the aforementioned DF, 3) the distinct names from said DF, and 4) the DF created from “people.txt”.

Text

Description automatically generated

2

3

4

1