PANDAS Tutorial Presentation

Emily

January 15, 2018

Introduction

- This tutorial will guide you through the basics of PANDAS, a package that works in tandem with num.py and python
- After completing this tutorial you should have a better understanding of the foundations of PANDAS

How to use this tutorial

- Before starting the tutorial you should pip install the requirements.txt a
- To get the most out of this tutorial start with the README, then these slides, then whichever notebooks are most relevant to your needs.
- One could either do the individual tutorials and walkthrough first, or if you are looking to solve a specific problem with PANDAS, look to see if there is a notebook on that issue.

Graph

```
df.describe()
```

	Miles	Seconds
count	168.000000	52.000000
mean	1.512738	1901.788462
std	2.966583	2560.424171
min	0.000000	376.000000
25%	0.000000	583.750000
50%	0.000000	1343.500000
75%	2.002500	1743.250000
max	26.200000	15643.000000

Figure 1: Above is an example of a dataset being analyzed using PANDAS' describe feature.

Code

Here is an example of code from the "Rolling Apply" notebook. Rolling apply is especially useful as it applies a function across a DataFrame or a series.

```
def times_3 (x):
    if x >4:
        return x * 3
    else:
        return x)
df3['A'].apply(times_3)
```

Summary

- PANDAS can be used in the lab for manipulating data in many different ways
- It has particular strengths in organizing and visualizing large sets of data
- If you want to learn more about PANDAS here are some useful places to look:
  - https://pandas.pydata.org
  - https://www.datacamp.com/community/blog/python-pandas-cheat-sheet