Assignment Questions:

**Instruction:** Please submit the PySpark code in jupyter notebook

**Questions:**

1. Write a Pyspark code for decision tree like exercise you have completed in pyhton.
2. Create array of names and publish name and length of each name.
3. Create array of names and find max, min, standard deviation, variance of length of each element in array.
4. Use data set “ocp\_data.tsv” attached in sheet (data is delimited with tab space)
   1. List no. of shows per year.
   2. List no. of shows per year for every occupation
   3. Find top occupation for every year.
   4. Which guest have attended maximum shows
   5. List occupation under each group
   6. How many show per group of year and occupation

