vector_auto_regression_participant1_results

November 14, 2020

1 VAR Results Participant1

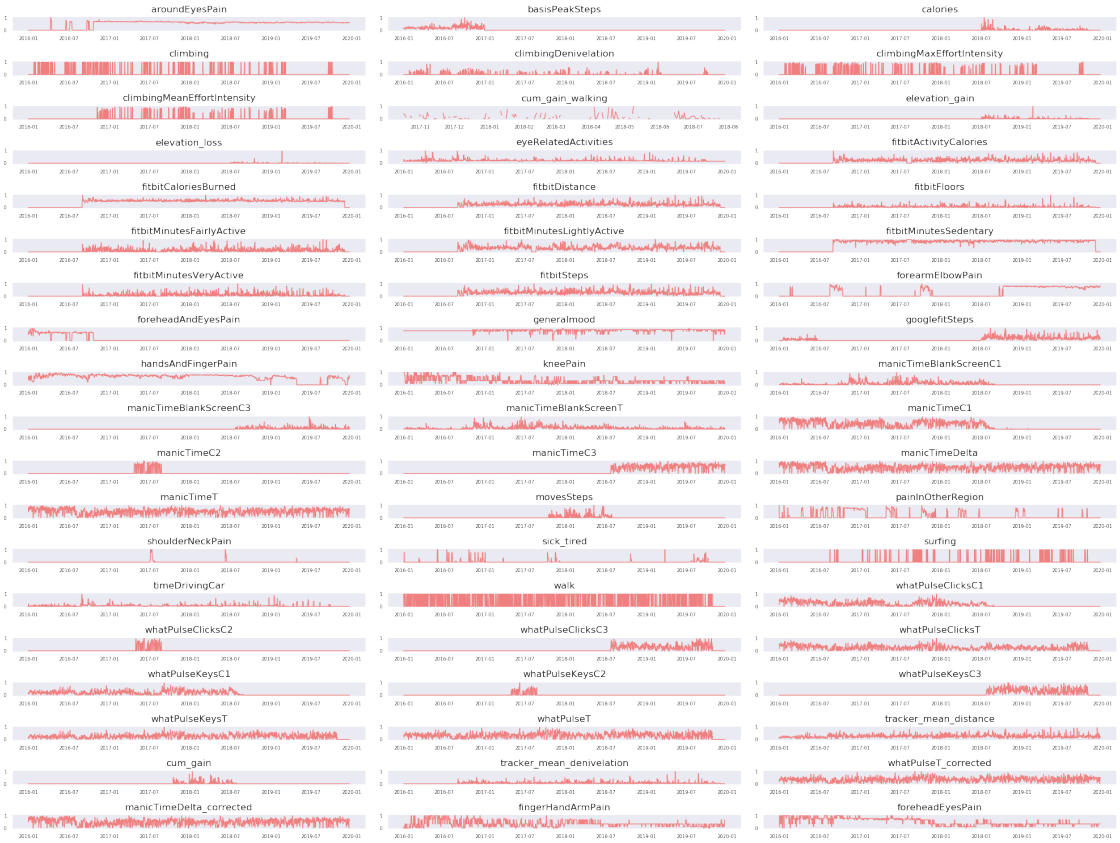
Vector Autoregression (VAR) is a multivariate forecasting algorithm that is used when two or more time series influence each other. It is considered as an Autoregressive model because, each variable (time series) is modeled as a function of the past values. The VAR model includes the variable's lagged (past) values, the lagged values of the other variables in the model, and an error term.

2 Preprocessing

The original shape of the dataframe was 1452 rows with 70 variables. After the removal of variables with missing data (columns where the number of missing data $>$ 1400 rows) and those with no variation, the number of variables left was 57. The data was normalized so that all values were between 0 and 1.

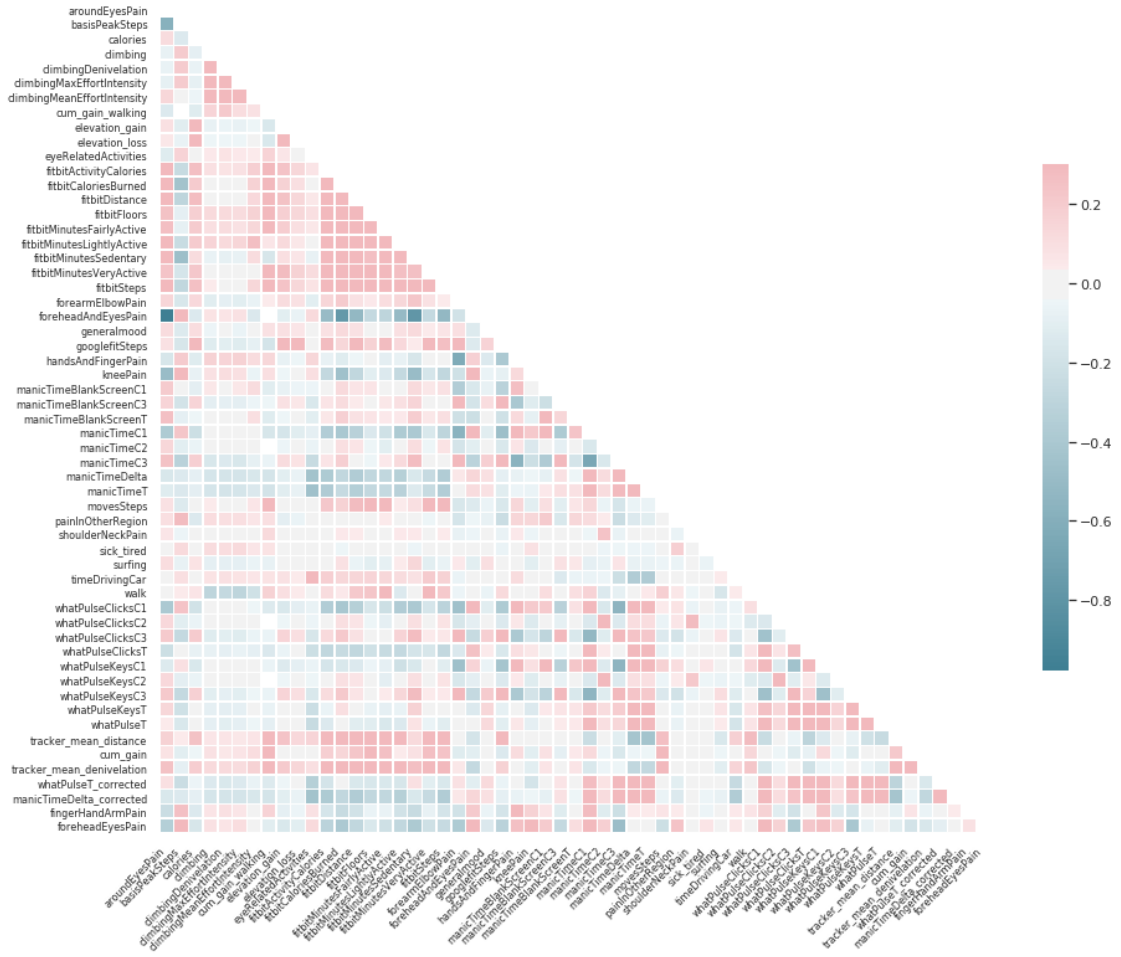
2.1 Visualization

We plot the variables to detect trends, seasonality and possible correlations.



2.2 Correlation

Correlation plot for all variables



2.3 VIF Analysis

A variance inflation factor (VIF) detects multicollinearity in regression analysis. Multicollinearity is when there's correlation between predictors (i.e. independent variables) in a model and can distort the results of the model. The VIF estimates how much the variance of a regression coefficient is inflated due to multicollinearity in the model.

Interpretation of the VIF value: - 1 = not correlated - Between 1 and 5 = moderately correlated - Greater than 5 = highly correlated

All variables with VIF factor greater than 5 were removed. After the VIF analysis, 27 variables remain:

```
['basisPeakSteps', 'calories', 'climbingDenivelation', 'climbingMeanEffortIntensity',
'elevation_loss', 'fitbitMinutesFairlyActive', 'fitbitMinutesLightlyActive',
'fitbitMinutesVeryActive', 'forearmElbowPain', 'foreheadAndEyesPain',
'googlefitSteps', 'kneePain', 'manicTimeBlankScreenC1', 'manicTimeBlankScreenC3',
'movesSteps', 'painInOtherRegion', 'shoulderNeckPain', 'sick_tired',
'surfing', 'timeDrivingCar', 'walk', 'whatPulseClicksC3', 'whatPulseKeysC2',
```

```
'whatPulseKeysT', 'cum_gain', 'tracker_mean_denivelation', 'fingerHandArmPain']
```

3 VAR analysis

3.1 Stationary assumption

VAR assumes that the data is stationary: i.e. their means and variances are constant over time and do not show any trending behaviour. An ADFuller test was performed to test for stationarity. The test showed that some series did not meet the stationarity assumption. First order differencing was applied in order to make all the series stationary (differencing is performed by subtracting the previous observation from the current observation).

3.2 Select the Order (P) of VAR model

To select the right order of the VAR model, we iteratively fit increasing orders of VAR model and pick the order that gives a model with least AIC (Akaike Information Critera). We continue with a model with lag order = 6 as that has the lowest AIC value.

We continue with lag order = 6

3.3 VAR Results

See the notebook on github for details (output too long). Results for the various areas of pain are as follows.

- Kneepain: climbingDenivelation (the number of meters climbed while rock climbing) has a negative impact on kneepain, tracker_mean_denivelation has a positive impact on kneepain (the number of meters climbed while walking). This could indicate that on days with rock climbing the participant involved less in other activities that put more stress on the knees. Walking on the days before (see L2.walk, L3.walk (not significant though), L4.walk) is likely to decrease kneepain.
- ShoulderNeckPain: ClimbingMeanEffortIntensity has a negative effect on ShoulderNeckPain. This last up to six days before, with L1 to L6 all significant. In addition, L1.surfing, L1.timeDrivingCar, and L4.manicTimeBlankScreenC1 + L5.manicTimeBlankScreenC1 increases ShoulderNeckPain. It seems that spending time behind the computer and driving increases neck and shoulder pain. The following variables have a negative impact on neck and shoulder pain: L1.tracker_mean_denivelation, L2.fitbitMinutesLightlyActive and L5.fitbitMinutesLightlyActive. This may indicate that being slightly active in the days before may decrease ShoulderNeckPain.
- fingerHandArmPain: L1.whatsPulseKeysT is positive. Key pressures have effect on fingerHandArmPain. L5.kneePain is positive too. This finding is counterintuitive and might be a false positive.
- forearmElbowPain: L1.surfing increase forarmElbowPain. No other significant results found.
- foreheadAndEyespain, L2.painInOtherRegion, L5.painInOtherRegion and L5.fingerHandArmPain are all positive and significant. foreheadAndEyespain seems to be positively related to pain in other regions.