BOMI SO

Time Series Analysis

1. Introduction

In this research, it analyses a dataset regarding private car registration in Ireland, which is collected between January 1995 and January 2022. Firstly, it explores the descriptive statistics of the dataset using Python. The results show that the dataset contains a total of 325 values, representing the number of car registrations per month. The maximum registration is 32961, while the minimum is 474. A normality test and examination of the histogram indicates that the registration data is not normally distributed.

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 325 entries, 1995-01-31 to 2022-01-31
Freq: M
Data columns (total 2 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Registration  325 non-null    int64
 1   Unnamed: 2    325 non-null    float64
dtypes: float64(1), int64(1)
memory usage: 7.6 KB
```

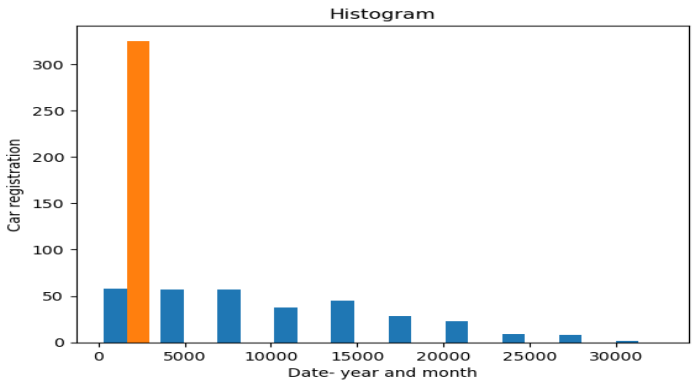
	Registration	Unnamed: 2
count	325.00000000	325.00000000
mean	10510.54153846	0.00000000
std	7175.56340901	0.00000000
min	474.00000000	0.00000000
25%	4355.00000000	0.00000000
50%	9359.00000000	0.00000000
75%	15384.00000000	0.00000000
max	32961.00000000	0.00000000

In [35]:
```
from scipy.stats import normaltest
import numpy as np

data = df['Registration']
stat, p = normaltest(data)

if p > 0.05:
    print("The car registration data is normally distributed (fail to reject H0)")
else:
    print("The car registration data is not normally distributed (reject H0)")
```

```
The car registration data is not normally distributed (reject H0)
```



Based on the results, the dataset will be tested for time series analysis to forecast the future of car registrations in Ireland per month. In order to conduct the analysis, dataset is assessed to identify seasonality and trends in Section 3 as a first step.

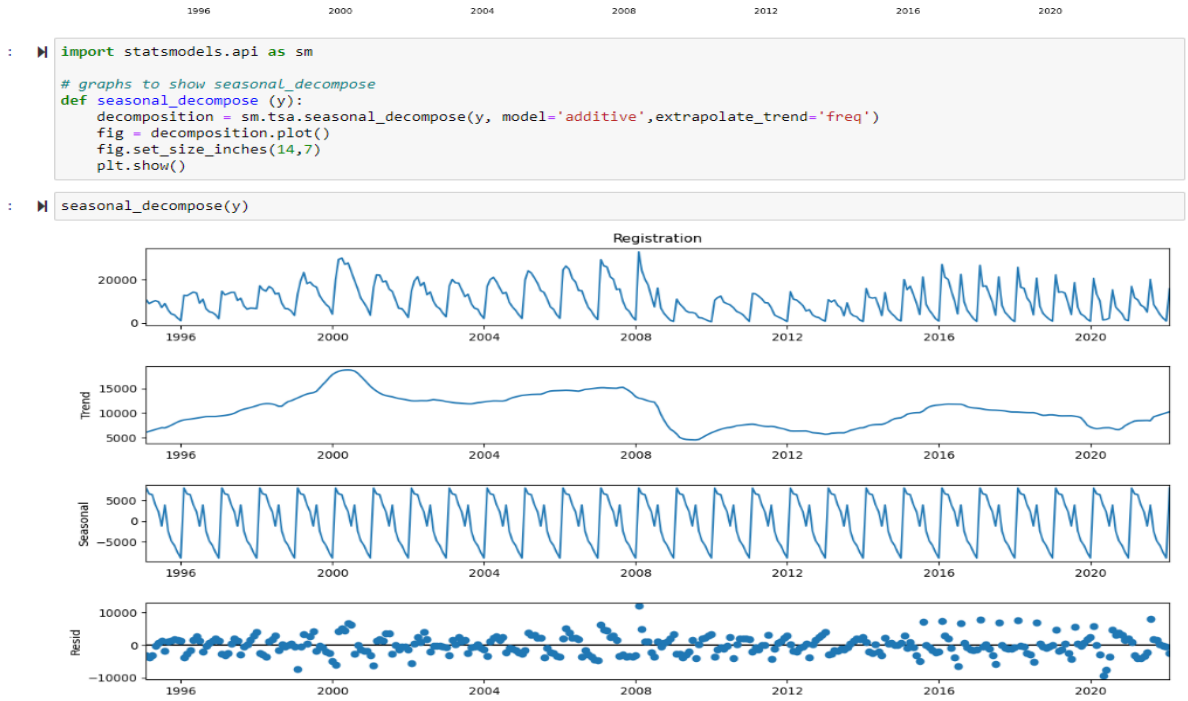
2. Background & Introduction

There are three types of models for time series analysis. Firstly, the Exponential Smoothing (ES) method is a technique commonly used for data that has a trend or/and seasonality. It is an appropriate model for forecasting non-stationary data [1]. In this paper, the Holt-Winter's Exponential Smoothing method is utilized. Holt-Winter's exponential smoothing is a technique which forecasts data containing level, trend, and seasonality [1]. Secondly, the Autoregressive Integrated Moving Average (ARIMA) model is a technique for forecasting future values in a time series by using past data and an error term [2]. It assumes the stationarity of the dataset and is therefore appropriate for datasets that have stationarity [2]. Lastly, the simple time series model is a model that is often utilized for underlying models before building ES or ARIMA models and assume a trend for the data [3]. More specifically, mean models, linear trend models, and random walk models are typical simple time series models [3].

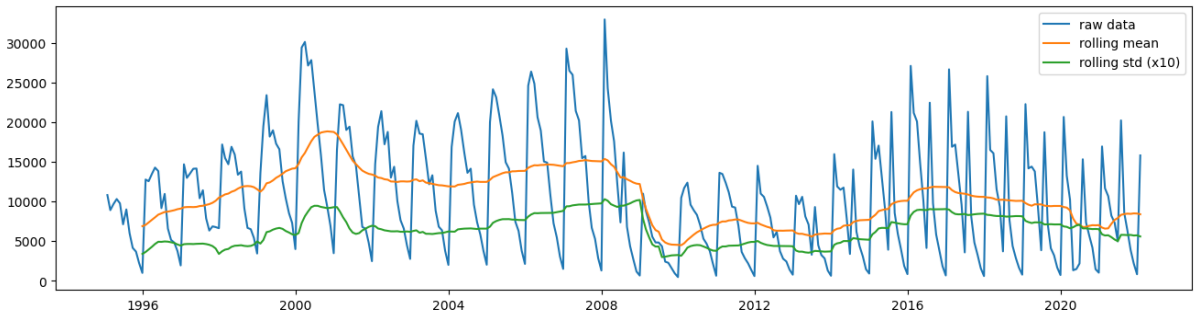
3. Selected model test result

a. Identifying appropriate model for time series analysis

Firstly, in order to process the data using Python, the date format was changed to yyyy-mm-dd using Excel. After this process, secondly, Python was run to identify the seasonality and trend in the data.



```
pd.options.display.float_format = '{:.8f}'.format
test_stationarity(y,'raw data')
```



According to the decomposition results, it is identified that there is a clear seasonality in this data. Additionally, it is observed that there is a trend presented. Lastly, the stationarity of the dataset is analysed. Generally, stationarity is recognized by three conditions: 1) a constant mean of the data, 2) a constant standard deviation, and 3) the absence of seasonality. In this data, the graph indicates seasonality. Also, the mean and standard deviation are not constant. Therefore, it can be concluded that there is no stationarity in this data. According to the test result, 'Holt-Winter method (Exponential Smoothing model) will be used as the data has no stationarity but has trend and seasonality.

b. Test process and result

Test is conducted via Python.

```
In [89]: from statsmodels.tsa.holtwinters import ExponentialSmoothing

In [90]: # Load your time series data into a pandas DataFrame
         data = pd.read_csv('data1.csv')
         df= data['Registration']

In [91]:
         # Define the model
         model = ExponentialSmoothing(df, trend='add', seasonal='add', seasonal_periods=12)
```

As a result, six periods prediction is as below.

```
In [94]: print(forecast)

2022-02-28   13775.96233285
2022-03-31   13156.91126004
2022-04-30    9678.88525895
2022-05-31    8564.52132759
2022-06-30    5954.27926952
2022-07-31   19954.74316334
Freq: M, dtype: float64
```

3. Conclusion

In this research, the aim is to analyse private car registration in Ireland between January 1995 and January 2022 and forecast the future number of registrations based on the dataset. To achieve this, time series analysis is conducted.

As a first step, three time series models - ETS, ARIMA, and simple time series analysis - are compared. Additionally, the dataset's trend, seasonality, and stationarity are visualized as graphs using Python. Based on the results, it is identified that the dataset has clear seasonality and trend, however, it lacks stationarity, making it ineligible for the selection of the ARIMA model. Therefore, it is decided to select the ETS model as ETS model is applicable for data which has no stationarity. Among ETS models, the Holt-Winter method is selected and run via Python as the data has seasonality and trend. As a conclusion, the next six periods of private car registration are predicted as follows. (The decimal point was rounded to an integer).

Month	Registration (forecasting)
Feb 2022	13776
Mar 2022	13157
Apr 2022	9679
May 2022	8565
June 2022	5954
July 2022	19955

References

[1] R. J. Hyndman and G. Shmueli, "Forecasting: principles and practice," [Online]. Available: http://www.forecastingbook.com. [Accessed: 01-Jan-2023].

[2] S. Glen, "ARIMA (Box-Jenkins Models): Autoregressive Integrated Moving Average," [Online]. Available: https://www.statisticshowto.com/arima/. [Accessed: 01-Jan-2023].

[3] S.Tyagi, "Introduction to Time Series Forecasting Part 1: Average and Smoothing Models,", Towards Data Science, [Online] Available: https://towardsdatascience.com/introduction-to-time-series-forecasting-part-1-average-and-smoothing-models-a739d832315 [Accessed: 02-Jan-2023]