

Job Vacancy over the years and University Graduates' Prospect Analysis

Presented by: Tan Yi Ting (P7400025)

TABLE OF CONTENTS

01 Dataset



Source, Data Overview &
Descriptive Statistic

03 Analysis



Visualisation of data and
insights

02 Process



Setting up Data, Data
Overview, Data Sorting
and Graph Ploting

04 SQL to Python



Setup, Process and
Result

01

DATASET

Source, Data Overview &
Descriptive Statistic

Dataset

Datasets were extracted from Data.gov.sg:

1. Job Vacancy by Industry and Occupational Group, Annual

Job Vacancy, Topline

Job Vacancy by Industry (Level 2)

https://data.gov.sg/dataset/job-vacancy-by-industry-and-occupational-group-annual?view_id=af19c76d-52af-4840-af44-c6c02c087ea3&resource_id=5d429272-3383-442e-b20b-5306daf55229

2. Per Capita GNI and per Capita GDP at Current Prices, Annual

https://data.gov.sg/dataset/per-capita-gni-and-per-capita-gdp-at-current-market-prices-annual?view_id=fbe9781f-7a5c-4372-a6ec-1c122edf4d65&resource_id=0ba721b8-d290-44f0-a3a6-748f5fd74dc5

3. Graduate Employment Survey – NTU, NUS, SIT, SMU, SUSS & SUTD

<https://data.gov.sg/dataset/graduate-employment-survey-ntu-nus-sit-smu-suss-sutd>

Overview

Job Vacancy Dataset

- **Columns:** Year, Job vacancy number (int64)
- **Shape:** (30,1)
- **Data are from the year of 1990 to 2019**
- **No missing values**
- **Descriptive Statistic:**
 - Mean: 38,356
 - Std: 38,356
 - Min: 12,600
 - 25%: 28,300
 - 50%: 40,350
 - 75%: 48,750
 - Max: 63,300

```
***Job Vacancy Dataset Overview***  
The shape of the data is (30, 1)  
The Job Vacancy data is from the year of 1990 to 2019  
The columns are: ['job_vacancy']  
There are no missing values.
```

```
***Summary of Dataset:***  
<class 'pandas.core.frame.DataFrame'  
Int64Index: 30 entries, 1990 to 2019  
Data columns (total 1 columns):  
#   Column      Non-Null Count  Dtype  
---  ---  
0   job_vacancy  30 non-null     int64  
dtypes: int64(1)  
memory usage: 480.0 bytes  
None
```

```
***Descriptive Statistic of Dataset:***  
      job_vacancy  
count      30.000000  
mean    38356.666667  
std     14607.123518  
min     12600.000000  
25%     28300.000000  
50%     40350.000000  
75%     48750.000000  
max     63300.000000
```

Overview

GDP Dataset

- **Columns:** type (str), value (int64)
- **Shape:** (59,2)
- **Data are from the year of 1960 to 2018**
- **No missing values**
- **Descriptive Statistic:**
 - Mean: 28,842
 - Std: 325,858
 - Min: 1,310
 - 25%: 5,912
 - 50%: 20,918
 - 75%: 44,004
 - Max: 87,108

```
***GDP Dataset Overview***  
The shape of the data is (59, 2)  
The GDP data is from the year of 1960 to 2018  
The columns are: ['type', 'value']  
There are no missing values.
```

```
***Summary of Dataset:***  
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 59 entries, 1960 to 2018  
Data columns (total 2 columns):  
#   Column  Non-Null Count  Dtype  ---  ---  ---  
0   type    59 non-null    object  
1   value   59 non-null    int64  
dtypes: int64(1), object(1)  
memory usage: 1.4+ KB  
None
```

```
***Descriptive Statistic of Dataset:***  
          value  
count    59.000000  
mean    28841.966102  
std     25858.062993  
min      1310.000000  
25%      5911.500000  
50%     20918.000000  
75%     44003.500000  
max      87108.000000
```

Overview

Job Vacancy by Industry Dataset

- **Columns:** year(int64), industry1(str), industry2(str), job_vacancy(int64)
- **Shape:** (584,4)
- **Data are from the year of 1990 to 2019**
- **7 Missing values in job_vacancy**
- **Descriptive Statistic of job_vacancy:**
 - Mean: 1994
 - Std: 2109
 - Min: 100
 - 25%: 500
 - 50%: 1,400
 - 75%: 2,700
 - Max: 13,700

```
***Job Vacancy by Industry Dataset Overview***
The shape of the data is (584, 4)
The dataset is from the year of 1990 to 2019
The columns are: ['year', 'industry1', 'industry2', 'job_vacancy']

***Number of non NA Values***
year          584
industry1     584
industry2     584
job_vacancy   577
dtype: int64

***Summary of Job Vacancy by Industry Dataset:***
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 584 entries, 0 to 583
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   year        584 non-null    int64
1   industry1   584 non-null    object
2   industry2   584 non-null    object
3   job_vacancy 577 non-null    float64
dtypes: float64(1), int64(1), object(2)
memory usage: 18.4+ KB
None

***Descriptive Statistic of Job Vacancy by Industry Dataset:***
              year      job_vacancy
count  584.000000    577.000000
mean   2004.075342    1993.760832
std      8.556210    2108.933355
min     1990.000000     100.000000
25%     1997.000000     500.000000
50%     2004.000000    1400.000000
75%     2011.000000    2700.000000
max     2019.000000   13700.000000
```

Overview

Local University Graduate Employment Dataset

- **Columns:** year(int64), university(str), employment_rate_overall(float64), gross_monthly_mean(int64),
- **Shape:** (703,12)
- **Data are from the year of 2013 to 2018**
- **There are missing values**
- **Descriptive Statistic of employment rate and gross monthly mean income:**

- Mean: 91	3,515
- Std: 7	522
- Min: 67	2,100
- 25%: 86	3,207
- 50%: 91	3,431
- 75%: 96	3,724
- Max: 100	5,617

```
***Local University Graduate Employment Dataset Overview***
The shape of the data is (703, 12)
The dataset is from the year of 2013 to 2018
The columns are: ['year', 'university', 'school', 'degree', 'employment_rate_overall',
'employment_rate_ft_perm', 'basic_monthly_mean', 'basic_monthly_median',
'gross_monthly_mean', 'gross_monthly_median', 'gross_mthly_25_percentile',
'gross_mthly_75_percentile']

***Summary of Local University Graduate Employment Dataset:***
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 703 entries, 0 to 702
Data columns (total 12 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   year                                703 non-null    int64
1   university                          703 non-null    object
2   school                             691 non-null    object
3   degree                             703 non-null    object
4   employment_rate_overall             630 non-null    float64
5   employment_rate_ft_perm             630 non-null    float64
6   basic_monthly_mean                  630 non-null    float64
7   basic_monthly_median                630 non-null    float64
8   gross_monthly_mean                  630 non-null    float64
9   gross_monthly_median                630 non-null    float64
10  gross_mthly_25_percentile            630 non-null    float64
11  gross_mthly_75_percentile            630 non-null    float64
dtypes: float64(8), int64(1), object(3)
memory usage: 66.0+ KB
None

***Descriptive Statistic of Local University Graduate Employment Dataset:***
      employment_rate_overall  gross_monthly_mean
count      630.000000      630.000000
mean       90.544921      3515.128571
std        6.911726      521.973887
min        66.700000      2100.000000
25%        86.025000      3206.500000
50%        91.300000      3431.000000
75%        96.200000      3723.750000
max        100.000000      5617.000000
```


02

Process

Setting up Data, Data
Overview, Data Sorting and
Graph Plotting

Process



Scatter plot of Job Vacancy vs Current GDP per capita

1. Setting up and Data Overview

- Import packages – Pandas, matplotlib.pyplot, seaborn
- Import data: `pd.read_csv(file, na_values = ['-','NA', 'N/A'])`
- Data overview: `df.shape()`, `df.columns`, `df.describe()`, `df.info()`

2. Data sorting and Graph plotting

- Slicing to remove 2019 data from Job vacancy to match with GDP data
- `pd.merge()` by year to combine 2 datasets → `df.reset_index` to get year as column
- **Change colour:** `palette = sns.color_palette("mako_r", 4)` → 4 colour types to distinguish every 10 years from 1990 to 2020.
- **Scatter plot:** `sns.scatter(x,y, hue = 'year', size = 'year', palette)`
- **Legend:** `plt.legend(bbox_to_anchor=(1, 1), loc=2)` to move legend outside the graph plot
- **Labelling:** `ax.set(xlabel,ylabel); ax.set_title()`

Process



Line Plot of Job Vacancy by Industry from 1990 to 2019

1. Setting up and Data Overview

- Import packages – Pandas, matplotlib.pyplot, seaborn
- Import data: `pd.read_csv()`
- Data overview: `df.shape()`, `df.columns`, `df.describe()`, `df.info()`
- Find data year range → `df.year.max()`, `df.year.min()`
- Missing values: `df.count()` → number of non-na values does not match with other columns → missing values

2. Data sorting and Graph plotting

- **Remove unused column:** `df.drop(columns = 'industry2')`
- **Groupby data and aggregate:** `df.groupby(['year', 'industry1']).sum()`
- **Change colour for 4 line plots:** `palette = sns.color_palette("mako_r", 4)`
- **Line plot:** `sns.lineplot(data=df, x="year", y="job_vacancy", style = 'industry1', markers=True, hue="industry1", palette=palette)`
- **Labelling:** `ax.set(xlabel, ylabel); ax.set_title()`
- **Moving legend outside chart and renaming legend title:**
`handles, labels = ax.get_legend_handles_labels()`
`ax.legend(handles=handles[1:], labels=labels[1:], title = 'Industry', bbox_to_anchor = (1, 1), loc = 2)`

Process

- ✍ Box plot of Mean Monthly Income of Local University Graduates in 2018
- ✍ Bar Plot of Employment Rate of Local University Graduates from 2013–2018

1. Setting up and Data Overview

- Import packages – Pandas, matplotlib.pyplot, seaborn
- Import data: `pd.read_csv()`
- Data overview: `df.shape()`, `df.columns`, `df.describe()`, `df.info()`
- Find data year range → `df.year.max()`, `df.year.min()`
- Missing values: `df.count()` → number of non-na values does not match with other columns → missing values

2. Data sorting and Graph plotting

- **Creating shortform for the 6 universitys in a new colums:** `pd.unique(raw['university'].tolist())` to get the list of universities → create shortform in a list → create dictionary with `zip()` → `df['university'].replace(dict, regex=True)`
- Slicing to get 2018 data from mean monthly gross income
- **Groupby data and mean for employment rate:** `df.groupby(['year', 'university']).mean()` → `df.reset_index()` to change index to column
- **Box plot:** `sns.boxplot(x,y,data, hue = 'university')`
- **Bar plot:** `sns.barplot(x,y,data,hue = 'university')`
- **Labelling:** `ax.set(xlabel,ylabel); ax.set_title()`
- **Moving legend outside chart and renaming legend title:**
`handles, labels = ax.get_legend_handles_labels()`
`ax.legend(handles=handles, labels=labels,`
`title = 'University', bbox_to_anchor = (1,1), loc = 2)`

03

Analysis

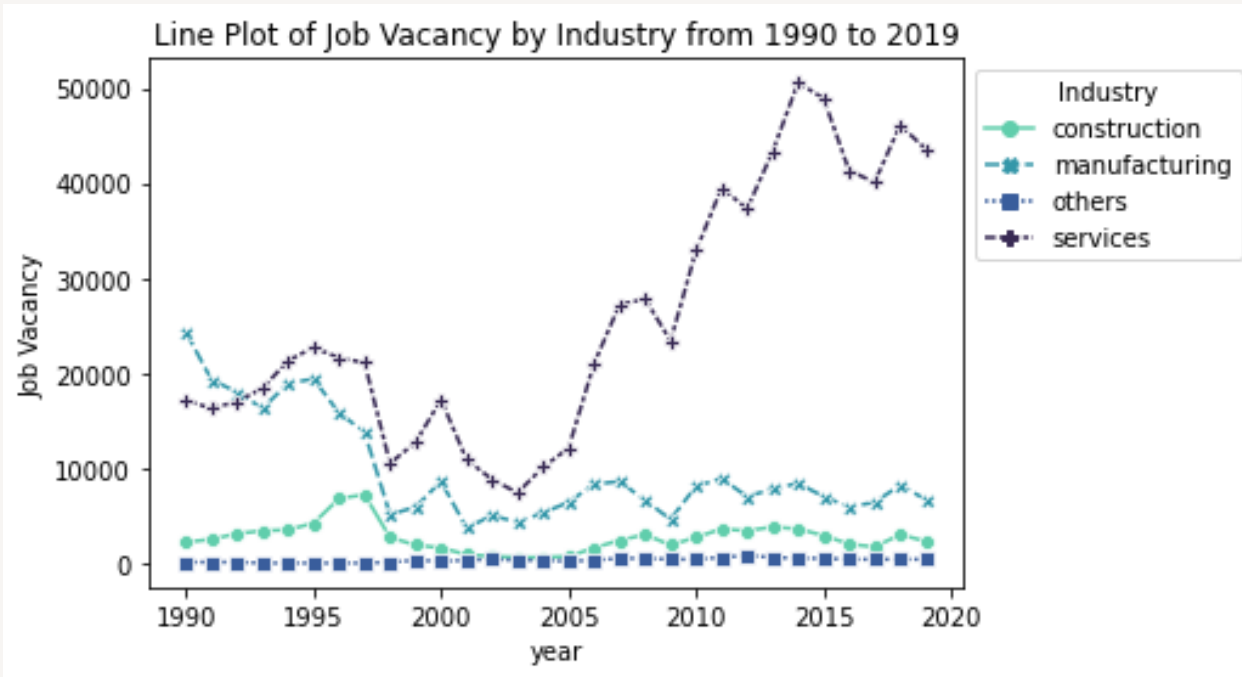
Visualization of data and Insights

Analysis



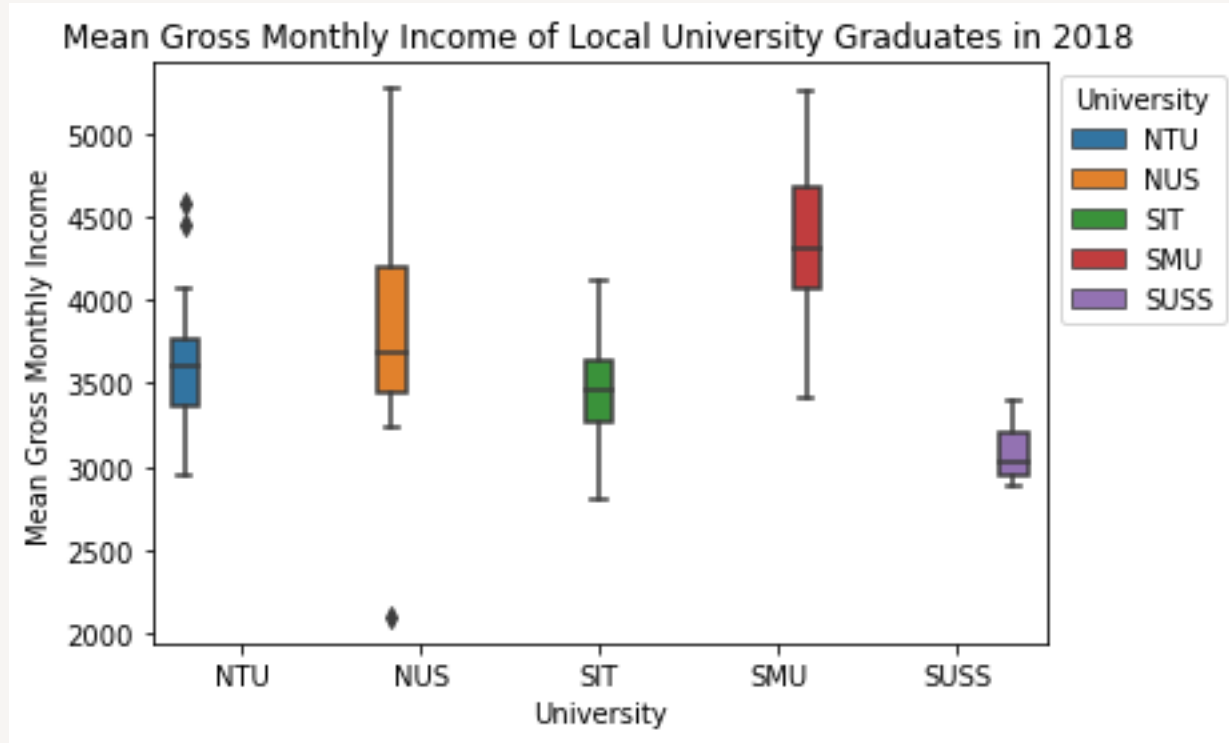
- Generally a positive relationship between job vacancy and GDP from year 2000 onwards
- Abnormality in the early year of 2000-2010. Despite higher GDP, job vacancy dip with the effect of 1997 Asian Financial Crisis
- Global Financial Crisis (2007-2008) does not have a strong effect on the growth of GDP and job market

Analysis



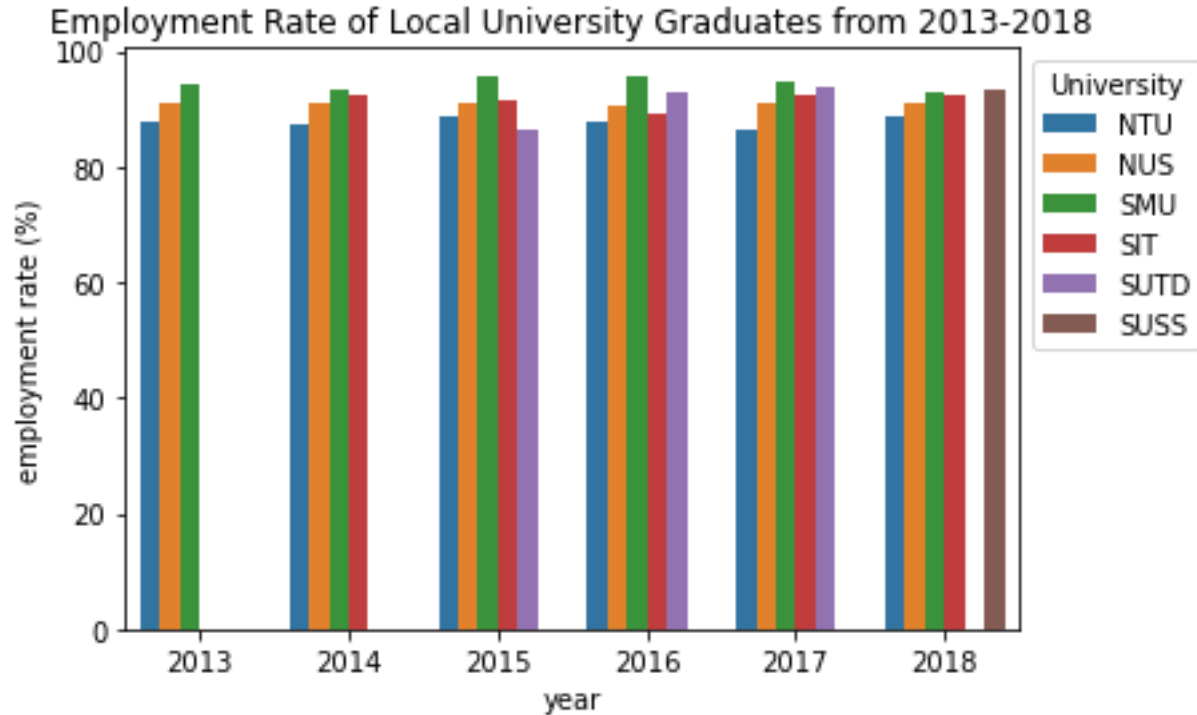
- Increasing trend for service sector and decreasing trend for manufacturing sector over the years
- Construction sector stay relatively constant with peak in 1995-1998
- Sharp decrease across all 3 main industry in 1997 and stayed low with peaks and falls
- Job market picked up in 2005 with sharp increase by the service industry

Analysis



- Generally, graduates from SMU earns a higher income and graduates from SUSS earns the least
- Large variability in earnings from NUS and it is positively skewed

Analysis



- SMU graduates remain the most employable over the years
- Despite being a new university, SUTD employment rate surpassed the other universities in 2016 except SMU
- Majority of the employment rate remains relatively similar over the years except for SIT
- SIT drop in employment rate from 2014-2016 and then increase in 2017

O4

SQL to Python

Setup, Process and Result

SQL to Python





SQL

Local University Graduate Employment Dataset

- CREATE DATABASE mydatabase
- Upload dataset into table 'grad_survey'

```
use mydatabase;
```

```
select year, university, school, degree from grad_survey where basic_monthly_mean > 3800
```

Result Grid				
  Filter Rows: <input type="text"/> Export:  Wrap Cell Content: 				
	year	university	school	degree
	2013	National University of Singapore	School of Computing	Bachelor of Computing (Computer Science)
	2013	National University of Singapore	Faculty of Dentistry	Bachelor of Dental Surgery
	2013	National University of Singapore	Faculty of Law	Bachelor of Laws (LLB) (Hons) #
	2013	National University of Singapore	YLL School of Medicine	Bachelor of Medicine and Bachelor of Surgery (...)
	2013	Singapore Management University	School of Business (4-years programme) *	Business Management (4-years programme) Cu...
	2013	Singapore Management University	School of Law (4-years programme) *	Law (4-years programme) #
	2013	Singapore Management University	School of Law (4-years programme) *	Law (4-years programme) Cum Laude and abov...

SQL to Python

Python

Local University Graduate Employment Dataset

- **Import packages:** `mysql.connector`, `pandas`
- **Import data:** `pd.read_csv(file, na_values = ['na'], encoding = 'latin-1')`
- **Connect to SQL:** `mysql.connector.connect(user='root', password='1234', host='127.0.0.1', database='mydatabase')`
- **Retrieve data from SQL where basic_monthly_mean > 3,800 and store:** `sql = "SELECT year,university, school, degree FROM grad_survey where basic_monthly_mean > 3800"`
`df = pd.read_sql(sql, con = cnx)`
- **Show result:** `print(df), shape(18,5)`

Majority of the graduates with monthly earning > \$3,800 are from Law

SQL to Python

Result:

**The following are the degrees that has a basic monthly mean salary > \$3,800.
(non-exclusive)**

2013	National University of Singapore	School of Computing	Bachelor of Computing (Computer Science)	3933
2013	National University of Singapore	Faculty of Dentistry	Bachelor of Dental Surgery	4106
2013	National University of Singapore	Faculty of Law	Bachelor of Laws (LLB) (Hons) #	4922
2013	National University of Singapore	YLL School of Medicine	Bachelor of Medicine and Bachelor of Surgery (MBBS) #	4406
2013	Singapore Management University	School of Business (4-years programme) *	Business Management (4-years programme) Cum Laude and above	3825
2013	Singapore Management University	School of Law (4-years programme) *	Law (4-years programme) #	5023
2013	Singapore Management University	School of Law (4-years programme) *	Law (4-years programme) Cum Laude and above #	5329
2014	National University of Singapore	NUS Business School	Bachelor of Business Administration (Hons)	3979
2014	National University of Singapore	Faculty of Dentistry	Bachelor of Dental Surgery	4054
2014	National University of Singapore	School of Design & Environment	Bachelor of Arts (Architecture) #	4290
2014	National University of Singapore	Faculty of Law	Bachelor of Laws (LLB) (Hons) #	5027
2014	National University of Singapore	YLL School of Medicine	Bachelor of Medicine and Bachelor of Surgery #	4404
2014	Singapore Management University	School of Accountancy (4-year programme) *	Accountancy (4-year programme) Cum Laude and above	3871
2014	Singapore Management University	School of Business (4-year programme) *	Business Management (4-year programme) Cum Laude and above	4019

Thank you