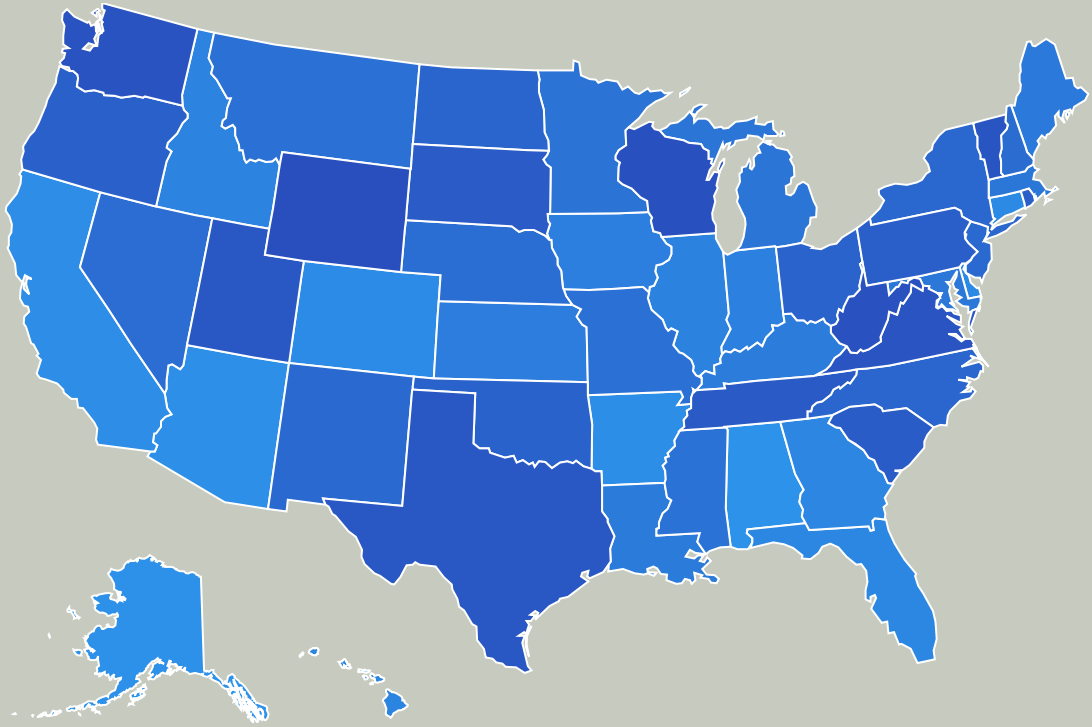
Project Proposal for Data Incubator

Weekly initial claims for unemployment in the US and google search data



https://mfy.netlify.com

Project Proposal

Project Definition:

Title of this project is "Weekly initial claims for unemployment in the US and google search data.

Objective and objective:
The objective is to predict the rate of weekly initial claims for unemployment by google search keywords. In this project, it was aimed to examine how google data search can be used to improve short term forecasts

Data:
The federal reserve economic data set was obtained from economic research division of Federal Reserve Bank of St Louis, Link: https://fred.stlouisfed.org. The data consist of the weekly initial claims for unemployment insurance in the US, as reported by the US Federal Reserve. For economic decisions based on these and similar numbers, it would help to have an early forecast of the current week's number as of the close of the week.

Methodology:
For this project I used Bayesian structural time series to fit the model.

Analysis:
Bayesian structural time series method was used to fit time series models. structural time series model was used to show how Google search data can be used to improve short term forecasts of economic time series. Structural time series models are useful because they are flexible and modular. For economic decisions based on these and similar numbers, it would help to have an early forecast of the current week's number as of the close of the week.

The data was divided in to two parts (train, test). In the first model (Model 1), I tried to fit a bsts model with just the trend and seasonal components on the weekly claims without other components. Subsequently, I used to predict method to predict future the next 52 time points. After that, test data was used for validation of the prediction.

Finally, regression components (michigan unemployment, military bah, pennsylvania unemployment, unemployment offices, unemployment

Model 1

Model 1

Codes

```
> library(readxl)
> iclaims <- read_excel("C:/Users/MFY/Desktop/data inc
project/Bayesian Structured Time Series Data.xlsx",
+     sheet = "Train")
> View(iclaims)
> library("bsts")
> data(iclaims)
> ss <- AddLocalLinearTrend(list(), iclaims$ICNSA)
> ss <- AddSeasonal(ss, iclaims$ICNSA, nseasons = 52)
> model1 <- bsts(iclaims$ICNSA,
+                state.specification = ss,
+                niter = 1000)
> plot(model1)
> plot(model1, "components")
```

01

Figure 1: Distribution of train data

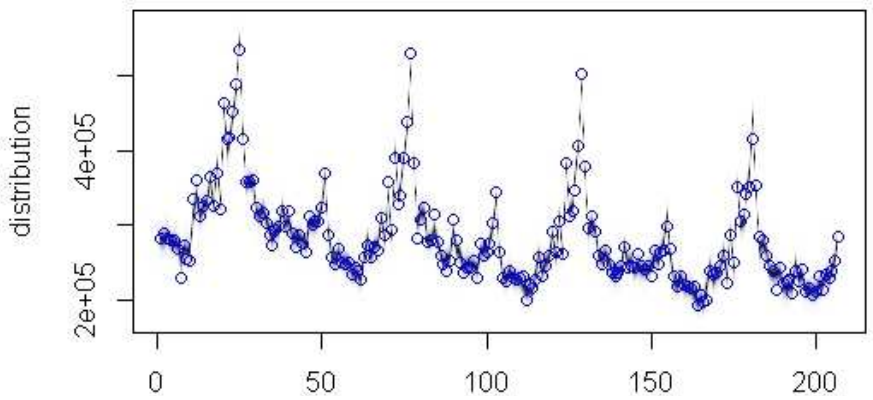
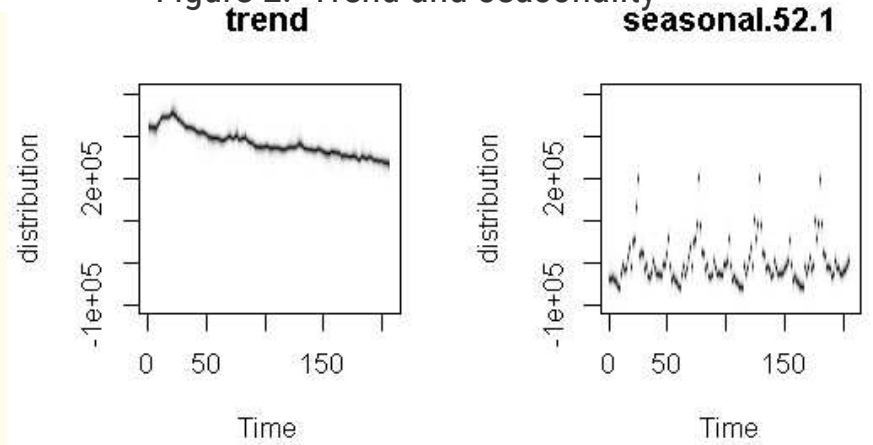


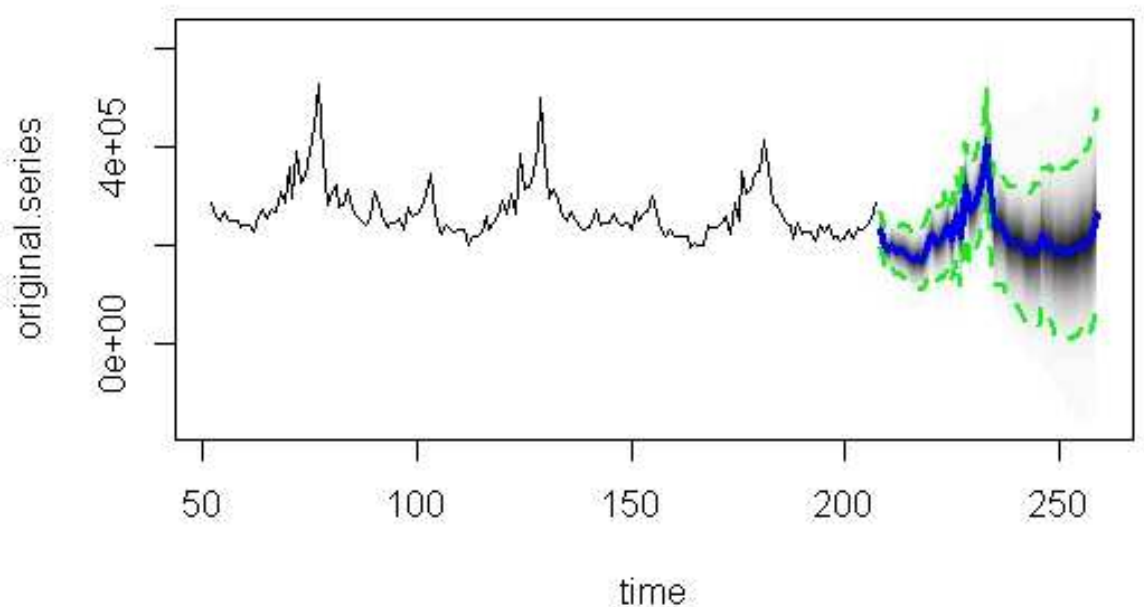
Figure 2: Trend and seasonality



Prediction

```
> pred1 <- predict(model1, horizon = 52)
> plot(pred1, plot.original = 156)
```

Figure 3: Predictive distribution for the next 52 weeks of initial claims.

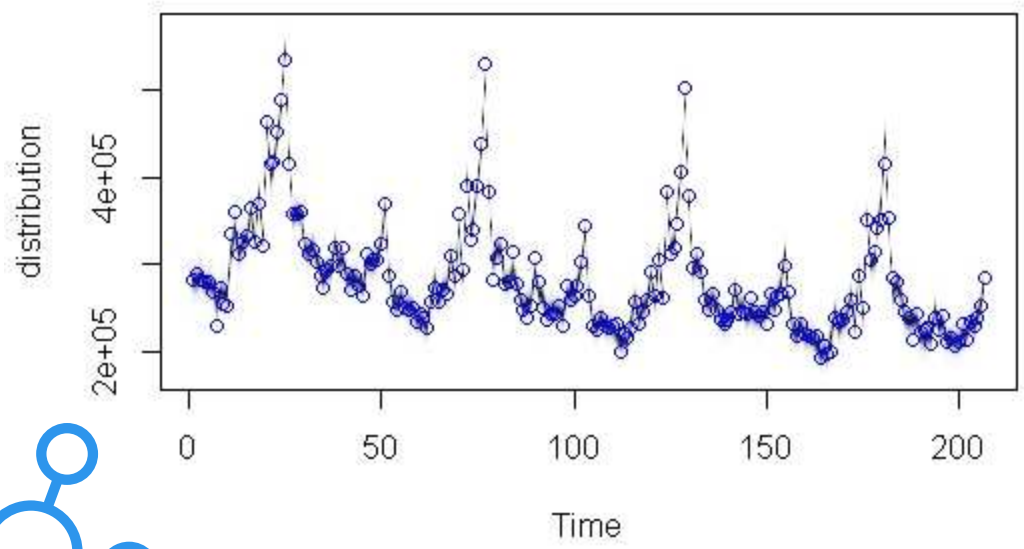


Model 2

Model 2 Codes

```
library(bsts)
> data(iclaims)
> ss <- AddLocalLinearTrend(list(), iclaims$ICNSA)
> ss <- AddSeasonal(ss, iclaims$ICNSA, nseasons = 52)
> model2 <- bsts(ICNSA ~ .,
+                state.specification = ss,
+                niter = 1000,
+                data = iclaims)
plot(model2)
```

Figure 4: Distribution of the data with regression components

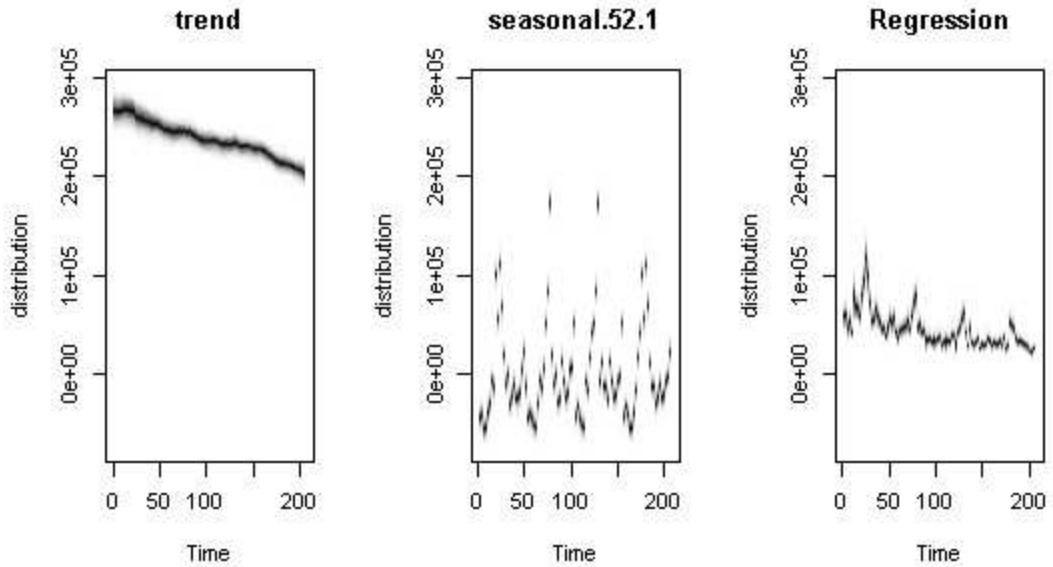


Google search data and regression

In this part, regression component will be add to the model described, so that we can use Google search data to improve the forecast

```
plot(model2,"comp")
```

Figure 5: Contribution of each state component to the initial claims data, assuming a regression component with default prior. Compare to Figure 2

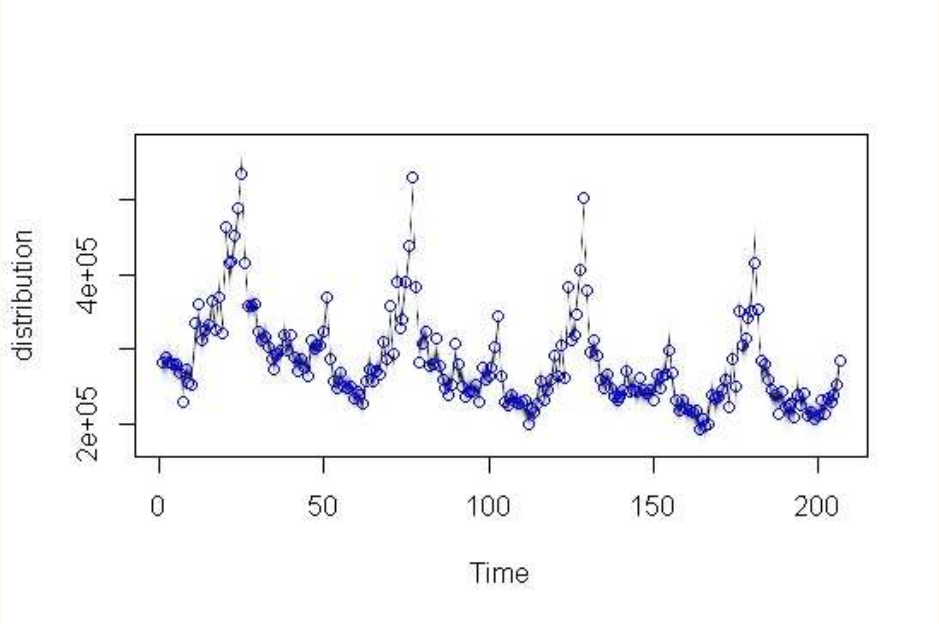


MODEL 2

Test data

```
library(readxl)
> iclaimstest <- read_excel("C:/Users/MFY/Desktop/data
inc project/Bayesian Structured Time Series Data.xlsx",
+     sheet = "Test")
> View(iclaimstest)
> newdata<-iclaimstest
> pred2 <- predict(bsts.model2,
+                  newdata=newdata)
pred2 <- predict(model2,
+                  newdata=newdata)
> plot(model2)
```

Figure 2: distribution of the test data



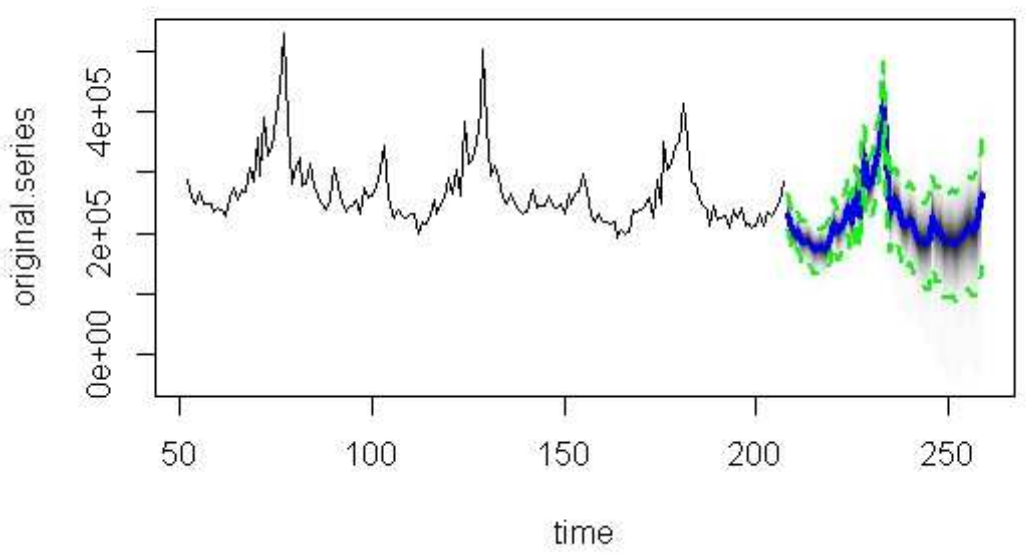
Model 2

Model 2 Prediction

```
> plot(pred2,
+      plot.original=156)
```

Figure 7 reveals the predictive distribution of the initial claims with regression coefficients.

Figure 7: Predictive distribution for the next 52 weeks of initial claims with regression coefficient.



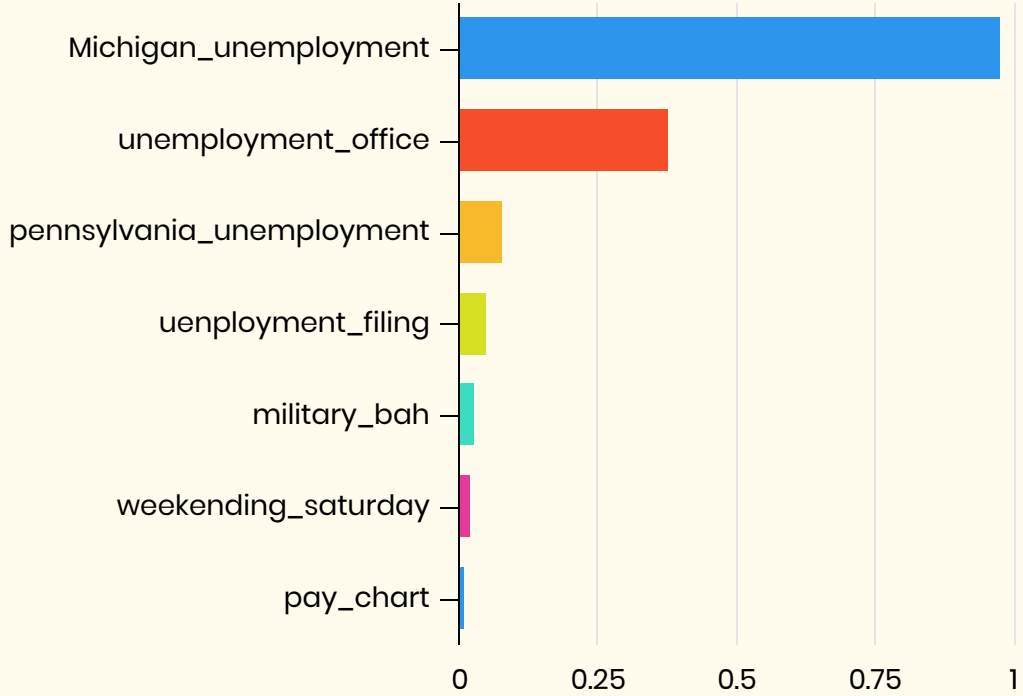
Model 2

Coefficients

- In this section, the regression coefficients visualized

```
> plot(model2, "coef")
```

Figure 8: Inclusion probabilities for predictors in the "initial claims"



Model 1
Model 2

Conclusion

The comparison of the two-model with the actual values revealed that, there is an improvement in the model accuracy.

For the model1 the mean absolute percentage errors (MAPE) scores were found 9.5%. Subsequently, MAPE2 scores for model2 was 5.7%.

As a result, there was 3.8% difference between the two models (model1 and model2) This result indicates 40% percent improvement in model2.

Figure 9: Initial claims, model1 and model2 comparison.