**Assignment 1:**

A survey was conducted to gauge an audience interest in different data science topics, namely:

Big Data (Spark / Hadoop)

Data Analysis / Statistics

Data Journalism

Data Visualization

Deep Learning

Machine Learning

The participants had three options for each topic: Very Interested, Somewhat interested, and not interested. 2,233 respondents completed the survey. The survey results have been saved in a CSV file and can be accessed through this link: https://cocl.us/datascience\_survey\_data.If you examine the CSV file, you will find that the first column represents the data science topics and the first row represents the choices for each topic. Use the pandas read\_csv method to read the CSV file into a pandas dataframe, that looks like the following:

**Code Solution:**

Import pandas as pd  
df=pd.read\_csv(‘https://cocl.us/datascience\_survey\_data’)  
df.head()  
df.coloums=[‘topic’, ‘very interested’, ‘somewhat Interested’, ‘Not Interested’]  
df.head()

|  |  |  |  |
| --- | --- | --- | --- |
|  | Very interested | Somewhat interested | Not interested |
| Big Data (Spark / Hadoop) | 1332 | 729 | 127 |
| Data Analysis / Statistics | 1688 | 444 | 60 |
| Data Journalism | 429 | 1081 | 610 |
| Data Visualization | 1340 | 734 | 102 |
| Deep Learning | 1263 | 770 | 136 |
| Machine Learning | 1629 | 477 | 74 |