Oleksandr Romanko, Ph.D.
Senior Research Analyst, Risk Analytics, Watson Financial Services, IBM Canada
Adjunct Professor, University of Toronto

MIE1624H – Introduction to Data Science and Analytics

Lecture 2 – Python Programming

University of Toronto
September 25, 2018

Lecture outline

Python essentials

- IPython notebooks
- Modules
- Variables and types
- Operators and comparisons
- Compound types - strings, tuples, lists and dictionaries
- Control flow - conditional statements (if, elif, else), loops
- Functions
- Classes
- Files and the operating system
- Exception handling

Lecture outline

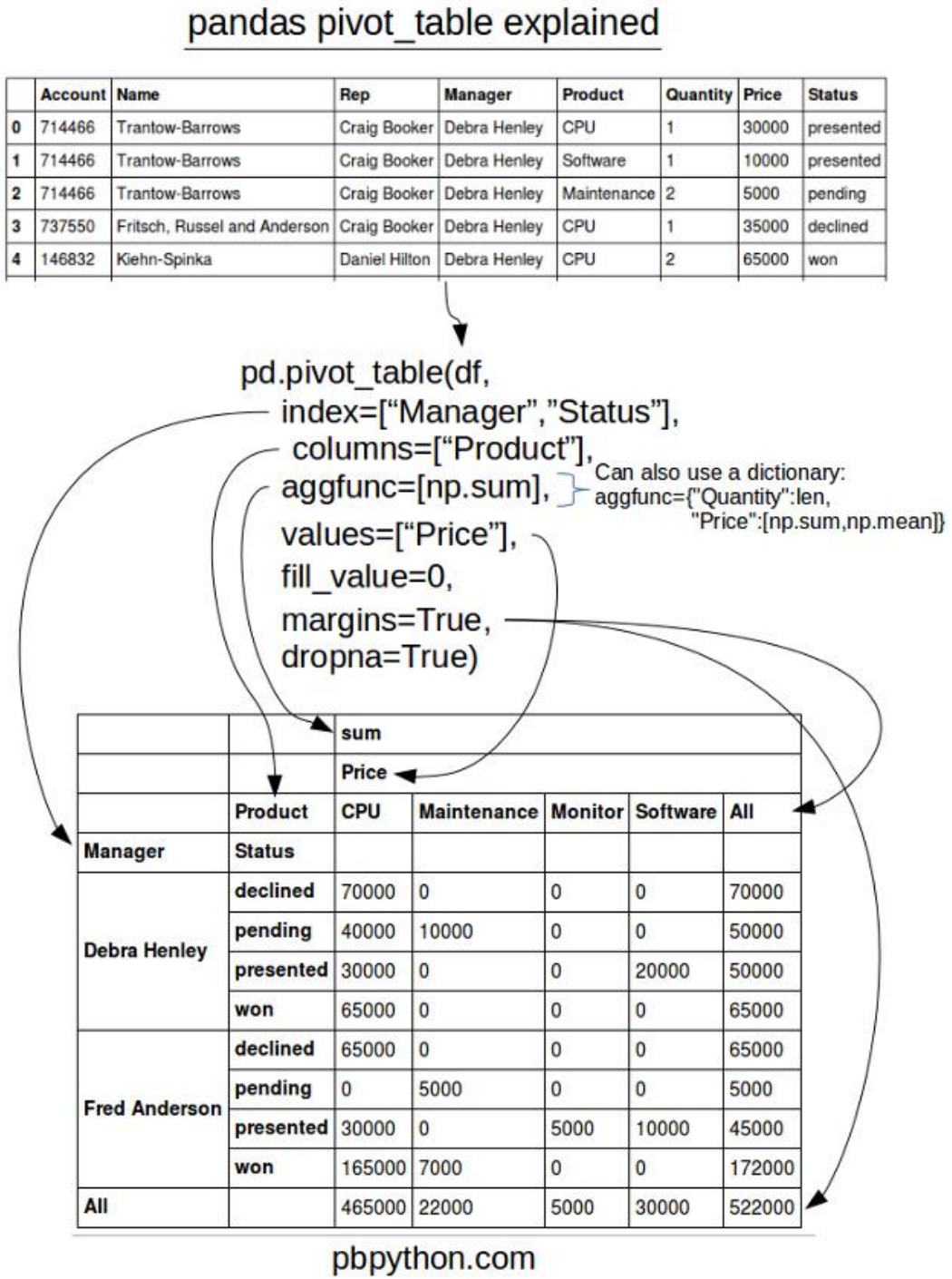
Introduction to Pandas

- Introduction to pandas data structures – DataFrame, index objects
- Pandas essential functionality
- Summarizing and computing descriptive statistics
- Pivot tables in pandas

Web-scrapping with Python

Introduction to Data Science and Analytics (continuing Lecture 1)

Pandas pivot_table cheat sheet



	Account	Name	Rep	Manager	Product	Quantity	Price	Status
0	714466	Trantow-Barrows	Craig Booker	Debra Henley	CPU	1	30000	presented
1	714466	Trantow-Barrows	Craig Booker	Debra Henley	Software	1	10000	presented
2	714466	Trantow-Barrows	Craig Booker	Debra Henley	Maintenance	2	5000	pending
3	737550	Fritsch, Russel and Anderson	Craig Booker	Debra Henley	CPU	1	35000	declined
4	146832	Kiehn-Spinka	Daniel Hilton	Debra Henley	CPU	2	65000	won

		sum				
		Price				
	Product	CPU	Maintenance	Monitor	Software	All
Manager	Status					
Debra Henley	declined	70000	0	0	0	70000
	pending	40000	10000	0	0	50000
	presented	30000	0	0	20000	50000
	won	65000	0	0	0	65000
Fred Anderson	declined	65000	0	0	0	65000
	pending	0	5000	0	0	5000
	presented	30000	0	5000	10000	45000
	won	165000	7000	0	0	172000
All		465000	22000	5000	30000	522000

📖 To Do before Lecture 3

Run IPython examples provided in class

- **Use Python on cloud via Data Scientist Workbench**
  - Register for CognitiveClass.ai MOOC portal https://cognitiveclass.ai to access 60+ free data science courses and to use Python on the DSW cloud
  - You can use Python on DSW cloud via https://datascientistworkbench.com
- **Install Python**
  - Recommended to use Python version 3.X
  - You may use your own Python distribution, e.g., Anaconda that can be downloaded from https://www.anaconda.com/download/
- **Form groups of six students for in-class presentations and course project**
  - Add all your group members to Group X on Q
  - All groups should have six members
  - In-class presentations will be done in the order of group numbers
  - Course project will be the same for all groups
  - Every group member get the same mark, independently on how you split responsibilities inside each group
- **Check class web-page on Q regularly**