VIRGINIA COMMONWEALTH UNIVERSITY

FORECASTING METHODS

ASSIGNMENT 3

SATYANARAYAN VENKAT NALDIGA

V01108247

SUBMITTED TO-

PROF.JASON MERRICK

Date of Submission: 09-25-2024

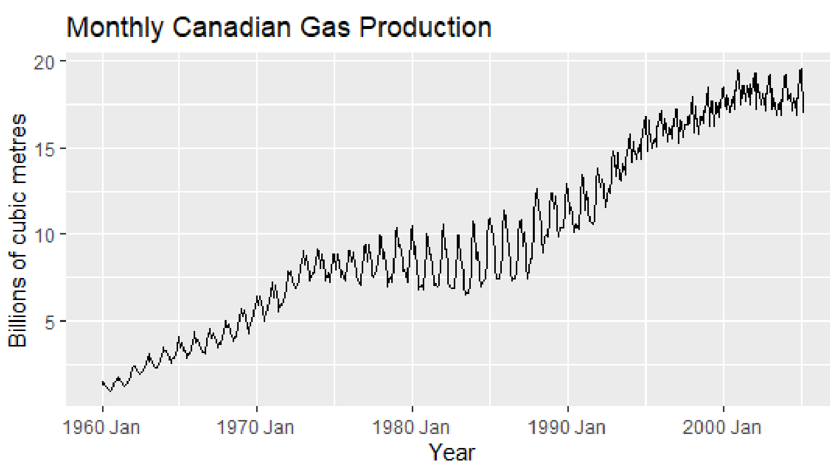
Introduction to Canadian Gas Production Data Analysis

In this exercise, we analyze monthly Canadian gas production data from January 1960 to February 2005, measured in billions of cubic meters. Time series data like this often exhibit patterns such as trends, seasonality, and cyclic behavior, making it ideal for advanced decomposition and visualization techniques. Understanding how seasonality and trends change over time in gas production can offer valuable insights into the dynamics of the energy sector and factors influencing production levels.

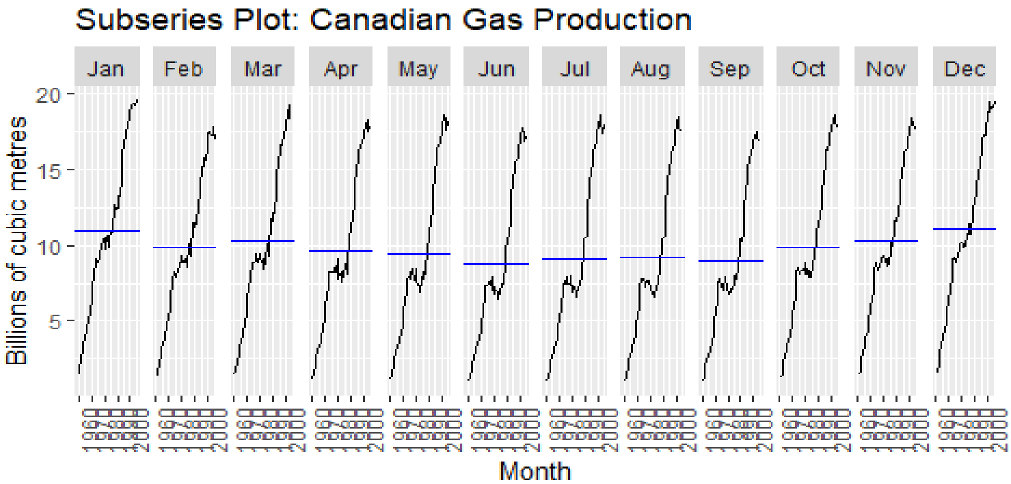
We will perform the following tasks:

1. **Visualization**: Using autoplot(), gg_subseries(), and gg_season() to explore the structure of the data and observe how production and seasonality evolve over time.
2. **STL Decomposition**: Applying the STL (Seasonal-Trend decomposition using LOESS) method with a flexible seasonal window to study the changing seasonality and trend components. The STL decomposition allows us to isolate these elements, providing a clearer picture of how gas production behaves over time.
3. **Comparison with SEATS and X-11 Decomposition**: We will compare the STL results with those obtained using classical decomposition methods—SEATS (Seasonal Extraction in ARIMA Time Series) and X-11. These methods are commonly used for time series decomposition and help us understand how different approaches model seasonality and trends.

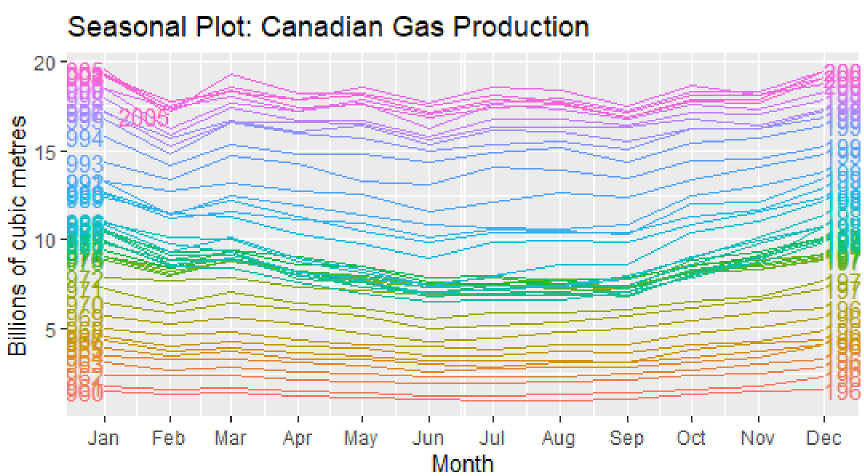
The results will provide insights into the long-term trends in Canadian gas production, the impact of seasonal factors, and how these factors evolve over time. This analysis is important for energy policy makers, market analysts, and businesses involved in the energy sector to anticipate future production patterns and make informed decisions.



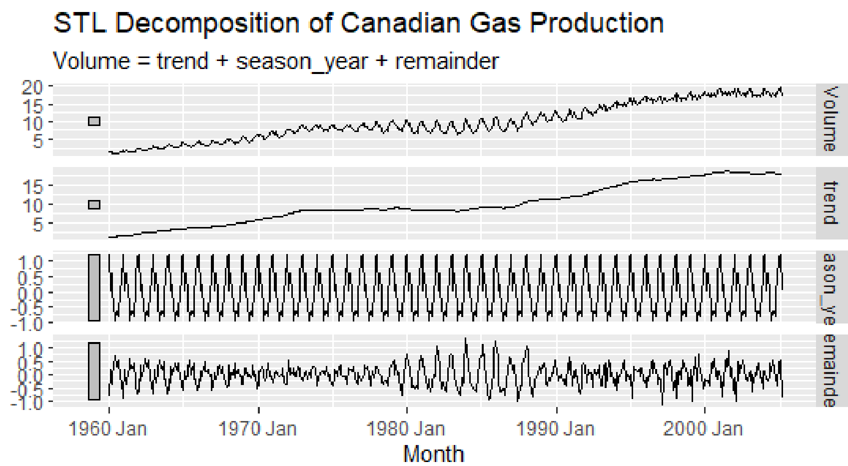
The graph shows the monthly Canadian gas production (in billions of cubic meters) from 1960 to 2005, highlighting a clear upward trend over time. Initially, production grows steadily from the 1960s to the late 1980s, after which it accelerates significantly, indicating increasing demand or technological advancements. The plot also exhibits a strong seasonal pattern, with regular fluctuations within each year, likely reflecting seasonal factors such as increased gas consumption during colder months. Additionally, the amplitude of the seasonal fluctuations appears to grow over time, suggesting that seasonal effects on gas production became more pronounced as overall production increased.



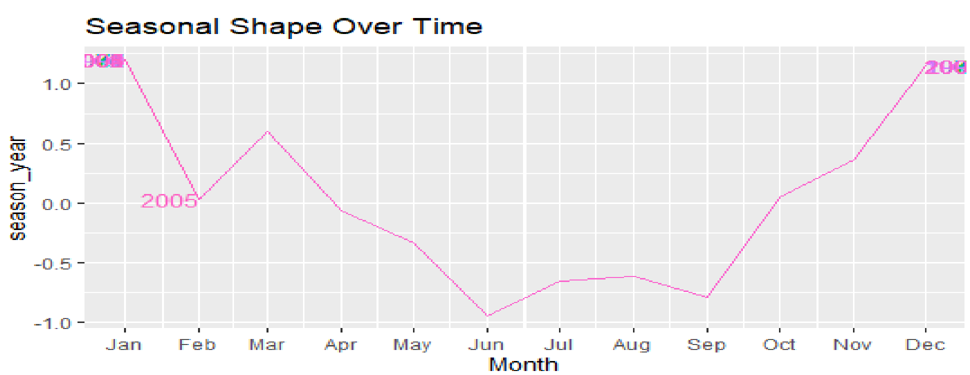
The subseries plot displays the monthly Canadian gas production from 1960 to 2005, separated by month to highlight seasonal patterns and changes over time. Each panel represents a month, showing how production tends to peak in the winter months (December to February) and dip in the summer months (June to August). The upward slope within each month across the years illustrates a consistent increasing trend in gas production. Additionally, the increasing spread within each month's subseries over time indicates that seasonal fluctuations have become more pronounced as production has grown, reflecting both changing seasonal dynamics and the overall trend.



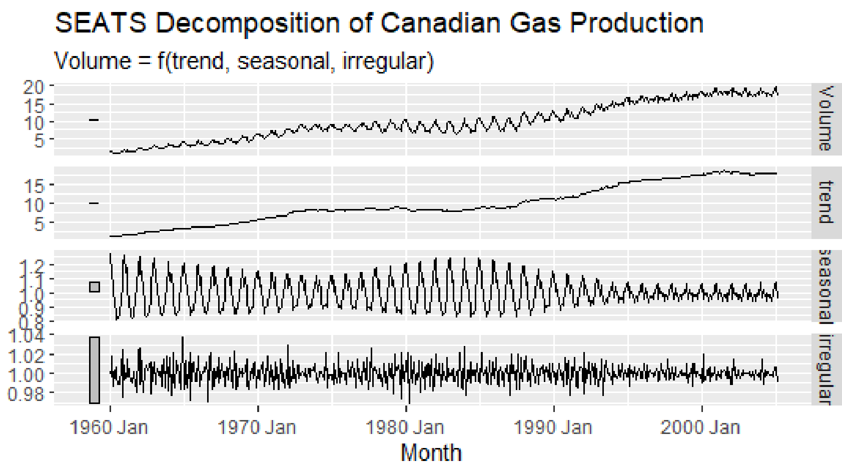
This seasonal plot illustrates the monthly patterns of Canadian gas production from 1960 to 2005, with each line representing a different year. It shows a clear seasonal pattern where production peaks in the winter months (January to March) and drops to its lowest levels in the summer (June to September). Over the years, the lines progressively move upward, indicating an overall increase in production. The increasing vertical spread between lines in later years (closer to 2005) suggests that not only has production grown, but the magnitude of seasonal fluctuations has also become more pronounced, reflecting both long-term trends and changing seasonal effects.



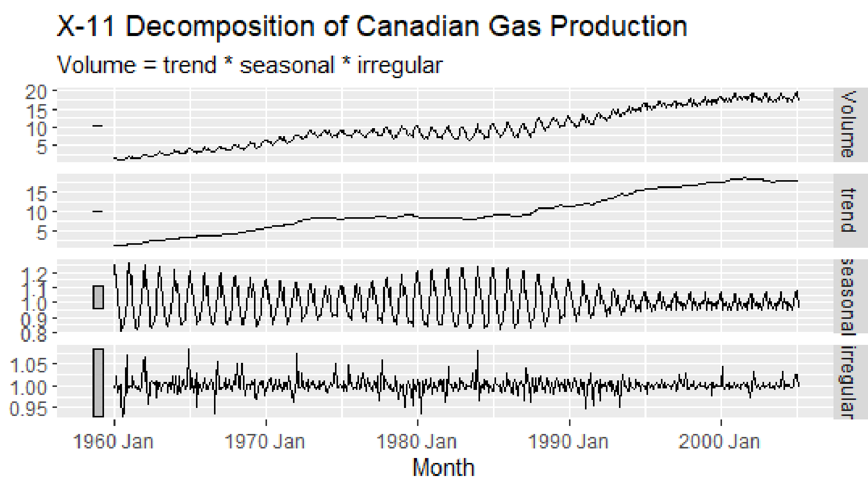
The STL decomposition of Canadian gas production from 1960 to 2005 breaks down the data into three components: trend, seasonality, and remainder. The trend component shows a steady increase in gas production, particularly accelerating during the 1990s before stabilizing at a higher level in the 2000s. The seasonal component reveals a consistent yearly pattern, with peaks during the winter months, and an increasing amplitude over time, indicating that seasonal effects have become more pronounced as production grew. The remainder component captures random short-term fluctuations around zero, with some periods of increased volatility, especially in the late 1980s and early 1990s. This decomposition highlights both the long-term upward trend and the changing seasonal dynamics of gas production.



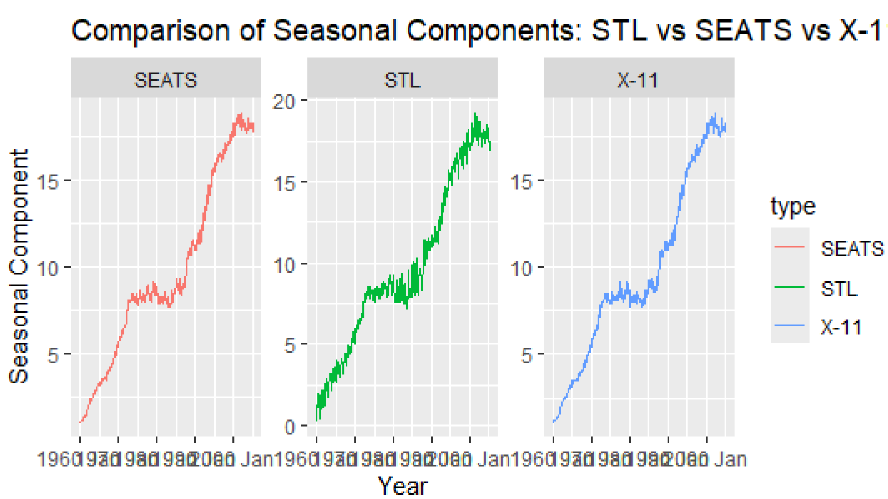
The plot shows the seasonal shape of Canadian gas production over a typical year, highlighting how production fluctuates across the months. There is a clear peak in January, indicating high production levels during the winter, likely due to increased heating demand. Production then rapidly declines through February and reaches its lowest point in the summer months (June to August). Starting in September, there is a steady increase in production, leading to another peak in December. This pattern reflects the strong seasonality of gas production, driven by seasonal variations in energy demand. The symmetrical rise and fall around mid-year emphasize the recurring nature of the seasonal cycle each year.



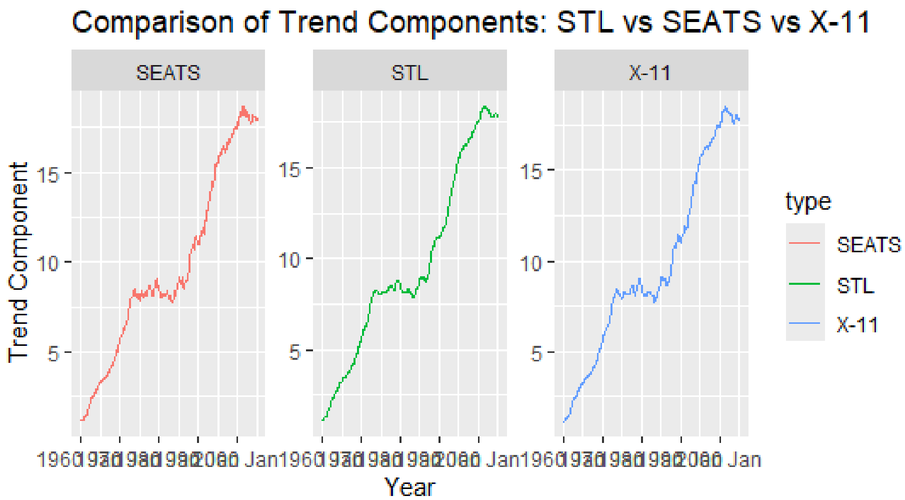
This plot shows the SEATS decomposition of Canadian gas production into its trend, seasonal, and irregular components. The first panel displays the original time series, which has a clear upward trend and repeating seasonal patterns. The second panel isolates the trend, revealing a steady increase in gas production, particularly after the 1980s. The third panel represents the seasonal component, showing a consistent annual cycle with higher production in the winter months and lower production in the summer. Unlike the STL decomposition, the seasonal pattern here appears relatively stable over time. The final panel shows the irregular (or residual) component, capturing short-term fluctuations and random noise around zero. The irregular component has varying levels of volatility, especially in the earlier years, indicating occasional anomalies in gas production that aren't explained by the trend or seasonality.



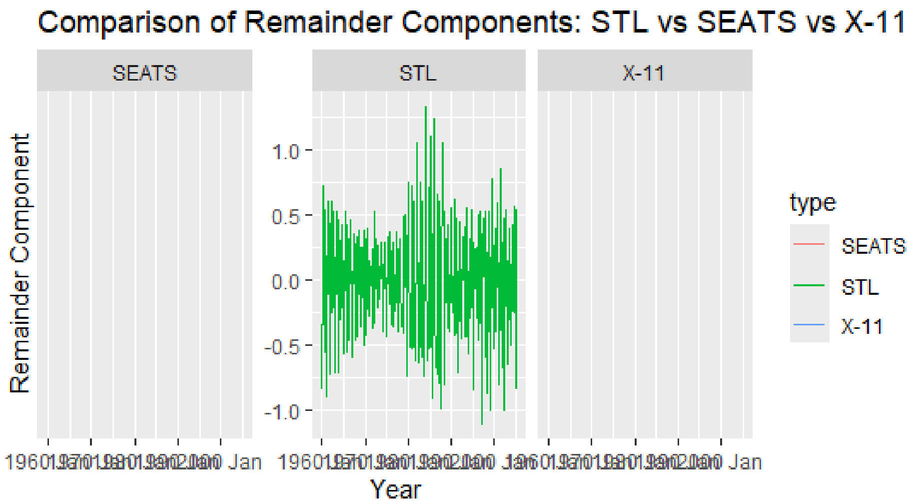
This plot illustrates the X-11 decomposition of Canadian gas production into its trend, seasonal, and irregular components. The first panel shows the original gas production data, featuring an upward trend and repeating seasonal patterns. The second panel isolates the trend, highlighting a steady increase in production, especially from the 1980s onwards. The third panel represents the seasonal component, displaying a consistent yearly cycle with higher production in the winter months and lower production in the summer. Notably, this decomposition maintains a relatively stable seasonal pattern throughout the entire period. The final panel shows the irregular component, capturing short-term variations and random fluctuations around zero. The irregular component suggests some variability in the data, particularly in the earlier years, but the fluctuations decrease slightly in more recent years, indicating fewer anomalies in production.



This comparison plot shows the seasonal components of Canadian gas production derived using three different decomposition methods: SEATS, STL, and X-11. All three panels depict similar seasonal patterns, but with subtle differences. The SEATS and X-11 methods (left and right panels) show relatively smooth seasonal components, indicating a more consistent seasonal effect over time. In contrast, the STL method (center panel) reveals more flexibility, with variations in the seasonal pattern becoming more apparent in recent years. This suggests that STL captures changes in the seasonality's shape and magnitude more dynamically compared to the more rigid seasonal components from SEATS and X-11. Overall, while all methods capture the annual cyclical pattern, STL is more responsive to evolving seasonal variations in the data.



This comparison plot shows the trend components of Canadian gas production using SEATS, STL, and X-11 decomposition methods. All three methods capture the overall upward trend in gas production from 1960 to 2005. However, the SEATS and X-11 methods (left and right panels) display relatively smoother trends, with a gradual increase over time. The STL method (center panel) reveals a slightly more flexible trend, capturing small fluctuations and changes in the growth rate more dynamically. Overall, while the general pattern is consistent across methods, STL provides a more detailed representation of variations in the trend.



This plot compares the remainder (or irregular) components of Canadian gas production using SEATS, STL, and X-11 decomposition methods. The STL method (center panel) shows visible fluctuations around zero, capturing short-term variations and noise in the data that are not explained by the trend or seasonal components. In contrast, the SEATS and X-11 methods (left and right panels) do not display any remainder component, likely because their decomposition process smoothed out most of the irregular variations, or there might be an issue with how the remainder was extracted in this plot. Overall, STL appears to provide a more detailed view of the residual fluctuations in the dataset.

CHAPTER 5

Introduction to Australian Takeaway Food Turnover Analysis

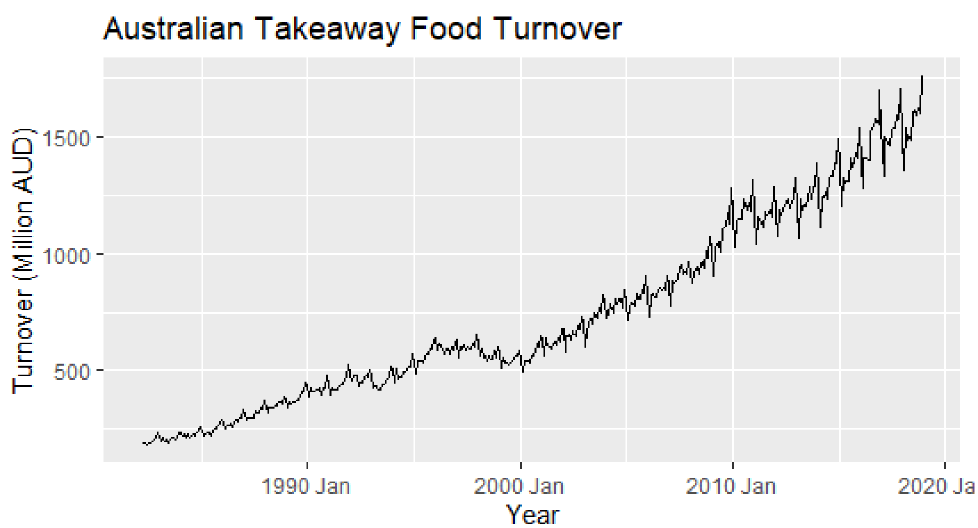
In this exercise, we explore the Australian takeaway food turnover data from the `aus_retail` dataset. This dataset contains monthly turnover data for various retail industries across Australia, and we focus specifically on the "Takeaway food services" industry. The objective is to analyze the turnover trends by first filtering and aggregating the data across all states. To assess and forecast the takeaway food turnover, we will split the data into a training set and a test set, with the training set containing all observations up to the last four years and the test set covering the final four years.

We will employ a range of benchmark forecasting methods—including Naive, Seasonal Naive, and Mean forecasts—to model the training set and predict the turnover for the test period. By comparing the accuracy of these forecasts using appropriate accuracy metrics, we aim to identify which method best captures the patterns in the takeaway food turnover data. This analysis will provide insights into the turnover trends and help evaluate the effectiveness of various forecasting techniques in a retail industry context.

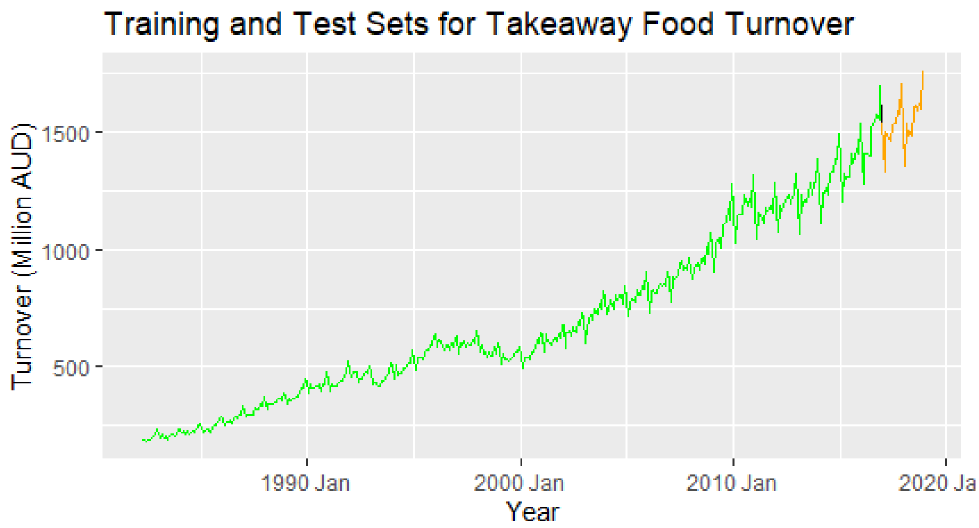
The dataset takeaway_data contains 441 rows and 2 columns:

1. **Month**: This column is of type <mth> and represents the monthly time series data, starting from April 1982.
2. **Turnover**: This column is of type <dbl> and contains numerical values representing the monthly turnover (likely in millions of dollars) for the "Takeaway food services" industry.

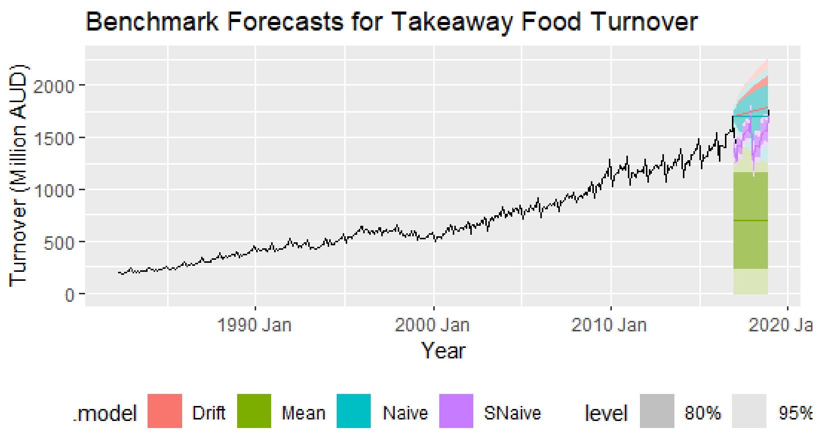
```
glimpse(takeaway_data)
Rows: 441
Columns: 2
$ Month    <mth> 1982 Apr, 1982 May, 1982 Jun, 1982 Jul, 1982 Aug, 1982 Sep, 1982 Oct, 198…
$ Turnover <dbl> 193.7, 193.9, 186.3, 189.5, 190.5, 195.3, 209.0, 212.5, 238.2, 224.7, 202…
```



This plot shows the monthly turnover for the Australian takeaway food industry from April 1982 to January 2020. The graph indicates a clear upward trend in turnover over the years, with noticeable growth accelerating around the mid-2000s. Seasonal fluctuations are present throughout, with periodic peaks and troughs that become more pronounced as turnover increases. This suggests that not only has the takeaway food industry expanded significantly in terms of revenue, but it also experiences regular seasonal effects, possibly related to consumer behavior during different times of the year. The increased volatility in recent years reflects the growing scale and changing dynamics of the industry.



This plot displays the Australian takeaway food turnover data split into training and test sets. The data from April 1982 to December 2015 is marked in green and represents the training set used for building forecasting models. The period from January 2016 onwards, shown in orange, represents the test set, which covers the last four years of data. This division allows for evaluating the model’s forecasting accuracy on recent turnover data. The graph continues to show the overall upward trend and seasonal fluctuations in turnover, indicating consistent growth in the industry with pronounced seasonal peaks and troughs.



This plot shows the benchmark forecasts for Australian takeaway food turnover using different forecasting methods: Drift, Mean, Naive, and Seasonal Naive (SNaive). The historical data is displayed in black, while the forecasts for the test period (2016 onwards) are marked in various colors, representing each method's predictions. The colored bands around each forecast indicate the prediction intervals, with both 80% and 95% confidence levels shown. The Seasonal Naive (purple) and Naive (blue) forecasts appear to follow the seasonality and trend more closely, while the Mean (green) and Drift (red) models show less adaptability to the turnover's seasonal pattern. This comparison helps identify which forecasting method best captures the trend and seasonal variations in the turnover data, with the Seasonal Naive method likely offering the most accurate predictions given the observed seasonal patterns.

MODEL	RMSE	MAE	MAPE	ACF1
DRIFT	222.0	207.0	13.8%	0.162
MEAN	849.0	844.0	54.6%	0.308
NAIVE	186.0	167.0	11.2%	0.308
SNAIVE	73.7	62.8	4.13%	0.408

- **RMSE** and **MAE** both measure the forecast errors in units of the turnover. The Seasonal Naive (SNaive) model has the lowest RMSE (73.7) and MAE (62.8), indicating that it provides the most accurate forecasts. In contrast, the Mean model has the highest errors (RMSE = 849.0 and MAE = 844.0), showing that it struggles to capture the data's patterns.
- **MAPE** reflects the average percentage error of the forecasts. The SNaive model has the smallest MAPE (4.13%), suggesting it is most accurate relative to the scale of the turnover data, while the Mean model has the highest MAPE (54.6%), indicating poor performance.
- **ACF1** represents the autocorrelation of the residuals at lag 1, which indicates how much error from one period is correlated with the next. The SNaive model has a higher ACF1 (0.408), possibly implying that some patterns remain in the residuals. However, its low errors in other metrics show that it still outperforms the other models overall.

The takeaway food turnover exhibits a clear and strong upward trend, reflecting consistent growth in the industry over time. Among the models, the Seasonal Naive (SNaive) model, which effectively captures both the trend and seasonal variations, has proven to be the most reliable for this dataset. This is further validated by its relatively narrow forecast intervals compared to those of the other models.