

MAIN
ACADEMY

Sales forecasting using machine learning models

K. Shportko

29.06.2021

Contents

- 1. Some descriptive statistics
 - a. General data set overview
 - b. Preliminary forecasting of sales
- 2. ABC analysis by Products
 - a. sklearn models sales forecasting
 - b. sktime models sales forecasting
- 3. ABC analysis by Customers
 - a. sktime models sales forecasting

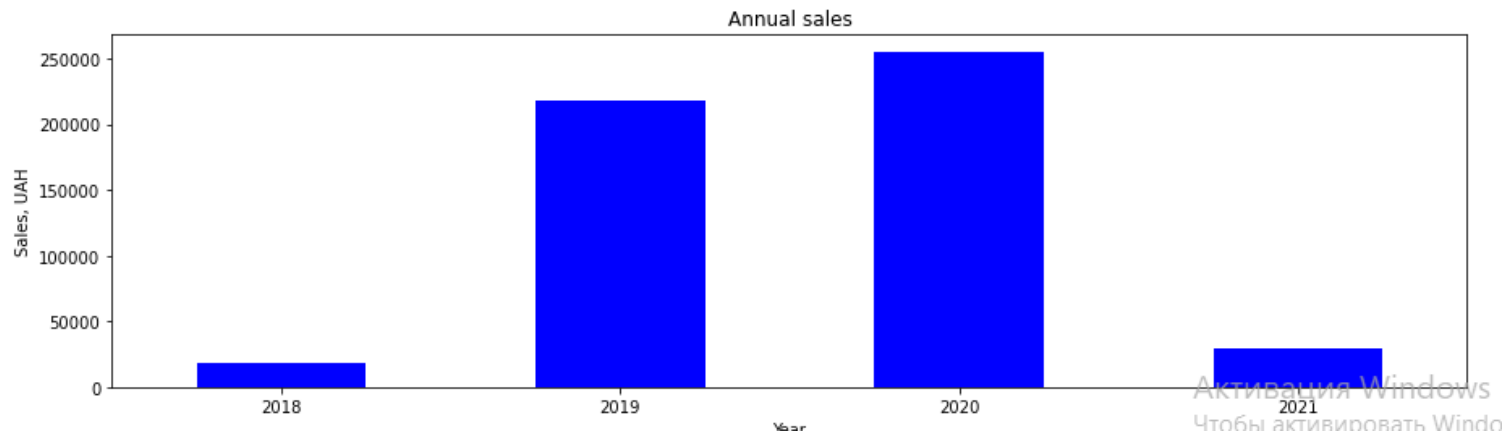
1. Some descriptive statistics

Data set:

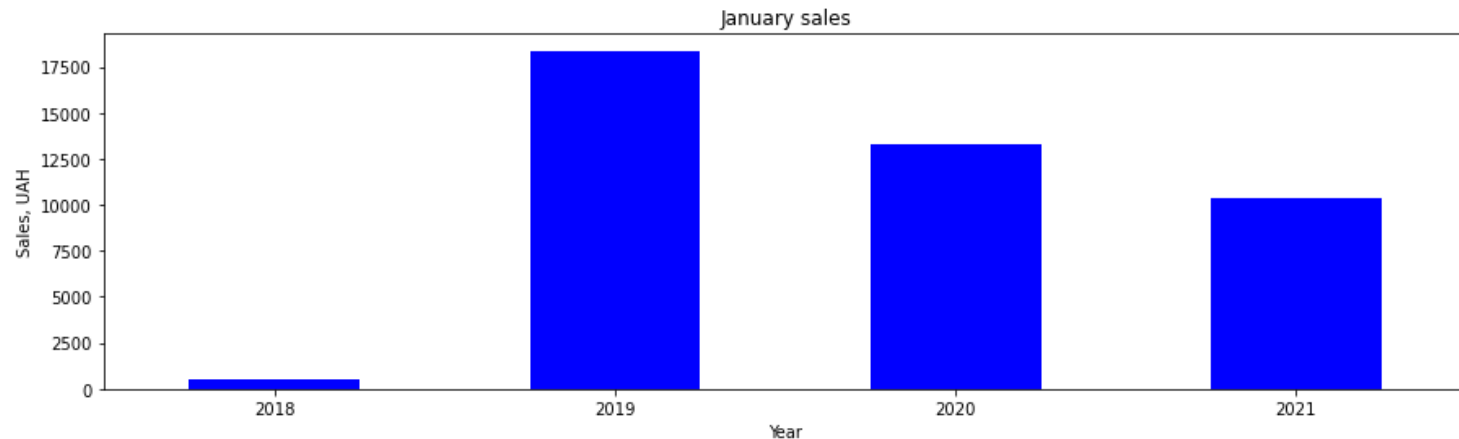
- 27994 entries;
- 307 products;
- 704 customers;
- Period of time: 3 years, 10 months, 21 days.

1	Product	Customer	Order	Qty	Sales	Period
2	Product 1	Customer 1	78542	1	3.34	01.07.2019 09:25:47
3	Product 2	Customer 2	78554	1	6.09	02.07.2019 09:42:14
4	Product 3	Customer 2	78554	1	3.19	02.07.2019 09:42:14
5	Product 4	Customer 2	78554	1	6.30	02.07.2019 09:42:14
6	Product 5	Customer 3	78578	1	9.56	01.07.2019 14:40:51
7	Product 6	Customer 3	78578	1	2.79	01.07.2019 14:40:51
8	Product 7	Customer 4	78593	1	28.94	01.07.2019 10:49:20
9	Product 8	Customer 4	78593	1	30.96	01.07.2019 10:49:20
10	Product 9	Customer 5	78597	1	0.96	01.07.2019 10:41:27
11	Product 10	Customer 5	78597	1	1.83	01.07.2019 10:41:27
12	Product 11	Customer 5	78597	1	1.37	01.07.2019 10:41:27
13	Product 2	Customer 6	78598	1	5.92	19.06.2019 23:59:59
14	Product 3	Customer 6	78598	1	3.10	19.06.2019 23:59:59
15	Product 4	Customer 6	78598	1	6.12	19.06.2019 23:59:59
16	Product 15	Customer 7	78603	1	29.68	01.07.2019 10:33:12
17	Product 16	Customer 7	78603	1	17.86	01.07.2019 10:33:12
18	Product 17	Customer 8	78618	1	13.55	01.07.2019 13:05:15
19	Product 18	Customer 8	78618	1	17.94	01.07.2019 13:05:15

1. Some descriptive statistics

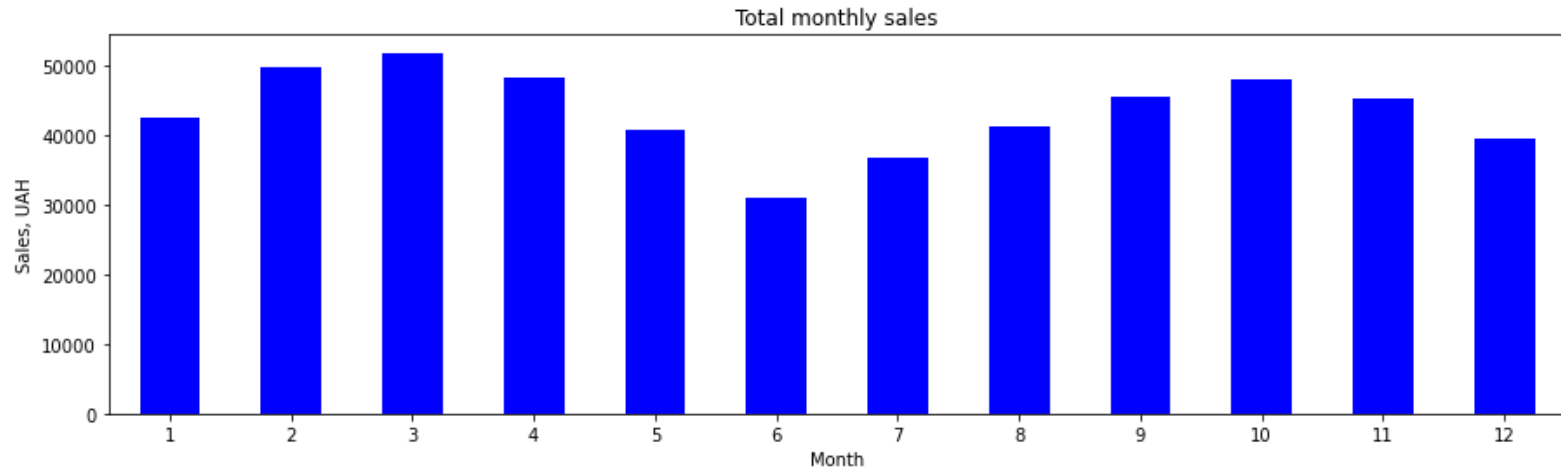


Conclusion 1. Annual sales grow starting from 2018 till 2020.

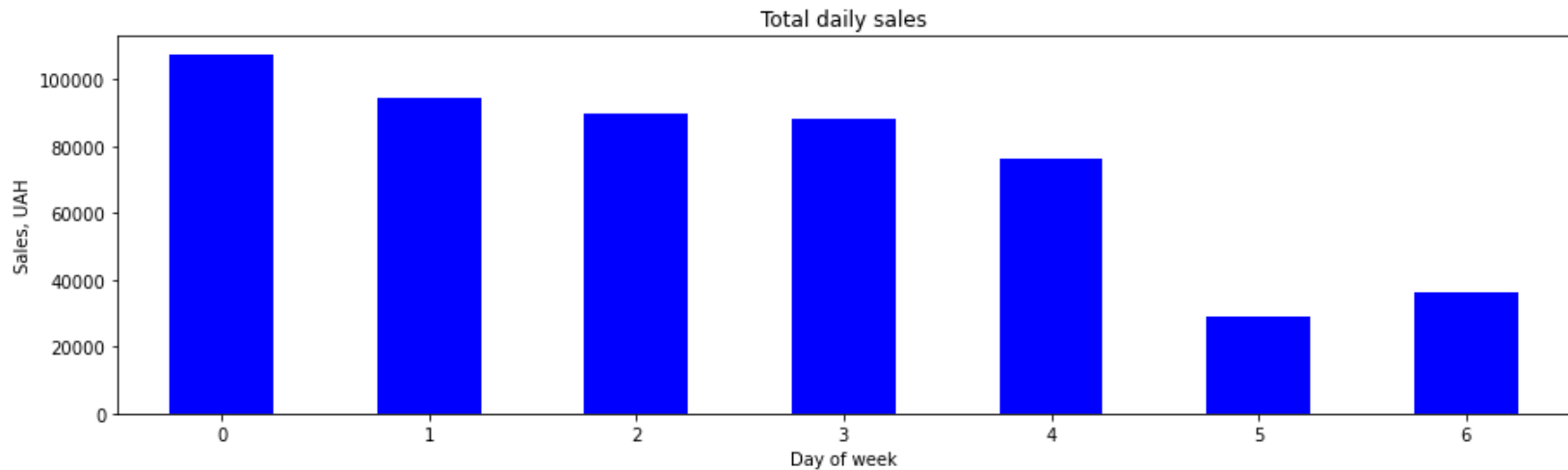


Conclusion 2. Despite the positive trend in annual sales in 2018-2020, January sales illustrate negative trend in 2019-2021.

1.a. General data set overview

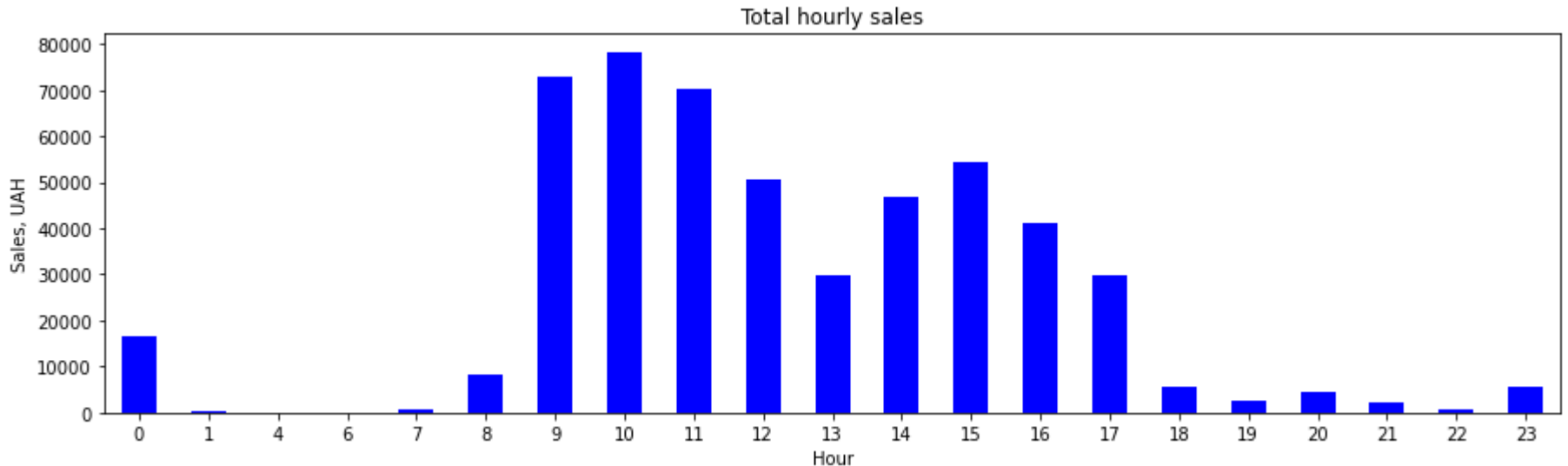


Conclusion 3. We clearly observe dependence of the sales on the months



Conclusion 4. At the weekends we observe lower sales than on workdays. Within the workday sales decrease from Monday to Friday

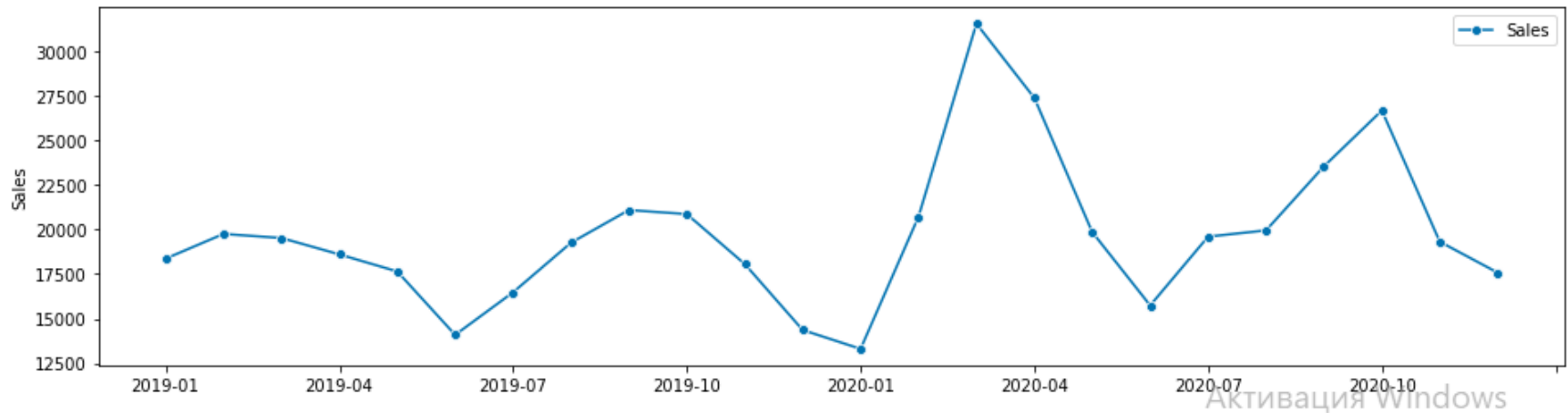
1.a. General data set overview



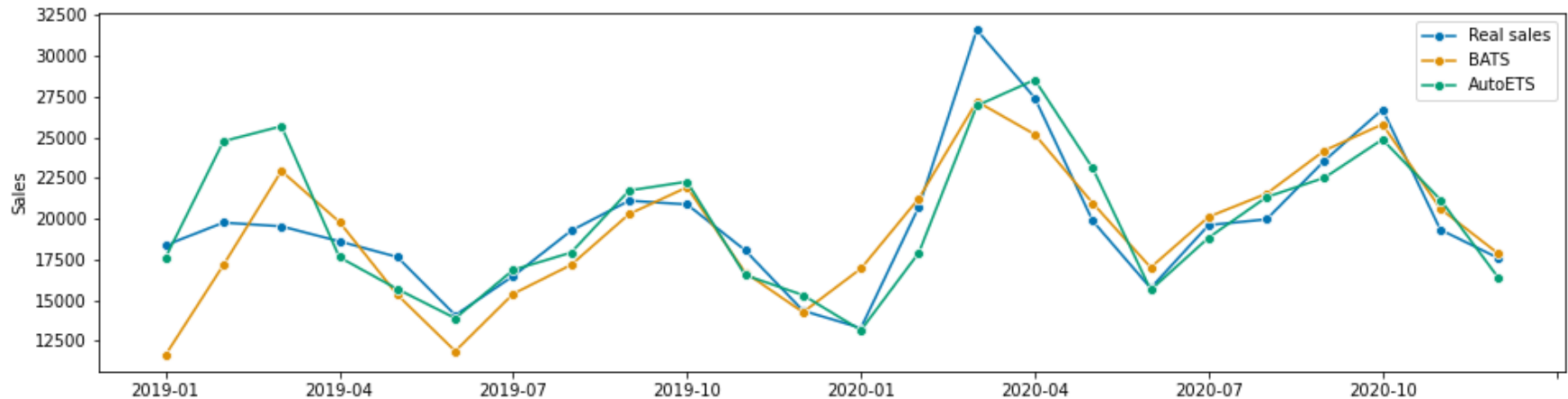
Conclusion 5. The hourly sales can be divided into 3 ranges: maximum sales before the afternoon with peak at 10 a.m., afternoon with peak at 3 p.m., and almost negligible sales in the period from 7 p.m. till 7 a.m. However, there is a local peak at midnight.

1.b. Preliminary forecasting of sales

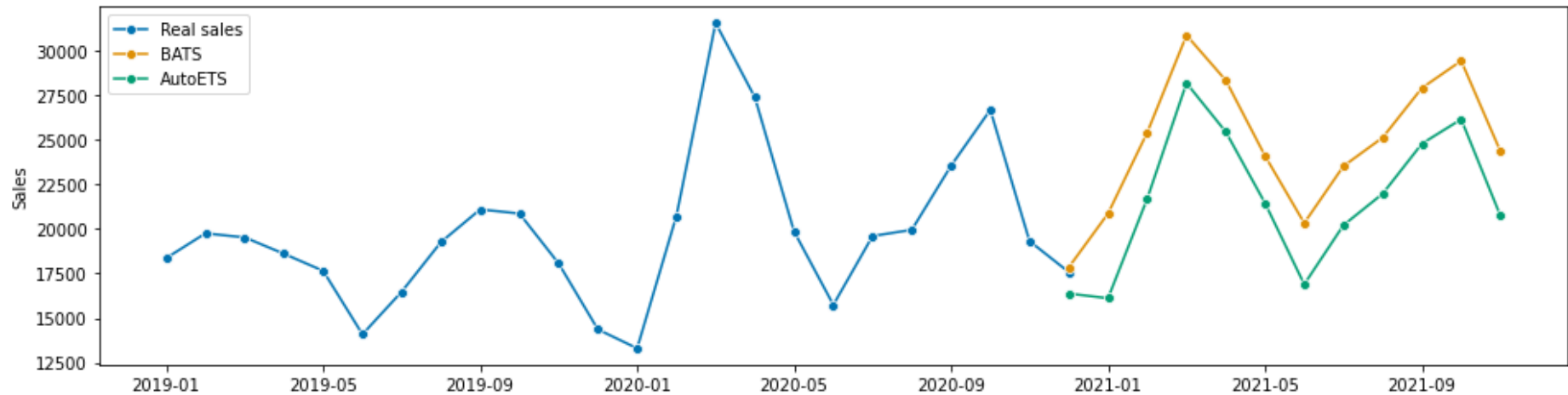
- To get the whole image of the trends in sales.



- Sktime library: AutoETS, BATS models.



1.b. Preliminary forecasting of sales

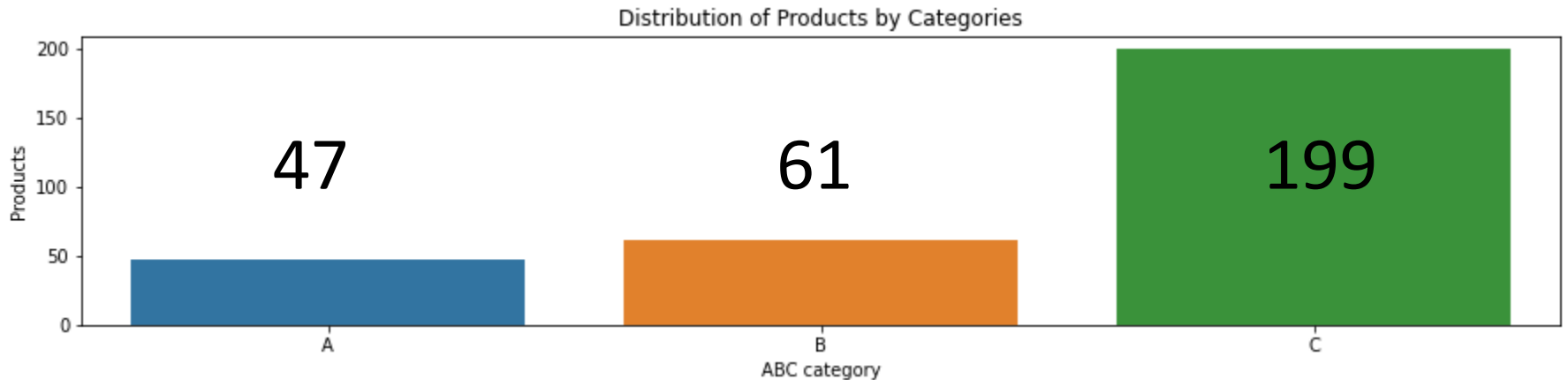


Conclusions.

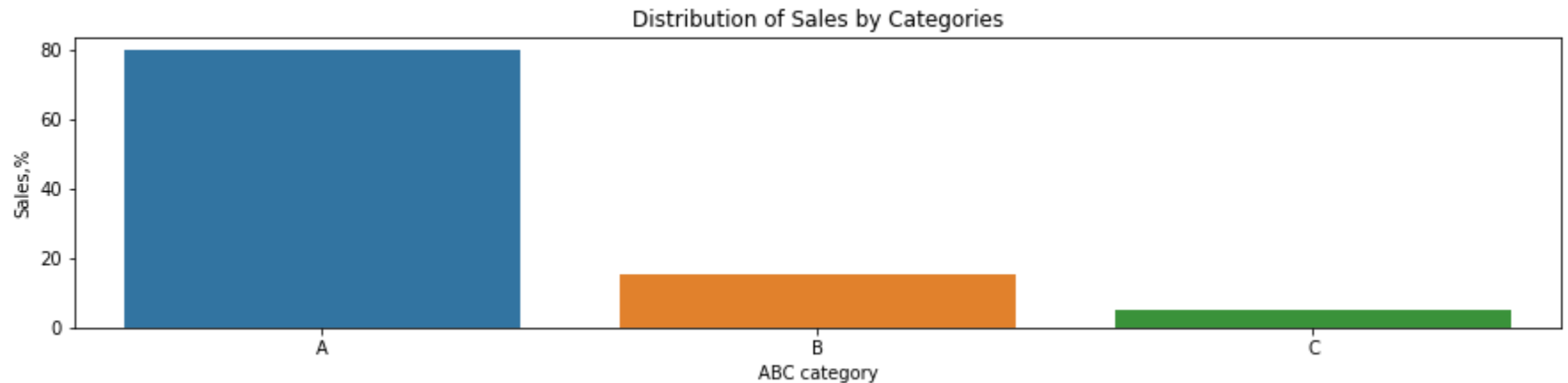
- In the obtained forecast we can see the general trend.
- However, 2 models applied provided different absolute values.
- Let us look in more details.

2. ABC analysis by Products

- What and how we sell

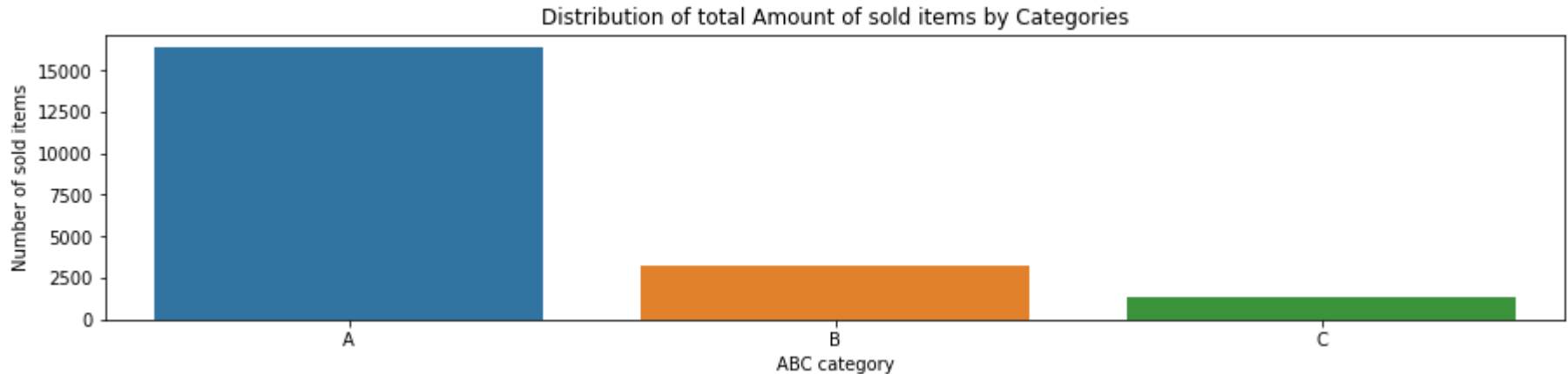


47 Products provide 80% of sales



2. ABC analysis by Products

High sales of products from the category A is not caused by high prices, but due to the amount of sold items



ABC (Products) analysis conclusions:

- 1. There are 47 products in the category A, which is 15% from the total number of products.
- 2. The category A provides almost 80% of annual sales, and, therefore, these products are of high importance for the company.
- 3. The high amount of sales in category A is not due to the high price of the Products, but because of high numbers of sold items.
- 4. We need to forecast the sales in the category A, which will be done in the next section

2.a. sklearn models sales forecasting

Category A products:

- Linear regression 0.806
- Random forest 0.942
- XGBoost 0.999
- Voting regressor ensemble 0.956



2.a. sklearn models sales forecasting



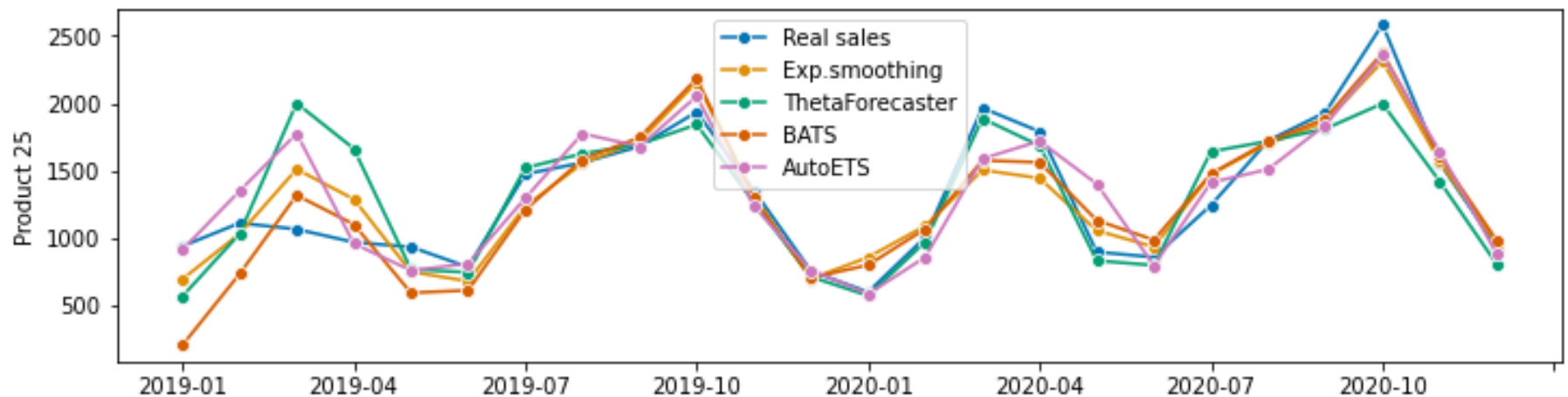
Conclusions:

1. We focused on the Products from A category.
2. For the sales of selected product we applied machine learning algorithms, such as Linear regression, Random forest, and XGBoost, as well as the Voting Regressor ensemble.
3. The obtained forecast demonstrates periodicity in the sales with positive inclination.

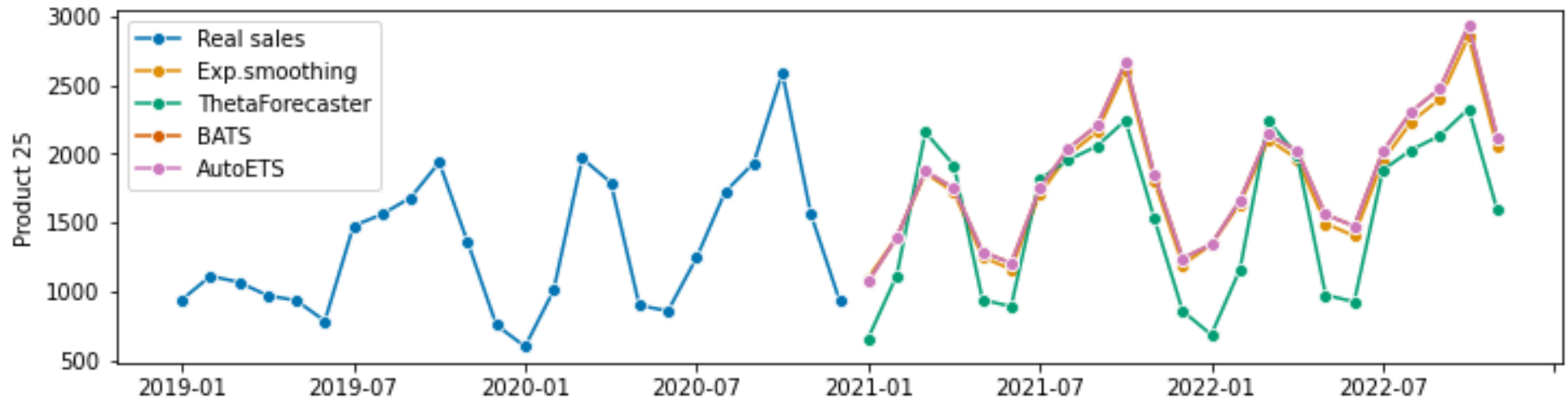
2.b. sktime models sales forecasting

Category A products:

- Exponential smoothing: 0.138;
- ThetaForecaster: 0.141;
- BATS: 0.194;
- AutoETS: 0.118.



2.b. sktime models sales forecasting

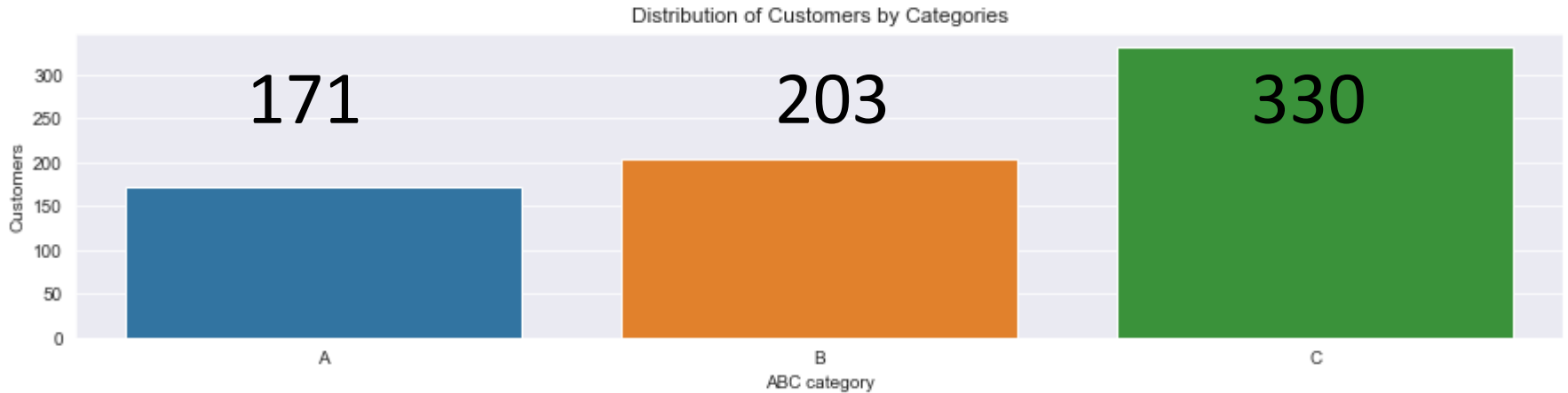


Conclusions:

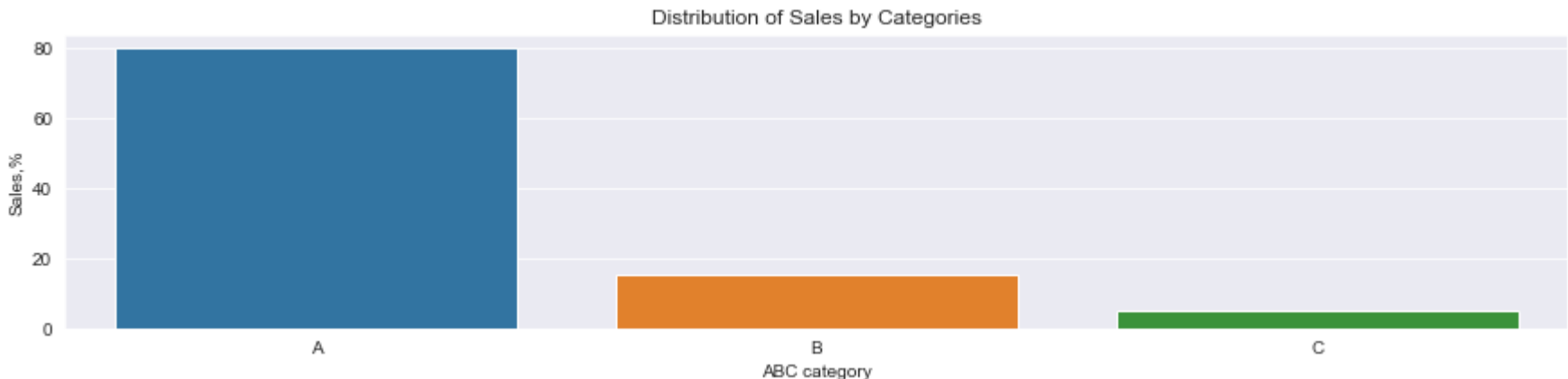
1. Again we focused on the Products from A category.
2. For our analysis we selected 1 Product from the group. For the rest products the procedure might be done in the same way.
3. Here we used models offered by sktime library, such as Exponential smoothing, Thetaforecaster, BATS, and Auto ETS. These models are aimed to work with time series, which is our case.
4. Qualitative analysis of the obtained forecasts shows the presence of periodic sales maxima with general positive inclination. These results are in the good agreement with forecasts obtained in the previous section.
5. Despite the similarity in learning results, models provided the different forecasts. Exponential smoothing, BATS, and Auto ETS provide consistent results. The forecast of Thetaforecaster differs by absolute values.

3. ABC analysis by Customer

- Who and how buys



171 Customers (24%) provide 80% of sales



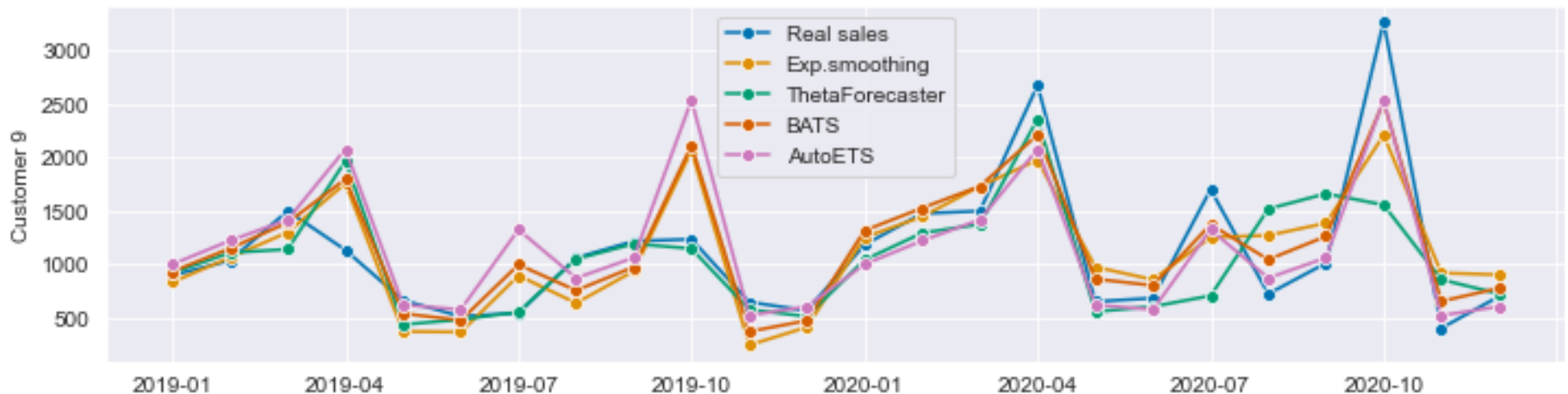
3. ABC analysis by Customer

- **ABC (Customers) analysis conclusions:**
- **1. There are 171 customers in the category A, who is 24% from the total number of customers.**
- **2. The category A provides almost 80% of annual sales, and, therefore, these customers are of high importance for the company.**
- **3. We need to focus on forecasting of sales provided by these customers.**

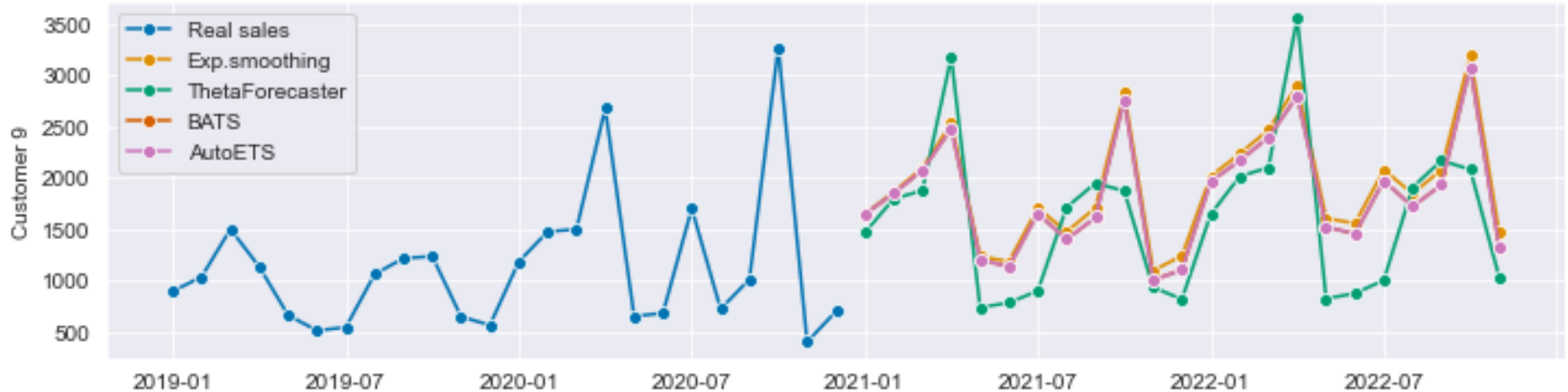
3.a sktime models sales forecating

Category A Customers:

- Exponential smoothing: 0.344;
- ThetaForecaster: 0.248;
- BATS: 0.245;
- AutoETS: 0.219.



3.a sktime models sales forecating



Conclusions:

- 1. Here we focused on the Customers from A category.
- 2. For our analysis we selected 1 Customer from the group. For the rest customers the procedure might be done in the same way.
- 3. Here we used models offered by sktime library, such as Exponential smoothing, Thetaforecaster, BATS, and Auto ETS. These models are aimed to work with time series, which is our case.
- 4. Qualitative analysis of the obtained forecasts shows the presence of periodic sales maxima with general positive inclination. These results are in the good agreement with forecasts obtained in the previous section.
- 5. Despite the similarity in learning results, models provided the different forecasts. Exponential smoothing, BATS, and Auto ETS provide consistent results. The forecast of Thetaforecaster differs by absolute values.

General conclusions

- In this work we built a forecast for the sales using the given data set.
- In the data set we found some periodic dependencies and we performed preliminary modelling of sales.
- To figure out what products and what customers provide the most impact, we performed ABC analysis by these parameters.
- For the most important products and customers we built sales forecasts using various models of machine learning, which provided us with consistent results.

Thank you for your attention!!