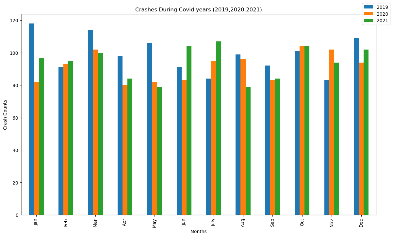
Anova test : Was the variance in States’ road fatalities month by month during covid lockdown following a similar variance?

Null Hypothesis: There was no variance across States’ road fatalities across the 3 years of Covid month by month.

This code is one independent block on jupyter notepad, after this graph



df\_2020cvd\_state\_month = df\_2020.groupby(['State', 'Month'])[['Crash ID']].count().reset\_index()

df\_2021cvd\_state\_month = df\_2021.groupby(['State', 'Month'])[['Crash ID']].count().reset\_index()

df\_2021cvd\_state\_month['Month'] = df\_2021cvd\_state\_month['Month'] + 12

df\_2022cvd\_state\_month = df\_2022.groupby(['State', 'Month'])[['Crash ID']].count().reset\_index()

df\_2022cvd\_state\_month['Month'] = df\_2022cvd\_state\_month['Month'] + 24

df\_statecvd\_month = pd.concat([df\_2020cvd\_state\_month, df\_2021cvd\_state\_month, df\_2022cvd\_state\_month])

df\_statecvd\_month = df\_statecvd\_month.reset\_index()

df\_statecvd\_month = df\_statecvd\_month.rename(columns={'Crash ID': 'Fatality'})

import statsmodels.api as sm

from statsmodels.formula.api import ols

# Fit the model

model = ols('Fatality ~ State + Month + State:Month', data=df\_statecvd\_month).fit()

# Perform the ANOVA test

table = sm.stats.anova\_lm(model, typ=2)

# Print the results

print(table)

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$

what do these results mean?

sum\_sq df F PR(>F)

State 19537.349500 7.0 186.744801 4.121640e-96

Month 51.785377 1.0 3.464879 6.384480e-02

State:Month 98.381721 7.0 0.940367 4.758301e-01

Residual 3781.286705 253.0 NaN NaN

Chat GPT says

These results are the output of the ANOVA test, which tests for differences in means across groups. The table shows the sum of squares (sum\_sq), degrees of freedom (df), F-statistic (F), and the p-value (PR(>F)) for each variable in the model.

* Sum of squares (sum\_sq): the amount of variation explained by each variable in the model.
* Degrees of freedom (df): the number of degrees of freedom associated with each variable. For the State and State:Month variables, there are 7 degrees of freedom because there are 8 states in the model. For the Month variable, there is 1 degree of freedom because there are 12 months in the model. For the Residual variable, there are 253 degrees of freedom, which is the total number of observations minus the number of parameters in the model.
* F-statistic (F): the ratio of the mean square of the variable to the mean square of the residual. It tests whether the group means are significantly different from each other.
* p-value (PR(>F)): the probability of observing the F-statistic or a more extreme value under the null hypothesis that the variable has no effect. A small p-value (< 0.05) indicates that the variable has a significant effect on the outcome.

In this case, the State variable has a very small p-value, indicating that there is a significant difference in fatality counts among the states. The Month variable has a p-value of 0.06, which is close to the significance threshold, suggesting that there may be a weak effect of month on fatality counts. The State:Month interaction variable has a p-value of 0.48, indicating that there is no significant interaction effect between State and Month. The Residual variable has no associated F-statistic or p-value because it represents the variation in the data that is not explained by the model.