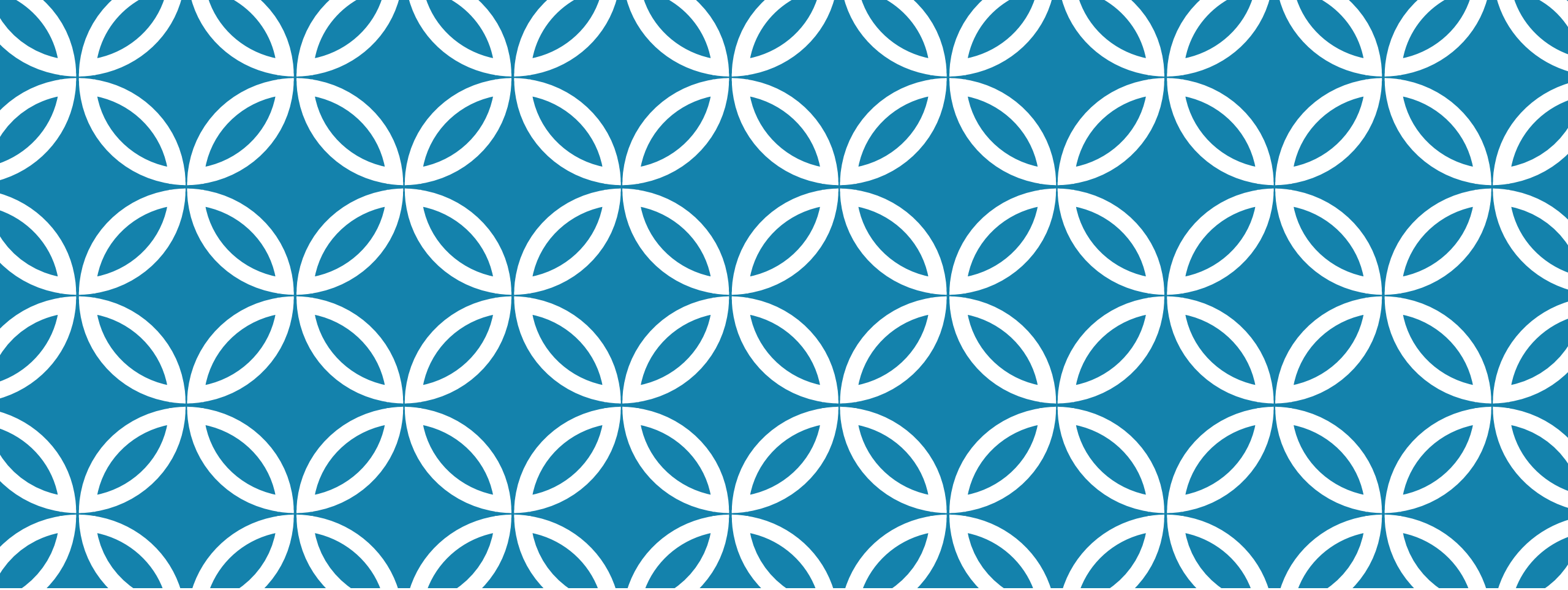


TIME SERIES FORECASTING

Python

PROBLEM

Forecast sales in the Time Series dataset

INTRODUCTION OF TIME SERIES ANALYSIS AND FORECASTING

Time series forecasting is essential to predict events in immediate future.

Time series analysis helps in identification individual components of time series – known as decomposition

Time series analysis and forecasting is based on summarized data based on a time dimension

Time series is usually equally spaced events. If not, the time dimension is usually modified

DATASET INTRODUCTION

The dataset is the famous tableau dataset Superstore.

There are 9,994 rows of data.

There are 21 attributes of the dataset

For the sake of simplicity, I used following columns:

- Order Date
- Category
- Sales

DATASET INTRODUCTION

Order Date has a range from 03Jan2014 to 30Dec2017. It has a format of dd/mm/yyyy

The category has 3 values:

- Furniture
- Office Supplies
- Technology

Sales is a continuous variable with the minimum value of 0.44 and maximum value of 22,638.48

GAME PLAN FOR SOLUTION

Use Python as the enabling platform for forecasting

Read the dataset

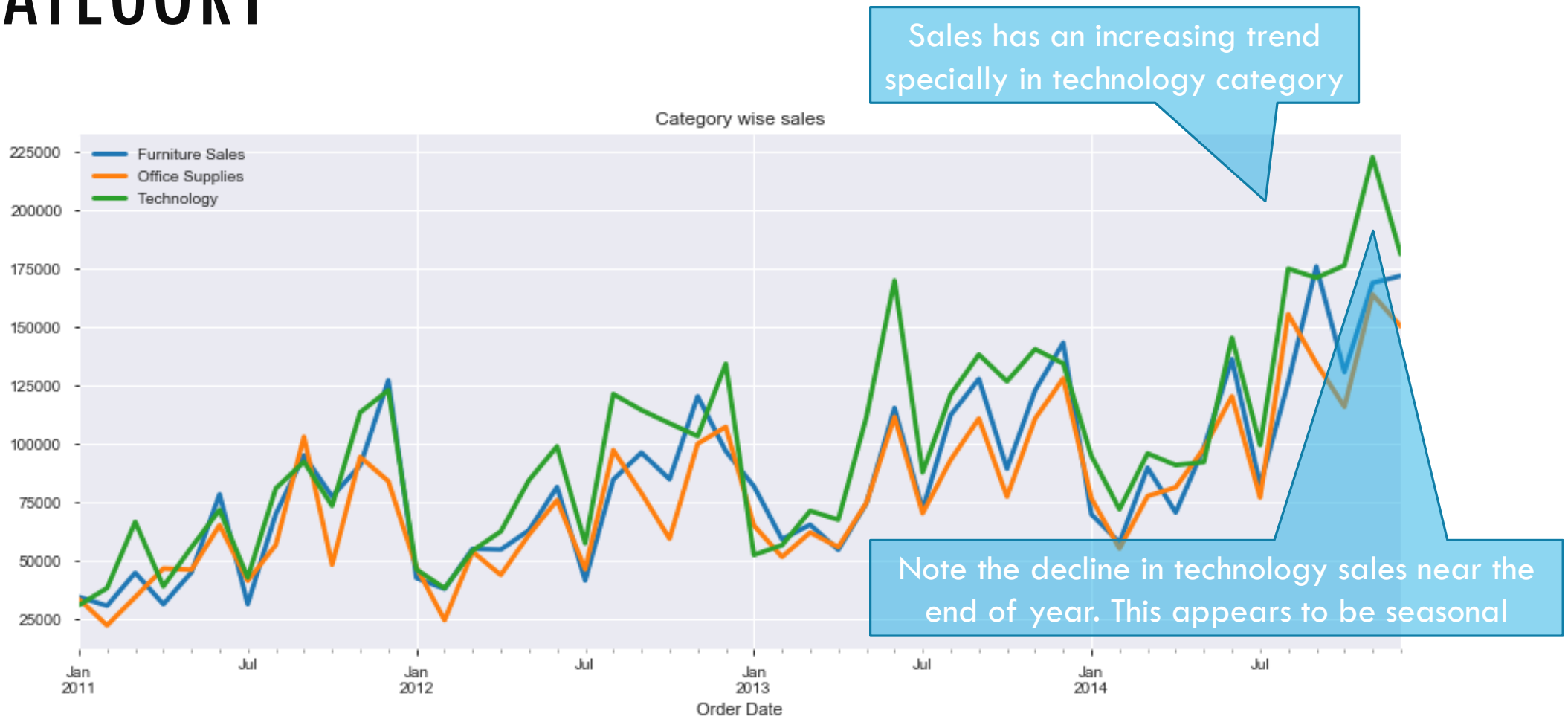
Bring the dataset in proper format to enable time series analysis

Use SARIMAX (**S**easonal **A**uto **R**egressive **I**ntegrated **M**oving **A**verage E**x**ogenous) method to forecast

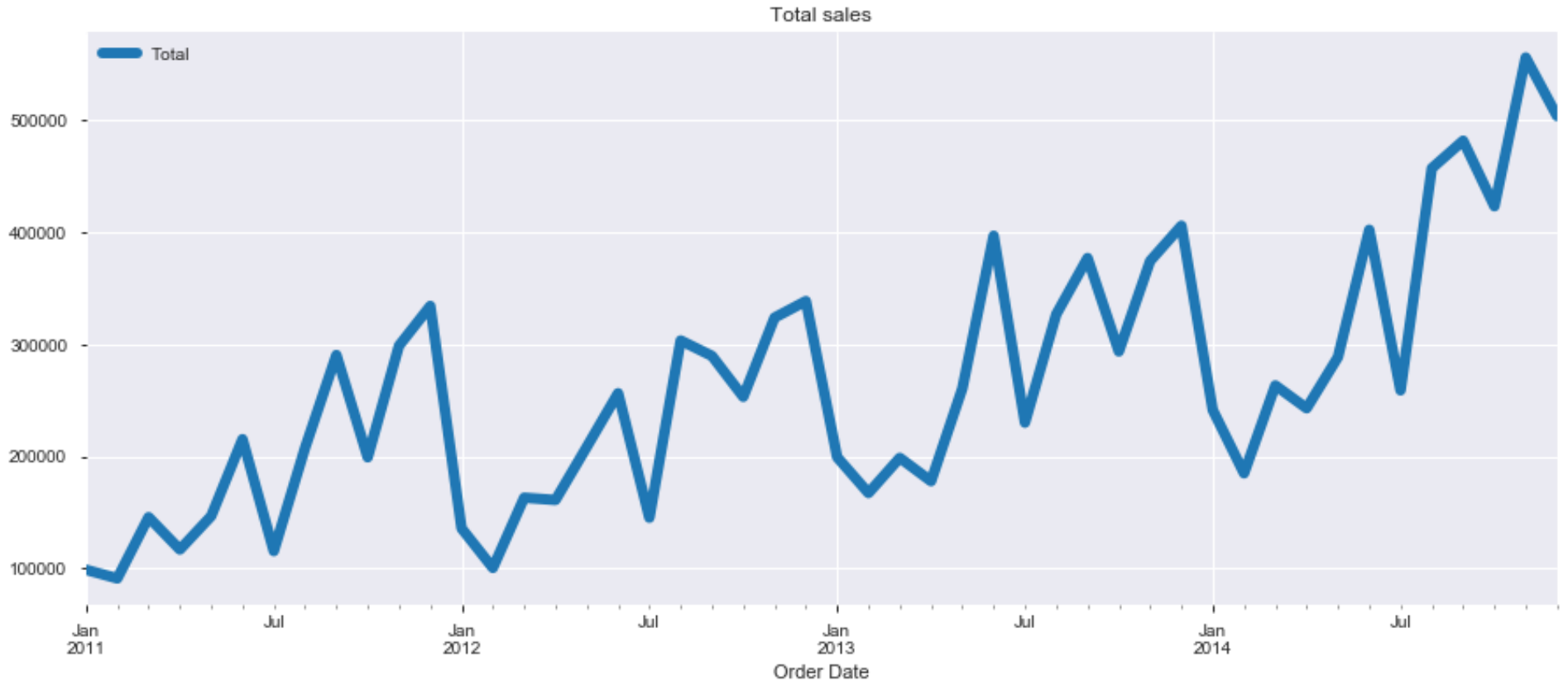
Visualize and analyse time series

Forecast sales

VISUALIZING SALES BY MONTH AND BY CATEGORY



VISUALIZING AGGREGATED SALES BY MONTH

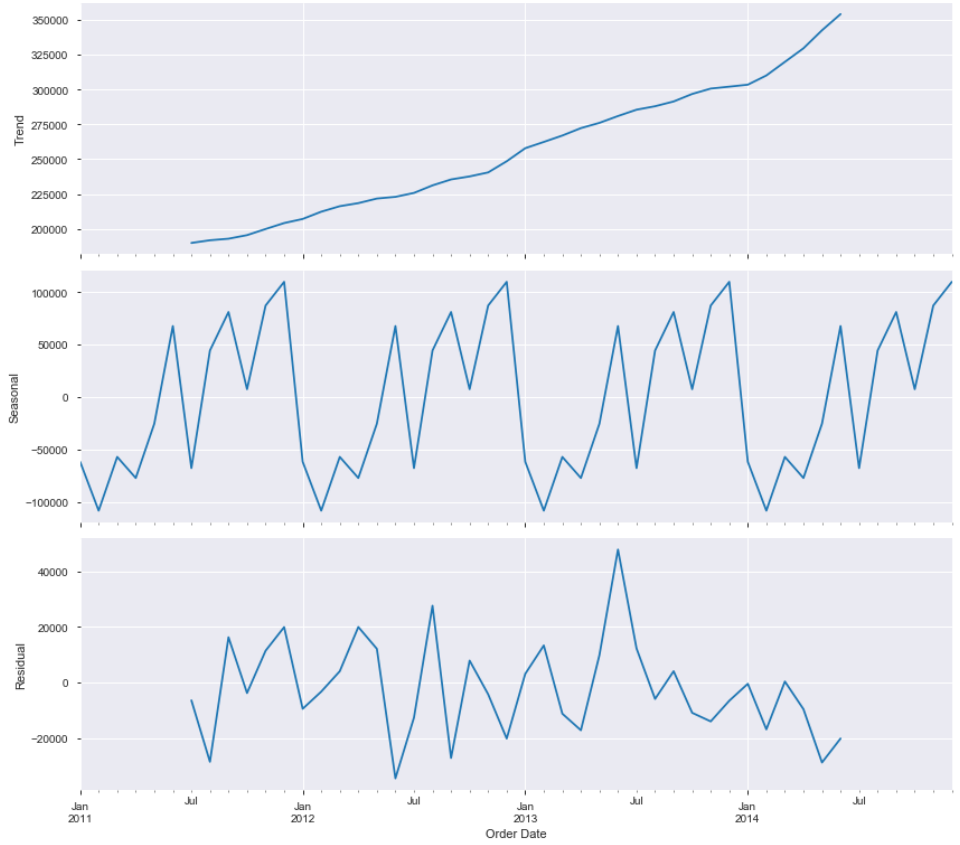


TIME SERIES DECOMPOSITION

Time series decomposition shows a visible strong trend in the top graph

The middle graph shows the seasonality in the sales data

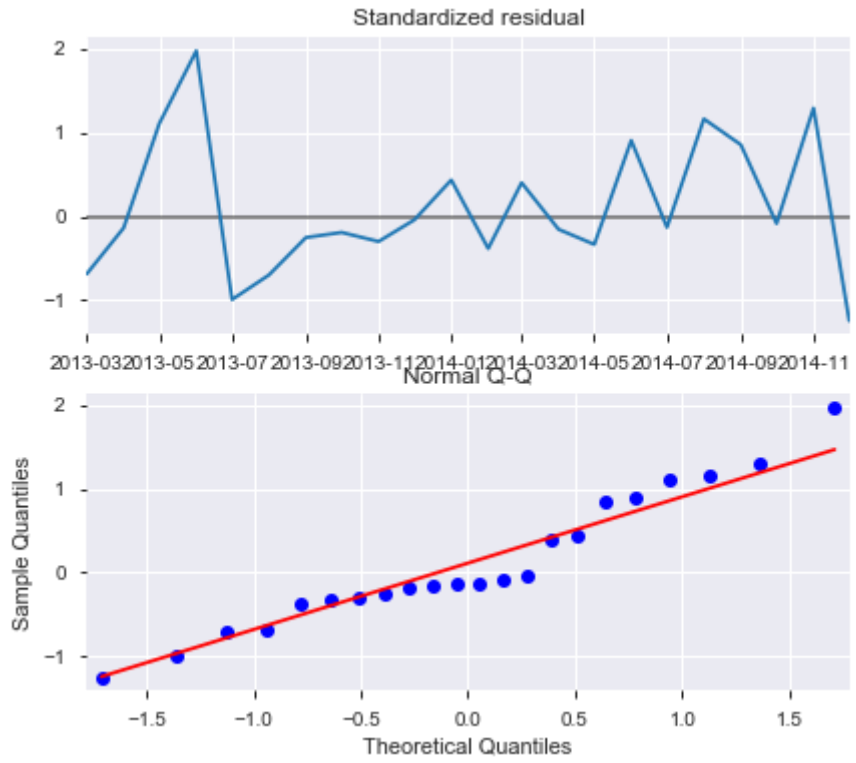
The last graph shows the randomness or the residuals in the time series data



VISUALIZING MODEL RESULTS

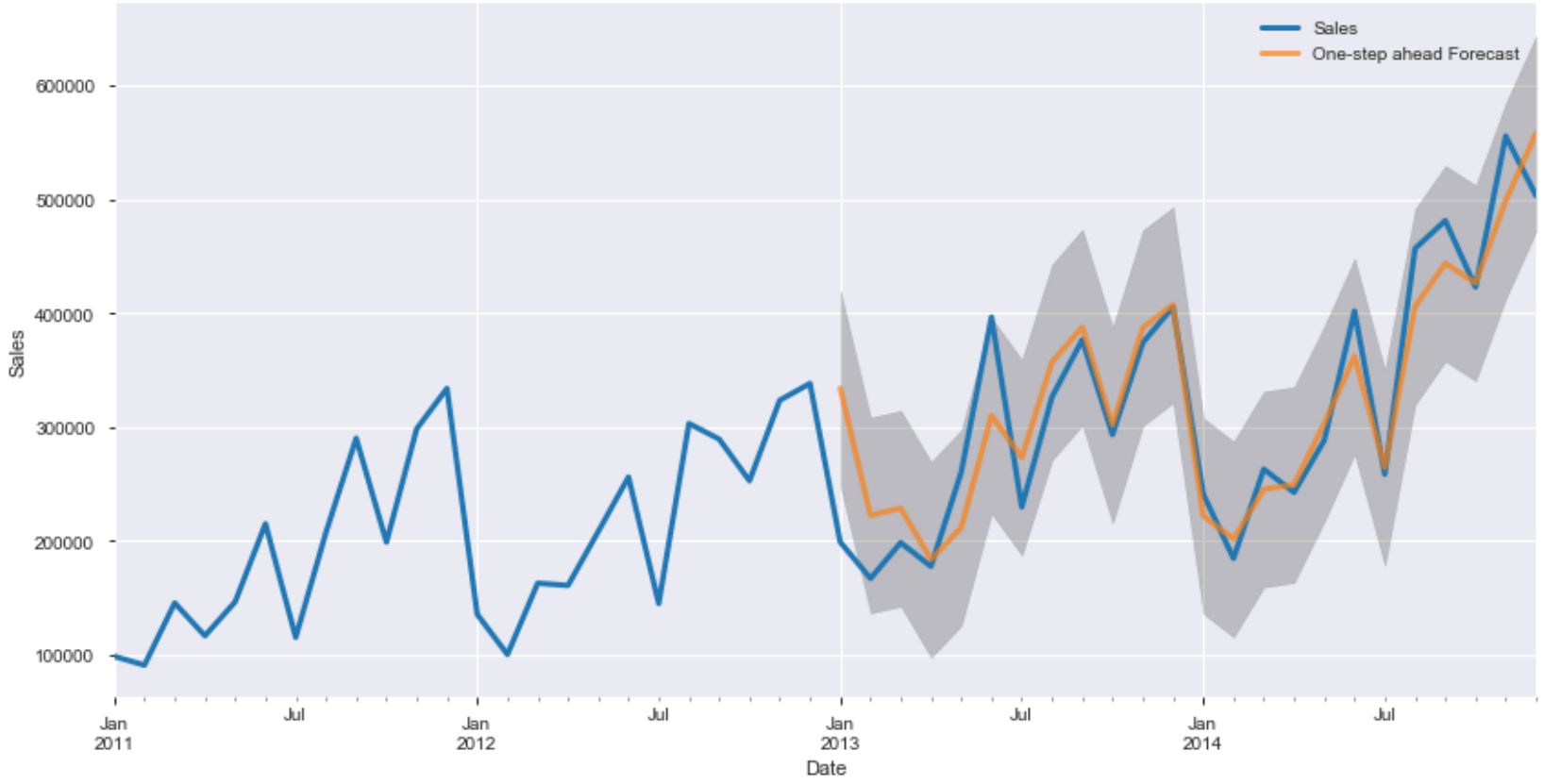
The model seems to be good with the top graph showing residuals almost equally distributed above and below the line

The bottom graph shows the qq plot which shows a fairly normal residual distribution

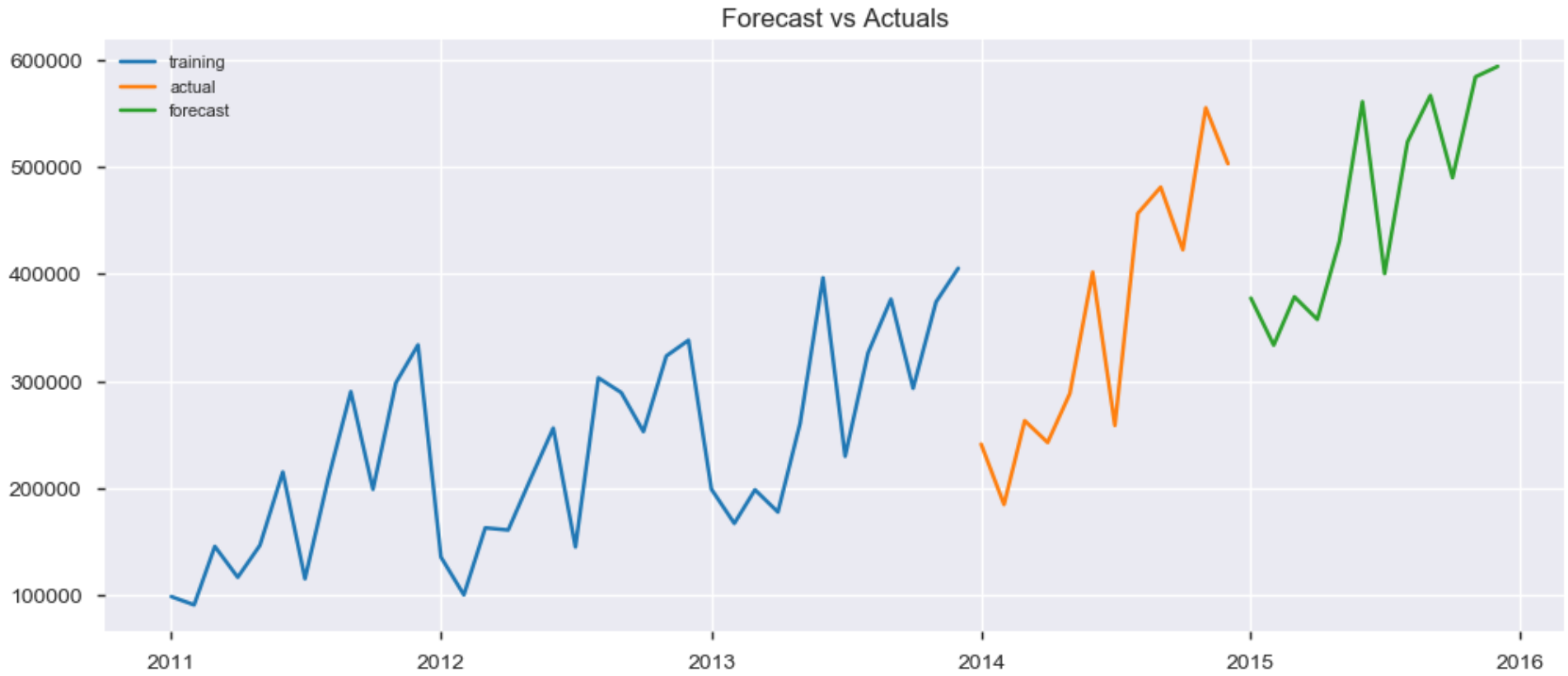


TESTING THE MODEL WITH THE EXISTING DATA

The values returned by model seems to follow closely the values in the dataset



VISUALIZING THE FORECAST



MODEL ACCURACY AND PLANS FOR FUTURE IMPROVEMENTS

My model's accuracy is 63%

Try improving the model accuracy by feeding more time duration

Use alternative methods to model forecast