HW-5-Product-demand-data-set

Pushkar Singh Negi

Ku ID: 2946319

EECS 731: Introduction to Data Science

Notebook: LabProject#5_Product-demand-Forecasting.ipynb
Purpose: Deduced Additional Information, visualization and ARIMA Time series forecasting.

Input raw dataset : Historical Product Demand.csv.zip
Processed dataset : AdditionalInfo#1_NoOfProductEachCategory.csv,
AdditionalInfo#3_WHouse_Category_product.csv,
AdditionalInfo#4_Demans_order_per_day.csv
Jupyter Notebook File name : LabProject#5_Product-demand-Forecasting.ipynb

1. Imported the Historical Product Demand.csv data file in dataframe using pandas.

2. Checked the total non-null entries to handle missing values.

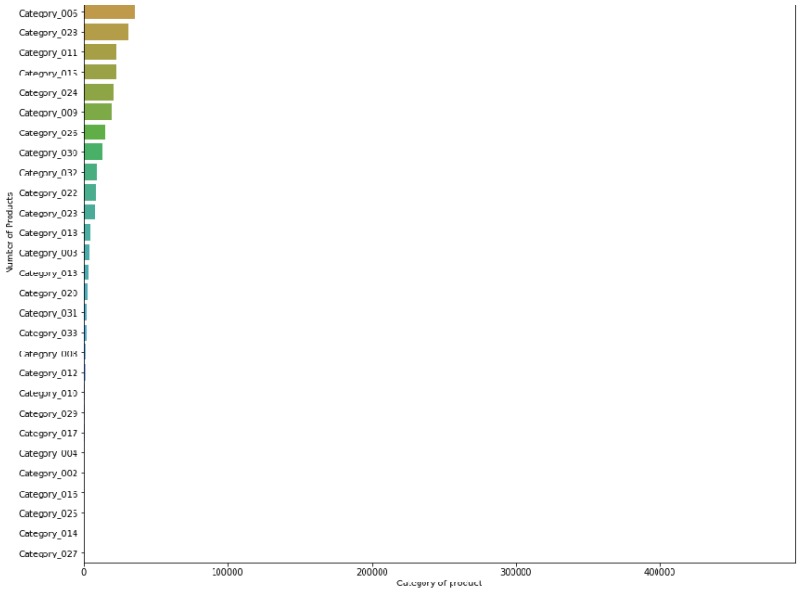
3. Handled the missing values.

4. Analyzed the dataset with the help of various commands, such as finding the total unique warehouse, category and product types.

5. Next, I found out the **Additional Information #1: To find total number of products that belong to each category.**

6. Next, I applied various pandas feature to convert the above result dataset into a panda frame.

7. **Viusalization:** Plotted a graph, number of products against each category of products.



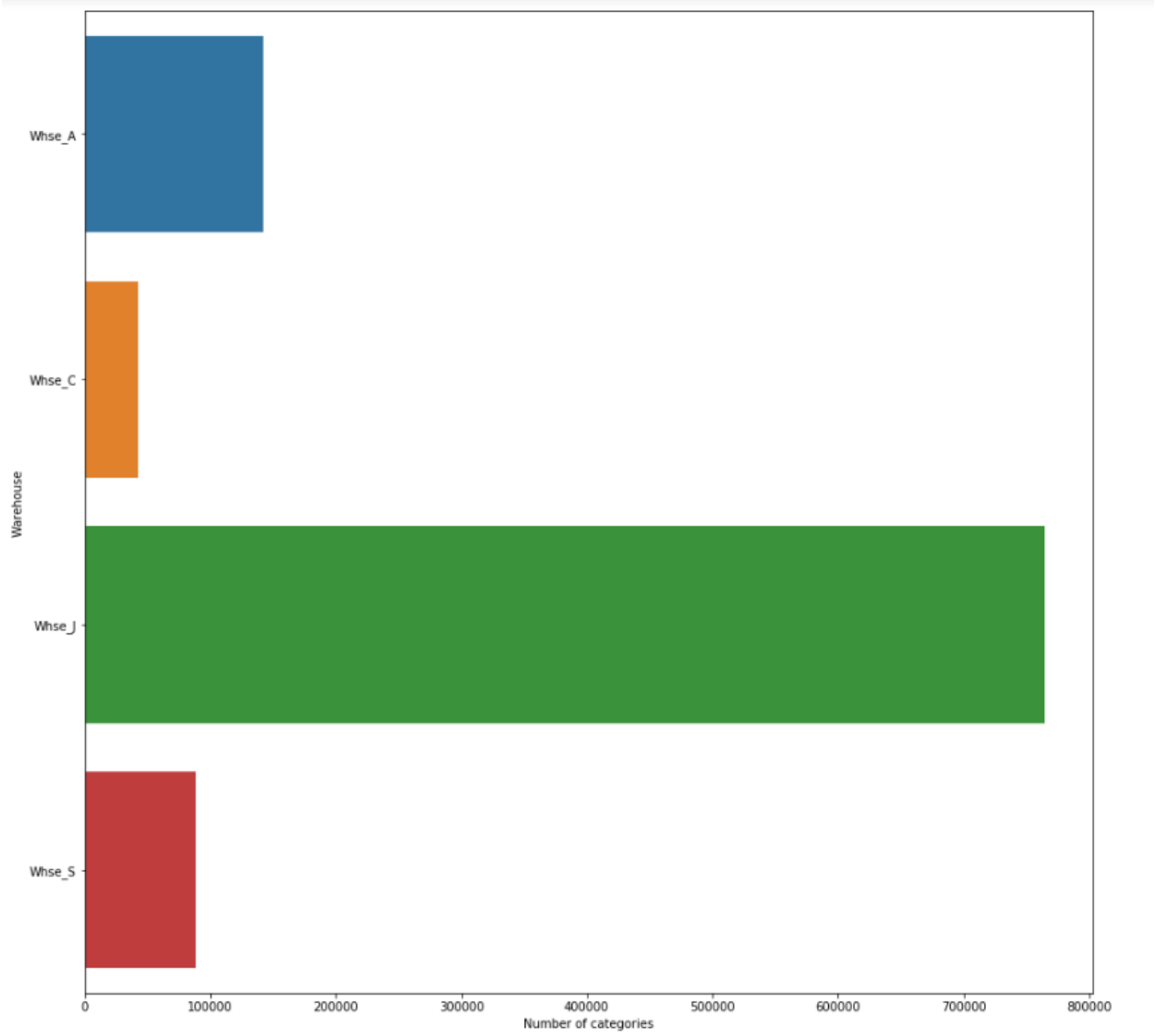
8. Converted the above data into a new data frame (df_prod_ProductPerCategory).

9. Saved the resulted dataframe in a csv
(AdditionalInfo#1_NoOfProductEachCategory.csv.csv)

10. **Additional Information #2:** To find total number of product category that belong to each warehouse.

11. Converted the above new data into a dataframe (df_prod_CategoryPerWHouse).

12. Viusalization: Plotted a graph, number of category that belongs to each warehouse.



13. **Additional Information #3:** For each Warehouse to find corresponding Product_Category and for each product category **find total product. **

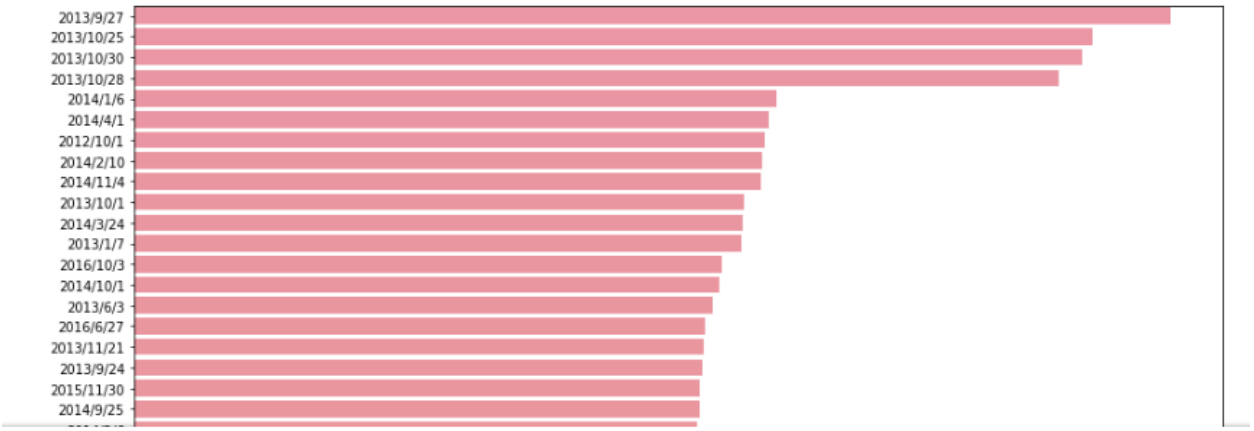
14. Converted the above new data into a dataframe (df_prod_WHouse_Category_product)

15. Saved the resulted dataframe in a csv (**AdditionalInfo#3_WHouse_Category_product.csv**)

16. **Additional Information #4:** To find the date on which maximum and minimum orders where placed.

17. Converted the above new data into a dataframe (df_prod_OrderPerDay)

18. **Visualization**: Date wise order demand distribution



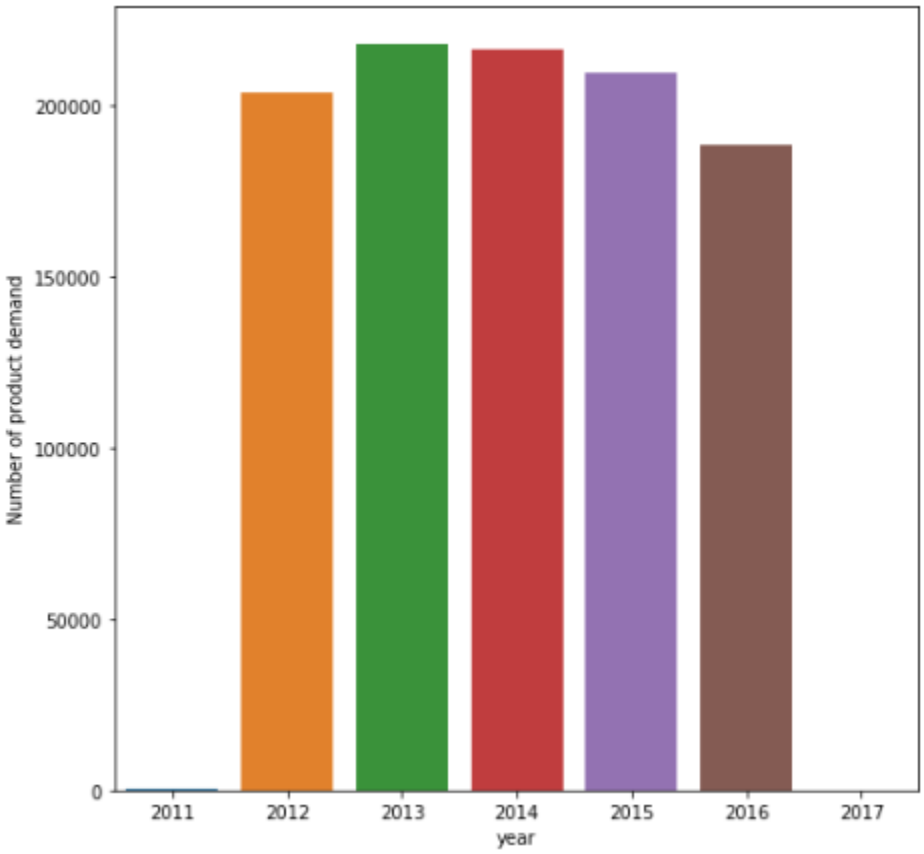
19. **Additional Information #5:** Created 3 new columns namely year, month and date for each of the order placed.

	Product_Code	Warehouse	Product_Category	Date	Order_Demand	year	month	date
0	Product_0993	Whse_J	Category_028	2012/7/27	100	2012	7	27
1	Product_0979	Whse_J	Category_028	2012/1/19	500	2012	1	19
2	Product_0979	Whse_J	Category_028	2012/2/3	500	2012	2	3
3	Product_0979	Whse_J	Category_028	2012/2/9	500	2012	2	9
4	Product_0979	Whse_J	Category_028	2012/3/2	500	2012	3	2

20. Additional Information #6: To find for each year how many orders were placed.

21. Converted the above new data into a dataframe (df_prod_new_Year)

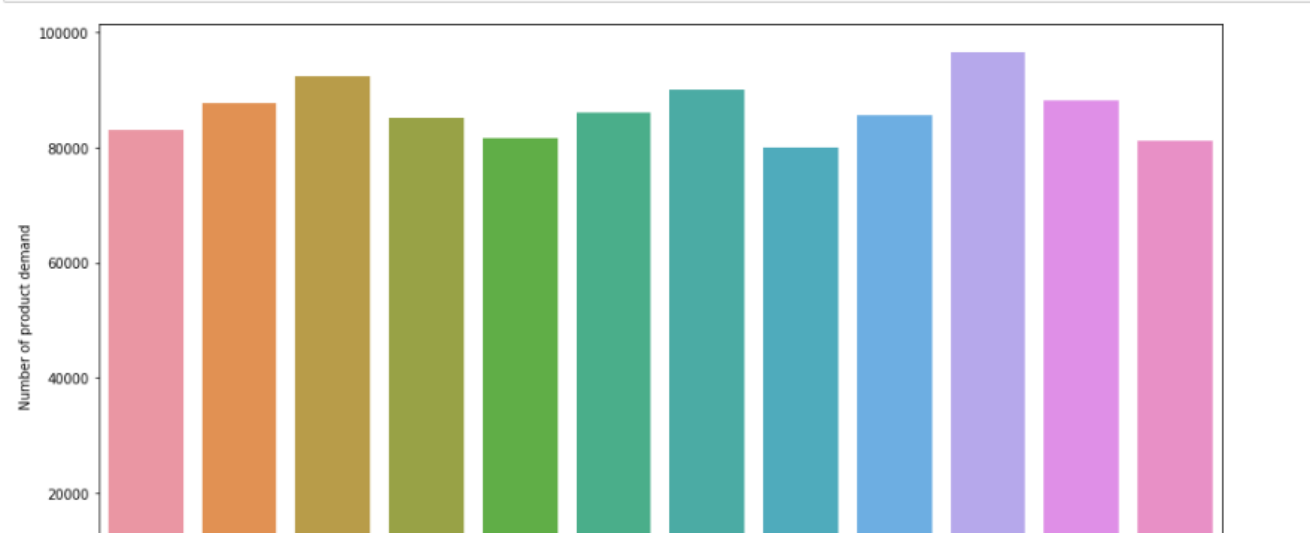
22. **Visualization**: Year wise order demand distribution



23. Additional Information #7: To find for each month how many orders were placed.

24. Converted the above new data into a dataframe (df_prod_new_Month)

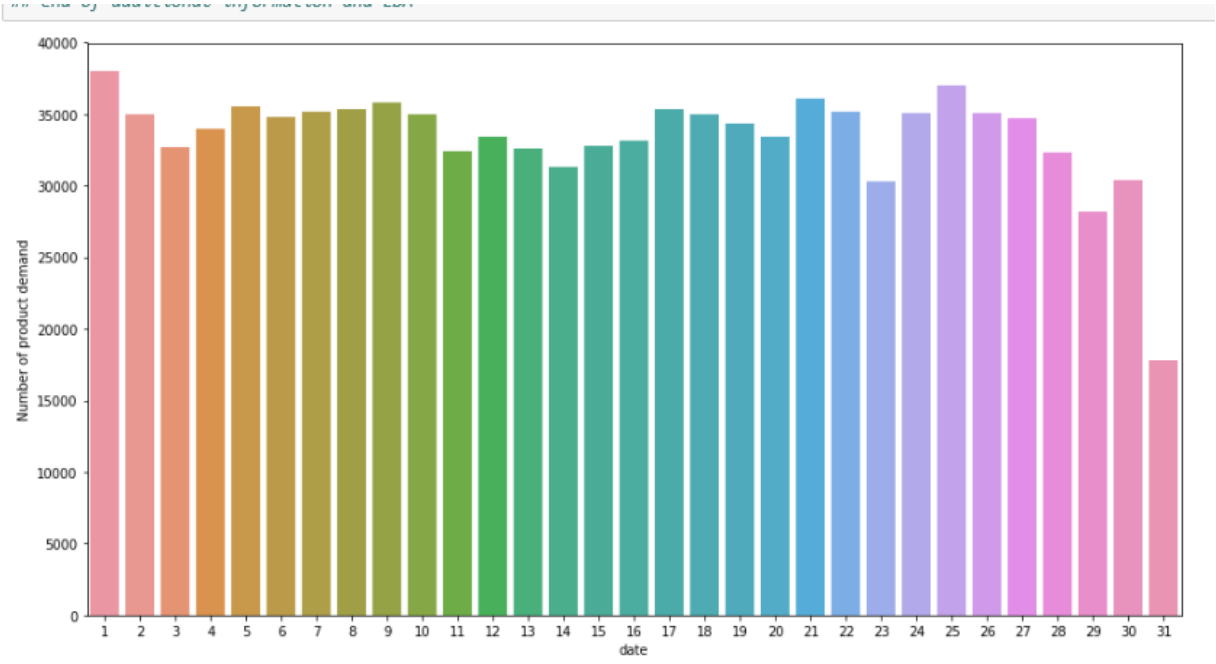
25. **Visualization**: Month wise order demand distribution



26. Additional Information #8: To find for each day how many orders were placed.

27. Converted the above new data into a dataframe (df_prod_new_date)

28. Visualization: Day wise order demand distribution

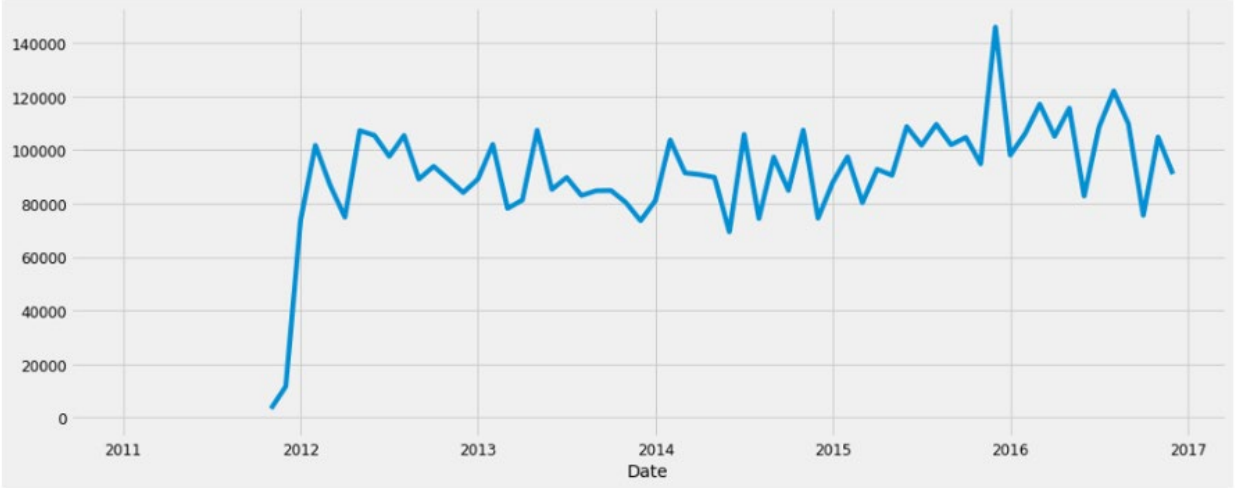


29. My Findings→

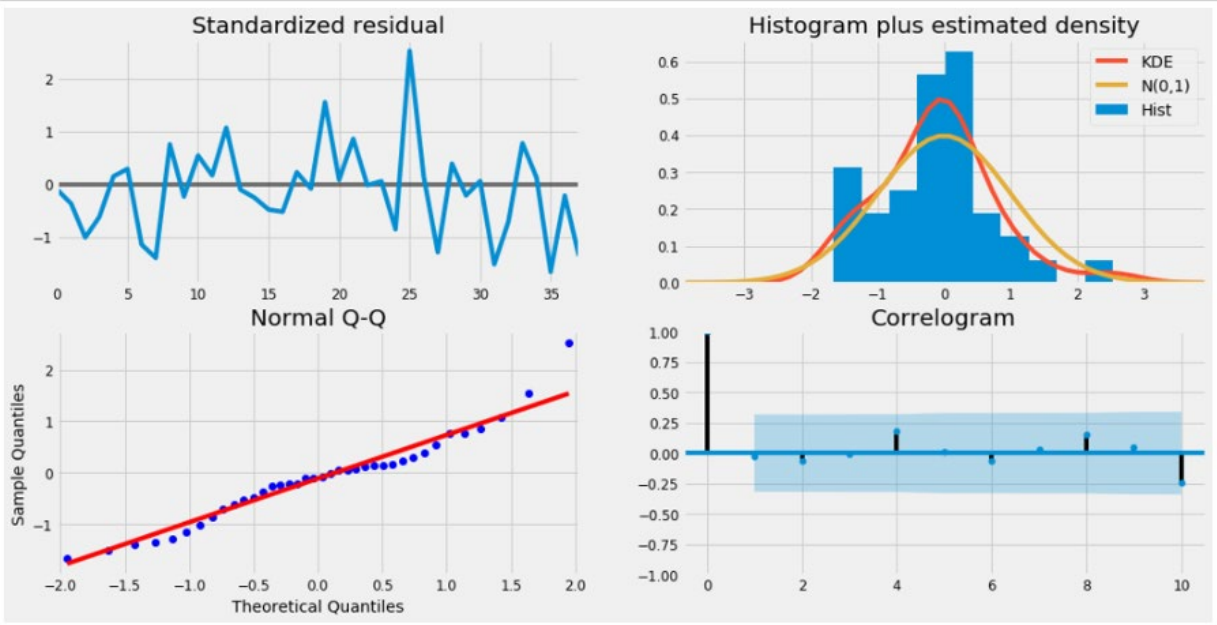
- There are total 2160 unique products.
- The most popular product is Product_1359 with 16936 frequency.
- Only 4 unqique warehouses
- All in all total 33 unique Product_Category
- Category_019 has maximum of 470266 produtcs.
- Whse_J has maximum and Whse_C has minimum products respectively.
- On date 2013/9/27 maximum of 2075 orders were placed.
- In year 2013, maximum number of 218298 orders were placed.
- Most/ maximum of 96619 orders were demand/placed in the month of October.
- On every 1st of month, maximum orders were demand.

ARIMA: Time Series Forecasting

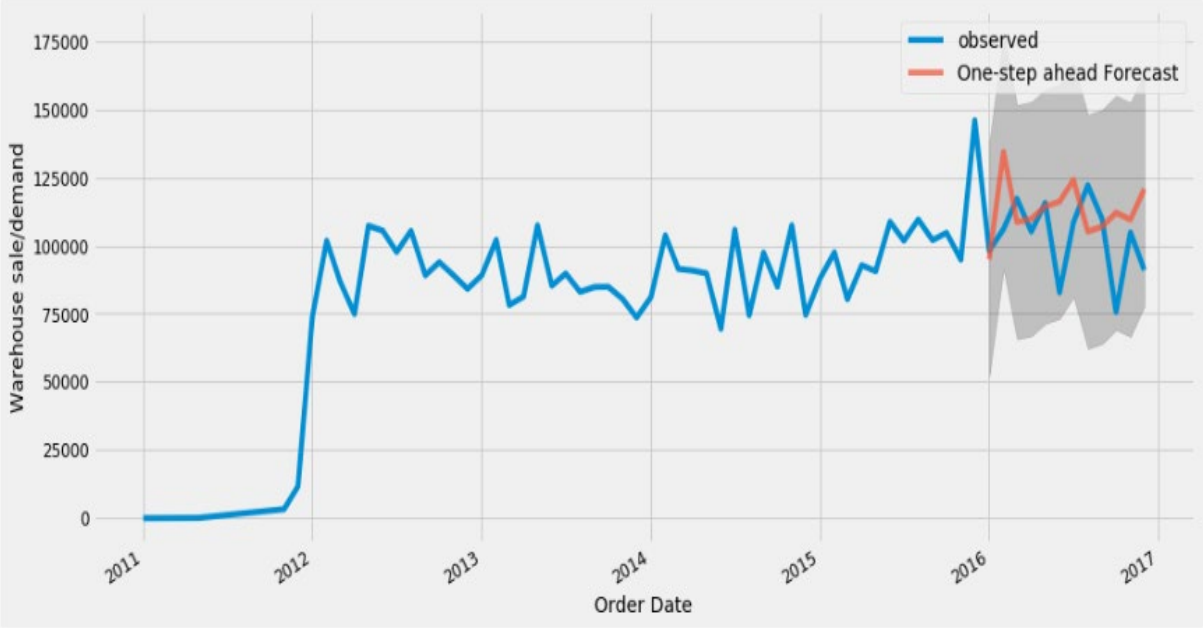
- Forecasted for the warehouseA demand.
- Removed the unnecessary columns from the dataframe.
- Indexing with Time Series Data
- Visualizing WarehouseA Sales Time Series Data



-
- Visualize the data using time-series decomposition: decompose our time series into three distinct components: trend, seasonality, and noise
- ARIMA, which stands for Autoregressive Integrated Moving Average.
- ARIMA(p, d, q): parameters account for seasonality, trend, and noise in data
- Parameter Selection for our warehouseA demand ARIMA Time Series Model.



- Fitting the ARIMA model
- Validating forecasts



- Mean Squared Error of our forecasts