**COMM5000**

**Data Literacy**

**Sandbox PwC Distribution Project**

**Milestone 2**

Student Name & ID: Ankit Agja & Z5400724

# Introduction

The given dataset contains various financial parameters of Australian companies for the years 2016 to 2021. This dataset includes important financial variables such as fixed assets, current assets, stock, debtors, cash and cash equivalents, total assets, non-current liabilities, long-term debt, current liabilities, working capital, number of employees, operating revenue, costs of goods sold, gross profit, other operating expenses, operating profit/loss, profit/loss before tax, and depreciation & amortization. The primary objective of analyzing this dataset is to identify the factors that affect the financial performance of companies in Australia. For this purpose, we can perform hypothesis testing to determine if there is a statistically significant relationship between the financial variables and the financial performance of the companies. In this analysis, we can identify the key financial drivers of the company's profitability, liquidity, and solvency by exploring the relationship between various financial parameters. By analyzing this dataset, we can determine the strengths and weaknesses of companies in terms of financial performance, which can be used to make better investment decisions. Therefore, the analysis of this dataset can be beneficial for investors, financial analysts, and other stakeholders interested in Australian companies.

# Hypothesis Testing

We choose Operating revenue (Turnover), Costs of goods sold, and Gross Profit to determine our hypothesis testing in different-different countries.

1. **Operating Revenue**

**Result of Australia-Korea**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances |  | 0.05 |  | 1 | Column Order |
| Unequal Sample Sizes |  |  |  |  |  |
|  | *Data1* | *Data2* | diff | 95% Confidence Interval | |
| Mean | 128935.5184 | 8768.284128 | 120167.234 | 93318.893 | 147015.576 |
| Variance | 3.20741E+11 | 806619079.5 |  |  |  |
| Observations | 1716 | 1715 |  |  |  |
| Hypothesized Mean Difference | 0 |  |  |  |  |
| df | 1723 |  |  |  |  |
| t Stat | 8.779 |  |  |  |  |
| P(T<=t) one-tail - Difference < Hypothesized Difference | 0.000 | 1.000 | Difference > Hypothesized Difference | | |
| T Critical one-tail | 1.646 |  |  |  |  |
| P(T<=t) two-tail | 0.000 |  | Reject Null Hypothesis because p < 0.05 (Means are Different) | | |
| T Critical Two-tail | 1.961 |  |  |  |  |

The null hypothesis being tested is that the mean difference between the two sets of data is zero, while the alternative hypothesis is that the mean difference is not zero. The results indicate that the mean of Data1 is 128,935.5184, and the mean of Data2 is 8,768.284128. The difference between the two means is 120,167.234. The 95% confidence interval for this difference is (93,318.893, 147,015.576), meaning that we can be 95% confident that the true difference between the two population means falls within this range. The t-test statistic is 8.779, and the p-value for a two-tailed test is 0.000, which means that we reject the null hypothesis. The degrees of freedom for the test are 1723, and the critical values for a two-tailed test at the 0.05 level of significance are -1.961 and +1.961. Since the calculated t-value is greater than the critical value, we conclude that there is strong evidence that the means of the two populations are different. Based on these results, the null hypothesis is rejected because the p-value is less than 0.05. Therefore, it can be concluded that the means of the two populations are different.

* **Country Comparison:**

Test = 1.55199E-05

* **Covid Effect Test:**

Test value = 4.53983E-06

**Result of Korea-New Zealand**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances |  | 0.05 |  | 1 | Column Order |
| Unequal Sample Sizes |  |  |  |  |  |
|  | *Data1* | *Data2* | diff | 95% Confidence Interval | |
| Mean | 8768.284 | 106796.8 | -98028.525 | -146638.739 | -49418.310 |
| Variance | 8.07E+08 | 1.9E+11 |  |  |  |
| Observations | 1715 | 312 |  |  |  |
| Hypothesized Mean Difference | 0 |  |  |  |  |
| df | 311 |  |  |  |  |
| t Stat | -3.968 |  |  |  |  |
| P(T<=t) one-tail - Difference < Hypothesized Difference | 0.000 | 1.000 | Difference > Hypothesized Difference | | |
| T Critical one-tail | 1.650 |  |  |  |  |
| P(T<=t) two-tail | 0.000 |  | Reject Null Hypothesis because p < 0.05 (Means are Different) | | |
| T Critical Two-tail | 1.968 |  |  |  |  |

The null hypothesis being tested is that the mean difference between the two sets of data is zero, while the alternative hypothesis is that the mean difference is not zero. The results indicate that the mean of Data1 is 8,768.284, and the mean of Data2 is 106,796.8. The difference between the two means is -98,028.525, which means that Data2 has a significantly larger mean than Data1. The 95% confidence interval for this difference is (-146,638.739, -49,418.310), meaning that we can be 95% confident that the true difference between the two population means falls within this range. The t-test statistic is -3.968, and the p-value for a two-tailed test is 0.000, which means that we reject the null hypothesis. The degrees of freedom for the test are 311, and the critical values for a two-tailed test at the 0.05 level of significance are -1.968 and +1.968. Since the calculated t-value falls outside this range, we conclude that there is strong evidence that the means of the two populations are different.

* **Country Comparison:**

Test Value = 9.09991E-06

* Covid Effect Test:

Test Value = 2.88358E-05

**Result of New Zealand-Australia**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances |  | 0.05 |  | 1 | Column Order |
| Unequal Sample Sizes |  |  |  |  |  |
|  | *Data1* | *Data2* | diff | 95% Confidence Interval | |
| Mean | 128935.5 | 106796.8 | 22138.710 | -33314.655 | 77592.075 |
| Variance | 3.21E+11 | 1.9E+11 |  |  |  |
| Observations | 1716 | 312 |  |  |  |
| Hypothesized Mean Difference | 0 |  |  |  |  |
| df | 521 |  |  |  |  |
| t Stat | 0.784 |  |  |  |  |
| P(T<=t) one-tail - Difference < Hypothesized Difference | 0.217 | 0.783 | Difference > Hypothesized Difference | | |
| T Critical one-tail | 1.648 |  |  |  |  |
| P(T<=t) two-tail | 0.433 |  | Cannot Reject Null Hypothesis because p > 0.05 (Means are not different) | | |
| T Critical Two-tail | 1.965 |  |  |  |  |

The null hypothesis being tested is that the mean difference between the two sets of data is zero, while the alternative hypothesis is that the mean difference is not zero. The results indicate that the mean of Data1 is 128,935.5, and the mean of Data2 is 106,796.8. The difference between the two means is 22,138.710, which suggests that Data1 has a slightly larger mean than Data2. The 95% confidence interval for this difference is (-33,314.655, 77,592.075), which means that we cannot be 95% confident that the true difference between the two population means falls within this range. The t-test statistic is 0.784, and the p-value for a two-tailed test is 0.433, which means that we cannot reject the null hypothesis. The degrees of freedom for the test is 521, and the critical values for a two-tailed test at the 0.05 level of significance are -1.965 and +1.965. Since the calculated t-value falls within this range, we conclude that there is not enough evidence to suggest that the means of the two populations are different. Based on these results, the null hypothesis is rejected because the p-value is less than 0.05. Therefore, it can be concluded that the means of the two populations are different.

* **Country Comparison:**

Test Value = 1.66879E-06

* **Covid Effect Test:**

Test Value = 5.45E-07

1. **Costs of goods sold**

**Result of Australia-Korea**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances |  | 0.05 |  | 1 | Column Order |
| Equal Sample Sizes |  |  |  |  |  |
|  | *Data1* | *Data2* | diff | 95% Confidence Interval | |
| Mean | 102024.1 | 6891.822 | 95132.231 | 72634.326 | 117630.136 |
| Variance | 2.25E+11 | 4.31E+08 |  |  |  |
| Observations | 1716 | 1716 |  |  |  |
| Hypothesized Mean Difference | 0 |  |  |  |  |
| df | 1721 |  |  |  |  |
| t Stat | 8.294 |  |  |  |  |
| P(T<=t) one-tail - Difference < Hypothesized Difference | 0.000 | 1.000 | Difference > Hypothesized Difference | | |
| T Critical one-tail | 1.646 |  |  |  |  |
| P(T<=t) two-tail | 0.000 |  | Reject Null Hypothesis because p < 0.05 (Means are Different) | | |
| T Critical Two-tail | 1.961 |  |  |  |  |

The null hypothesis being tested is that the mean difference between the two sets of data is zero, while the alternative hypothesis is that the mean difference is not zero. The results indicate that the mean of Data1 is 102,024.1, and the mean of Data2 is 6,891.822. The difference between the two means is 95,132.231, which suggests that Data1 has a significantly larger mean than Data2. The 95% confidence interval for this difference is (72,634.326, 117,630.136), which means that we can be 95% confident that the true difference between the two population means falls within this range. The t-test statistic is 8.294, and the p-value for a two-tailed test is 0.000, which means that we reject the null hypothesis. The degrees of freedom for the test is 1721, and the critical values for a two-tailed test at the 0.05 level of significance are -1.961 and +1.961. Since the calculated t-value falls outside this range, we conclude that there is enough evidence to suggest that the means of the two populations are different. Based on these results, the null hypothesis is rejected because the p-value is less than 0.05. Therefore, it can be concluded that the means of the two populations are different.

* **Country Comparison:**

Test Value = 1.74872E-05

* **Covid Effect Test:**

Test value = 2.33E-05

**Result of Korea-New Zealand**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances |  | 0.05 |  | 1 | Column Order |
| Unequal Sample Sizes |  |  |  |  |  |
|  | *Data1* | *Data2* | diff | 95% Confidence Interval | |
| Mean | 6891.822 | 81562.76 | -74670.937 | -115872.814 | -33469.060 |
| Variance | 4.31E+08 | 1.37E+11 |  |  |  |
| Observations | 1716 | 312 |  |  |  |
| Hypothesized Mean Difference | 0 |  |  |  |  |
| df | 311 |  |  |  |  |
| t Stat | -3.566 |  |  |  |  |
| P(T<=t) one-tail - Difference < Hypothesized Difference | 0.000 | 1.000 | Difference > Hypothesized Difference | | |
| T Critical one-tail | 1.650 |  |  |  |  |
| P(T<=t) two-tail | 0.000 |  | Reject Null Hypothesis because p < 0.05 (Means are Different) | | |
| T Critical Two-tail | 1.968 |  |  |  |  |

The output shows the results of a two-sample t-test assuming unequal variances to determine whether the means of two independent groups are statistically different. The analysis shows that there is a significant difference between the means of the two datasets, with Data2 having a significantly higher mean value than Data1. The 95% confidence interval for the mean difference is also provided, along with the probability values for a one-tailed and two-tailed test, which both indicate a statistically significant difference in means. Therefore, the null hypothesis of no difference in means is rejected, and it is concluded that the means of the two datasets are statistically different. Based on these results, the null hypothesis is rejected because the p-value is less than 0.05. Therefore, it can be concluded that the means of the two populations are different.

* Country Comparison:

Test value = 9.64654E-06

* Covid Effect Test:

Test value = 3.3977E-05

**Result of New Zealand-Australia**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances |  | 0.05 |  | 1 | Column Order |
| Unequal Sample Sizes |  |  |  |  |  |
|  | *Data1* | *Data2* | diff | 95% Confidence Interval | |
| Mean | 102024.1 | 81562.76 | 20461.293 | -26423.928 | 67346.515 |
| Variance | 2.25E+11 | 1.37E+11 |  |  |  |
| Observations | 1716 | 312 |  |  |  |
| Hypothesized Mean Difference | 0 |  |  |  |  |
| df | 516 |  |  |  |  |
| t Stat | 0.857 |  |  |  |  |
| P(T<=t) one-tail - Difference < Hypothesized Difference | 0.196 | 0.804 | Difference > Hypothesized Difference | | |
| T Critical one-tail | 1.648 |  |  |  |  |
| P(T<=t) two-tail | 0.392 |  | Cannot Reject Null Hypothesis because p > 0.05 (Means are not different) | | |
| T Critical Two-tail | 1.965 |  |  |  |  |

The mean value of Data1 is 102024.1, and the mean value of Data2 is 81562.76. The difference between the means is 20461.293, indicating that Data1 has a slightly higher mean value than Data2. The 95% confidence interval for the mean difference is between -26423.928 and 67346.515, which includes the value of 0. Therefore, there is no statistically significant difference between the means of the two datasets at the 0.05 significance level. The t-statistic is 0.857, which is lower than the critical value of 1.965 for a two-tailed test at the 0.05 significance level, and the p-value is 0.392, which is greater than 0.05. Therefore, the null hypothesis of no difference in means is not rejected, and it is concluded that the means of the two datasets are not statistically different. The two-tailed p-value is 0.392, greater than the significance level of 0.05. The critical t-value for a two-tailed test with 516 degrees of freedom is 1.965. In this, the Null Hypothesis cannot be rejected.

* Country Comparison:

Test value = 2.16267E-06

* Covid Effect Test:

Test value = 6.15E-07

1. **Gross Profit**

**Result of Australia-Korea**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances |  | 0.05 |  | 1 | Column Order |
| Equal Sample Sizes |  |  |  |  |  |
|  | *Data1* | *Data2* | diff | 95% Confidence Interval | |
| Mean | 26625.18 | 1872.587 | 24752.592 | 19443.372 | 30061.813 |
| Variance | 1.25E+10 | 93370722 |  |  |  |
| Observations | 1716 | 1716 |  |  |  |
| Hypothesized Mean Difference | 0 |  |  |  |  |
| df | 1740 |  |  |  |  |
| t Stat | 9.144 |  |  |  |  |
| P(T<=t) one-tail - Difference < Hypothesized Difference | 0.000 | 1.000 | Difference > Hypothesized Difference | | |
| T Critical one-tail | 1.646 |  |  |  |  |
| P(T<=t) two-tail | 0.000 |  | Reject Null Hypothesis because p < 0.05 (Means are Different) | | |
| T Critical Two-tail | 1.961 |  |  |  |  |

The mean value of Data1 is 6891.822, and the mean value of Data2 is 81562.76. The difference between the means is -74670.937, indicating that Data2 has a significantly higher mean value than Data1. The 95% confidence interval for the mean difference is between -115872.814 and -33469.060, which does not include the value of 0. Therefore, there is a statistically significant difference between the means of the two datasets at the 0.05 significance level. The t-statistic is -3.566, which is lower than the critical value of 1.968 for a two-tailed test at the 0.05 significance level, and the p-value is less than 0.05. Therefore, the null hypothesis of no difference in means is rejected, and it is concluded that the means of the two datasets are statistically different, with Data2 having a higher mean value than Data1. Based on these results, the null hypothesis is rejected because the p-value is less than 0.05. Therefore, it can be concluded that the means of the two populations are different.

* Country Comparison:

Test value = 8.21535E-05

* Covid Effect Test:

Test value = 1.79E-06

**Result of Korea-New Zealand**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances |  | 0.05 |  | 1 | Column Order |
| Unequal Sample Sizes |  |  |  |  |  |
|  | *Data1* | *Data2* | diff | 95% Confidence Interval | |
| Mean | 1872.587 | 25210.83 | -23338.245 | -33413.388 | -13263.101 |
| Variance | 93370722 | 8.16E+09 |  |  |  |
| Observations | 1716 | 312 |  |  |  |
| Hypothesized Mean Difference | 0 |  |  |  |  |
| df | 312 |  |  |  |  |
| t Stat | -4.558 |  |  |  |  |
| P(T<=t) one-tail - Difference < Hypothesized Difference | 0.000 | 1.000 | Difference > Hypothesized Difference | | |
| T Critical one-tail | 1.650 |  |  |  |  |
| P(T<=t) two-tail | 0.000 |  | Reject Null Hypothesis because p < 0.05 (Means are Different) | | |
| T Critical Two-tail | 1.968 |  |  |  |  |

The mean value of Data1 is 1872.587, and the mean value of Data2 is 25210.83. The difference between the means is -23338.245, indicating that Data2 has a significantly higher mean value than Data1. The 95% confidence interval for the mean difference is between -33413.388 and -13263.101, which does not include the value of 0. Therefore, there is a statistically significant difference between the means of the two datasets at the 0.05 significance level. The t-statistic is -4.558, which is lower than the critical value of 1.968 for a two-tailed test at the 0.05 significance level, and the p-value is less than 0.05. Therefore, the null hypothesis of no difference in means is rejected, and it is concluded that the means of the two datasets are statistically different, with Data2 having a higher mean value than Data1.

* Country Comparison:

Test value = 5.04961E-05

* Covid Effect Test:

Test value = 8.28251E-05

**Result of New Zealand-Australia**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances |  | 0.05 |  | 1 | Column Order |
| Unequal Sample Sizes |  |  |  |  |  |
|  | *Data1* | *Data2* | diff | 95% Confidence Interval | |
| Mean | 26625.18 | 25210.83 | 1414.348 | -9946.868 | 12775.564 |
| Variance | 1.25E+10 | 8.16E+09 |  |  |  |
| Observations | 1716 | 312 |  |  |  |
| Hypothesized Mean Difference | 0 |  |  |  |  |
| df | 500 |  |  |  |  |
| t Stat | 0.245 |  |  |  |  |
| P(T<=t) one-tail - Difference < Hypothesized Difference | 0.403 | 0.597 | Difference > Hypothesized Difference | | |
| T Critical one-tail | 1.648 |  |  |  |  |
| P(T<=t) two-tail | 0.807 |  | Cannot Reject Null Hypothesis because p > 0.05 (Means are not different) | | |
| T Critical Two-tail | 1.965 |  |  |  |  |

The t statistic is 0.245, which is less than the critical t value of 1.965 at a significance level of 0.05 with 500 degrees of freedom. The p-value (two-tail) is 0.807, which is greater than 0.05. Therefore, we cannot reject the null hypothesis.

* Country Comparison:

Test value = 2.56359E-06

* Covid Effect Test:

Test value = 2.42E-06

# Conclusion

Based on the analysis of the financial data provided in the dataset, we performed hypothesis testing on three variables: operating revenue, costs of goods sold, and gross profit to determine the financial performance of Australian companies in comparison to other countries such as Korea and New Zealand. For the operating revenue, the results showed that there was a significant difference in the means of the two populations between Australia and Korea and Australia and New Zealand. However, there was no significant difference between New Zealand and Australia. This indicates that Australian companies perform differently in terms of operating revenue when compared to Korea and New Zealand. Similarly, for the costs of goods sold, there was a significant difference in the means of the two populations between Australia and Korea. This indicates that Australian companies have different costs of goods sold compared to Korea. However, there was no significant difference between New Zealand and Australia. The financial data provided in the dataset and the results of our hypothesis testing suggest that there are significant differences in the financial performance of Australian companies in comparison to other countries, particularly in operating revenue and costs of goods sold. This information can be useful for investors and analysts to make informed decisions about investing in Australian companies. However, further analysis and investigation may be required to determine the reasons for these differences and their implications for the companies' future performance.