

NZ employment March 2025

Finn de Lange

2025-06-29

```
data_employment <- read_csv("employment.csv") %>%
  mutate(
    year = floor(Period),
    month = round((Period - year) * 100),
    date = yearmonth(ymd(paste(year, month, "01", sep="-")))
  ) %>%
  select(-c(Series_reference, year, month, Series_title_4, Series_title_5, Period)) %>%
  filter(Series_title_3 == "Actual" & Series_title_1 == "Filled jobs")
```

```
## Rows: 24511 Columns: 14
## -- Column specification -----
## Delimiter: ","
## chr (9): Series_reference, Suppressed, STATUS, UNITS, Subject, Group, Series...
## dbl (3): Period, Data_value, Magnitude
## lgl (2): Series_title_4, Series_title_5
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
head(data_employment)
```

```
## # A tibble: 6 x 11
##   Data_value Suppressed STATUS UNITS Magnitude Subject Group Series_title_1
##   <dbl> <chr>    <chr> <chr>    <dbl> <chr>    <chr> <chr>
## 1    80078 <NA>        F      Number      0 Business D~ Indu~ Filled jobs
## 2    78324 <NA>        F      Number      0 Business D~ Indu~ Filled jobs
## 3    85850 <NA>        F      Number      0 Business D~ Indu~ Filled jobs
## 4    90743 <NA>        F      Number      0 Business D~ Indu~ Filled jobs
## 5    81780 <NA>        F      Number      0 Business D~ Indu~ Filled jobs
## 6    79261 <NA>        F      Number      0 Business D~ Indu~ Filled jobs
## # i 3 more variables: Series_title_2 <chr>, Series_title_3 <chr>, date <mth>
```

```
data_employment %>%
  count(date, sort = TRUE) %>%
  filter(n > 1)
```

```
## # A tibble: 56 x 2
##   date      n
##   <mth> <int>
## 1 2012 Sep  117
```

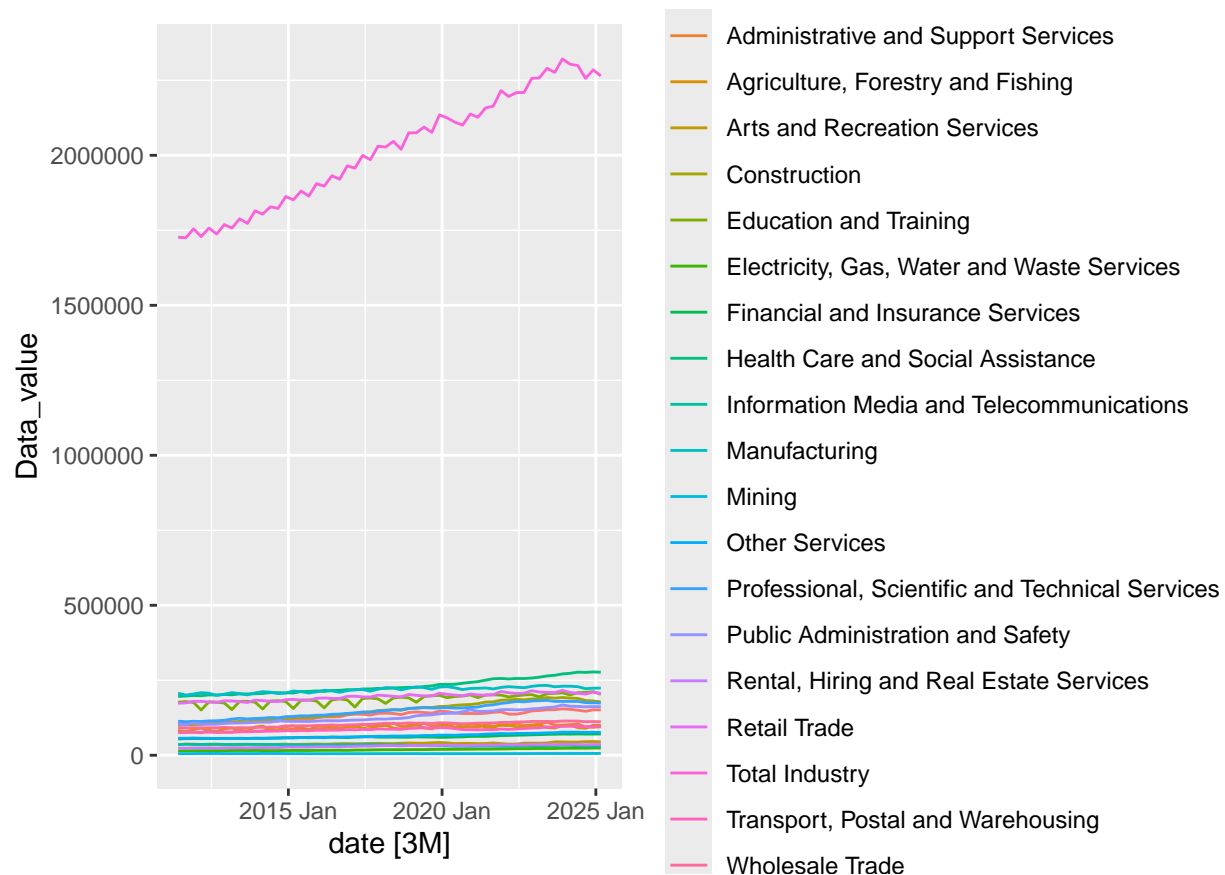
```
## 2 2011 Jun 116
## 3 2011 Sep 116
## 4 2011 Dec 116
## 5 2012 Mar 116
## 6 2012 Jun 116
## 7 2012 Dec 116
## 8 2013 Mar 116
## 9 2013 Jun 116
## 10 2013 Sep 116
## # i 46 more rows
```

```
selected_industries = unique(data_employment$Series_title_2)[1:20]

data_employment <- data_employment %>%
  filter(Series_title_2 %in% selected_industries) %>%
  as_tsibble(index = date, key = Series_title_2)
```

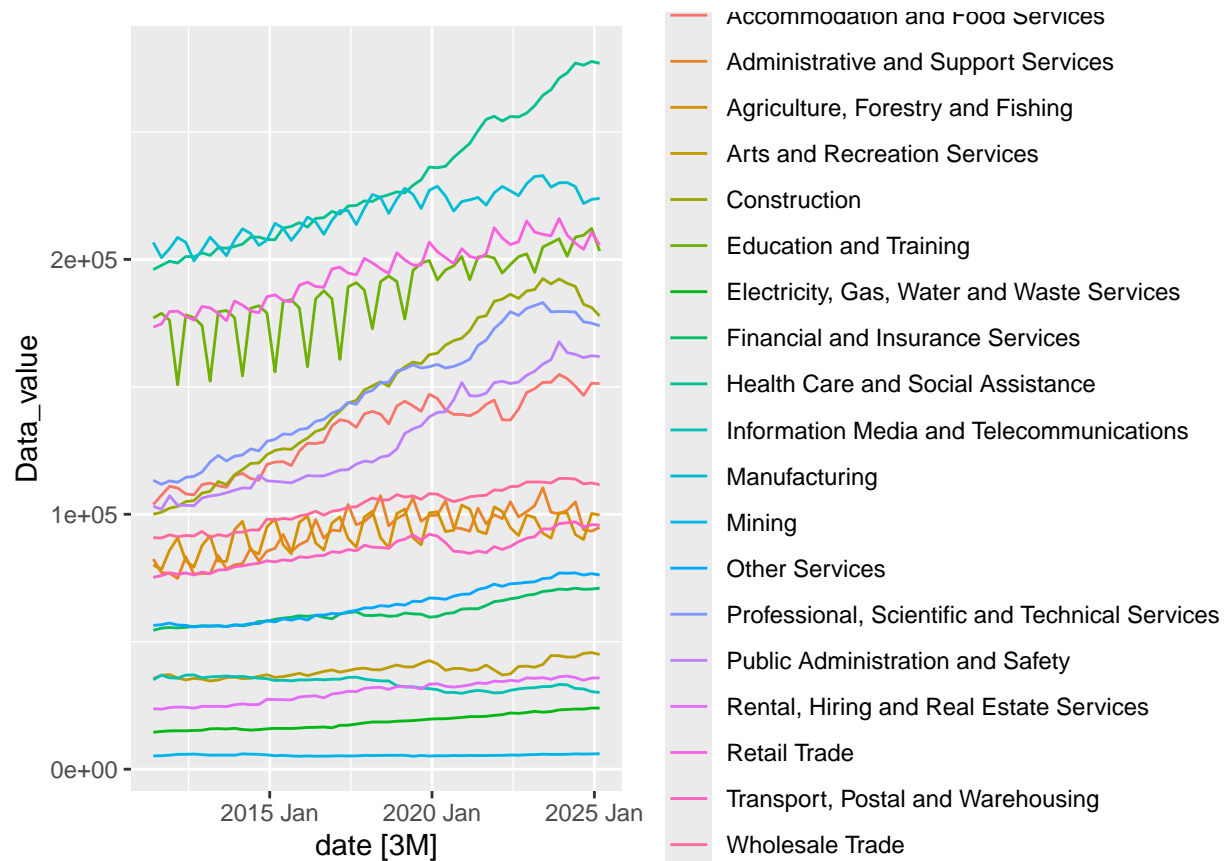
I want to say that I do not like how much I had to process that. While I appreciate StatsNZ for including so much information in a single file, this was an obscene amount of pre-processing I had to do.

```
data_employment %>%
  autoplot(Data_value)
```



We can see what is presumably total industry disrupting our ability to adequately view the other industries, so let us remove it for now.

```
data_employment %>%
  filter(Series_title_2 != "Total Industry") %>%
  autoplot(Data_value)
```

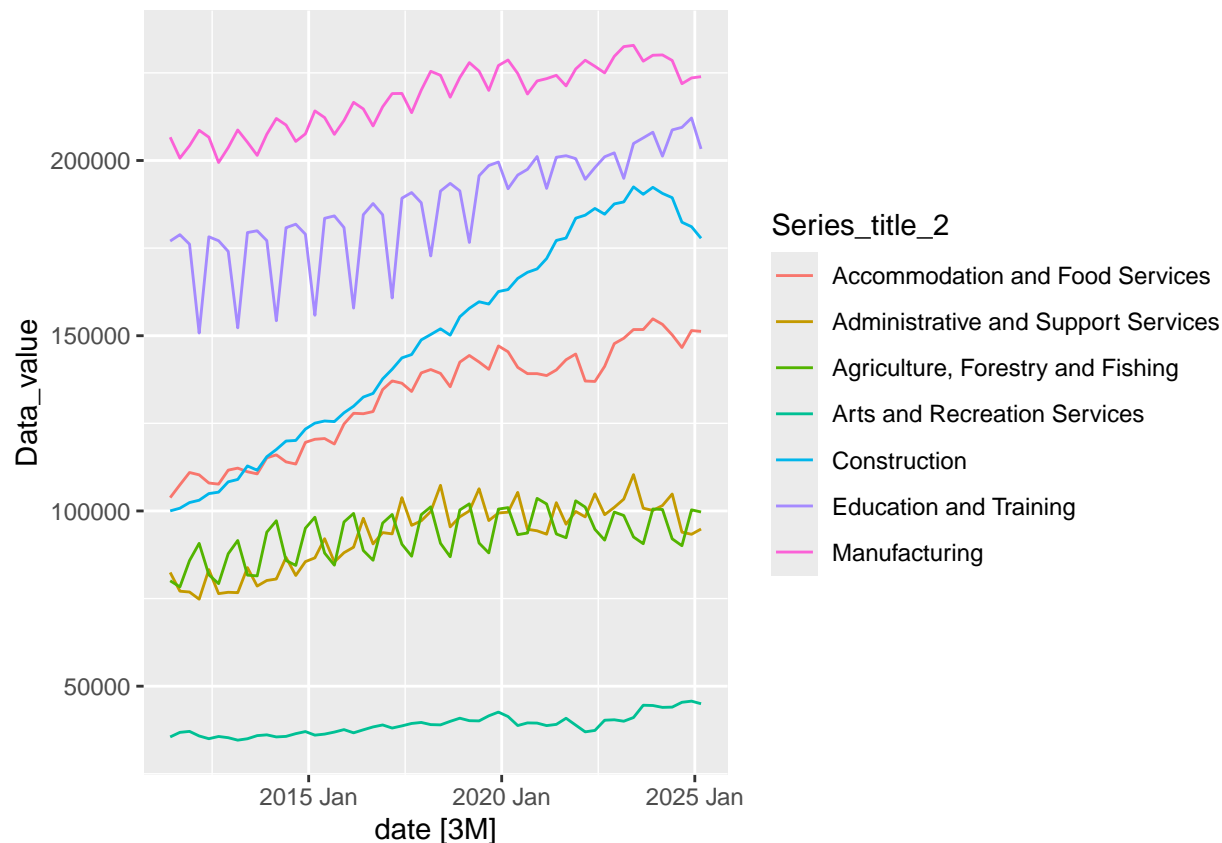


We can observe some much more interesting behaviours here. In particular, we can see that some industries used to show highly seasonal behaviour, before covid “tamed” these, while other industries removed mostly the same, at least two showed no change, and others peaked during 2023 before declining.

I want to pull the more interesting ones for further inspection.

```
interesting_ones <- c("Education and Training", "Accommodation and Food Services", "Administrative and Support Services")

data_employment %>%
  filter(Series_title_2 %in% interesting_ones) %>%
  autoplot(Data_value)
```



I wanted to zero in on these seven industries, for a few keys reasons. Let's discuss each of them:

Accommodation and Food Services: Prior to COVID, this sector seems to be one of the faster growing from this list, which is not surprising given that New Zealand was preparing to host the America's cup, but has also long been considered a major tourism destination, with consistent ad campaigns about travelling to New Zealand, and the flexibility of Airbnb combined with the global image of New Zealand food and drink such as Marlborough white wines and Central Otago reds likely drew more customers to these industries, resulting in greater levels of hiring. However, this appears to be one of the hardest hit industries of those chosen, with an immediate, noticeable decline that hits it's lowest in early 2022. However, after this we do see a fairly consistent growth, which would make sense as 2022 was when the last of the COVID restrictions were lifted domestically, encouraging both more domestic and international tourism. We can also see a clear seasonal pattern in this industry, which, as a former employee myself, is likely due to the tourists seasons requiring more staff on short-term contracts. This could be an interesting industry to model, so it shows some clear seasonal behaviour, but it has not yet returned, and the trend appears to have changed.

Administrative and Support Services: Prior to COVID, we observe a highly seasonal behaviour. This clearly increases until about 2020, when we see a clear dip from the time of the first lockdown, but this pattern fairly quickly returns to almost the level observed prior, although does show a significant decline at the end of 2024. There are a number of possible reasons for this, such as the hiring of more staff in the lead up to the election, a post-COVID economic rebound with more spending driving a temporary need for more staff, or a range of other reasons. This series could benefit from further modelling to see what underlying behaviours are present.

Agriculture, Forestry and Fishing: The largest corner stone of New Zealand exports are food or forestry related: Dairy Products, Meat and Edible Offal, Edible Fruits, Wood and Articles of Wood, Cereal, Flour, and many more not listed here. It is hardly surprising therefore that this industry continued to hire and not lay off staff in their usual behaviour. However, it is of interest that this was not a problem for 2023, as that year there was a well documented and publicised issue of orchards struggling to hire enough staff to harvest

fruit. This series is fairly constant in behaviour, so we could model it to understand the timings of hiring and firing, but this would not be of particular use.

Arts and Recreation Services: New Zealand is a hotspot for the international film industry, with a wide range of movies and artistic endeavours taking place within our borders. The film industry is heavily subsidised by the government as well, giving the industry a certain level of stability while continuing to attract international productions. Recreational Services includes tourism, another important player in the domestic economy, that should struggle without the large tourist population. Arts is notoriously unstable industry, and with the recreational services suffering from a lack of tourists both internationally and domestically I expected to see massive layoffs during this time, as well as a form of boom/bust cycle overall. However, the Arts and Recreation Services is almost as dull as the Agriculture, Forestry and Fishing industry. We see a dip towards the end of the COVID years, which comes after a very slowly increasing trend with a minor seasonal component, before a sizeable increase again from the end of 2022. The government subsidies of the arts sector were present during COVID years, and the tourism industry received substantial government support during this time, so it does make sense that as an industry layoffs were not as present as expected. What is more interesting is the number of jobs filled declined at the end of 2022. I myself know of several in the tourism industry who took the time with subsidised income to retrain into other sectors, but it is unclear what exactly drove this. The sudden increase after this time is likely due to the sudden increase of tourism internationally. Despite being fairly dull, this may be one of the more interesting sectors to investigate further and model.

Construction: Construction does not appear to have greatly changed much from the effects of COVID. Prior to COVID, many urban centres across New Zealand were having construction booms, with Auckland alone working on two new ultra-tall skyscrapers, a new shopping centre on the waterfront, new convention centre, new rail connection, and a host of other projects. I would expect this sector to show some cyclicity, likely in line with the business cycle, but it is not clear on this scale. Additionally, the industry seems to hit its peak in employment in late 2023. I cannot find any especially strong indications as to why this is, especially as it was at this time that interest rates began to decline, and the new elected government signaled a heavy investment in certain infrastructure issues. There is likely a more obvious reason I am missing, but it is worth further investigating this industry, especially given the claims of the incumbent government about the success of the construction industry.

Education and Training: A massive employer nation wide, that shows a very large and clear seasonal pattern, that COVID almost completely tames (although it did appear to be declining somewhat already), lasting far into the future including to present day. This is easily the most unusual behaviour, and definitely worth modelling. It does not make sense for education to be highly seasonal, so I assume that the seasonal behaviours observed pre-COVID are from seasonal training for those working in seasonal industries such as Arts and Recreational Services, Agriculture, Forestry and Fishing, and Accommodation and Food Services. However, this is speculation. Absolutely worth modelling.

Manufacturing: New Zealand is an agricultural economy, and most of our food processing is included in the manufacturing sector. Additionally, the Tiwai point aluminium smelter supports a lot of workers, but we also produce a large number of therapeutics, such as those of Fisher and Paykel Healthcare, which would theoretically benefit from COVID activity, as their respirators were in hot demand. However, what we instead observe is a highly seasonal pattern, which dips slightly in the beginning of COVID but returns to a mostly normal pattern, although there appears to be a more unusual dip at the end of 2024, similar to the Administrative and Support Services field. This series could be worth investigating with intervention effects.

Next steps I'd like to investigate the education series the most, as the pattern seen is very unusual. As the seasonality seems to disappear from the series, I will first fit an ARIMA model. It seems likely that the COVID-19 pandemic drove this change in behaviour, so I will later try a dynamic model with ARIMA errors, but for now let us try a plain ARIMA model and see what is fit.

First, let us check if a Box-Cox transformation is needed.

```
data_education <- data_employment %>%
  filter(Series_title_2 == "Education and Training")
lambda <- data_education %>%
  features(Data_value, features = guerrero)

lambda
```

```
## # A tibble: 1 x 2
##   Series_title_2      lambda_guerrero
##   <chr>              <dbl>
## 1 Education and Training      2.00
```

The lambda value is very close to 2, suggesting that we could use a Box-Cox transformation, although we could also almost use a square transformation. For now, let us fit both models.

```
data_education <- data_education %>%
  mutate(
    transformed = box_cox(Data_value, lambda$lambda_guerrero)
  )

data_education <- data_education %>%
  mutate(
    year = as.integer(str_extract(date, "\\d{4}")),
    quarter = case_when(
      str_detect(date, "Mar") ~ 1,
      str_detect(date, "Jun") ~ 2,
      str_detect(date, "Sep") ~ 3,
      str_detect(date, "Dec") ~ 4,
      TRUE ~ NA_integer_
    ),
    yearquarter_str = paste(year, quarter, sep = " Q"),
    yq = yearquarter(yearquarter_str)
  )

data_education <- as_tsibble(data_education, index = yq, key = Series_title_2)

first_fit <- data_education %>%
  model(
    first_model = ARIMA(Data_value),
    second_model = ARIMA(box_cox(Data_value, lambda$lambda_guerrero))
  )

first_fit %>% glance() %>%
  arrange(AICc)
```

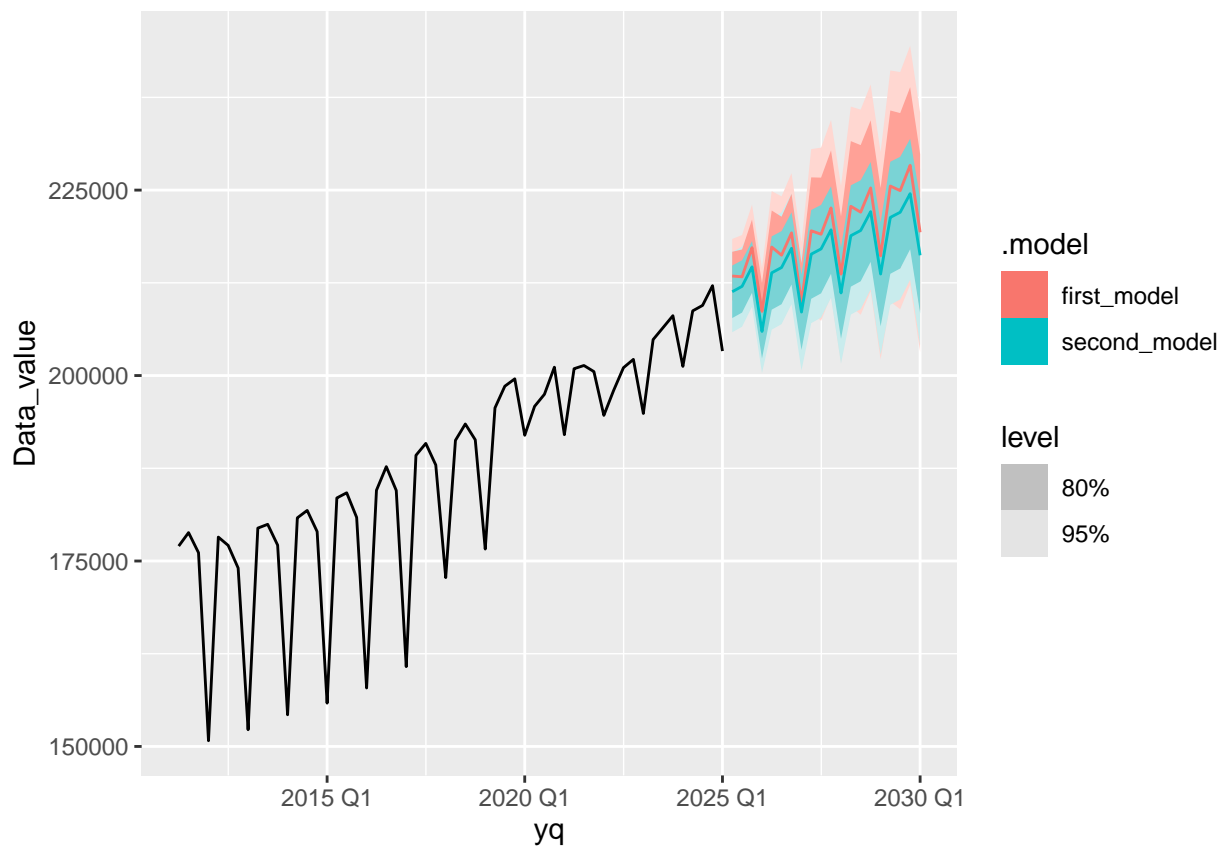
```
## # A tibble: 2 x 9
##   Series_title_2      .model sigma2 log_lik   AIC   AICc   BIC ar_roots ma_roots
##   <chr>            <chr>    <dbl>   <dbl> <dbl> <dbl> <dbl> <list>  <list>
## 1 Education and Trai~ first~ 6.49e 6   -481.  973.  975.  985. <cpl>    <cpl>
## 2 Education and Trai~ secon~ 3.39e17  -1123. 2250. 2250. 2253. <cpl>    <cpl>
```

```
first_fit %>%
  select(first_model) %>%
  report()
```

```
## Series: Data_value
## Model: ARIMA(1,0,0)(1,1,2)[4] w/ drift
##
## Coefficients:
##      ar1      sar1      sma1      sma2      constant
##      0.5202  -0.5265  0.4082  0.4858  2145.2180
## s.e.  0.1217   0.2683  0.2605  0.2016   608.4082
##
## sigma^2 estimated as 6494531:  log likelihood=-480.69
## AIC=973.39  AICc=975.26  BIC=985.1
```

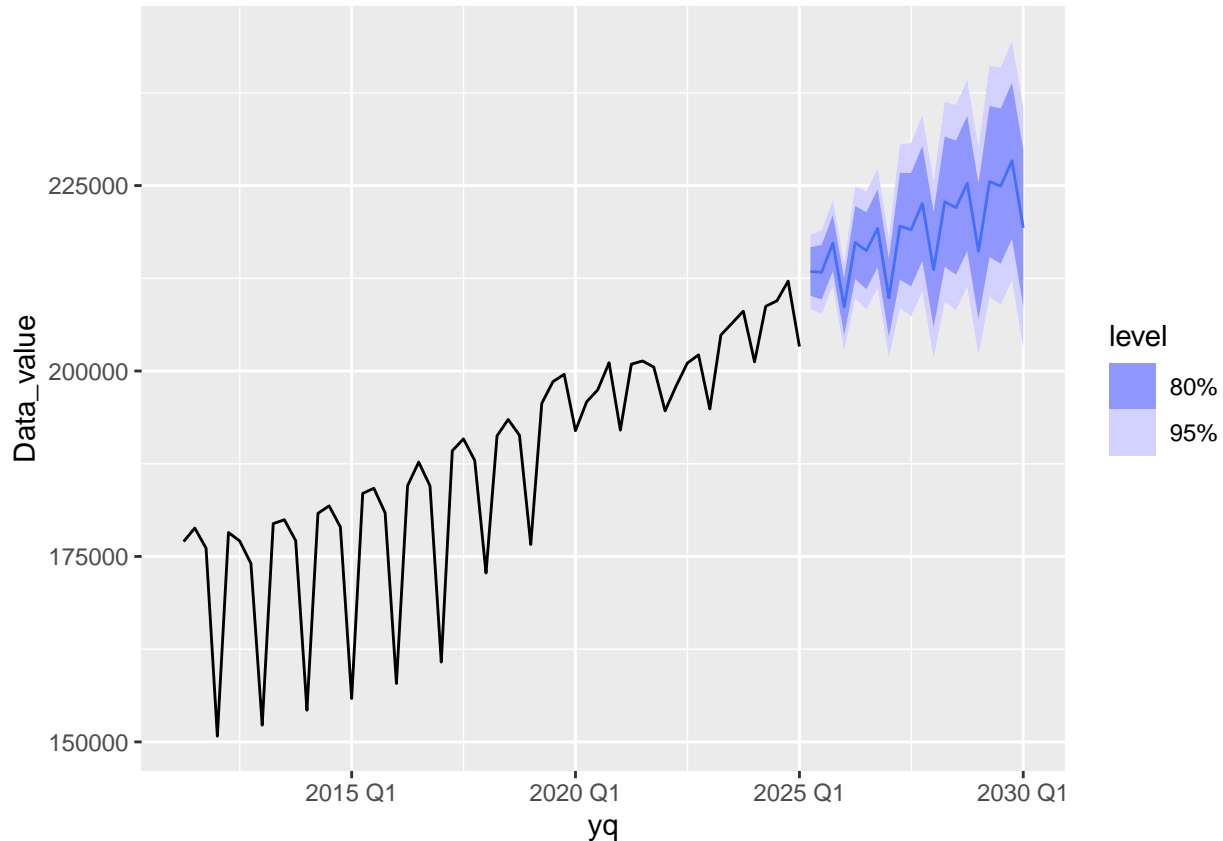
When comparing the models, we can see that the first model, fit on the original, untransformed data has a much, much lower AICc, indicating that it is better fitting.

```
first_fit %>%
  forecast(h = 20) %>%
  autoplot(data_education)
```



Removing the second model as the first was a better fit allows us to better inspect the forecasts.

```
first_fit %>%
  forecast(h = 20) %>%
  filter(.model == "first_model") %>%
  autoplot(data_education)
```



Clearly there is a slight disconnect from where the final observation is and the first forecasted value, but when we consider the behaviour of this forecast, it appears that it has mostly accounted for the weaker seasonal pattern seen more recently, but has returned to a fairly consistent additive seasonal trend and looks fairly consistent with what we have observed. Let us now try the same model, using cross-validation and compare the forecast accuracy.

We want to train on the majority of the data, so let us use 2021 Q2 as our cut-off.

```
training <- data_education %>%
  filter(yq <= yearquarter("2021-Q2"))

training %>%
  stretch_tsibble(.init = 20, .step = 1) %>%
  model(arima = ARIMA(Data_value ~ 1 + pdq(1, 0, 0) + PDQ(1, 1, 2))) %>%
  forecast(h=1) %>%
  accuracy(training) %>%
  select(MAPE, MASE)
```

```
## Warning: 4 errors (1 unique) encountered for arima
```

```
## [4] Not enough data to estimate a model with those options of P and Q. Consider allowing smaller values
```



```
## Warning: The future dataset is incomplete, incomplete out-of-sample data will be treated as missing.
## 1 observation is missing at 2021 Q3
```

```
## # A tibble: 1 x 2
##   MAPE MASE
##   <dbl> <dbl>
## 1  1.48 0.870
```

So our model has a very low forecast error, a great sign that this is a suitable model for making forecasts.

However, let us now consider the use of a dynamic regression models with ARIMA errors, as we may be able to better account for any effects of the COVID-19 pandemic. I also want to account for a change-of-government effect, so we will need to mutate some new variables.

The effects of COVID have been long lasting, but we can weight on the lagged predictors, so we will include a COVID effect from 2020 Q1 to 2022 Q4, and a change-of-government effect from 2023 Q4 onward. I would prefer to model the COVID effect with a decay or damping, but COVID was not consistent in its effect during the three years it was most active, so I have decided to simply use spike interventions for now, and will look at the use of lagged predictors to model the lasting damage it had later.

```
data_employment <- data_employment %>%
  mutate(
    year = as.integer(str_extract(date, "^\\d{4}")),
    quarter = case_when(
      str_detect(date, "Mar") ~ 1,
      str_detect(date, "Jun") ~ 2,
      str_detect(date, "Sep") ~ 3,
      str_detect(date, "Dec") ~ 4,
      TRUE ~ NA_integer_
    ),
    yearquarter_str = paste(year, quarter, sep = " Q"),
    yq = yearquarter(yearquarter_str)
  )

data_employment <- data_employment %>%
  mutate(
    covid_effect = case_when(
      yq >= yearquarter("2020 Q1") & yq <= yearquarter("2022 Q4") ~
        1,
      TRUE ~ 0
    ),
    governance_effect = if_else(yq >= yearquarter("2023 Q4"), 1, 0)
  )
```

Now that the mutations are complete, we can start thinking about our model for this. Aside from the governance and COVID effects, we shall leave the ARIMA aspect to the automatic setting to decide.

```
dynamic_models <- data_employment %>%
  model(
    arima = ARIMA(Data_value ~ governance_effect + covid_effect)
  )
```

Let us once again focus on the education model.

```
education_model <- dynamic_models %>%
  filter(Series_title_2 == "Education and Training")

education_model %>%
  report()
```

```
## Series: Data_value
## Model: LM w/ ARIMA(0,0,1)(1,1,2)[4] errors
##
## Coefficients:
##          ma1          sar1          sma1          sma2  governance_effect  covid_effect
##          0.4598      -0.4282      0.3654      0.7151             4553.246           18.2669
## s.e.      0.1126      0.2004      0.1632      0.2316             1668.101          1045.6095
##          intercept
##          2528.4240
## s.e.      654.3707
##
## sigma^2 estimated as 5798395:  log likelihood=-478.27
## AIC=972.53   AICc=975.88   BIC=988.14
```

What is very interesting here is that the AICc of this model is almost exactly the same as the first ARIMA model we fit on the untransformed data. I would like to perform cross-validation on this data, but software issues regarding rank deficiency have prevented me from doing so.