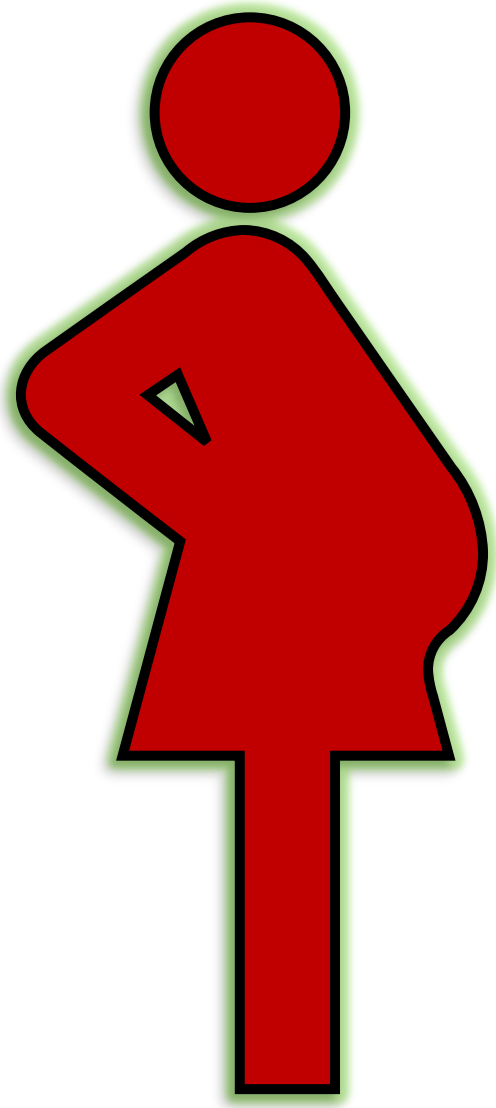
ABORTION STATISTICS IN NEW ZEALAND

(1980 – 2019)



R Shiny Application – Dashboard

Author: Raghuraman Srinivasan

1. Data

The data used in this application have been sourced from the website of *Stats NZ*.

Data Link:
https://www.stats.govt.nz/assets/Uploads/Abortion-statistics/Abortion-statistics-Year-ended-December-2019/Download-data/Abortion-statistics-year-ended-december-2019-csv.zip

App Link:
https://raghuraman-srinivasan.shinyapps.io/Abortion-Rates-NZ/

Format:
The data source contains 4 CSV files. Two of the CSV files contain data of Abortion Statistics for the years 1980 to 2019. These two are merged based on "year" value and the data frame is named "*AbortionCR*". The remaining two CSV files contain data of Abortion Statistics for the years 2000 to 2019 with further categorization based on the Age group of women. These two are merged based on "year" value and the data frame is named "*AbortionCRAge*".

Preparation:

- Datasets are loaded into the application console along with converting all the "N/A" values as Null values.
- Datasets are initially combined using an inner join.
- Duplicates found in the "*AbortionCRAge*" dataset is eliminated by combining both "*Period*" and "*Age_of_woman*" as the joint primary key.
- Data Summary is generated to identify the fields with incorrect data types. All such fields are defined with appropriate data types.
- New column "Date" is generated for performing Timeseries forecasting.
- The population of women overall and across different age groups in New Zealand over the years is extracted by using the Abortion rate and Number of Abortions by using the following formula

Women Population = (Number of Abortions * 1000) / (Abortion rate)

- Missing values are found in the *"Induced_abortions"* field for "*Period = 2019*" in the "*AbortionCRAge*" dataset. All such entities are imputed with appropriate values by creating a custom formula that uses corresponding values in *"AbortionCR"* dataset.

```
> head(AbortionCR)
  Period General_abortion_rate Induced_abortions Women_Population       Date
1   1980                  8.5              5945         699411.8 1980-12-31
2   1981                  9.6              6759         704062.5 1981-12-31
3   1982                  9.6              6903         719062.5 1982-12-31
4   1983                  9.7              7198         742061.9 1983-12-31
5   1984                  9.6              7275         757812.5 1984-12-31
6   1985                  9.3              7130         766666.7 1985-12-31
```

Image 1.1 Dataframe of "AbortionCR" data

```
> head(AbortionCRAge)
  Period Age_of_woman Abortion_rate Induced_abortions Women_Population       Date
1   2000        11-14           0.7                74         105714.3 2000-12-31
2   2000        15-19          23.1              3107         134502.2 2000-12-31
3   2000        20-24          35.7              4548         127395.0 2000-12-31
4   2000        25-29          24.5              3399         138734.7 2000-12-31
5   2000        30-34          16.4              2496         152195.1 2000-12-31
6   2000        35-39          11.4              1823         159912.3 2000-12-31
```

Image 1.2 Dataframe of "AbortionCRAge" data

2. Exploratory Analysis

Image 2.1 Quick Statistics on Different Parameters

From *Image 2.1* it can be observed that while the net abortion rate (averaged for a year) is 15.22, the standard deviation is 3.62 (approximately 24%). This is a considerably big difference.

Top 5 Years with Highest Abortion Rates

Period	General_abortion_rate	Induced_abortions
2003	20.80	18511
2004	20.20	18211
2007	20.10	18382
2002	19.90	17380
2008	19.70	17940

Image 2.2 Stats on Top 5 years with Highest Abortion Rates

From *Image 2.2* it can be seen that most years in the Highest Abortion Rates fall between 2002 – 2008. This conveys that in the current years the Abortion rate has dropped considerably which is a positive sign.

Top 5 Years with Lowest Abortion Rates

Period	General_abortion_rate	Induced_abortions
1980	8.50	5945
1985	9.30	7130
1981	9.60	6759
1982	9.60	6903
1984	9.60	7275

Image 2.3 Stats on Top 5 years with Lowest Abortion Rates

From *Image 2.3* it can be seen that most years in the Highest Abortion Rates fall between 1980 – 1985. This can convey two things. First, it needs to consider if all the cases back in the 1980s have been reported. If so second, research and identification on the positive factors that kept abortion rates in check could be useful.

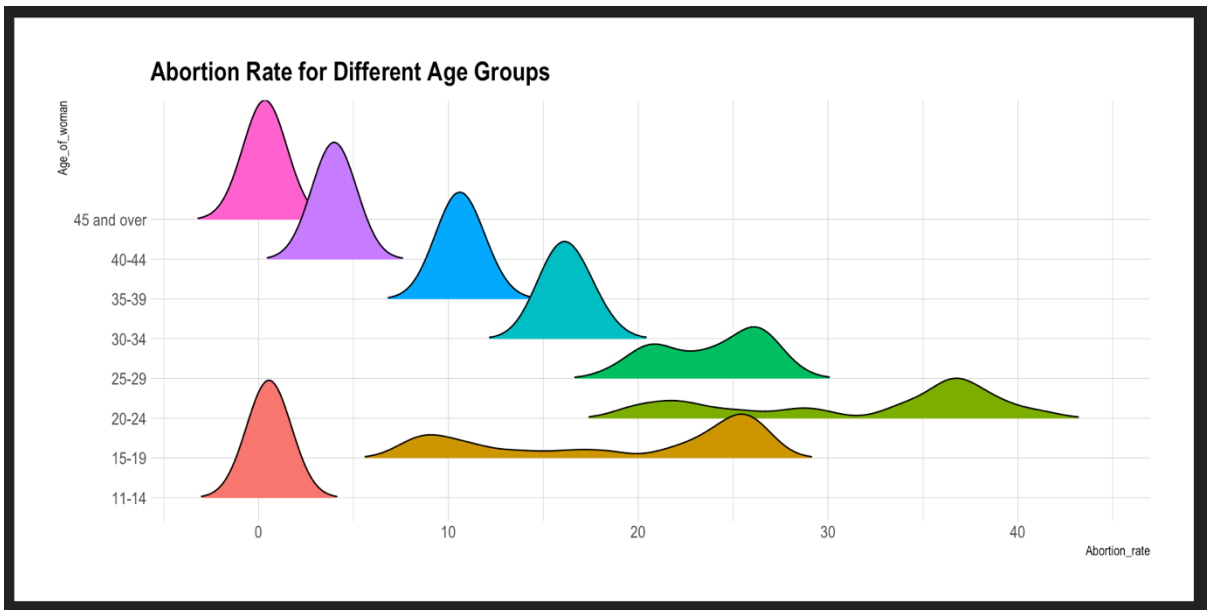


Image 2.4 Ridgeline plot on Abortion Rate vs Age Groups

To identify the age groups that have predominantly induced the function in the abortion rate, from *Image 2.4* it can be seen that drastic changes in abortion rate are predominant among women of age 20-24, 15-19 and 25-19. For the remaining age groups, the values seem to be consistent and lower than the vulnerable age groups.

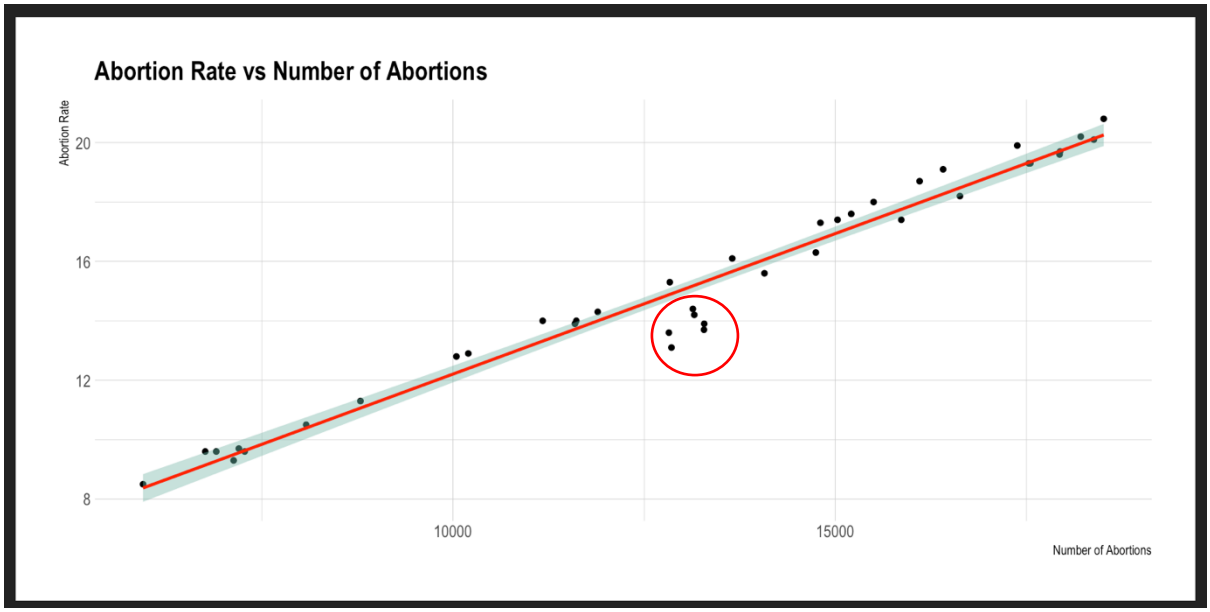


Image 2.5 Line Plot on Abortion Rate vs Number of Abortions

Further, from *Image 2.5* it can be seen that there is a linear relationship between the number of abortions and abortion rate. The data points circled red in the image indicates the years where the abortion rate has dropped considerably. A deeper investigation on the data points and from *Image 2.6* reveals the fact that those belong to the recent 6 years (2014 – 2019) which is a very positive sign.

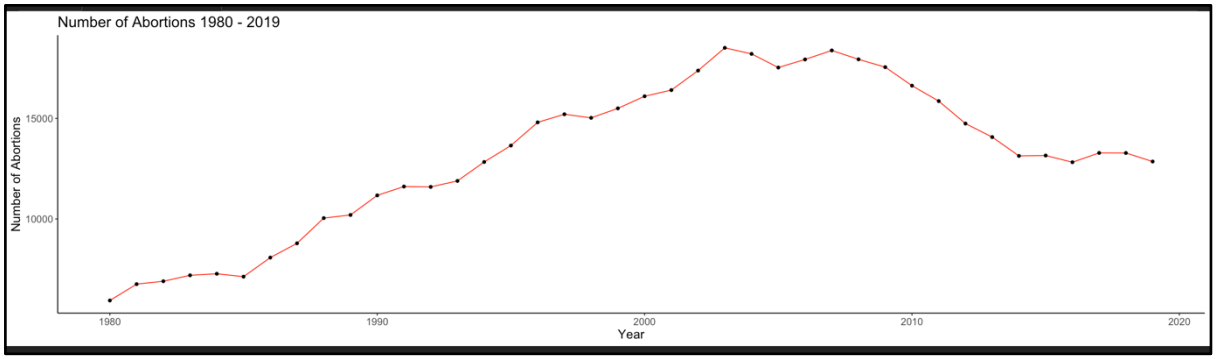


Image 2.6 Line Plot on Abortion Rate vs Number of Abortions

From *Image 2.7* it can be seen that the significant drop in the abortion rate among two of the vulnerable age groups 15-19 and 20-24 have induced the drop in overall abortion rate between the years 2014 and 2019.

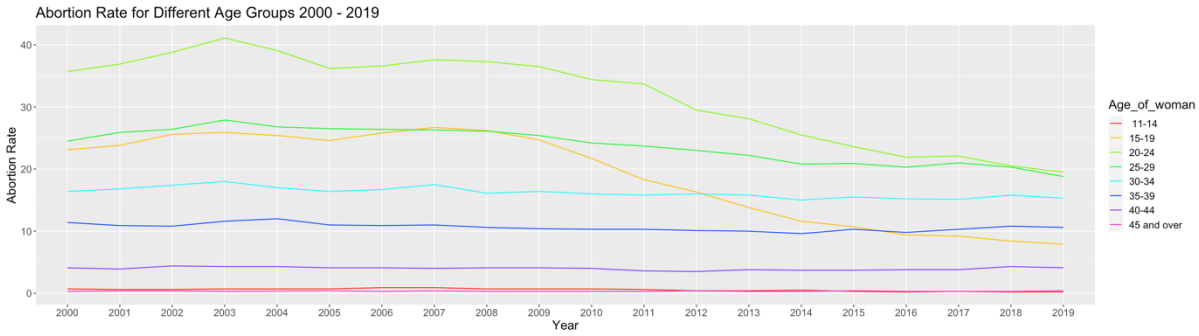


Image 2.7 Line Plot on Abortion Rate for Different Age Groups 2000 - 2019

3. Future Forecast

ARIMA model has been the most sought after for forecast analysis in time series data. Hence it has been decided to forecast the abortion rate in the future. To find the best ARIMA model, auto search identification has been implemented. *Image 3.1* shows the forecasted trend till 2025 (In x-axis: "0" denotes the year "1980" and "40" denotes the year "2019").

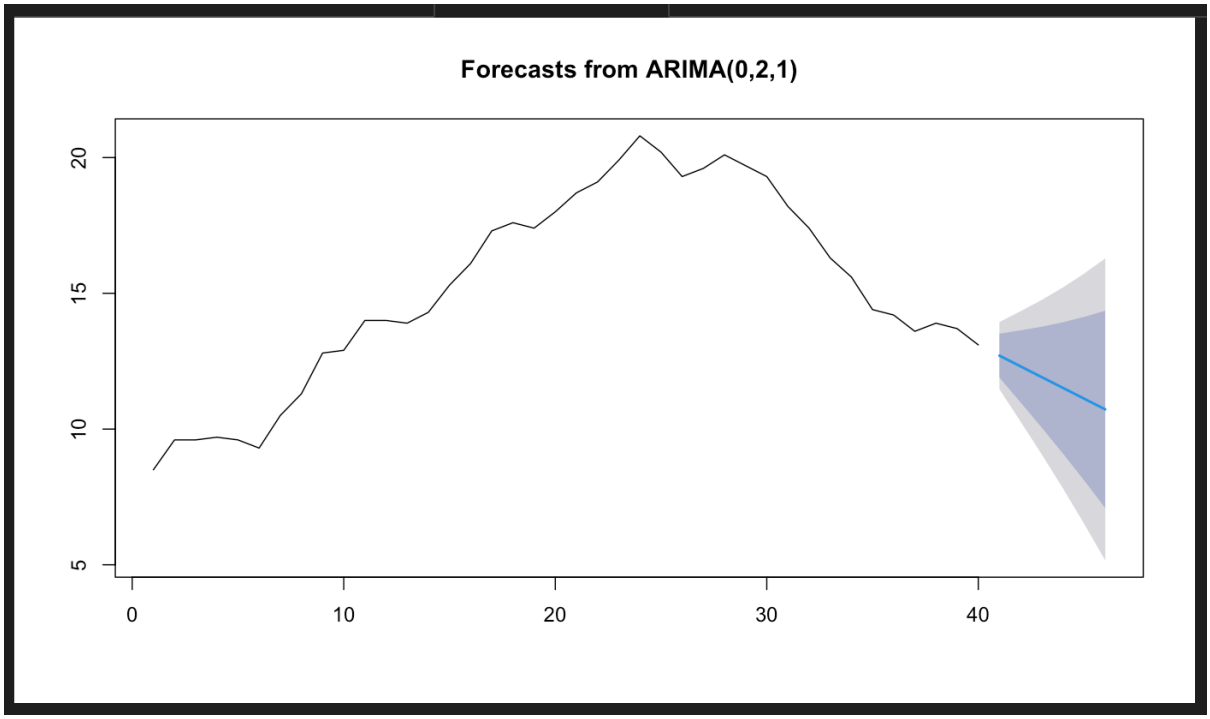


Image 3.1 Forecast of Abortion Rate using ARIMA 021 Model

The best model ARIMA (P, D, Q) model comes to be the ARIMA (0,2,1), Model. Here P denotes the number of seasonal auto aggressive terms. In this case, P = 0 meaning that the pattern of values is random and not affected seasonally. As the model has now been identified as non-seasonal D = 2 indicates that there are two non-seasonal differences. Q = 1 indicates the number of lagged forecast errors which occurs due to the change in abortion rate since 2007. This ARIMA (0,2,1) comes under the category of linear exponential smoothing models.

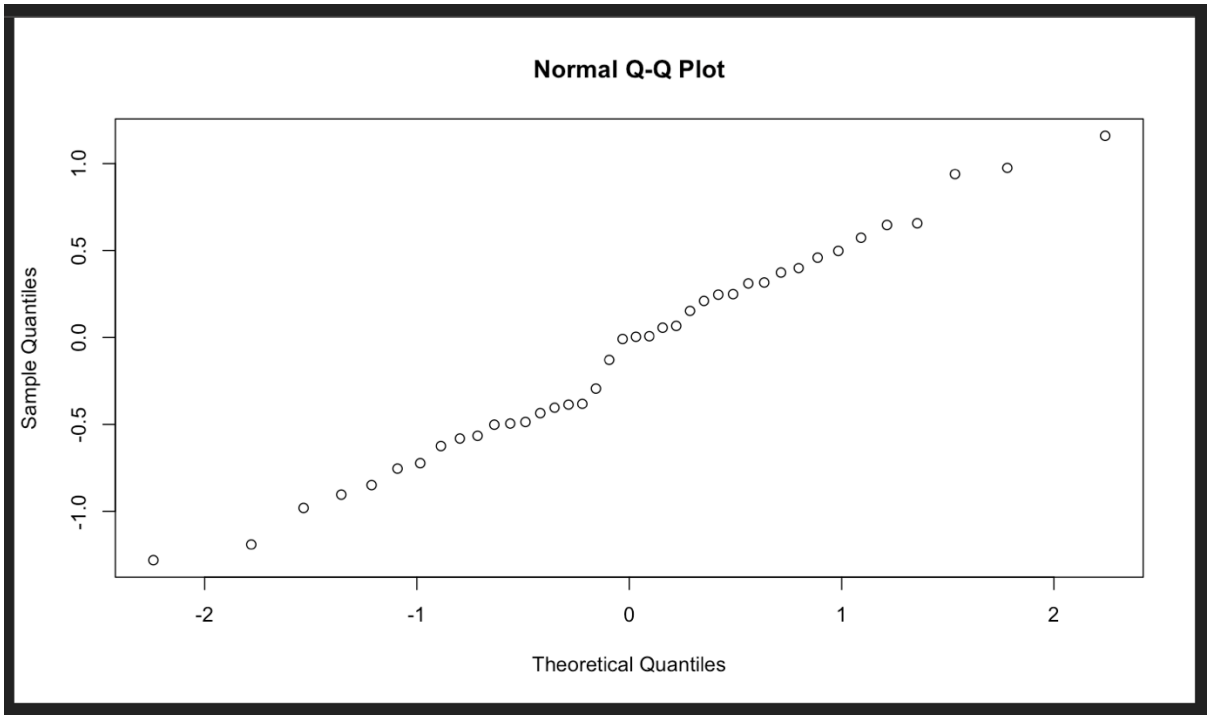


Image 3.2 Q-Q Residual Plot of the ARIMA 0,2,1 Model

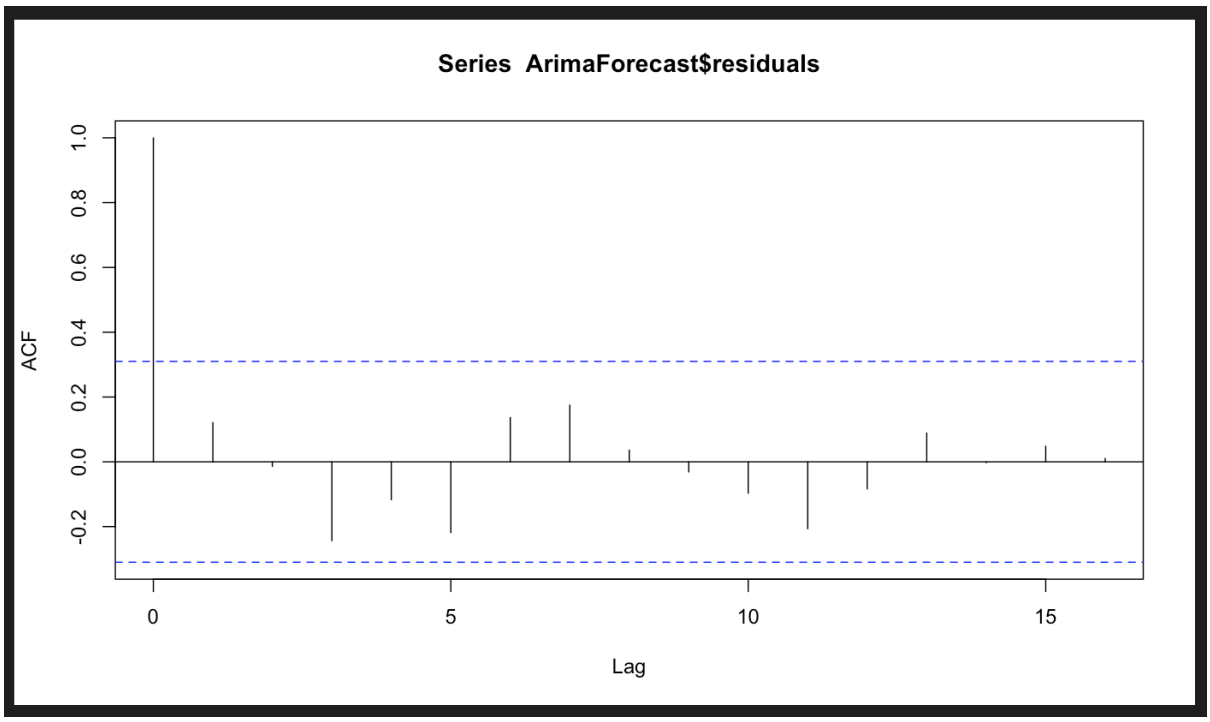


Image 3.3 ACF Residual Plot of the ARIMA 0,2,1 Model

The residual plots are generated. QQ Plot in *Image 3.2* helps to understand the Actual vs Predicted values. The close to a straight line graph of the residuals indicates the good performance of the model. ACF plot in *Image 3.3* says the relation of residual value with the past value. PACF plot in *Image 3.4 s*ays the relation of residual value with the next values. From the ACF and PACF plots it can be seen that the residuals stays well within their boundary except only one in the ACF plot. This indicates that the model has performed well and fit of the model is accurate.

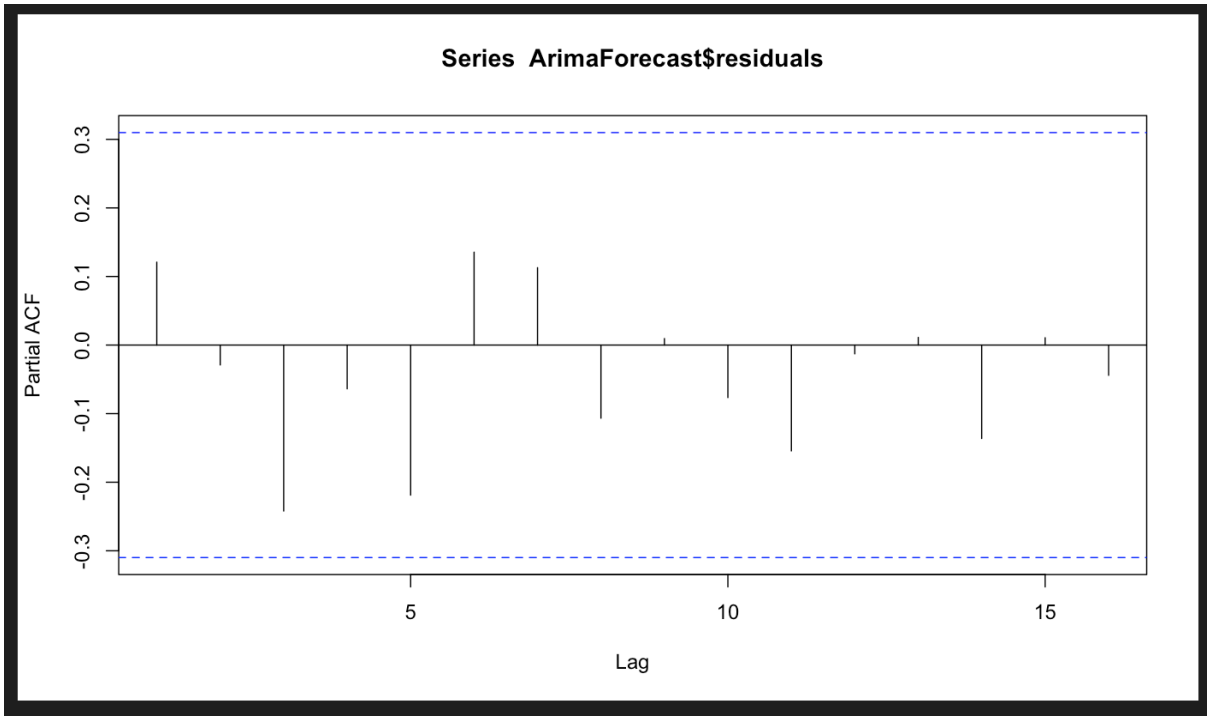


Image 3.4 PACF Residual Plot of the ARIMA 0,2,1 Model

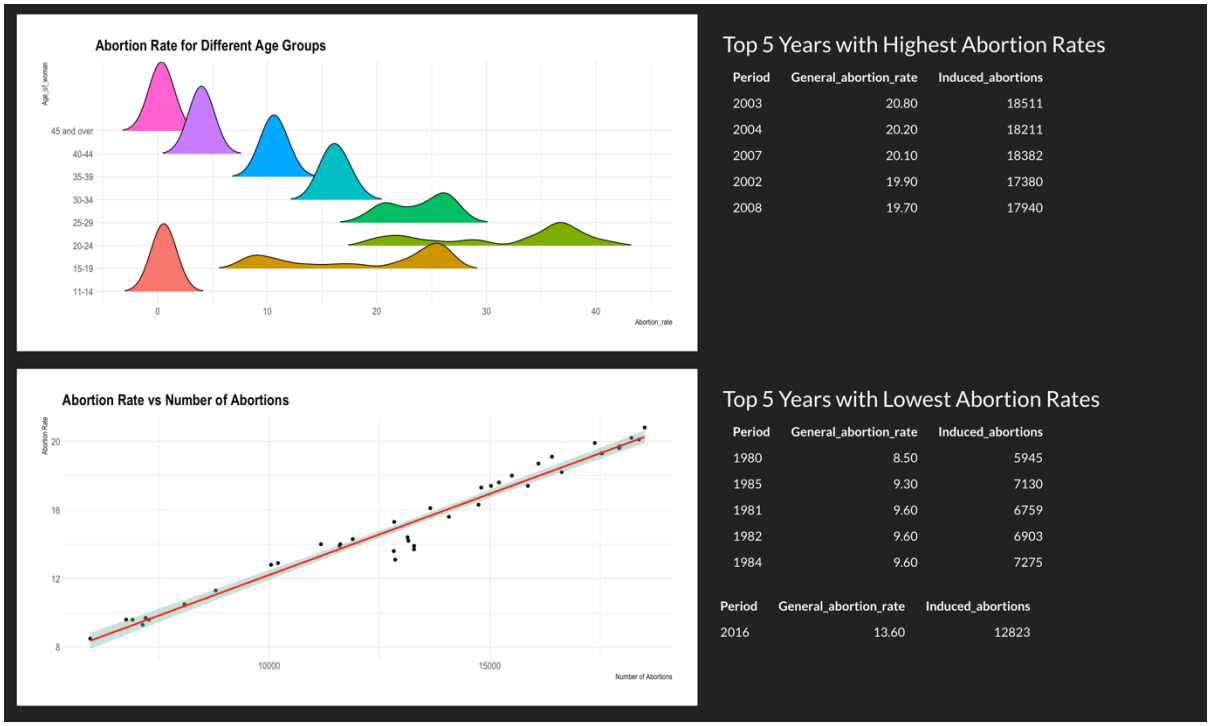
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-0.09198872	0.6057224	0.5066244	-0.5847859	3.48559	0.8372183	0.120942

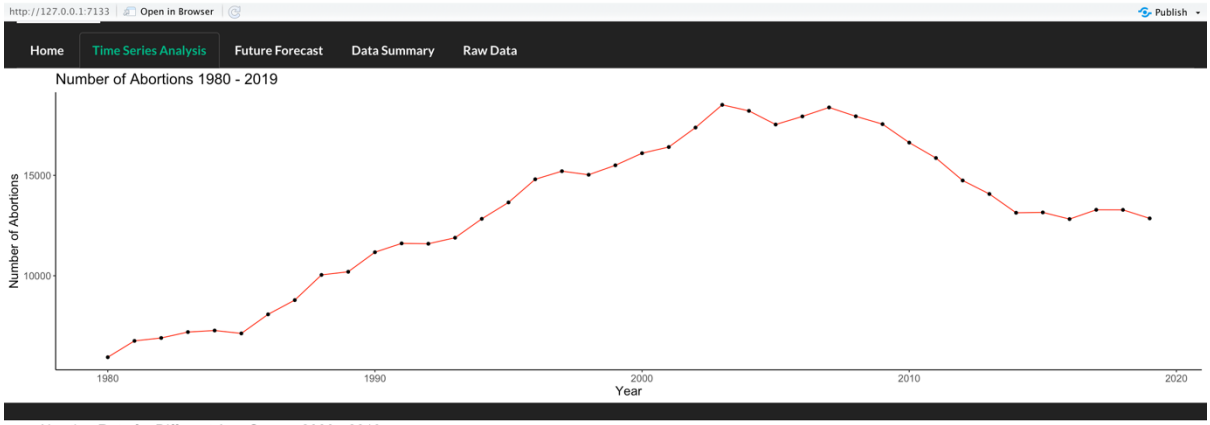
Image 3.5 Perfromance Metrics of the ARIMA 0,2,1 Model

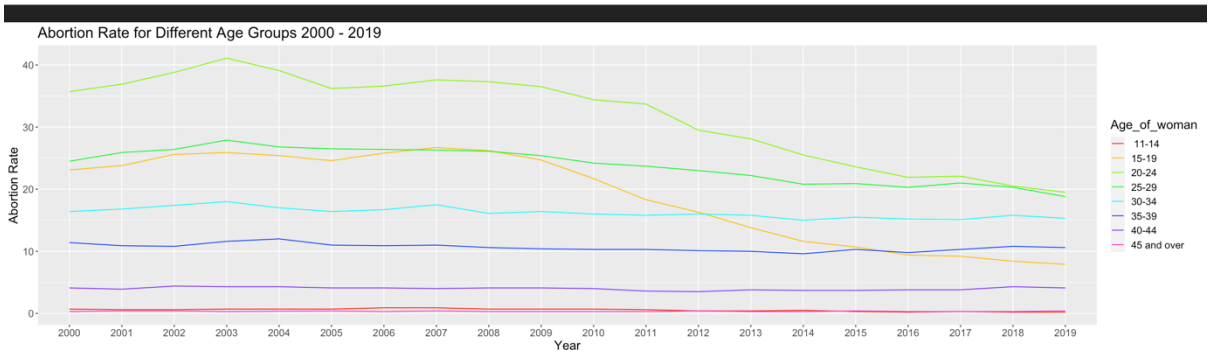
Finally analysing the performance metrics in *Image 3.5,* the RMSE value is 0.6 which is considerably low and the MAPE is 3.48 which means there only 3.48% error and hence the accuracy of the model comes out to be 96.52% which is good.

4. Appendix

This section contains the screen captures of the shiny application pages.







	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	−0.09198872	0.6057224	0.5066244	−0.5847859	3.48559	0.8372183	0.120942


http://127.0.0.1:7133
Open in Browser
Publish
Abortion Statistics in New Zealand 1980 - 2019
Home
Time Series Analysis
Future Forecast
Data Summary
Raw Data
Data Frame Summary
AbortionCR
Dimensions: 40 x 4
Duplicates: 0
No Variable Stats / Values Freqs (% of Valid) Graph Valid Missing
1 Period [factor] 1. 1980 1 (2.5%) 40 0
2. 1981 1 (2.5%) (100.0%) (0.0%)
3. 1982 1 (2.5%)
4. 1983 1 (2.5%)
5. 1984 1 (2.5%)
6. 1985 1 (2.5%)
7. 1986 1 (2.5%)
8. 1987 1 (2.5%)
9. 1988 1 (2.5%)
10. 1989 1 (2.5%)
[30 others] 30 (75.0%) IIIIIIIIIIIIIII
2 General_abortion_rate [numeric] Mean (sd) : 15.2 (3.6) 34 distinct values 40 0
min < med < max: (100.0%) (0.0%)
8.5 < 14.9 < 20.8
IQR (CV) : 5.3 (0.2)
3 Induced_abortions [integer] Mean (sd) : 13185.7 (3769.3) 40 distinct values 40 0
min < med < max: (100.0%) (0.0%)
5945 < 13283.5 < 18511
IQR (CV) : 5250 (0.3)
4 Date [Date] min : 1980-12-31 40 distinct values 40 0
med : 2000-07-01 (100.0%) (0.0%)
max : 2019-12-31


http://127.0.0.1:7133
Open in Browser
Publish
Abortion Statistics in New Zealand 1980 - 2019
Home
Time Series Analysis
Future Forecast
Data Summary
Raw Data
AbortionCR
AbortionCRAge
Search and Sort the Dataset
Show 10 entries
Search:
Period
Age_of_woman
Abortion_rate
Induced_abortions
1 2000 11-14 0.7 74
2 2000 15-19 23.1 3107
3 2000 20-24 35.7 4548
4 2000 25-29 24.5 3399
5 2000 30-34 16.4 2496
6 2000 35-39 11.4 1823
7 2000 40-44 4.1 613
8 2000 45 and over 0.3 43
9 2001 11-14 0.6 66
10 2001 15-19 23.8 3240
Showing 1 to 10 of 159 entries
Previous
1
2
3
4
5
16
Next