

Python for Economics

Zhenhua He
Richard Lawrence

Morning session, 10/1/2021

Table of Contents

This course is divided into numbered lessons

17. Web Scraping with BeautifulSoup

- Advanced exercise: Candlestick plot
- Wrap-up

Lesson 17

Web Scraping with BeautifulSoup

Use Python libraries to extract data from a website

Applying Lesson Learning Objectives

After this lesson, you will know:

- What are regular expressions
- What is BeautifulSoup library and how to use it

During this lesson, you will use:

- What is web scraping
- HTML Basics
- What is requests library and how to use it

What is web scraping

- **Scraping:** a process of extracting (from various means), copying and screening of data.
- **Web Scraping:** scraping or extracting data or feeds from the web (like from web-pages or websites)

Why does one scrape the web?

- Collect data for Research
- Sales and Marketing
- Products prices & popularity comparison
- ...

Regular Expression

- A sequence of characters forms a search **pattern**
- `findall()`: Check if a string contains the pattern
- Python built-in module **re**
- Example

```
import re
text = "I went to TAMU at 8 AM"
pattern = "AM"
result = re.findall(pattern, text)
["AM", "AM"]
```

Regular Expression

re.compile()

- convert a pattern string into a pattern **object**
- **Object** combines the pattern and `findall()` function

- Example

```
text = "I went to TAMU at 8 AM"  
pattern = re.compile("AM")  
result = pattern.findall(text)
```

- No need to enter the pattern string again

Regular Expression Examples and Exercises

Go to Google Classroom Lesson 17 “Regular Expressions”

Tasks

- Follow instructions for the examples
- Work on the exercises (**due** by 10/1 6:00 PM)

BeautifulSoup

- A Python library for pulling data out of HTML files.
- Works with a *parser* to provide idiomatic ways of navigating, searching, and modifying the source tree.
- Built-in parser available (we will use)

Review: JSON - JavaScript Object Notation

- A text format for storing data
- language-independent
- The data to extract could be in JSON format.

JSON string examples:

```
'{"name": "Jack", "age": 20, "major": "computer science"}'
```

```
'{"args": {}, "data": "", "files": {},  
  "form": { "soup": "hot soup"}, ... }'
```

from the Requests exercise
we did previously

JSON module

- Python built-in module **json**
- `json.loads()` : converts JSON string to Python dictionary
- Example

```
import json
text = "{ keys : values, ... }"
dict = json.loads(text)
```

Today's Exercise: Scraping a website

- Collect salary stats from <https://www.salary.com/>
- Salary stats for a single city
- Salary stats for multiple cities
- Save data in a file

Investigate salary stats

The screenshot shows the salary.com website in a browser window. The browser's address bar contains 'salary.com' and is highlighted with a red box. The website's navigation bar is blue and features the 'salary.com' logo on the left and several menu items: 'FOR YOU', 'FOR EMPLOYERS', 'SURVEY CATALOG', 'RESOURCES', and 'FIND A JOB'. The 'FOR YOU' menu item is highlighted with a red box. Below the navigation bar, the page is divided into several sections. On the left, there is a 'Know Your Worth' section with links for 'Salaries', 'Skills', 'Companies', 'Jobs', 'Benefits', 'Industries', and 'Locations'. Below this is a search input field with the placeholder text 'Enter a job title or keyword', which is also highlighted with a red box. To the right of the search field are three columns of links: 'SALARIES' (with a dropdown arrow), 'JOB SEARCH' (with a dropdown arrow), and 'CAREER' (with a dropdown arrow). The 'SALARIES' column includes links for 'Salary Wizard', 'Benefits', 'Cost of Living', 'Minimum Wage', 'Gender Gap', and 'How to Get a Raise'. The 'JOB SEARCH' column includes links for 'Find a Job', 'Research Companies', 'Create Job Alerts', 'Post Resume', 'Create an Account', 'How to Find a Job', and 'Job Seeker Advice'. The 'CAREER' column includes links for 'Skills', 'Career Development', 'Negotiation', 'Future of Work', 'Learning Center', and 'All Articles'. On the far right, there is a 'BLOG' section with a 'DEMO' button and a search icon. Below the blog section is a featured article titled 'Eight Factors That Can Affect Your Pay' with a sub-headline 'Search Salary.com now'.

Analyze the salary stats

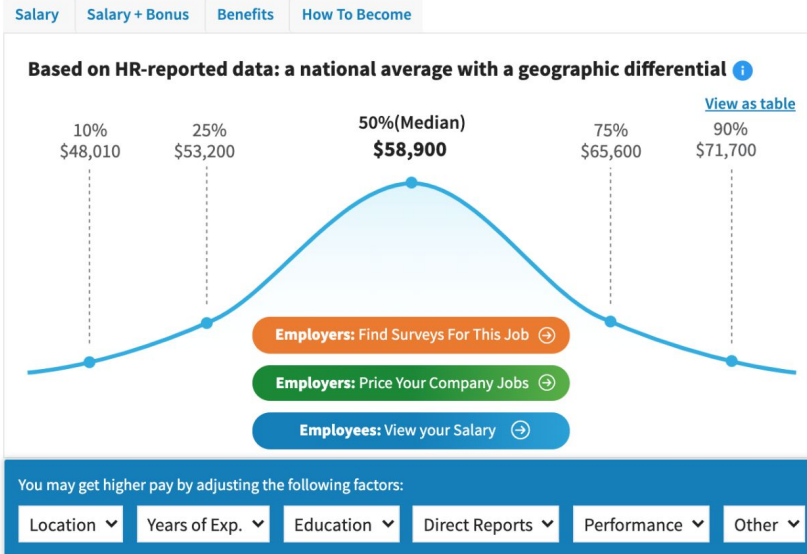
Tax Accountant - Entry Salary in the United States

Search More Jobs

Change City

Change Country

How much does a Tax Accountant - Entry make in the United States? The average Tax Accountant - Entry salary in the United States is **\$58,900** as of August 27, 2021, but the range typically falls between **\$53,200** and **\$65,600**. Salary ranges can vary widely depending on many important factors, including [education](#), certifications, additional skills, [the number of years](#) you have spent in your profession. With more online, real-time compensation data than any other website, Salary.com helps you determine your exact pay target.



Check out Tax Accountant - Entry jobs in Texas

[Entry Level Certified Public Accountant](#)

Established CPA firm - Arlington, TX

Apply Now

upper level job

[Staff Accountant - Remote](#)

TGG Accounting - Austin, TX

Apply Now

[Staff Accountant](#)

Motor Controls Inc. - Dallas, TX

Apply Now

[Staff Accountant Entry Level](#)

Sentoria Staffing Solutions - Houston, TX

Collect the salary stats for Dallas, TX

Tax Accountant - Entry Salary in Dallas, Texas

Search More Jobs

Change City

Change Country

How much does a Tax Accountant - Entry make in Dallas, TX? The average Tax Accountant - Entry salary in Dallas, TX is **\$58,500** as of August 27, 2021, but the range typically falls between **\$52,800** and **\$65,200**. Salary ranges can vary widely depending on many important factors, including education, certifications, additional skills, the number of years you have spent in your profession. With more online, real-time compensation data than any other website, Salary.com helps you determine your exact pay target.

Salary Salary + Bonus Benefits How To Become

Based on HR-reported data: a national average with a geographic differential



You may get higher pay by adjusting the following factors:

Location Years of Exp. Education Direct Reports Performance Other

Check out Tax Accountant - Entry jobs in Dallas, Texas

Tax Accountant - Corporate, Entry/Staff level, Remote

The Bolton Group - Dallas, TX

Apply Now

Staff Accountant

Motor Controls Inc. - Dallas, TX

Apply Now

Staff Accountant - Remote

TGG Accounting - Dallas, TX

Apply Now

Tax Director - REMOTE OPTIONS for local candidate

Kutchins, Robbins & Diamond, Ltd. - Dallas, TX

Salary Salary + Bonus Benefits How To Become

Based on HR-reported data: a national average with a geographic differential

View as graph

Tax Accountant - Entry Salaries by Percentile

PERCENTILE	SALARY	LOCATION	LAST UPDATED
10th Percentile Tax Accountant - Entry Salary	\$47,610	Dallas, TX	August 27, 2021
25th Percentile Tax Accountant - Entry Salary	\$52,800	Dallas, TX	August 27, 2021
50th Percentile Tax Accountant - Entry Salary	\$58,500	Dallas, TX	August 27, 2021
75th Percentile Tax Accountant - Entry Salary	\$65,200	Dallas, TX	August 27, 2021
90th Percentile Tax Accountant - Entry Salary	\$71,300	Dallas, TX	August 27, 2021

Check the page source

Tax Accountant - Entry Salary in Dallas, Texas

How much does a Tax Accountant - Entry make in Dallas, TX? The average Tax Accountant - Entry salary in Dallas, falls between **\$52,800** and **\$65,200**. Salary ranges can vary widely depending on many important factors, including ϵ you have spent in your profession. With more online, real-time compensation data than any other website, Salary.co

Salary	Salary + Bonus	Benefits	How
Based on HR-reported data: a national			
Tax Accountant - Entry Salary in Dallas, TX View as graph			
PERCENTILE			LAST UPDATED
10th Percentile Tax Accountant - Entry S			TX August 27, 2021
25th Percentile Tax Accountant - Entry S			TX August 27, 2021
50th Percentile Tax Accountant - Entry S			TX August 27, 2021
75th Percentile Tax Accountant - Entry S			TX August 27, 2021
90th Percentile Tax Accountant - Entry S			TX August 27, 2021

- Back
- Forward
- Reload
- Save As...
- Print...
- Cast...
- Create QR Code for this Page
- Translate to English
- EditThisCookie
- View Page Source**
- Inspect

```
283 <!-- Occupation -->
284
285 <script type="application/ld+json">
286 {
287   "@context": "http://schema.org",
288   "@type": "Occupation",
289   "name": "Tax Accountant - Entry",
290   "mainEntityOfPage": {
291     "@type": "WebPage",
292     "lastReviewed": "2021-08-27T00:00:00Z"
293   },
294   "description": "The Tax Accountant - Entry helps with
295   "estimatedSalary": [{
296     "@type": "MonetaryAmountDistribution",
297     "name": "base",
298     "currency": "USD",
299     "duration": "P1Y",
300     "percentile10": "47610",
301     "percentile25": "52800",
302     "median": "58500",
303     "percentile75": "65200",
304     "percentile90": "71300"
305   }],
306   {
307     "@type": "MonetaryAmountDistribution",
308     "name": "total compensation",
309     "currency": "USD",
310     "duration": "P1Y",
311     "percentile10": "48473",
312     "percentile25": "54300",
313     "median": "60700",
314     "percentile75": "67800",
315     "percentile90": "74264"
316   }]
317 },
318   "occupationLocation": [{
319     "@type": "City",
320     "name": "Dallas, TX"
321   }]
322 }
323 </script>
```



Collect stats for multiple cities – structured URL

Senior accountant
Dallas, Texas

<https://www.salary.com/research/salary/alternate/senior-accountant-salary/dallas-tx>

Entry data analyst
Miami, Florida

<https://www.salary.com/research/salary/alternate/entry-data-analyst-salary/miami-fl>

Tax accountant entry
Washington, DC

<https://www.salary.com/research/salary/alternate/tax-accountant-entry-salary/washington-dc>



```
url_template = 'https://www.salary.com/research/salary/alternate/{}-salary/{}'
```

Goal: CSV file with collected data

salary_data_multi_city.csv ×



Title	Location	Description	1 to 5 of 5 entries					Filter
			10%	25%	50%	75%	90%	
Tax Accountant - Entry	New York, NY	The Tax Accountant - Entry helps with the preparation of local, state and federal level returns to be submitted within specified tax deadlines. Assists in the maintenance and preparation of tax-related items to include tax records, tax returns, tax schedules, and related tax reports. Being a Tax Accountant - Entry typically reports to a supervisor or a manager. May require a bachelor's degree. Being a Tax Accountant - Entry works on projects/matters of limited complexity in a support role. Work is closely managed. Working as a Tax Accountant - Entry typically requires 0-2 years of related experience.	57618	63900	70800	78900	86275	
Tax Accountant - Entry	Los Angeles, CA	The Tax Accountant - Entry helps with the preparation of local, state and federal level returns to be submitted within specified tax deadlines. Assists in the maintenance and preparation of tax-related items to include tax records, tax returns, tax schedules, and related tax reports. Being a Tax Accountant - Entry typically reports to a supervisor or a manager. May require a bachelor's degree. Being a Tax Accountant - Entry works on projects/matters of limited complexity in a support role. Work is closely managed. Working as a Tax Accountant - Entry typically requires 0-2 years of related experience.	54182	60100	66600	74200	81119	
Tax Accountant - Entry	Chicago, IL	The Tax Accountant - Entry helps with the preparation of local, state and federal level returns to be submitted within specified tax deadlines. Assists in the maintenance and preparation of tax-related items to include tax records, tax returns, tax schedules, and related tax reports. Being a Tax Accountant - Entry typically reports to a supervisor or a manager. May require a bachelor's degree. Being a Tax Accountant - Entry works on projects/matters of limited complexity in a support role. Work is closely managed. Working as a Tax Accountant - Entry typically requires 0-2 years of related experience.	50937	56400	62400	69600	76155	
Tax Accountant - Entry	Houston, TX	The Tax Accountant - Entry helps with the preparation of local, state and federal level returns to be submitted within specified tax deadlines. Assists in the maintenance and preparation of tax-related items to include tax records, tax returns, tax schedules, and related tax reports. Being a Tax Accountant - Entry typically reports to a supervisor or a manager. May require a bachelor's degree. Being a Tax Accountant - Entry works on projects/matters of limited complexity in a support role. Work is closely managed. Working as a Tax Accountant - Entry typically requires 0-2 years of related experience.	49219	54500	60300	67200	73482	
Tax Accountant - Entry	Phoenix, AZ	The Tax Accountant - Entry helps with the preparation of local, state and federal level returns to be submitted within specified tax deadlines. Assists in the maintenance and preparation of tax-related items to include tax records, tax returns, tax schedules, and related tax reports. Being a Tax Accountant - Entry typically reports to a supervisor or a manager. May require a bachelor's degree. Being a Tax Accountant - Entry works on projects/matters of limited complexity in a support role. Work is closely managed. Working as a Tax Accountant - Entry typically requires 0-2 years of related experience.	47401	52500	58100	64800	70900	

Show per page

Examples and Exercises

Go to Google Classroom assignment Lesson 17 “Beautiful Soup”

Tasks

- Follow instructions for the examples
- Work on the exercises (**due** by 10/1 6:00 PM)

Lesson 18

Candlestick Plot

Use Matplotlib finance module (mplfinance) to plot stock price movement

Candlestick plot



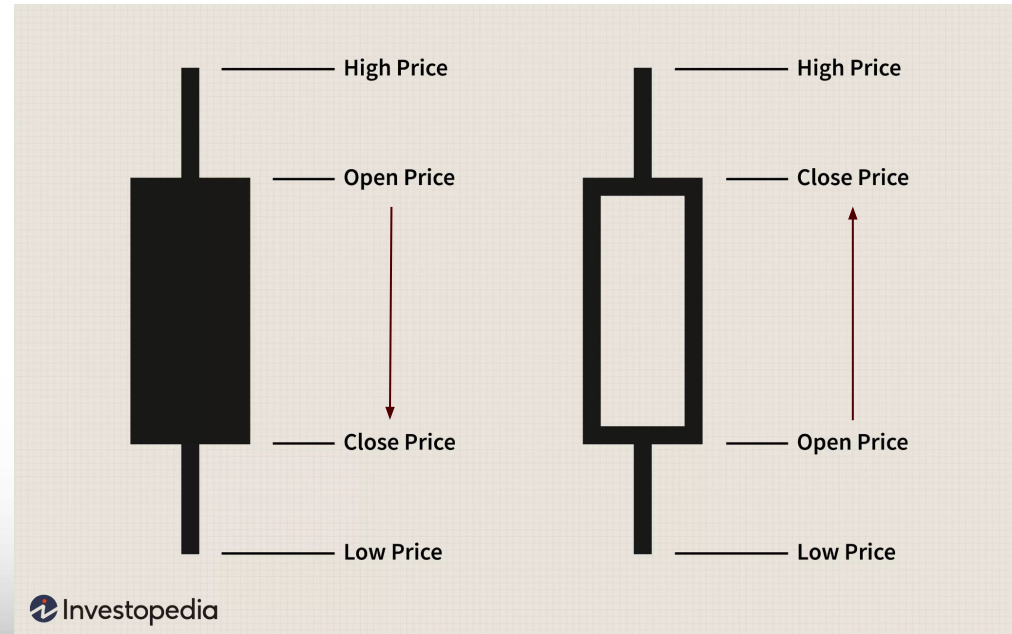
Candlestick Plot

The color of candlestick rectangle area depends on the relationship between the opening and closing price of a stock.

The color of the rectangle shows increase/decrease

Price Decrease

Price Increase



Candlestick Plot

Keyword arguments:

- type
- style
- volume (trade volume)
- mav (moving averages)

Examples and Exercises

Go to Google Classroom assignment Lesson 18 “Candlestick Plot”

Tasks

- Follow instructions for the examples
- Work on the exercises (**due** by 10/1 6:00 PM)

References

- https://www.tutorialspoint.com/beautiful_soup/beautiful_soup_quick_guide.htm
- <http://www.compjour.org/warmups/govt-text-releases/intro-to-bs4-lxml-parsing-wh-press-briefings/#converting-html-text-into-a-data-object>
- <https://www.w3schools.com/html/>
- <https://github.com/matplotlib/mplfinance>