

Python for Data Analysis, 3rd edition

Data Wrangling with pandas, NumPy, and Jupyter

Stephen CUI¹

January 4, 2022

¹cuixuanStephen@gmail.com

Contents

1	数据聚合与分组运算	3
1.1	Data Aggregation	3
1.2	apply: 一般性的“拆分-应用-合并”	3
2	Appendix A	5

Chapter 1

数据聚合与分组运算

1.1 Data Aggregation

1.2 apply：一般性的“拆分-应用-合并”

Table 1.1: Optimized groupby methods

Function name	Description
any, all	Return True if any (one or more values) or all non-NA values are “truthy”
count	Number of non-NA values
cummin, cummax	Cumulative minimum and maximum of non-NA values
cumsum	Cumulative sum of non-NA values
cumprod	Cumulative product of non-NA values
first, last	First and last non-NA values
mean	Mean of non-NA values
median	Arithmetic median of non-NA values
min, max	Minimum and maximum of non-NA values
nth	Retrieve value that would appear at position n with the data in sorted order
ohlc	Compute four “open-high-low-close” statistics for time series-like data
prod	Product of non-NA values
quantile	Compute sample quantile
rank	Ordinal ranks of non-NA values, like calling Series.rank
size	Compute group sizes, returning result as a Series
sum	Sum of non-NA values
std, var	Sample standard deviation and variance

Chapter 2

Appendix A