Pandas Tutorial

Quach Huy Tung

November 6, 2020

Table of contents

Pandas
Tutorial

Quach Huy
Tung

Hands on
examples
Introduction to
Pandas
Pandas objects
Indexing and
selection
Data operations
Handling missing
data
Groupby &
Aggregation
Merge, join,
concatenate
Vectorized string
operations

Exercises
SF Salaries
Ecommerce
Purchases

Hmmm, Pandas ?

1 **What it is?**:
- Data analysis
- Data manipulation
- Fast, flexible, easy to use, built on top of Python

2 **What is it for?**:
- Calculate statistics and answer questions about the data
- Clean the data
  - Removing missing values
  - Filtering rows or columns by some criteria
  - Visualize the data
  - Store the cleaned, transformed data back into a CSV, other file or database

Series

1 Series object: values, index

2 Series object construction
  - From array
  - From dictionary

3 Some methods of Series object:
  - Agregation: sum(), min(), max(), mean(), std(), ...
  - describe(), ...

Dataframe

1. Dataframe object: index, columns
2. Dataframe contruction
3. Some methods on Dataframe object:
   - describe(), info(), ...
   - Agregration: sum(), min(), max(), mean(), std(), ...
   - Drop columns
   - Rename columns

1. Series as dictionary
2. Series as one-dimensional array
   - Slicing by explicit index
   - Slicing by implicit integer index
   - Masking
   - Fancy indexing
3. loc[], iloc[]

Indexing & selection in DataFrame

1. DataFrame as dictionary
2. DataFrame as two-dimensional array

Data operations

Pandas
Tutorial

Quach Huy
Tung

Hands on
examples
Introduction to
Pandas
Pandas objects
Indexing and
selection
Data operations
Handling missing
data
Groupby &
Aggregation
Merge, join,
concatenate
Vectorized string
operations
Exercises
SF Salaries
Ecommerce
Purchases

1. Element-wise operations: applymap()
2. Row(column)-wise operations: apply() with axis

Handling missing data

Pandas
Tutorial

Quach Huy
Tung

Hands on
examples

Introduction to
Pandas

Pandas objects

Indexing and
selection

Data operations

Handling missing
data

Groupby &
Aggregation

Merge, join,
concatenate

Vectorized string
operations

Exercises

SF Salaries

Ecommerce
Purchases

1. Check missing data: isnull(), notnull()
2. Drop missing values:
   - Drop by row
   - Drop by column
   - Some advanced options
3. Fill missing values:
   - Fill by values
   - Fill by methods: ffill, bfill
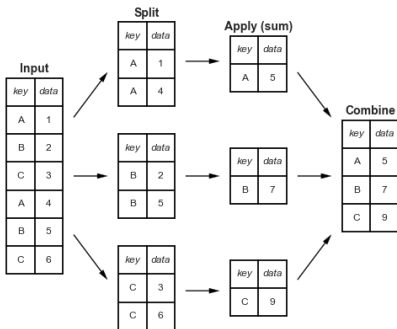
Groupby demystified

Pandas
Tutorial

Quach Huy
Tung

Hands on
examples

Introduction to
Pandas

Pandas objects

Indexing and
selection

Data operations

Handling missing
data

**Groupby &
Aggregation**

Merge, join,
concatenate

Vectorized string
operations

Exercises

SF Salaries

Ecommerce
Purchases

Aggregation - Filter - Transform

Pandas
Tutorial

Quach Huy
Tung

Hands on
examples
Introduction to
Pandas
Pandas objects
Indexing and
selection
Data operations
Handling missing
data
Groupby &
Aggregation
Merge, join,
concatenate
Vectorized string
operations

Exercises
SF Salaries
Ecommerce
Purchases

1 **Aggregation**:
- count()
- first(), last()
- mean(), median(), std(), var()
- min(), max()

2 **Filter**

3 **Transform**

1 Concatenate:
  - Concatenate by axis
  - Some advanced options: join, ignore_index

Merge - Join

1. Relational Algebra: Descartes product
2. Join:
   - Inner
   - Outer
   - Left, Right
   - Some advanced options: on, left_on, right_on, suffixes

Vectorized string operations

Some methods on vectorized string operations:

- Length: .str.len()
- Slicing: .str[:]
- Regex

Questions - Part 1

- Import pandas
- Read Salaries.csv as a dataframe called sal.
- Check the head of the DataFrame.
- Use the .info() method to find out how many entries there are.
- What is the average BasePay ?
- What is the highest amount of OvertimePay in the dataset ?
- What is the job title of JOSEPH DRISCOLL ?
- How much does JOSEPH DRISCOLL make (including benefits)?
- What is the name of highest paid person (including benefits)?

- What is the name of lowest paid person (including benefits)? Do you notice something strange about how much he or she is paid?
- What was the average (mean) BasePay of all employees per year? (2011-2014) ?
- How many unique job titles are there?
- What are the top 5 most common jobs?
- How many Job Titles were represented by only one person in 2013? (e.g. Job Titles with only one occurence in 2013?)
- How many people have the word Chief in their job title?
- Is there a correlation between length of the Job Title string and Salary?

Questions - Part 1

- Load data from csv and set it to Pandas DF ecom
- Check the head, the tail of ecom DF
- How many rows and columns are there?
- What is the average Purchase Price?
- What were the highest and lowest purchase prices?
- How many people have English 'en' as their Language of choice on the website?
- How many people have the job title of "Lawyer" ?
- How many people made the purchase during the AM and how many people made the purchase during PM ?

- What are the 5 most common Job Titles?
- Someone made a purchase that came from Lot: "90 WT" , what was the Purchase Price for this transaction?
- What is the email of the person with the following Credit Card Number: 4926535242672853
- How many people have American Express as their Credit Card Provider and made a purchase above $95 ?
- **How many people have a credit card that expires in 2025?**
- **What are the top 5 most popular email providers/hosts (e.g. gmail.com, yahoo.com, etc...)?**