Task A: Investigating Job Vacancy and Unemployment Rate Data

A1. Investigating the Population Data

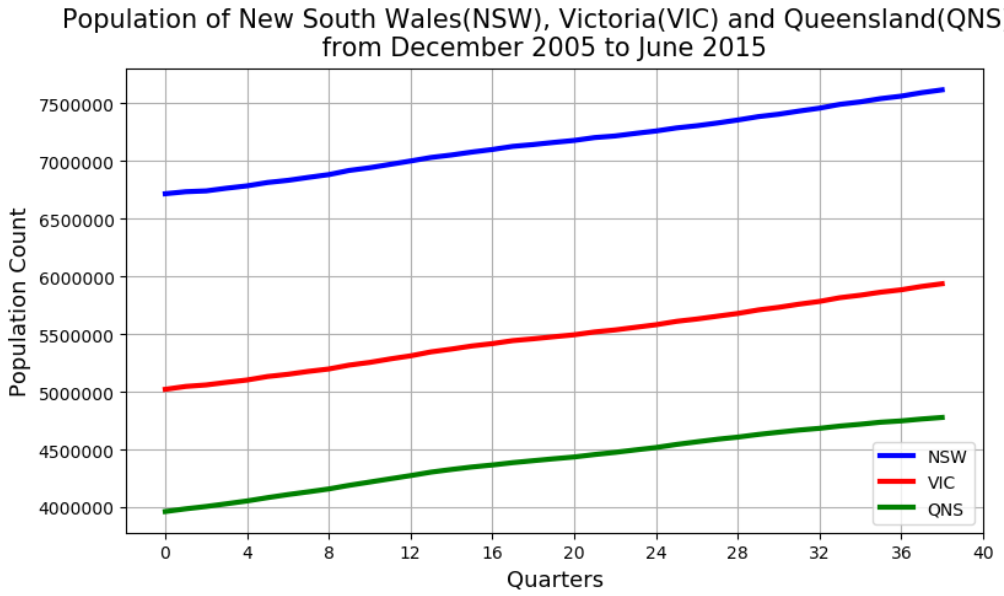
Have a look at the resident population data. You will see many columns. We are interested only in the total values for each state (marked "Persons"), so you can drop the other columns and rename the columns for each state if you wish.

(HINT: The file isn't very big so you can make the changes in Excel if you want.)

1. In Python (or R) plot the population of Victoria, New South Wales and Queensland over time. (HINT: You don't need to put the dates on the x-axis, just showing the index of each quarter is fine)

a) Are the population values increasing or decreasing over time?
b) Does the population data exhibit a trend and if so, what type?

Answer: The below relation is obtained while tracing the count of the population for the three states viz Victoria, New South Wales and Queensland over the time.

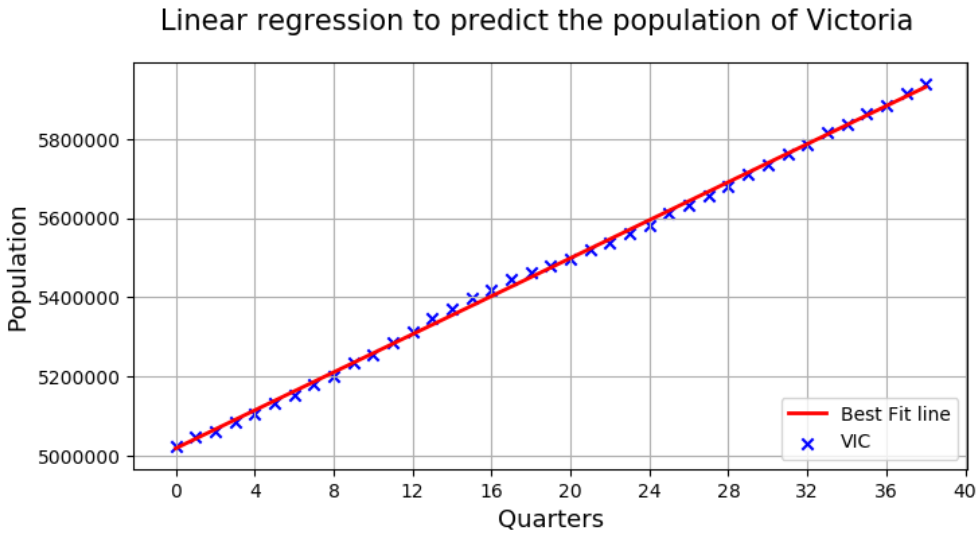


As the graphs are plotted it is evident that the count of the population is gradually increasing for the three states over the time. Queensland has the least population among the three states while New South Wales has the maximum population. The trend is linearly increasing one with a positive slope over the time.

2. Fit a linear regression using Python (or R) to the Victorian population data and plot the linear fit. (HINT: In Python, you can use the "range (1, n)" function to generate a sequence of integer values: 1, 2..., n)

a) *Does the linear fit look good?*
b) *Use the linear fit to predict the resident population in Victoria for the dates: 1/9/15, 1/12/15, 1/12/16, and 1/12/17.*

Answer: The values of the Victorian population is first scattered plotted and then linear regression is applied on the data for best fit line. The linear fit looks definitely good. The graph is as follows:



The predicted population for the given dates are as below:

Date	Population
1/9/15	*5739516.54838*
1/12/15	*5979953.5504*
1/12/16	*6076128.35121*
1/12/17	*6172303.15202*

A2. Investigating the Job Vacancies Data

Now have a look at the job vacancies data.

1. Use Python (or R) to plot the job vacancy counts for Victoria over time. (HINT: Pandas contains a "transpose ()" method and Excel can also be used to transpose data.)

a) *What are maximum and minimum values for job vacancies in Victoria over time period?*

Answer: The vacancy count of Victoria is plotted over time. The graph is as follows:

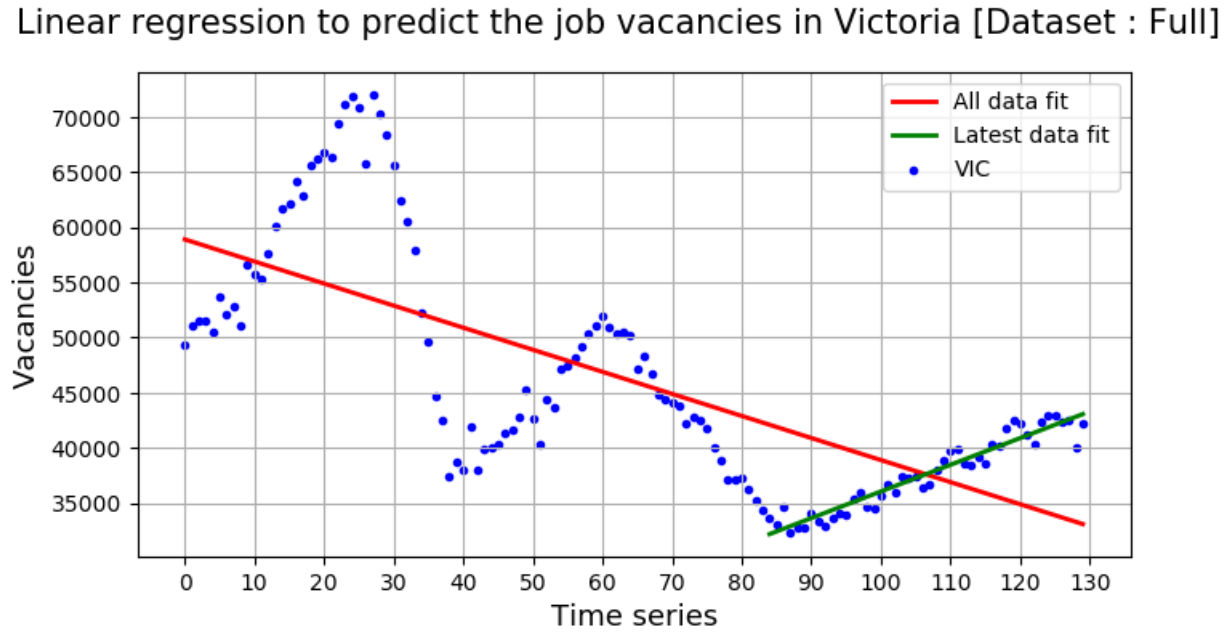
The maximum and the minimum values of the population are *71971* and *32322* respectively.

2. Fit a linear regression to the data and plot it.

a) Does it look like a good fit to you? Would you believe the predictions of the linear model going forward?

b) Instead of fitting the linear regression to all of the data, try fitting it to just the most recent data points (say from the 85th data point onwards). How is the fit? Which model would give better predictions of future vacancies do you think?

Answer: Firstly, the linear regression is implemented on the total Victorian population data. Then the linear regression is implemented on the 85th data onwards. The below graphs are obtained.



The line is definitely not a good fit. The data is arranged as a function of polynomial equation rather than a linear one. In this case a linear fit line will not be able to provide correct estimations of the data. Hence, the linear model based on all the data is not plausible for any prediction.

Choosing the data from the 85th row onwards provides a linear arrangement of data. In this scenario, a linear fit line is desirable. As per the plotted graph, it can be seen that the line fits very close to all the data linearly. Hence to predict a data ***WITHIN*** the time interval [85th Row] to [130th Row], the second model suits the best.

However, to predict the ***FUTURE*** data, none of the above models fits best as it is evident from the history value, the data shows linear trend (both positive and negative slopes) at certain intervals only. It might be the case that the interval from the 131th row onwards shows a linear trend with a downward slope. In this case, the second model fails as well, to predict the data correctly. Here, regression using a polynomial model definitely holds an upper hand than the linear model.

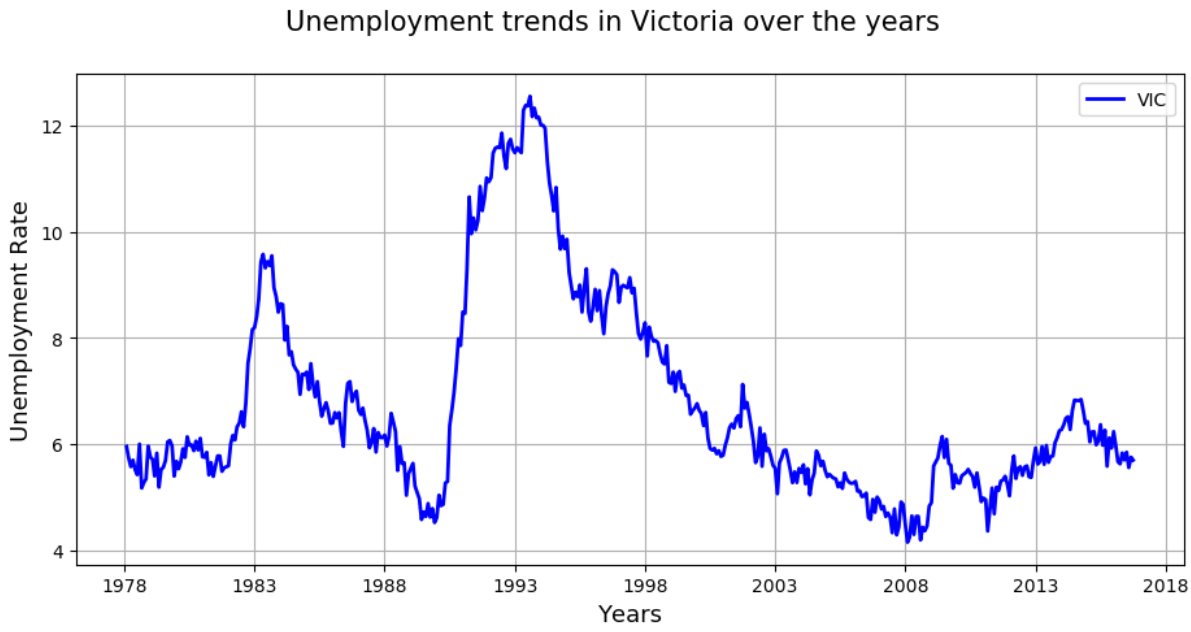
A3. Investigating the Unemployment Data

Now have a look at the unemployment data.

1. Use Python (or R) to plot the Unemployment Rate for Victoria over time.

a) It looks like the rate has been very high at times in the past. What was the maximum unemployment rate in Victoria recorded in the dataset and when did that occur?

Answer: *Next Page (Contd.)*



The maximum unemployment rate was: 12.5533377 during the year 1993 in the month of August.

A4. Visualising the Relationship between Unemployment and Job Vacancies

Now let's look at the relationship between unemployment levels and job vacancies.

1. *Python (or R) to combine the data from the different files into a single table. The table should contain population values, job vacancy counts and unemployment rates for the different dates and different States/Territories.*

a) *What is the first date and last date for the combined data?'*

Answer: The first date and the last date for the combined data are as below:

Argument	Value
Min Date	*2015/03/01*
Max Date	*2015/06/01*

2. Now that you have the data aggregated, we can see whether there is a relationship between unemployment and the number of job vacancies. Plot the values against each other.

a) *Can you see a relationship there?*

Answer: The merged data is now used to plot the unemployment and the vacancy of all the states. A scatter plot has been used instead of a line plot as the graph generated from the scatter plot is more legible in this case. The graph is as below*:*

The above picture shows that the vacancies are quite high when the unemployment rate is between 4 and 6. However the graph fails to produce any meaningful insight. This can be due to the fact that the plotted data contains vacancy rate and unemployment rate of all the States for all quarters in an unstructured way, without any correlation among them.

An approach to deduce a more meaningful relation between unemployment rate and vacancies would be to group the cumulative values (for all states) based on each quarter. On plotting the data, it produces the following graph:

This graph clearly shows that the Vacancy and Unemployment has an inverse relation. As the Vacancy increases gradually the unemployment decreases. This is in accordance to the real-life scenario.

3. Try selecting and plotting only the data from Victoria.

a) Can you see a relationship now? If so, what relationship is there?

Answer: Unlike the previous graph to establish relationship for all states, in this case, the unemployment and the vacancy data is plotted against the state of Victoria only. The below graph is obtained.

The graph correlates to the previous finding of grouped data. Here the Vacancy for the state of Victoria is gradually decreasing as the Unemployment Rate grows. Noteworthy, the vacancies for the state of Victoria are quite high and seemingly unaffected until the unemployment rate reaches the value of 5.

4. The different populations across the states will influence the number of job vacancies in each. Remove this effect by introducing a new column called 'Vacancy Rate' which contains the vacancy count divided by the population size, multiplied by 100.

a) Is there a relationship between the unemployment rate and the job vacancy rate across all the data?

Answer: The column is added to the source data. Now, the vacancy rate and the unemployment rate are plotted for both type of data (Grouped and Ungrouped).

Next Page (Contd.)

Both the above methodology suggests that the Vacancy rate is inversely related to the Unemployment Rate. The Vacancy Rate has clearly shaped the trend in to a more linearly degrading form by omitting the effect of population count.

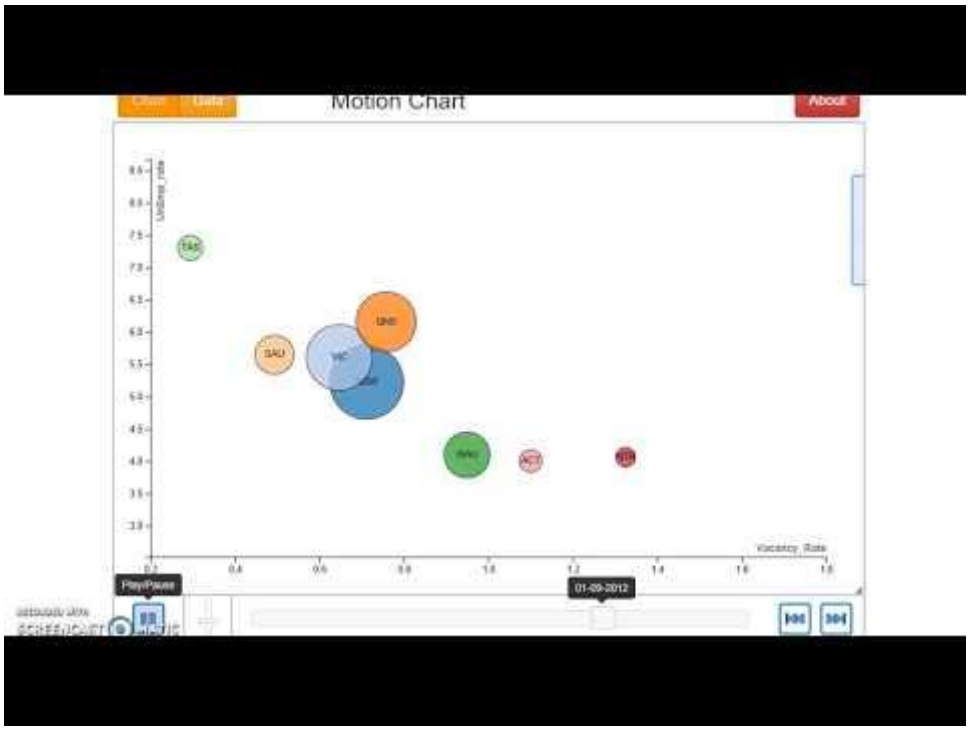
Mention worthy, in all the above cases the vacancies are not impacted by the unemployment rate until it reaches a certain threshold unemployment rate of around 4.5

A5. Visualising the Relationship over Time

Now let's look at the relationship between unemployment levels and job vacancies over time.

1. Use Python (or R) to build a Motion Chart comparing the job vacancy rate, the unemployment rate, and the population of each state over time. The motion chart should show the job vacancy rate on the x-axis, the unemployment rate on the y-axis, and the bubble size should depend on the population. (HINT: A Jupyter notebook containing a tutorial on building motion charts in Python is available here.)

Answer: The motion chart is in the video below:



2. Run the visualisation from start to finish. (Hint: In Python, to speed up the animation, set timer bar next to the play/pause button to the minimum value.) And then answer the following questions:

a) *Which state generally has the lowest job vacancy rate?*
b) *Is the economy generally getting better or worse? I.e. was the Australian economy better in 2006/7 or 2014/5? Explain your answer.*
c) *Compared to the states, does the Northern Territory generally have higher or lower unemployment and higher or lower job vacancy rates? What might cause this? Would it make sense economically to move to NT?*
d) *According to the graph what happened at the end of 2008 and start of 2009? What might have caused this?*
e) *Any other interesting things you notice in the data?*