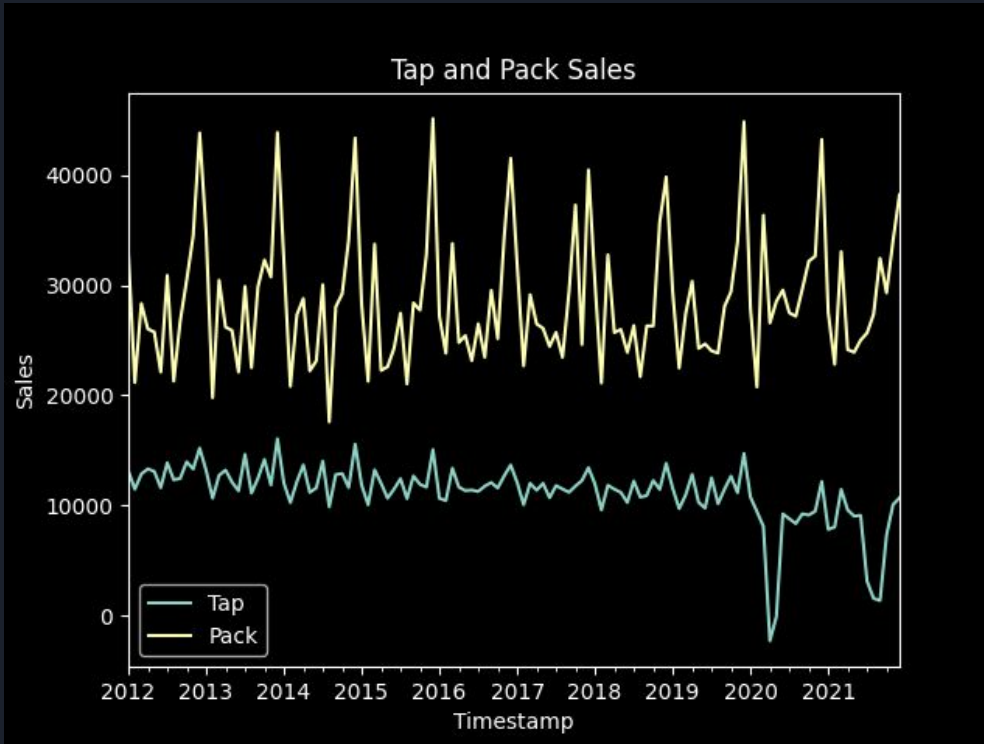




Forecasting Challenge

Forecasting Tap and Pack Sales

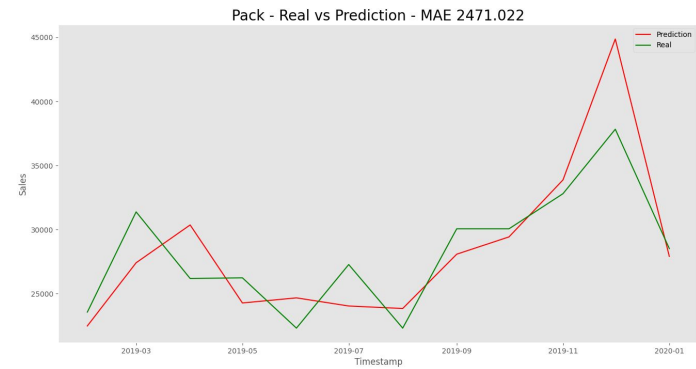
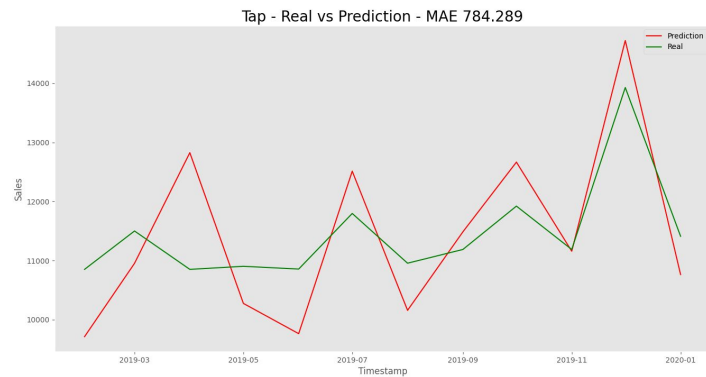
Current Trend



Key points:

- There is a repeating trend
- Covid period has caused the trend to deviate
- Tap has deviated significantly
- Pack has slightly more sales during covid

Model Test Results

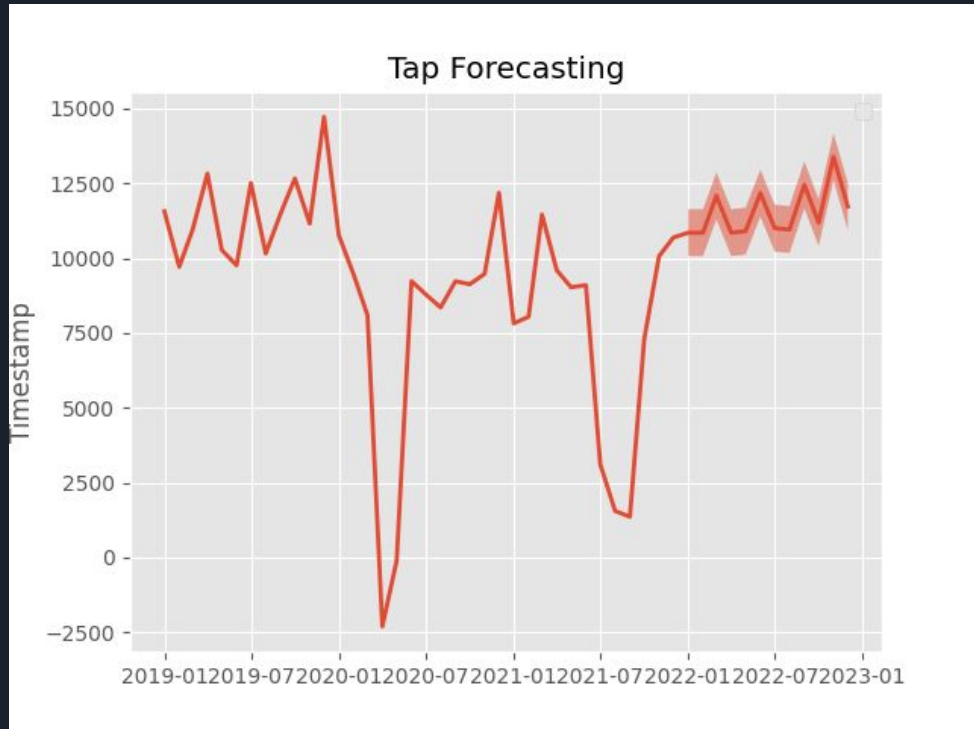




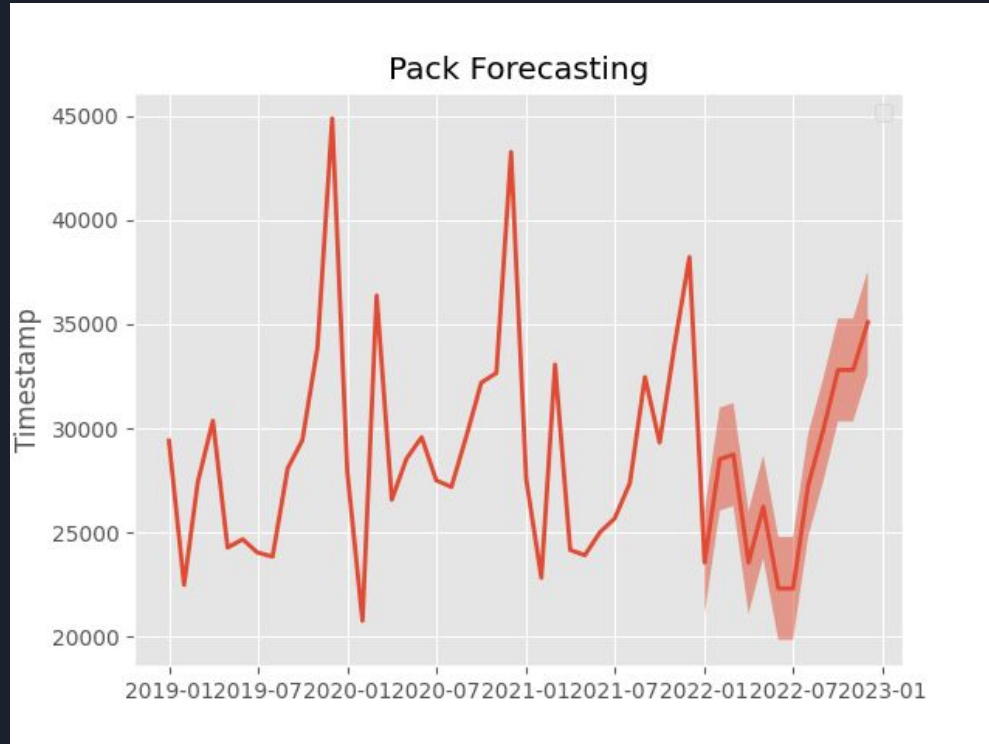
Model Notes

- The model used is LightGBM
- It is a decision tree modelling solution
- Model results look to follow the trend closely
- Tap has a ~7% error margin
- Pack has a ~8% error margin
- Features used:
 - Month
 - Season
 - Previous year-month value

Tap Forecasted Trend



Pack Forecasted Trend





Forecast Notes

- The shaded red area is the mean absolute error to account for possible upper and lower limits
- Covid period was excluded
- With this data we can forecast future sales and also match against marketing campaigns
- We can see there Tap is trending downwards showing more resources should be focused in that area improve sales
- Pack is trending upwards and it should be investigated why this is the case
- The Tap forecast can be improved, and this is outlined in the “Modelling Improvements” slide



Data Analysis

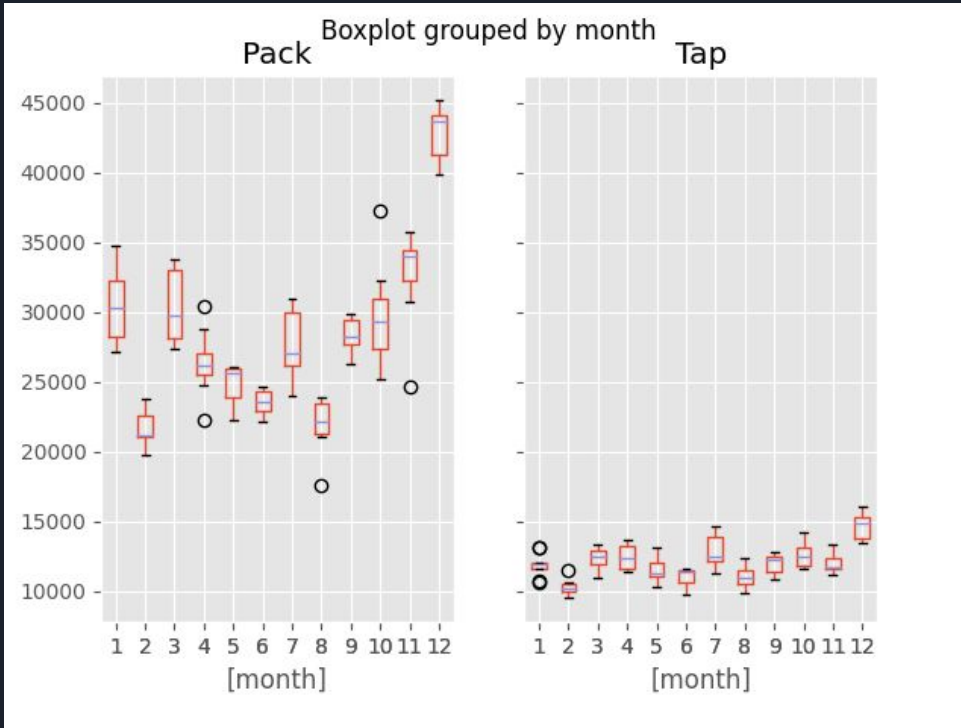
Below are the tools to analyse the data to find features for modelling.

Some analysis concepts:

- Group data by month
- Boxplot analysis
- Autocorrelation analysis
- Feature analysis
- Histogram analysis
- Mean trend analysis

Only Boxplot and Autocorrelation are shown below for simplicity.

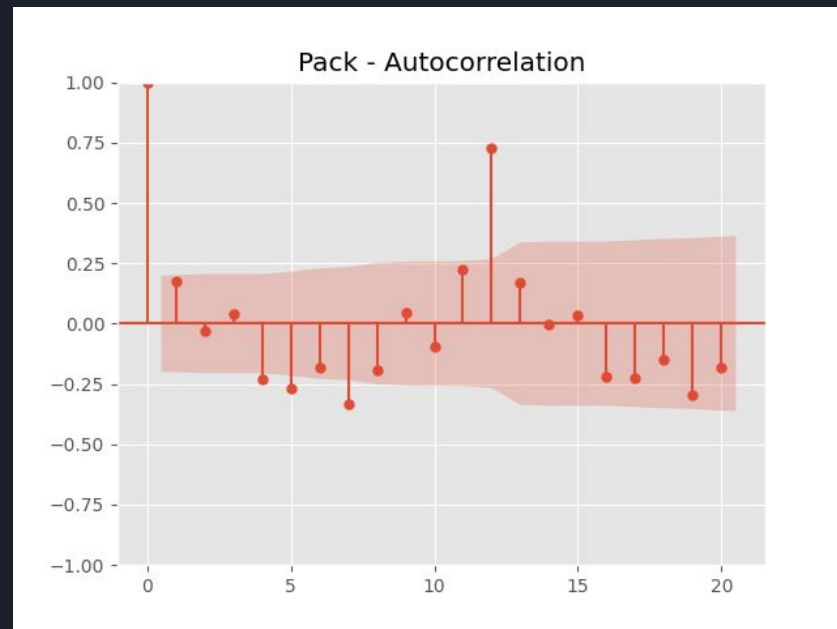
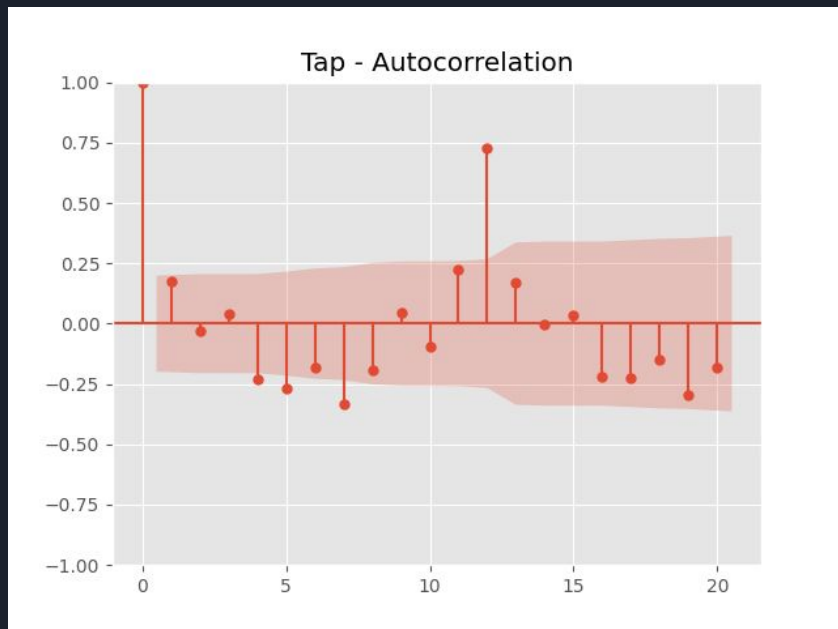
Boxplot



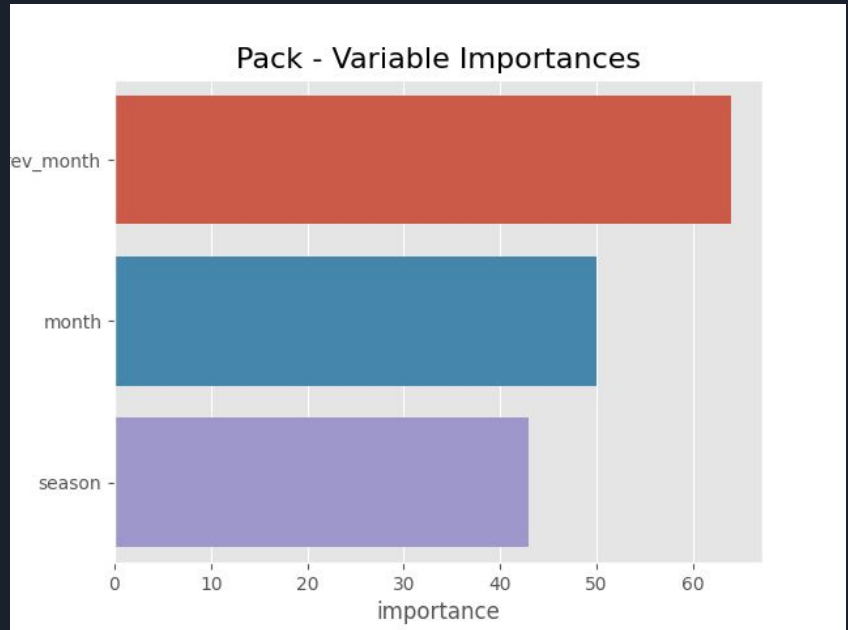
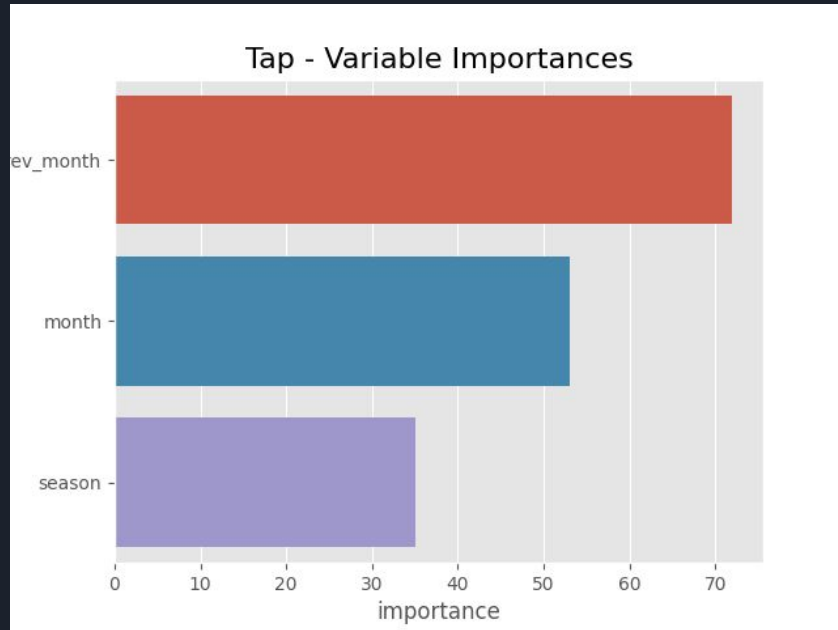
Key points:

- Pack sales fluctuate greatly depending on the month
- Pack has more outliers
- Tap sales tends to follow the same amount
-

Autocorrelation



Feature importance





Modelling Improvements

- Covid period should not be removed but should be cleaned to remove outliers and use averages to replace them
- Different models should be trialed to check different results and accuracy
- A hyperparameter optimisation should be used to find the best parameters
- An ensemble model system should be trialed for better results
- More feature engineering and discovery should be done
- More data sources for features should be used
- A value weighting should be applied to Tap to adjust for the lower trend in sales.
- Weighting should be applied to the prediction to account for upward or downward trends