Sales Forecast Analytics

DATA SCIENCE PRODEGREE PROJECT

DSP -18 (Group 1)

- Komal Ise
- Niva Srivastava
- Kishore Kumar
- Dhivya Anandan
- Sakshi Yadav

Agenda

- Introduction
- Problem Statement
- Dataset Overview
- EDA (Exploratory Data Analysis)
- Data Cleaning and Preparation
- Feature Engineering
- Model Building
- Testing and Cross-validation
- Visualizing Forecasts
- Conclusion

Introduction

- Pest control is the major part of agriculture. There are different types of pesticides available in the market.
- Some external factors may affect the sales trend.
- Here, we are analyzing and forecasting the sales of different pesticides for particular company named "ABC manufacturing ".

Problem Statement

Statement

State Level Sales Data: containing sales data by various Pesticide manufacturing companies. This data is available for state and district level ranging from 2014-2018.

Data and Problem Details

- Conduct statewise trend analysis of the given sales
- To identify other reasons if they are influencing Pesticide sales
- To provide statewise forecast of Pesticide sales for ABC Manufacturing.

Dataset Overview

```
In [5]: sale_forecast_df=pd.read_excel("Sales_Forecast_Data.xlsx")
```

```
In [6]: sale_forecast_df.head(10)
```

Out[6]:

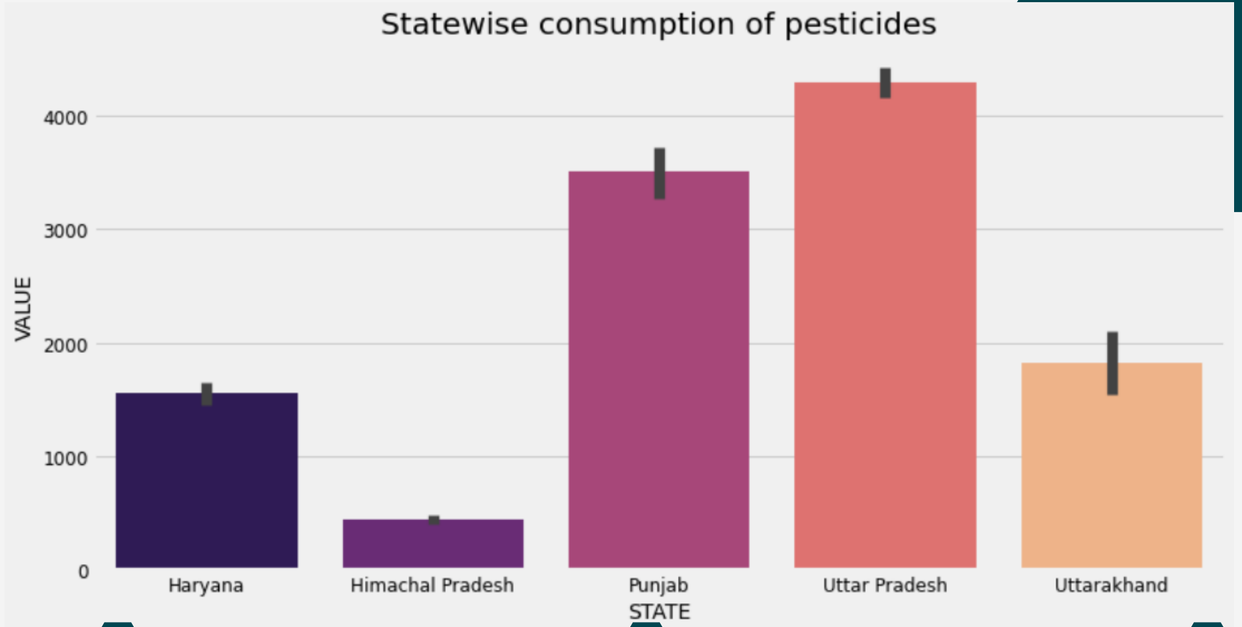
	FIN_YEAR	MONTH	STATE	FG	DISTRICT	COMPANY	VALUE
0	2014-2015	Nov	Haryana	Insecticides	Rohtak	monous	675.00
1	2014-2015	Nov	Haryana	Insecticides	Rewari	monous	73.71
2	2014-2015	Nov	Haryana	Insecticides	Gurgaon	monous	405.00
3	2014-2015	Nov	Haryana	Insecticides	Mewat	monous	1188.00
4	2014-2015	Nov	Haryana	Insecticides	Palwal	monous	5188.59
5	2014-2015	Nov	Haryana	Insecticides	Jind	kadoo	0.00
6	2014-2015	Nov	Haryana	Insecticides	Sirsa	kadoo	0.00
7	2014-2015	Nov	Haryana	Insecticides	Hisar	kadoo	0.00
8	2014-2015	Nov	Haryana	Insecticides	Bhiwani	kadoo	0.00
9	2014-2015	Nov	Haryana	Insecticides	Rohtak	kadoo	0.00

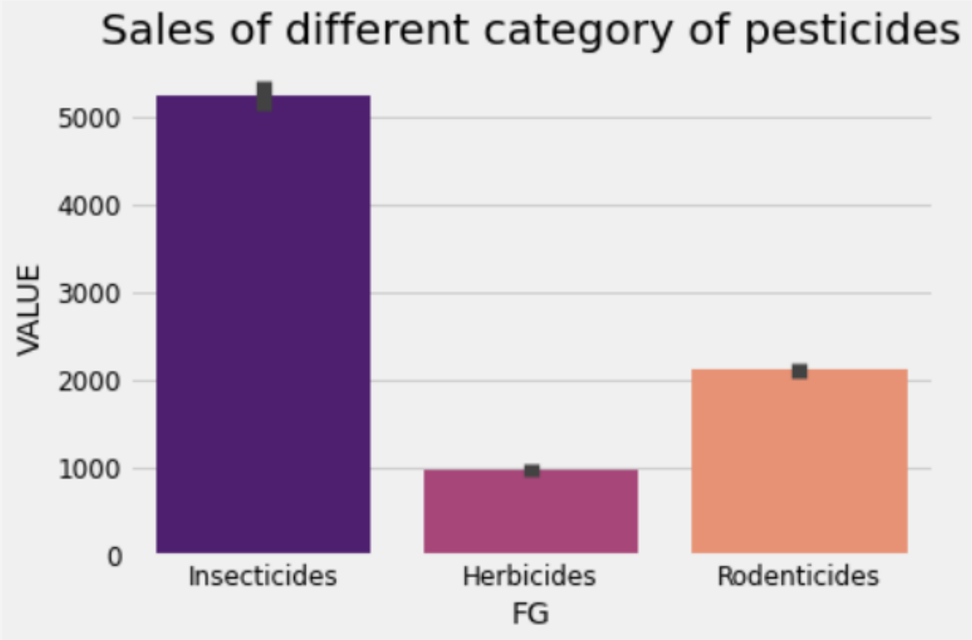
EDA

(Exploratory Data Analysis)

As per the graph,

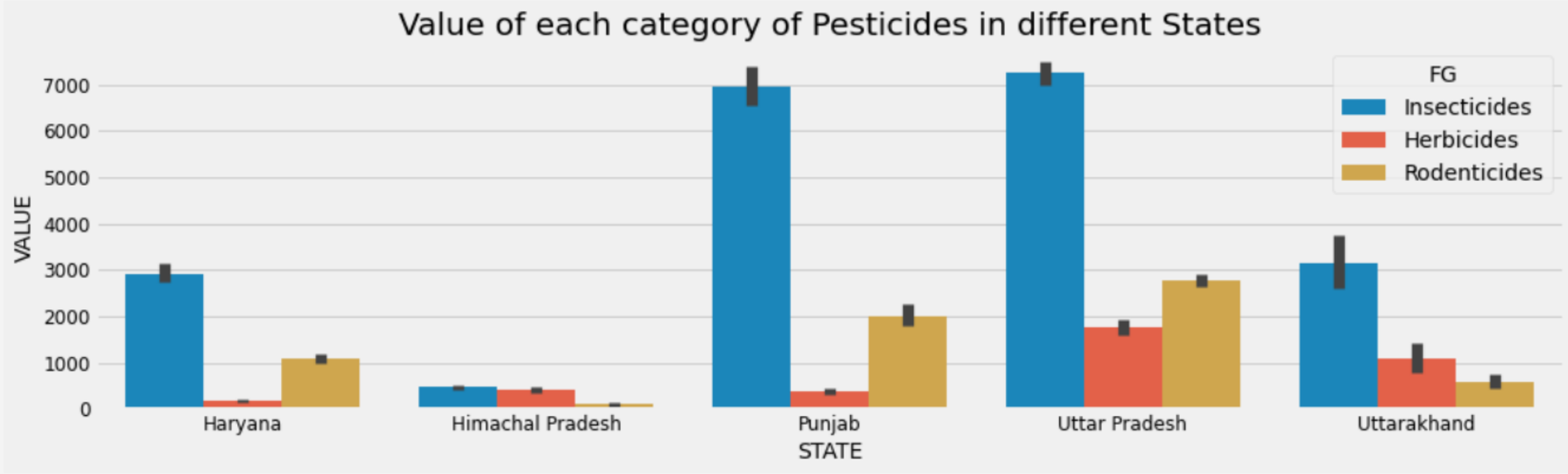
1. Consumption of pesticide is high in Uttar Pradesh.
2. Punjab is the second highest state in pesticide consumption.
3. Himachal Pradesh consumes least while comparing all other 5 Northern States





As per the graph,

- For ABC_Manufacturing, Insecticides accounts for the highest portion of the sales whereas Herbicides are the least .



From the above graph,

- **Punjab and UP are consuming higher proportion of Insecticides.**
- **Haryana consumes least Herbicides**
- **Himachal Pradesh has no consumption of Rodenticides.**

Plotting the graph to check stationarity

The plot indicates that the time series has seasonality pattern. There is no upward or downward trend. So, it looks stationary.

Performing the Augumented Dickey Fuller test

```
from pandas import Series
from statsmodels.tsa.stattools import adfuller

result = adfuller(df)

print('ADF Statistic: %f' % result[0])

print('p-value: %f' % result[1])

print('Critical Values:')

for key, value in result[4].items():
       print(key, value)
```

```
ADF Statistic: -7.431211
p-value: 0.000000
Critical Values:
1% -3.584828853223594
5% -2.9282991495198907
10% -2.6023438271604937
```

Here, the p-value is also less than 0.05. So we can reject the null hypothesis and confirm that our time series is stationary.

Model Building

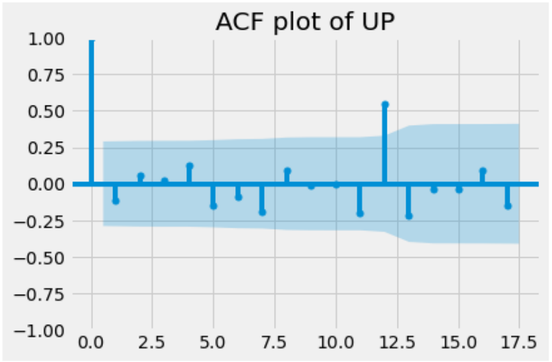
ACF and PACF Plots

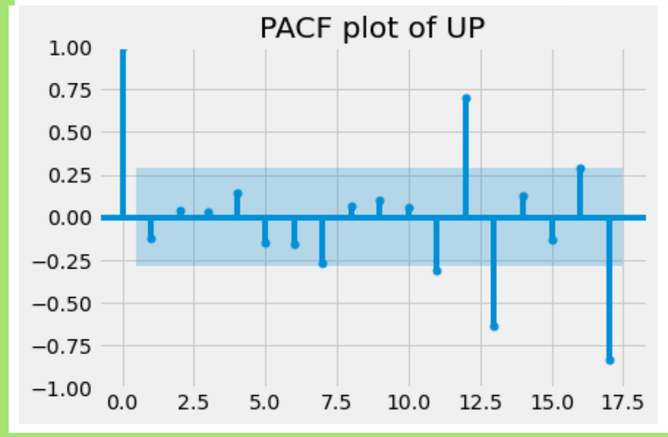
```python
from statsmodels.graphics.tsaplots import plot_acf,plot_pacf
```

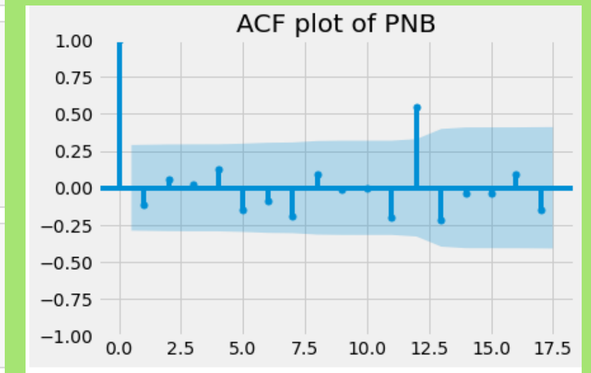
```python
#function to create acf plot

def show_acf(df,state):
    plot_acf(df)
    plt.title("ACF plot of {}". format(state))
    plt.show()
```

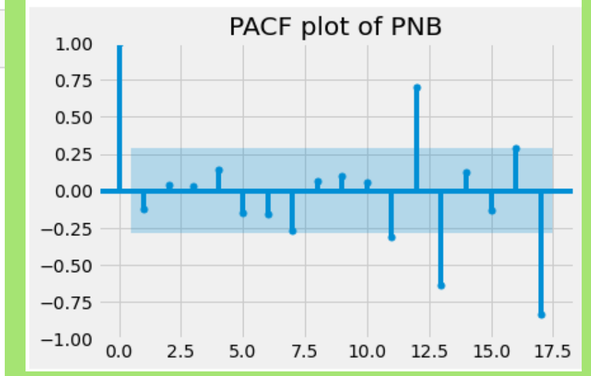
```python
#function to show pacf plot
def show_pacf(df,state):
    plot_pacf(df)
    plt.title("PACF plot of {}". format(state))
    plt.show()
```

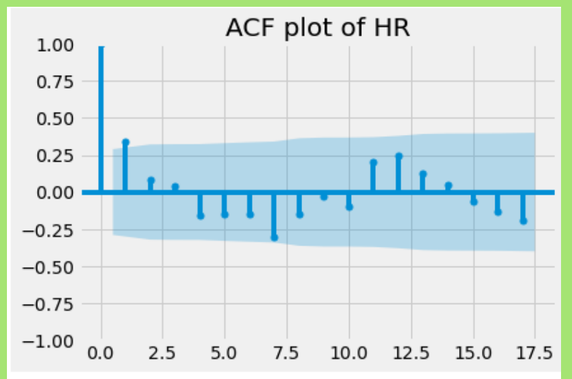
```python
show_acf(UP,"UP")
show_pacf(UP,"UP")
```

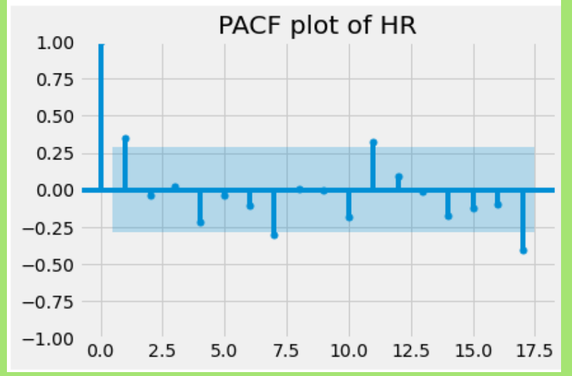


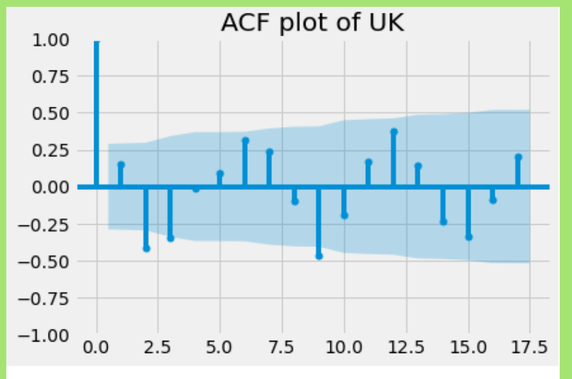


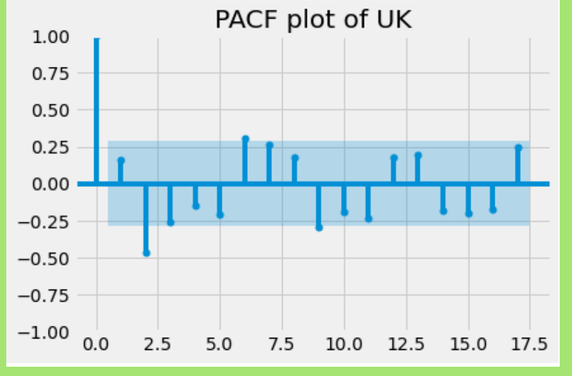


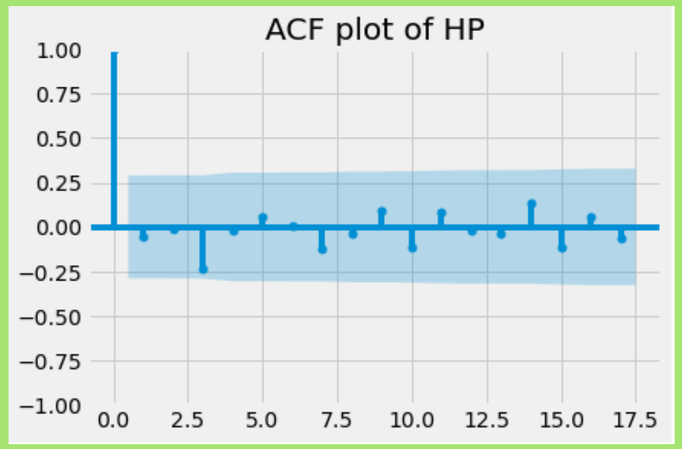


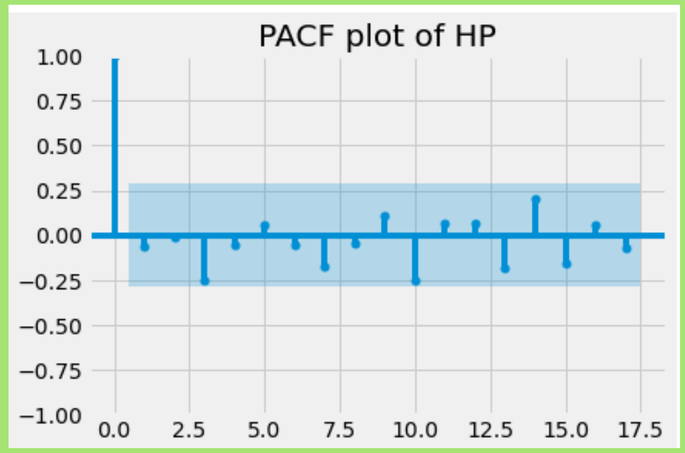












Splitting the dataset as Training and Testing

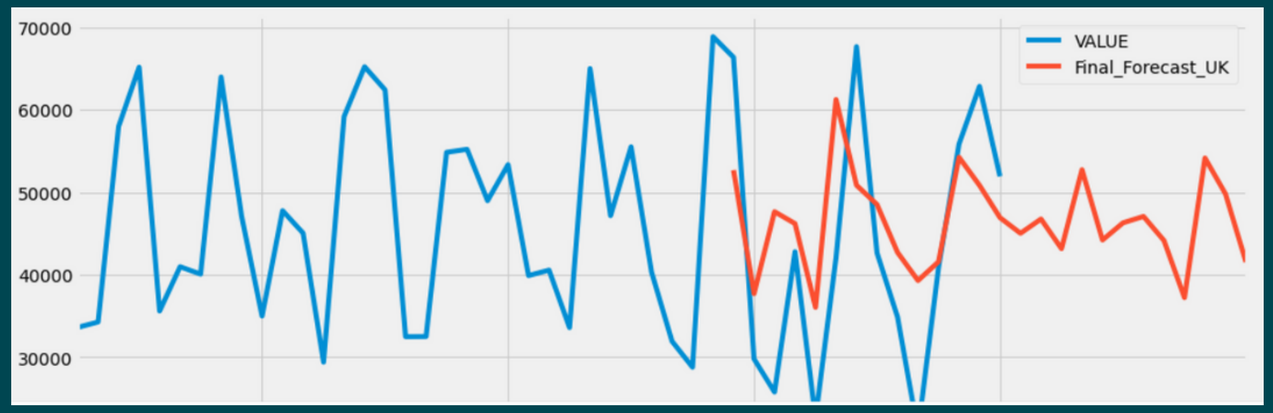
```
#since there are 46 entries we using 70% in train and remianing on test i,e 30% we validate on test
train_UP=UP.iloc[0:32]
test_UP=UP.iloc[32:]

train_HR=HR.iloc[0:32]
test_HR=HR.iloc[32:]

train_PB=PNB.iloc[0:32]
test_PB=PNB.iloc[32:]

train_HP=HP.iloc[0:32]
test_HP=HP.iloc[32:]

train_UK=UK.iloc[0:32]
test_UK=UK.iloc[32:]
```

SARIMA Model

```
p,d,q,P,D,Q=range(3),range(1),range(3),range(3),range(1),range(3)
#p,d,q,P,D,Q=[4,5,6],range(1),range(4,6),range(3),range(1),[4,5,6]
pList,dList,qList,PList,DList,QList=[],[],[],[],[],[]

#for storing the MAPE values of the different model
mapeList=[]
```

```
def perform_gridSearch(original_df,train_df):
    for i in p:
        for j in d:
            for k in q:
                for l in P:
                    for m in D:
                        for n in Q:

                            print(i,j,k,l,m,n)

                            temparima_model=ARIMA((i,j,k),(l,m,n,12)).fit(train_df)

                            Forecast=pd.Series(temparima_model.predict(14)).rename('Forecast')
                            Forecast.index=test_UP.index
                            Actual_Forecast_Df=pd.concat([original_df,Forecast],axis=1)
                            Validation_df=Actual_Forecast_Df[-14:].copy()
                            tempMAPE=np.mean(abs(Validation_df['VALUE']-Validation_df['Forecast'])/Validation_df['VALUE'])*100

                            pList.append(i)
                            dList.append(j)
                            qList.append(k)
                            PList.append(l)
                            DList.append(m)
                            QList.append(n)
                            mapeList.append(tempMAPE)
```

- By performing the grid search, we have found the best P,D,Q/p,d,q values.
- Based on the values, we have performed the SARIMA model for statewise.
- Statewise forecast for next 12 months was shown in the next slide.



Models	MAPE values				
	UP	HR	PNB	HP	UK
ARIMA	27.65	94.31	58.81	253.70	34.46
SARIMA	33.33	60.64	69.81	96.69	30.57

Conclusion

- Overall, our forecasts align with the true values very well.
- So, we can proceed with SARIMA model.
- By forecasting the data for the next 12 months, we can say that ABC Manufacturing company is good to grow in all states.

THANK YOU