DA 331 - Big Data Analytics : Tools & Techniques

Lab 1

Instructor:	Dr. Chiranjib Sur	**Time:**	Monday - 3:00-4:00 (5201)
			Tuesday - 2:00-3:00 (5201)
			Wednesday - 9:00-11:00 (MDSAI Lab)
Email:	chiranjib@iitg.ac.in	**Place:**	MDSAI Lab.

Problem 1:

Given the dataset, find the followings:

1. Find the average number of household for different geolocation in Spark?

2. Compare the time for execution for the above in Spark and in Pandas? Use the "import time; # record start time ; start = time.time()" code to record the time.

3. Which geolocation has the highest median income in Spark?

4. Compare the time for execution for the above in Spark and in Pandas? Use the "import time; # record start time ; start = time.time()" code to record the time.