

# Python 0 到 1 基礎商業分析實戰

**TMR**

Html爬蟲

# 爬蟲概念

```

<!doctype html>
<html lang="zh-TW" data-react-helmet="lang">
  <head>...</head>
  <body>
    <div id="root" courselistrandperiod="1800" courselistrandwindow="96" recsectionrandperiod="1800" recsectionrandwindow="32" homepageandperiod="1800" homepageandwindow="32">
      <div class="sc-jqCOKK jYEAXq">
        <div class=" flex justify-center sc-kvZOFW bKHYXY" id="fixed-bottom-warning"></div>
        <header class="lecture-editor-no-clickoutside">
          <div class="header-down sc-Rmtcm bZpEyQ">...</div>
        </header>
        <main id="main-screen" class="content marg-t-50 padding-for-dismissable-fixed-top-message" style="min-height: calc(100vh - 100px);">
          <div class="sc-eTuwsz dRaaOU">...</div>
        </main>
        <footer class="sc-gipzik VHQd">
          <div class="container footer-container">...</div>
        </footer>
      </div>
    </div>
    <script src="//fast.wistia.com/assets/external/E-v1.js" defer="defer"></script>
    <script> == $0
      [Element.prototype,Document.prototype,DocumentFragment.prototype].forEach(function(e){e.hasOwnProperty("prepend")||Object.defineProperty(e,"prepend",{configurable:!0,enumerable:
      e=Array.prototype.slice.call(arguments),n=document.createDocumentFragment();e.forEach(function(e){var t=e instanceof Node;n.appendChild(t);e=document.createTextNode(String(e))
      loadDeferredStyles=function(){var e=document.getElementById("deferred-
      styles"),t=document.createElement("div");t.innerHTML=e.textContent,document.head.prepend(t),e.parentElement.removeChild(e)},raf=window.requestAnimationFrame||window.mozReques
      estAnimationFrame;raf?raf(function(){window.setTimeout(loadDeferredStyles,0)}):window.addEventListener("load",loadDeferredStyles)
      </script>
    <script>...</script>
    <script>...</script>
    <script src="https://hahow.in/static/js/21.c5374323.chunk.js"></script>
    <script src="https://hahow.in/static/js/main.2647e423.chunk.js"></script>
    <iframe name="_hjRemoteVarsFrame" title="_hjRemoteVarsFrame" id="_hjRemoteVarsFrame" src="https://vars.hotjar.com/box-90f3a29...html" style="display: none !important; width: 1px
    pointer-events: none !important;">...</iframe>
    <div style="width:0px; height:0px; display:none; visibility:hidden;" id="batBeacon0.17876630599968468">...</div>
    <script type="text/javascript" id">...</script>
    <script type="text/javascript" id src="https://www.googletagmanager.com/gtag/js?id=AW-775545404"></script>
    <script type="text/javascript" id">...</script>
    <iframe id="intercom-frame" style="position: absolute !important; opacity: 0 !important; width: 1px !important; height: 1px !important; top: 0 !important; left: 0 !important; b
    !important;" aria-hidden="true" tabindex="-1">...</iframe>
    <div id="intercom-css-container">...</div>
    <div id="intercom-container" class="intercom-namespace">...</div>
    <style id="wistia_18_style" type="text/css" class="wistia_injected_style">...</style>
    <script type="text/javascript" id">...</script>
    <script type="text/javascript" id">...</script>
    <script type="text/javascript" id">...</script>
    <script type="text/javascript" id">...</script>
    <script type="text/javascript" id">...</script>
    <script type="text/javascript" id">...</script>
    <script type="text/javascript" id">...</script>
    <script type="text/javascript" id">...</script>
    <script type="text/javascript" id">...</script>
    <div id="react-beautiful-dnd-announcement-0" aria-live="assertive" role="log" aria-atomic="true" style="position: absolute; width: 1px; height: 1px; margin: -1px; border: 0px;
    clip-path: inset(100%);">...</div>
  </body>
</html>

```

# 標籤總攬

標籤	用途	範例
div	標籤	<div>裡面可以放標籤或文字</div>
a	連結	<a href="網址">
h1~6	文字	<h1>你想說的話</h1>
span	文字	<span>你想說的話</span>
p	文字	<p>你想說的話</p>
table	表格	<table><tr><td></td></tr></table>

# 內文總攬

內文	用途	範例
id	唯一值	<div id='only'>
class	宣告CSS	<div class='myclass'>
style	CSS樣式	<div style='font-size:20px;'>
href	連結	<a href='http://abc.com'>
width	寬度	<table width='2'>
border	框線	<table border='3'>