



Digital Receipt

This receipt acknowledges that **Turnitin** received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Yifan Zhang
Assignment title: Case Study: Submit to Turnitin
Submission title: caseStudy_3_6.pdf
File name: caseStudy_3_6.pdf
File size: 86.66K
Page count: 3
Word count: 1,120
Character count: 5,557
Submission date: 11-Mar-2023 02:09PM (UTC+1000)
Submission ID: 2034456438

Name: Yifan Zhang
Student id: s4686474

DATA7001: Introduction to Data Science, S1, 2023
Case Study Assignment

For this case study, I decided to consider Reports 3 and 6.

Report 3: Unemployment in Queensland

A. Why was this analysis carried out? What value can it bring to stakeholders? Who are the stakeholders?

A study of this issue will allow the Queensland government and the public to understand why this problem is occurring and will allow the government to put policies in place to prevent possible problems and allow the people to know how long the effects of unemployment will probably last.

For the different stakeholders, they can adjust the arrangement in time and make appropriate and positive response.

1. Individuals: This analysis can be used to understand the situation and trend of the job market and make better choices.
2. Government: Use this analysis to make more efficient policies to promote economic development and provide job opportunities.
3. Enterprises: to better understand the supply and demand in the job market and to adjust their recruitment plans.

The main stakeholder may be working or looking for work in Queensland, the Queensland Government and various businesses are also stakeholders.

B. What was the form and format of the data used? What can you say about the data size and specific format?

The main forms of data used are numeric data, image data, time series data and textual data.

For the format of the data used is TXT, Excel (charts).

For the size of the data: I think a large amount of numerical data was collected, perhaps more than ten thousand relevant data.

For the format of the data: I think it is the year, month, number of people, age of people, etc., to unify the data to be processed and to make it easy to integrate the processed data.

C. Based on what you can understand – what tools were used in the analysis? You don't have to understand all of the tools used – but look some of them up to see what is the purpose?

R was used.

R: Checked data integrity and processed missing data. Filtered out the requiring data and used graphing tools to make the processed data clearer and more concise.

Used line analysis and logistic regression analysis to analyze the data.

Line analysis: Line analysis was performed on three different sets of data, and r-squared was used to determine which set of data was more correlated.

Logistic regression analysis: logistic regression was performed by setting up two different groups to determine which group had the larger percentage.

D. How can such analysis affect decisions? What are important questions that need to be solved?