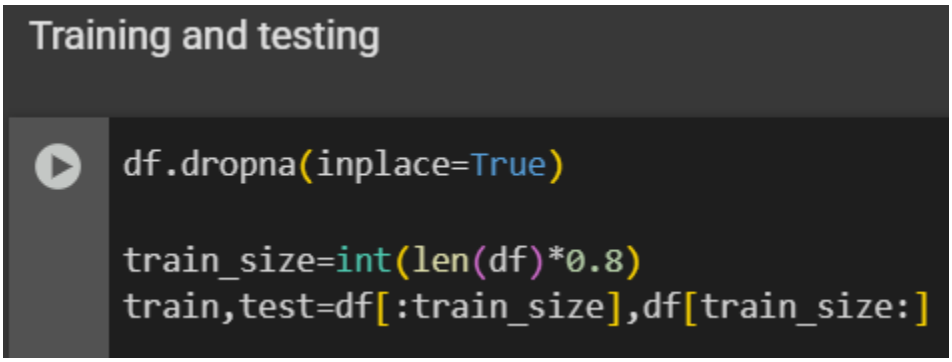
Model Development Phase Template

Date	23 June 2025
Team	**AS PS VS VV**
Project Title	Unemployed Insurance Beneficiary Forecasting
Maximum Marks	10 Marks

Initial Model Training Code, Model Validation and Evaluation Report

The initial model training code will be showcased in the future through a screenshot. The model validation and evaluation report will include a summary and training and validation performance metrics for multiple models, presented through respective screenshots.

Initial Model Training Code (5 marks):



Model building

```
[40] #Augmented Dickey-Fuller test(used to convert non-stationary data to
     #stationary data)

     adf=adfuller(df['Beneficiaries'],autolag='AIC')
     print("P-Value",adf[1])

     adf=adfuller(train['Beneficiaries_diff'],autolag='AIC')
     print("P-Value",adf[1])

     #ACF and PACF(to check how our data is correlated with ACF &PACF)
     plot_acf(train['Beneficiaries'], lags=30, title='Original ACF')
     plot_pacf(train['Beneficiaries'], lags=30, title='Original PACF')
     plt.show()
     #differenced ACF and PCAF
     plot_acf(train['Beneficiaries_diff'], lags=30, title='Differenced ACF')
     plot_pacf(train['Beneficiaries_diff'], lags=30, title='Differenced PACF')
     plt.show()
     #Augmented Dickey-Fuller test(used to convert non-stationary data to
     #stationary data)

     adf=adfuller(df['Beneficiaries'],autolag='AIC')
     print("P-Value",adf[1])

     adf=adfuller(train['Beneficiaries_diff'],autolag='AIC')
     print("P-Value",adf[1])

     #ACF and PACF(to check how our data is correlated with ACF &PACF)
     plot_acf(train['Beneficiaries'], lags=30, title='Original ACF')
     plot_pacf(train['Beneficiaries'], lags=30, title='Original PACF')
     plt.show()
     #differenced ACF and PCAF
     plot_acf(train['Beneficiaries_diff'].dropna(), lags=30, title='Differenced ACF')
     plot_pacf(train['Beneficiaries_diff'].dropna(), lags=30, title='Differenced PACF')
     plt.show()


     #smoothing out our data(visual representation)
     plt.plot(train['Beneficiaries'])
     plt.plot(train['Beneficiaries_diff'])
     plt.show()
```

```
#ARIMA
from statsmodels.tsa.arima.model import ARIMA
model = ARIMA(train['Beneficiaries_diff'].dropna(), order=(5,0,0)) # Pass the series with the correct index
model_arima=model.fit()
model_arima.summary()
```

```
#SARIMA
model=SARIMAX(train['Beneficiaries_diff'].dropna(),order=(5,0,0),seasonal_order=(0,1,2,3))
model_sarima=model.fit()
model_sarima.summary()
```

```
#Auto Regression
model_ar=AutoReg(train['Beneficiaries_diff'].dropna(), lags=10).fit()
model_ar.summary()
```

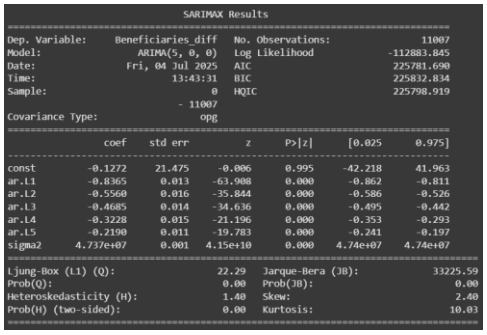
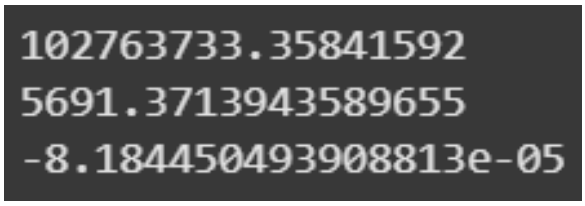
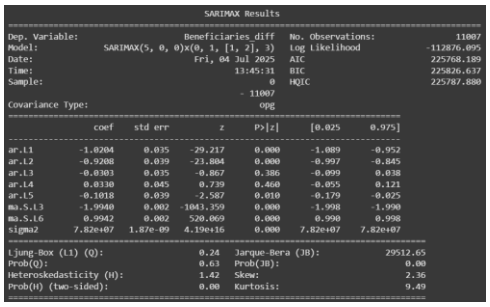
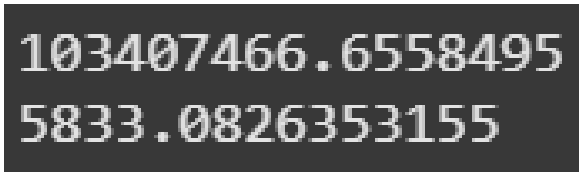
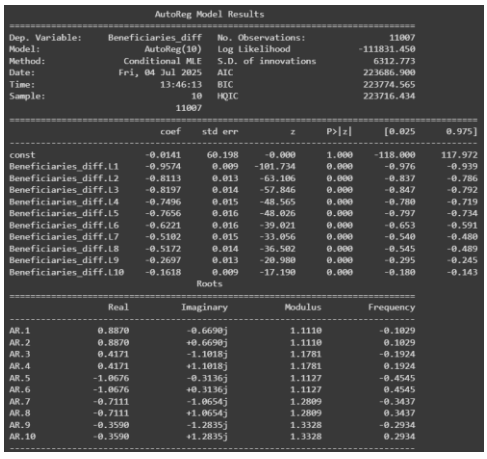
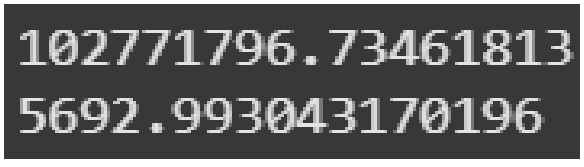
```
#VAR
model =VAR(train[['Beneficiaries_diff','Benefit Amounts (Dollars)']].dropna()) # Add dropna()
model_AR = model.fit(maxlags=10)
model_AR.summary()
```

```
import prophet

# Prepare the data for Prophet
# Prophet requires the dataframe to have columns named 'ds' and 'y'
prophet_df = df[['Year', 'Month', 'Beneficiaries']].copy()
prophet_df['ds'] = pd.to_datetime(prophet_df['Year'].astype(str) + '-' + prophet_df['Month'].astype(str))
prophet_df = prophet_df[['ds', 'Beneficiaries']].rename(columns={'Beneficiaries': 'y'})

# Instantiate and fit the Prophet model
model_prophet = prophet.Prophet()
model_prophet.fit(prophet_df)
```

Model Validation and Evaluation Report (5 marks):

Model	Summary	Training and Validation Performance Metrics (MSE,MAE,R2 score)
ARIMA		
SARIMA		 R2-score: NA
AutoReg		 R2-score: NA

SARIMAX Results

Dep. Variable:	Beneficiaries_diff	No. Observations:	11007
Model:	ARIMA(5, 0, 0)	Log Likelihood	-112883.845
Date:	Fri, 04 Jul 2025	AIC	225781.690
Time:	13:43:31	BIC	225832.834
Sample:	0	HQIC	225798.919
	- 11007		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.1272	21.475	-0.006	0.995	-42.218	41.963
ar.L1	-0.8365	0.013	-63.908	0.000	-0.862	-0.811
ar.L2	-0.5560	0.016	-35.844	0.000	-0.586	-0.526
ar.L3	-0.4685	0.014	-34.636	0.000	-0.495	-0.442
ar.L4	-0.3228	0.015	-21.196	0.000	-0.353	-0.293
ar.L5	-0.2190	0.011	-19.783	0.000	-0.241	-0.197
sigma2	4.737e+07	0.001	4.15e+10	0.000	4.74e+07	4.74e+07

Ljung-Box (L1) (Q):	22.29	Jarque-Bera (JB):	33225.59
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	1.40	Skew:	2.40
Prob(H) (two-sided):	0.00	Kurtosis:	10.03

```
102763733.35841592
5691.3713943589655
-8.184450493908813e-05
```

SARIMAX Results

Dep. Variable:	Beneficiaries_diff	No. Observations:	11007
Model:	SARIMAX(5, 0, 0)x(0, 1, [1, 2], 3)	Log Likelihood	-112876.095
Date:	Fri, 04 Jul 2025	AIC	225768.189
Time:	13:45:11	BIC	225826.637
Sample:	0	HQIC	225787.880
	- 11007		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-1.0204	0.035	-29.217	0.000	-1.089	-0.952
ar.L2	-0.9208	0.039	-23.804	0.000	-0.997	-0.845
ar.L3	-0.0303	0.035	-0.867	0.386	-0.099	0.038
ar.L4	0.0338	0.045	0.739	0.460	-0.055	0.121
ar.L5	-0.1018	0.039	-2.587	0.010	-0.179	-0.025
ma.S.L3	-1.9940	0.002	-1043.359	0.000	-1.998	-1.990
ma.S.L6	0.9942	0.002	520.069	0.000	0.990	0.998
sigma2	7.82e+07	1.87e-09	4.19e+16	0.000	7.82e+07	7.82e+07

Ljung-Box (L1) (Q):	0.24	Jarque-Bera (JB):	29512.65
Prob(Q):	0.63	Prob(JB):	0.00
Heteroskedasticity (H):	1.42	Skew:	2.36
Prob(H) (two-sided):	0.00	Kurtosis:	9.49

```
103407466.6558495
5833.0826353155
```

AutoReg Model Results

Dep. Variable:	Beneficiaries_diff	No. Observations:	11007
Model:	AutoReg(10)	Log Likelihood	-111831.450
Method:	Conditional MLE	S.D. of innovations	6312.773
Date:	Fri, 04 Jul 2025	AIC	223686.900
Time:	13:46:13	BIC	223774.565
Sample:	10	HQIC	223716.434
	11007		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0141	60.198	-0.000	1.000	-118.000	117.972
Beneficiaries_diff.L1	-0.9574	0.009	-101.734	0.000	-0.976	-0.939
Beneficiaries_diff.L2	-0.8113	0.013	-63.106	0.000	-0.837	-0.786
Beneficiaries_diff.L3	-0.8197	0.014	-57.846	0.000	-0.847	-0.792
Beneficiaries_diff.L4	-0.7496	0.015	-48.565	0.000	-0.780	-0.719
Beneficiaries_diff.L5	-0.7656	0.016	-48.026	0.000	-0.797	-0.734
Beneficiaries_diff.L6	-0.6221	0.016	-39.021	0.000	-0.653	-0.591
Beneficiaries_diff.L7	-0.5102	0.015	-33.056	0.000	-0.540	-0.480
Beneficiaries_diff.L8	-0.5172	0.014	-36.502	0.000	-0.545	-0.489
Beneficiaries_diff.L9	-0.2697	0.013	-20.980	0.000	-0.295	-0.245
Beneficiaries_diff.L10	-0.1618	0.009	-17.190	0.000	-0.180	-0.143

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	0.8870	-0.6690j	1.1110	-0.1029
AR.2	0.8870	+0.6690j	1.1110	0.1029
AR.3	0.4171	-1.1018j	1.1781	-0.1924
AR.4	0.4171	+1.1018j	1.1781	0.1924
AR.5	-1.0676	-0.3136j	1.1127	-0.4545
AR.6	-1.0676	+0.3136j	1.1127	0.4545
AR.7	-0.7111	-1.0654j	1.2809	-0.3437
AR.8	-0.7111	+1.0654j	1.2809	0.3437
AR.9	-0.3590	-1.2835j	1.3328	-0.2934
AR.10	-0.3590	+1.2835j	1.3328	0.2934

```
102771796.73461813
5692.993043170196
```

VAR	Summary of Regression Results Model: VAR Method: OLS Date: Fri, 04, Jul, 2025 Time: 13:46:48 No. of Equations: 2.00000 BIC: 44.6396 Nobs: 10997.0 HQIC: 44.6211 Log likelihood: -276464. FPE: 2.36930e+19 AIC: 44.6117 Det(Omega_mle): 2.36028e+19 Results for equation Beneficiaries_diff coefficient std. error t-stat prob const 2437.660425 113.470188 21.483 0.000 L1.Beneficiaries_diff -0.958505 0.071387 -13.427 0.000 L1.Benefit Amounts (Dollars) -0.000067 0.000068 -0.992 0.321 L2.Beneficiaries_diff -2.515005 0.091166 -27.587 0.000 L2.Benefit Amounts (Dollars) 0.001572 0.000065 24.334 0.000 L3.Beneficiaries_diff -2.429829 0.103921 -23.381 0.000 L3.Benefit Amounts (Dollars) -0.000167 0.000067 -2.514 0.012 L4.Beneficiaries_diff -2.233172 0.113941 -19.599 0.000 L4.Benefit Amounts (Dollars) -0.000159 0.000066 -2.409 0.016 L5.Beneficiaries_diff -3.068403 0.114709 -26.749 0.000 L5.Benefit Amounts (Dollars) 0.000730 0.000066 10.985 0.000 L6.Beneficiaries_diff -2.410467 0.115130 -20.937 0.000 L6.Benefit Amounts (Dollars) -0.000528 0.000067 -7.921 0.000 L7.Beneficiaries_diff -1.986027 0.106088 -18.721 0.000 L7.Benefit Amounts (Dollars) -0.000319 0.000066 -4.844 0.000 L8.Beneficiaries_diff -1.360005 0.093685 -14.517 0.000 L8.Benefit Amounts (Dollars) -0.000688 0.000065 -10.528 0.000 L9.Beneficiaries_diff -0.058414 0.072965 -0.801 0.423 L9.Benefit Amounts (Dollars) -0.001073 0.000065 -16.440 0.000 L10.Beneficiaries_diff -0.061638 0.009341 -6.598 0.000 L10.Benefit Amounts (Dollars) 0.000060 0.000069 0.873 0.383 Results for equation Benefit Amounts (Dollars) coefficient std. error t-stat prob const 2527995.102185 119628.595562 21.132 0.000 L1.Beneficiaries_diff -55.341014 75.261890 -0.735 0.462 L1.Benefit Amounts (Dollars) 0.024889 0.071381 0.349 0.727 L2.Beneficiaries_diff -1822.686324 96.113910 -18.964 0.000 L2.Benefit Amounts (Dollars) 1.775600 0.068097 26.075 0.000 L3.Beneficiaries_diff -1718.386647 109.561484 -15.684 0.000 L3.Benefit Amounts (Dollars) -0.190232 0.070229 -2.709 0.007 L4.Beneficiaries_diff -1683.508188 120.125480 -14.015 0.000 L4.Benefit Amounts (Dollars) -0.018831 0.069716 -0.270 0.787 L5.Beneficiaries_diff -2611.555331 120.934946 -21.595 0.000 L5.Benefit Amounts (Dollars) 0.808297 0.070082 11.534 0.000 L6.Beneficiaries_diff -2010.644795 121.378814 -16.565 0.000 L6.Benefit Amounts (Dollars) -0.454442 0.070274 -6.467 0.000 L7.Beneficiaries_diff -1625.429682 111.846027 -14.533 0.000 L7.Benefit Amounts (Dollars) -0.276459 0.069366 -3.985 0.000 L8.Beneficiaries_diff -986.981249 98.769580 -9.993 0.000 L8.Benefit Amounts (Dollars) -0.716668 0.068939 -10.396 0.000 L9.Beneficiaries_diff 69.904650 76.924990 0.909 0.363 L9.Benefit Amounts (Dollars) -0.815057 0.068831 -11.841 0.000 L10.Beneficiaries_diff -61.923830 9.648272 -6.288 0.000 L10.Benefit Amounts (Dollars) 0.199110 0.072836 2.734 0.006 Correlation matrix of residuals Beneficiaries_diff Benefit Amounts (Dollars) Beneficiaries_diff 1.000000 0.991185 Benefit Amounts (Dollars) 0.991185 1.000000	NA
Prophet	ds yhat yhat_lower yhat_upper 0 2001-01-01 4850.952377 -2818.407480 13519.833896 1 2001-02-01 4636.011708 -3680.946446 12425.524121 2 2001-03-01 4531.799997 -3757.700050 12761.117768 3 2001-04-01 4423.377526 -4139.662938 13044.800193 4 2001-05-01 3827.318423 -4071.144641 12201.204735	49080877.302183054 4541.665548954107 0.0033392111908991 94