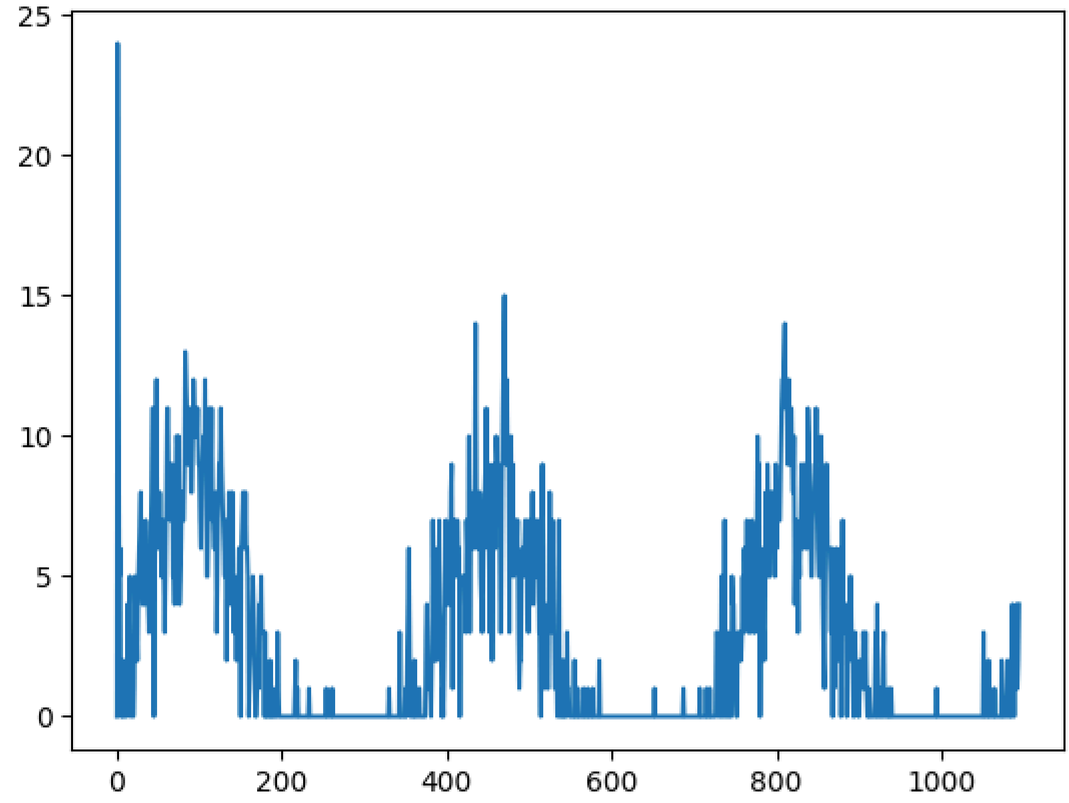
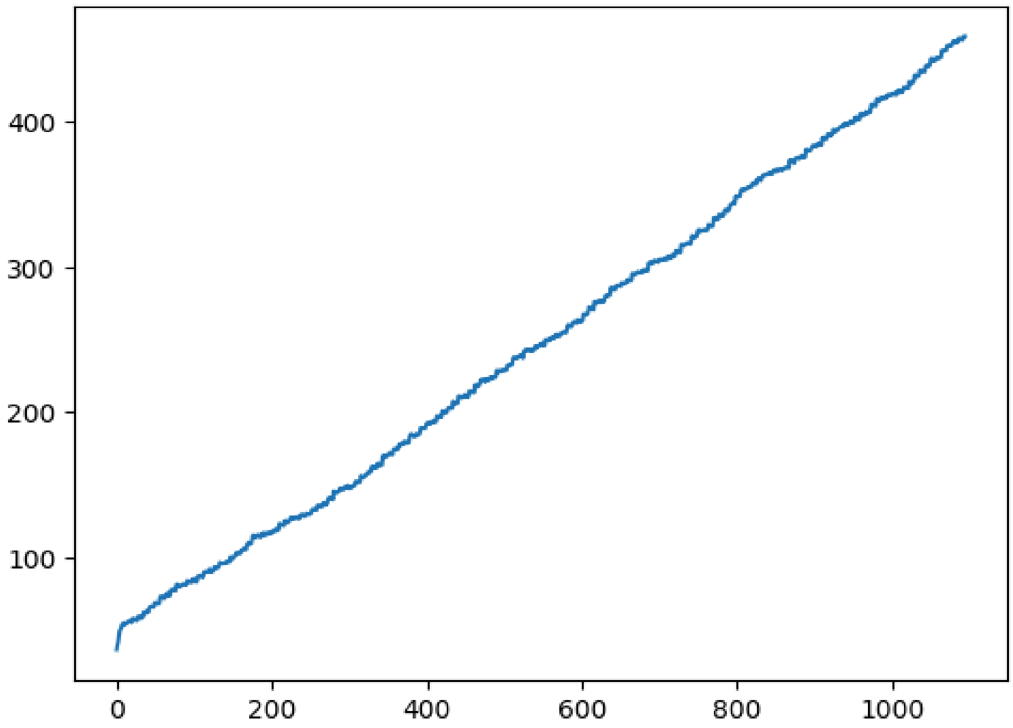
1. Present a thorough analysis of the usage data and your findings

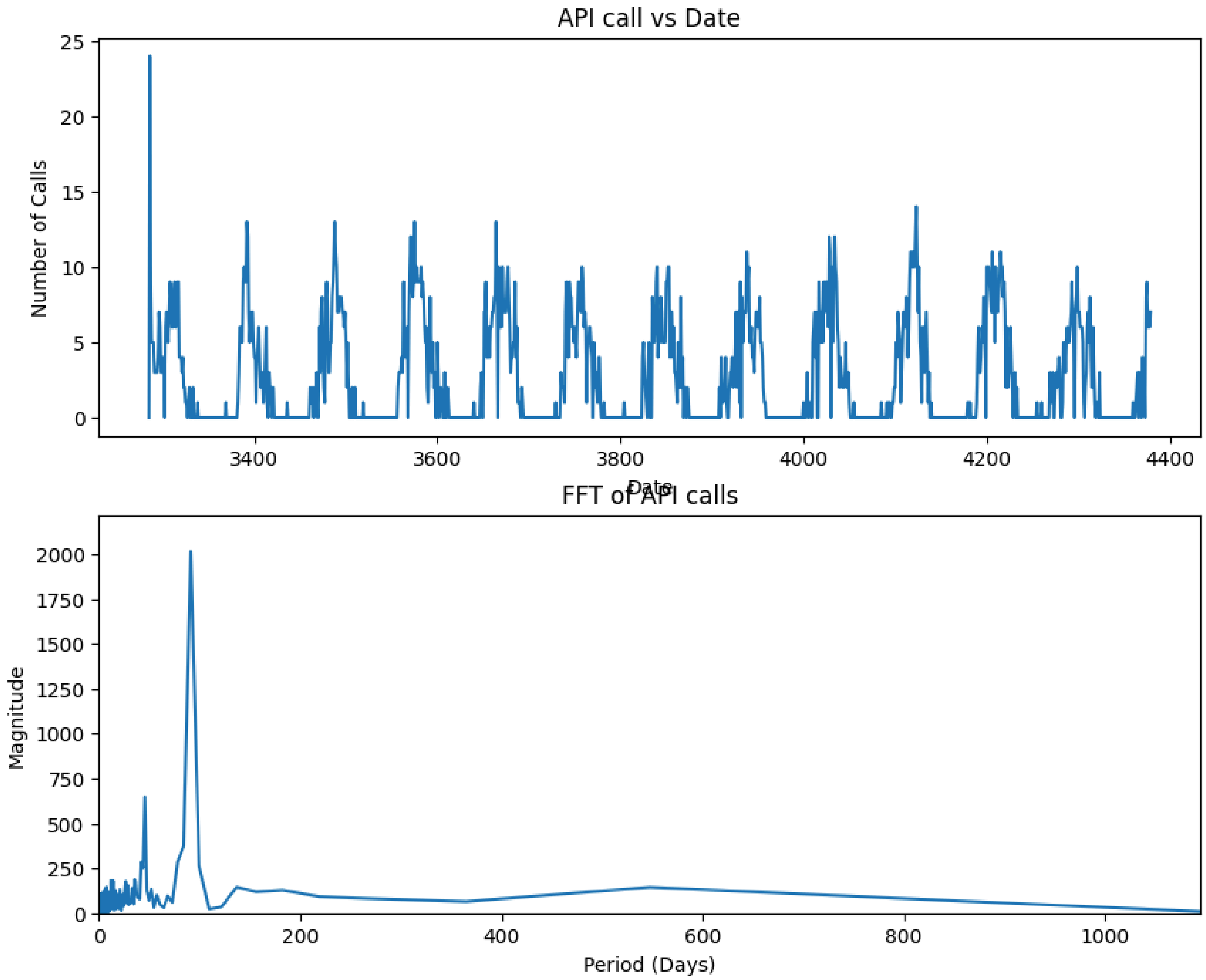
After loading the data, I did exploratory data analysis (EDA) and visualized the number of API calls vs. date for each customer ID. We noticed a cyclic pattern for each customer with varying periods. To further analyze this data, we used the Fast Fourier transform (FFT), which shows the time period of each customer with its spike in magnitude. After this, we did a plot for the memory usage vs. date for each customer, and the plot came out to be linear, so we will use different algorithms for both the API calls and memory usage prediction. For forecasting, I have combined the total API calls and memory usage, respectively, because when we have data on the consumption of each day for three years, we can predict it for the upcoming days. The memory storage plot simply shows that memory storage is increasing linearly with time but api calls plot does not showing any simple pattern but monthly plot is telling that in start of year there is highest api calls but in end of year the api calls are less. So we are generating more revenue by api calls in starting 6 months as compared to last 6 months of year that can also be seen by 6 months of usage plots.



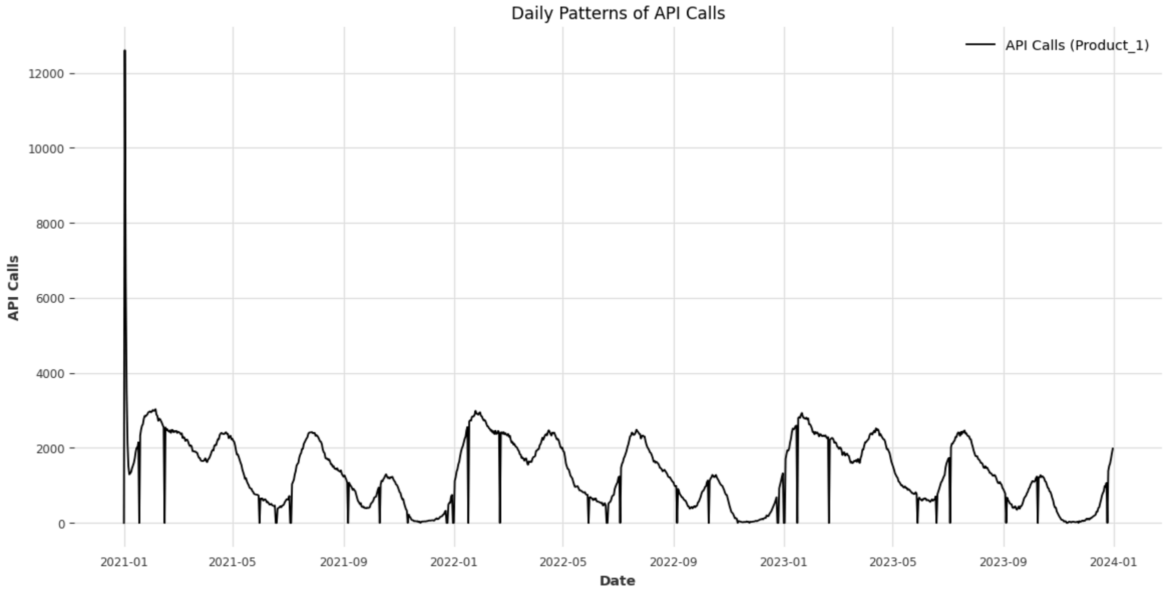
Customer 1 data plot for API_calls

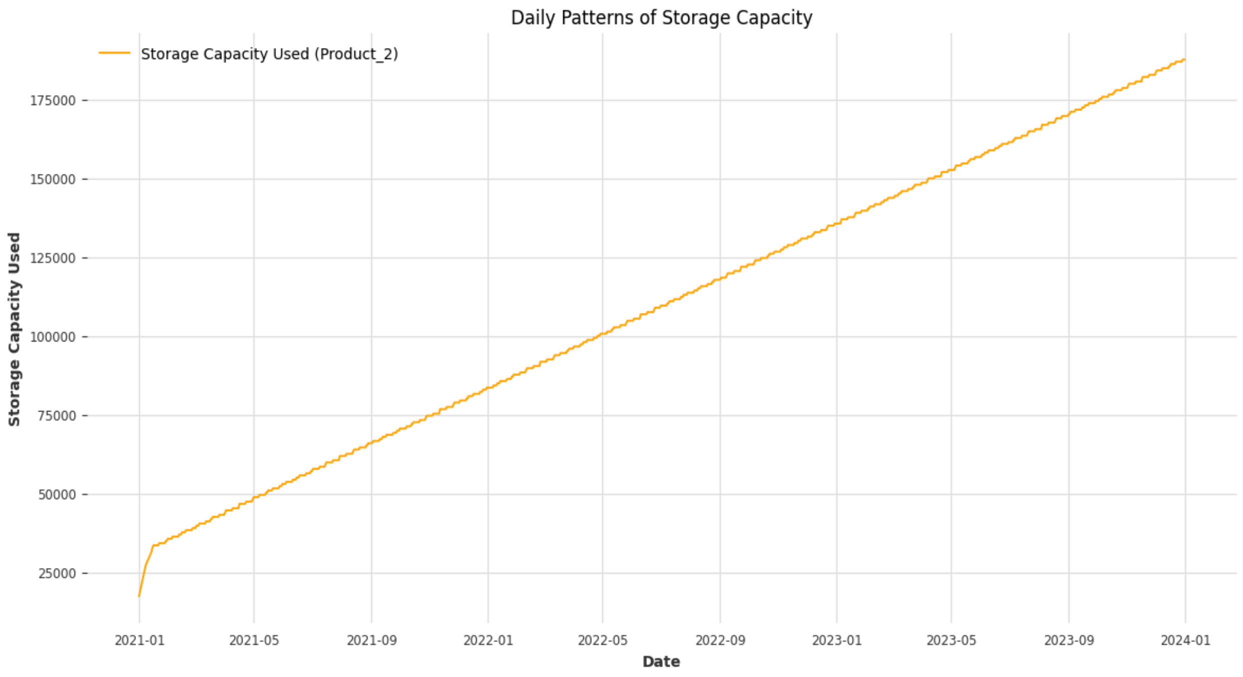


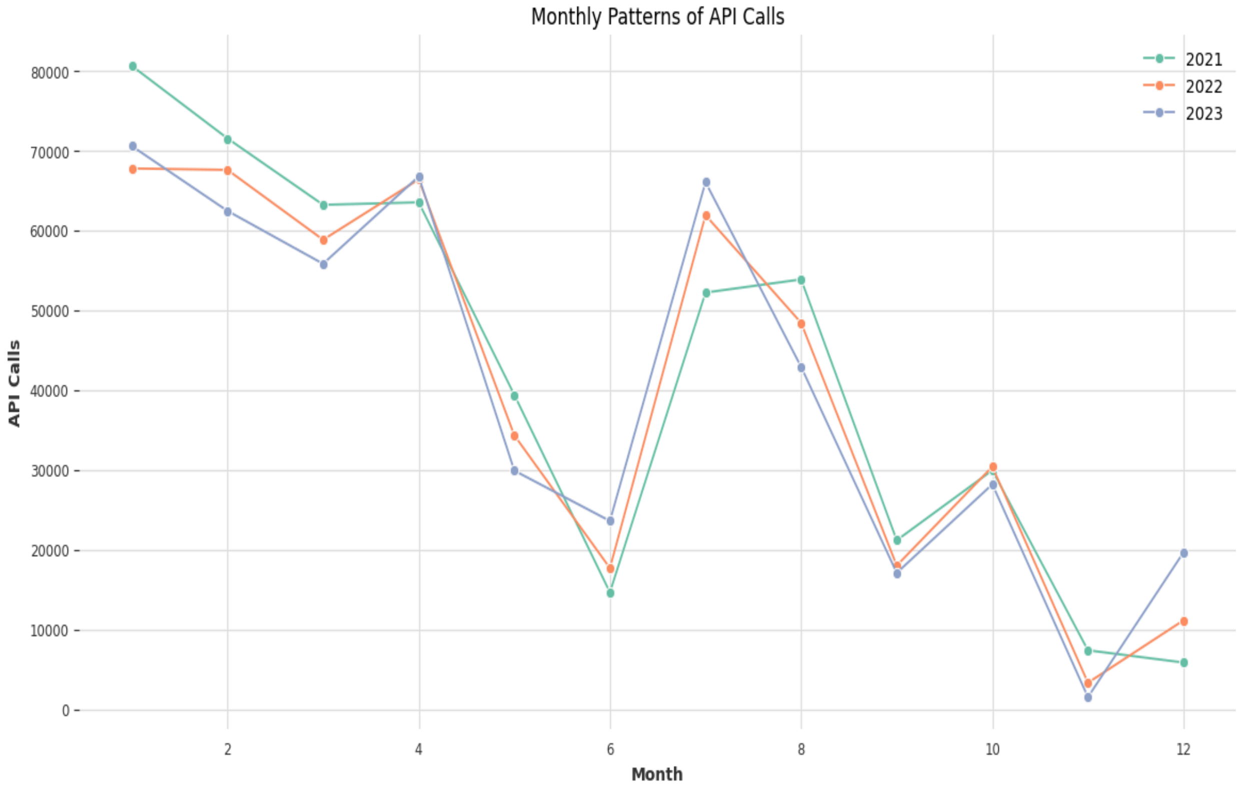
Customer 1 data plot for Memory_storage

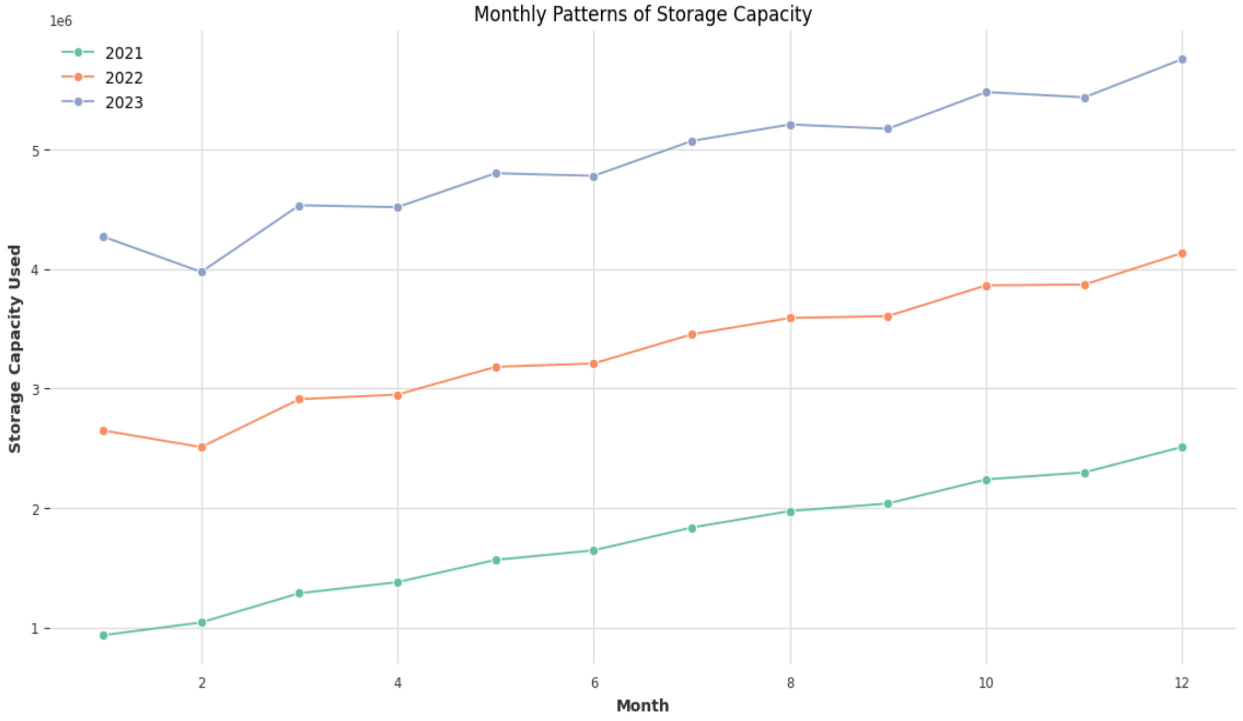


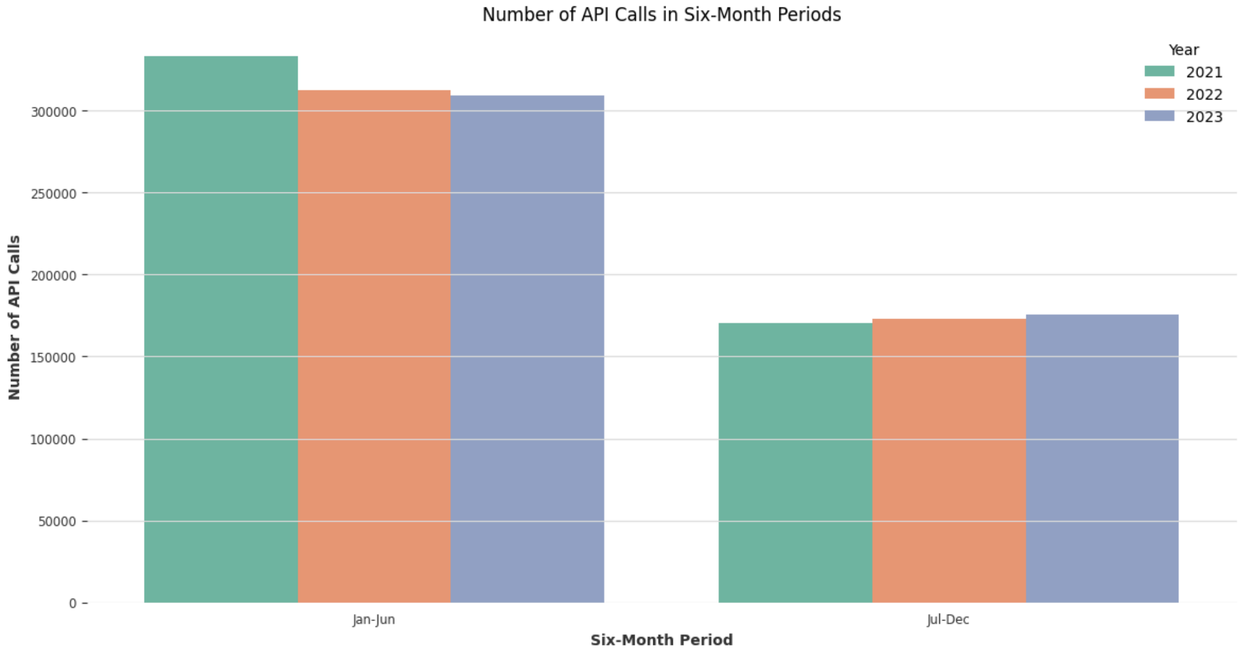
Plot for Customer 2 for API Call and Magnitude of the FFT plotted with period

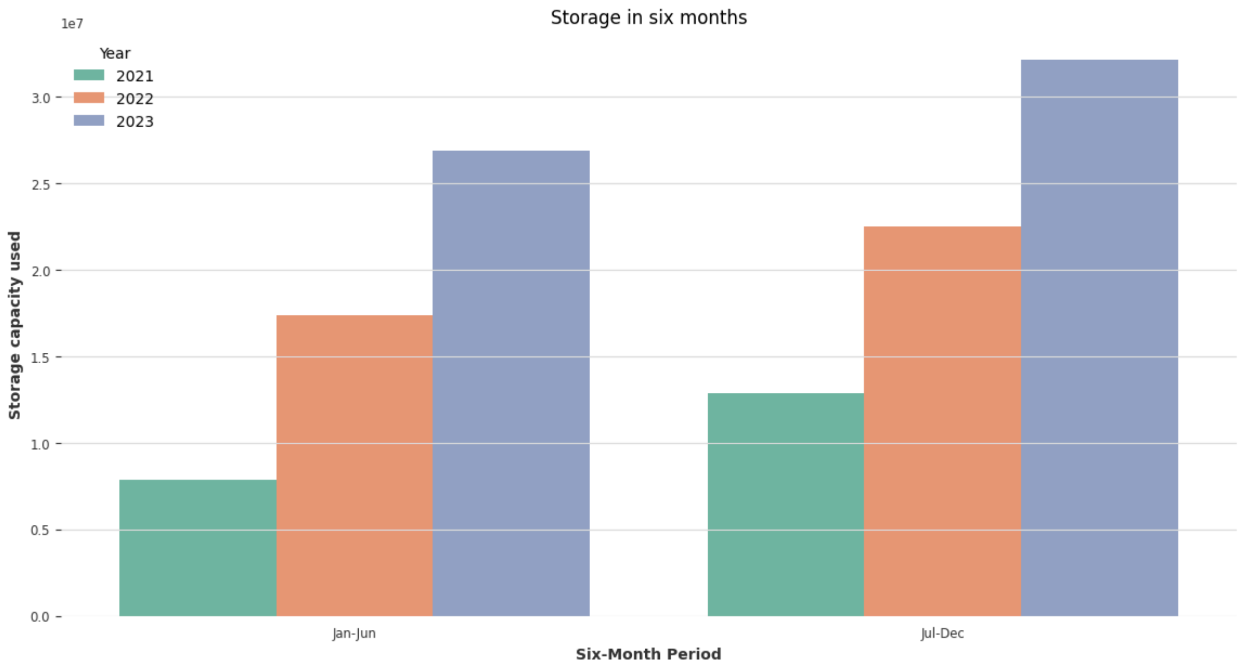


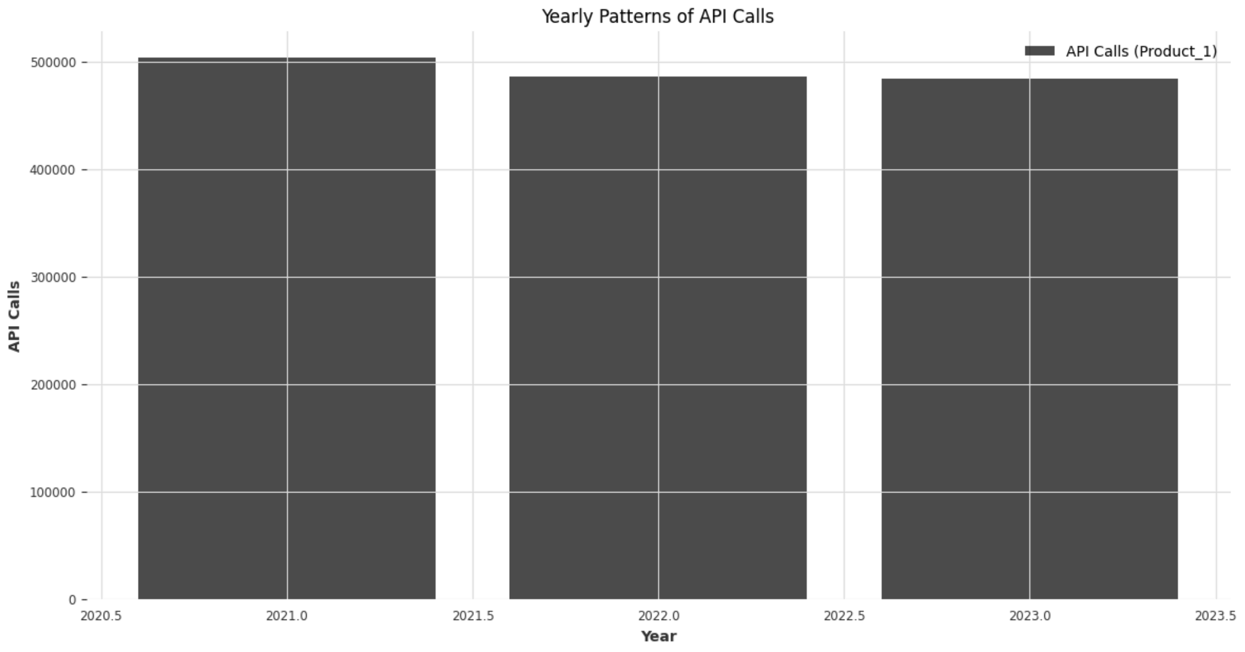

Daily Patterns of Storage Capacity
Storage Capacity Used (Product_2)
Storage Capacity Used
175000
150000
125000
100000
75000
50000
25000
2021-01
2021-05
2021-09
2022-01
2022-05
2022-09
2023-01
2023-05
2023-09
2024-01
Date

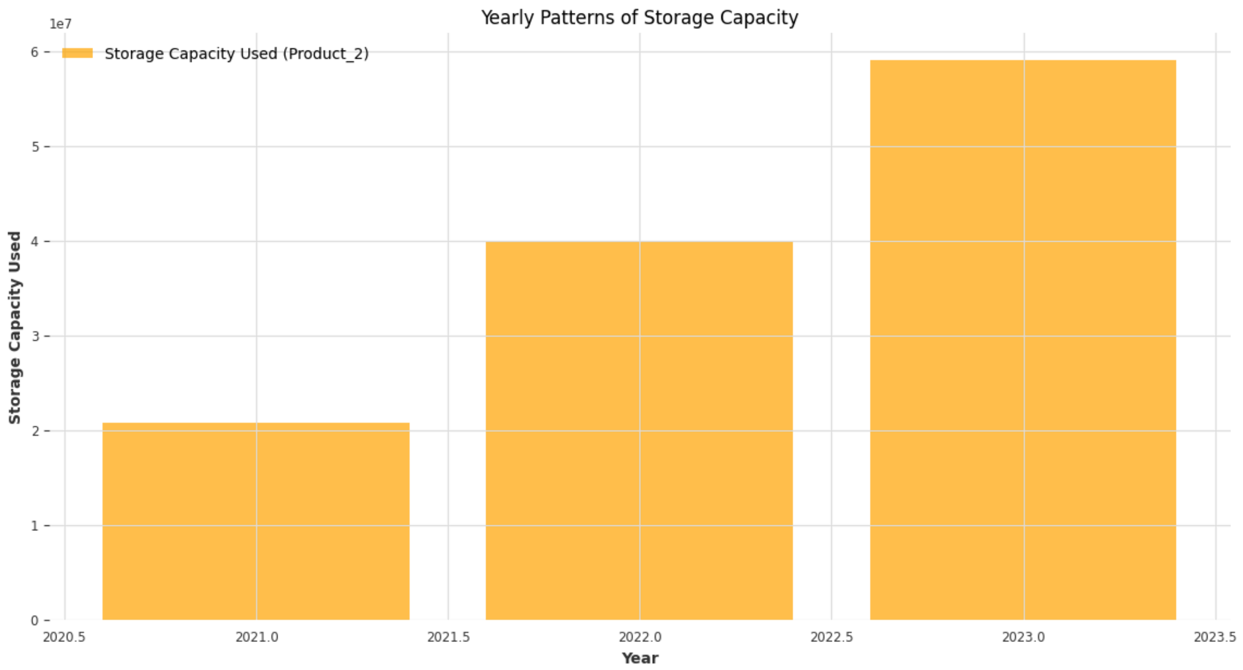

Monthly Patterns of API Calls
2021
2022
2023
API Calls
80000
70000
60000
50000
40000
30000
20000
10000
0
2
4
6
8
10
12
Month

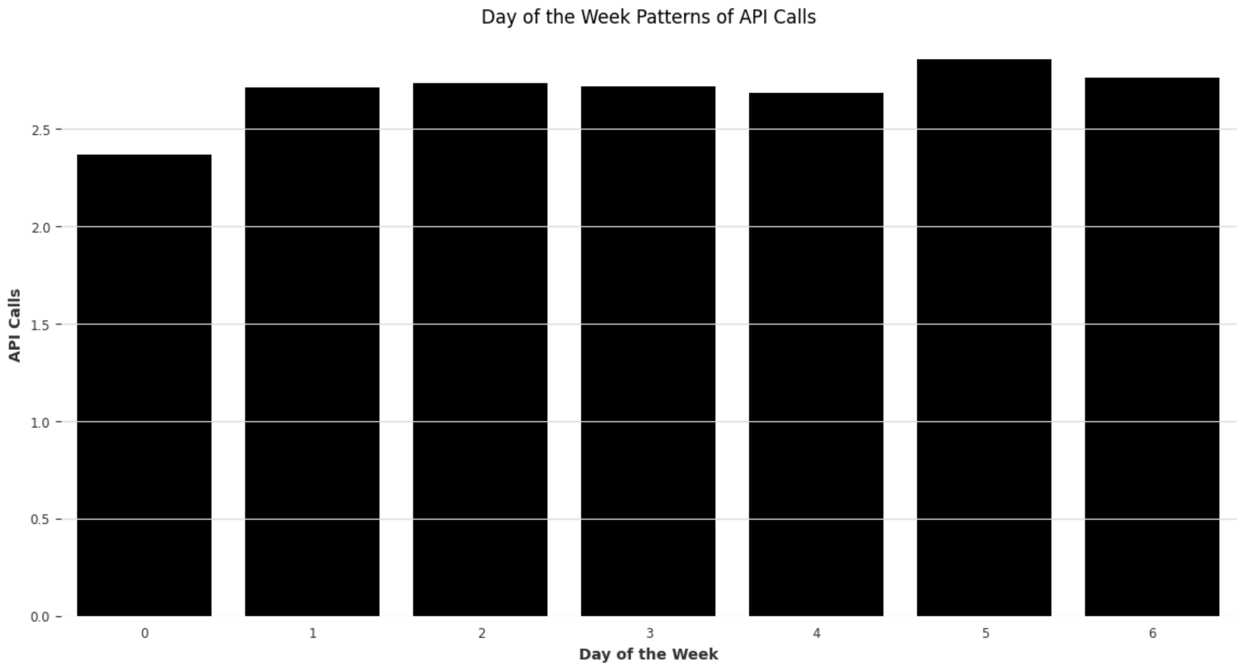

Monthly Patterns of Storage Capacity
1e6
2021
2022
2023
Storage Capacity Used
1
2
3
4
5
2
4
6
8
10
12
Month

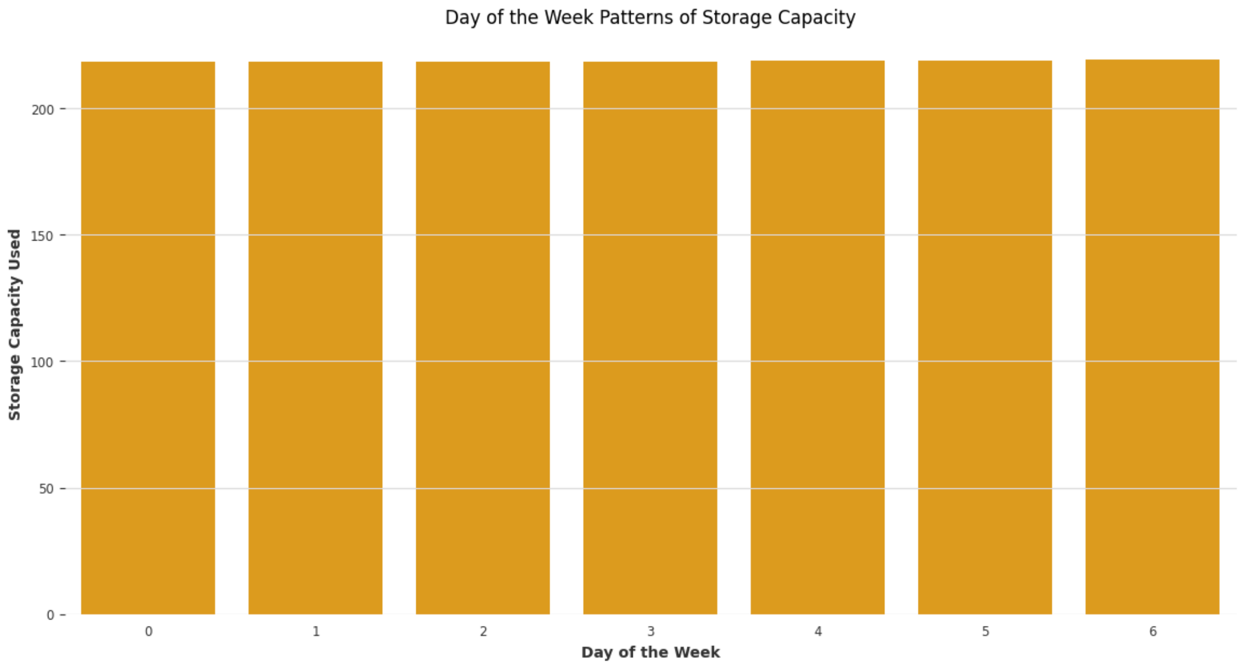

Number of API Calls in Six-Month Periods
Year
2021
2022
2023
Number of API Calls
0
50000
100000
150000
200000
250000
300000
Jan-Jun
Jul-Dec
Six-Month Period


Storage in six months
1e7
Year
2021
2022
2023
Storage capacity used
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Jan-Jun
Jul-Dec
Six-Month Period


Yearly Patterns of API Calls
API Calls (Product_1)
API Calls
500000
400000
300000
200000
100000
0
2020.5
2021.0
2021.5
2022.0
2022.5
2023.0
2023.5
Year


Yearly Patterns of Storage Capacity
1e7
Storage Capacity Used (Product_2)
Storage Capacity Used
0
1
2
3
4
5
6
2020.5
2021.0
2021.5
2022.0
2022.5
2023.0
2023.5
Year


Day of the Week Patterns of API Calls
API Calls
0.0
0.5
1.0
1.5
2.0
2.5
0
1
2
3
4
5
6
Day of the Week



2. Explain the choices you make for building the usage forecasting model

I have tried multiple algorithms, starting with RNN. In RNN, at first, I split the data for the training and testing for each user, but I didn't get the results as expected. After that, I built another RNN model by taking the features as periods and magnitudes of FFT that I calculated from fast Fourier transform but in that, we need to make a module for each customer, and then we need to combine for all to predict the future sales that will be a time taking process so I switched to my 3rd trial of using ARIMA model, but ARIMA also didn't perform well on it. After that, I grouped the data date wise of all the customers and then plotted them. I have seen there is a cyclic pattern that it is following. So, I thought of forecasting that pattern using the Fourier transform and the Inverse Fourier transform. After that, I used Prophet forecasting to forecast API calls. This model is working somewhat better than the other models but still it is not the best model that is even close to my expected results. After this I have tried SARIMA, stationery test, seasonal, trend and residual visualization,LSTM, PMD ARIMA model. After running all the algos, I reached to a conclusion that all these algos are not performing well because they need good amount of data and pattern to learn else either they are underfitting or overfitting.
Finally for Api calls revenue prediction I have used the analysis of Adfuller test that our model is stationary so the plot from Jan-2024 to Jun-2024 should also be stationary. So to predict that approximate curve I have used the mean of the first three cycles of 6 months from the plot of Jan2020 to Dec 2023.
For memory usage, The plot is coming linear so, I am using the linear regression model for prediction.

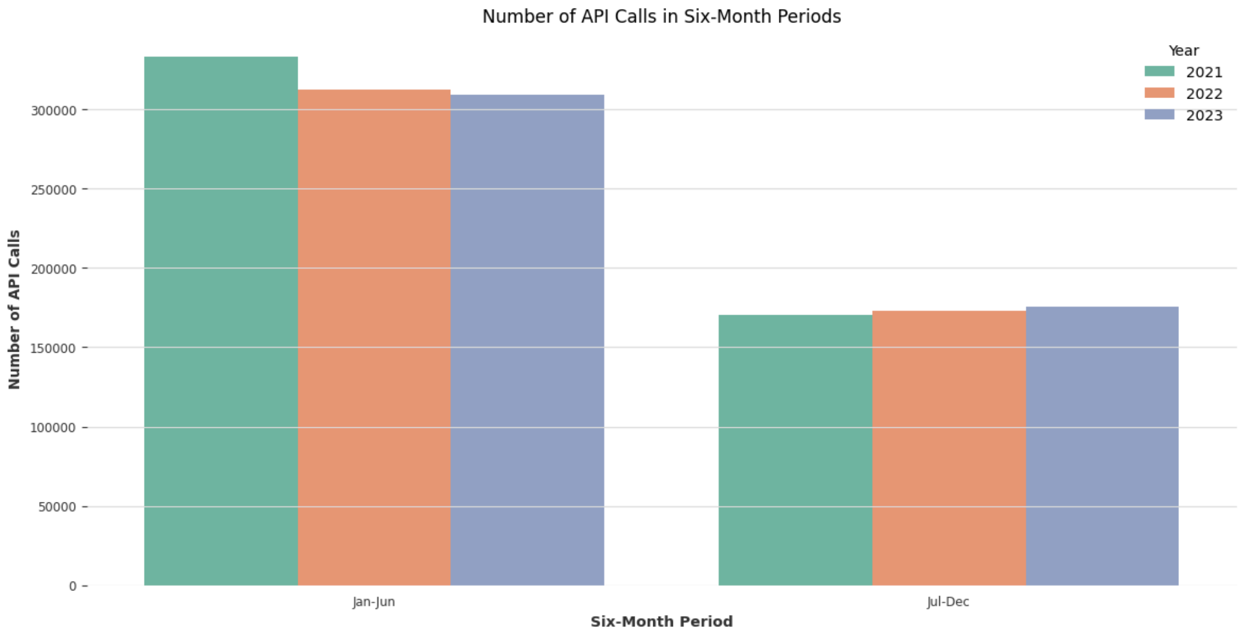
3. **Build a forecasting model that can help predict the Revenue of the cohort of customers shared for the time-period between Jan-June**

   I built the model, and its code is available in the collab file that I have shared. For api call revenue prediction the mathematical model by Adfuller analysis is performing best and for memory usage linear regression the only model I used is performing best.

4. **Provide an understanding of the performance of the forecasting model such that it can be easily understood by VP of FP&A**

Calculation and explanation for api usage:

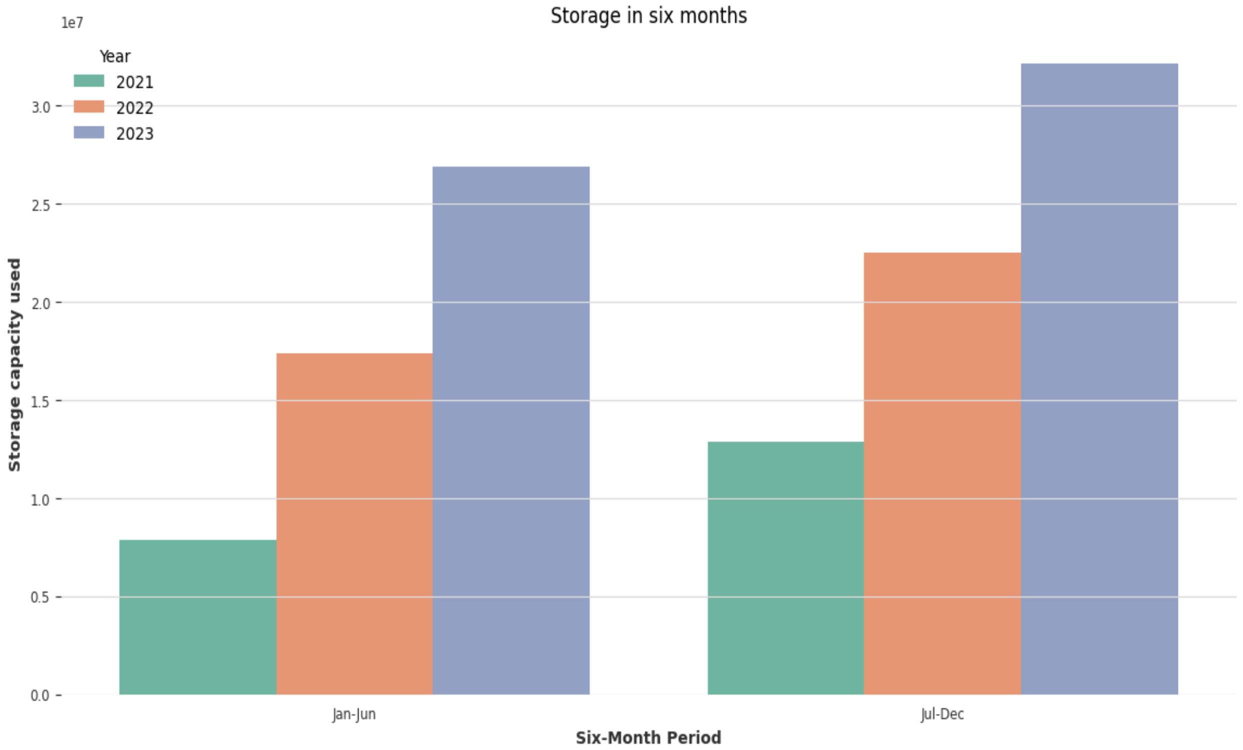
In the below bar graph we can see that the number of Api calls in the starting 6 months of the year is on average around 320000. So, the number of API calls in the upcoming year in the starting 6 months we can expect to be around 315000 because from the graph we can see the number of API calls is slightly decreasing. So the expected revenue is 315000*0.2= 63000. And as our calculated revenue from the model is also 62354 so it is very close to the expected revenue. So we can say our model will work fine.



Calculation and explanation for memory usage:

Calculated Revenue generated from memory_usage = 3658322

The Linear regression model that I have built has an R-2 score of 0.9994 which is showing it is performing well. The below graph shows that on average in the starting 6 months of the year our storage capacity is increasing by 0.95 which implies that in 2024 the memory usage from Jan to June will be around 2.65+0.95. So the revenue generated from it will be 3.6*1e7*0.1= 3600000 which is close to our calculated revenue. Hence our linear regression model is also performing well.



The below plot is showing the forecast data from Jan 2024 to June 2024 for api_usage.

5. Provide a strategy for the deployment/ maintenance of the model.

We can use cloud services like AWS, GCP, or Azure to provide scalable and flexible infrastructure. We can package the model using Docker to ensure consistency across different environments. For maintenance, we can use tools like Prometheus or ELK stack to monitor model performance (e.g., latency, throughput). And we can also schedule regular retraining of the model using the latest data to keep it up-to-date.

6. List the assumptions that you have made and challenges that you foresee in the rollout of the new forecasting model.

I have checked for the data of around 10 to 12 customers and found that they are periodic for api_calls, and based on that, I have taken the assumption that the data of all the customers is periodic for api_calls. Another assumption is that for Memory usage the plot seems to be linear so there we have assumed it will behave in a similar manner in future.
And the last assumption is that users will behave in the same manner in the future as they are behaving now.