



Kung-fu Pandas 101
DVTO #5 – 13 Feb 2014

founders@vantageanalytics.com

Panda != Cute furry animal

Panda == Panel Data

**** There will be no pictures of cute furry animals tonight ****

Who am I?

- Business Degree from Military University
- 8 Years in Software Development @ Workbrain, NexJ
- 4 Years running R&D at predictive analytics software vendor Angoss
- Currently CTO @ Vantage Analytics

Python Data Science Landscape



Pandas in the Vantage Stack

- We use Pandas on the server (no Ipython...)
- Data stored in MySQL, which is better at filtering and managing large datasets
- Read from MySQL into Pandas for analysis and manipulation of small(ish) datasets
- Use AngularJS, Django, TastyPie to interact with results of analysis

3 MINUTE DEMO OF HOW AWESOME VANTAGE IS

Pandas Curriculum

- 101 – Terminology, Intro to Data Access, Recoding, Basic Filtering and Summarization
- 102 – Data Import/Export, Advanced Summarization
- 103 – Advanced Boolean Indexing, datatypes, missing data, pivots, crosstabs, and more
- 201 – Time Series, Forecasting
- 202 – Integration with R via Rpy2
- 30x – Lots more we can do!

Terminology

DataFrame

Columns			
	Age	Gender	City
Index			
1	34.3	1	Toronto
2	35.6	1	New York
3	43.3	2	Toronto

Getting at the Data

If you use	You get a
<code>df['column']</code>	Series
<code>df[['col1','col2']]</code>	DataFrame with two columns
<code>df[0]</code>	big nasty error
<code>df.iloc[0,0]</code>	The value for a single cell
<code>df.ix[index, 'col1']</code>	The value for a single cell

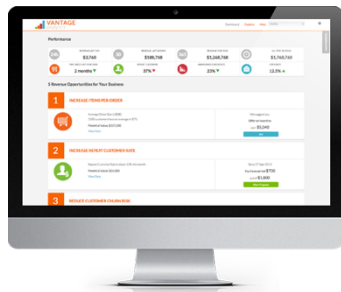
Numbers! And other stuff...

- nan
- inf
- $1 / 2 = ?$
- $1.0 / 2 = ?$
- `DataFrame([1.0, 2.0, nan]).mean() = ?`
- `DataFrame([1.0, 2.0, inf]).mean() = ?`

Ok, enough talk.
Let's get to some code...

“Vantage not only made our data accessible and useable, it helped us to better understand our customers and grow our business.”

- Brian Sharwood, President, HomeStars



Data >> Insight >> Action