<https://www.youtube.com/watch?v=Q73ADVZCqSU>

**Data Analysis of Uber trip data using Python, Pandas, and Jupyter Notebook**

* In jupyter notebook ->
  + "**data = pandas."** and the press **Tab** -> this shows you all the options under that library
  + "data = pandas.read\_csv**?**" -> the question mark shows you the documentation of the command line

* %**pylab** is a "magic function" that you can call within IPython, or Interactive Python. By invoking it, the IPython interpreter will import matplotlib and NumPy modules such that you'll have convenient access to their functions.

* For commenting and documentation change cell from code to **Markdown**
  + After **# or ## leave a space**

* **"dt.weekday()"** -> unlike others (e.g. dt.day, dt.month, dt.hour etc.) variable.weekday() is a function rather than a variable
  + Weekday() return a number from 0(Monday) to 6(Sunday)

* How to create functions :
  + **def get\_dom(dt):** #function is created using def function\_name(variable):

return dt.day #function definition

* **xticks(range(7),'Mon Tue Wed Thu Fri Sat Sun'.split())** -> To create your own elements for x-axis

* by\_date = data.groupby('dom').**apply(count\_rows)** -> applying function count\_rows to the data set

* data['Date/Time'] = data['Date/Time'].**map(pandas.to\_datetime)** -> modifying data in the table
* **data['hour']** = data['Date/Time'].map(get\_hour) -> adding new data in the table

* bt\_h\_d = **data.groupby('weekday hour'.split())**.apply(count\_rows) -> grouping multiple categories together
  + by\_cross = bt\_h\_d.**unstack()** -> weekdays as rows(y-axis) and hours in column section(x-axis)
  + **seaborn.heatmap(by\_cross)** -> heat map for the correlated data
* To show two histograms in the same chart -> execute both the histogram commands in the same cell
  + hist(data['Lat'], bins=100, range=(40.5,41))

hist(data['Lon'],bins=100,range=(-74.1,-73.9)**)**

* hist(data['Lat'], bins=100, range=(40.5,41), **color='g', alpha=.5**)

**twiny() #for upper and lower x-axis**

hist(data['Lon'],bins=100, range=(-74.1,-73.9), **color='r', alpha=.5)**

* **grid()** #for gridlines

* hist(data['Lat'], bins=100, range=(40.5,41), color='g', alpha=.5, **label = 'latitude'**)

**legend(loc='upper left')** -> loc indicates location

twiny() #for upper and lower x-axis

hist(data['Lon'],bins=100, range=(-74.1,-73.9), color='r', alpha=.5, **label = 'longitude'**)

**legend(loc='best')**

* plot(data['Lat'] ,'\*', ms=10, color='red', label='lat')
  + **xlim(0,100)** -> sets x-axis limits
  + **'\*'** -> denotes a star be used instead of lines on the graph ('.' , '^', etc. can also be used)
  + **ms** -> denotes size of the symbol

* **figure(figsize=(20,20))**
  + To change the size of a graph
  + To be used before plotting the chart