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

Y-axis Housing Price

X-axis – area in sq. feet

$0.1$ million

1000 square feet
Linear regression model

• Import libraries

• Use Pandas to load the data and view the first 5 rows
Linear regression model

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

```python
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
Linear regression model

- Predict the price of a house with a certain area

```python
area = 2000
price = regressor.predict([[area]])
# print(price)
print('The price of a house of {0:d} sqft is about ${0:0.0f}'.format(area, price[0]))
The price of a house of 2000 sqft is about $518564
```

- Visualize the test data

```python
# 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
- ...

19
Web scraping with Python

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

HTML is the standard markup language to create Web pages.
HTML elements

- 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 \texttt{<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=abcdefghijklmnopqrstuvwxyz123456&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/](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
Salary stats for a single city

Senior Accountant Salary in the United States

How much does a 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.

Check out Senior Accountant jobs in Dallas, Texas

- CyberCoders - Dallas, TX
- Aduro Advisors - Dallas, TX
- Undisclosed - Dallas, TX
- Dal-Tile Corporation - Dallas, TX

Search More Senior Accountant Jobs in Dallas, Texas
Salary stats for a single city
Salary stats for a single city

```javascript
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "Occupation",
  "name": "Senior Accountant",
  "mainEntityOfPage": {
    "@type": "WebPage",
    "lastReviewed": "2021-07-28T00:00:00Z"
  },
  "description": "The Senior Accountant ensures the accuracy of entries estimatedSalary": [
  {
    "@type": "MonetaryAmountDistribution",
    "name": "base",
    "currency": "USD",
    "duration": "P1Y",
    "percentile10": "66589",
    "percentile25": "73690",
    "median": "81490",
    "percentile75": "90290",
    "percentile90": "98302"
  },
  {
    "@type": "MonetaryAmountDistribution",
    "name": "total compensation",
    "currency": "USD",
    "duration": "P1Y",
    "percentile10": "67369",
    "percentile25": "75290",
    "median": "83990",
    "percentile75": "93990",
    "percentile90": "103094"
  }
],

'occupationLocation': [{
  "@type": "City",
  "name": "Dallas, TX"
}]
}
</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

<table>
<thead>
<tr>
<th>Title</th>
<th>Location</th>
<th>Description</th>
<th>nTile10</th>
<th>nTile25</th>
<th>nTile50</th>
<th>nTile75</th>
<th>nTile90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior Accountant</td>
<td>New York, NY</td>
<td>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.</td>
<td>80823</td>
<td>89290</td>
<td>98590</td>
<td>109290</td>
<td>119032</td>
</tr>
<tr>
<td>Senior Accountant</td>
<td>Los Angeles, CA</td>
<td>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.</td>
<td>75878</td>
<td>83890</td>
<td>92690</td>
<td>102790</td>
<td>111986</td>
</tr>
<tr>
<td>Senior Accountant</td>
<td>Chicago, IL</td>
<td>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</td>
<td>71133</td>
<td>78690</td>
<td>86990</td>
<td>96390</td>
<td>104948</td>
</tr>
</tbody>
</table>
References

- https://www.w3schools.com/html/