

Spring 2019

Homework: Probability and Statistics - Week 2

Evan Agovino
CUNY City College

NYC Tech-in-Residence Corps

Follow this and additional works at: https://academicworks.cuny.edu/cc_oers



Part of the [Computer Sciences Commons](#)

[How does access to this work benefit you? Let us know!](#)

Recommended Citation

Agovino, Evan and NYC Tech-in-Residence Corps, "Homework: Probability and Statistics - Week 2" (2019).
CUNY Academic Works.

https://academicworks.cuny.edu/cc_oers/169

This Assignment is brought to you for free and open access by the City College of New York at CUNY Academic Works. It has been accepted for inclusion in Open Educational Resources by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.

```
In [ ]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

1. Answer Questions 1 through 5 of part 6 of the classwork. You can work right in the notebook from class if you wish, or create a new notebook (or paste it below). Please copy and paste each question into its own cell with answers below it.

2. Answer Questions 1 through 10 from [here \(https://www.kaggle.com/kashnitsky/a1-demo-pandas-and-uci-adult-dataset\)](https://www.kaggle.com/kashnitsky/a1-demo-pandas-and-uci-adult-dataset) below (please copy and paste each question into its own cell with answers below it). I've loaded in the code for you and cleaned it up a bit.

```
In [ ]: #AdultDataSet
df = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data')
df.columns = ['age', 'workingclass', 'fnlwgt', 'education', 'education-num',
'martial-status', 'occupation', 'relationship', 'race', 'sex', 'capital-gain',
'capital-loss', 'hours-per-week', 'native-country', 'salary']
df['salary'] = df['salary'].str.strip()
df['native-country'] = df['native-country'].str.strip()
```

3. Think of some datasets you might be interested in analyzing and post two below. You can either use the `pd.read_table` to pull a table off of a page like Wikipedia or `pd.read_csv` to read a CSV from a site. Get creative! If you're not sure where to start, google 'free datasets' to get a sense of what's out there. Everything is only a search away. If there are datasets you're interested in working with but can't find the data you're looking for or are having trouble finding them

```
In [ ]: 
```