

DATA SCIENCE FROM SCRATCH

Workshop 1

Women in Data Science Meetup

March 17, 2016

Host: HomeAway

INTRODUCTIONS

- Thank you to HomeAway for providing the hosting site and food, drinks!
Thank you Jill!
- Thank you to Intersys for providing the meetup fees this year!
- Logistics - we will be meeting for about 6-9 months to go over this workshop. Notes will be posted in the meetup group.
- Book we are using : “Data Science from Scratch” - check for discount code on the meetup page
- Using slack to discuss - please join if you are a woman in data science - we will restrict the slack page to those who identify as women:
- <https://women-in-ds-atx.slack.com/>

AGENDA

Introduction/Logistics

Quick guide to command line - Randi

Getting into groups

Python Tutorial - Chai and Meghann

Download and start playing around with the Titanic dataset

Assignment

DATASET

Dataset for this workshop:

<https://www.kaggle.com/c/titanic>

We will use the Titanic dataset because SO much work has been done with this dataset and there are plenty of supplemental articles and tutorials on this dataset.

Please use a dataset from the following slide in addition if you want more of a challenge where you will have to work on the data independently or with your small group.

OTHER DATASETS

So many free fun datasets, some of my favorite:

<https://www.drivendata.org/>

<https://www.kaggle.com/datasets>

<http://archive.ics.uci.edu/ml/> (old school!)

Local:

<https://data.texas.gov/browse?limitTo=datasets>

<https://data.austintexas.gov/>

Other lists:

<http://www.datasciencecentral.com/profiles/blogs/big-data-sets-available-for-free>

<https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public>

<http://www.kdnuggets.com/datasets/index.html>

EXPECTATIONS

You will get out of this workshop what you put in! So dig into the datasets and get your hands dirty working out the code!

Today:

Make sure you have Python/R/Ipython running on your machine.

Have access to Github

Can use the command line

Download Titanic Data Set

Meet 2-4 people who will be your go-to for getting through this material.

Know what different Python data structures are about.

RANDI!

GET INTO GROUPS

Connect with 2-4 people who have the same color sticker on their nametag as you.

PYTHON TUTORIAL

Using Python 2.7 (or Python 3.5 for the brave) in this class.

DOWNLOAD AND PLAY AROUND WITH TITANIC DATASET

Next month we will go into data exploration with Pandas – but for now let's download the data locally and load the data into python using the `pandas.read_csv()` function.

ASSIGNMENT

Expectations - Did you get Python/R on your machine? Did you use the command line? Connect with 2-4 other people in the group?

Read through Chapter 1, 2, and Pandas Docs:

<http://pandas.pydata.org/pandas-docs/stable/10min.html>

We will be going over this next time: <https://www.kaggle.com/c/titanic/details/getting-started-with-python-ii>