

AGENDA

- **DATASET OVERVIEW**
- **DESCRIPTIVE STATISTICS**
- **ORDER ANALYSIS**
- **DATA EXPLORATION**
- **PREDICTION MODELS**
- **FORECASTING AND RESULTS**

DATASET OVERVIEW



HISTORICAL DATA



PREDICT THE TOTAL NUMBER OF ORDERS



FEATURES AVAILABLE: 12

DESCRIPTIVE STATISTICS

	Week_of_Month	Day_of_Week	Non_Urgent_Order	Urgent_Order	Order_Type_A	Order_Type_B	Order_Type_C	Fiscal_Sector_Orders
count	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000
mean	3.016667	4.033333	172.554933	118.920850	52.112217	109.229850	139.531250	77.396133
std	1.282102	1.401775	69.505788	27.170929	18.829911	50.741388	41.442932	186.502470
min	1.000000	2.000000	43.651000	77.371000	21.826000	25.125000	74.372000	0.000000
25%	2.000000	3.000000	125.348000	100.888000	39.456250	74.916250	113.632250	1.243250
50%	3.000000	4.000000	151.062500	113.114500	47.166500	99.482000	127.990000	7.831500
75%	4.000000	5.000000	194.606500	132.108250	58.463750	132.171000	160.107500	20.360750
max	5.000000	6.000000	435.304000	223.270000	118.178000	267.342000	302.448000	865.000000

Traffic_Controller_Orders	Banking_Orders_1	Banking_Orders_2	Banking_Orders_3	Total_Orders
60.000000	60.000000	60.000000	60.000000	60.000000
44504.350000	46640.833333	79401.483333	23114.633333	300.873317
12197.905134	45220.736293	40504.420041	13148.039829	89.602041
11992.000000	3452.000000	16411.000000	7679.000000	129.412000
34994.250000	20130.000000	50680.500000	12609.750000	238.195500
44312.000000	32527.500000	67181.000000	18011.500000	288.034500
52111.750000	45118.750000	94787.750000	31047.750000	334.237250
71772.000000	210508.000000	188411.000000	73839.000000	616.453000

DATA EXPLORATION

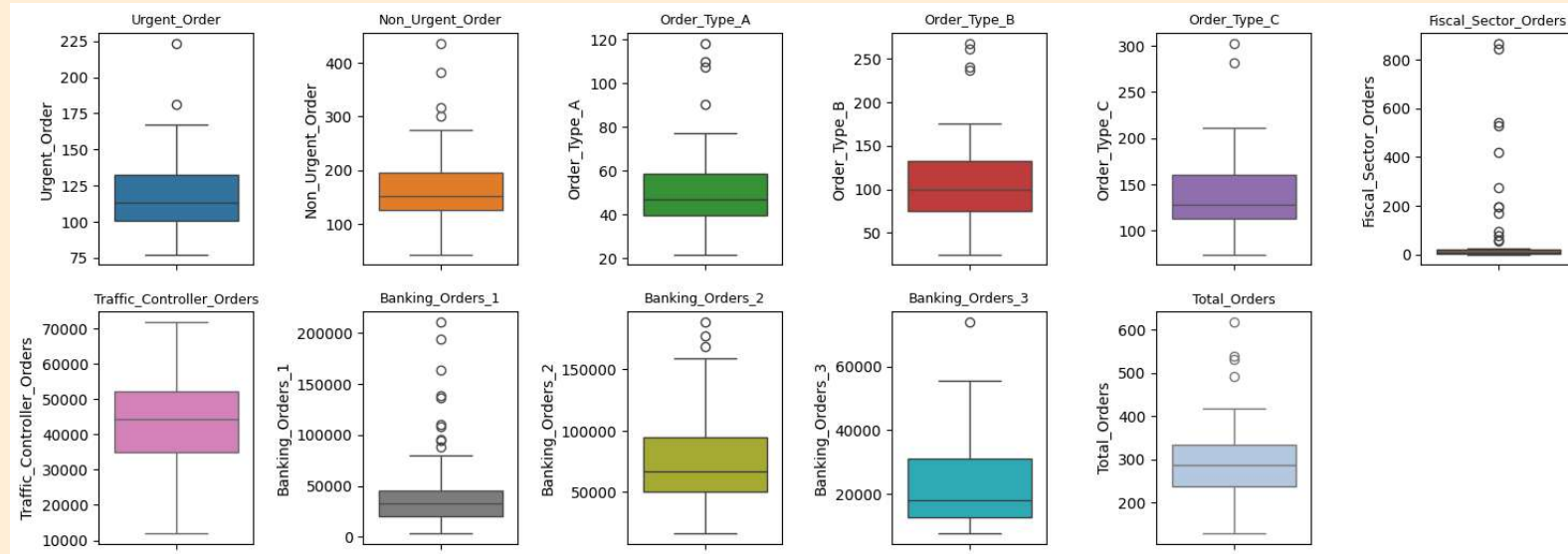
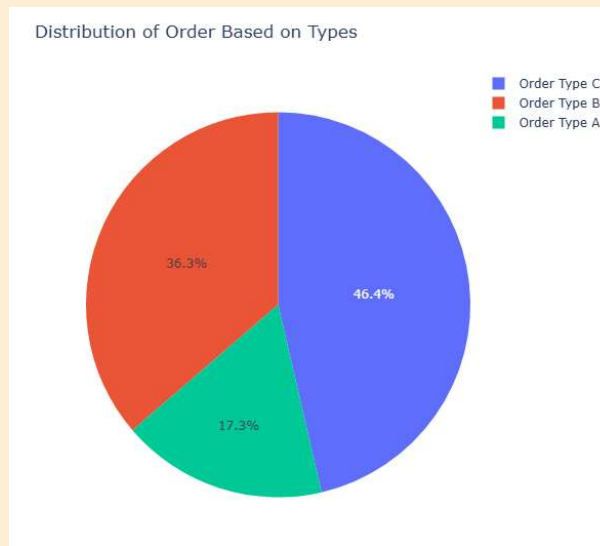
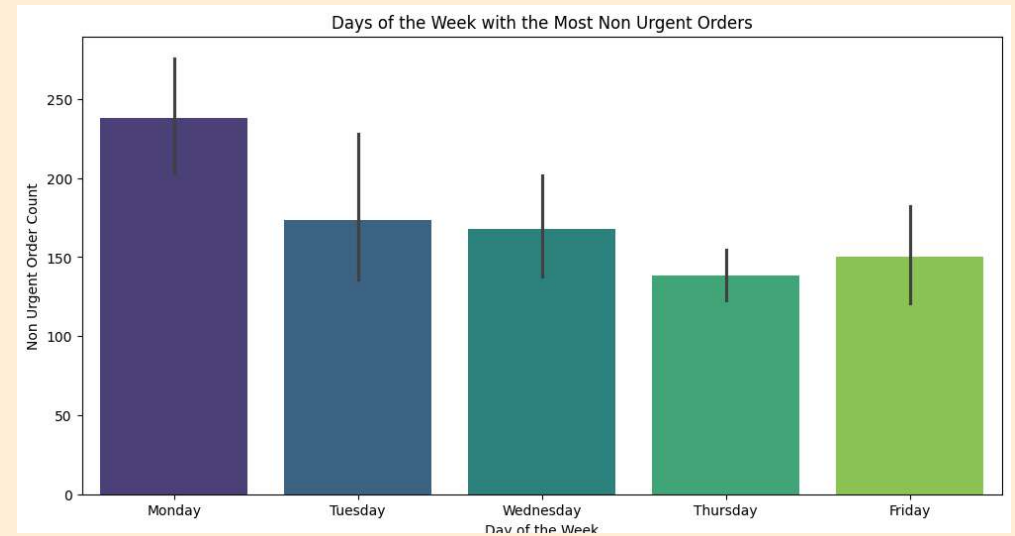
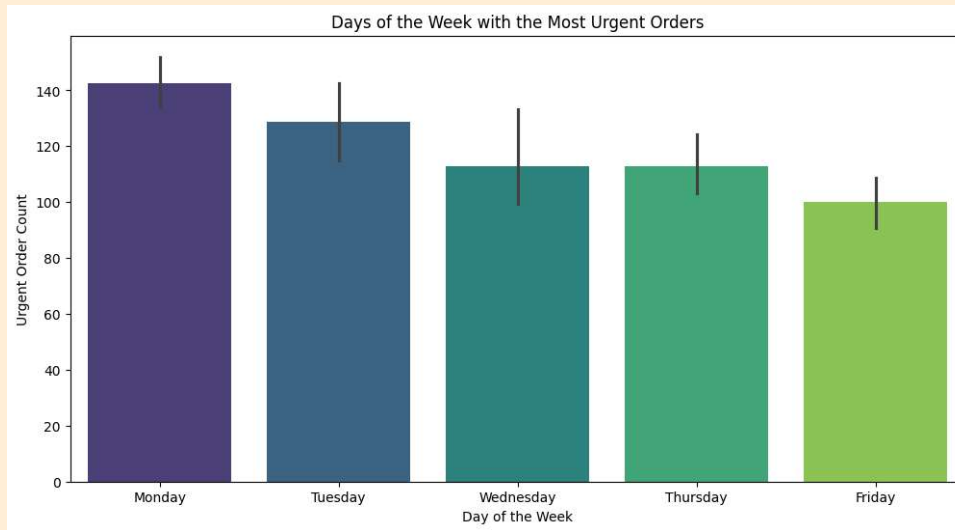


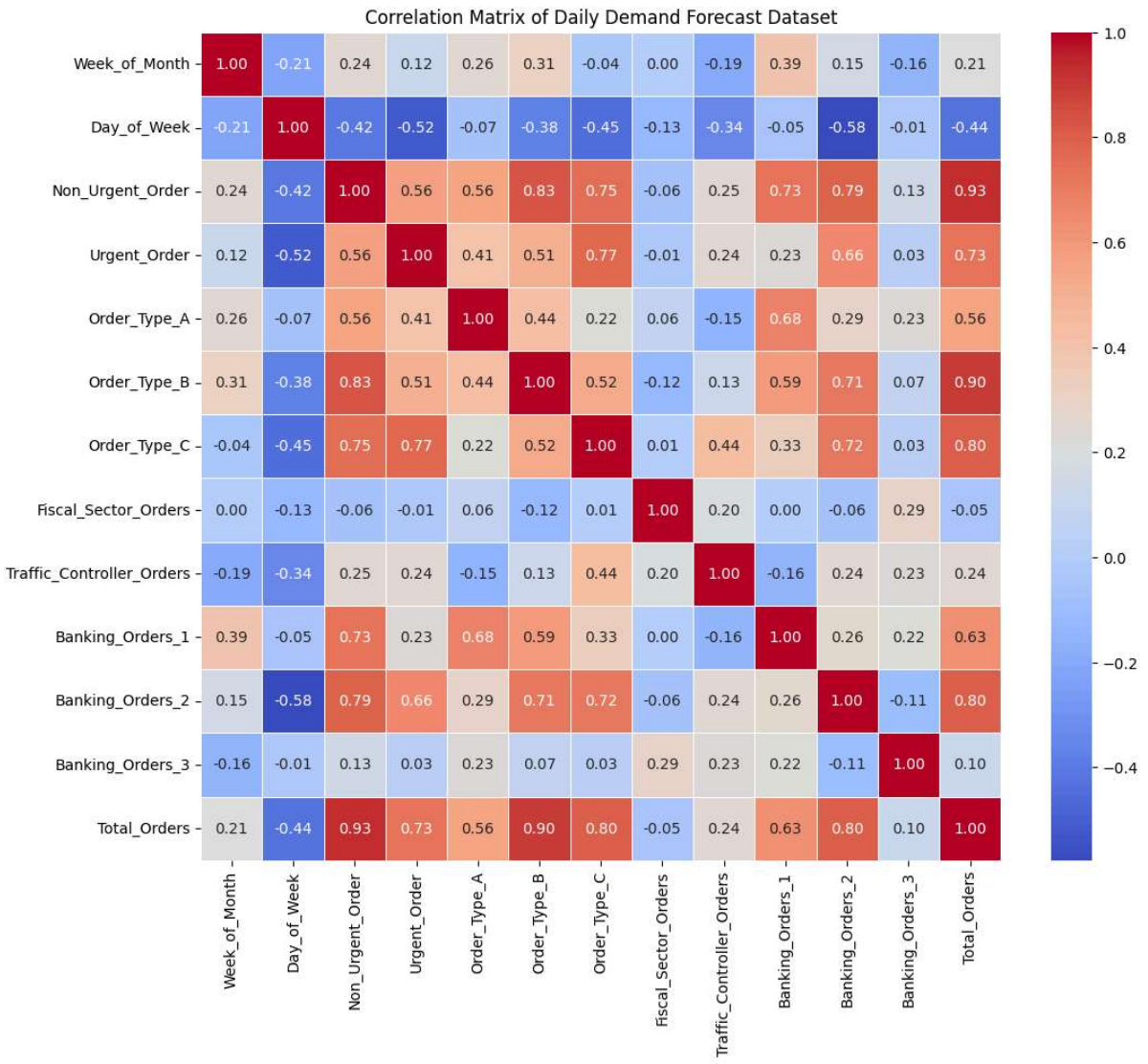
FIGURE: BOX PLOT

1. COUNT ALL ARE SAME: NO MISSING VALUES.
2. WE CAN SEE THAT FISCAL SECTORS ORDERS SHOWING HIGH VARIANCE WITH LOT OF OUTLIERS.
3. SAME WITH BANKING ORDERS_1 COLUMN.

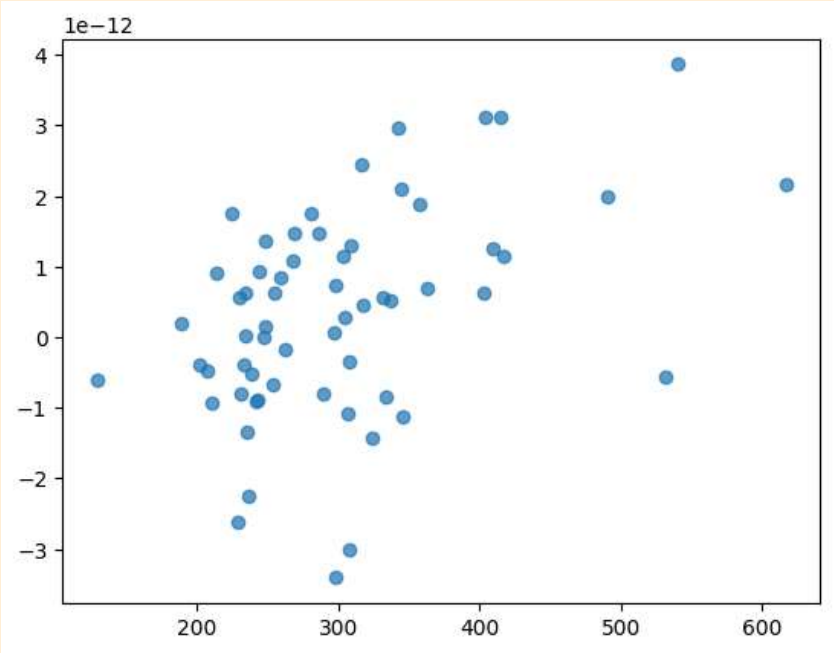
1. TOTAL ORDERS HAVE MEAN AROUND ~300
2. WE CAN SEE THE PATTERN OF INCREASING MEAN IN ORDER TYPES.
3. REST OF THE DATA SEEMS PRETTY FINE.

ORDERS ANALYSIS





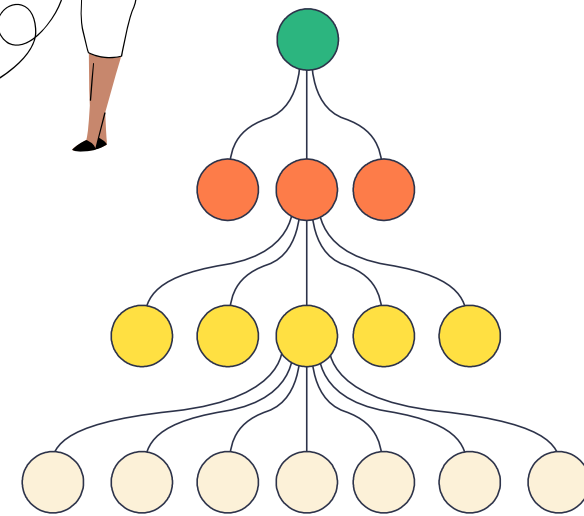
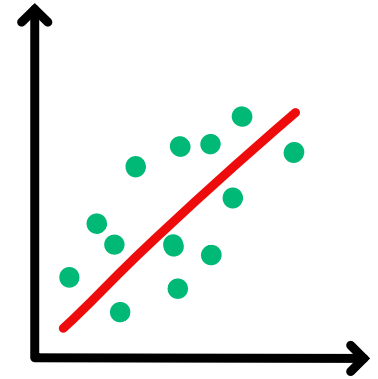
CORRELATION MATRIX



RESIDUAL GRAPH

PREDICTION MODELS

LINEAR REGRESSION



--- OLS Regression Summary (Includes All Predictors) ---

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Total_Orders      R-squared:                1.000
Model:                  OLS              Adj. R-squared:           1.000
Method:                 Least Squares     F-statistic:             1.038e+27
Date:                   Sun, 12 Oct 2025   Prob (F-statistic):       0.00
Time:                   13:38:51          Log-Likelihood:          1487.2
No. Observations:       60               AIC:                    -2936.
Df Residuals:           41               BIC:                    -2897.
Df Model:               18
Covariance Type:        nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	5.23e-12	4.81e-12	1.087	0.284	-4.49e-12	1.49e-11
Non_Urgent_Order	-5.773e-15	7.7e-14	-0.075	0.941	-1.61e-13	1.5e-13
Urgent_Order	-5.351e-14	7.78e-14	-0.687	0.496	-2.11e-13	1.04e-13
Order_Type_A	1.0000	8.57e-14	1.17e+13	0.000	1.000	1.000
Order_Type_B	1.0000	3.24e-14	3.08e+13	0.000	1.000	1.000
Order_Type_C	1.0000	6.98e-14	1.43e+13	0.000	1.000	1.000
Fiscal_Sector_Orders	4.559e-15	5.13e-15	0.889	0.379	-5.79e-15	1.49e-14
Traffic_Controller_Orders	5.226e-17	1.1e-16	0.474	0.638	-1.7e-16	2.75e-16
Banking_Orders_1	7.503e-17	6.26e-17	1.198	0.238	-5.15e-17	2.02e-16
Banking_Orders_2	4.749e-17	6.17e-17	0.770	0.446	-7.7e-17	1.72e-16
Banking_Orders_3	-9.259e-17	6.87e-17	-1.349	0.185	-2.31e-16	4.61e-17
Day_Name_Monday	-2.842e-14	3.07e-12	-0.009	0.993	-6.23e-12	6.17e-12
Day_Name_Thursday	7.105e-15	2.28e-12	0.003	0.998	-4.61e-12	4.62e-12
Day_Name_Tuesday	7.816e-14	3.08e-12	0.025	0.980	-6.14e-12	6.29e-12
Day_Name_Wednesday	-1.421e-14	2.55e-12	-0.006	0.996	-5.16e-12	5.13e-12
Week_Label_Week_2	-6.395e-14	2.55e-12	-0.025	0.980	-5.21e-12	5.08e-12
Week_Label_Week_3	-6.395e-14	2.79e-12	-0.023	0.982	-5.69e-12	5.56e-12
Week_Label_Week_4	-7.15e-14	2.62e-12	-0.027	0.978	-5.37e-12	5.23e-12
Week_Label_Week_5	-2.842e-14	3.4e-12	-0.008	0.993	-6.9e-12	6.84e-12

```

=====
Omnibus:                8.196      Durbin-Watson:           0.467
Prob(Omnibus):          0.017      Jarque-Bera (JB):         8.409
Skew:                   -0.917      Prob(JB):                 0.0149
Kurtosis:               2.992      Cond. No.                 1.08e+06
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.08e+06. This might indicate that there are strong multicollinearity or other numerical problems.

• NORMALITY

- both Prob(omnibus) and Prob(JB) values both are greater than 0.05 so we can say data is normal.

• Multicollinearity

- Cond No is greater than 10 so there is high multicollinearity

• Independence of errors.

- DB values are less than 2 so no autocorrelation, but there are categorical variable we need to treat that first.

• From P values

- Order_Type_A, Order_Type_B,
- Order_type_C, Banking_order_2 and Banking_Order_3 are significant.

One hot encoding to convert categorical to numeric(Days_of_Week and week_of month).

```
▶ # dictionary to map numbers to day names
day_of_week_map = {
    2: 'Monday',
    3: 'Tuesday',
    4: 'Wednesday',
    5: 'Thursday',
    6: 'Friday'
}

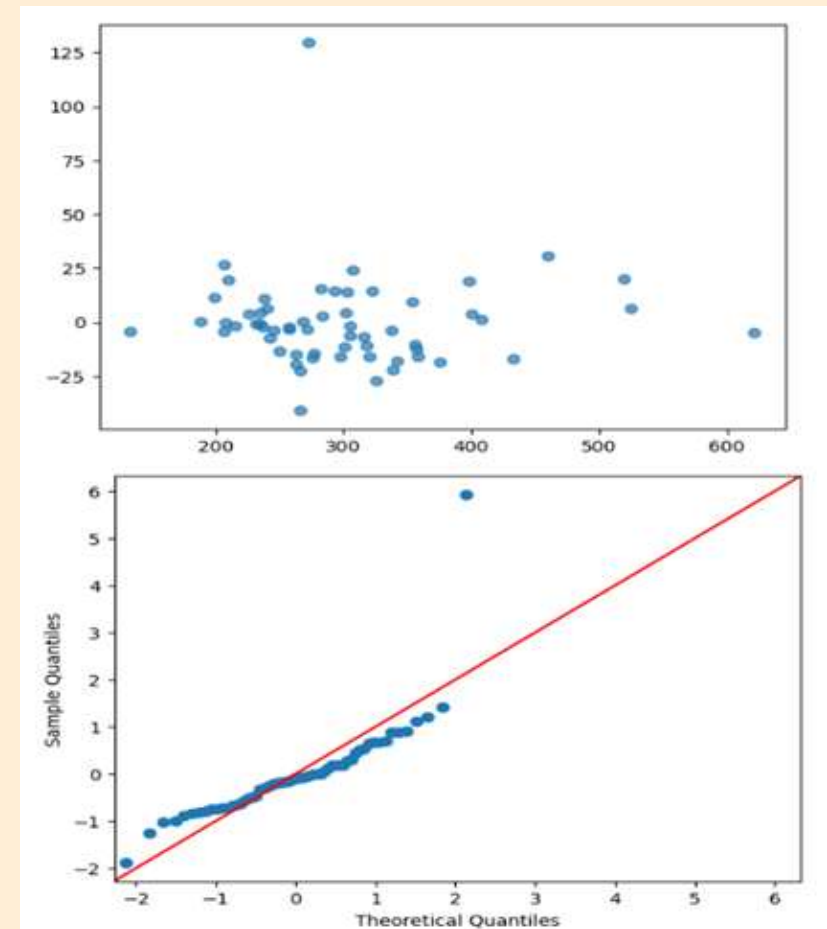
# dictionary to map numbers to week names
week_of_month_map = {
    1: 'Week_1',
    2: 'Week_2',
    3: 'Week_3',
    4: 'Week_4',
    5: 'Week_5'
}

# Create new columns with the descriptive names
df['Day_Name'] = df['Day_of_Week'].map(day_of_week_map)
df['Week_Label'] = df['Week_of_Month'].map(week_of_month_map)
```

PCA + OLS

OLS Regression Results						
=====						
Dep. Variable:	Total_Orders	R-squared:	0.940			
Model:	OLS	Adj. R-squared:	0.926			
Method:	Least Squares	F-statistic:	67.85			
Date:	Sun, 12 Oct 2025	Prob (F-statistic):	2.16e-25			
Time:	13:38:51	Log-Likelihood:	-270.16			
No. Observations:	60	AIC:	564.3			
Df Residuals:	48	BIC:	589.5			
Df Model:	11					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	300.8733	3.152	95.441	0.000	294.535	307.212
PC_1	44.7915	1.733	25.845	0.000	41.307	48.276
PC_2	-0.8888	2.269	-0.392	0.697	-5.450	3.673
PC_3	-0.5334	2.409	-0.221	0.826	-5.377	4.311
PC_4	8.9214	2.662	3.351	0.002	3.568	14.274
PC_5	-1.8695	2.738	-0.683	0.498	-7.374	3.635
PC_6	-3.7675	2.790	-1.350	0.183	-9.377	1.842
PC_7	16.8325	2.965	5.677	0.000	10.871	22.794
PC_8	13.1179	3.651	3.593	0.001	5.776	20.460
PC_9	9.3805	3.764	2.492	0.016	1.813	16.948
PC_10	-13.9901	4.526	-3.091	0.003	-23.090	-4.890
PC_11	9.2650	4.753	1.949	0.057	-0.291	18.821
=====						
Omnibus:	74.820	Durbin-Watson:	2.198			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	931.970			
Skew:	3.365	Prob(JB):	4.22e-203			
Kurtosis:	21.097	Cond. No.	2.74			
=====						



INFLUENCE VARIABLE CHECK

OLS Regression Results

```
=====
Dep. Variable:      Total_Orders    R-squared:              0.981
Model:              OLS             Adj. R-squared:         0.979
Method:             Least Squares   F-statistic:           379.4
Date:               Sun, 12 Oct 2025 Prob (F-statistic):     1.15e-41
Time:               13:38:52        Log-Likelihood:        -231.12
No. Observations:   59             AIC:                     478.2
Df Residuals:       51             BIC:                     494.9
Df Model:           7
Covariance Type:    nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	298.3402	1.704	175.092	0.000	294.919	301.761
PC_1	45.9489	0.934	49.188	0.000	44.073	47.824
PC_4	8.7497	1.426	6.135	0.000	5.886	11.613
PC_7	15.6291	1.592	9.817	0.000	12.433	18.825
PC_8	13.6614	1.957	6.982	0.000	9.733	17.590
PC_9	7.6481	2.022	3.782	0.000	3.588	11.708
PC_10	-10.2112	2.448	-4.171	0.000	-15.126	-5.297
PC_11	5.3125	2.570	2.067	0.044	0.152	10.473

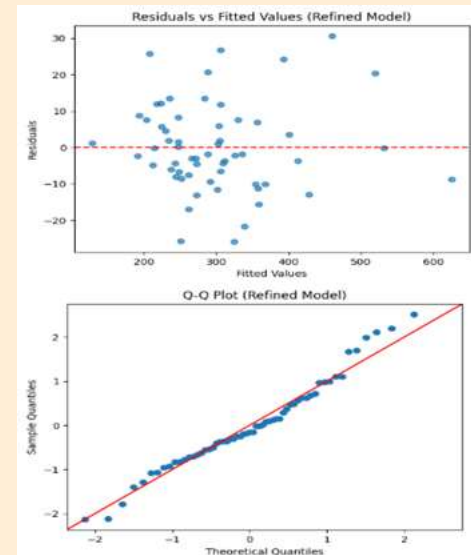
```
=====
Omnibus:           2.217    Durbin-Watson:           2.057
Prob(Omnibus):     0.330    Jarque-Bera (JB):         1.606
Skew:              0.396    Prob(JB):                 0.448
Kurtosis:          3.164    Cond. No.:                 2.76
=====
```

Removing Influencing Variable

```
# Get studentized residuals
influence = pcr_model.get_influence()
studentized_resid = influence.resid_studentized_internal

# Find index of extreme outlier (absolute z-score > 3 often indicates outlier)
outlier_index = np.argmax(np.abs(studentized_resid))
print("Outlier index:", outlier_index, " | Studentized Residual:", studentized_resid[outlier_index])
```

Outlier index: 48 | Studentized Residual: 5.801494291093002



FINAL MODEL

OLS Regression Results

Dep. Variable:	Total_Orders	R-squared:	0.981			
Model:	OLS	Adj. R-squared:	0.979			
Method:	Least Squares	F-statistic:	379.4			
Date:	Sun, 12 Oct 2025	Prob (F-statistic):	1.15e-41			
Time:	13:38:52	Log-Likelihood:	-231.12			
No. Observations:	59	AIC:	478.2			
Df Residuals:	51	BIC:	494.9			
Df Model:	7					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	298.3402	1.704	175.092	0.000	294.919	301.761
PC_1	45.9489	0.934	49.188	0.000	44.073	47.824
PC_4	8.7497	1.426	6.135	0.000	5.886	11.613
PC_7	15.6291	1.592	9.817	0.000	12.433	18.825
PC_8	13.6614	1.957	6.982	0.000	9.733	17.590
PC_9	7.6481	2.022	3.782	0.000	3.588	11.708
PC_10	-10.2112	2.448	-4.171	0.000	-15.126	-5.297
PC_11	5.3125	2.570	2.067	0.044	0.152	10.473
=====						
Omnibus:	2.217	Durbin-Watson:	2.057			
Prob(Omnibus):	0.330	Jarque-Bera (JB):	1.606			
Skew:	0.396	Prob(JB):	0.448			
Kurtosis:	3.164	Cond. No.	2.76			
=====						

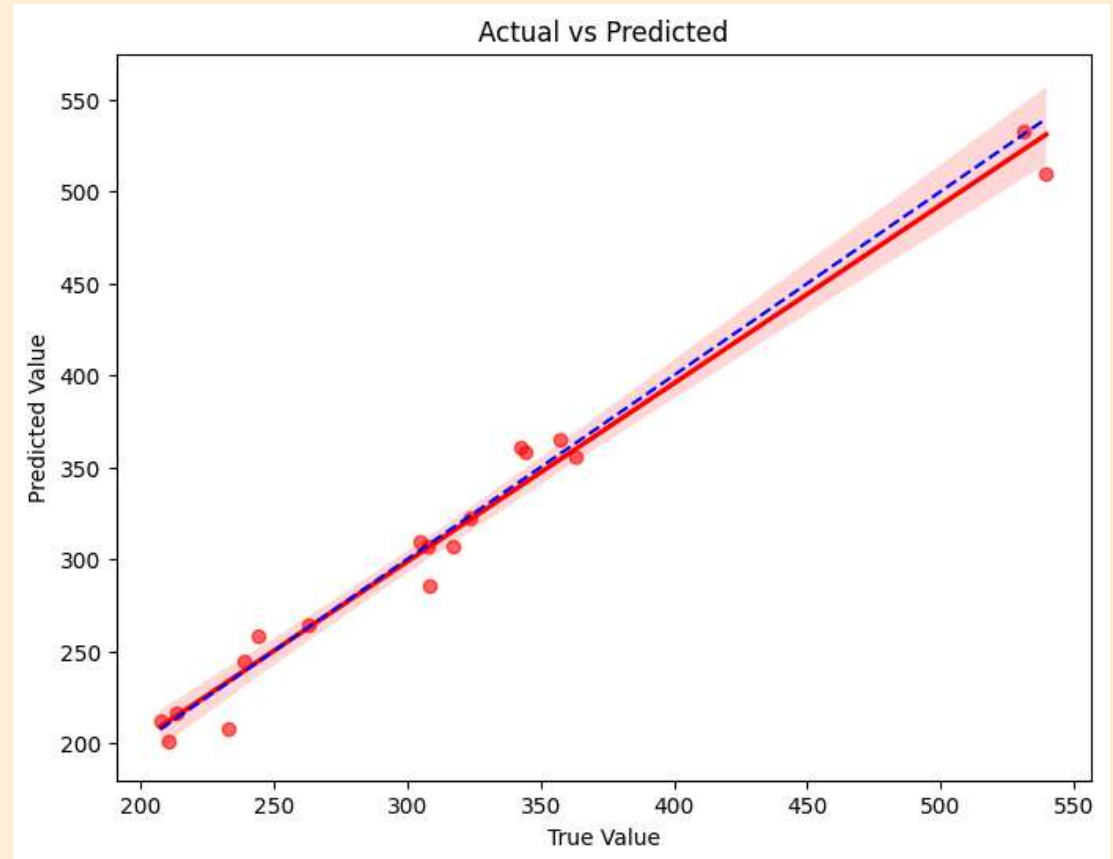
--- Influence of Original Variables on Total Orders ---

Non_Urgent_Order	31.457419
Banking_Orders_1	25.011917
Banking_Orders_2	23.872573
Urgent_Order	21.839431
Traffic_Controller_Orders	7.275800
Day_Name_Monday	5.649110
Week_Label_Week_3	4.126643
Banking_Orders_3	3.389276
Day_Name_Thursday	3.270676
Week_Label_Week_2	0.209558
Day_Name_Tuesday	-0.112091
Day_Name_Wednesday	-0.303472
Week_Label_Week_4	-1.746747
Week_Label_Week_5	-2.666710
Fiscal_Sector_Orders	-3.652936

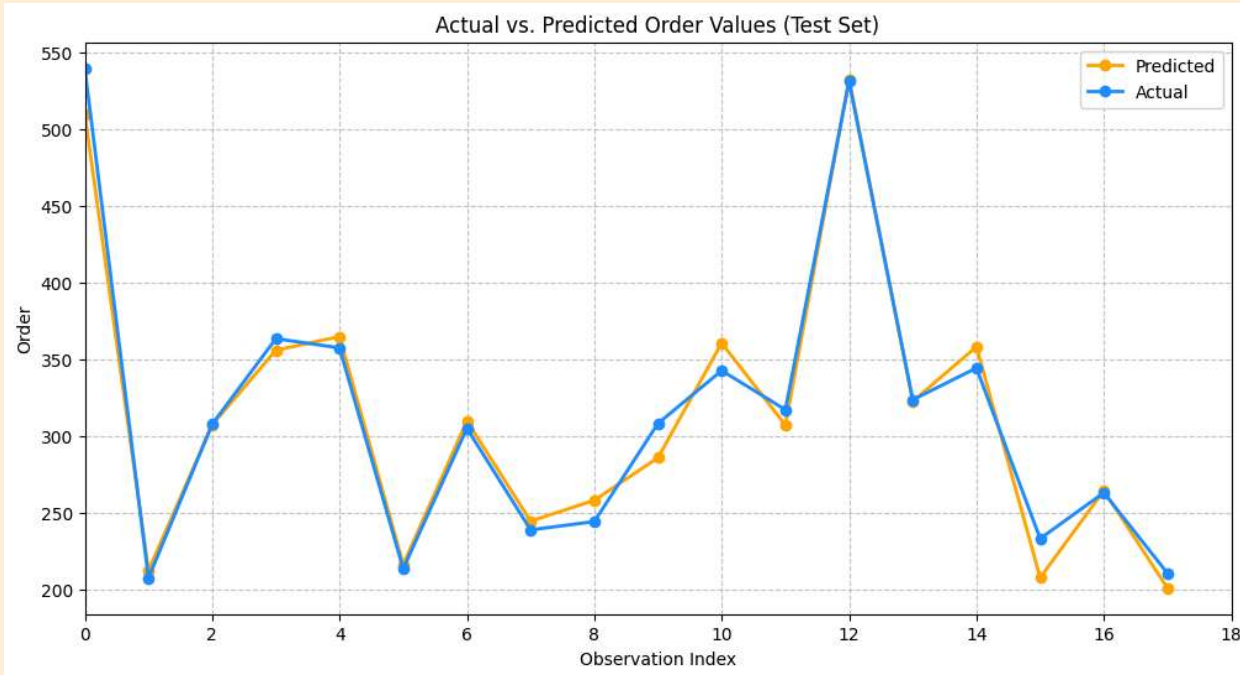
FORECASTING RESULTS

$$RMSE = 13.1988\%$$

$$R^2 = 98.1\%$$



FORECASTING RESULTS



	True Value	Predicted Value	Residual
0	539.577	509.852496	29.724504
5	207.364	212.333196	-4.969196
34	307.645	307.260429	0.384571
13	363.402	355.852349	7.549651
45	357.394	364.763027	-7.369027

**ANY
QUESTIONS..?**

THANK YOU