

# Python for Economics

Zhenhua He  
Ridham Patoliya  
Richard Lawrence

Afternoon session, 9/24/2021

# Table of Contents

This course is divided into numbered lessons

- 16. Data Analysis with Scikit-learn
- 17. Web Scraping
- Wrap-up

# Lesson 16

## Data Analysis with Scikit-learn

Use Python Scikit-learn library to train a linear regression model

# Lesson Learning Objectives

Train a linear regression model to predict house prices

After this lesson, you will know how to:

- Import libraries
- Load the data
- Split the data into train and test dataset
- Train the model
- Make predictions

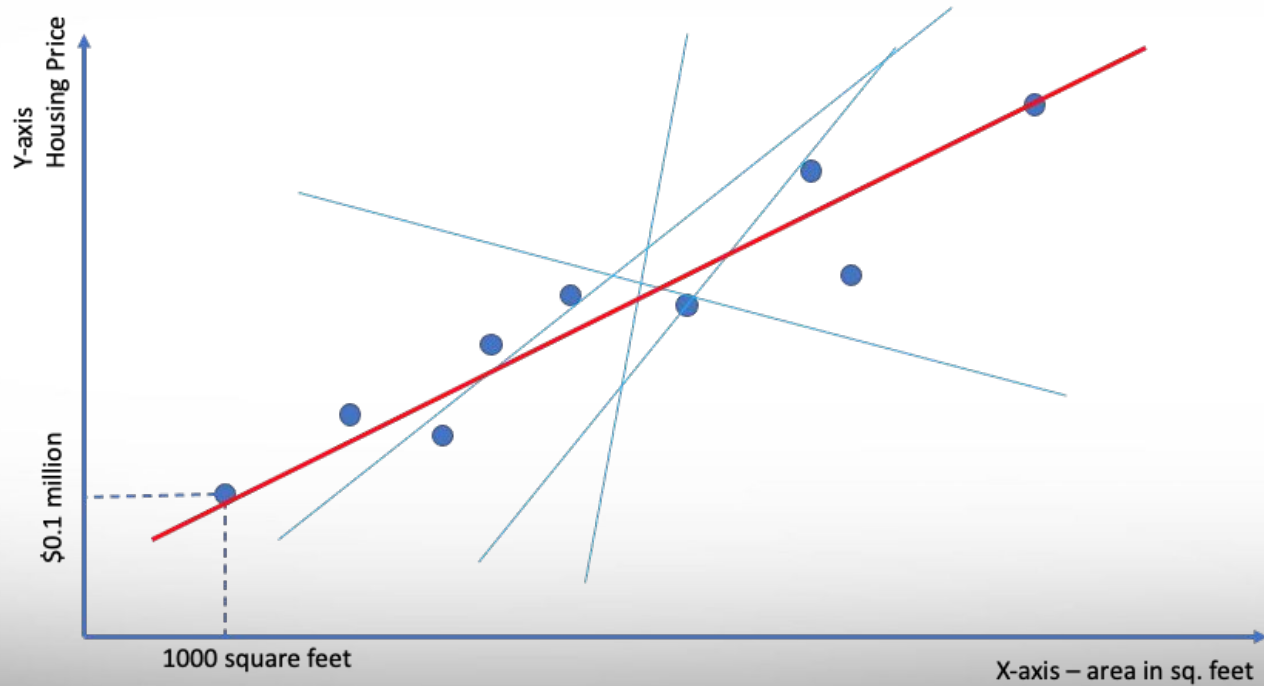
# What is linear regression-1

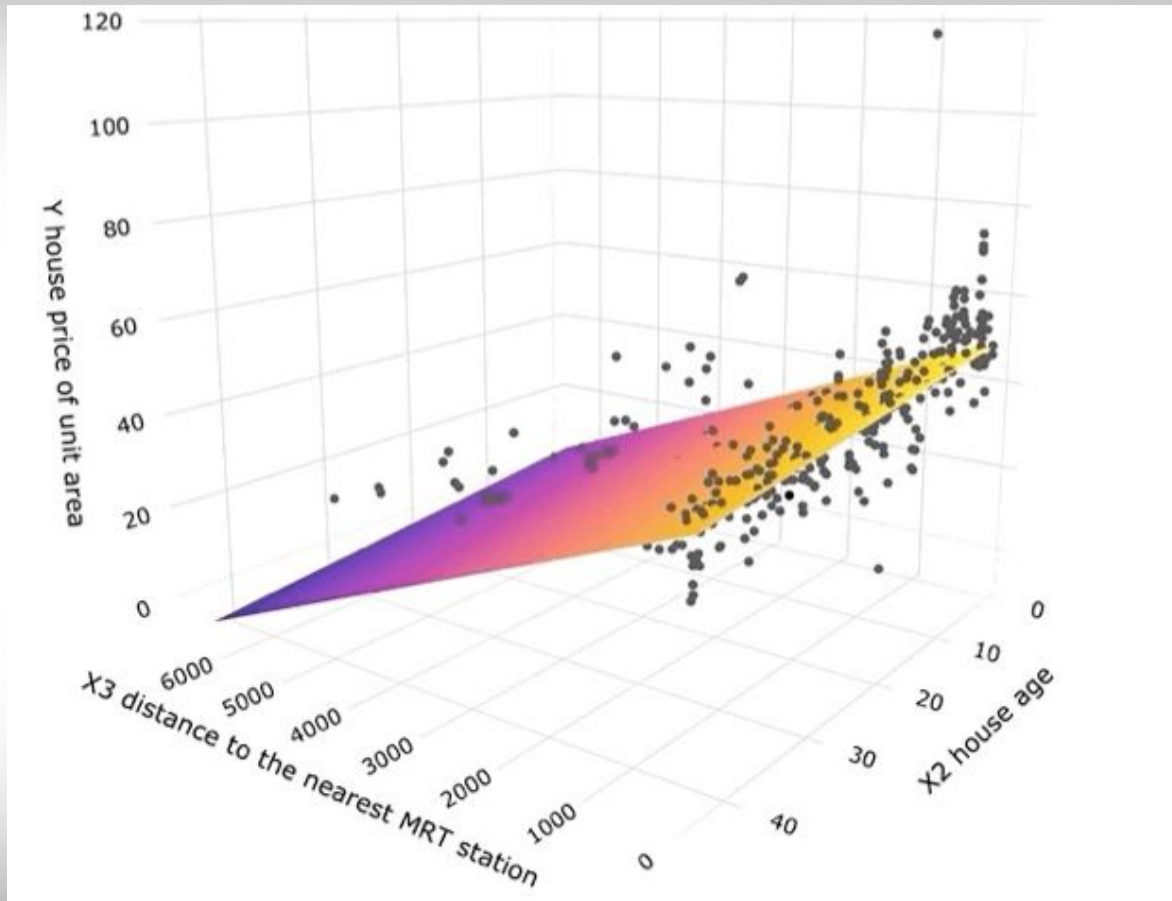
- **Linear regression** attempts to model the relationship between two variables by fitting a linear equation to observed data.
- One variable is called **explanatory variable**, and the other is called **dependent variable**.
- For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.

# What is linear regression-2

- In this exercise,  $x$  is the living area of a certain house, and  $y$  is its sale price. Linear regression aims to compute the slope and intercept of the line that minimizes error.
- This is a kind of machine learning, where a machine is learning the values for the slope and intercept given a set of training data. Once trained, the model can be used to predict the value of a house in King County given its living area.

# Linear regression







# Linear regression model

- **Import libraries**

```
# Import libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
```

- Use Pandas to **load the data** and view the first 5 rows

```
# Load data and view the first 5 rows
data = pd.read_excel("king_county_house_data.xlsx")

data.head(5)
```

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors
0	7129300520	20141013T000000	221900	3	1.00	1180	5650	1.0
1	6414100192	20141209T000000	538000	3	2.25	2570	7242	2.0
2	5631500400	20150225T000000	180000	2	1.00	770	10000	1.0
3	2487200875	20141209T000000	604000	4	3.00	1960	5000	1.0
4	1954400510	20150218T000000	510000	3	2.00	1680	8080	1.0

# Linear regression model

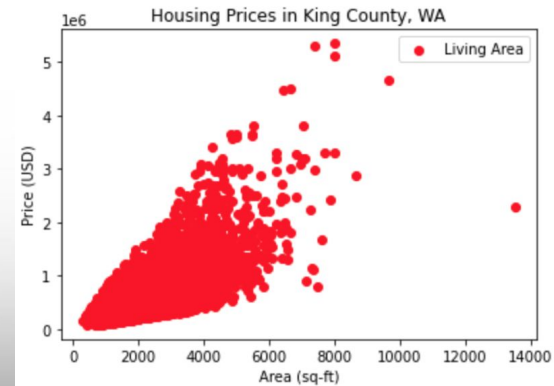
- Choose the columns from the data
- Split the data into train and test sets
- Visualize the train set

```
space = data['sqft_living']
price = data['price']

# Change X into 2D array
X = np.array(space).reshape(-1, 1)
Y = np.array(price)

# Split data into train sets and test sets
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=1/3,random_state=0)
```

```
# Visualize training set
plt.scatter(X_train,Y_train,color="red",label="Living Area")
plt.title("Housing Prices in King County, WA")
plt.xlabel("Area (sq-ft)")
plt.ylabel("Price (USD)")
plt.legend()
plt.show()
```



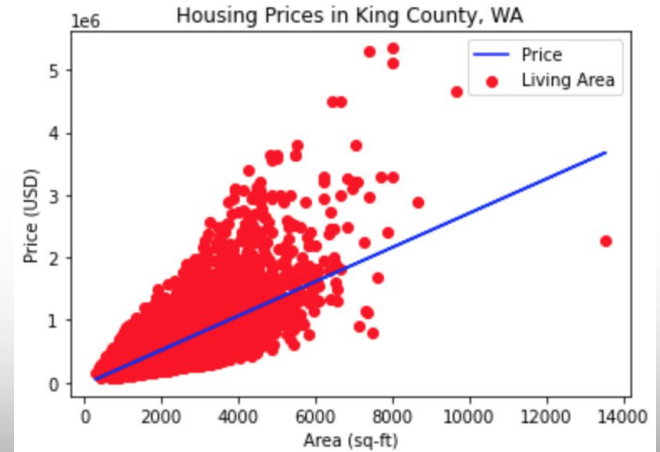
# Linear regression model

- Train the model with train set
- Predict on test set
  
- Visualize the train data and the best fit line

```
# Train
regressor = LinearRegression()
regressor.fit(X_train, Y_train)

# Prediction
y_pred = regressor.predict(X_test)
```

```
# Visualize the data and the bestfit line
plt.scatter(X_train, Y_train, color="red", label="Living Area")
plt.title("Housing Prices in King County, WA")
plt.plot(X_train, regressor.predict(X_train), color="blue", label="Price")
plt.xlabel("Area (sq-ft)")
plt.ylabel("Price (USD)")
plt.legend()
plt.show()
```



# Linear regression model

- Predict the price of a house with a certain area

```
area = 2000

price = regressor.predict([[area]])

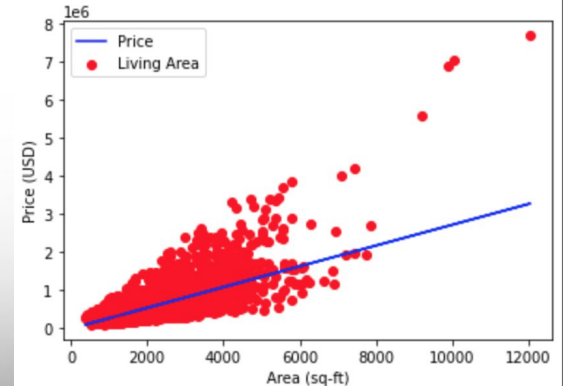
# print(price)

print('The price of a house of {d} sqft is about ${:0.0f}'.format(area, price[0]))

The price of a house of 2000 sqft is about $518564
```

- Visualize the test data

```
# Visualize test set
plt.scatter(X_test,Y_test,color='red',label="Living Area")
plt.plot(X_test,regressor.predict(X_test),color="blue",label="Price")
plt.xlabel("Area (sq-ft)")
plt.ylabel("Price (USD)")
plt.legend()
plt.show()
```



# Examples and Exercises

Go to Classroom assignment Lesson 16 “Linear Regression”

## Tasks

- Follow instructions for the examples
- Work on the exercises (**due** by 9/24 6:00 PM )

# Break Time Reminder Slide

10 minutes break



# Lesson 17

## Web Scraping

Use Python libraries to extract data from a website

# Lesson Learning Objectives

After this lesson, you will know how to:

- What is web scraping
- HTML Basics
- What is requests library and how to use it
- What is BeautifulSoup library and how to use it
- How to extract salary stats data from <https://www.salary.com/> and save it in a file



# 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 web scraping

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

# Why Python for web scraping

- Ease of Use
- Huge Library Support
- Huge Community
- ...

# Web scraping with Python

- HTML basics
- Browser “Inspect” / “View Page Source”
- Pandas read html
- Requests library for APIs
- BeautifulSoup for arbitrary Web Pages

# HTML

HTML is the standard markup language to create Web pages.

```
1 <!doctype html>
2 <html dir="ltr" lang="en">
3   <head>
4     <meta charset="utf-8">
5     <title>New Tab</title>
6     <style>
7       body {
8         background: #FFFFFF;
9         margin: 0;
10      }
11
12      #backgroundImage {
13        border: none;
14        height: 100%;
15        pointer-events: none;
16        position: fixed;
17        top: 0;
18        visibility: hidden;
19        width: 100%;
20      }
21
22      [show-background-image] #backgroundImage {
23        visibility: visible;
24      }
25    </style>
26  </head>
27  <body>
28    <iframe id="backgroundImage"
29      src="chrome-untrusted://new-tab-page/custom_background_image?url=">
30    </iframe>
31    <ntp-app></ntp-app>
32    <script type="module" src="new_tab_page.js"></script>
33    <link rel="stylesheet" href="chrome://resources/css/text_defaults_md.css">
34    <link rel="stylesheet" href="shared_vars.css">
35  </body>
36 </html>
37
```



🔍 Search Google or type a URL



# HTML elements

```
<!DOCTYPE html>  
<html></html>  
<head></head>  
<title></title>  
<body></body>  
<h1></h1>  
<p></p>
```

- HTML elements are *pairs* of **tags**.
  - `<!DOCTYPE>` is an exception
- The names match
- Tags live in angle brackets `<>`
- The first one is called “open” tag
- The second one is called “close” tag
  - It starts with a forward slash /

# Examples and Exercise

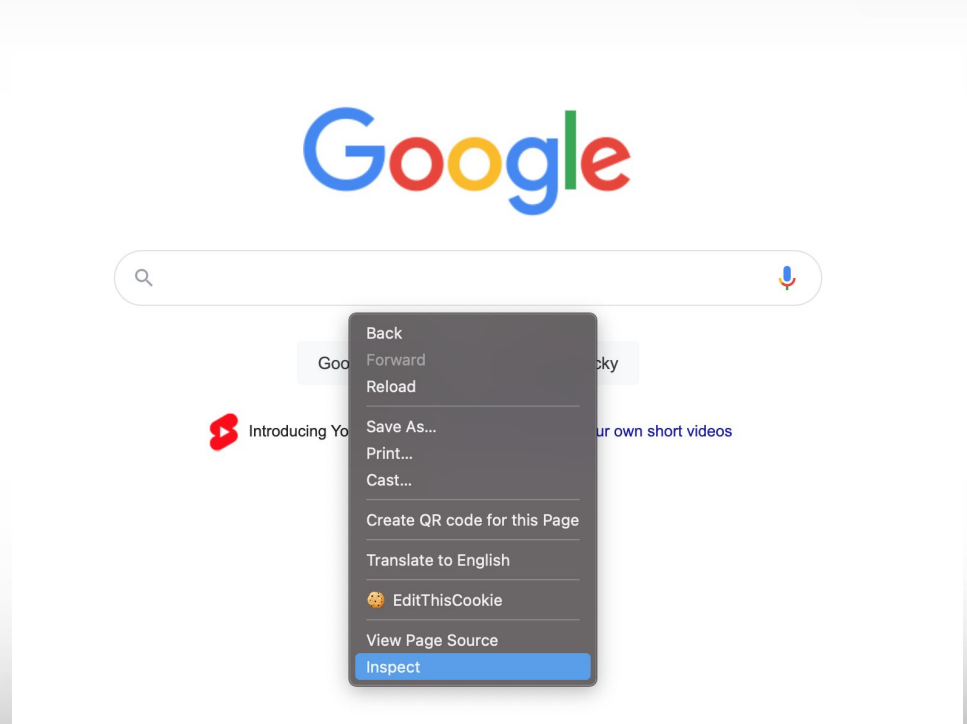
Go to Classroom assignment Lesson 17 “HTML”

Tasks:

- Use HTML in a Jupyter Notebook
- Practice with the `<table>` element.

# Browser “Inspect” elements

In Chrome: right click,  
choose “view page source”  
or “inspect” to see the code





# Exercise: Inspect a Website

Navigate to

- <https://www.federalreserve.gov/releases/h15/>

Find where the `<table>` tags are located in the code.

Find where the first data value lives in the table.

# Examples and Exercise

Go to Classroom assignment Lesson 17 “Pandas HTML”

Tasks:

- Fetch a table directly into a Pandas Dataframe
- Clean it up a bit

# Break Time Reminder Slide

10 minutes break



# Requests

**Requests** library for HTTP activities.

Replicate the experience of visiting a web page,  
but in a Notebook instead of a Browser.

# Examples and Exercise

Go to Classroom assignment Lesson 17 “Requests”

Tasks:

- Fetch a web page
- Interact with a website
- Practice: how to learn a simple API.

# Web Scraping with API

Reference: <https://fred.stlouisfed.org/docs/api/fred/>

## **Fred API**

retrieve economic data from the FRED® and ALFRED® websites hosted by the Economic Research Division of the Federal Reserve Bank of St. Louis

# Web Scraping with API

[Register](#) and log into your fredaccount.stlouisfed.org user account and request your API Key.

All web service requests require an API key to identify requests.

# Web Scraping with API

**Example:** Get an economic data series

[https://api.stlouisfed.org/fred/series?series\\_id=GNPCA&api\\_key=abcdefghijklmnpqrstuvwxyz123456&file\\_type=json&...](https://api.stlouisfed.org/fred/series?series_id=GNPCA&api_key=abcdefghijklmnpqrstuvwxyz123456&file_type=json&...)

## Parameters

- api\_key
- file\_type
- series\_id
- realtime\_start
- realtime\_end



# Examples and Exercises

Go to Google Classroom assignment “Web Scraping API”

## Tasks

- Follow instructions for the examples
- Work on the exercises (**due** by 9/24 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 parse tree.

# Scraping a website for salary stats

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

# Salary stats for a single city

The screenshot shows the salary.com website interface. At the top, the browser address bar displays 'salary.com' with a red box and arrow pointing to it. The website's navigation bar includes 'FOR YOU' (highlighted with a red box and arrow), 'FOR EMPLOYERS', 'SURVEY CATALOG', 'RESOURCES', and 'FIND A JOB'. On the left, there are sections for 'Know Your Worth' (with links for Salaries, Skills, Companies, Jobs, Benefits, Industries, Locations) and 'Get a Raise' (with a 'Get a Raise' button). A search bar with the placeholder 'Enter a job title or keyword' is highlighted with a red box and arrow. The main content area features three columns of links: 'SALARIES' (Salary Wizard, Benefits, Cost of Living, Minimum Wage, Gender Gap, How to Get a Raise), 'JOB SEARCH' (Find a Job, Research Companies, Create Job Alerts, Post Resume, Create an Account, How to Find a Job, Job Seeker Advice), and 'CAREER' (Skills, Career Development, Negotiation, Future of Work, Learning Center, All Articles). On the right, there is a featured article titled 'Eight Factors That Can Affect Your Pay' with a 'Search Salary.com now' link.

# Salary stats for a single city

## Senior Accountant Salary in the United States

Search More Jobs

Change City

Change Country

How much does an **Senior Accountant** make in the **United States**? The average Senior Accountant salary in the United States is **\$81,990** as of July 28, 2021, but the range typically falls between **\$74,190** and **\$90,890**. 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

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



lower level job

Employers: Find Surveys For This Job

Employers: Price Your Company Jobs

Employees: View your Salary

You may get higher pay by adjusting the following factors:

Location Years of Exp. Education Direct Reports Performance Other

Check out **Senior Accountant** jobs in **Dallas, Texas**

Senior Accountant

CyberCoders - Dallas, TX

Apply Now

Senior Accountant

Aduro Advisors - Dallas, TX

Apply Now

Senior Accountant

Undisclosed - Dallas, TX

Apply Now

upper level job

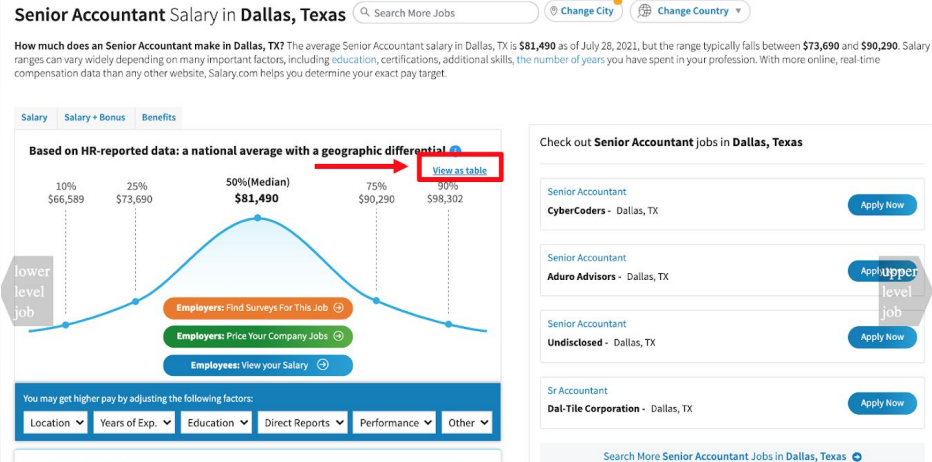
Sr Accountant

Dal-Tile Corporation - Dallas, TX

Apply Now

Search More Senior Accountant Jobs in Dallas, Texas

# Salary stats for a single city



Salary Salary + Bonus Benefits

**Based on HR-reported data: a national average with a geographic differential** [View as graph](#)

**Senior Accountant Salaries by Percentile**

PERCENTILE	SALARY	LOCATION	LAST UPDATED
10th Percentile Senior Accountant Salary	\$66,589	Dallas,TX	July 28, 2021
25th Percentile Senior Accountant Salary	\$73,690	Dallas,TX	July 28, 2021
50th Percentile Senior Accountant Salary	\$81,490	Dallas,TX	July 28, 2021
75th Percentile Senior Accountant Salary	\$90,290	Dallas,TX	July 28, 2021
90th Percentile Senior Accountant Salary	\$98,302	Dallas,TX	July 28, 2021

lower level job

You may get higher pay by adjusting the following factors:

Location  Years of Exp.  Education  Direct Reports  Performance  Other

# Salary stats for a single city

## Senior Accountant Salary in Dallas, Texas

Search More Jobs

Change City

Change Country

How much does an Senior Accountant make in Dallas, TX? The average Senior Accountant salary in Dallas, TX is **\$81,490** as of July 28, 2021, but the range typically falls between **\$73,690** and **\$90,290**. 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

Based on HR-reported data: a national

differential

View as graph

lower level job

PERCENTILE
10th Percentile Senior Accountant Salary
25th Percentile Senior Accountant Salary
50th Percentile Senior Accountant Salary
75th Percentile Senior Accountant Salary
90th Percentile Senior Accountant Salary

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

SALARY	LOCATION	LAST UPDATED
\$66,589	Dallas,TX	July 28, 2021
\$73,690	Dallas,TX	July 28, 2021
\$81,490	Dallas,TX	July 28, 2021
\$90,290	Dallas,TX	July 28, 2021
\$98,302	Dallas,TX	July 28, 2021

upper level job

```
293 <!-- Occupation -->
294
295 <script type="application/ld+json">
296 {
297   "@context": "http://schema.org",
298   "@type": "Occupation",
299   "name": "Senior Accountant",
300   "mainEntityOfPage": {
301     "@type": "WebPage",
302     "lastReviewed": "2021-07-28T00:00:00Z"
303   },
304   "description": "The Senior Accountant ensures the accuracy of entries
305   "estimatedSalary": [{
306     "@type": "MonetaryAmountDistribution",
307     "name": "base",
308     "currency": "USD",
309     "duration": "PIY",
310     "percentile10": "66589",
311     "percentile25": "73690",
312     "median": "81490",
313     "percentile75": "90290",
314     "percentile90": "98302"
315   }],
316   {
317     "@type": "MonetaryAmountDistribution",
318     "name": "total compensation",
319     "currency": "USD",
320     "duration": "PIY",
321     "percentile10": "67369",
322     "percentile25": "75290",
323     "median": "83990",
324     "percentile75": "93990",
325     "percentile90": "103094"
326   }
327
328   ,
329   "occupationLocation": [{
330     "@type": "City",
331     "name": "Dallas, TX"
332   }]
333
334 }
335 </script>
```



# Salary 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/{}'
```



# Salary stats for multiple cities

	A	B
1	New-York-NY	
2	Los-Angeles-CA	
3	Chicago-IL	
4	Houston-TX	
5	Phoenix-AZ	
6	Philadelphia-PA	
7	San-Antonio-TX	
8	San-Diego-CA	
9	Dallas-TX	
10	San-Jose-CA	

Title	Location	Description	1 to 10 of 10 entries <input type="button" value="Filter"/>				
			nTile10	nTile25	nTile50	nTile75	nTile90
Senior Accountant	New York, NY	The Senior Accountant ensures the accuracy of entries to ledger accounts and reconciles subsidiary ledger accounts to the general ledger. Maintains financial records and ensures that financial transactions are properly recorded. Being a Senior Accountant analyzes current costs, revenues, financial commitments, and obligations incurred to predict future revenues and expenses. Prepares complex balance sheets, profit and loss statements and other financial reports. In addition, Senior Accountant may supervise and guide lower-level accountants. Requires a bachelor's degree. May require CPA certification. Typically reports to a supervisor or manager. Being a Senior Accountant contributes to moderately complex aspects of a project. Work is generally independent and collaborative in nature. Working as a Senior Accountant typically requires 4 to 7 years of related experience.	80823	89290	98590	109290	119032
Senior Accountant	Los Angeles, CA	The Senior Accountant ensures the accuracy of entries to ledger accounts and reconciles subsidiary ledger accounts to the general ledger. Maintains financial records and ensures that financial transactions are properly recorded. Being a Senior Accountant analyzes current costs, revenues, financial commitments, and obligations incurred to predict future revenues and expenses. Prepares complex balance sheets, profit and loss statements and other financial reports. In addition, Senior Accountant may supervise and guide lower-level accountants. Requires a bachelor's degree. May require CPA certification. Typically reports to a supervisor or manager. Being a Senior Accountant contributes to moderately complex aspects of a project. Work is generally independent and collaborative in nature. Working as a Senior Accountant typically requires 4 to 7 years of related experience.	75878	83890	92690	102790	111986
Senior Accountant	Chicago, IL	The Senior Accountant ensures the accuracy of entries to ledger accounts and reconciles subsidiary ledger accounts to the general ledger. Maintains financial records and ensures that financial transactions are properly recorded. Being a Senior Accountant analyzes current costs, revenues, financial commitments, and obligations incurred to predict future revenues and expenses. Prepares complex balance sheets, profit and loss statements and other financial reports. In addition, Senior Accountant may supervise and guide lower-level accountants. Requires a	71133	78690	86990	96390	104948

# References

- [https://www.tutorialspoint.com/beautiful\\_soup/beautiful\\_soup\\_quick\\_guide.htm](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/>