Data Analysis and Visualization of Taiwanese Job Market

Presented By
B11303091   Fausto Urrutia
B11303087   Patrick Sun
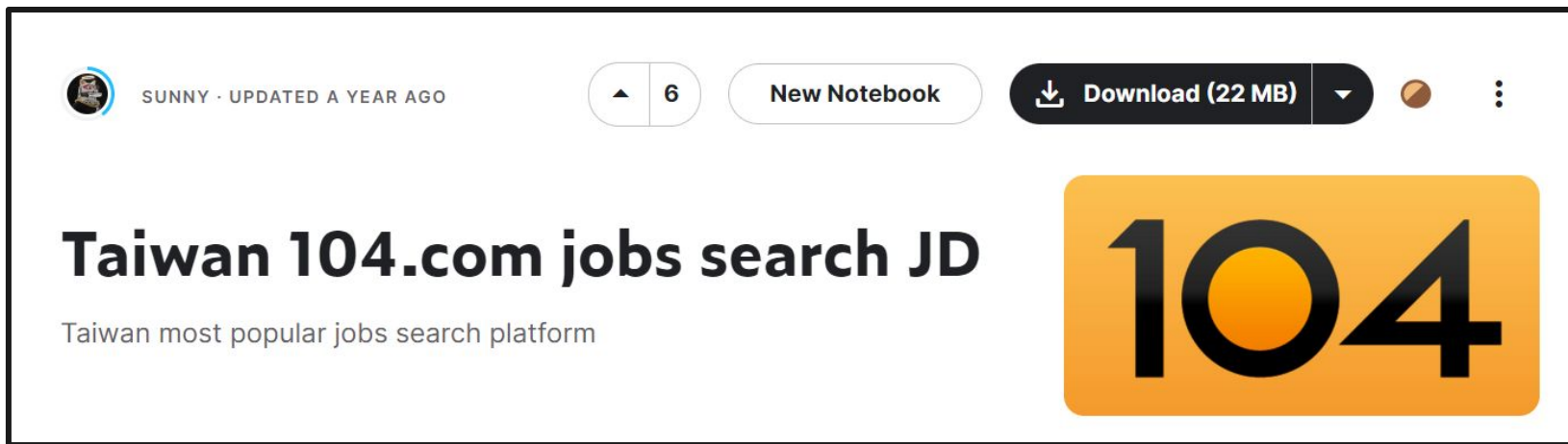B11303103   Hanaho Komatsu

May 28th, 2024

Project Background

- Difficulty of our market job searching in Taiwan
  - Limited resource in English
  - Different positions require different programming languages (like R and Python), although the job title (like "Data Analyst") is the same.

- Which programming language should I specialize in to have the best prospects when job hunting?

Project Background

- Database: Taiwan 104.com jobs search result (.csv) found in kaggle

Project Background

- The dataset consists of 26,000 job listings that focuses on software-related careers from 104 Website.
- 11 Columns have been extracted for this project, which includes:
  - **職位** (Position)
  - **縣市** (City/County)
  - **地區** (District/Area)
  - **公司名稱** (Company Name)
  - **工作內容** (Job Description)
  - **工作性質** (Nature of Work)
  - **需求人數** (Number of Positions)
  - **學歷要求** (Educational Requirements)
  - **擅長工具** (Tools Proficiency)
  - **員工人數** (Number of Employees)
  - **工作經歷** (Work Experience)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	職缺類別	職位類別	職位	縣市	地區	供需人數	公司名稱	職缺名稱	工作內容	職務類別	工作待遇	工作性質	上班地點	管理責任	上班時段	需求人數	工作經歷	學歷要求	科系要求	擅長工具	工作技能	其他條件	資本額	員工人數	公司標籤
2	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	國際厚生	系統分析i	1.系統功能需求分析設計。2.系統功能規劃與架構建佈署。3.具出色的溝通能力與組織協調能力,邏輯分析能力強,具備國際合作精神,且有專案經理經驗更佳。4.熟悉c#工具應用更佳。	軟體專案	月薪45,00	全職	台北市中正	管理4人以	日班, 9:0	2~3人	3年以上	大學以上	資訊工程	不拘	不拘	1.具軟體專案開發及軟體產品研發之系統分析及系統設計經驗,有UML設計經驗尤佳。2.具備軟體專案管理及需求導引之溝通技巧能力等經驗。3.熟悉各種資料庫整合開發技術,如MSSQL或Oracle	['1億5000	['75人']	['"]
3	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	三商電腦	專案經理	1、負責NI	軟體專案	待遇面議	全職	台北市中I	不需負擔I	日班	1人	不拘	大學、碩	不拘	Shell、UN	專案溝通i		['18億430	['768人']	['"]
4	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	威旭資訊	資深軟體	VICI Holdi	軟體設計	月薪100,0	全職	台北市中I	不需負擔I	日班	1人	5年以上	大學以上	資訊工程	Linux、C-	系統架構i		['2000萬元	['60人']	['"]
5	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	精誠資訊	AML反洗	1.熟悉反洗	軟體設計	月薪100,0	全職	台北市中I	管理4人以	日班	1人	10年以上	大學以上	資訊工程	不拘		1. 具金融	['26億700	['4000人']	['內湖科'
6	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	奔騰網路	SRE 專案	潤寨和量	軟體設計	待遇面議	全職	台北市中I	不需負擔I	日班, 9:0	1人	6年以上	大學以上	資訊工程	Mac OS、	不拘	你必須具備	['暫不提供	['暫不提供	['"]
7	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	財團法人t	資深軟體	【申請方!	軟體設計	月薪70,00	全職	台北市中I	不需負擔I	日班, 10:	1人	不拘	大學以上	不拘	Git、Pyth	軟體程式論	【必備條件	['暫不提供	['31人']	['"]
8	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	17Life_康	專案經理	1.產品導入	軟體相關	待遇面議	全職	台北市中I	不需負擔I	日班	1人	3年以上	大學以上	企業管理	Axure RP!	專案規劃	#NAME?	['暫不提供	['暫不提供	
9	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	香港商書	Software	At Kicksc	電子商務	月薪60,00	全職	台北市中I	不需負擔I	日班, 09:	1~2人	3年以上	大學以上	資訊工程	UNIX、Gi	系統整合	未填寫	['5億']	['100人']	['"]
10	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	香港商書	Data engi	At Kicksc	電子商務	月薪60,00	全職	台北市中I	不需負擔I	日班, 09:	1~2人	3年以上	大學以上	資訊工程	Linux、PF	系統整合	未填寫	['5億']	['100人']	['"]
11	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	香港商書	Software	At Kicksc	電子商務	月薪60,00	全職	台北市中I	不需負擔I	日班, 09:	1~2人	3年以上	大學以上	資訊工程	Linux、UN	系統整合	未填寫	['5億']	['100人']	['"]
12	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	疊代人才	Data Engi	身為 data	軟體專案	月薪60,00	全職	台北市中I	不需負擔I	日班	1人	3年以上	大學以上	不拘	Git、Pyth	軟體工程	未填寫	['2550萬元	['15人']	['"]
13	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	疊代人才	Front-End	身為 Front	軟體專案	月薪60,00	全職	台北市中I	不需負擔I	日班	不限	3年以上	大學以上	不拘	AWS、HT	專案溝通	未填寫	['2550萬元	['15人']	['"]
14	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	疊代人才	Back-End	身為 Back	軟體專案	月薪60,00	全職	台北市中I	不需負擔I	日班	不限	3年以上	大學以上	不拘	Git、C#、	軟體工程	未填寫	['2550萬元	['15人']	['"]
15	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	三商電腦	技術經理	1、負責!	軟體設計	待遇面議	全職	台北市中I	不需負擔I	日班	1人	不拘	大學、碩	不拘	Shell、UN	其他溝通	1. 具備保	['18億430	['768人']	['"]
16	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	南山產物	商業險系	1. 熟悉商	系統分析	待遇面議	全職	台北市中I	不需負擔I	日班	1人	3年以上	大學以上	資訊管理	不拘	不拘	1.具備保	['40億']	['740人']	['"]
17	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			7.5	威旭資訊	軟體工程	VICI Holdi	軟體設計	月薪75,00	全職	台北市中I	不需負擔I	日班	1人	不拘	大學以上	資訊工程	Linux、C-	系統架構	投遞請附	['2000萬元	['60人']	['"]
18	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	香港商書	Software	KICKS CF	軟體設計	月薪60,00	全職	台北市中I	不需負擔I	日班, 09:	1~2人	3年以上	大學以上	資訊工程	UNIX、Gi	系統整合		['5億']	['100人']	['"]
19	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	國際康健	系統分析	保單服務	系統分析	待遇面議	全職	台北市中I	不需負擔I	日班, 09:	1~2人	5年以上	大學以上	不拘	AJAX、AS	系統整合	必備條件:	['20億']	['1300人']	['外商']
20	資訊軟體及軟體_工程 軟體專案	臺北市	中正區			2.5	哈瑪星科	專案經理	1.有3年以	軟體專案	待遇面議	全職	台北市中I	管理人數以	日班, 09:	1人	3年以上	大學、碩	資訊工程	不拘	不拘	1.有帶領	['1億5000	['170人']	['"]

Project Goal

- Create a UI for Int'l students' job hunting to help them stay informed and up to the trend

- Show and analyze datasets to suggest which programming language to learn/focus, as well as determining companies that matches employee's masteries and qualifications

Methodology

Three Important Steps

1. **Translation of the Dataset**

2. **Preparation of the Dataset**

3. **Data Visualization**

Step1:
Translation of the dataset

Efficient data handling using .unique() to translate every single unique data in each column


```python
import pandas as pd
from googletrans import Translator, LANGUAGES

translator = Translator()
# Read the CSV file
df = pd.read_csv("career job description.csv")

# Extract the '職缺類別' column as a list of strings
county = df.地區.unique()
print(county)
type(county)
```

```
['中正區' '大同區' '中山區' '松山區' '大
'文山區' '金山區' '板橋區' '汐止區' '石
'林口區' '五股區' '八里區' '淡水區' '蘆
'龜山區' '八德區' '大溪區' '大園區' '蘆
'苗栗市' '中區' '南區' '西區' '北區' '
'大雅區' '梧棲區' '清水區' '大甲區' '外
'東區' '安平區' '安南區' '永康區' '歸仁
'前鎮區' '三民區' '楠梓區' '左營區' '仁
'三峽區' '蘆洲區' '五結鄉' '後龍鎮' '豐
'和美鎮' '草屯鎮' '麥寮鄉' '大林鎮' '安
'泰山區' '宜蘭市' '頭城鎮' '冬山鄉' '仁
```

This allows Python to only translate 108 unique data instead of the entire 26,000 data individually

Step1:
Translation of the dataset

The library used is the Google Translate API, but it does not work as required when translating columns with many unique names

```python
# Extract the '職缺類別' column as a list of strings
county = df.地區.unique()
print(county)
type(county)
for i in range(len(county)):
  translated = translator.translate(county[i], dest="en")
  print(translated.text)
```

```
Shilin District
Beitou District
Neihu District
Nangang District
Wenshan District
Jinshan District
Banqiao District
Xizhi District
Shiji District
```

Step2: Preparation of dataset

- Unify the format in the dataset e.g. education requirement

- Filter the company according to the job vacancy and availability for each language

- Create a bar chart to visualize the most relevant programming language used in the job market

```python
jobs_translated['Education Requirements'].value_counts()
```

```
Education Requirements
Above University                                                         14970
University and master's degree                                          11983
Specialty ABove                                                          9689
College, University, Master's Degree                                     7157
University                                                               2866
Informal                                                                 2746
Above a master's degree                                                  2713
College, university                                                      2705
ABove High School                                                        1311
master                                                                   1200
High school, college, university                                          475
High school, college, university, master's degere                         397
Specialty                                                                 284
PHD                                                                       100
Not specified                                                             75
Below high school, high school, college, university                       42
Below high school, high school, college, university, master's degree      40
high school                                                               26
High school, spiecilty                                                    26
High school, university, master's degree, doctral doctor                   8
Below high school, high school                                             8
College, Master's Degree, DOCTOR                                           3
University, doctor                                                         3
College, Master's Degree                                                   2
High school, university, master's degree                                   1
College, university, doctorate                                             1
High School, University                                                    1
Below high school                                                          1
Name: count, dtype: int64
```

BEFORE

12

```python
high = "Highschool, Technical school, Undergraduate, Master's, PhD"
tech = "Technical school, Undergraduate, Master's, PhD"
uni = "Undergraduate, Master's, PhD"
master = "Master's, PhD"
phd = "PhD"

education_replacement = {
    "Above University": uni,
    "University and master's degree": uni,
    "College, University, Master's Degree": tech,
    "Specialty ABove": tech,
    "College, university": tech,
    "Informal": 'Not specified',
    "Above a master's degree": master,
    "master": master,
    "ABove High School": high,
    "University": uni,
    "high school": high,
    "High school, college, university": high,
    "High school, college, university, master's degere": high,
    "Specialty": tech,
    "Not specified": 'Not specified',
    "PHD": phd,
    "College, Master's Degree": tech,
    "Below high school, high school, college, university": high,
    "Below high school, high school, college, university, master's degree": high,
    "College, Master's Degree, DOCTOR": tech,
    "High school, university, master's degree, doctral doctor": high,
    "High school, spiecilty":  high,
    "High school, university, master's degree": high,
    "Below high school, high school": high,
    "University, doctor": uni,
    "College, university, doctorate": tech,
    "High School, University": high,
    "Below high school": high,
}
```
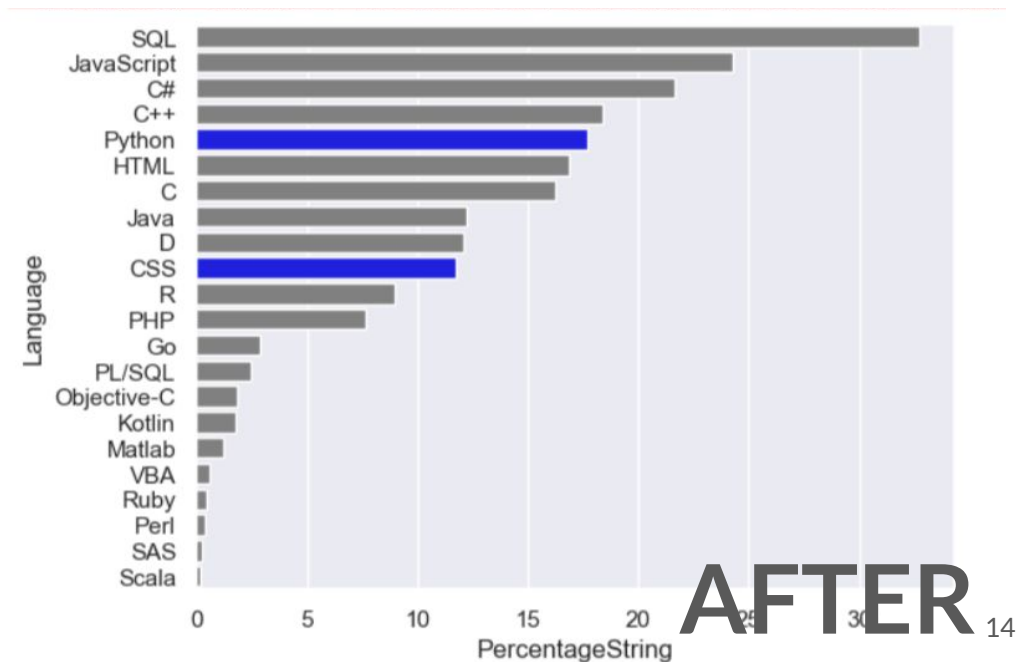
AFTER

13

In [30]: `jobs_translated["Proficiency in tools"].value_counts()`

Out[30]:
```
Proficiency in tools
不拘                                              21200
C、C++贊助提升專業能力                                     565
Excel、PowerPoint、Word贊助提升專業能力                     450
Python贊助提升專業能力                                    370
ASP.NET、C#贊助提升專業能力                                318
                                                 ...
MCU、C、C#、C++、FPGA贊助提升專業能力                          1
MCU、C、C++、Bluetooth、RF贊助提升專業能力                     1
MCU、C、C++、Java贊助提升專業能力                             1
EDA、FPGA、RTL、Verilog、VHDL贊助提升專業能力                  1
Linux、Python、TCP/IP、IPS贊助提升專業能力                    1
Name: count, Length: 9318, dtype: int64
```
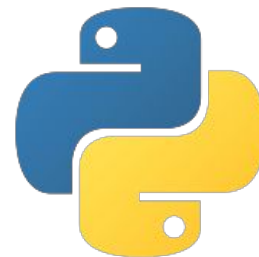
BEFORE

AFTER

14

STEP3: Data Visualization Using "Shiny for python"

- **Purpose: A framework designed to create interactive web applications for data visualization and analysis.**
- **Components:**
  - **UI (User Interface): Defines the layout and look of the application.**
  - **Server: Contains the logic to handle user input and update outputs.**
  - **App: Combines the UI and server into a single Shiny app object.**
- **Basic Structure:**
  - **Define the UI using layout functions.**
  - **Implement server logic with reactive functions that update outputs based on inputs.**
  - **Create the app by combining the UI and server**

Demonstration

https://faustourrutiareyes.github.io/104-jobs-dashboard/

Conclusion

Further Implementing 1

- Real Time Update of the dataset using Python Crawler → More reliable data analysis
  - Use "requests" or "selenium" to interact with website
  - Increase sources e.g. 1111.com, FB, linkedin and etc.
  - Improve data preparation

18

Further Implementing 2

- Smooth application from the app
  - Link the url of the job vacancy post to our app
  - Autofill of personal information
  - etc

Thank you!