Group Project no 13 – Police gun shooting (2000-2019)
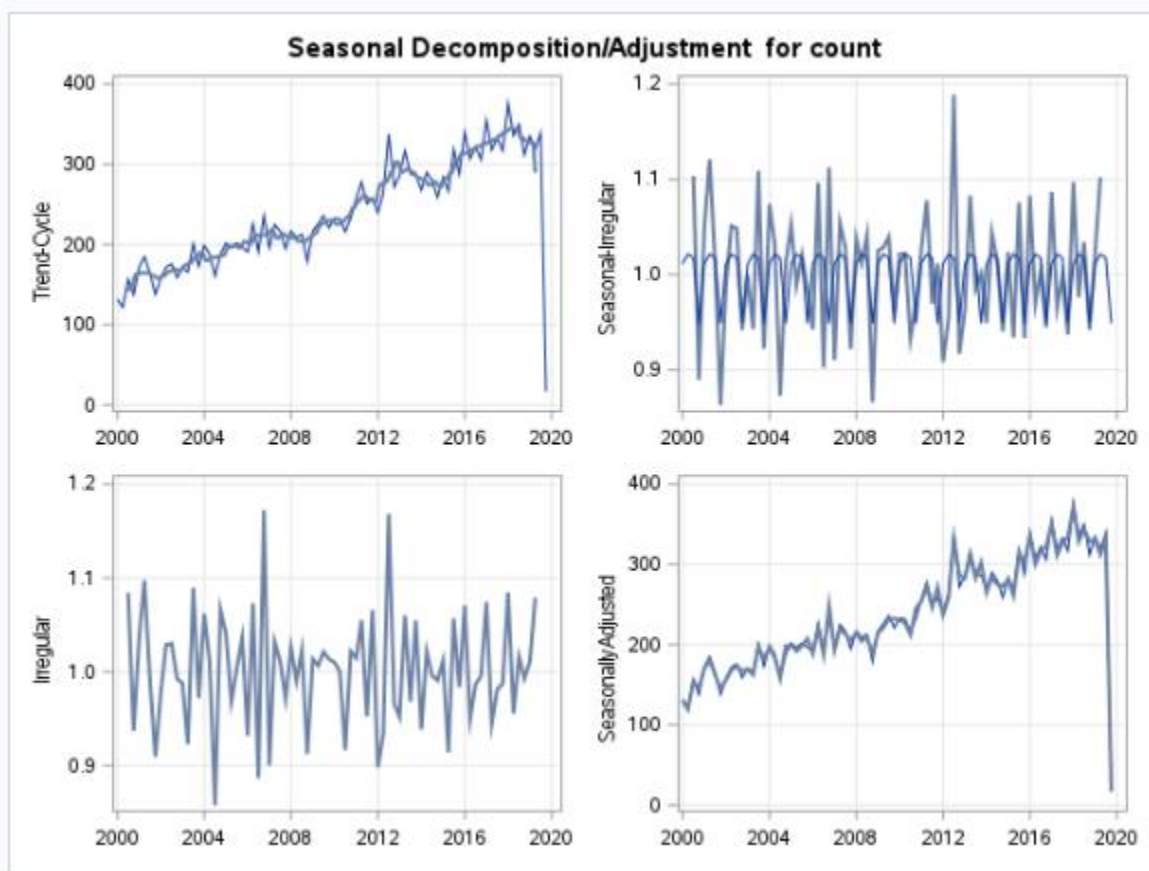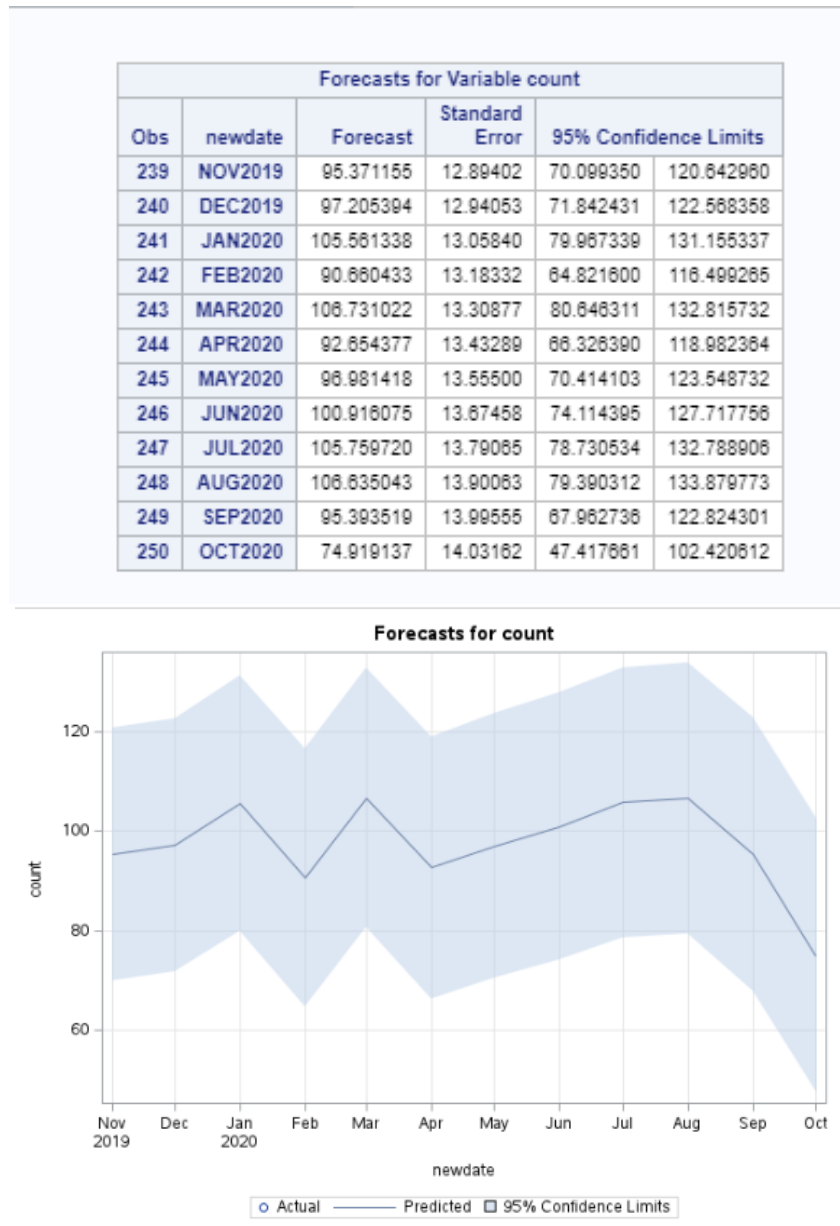
/* Explored the time series signal components for complete data from 2000 – 2019 */

Trend, Seasonality and Cycle charts

Simple Forecast based on the available data – 12 months forecast

			Forecasts for Variable count		
Obs	newdate	Forecast	Standard Error	95% Confidence Limits	
239	NOV2019	95.371155	12.89402	70.099350	120.642960
240	DEC2019	97.205394	12.94053	71.842431	122.568358
241	JAN2020	105.561338	13.05840	79.967339	131.155337
242	FEB2020	90.660433	13.18332	64.821600	116.499265
243	MAR2020	106.731022	13.30877	80.646311	132.815732
244	APR2020	92.654377	13.43289	66.326390	118.982364
245	MAY2020	96.981418	13.55500	70.414103	123.548732
246	JUN2020	100.916075	13.67458	74.114395	127.717756
247	JUL2020	105.759720	13.79065	78.730534	132.788906
248	AUG2020	106.635043	13.90063	79.390312	133.879773
249	SEP2020	95.393519	13.99555	67.962738	122.824301
250	OCT2020	74.919137	14.03162	47.417661	102.420612

Forecasts for count

Things to do this week - To create fit and holdout sample and understand the differences in output between models like ARIM, ARIMAX and Exponential smoothing

<mark>Segment 2</mark>

/* Created age groups to see the various signal components like Trend, seasonality and cycle */

Age below 19 are categorized as agegroup1 and created 9 groups to analyze based on age and forecast the expected incidents for next 12 months

Number of incidents based on Age group

agegrp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1308	6.88	1308	6.88
2	5728	30.15	7036	37.03
3	4623	24.33	11659	61.36
4	3147	16.56	14806	77.93
5	1824	9.60	16630	87.53
6	645	3.39	17275	90.92
7	207	1.09	17482	92.01
8	50	0.26	17532	92.27
9	1468	7.73	19000	100.00

Trend, Seasonality and Cycle charts for group 2 and 3 which has high number of incidents
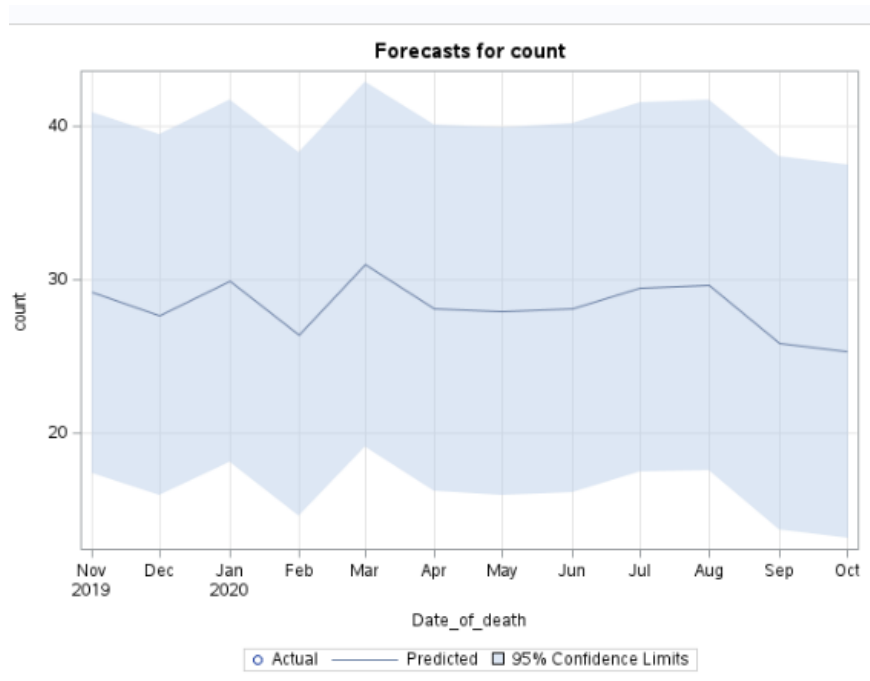
Age group - 2

Age group - 3

Seasonal Decomposition/Adjustment for count
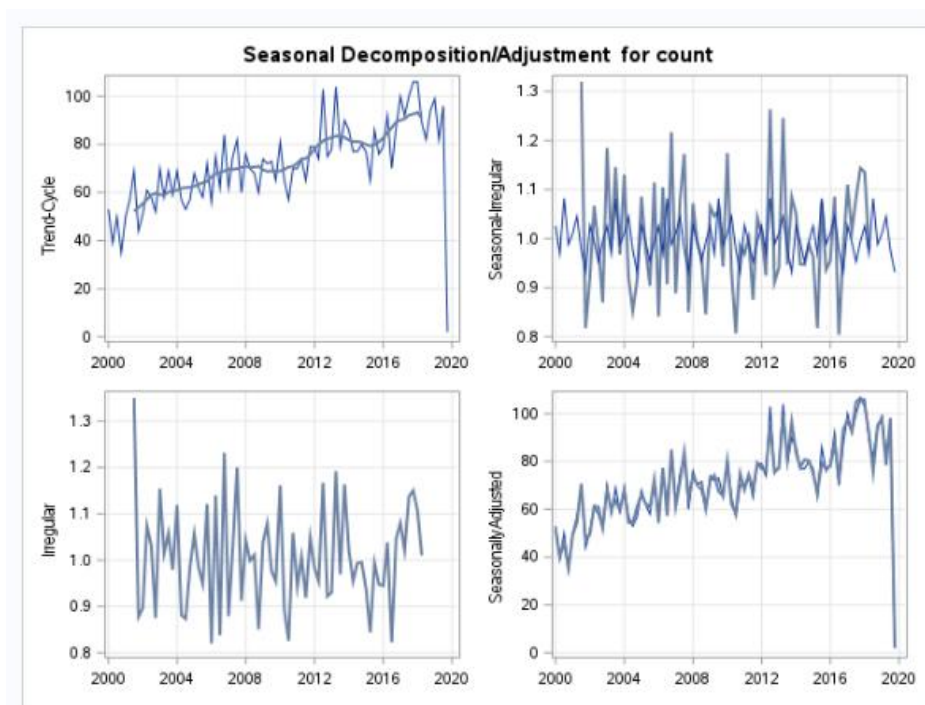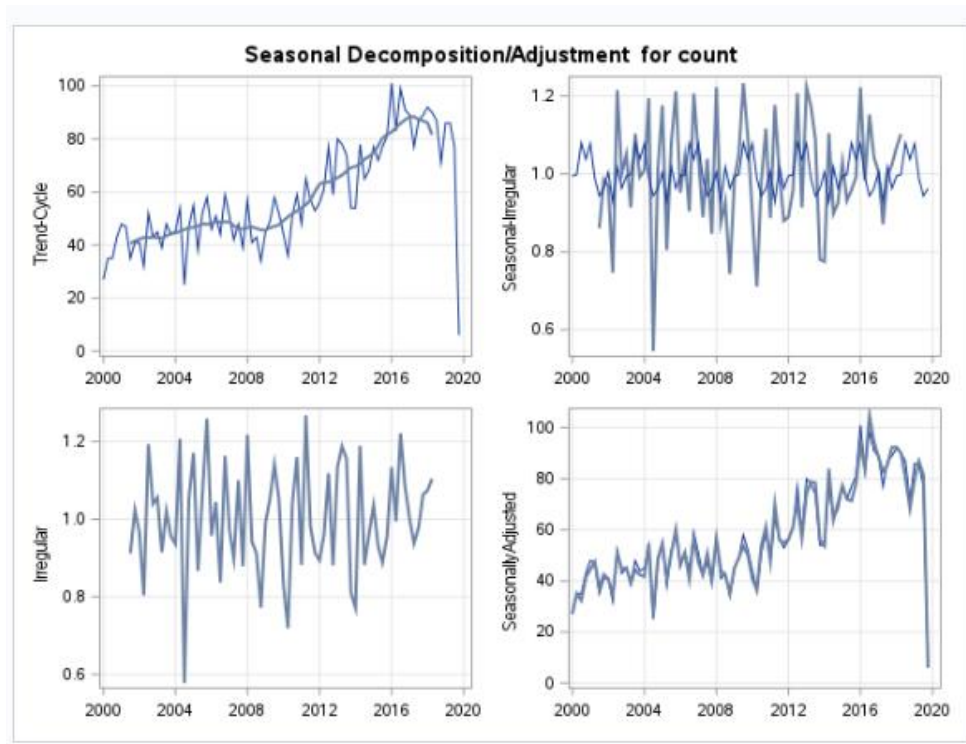
Forecast based on the age group data (Group 2 and 3) - 12 months forecast

Age group -2

		Forecasts for Variable count			
Obs	Date_of_death	Forecast	Standard Error	95% Confidence Limits	
239	NOV2019	29.170912	5.98064	17.449069	40.892755
240	DEC2019	27.694712	5.99585	15.943066	39.446357
241	JAN2020	29.934695	6.01601	18.143523	41.725868
242	FEB2020	26.414203	6.03886	14.578257	38.250150
243	MAR2020	31.001481	6.06147	19.121212	42.881750
244	APR2020	28.163329	6.08381	16.239286	40.087373
245	MAY2020	27.950138	6.10588	15.982840	39.917436
246	JUN2020	28.138591	6.12768	16.128550	40.148633
247	JUL2020	29.510457	6.14923	17.458192	41.562722
248	AUG2020	29.654364	6.17045	17.560504	41.748224
249	SEP2020	25.877738	6.19083	13.743934	38.011542
250	OCT2020	25.321945	6.20399	13.162344	37.481547

Age group – 3

Forecasts for Variable count					
Obs	Date_of_death	Forecast	Standard Error	95% Confidence Limits	
239	NOV2019	22.202659	4.96213	12.477057	31.928262
240	DEC2019	25.355591	5.00586	15.544282	35.166901
241	JAN2020	25.239560	5.04401	15.353484	35.125636
242	FEB2020	24.634822	5.08625	14.665957	34.603686
243	MAR2020	25.180141	5.12780	15.129829	35.230453
244	APR2020	22.925421	5.16869	12.794970	33.055871
245	MAY2020	24.970725	5.20893	14.761414	35.180036
246	JUN2020	25.016041	5.24853	14.729119	35.302963
247	JUL2020	27.111330	5.28750	16.748017	37.474643
248	AUG2020	23.556624	5.32587	13.118114	33.995133
249	SEP2020	23.601933	5.36364	13.089395	34.114471
250	OCT2020	23.597191	5.40082	13.011769	34.182612
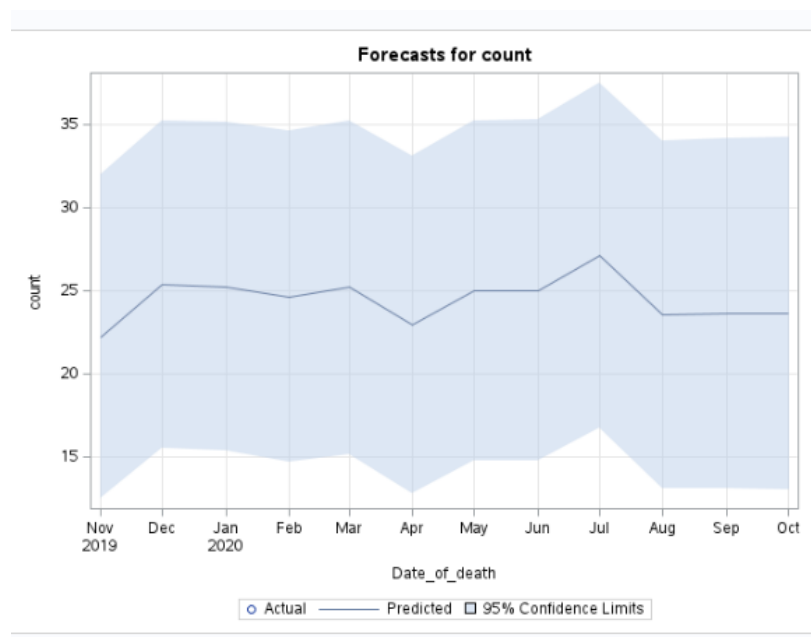
Forecasts for count

Things to do this week - To create fit and holdout sample and understand the differences in output between models like ARIM, ARIMAX and Exponential smoothing

/* Explored the data based on gender to see the various signal components like Trend, seasonality and cycle */

Gender – Female
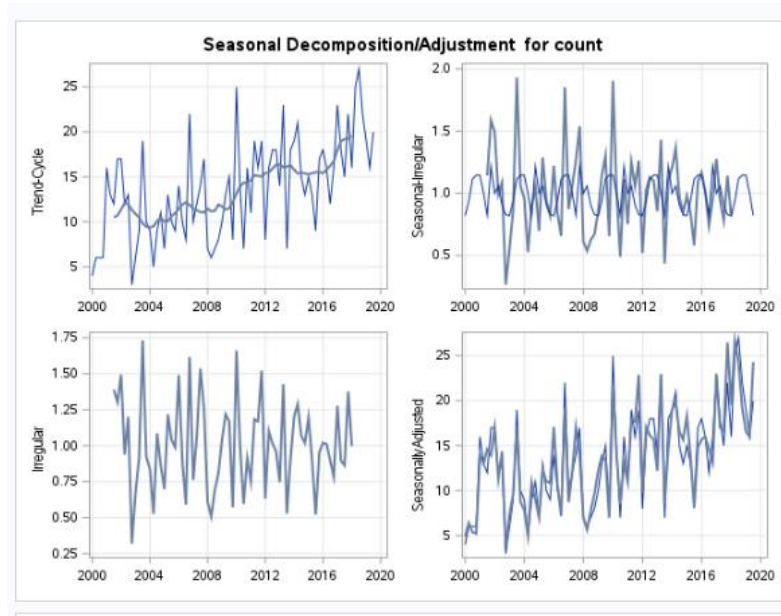
Gender – Male
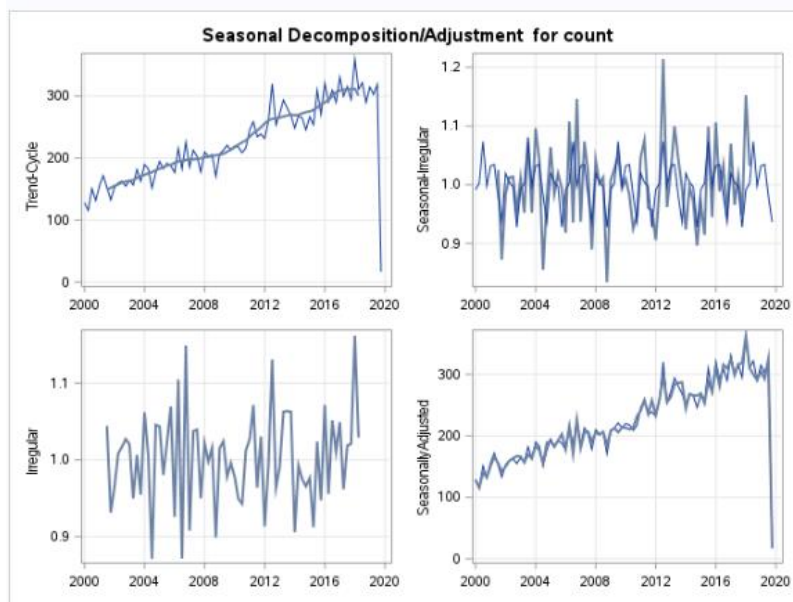
Forecast based on the gender - 12 month forecast

Gender – Female

			Standard	95% Confidence Limits	
Obs	Date_of_death	Forecast	Error		
238	OCT2019	6.999419	2.67030	1.765720	12.233117
239	NOV2019	6.089466	2.67376	0.848994	11.329938
240	DEC2019	7.021741	2.67718	1.774563	12.268920
241	JAN2020	6.528969	2.67785	1.280484	11.777455
242	FEB2020	6.364569	2.68120	1.109507	11.619631
243	MAR2020	6.400197	2.68453	1.138621	11.661773
244	APR2020	6.135782	2.68782	0.867752	11.403811
245	MAY2020	7.471406	2.69108	2.196983	12.745828
246	JUN2020	6.156980	2.69431	0.876226	11.437733
247	JUL2020	6.742638	2.69751	1.455613	12.029664
248	AUG2020	5.528197	2.70068	0.234961	10.821434
249	SEP2020	6.313782	2.70382	1.014395	11.613169

Forecasts for Variable count

Forecasts for count
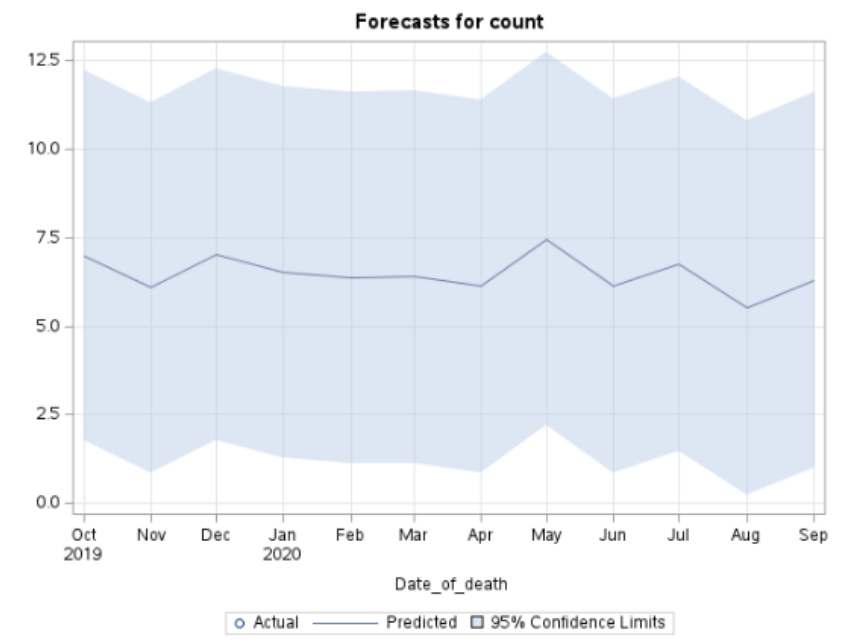
Gender – Male

		Forecasts for Variable count			
Obs	Date_of_death	Forecast	Standard Error	95% Confidence Limits	
239	NOV2019	90.720188	11.91644	67.364402	114.075974
240	DEC2019	90.279693	11.95772	66.842986	113.716401
241	JAN2020	100.776033	12.06616	77.126803	124.425264
242	FEB2020	85.271607	12.18126	61.396784	109.146430
243	MAR2020	100.861976	12.29693	76.760443	124.963510
244	APR2020	87.916138	12.41141	63.590219	112.242057
245	MAY2020	90.170202	12.52404	65.623532	114.716871
246	JUN2020	96.227195	12.63431	71.464410	120.989979
247	JUL2020	98.646689	12.74127	73.674265	123.619112
248	AUG2020	102.403332	12.84244	77.232612	127.574053
249	SEP2020	90.021358	12.92938	64.680231	115.362486
250	OCT2020	69.742949	12.96145	44.338970	95.146929
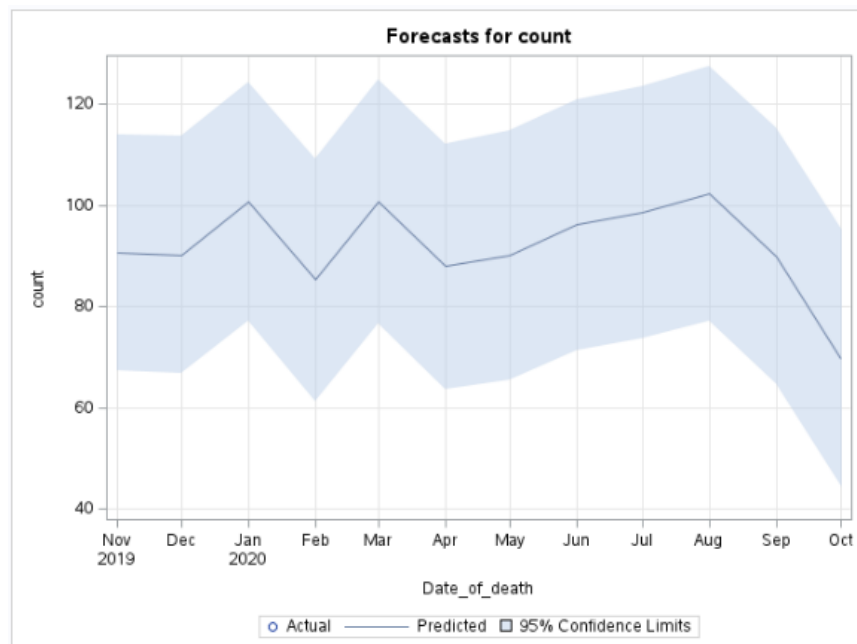
Things to be done:
- To create fit and holdout sample and understand the differences in output between models like ARIM, ARIMAX and Exponential smoothing
- Results of the Text analysis on the description part must be channelized with the output of final forecasts

Team: tasks to be completed by this week

1. **Alex** – Analysis on state and county wise proportion to forecast incidents across counties
2. **Aravind** – Forecast the Segment 1,2,3 across different models (ARIMA, ARIMAX) and to forecast the final model across the top counties with maximum incidents
3. **Mason** – Analysis on the age group and census data to channelize with the output of the final forecasts

Decision tree

Split Node 1   ×

Target Variable: Disposition

Variable	Variable Description	-Log(p)	Branches
Race	Race	97.8048	2
State	State	66.14	2
IMP_Mental_illness	Imputed Mental_illness	22.2092	2
County	County	10.4461	2
Gender	Gender	2.7138	2
agegrp	agegrp	1.0421	2
City	City	0.2614	2

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Disposition	Disposition	_NOBS_	Sum of Frequencies	9145	3921	.
Disposition	Disposition	_MISC_	Misclassification Rate	0.364571	0.364193	.
Disposition	Disposition	_MAX_	Maximum Absolute Error	0.828829	0.828829	.
Disposition	Disposition	_SSE_	Sum of Squared Errors	4042.405	1728.685	.
Disposition	Disposition	_ASE_	Average Squared Error	0.221017	0.220439	.
Disposition	Disposition	_RASE_	Root Average Squared Error	0.470125	0.46951	.
Disposition	Disposition	_DIV_	Divisor for ASE	18290	7842	.
Disposition	Disposition	_DFT_	Total Degrees of Freedom	9145	.	.

```
Event Classification Table

Data Role=TRAIN Target=Disposition Target Label=Disposition

  False        True         False        True
Negative     Negative     Positive     Positive

  1605         3104         1747         2689


Data Role=VALIDATE Target=Disposition Target Label=Disposition

  False        True         False        True
Negative     Negative     Positive     Positive

   659         1338          742         1182
```

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
State	State	4	1.0000	0.9815	0.9815
Race	Race	1	0.9987	1.0000	1.0013
IMP_Mental_illness	Imputed Mental_illness	3	0.4551	0.2416	0.5309
agegrp		1	0.2319	0.1780	0.7674
Gender	Gender	1	0.1190	0.0094	0.0792
County	County	0	0.0000	0.0000	.
City	City	0	0.0000	0.0000	.