## **Final Exam**

## 

## **Case: Forecasting quarterly net sales for Tiffany & Co**

The 1830s in New York City were a time of dynamic growth and golden opportunity for anyone with a little capital and an abundance of imagination. In 1837 New York became the proving ground for 25-year-old Charles Lewis Tiffany and John B. Young, who opened a “stationery and fancy goods” store with a $1,000 advance from Tiffany’s father. Tiffany retails and distributes a selection of Tiffany & Co. brand jewelry at a range of prices. Today more than 150 Tiffany & Co. stores sell to customers in US and international markets. In addition to jewelry, it sells Tiffany & Co brand merchandise in the following categories: stainless steel flatware; crystal, glassware, china, tableware; custom engraved stationery; writing instruments; and fashion accessories. Fragrance products are sold under trademarks Tiffany, Pure Tiffany, and Tiffany for Men. Tiffany also sells other brands of timepieces and tableware in its US stores.

Data: File

<https://github.com/robertasgabrys/Forecasting/blob/main/TiffanySales.xlsx?raw=true>

Tiffany’s quarterly sales for 2005 – 2013 along several economic indicators

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Sales | Tiffany's Net sales |
| CPI | Gross Domestic Product |
| GDP | Consumer Price Index |
| DPI | Disposable Personal Income |
| Unemp | Unemployment |
| Interst Rate | Interest Rate |

1. Install necessary libraries

Remark: you may or may not need to downgrade matplotlib. Just in case you need the following command may help:

!pip install matplotlib==3.5

2. Load necessary libraries

3. Import the data.

4. Create a graph of data and interpret it.

5. Divide data into training and testing sets: last 4 data points represent your testing set.

Build the following models to predict Tiffany’s quarterly sales:

6. Model 1: seasonal naïve model: Q1 prediction = previous year’s Q1 value, Q2 prediction = previous year’s Q2 value, etc.

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

7. Model 2 : regression model that captures linear trend and annual seasonality. You need to create these features and build a model.

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

8. Model 3 : regression model that captures linear trend, annual seasonality and abrupt change starting 2008-04-01 (You noticed that there was a decline in sales because of financial crisis).

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

9. Model 4 : regression model that captures linear trend, annual seasonality and temporary abrupt change starting 2008-04-01 and ending 2009-04-01 (both end points are included).

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

10. Model 5 : Model 3 + Ramp starting 2008-04-01.

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

11. Model 6 : Model 4 + Ramp starting 2008-04-01.

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

12. Model 7 : Model 5 + Ramp starting 2008-04-01.

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

13. Model 8 : Model 4 + Ramp 1 starting 2008-04-01 + Ramp 2 starting 2009-07-01

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

14. Model 9 : Model 5 + Ramp 1 starting 2008-04-01 + Ramp 2 starting 2009-07-01

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

15. Model 10 : use exponential smoothing. Determine the optimal setting that gives most accurate prediction.

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

16. Model 11: auto arima model: max p, q, P, Q ≤ 2, d+D ≤ 2.

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

17. Model 12: select the most accurate model among Models 2 – 9 and dump its features into auto arima.

Remark:<https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html>

The exogenous features are denoted X, i.e. X = data frame whose columns are exogenous regressors x1, x2, etc. where x1, x2, .. are column names of the features you want to include in the model.

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

18. Model 13: regression that predicts sales from exogenous variables, CPI,GDP, DPI, Unemp, Interest rate.

Remarks: 1. since you cannot use the values of the exogenous variables on the testing set, use auto arima function to predict them; 2. You may explore and identify the best variables for the model: maximum combinations is 2^5 = 32. Try a few combinations. If you feel comfortable with the concept of loop, you can perform an exhaustive search. Answer a.b. and c for your best model.

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

19. Model 14: dump exogenous variables that you identified in previous question into auto arima.

Remark:<https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html>

The exogenous features are denoted X, i.e. X = data frame whose columns are exogenous regressors x1, x2, etc. where x1, x2, .. are column names of the features you want to include in the model.

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.

20. Model 15: use pycaret and based on a 2 fold time series cross-validation select the best model and then generate the prediction on the testing set. It is up to you if you include or not exogenous variables. You can try both options.

a. Create a graph of the entire series of sales and overlay the fitted values and the forecast on the testing set.

b. Report MAPE on the training and testing sets.

c. Is your model adequate? Briefly explain.