**Hypothesis Testing**

For all the hypotheses, mention the following:

Hypothesis testing of Mean( annual\_inc) for loan status as “default\_new” and mean(annual\_inc) for loan status as “current\_new”

1. **Null hypothesis**: Mean( annual\_inc) for loan status as “default\_new” = mean(annual\_inc) for loan status as “current\_new”
2. **Alternate hypothesis:** : Mean( annual\_inc) for loan status as “default\_new” ! = mean(annual\_inc) for loan status as “current\_new”
3. **Test statistic: -**10.639
4. **p value corresponding to the test statistic:** 1.308887e-25
5. **Insight (explain in less than 100 words) :** For 95% confidence interval and 2 tail test the test statistic t should lie between -1.96 to +1.96. The t value -10.639 does not fall in this interval. Hence, we can successfully reject the Null Hypothesis. Also from the mean values we can deduce that annual\_inc of people falling in current\_new category is more than that of default\_new. But the number of loan defaulters in the default new category is more than that of current\_new. **Hence, we can say that people with less annual\_inc have a tendency to default more.**

Hypothesis testing of Mean( loan\_amnt) for loan status as “default\_new” and mean(loan\_amnt) for loan status as “current\_new”

1. **Null hypothesis**: Mean(loan\_amnt) for loan status as “default\_new” = mean(loan\_amnt) for loan status as “current\_new”
2. **Alternate hypothesis:** : Mean(loan\_amnt) for loan status as “default\_new” ! = mean(loan\_amnt) for loan status as “current\_new”
3. **Test statistic:** -19.407
4. **p value corresponding to the test statistic:** 2.958715e-75
5. **Insight (explain in less than 100 words) :** For 95% confidence interval and 2 tail test the test statistic t should lie between -1.96 to +1.96. The t value -19.407 does not fall in this interval. Hence, we can successfully reject the Null Hypothesis. Also from the mean values we can deduce that loan\_amnt of people falling in current\_new category is more than that of default\_new. But the number of loan defaulters in the default new category is more than that of current\_new. **Hence, we can say that people with less annual\_inc have a lesser loan amount and have a tendency to default more.**

Hypothesis testing of Mean(funded\_amnt) for loan status as “default\_new” and mean(funded\_amnt) for loan status as “current\_new”

1. **Null hypothesis**: Mean(funded\_amnt) for loan status as “default\_new” = mean(funded\_amnt) for loan status as “current\_new”
2. **Alternate hypothesis:** : Mean(funded\_amnt) for loan status as “default\_new” ! = mean(funded\_amnt) for loan status as “current\_new”
3. **Test statistic:**  -19.016
4. **p value corresponding to the test statistic**: 7.948493e-73
5. **Insight (explain in less than 100 words) :** For 95% confidence interval and 2 tail test the test statistic t should lie between -1.96 to +1.96. The t value -19.016 does not fall in this interval. Hence, we can successfully reject the Null Hypothesis. Also from the mean values we can deduce that funded\_amnt of people falling in current\_new category is more than that of default\_new. But the number of loan defaulters in the default new category is more than that of current\_new. **Hence, we can say that people with less annual\_inc have a lesser funded\_amnt** **and have a tendency to default more.**

Hypothesis testing of Mean(dti) for loan status as “default\_new” and mean(dti) for loan status as “current\_new”

1. **Null hypothesis**: Mean(dti) for loan status as “default\_new” = mean(dti) for loan status as “current\_new”
2. **Alternate hypothesis:** : Mean(dti) for loan status as “default\_new” ! = mean(dti) for loan status as “current\_new”
3. **Test statistic:**  -3.093
4. **p value corresponding to the test statistic**: 0.002015514
5. **Insight (explain in less than 100 words) :** For 95% confidence interval and 2 tail test the test statistic t should lie between -1.96 to +1.96. The t value -3.093 does not fall in this interval. Hence, we can successfully reject the Null Hypothesis. Also from the mean values we can deduce that dti of people falling in current\_new category is almost same as default\_new. But the number of loan defaulters in the default new category is more than that of current\_new. **Hence, we can say that people with less annual\_inc have a lesser funded\_amnt** **and have a tendency to default more.**

Hypothesis testing of Mean(annual\_inc) for int\_rate\_grp as “High” and mean(annual\_inc) for int\_rate\_grp as “Low”

1. **Null hypothesis**: Mean(annual\_inc) for int\_rate\_grp as “High” = Mean(annual\_inc) for int\_rate\_grp as “Low”
2. **Alternate hypothesis:** : Mean(annual\_inc) for int\_rate\_grp as “High” != Mean(annual\_inc) for int\_rate\_grp as “Low”
3. **Test statistic:**  12.034
4. **p value corresponding to the test statistic**: 3.179604e-32
5. **Insight (explain in less than 100 words) :** For 95% confidence interval and 2 tail test the test statistic t should lie between -1.96 to +1.96. The t value 12.034 does not fall in this interval. Hence, we can successfully reject the Null Hypothesis. Also from the mean values we can deduce that annual\_inc of people falling in High category is more than Low. The number of loan defaulters in the High category is more than that of Low. **Hence, we can say that people with high annual\_inc have int\_rate\_grp as High** **and have a tendency to default more.**

Hypothesis testing of Mean(loan\_amnt) for int\_rate\_grp as “High” and mean(loan\_amnt) for int\_rate\_grp as “Low”

1. **Null hypothesis**: Mean(loan\_amnt) for int\_rate\_grp as “High” = Mean(loan\_amnt) for int\_rate\_grp as “Low”
2. **Alternate hypothesis:** : Mean(loan\_amnt) for int\_rate\_grp as “High” != Mean(loan\_amnt) for int\_rate\_grp as “Low”
3. **Test statistic:**  29.92
4. **p value corresponding to the test statistic**: 1.434289e-160
5. **Insight (explain in less than 100 words) :** For 95% confidence interval and 2 tail test the test statistic t should lie between -1.96 to +1.96. The t value 29.92 does not fall in this interval. Hence, we can successfully reject the Null Hypothesis. Also from the mean values we can deduce that loan\_amnt of people falling in High category is more than Low. The number of loan defaulters in the High category is more than that of Low. **Hence, we can say that people with high loan\_amnt have int\_rate\_grp as High** **and have a tendency to default more.**

Hypothesis testing of Mean(funded\_amnt) for int\_rate\_grp as “High” and mean(funded\_amnt) for int\_rate\_grp as “Low”

1. **Null hypothesis**: Mean(funded\_amnt) for int\_rate\_grp as “High” = Mean(funded\_amnt) for int\_rate\_grp as “Low”
2. **Alternate hypothesis:** : Mean(funded\_amnt) for int\_rate\_grp as “High” != Mean(funded\_amnt) for int\_rate\_grp as “Low”
3. **Test statistic:**  26.684
4. **p value corresponding to the test statistic**: 3.981736e-134
5. **Insight (explain in less than 100 words) :** For 95% confidence interval and 2 tail test the test statistic t should lie between -1.96 to +1.96. The t value 26.684 does not fall in this interval. Hence, we can successfully reject the Null Hypothesis. Also from the mean values we can deduce that funded\_amnt of people falling in High category is more than Low. The number of loan defaulters in the High category is more than that of Low. **Hence, we can say that people with high funded\_amnt have int\_rate\_grp as High** **and have a tendency to default more.**

Hypothesis testing of Mean(dti) for int\_rate\_grp as “High” and mean(dti) for int\_rate\_grp as “Low”

1. **Null hypothesis**: Mean(dti) for int\_rate\_grp as “High” = Mean(dti) for int\_rate\_grp as “Low”
2. **Alternate hypothesis:** : Mean(dti) for int\_rate\_grp as “High” != Mean(dti) for int\_rate\_grp as “Low”
3. **Test statistic:**  3.9043
4. **p value corresponding to the test statistic**: 9.794598e-05
5. **Insight (explain in less than 100 words) :** For 95% confidence interval and 2 tail test the test statistic t should lie between -1.96 to +1.96. The t value 3.9043 does not fall in this interval. Hence, we can successfully reject the Null Hypothesis. Also from the mean values we can deduce that dti of people falling in High category is more than Low. The number of loan defaulters in the High category is more than that of Low. **Hence, we can say that people with high dti have int\_rate\_grp as High** **and have a tendency to default more.**