



# How Did the COVID-Pandemic Impact Water Usage in Long Beach, CA?

By: Ensieh Bahrami  
Data Science Intensive Capstone Project

# The Problem

Water usage by 6 type of customers within three years  
(2017 - 2020)

**Pick daily usage:**  
**June, July, August**



- Did COVID-Pandemic affect water usage?
- What factors affects water consumption?
- Can I predict water usage during the pandemic?

# Who might care?

- Water Department

Find effective ways to provide reliable water supply

- Customers

Manage their usage and save on water bills

- Government

Effectively allocate budget to provide the required Facilities and contribute into economy growth

## Residential



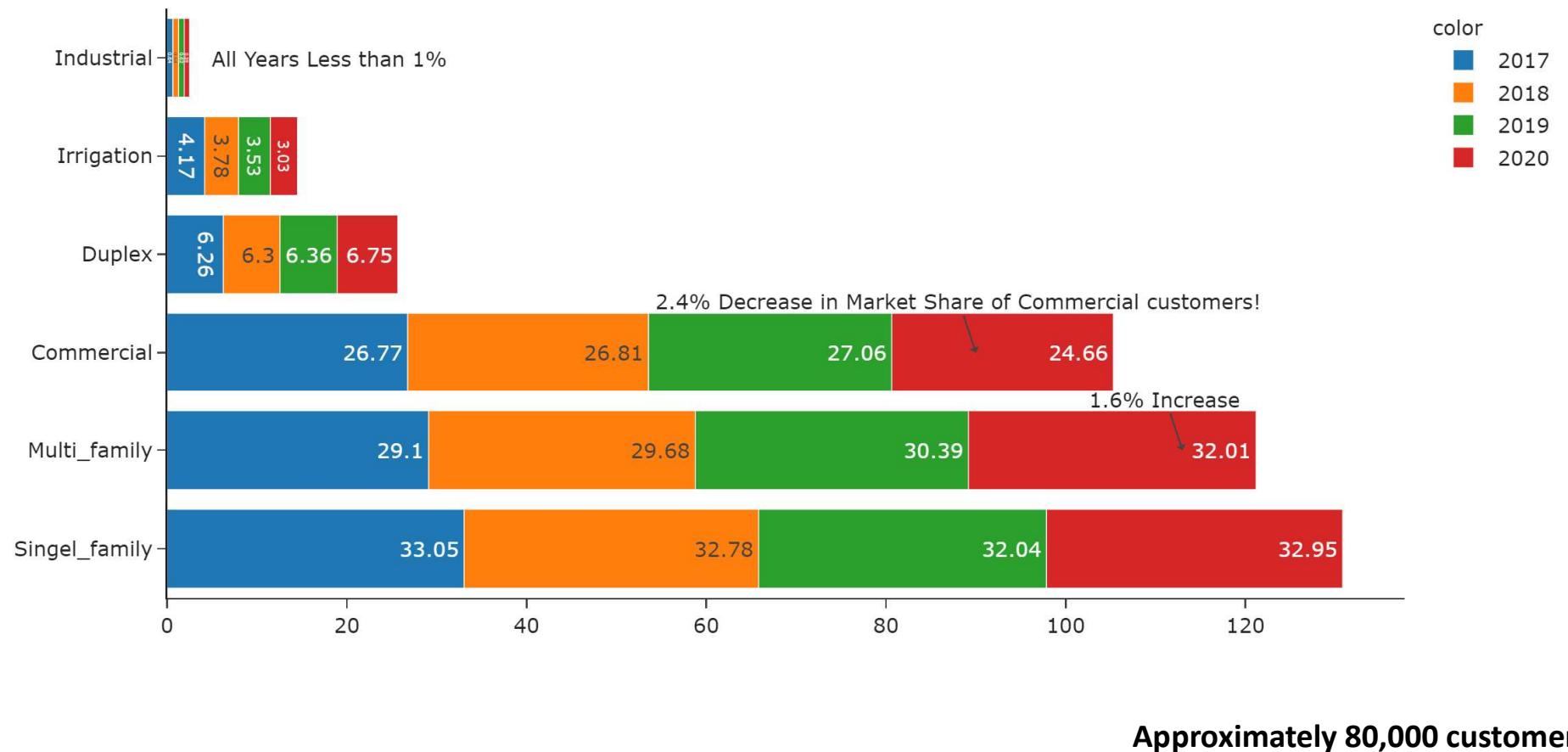
## Commercial



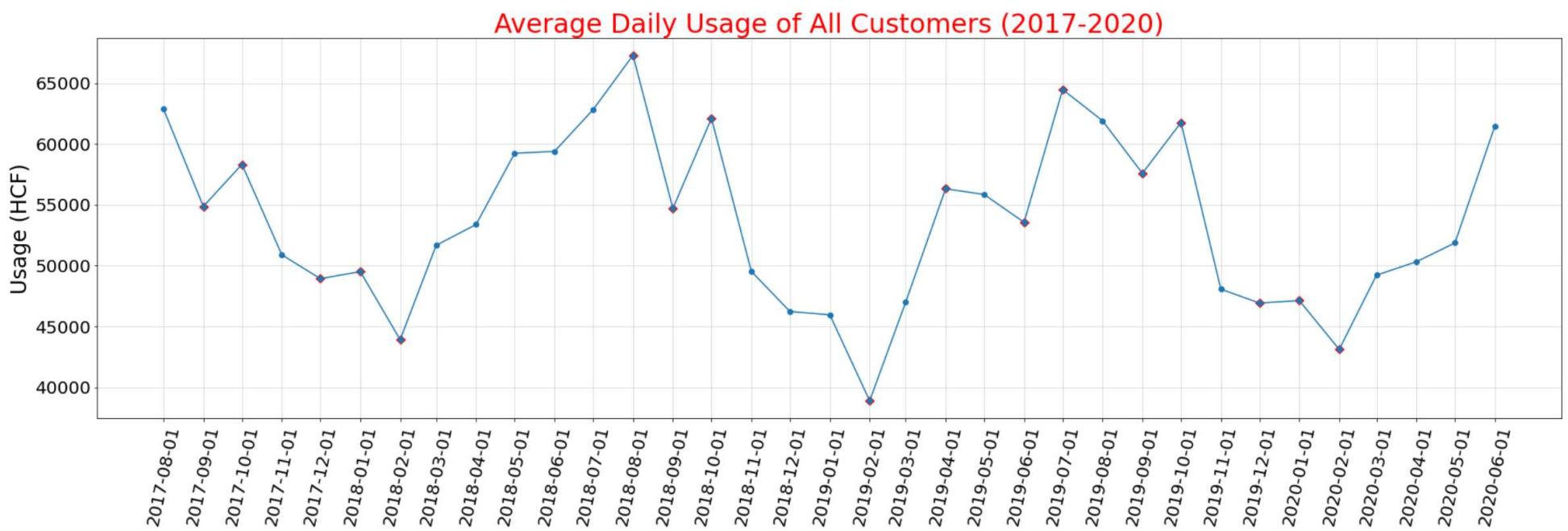
## Irrigation



## Market Share of Water Usage

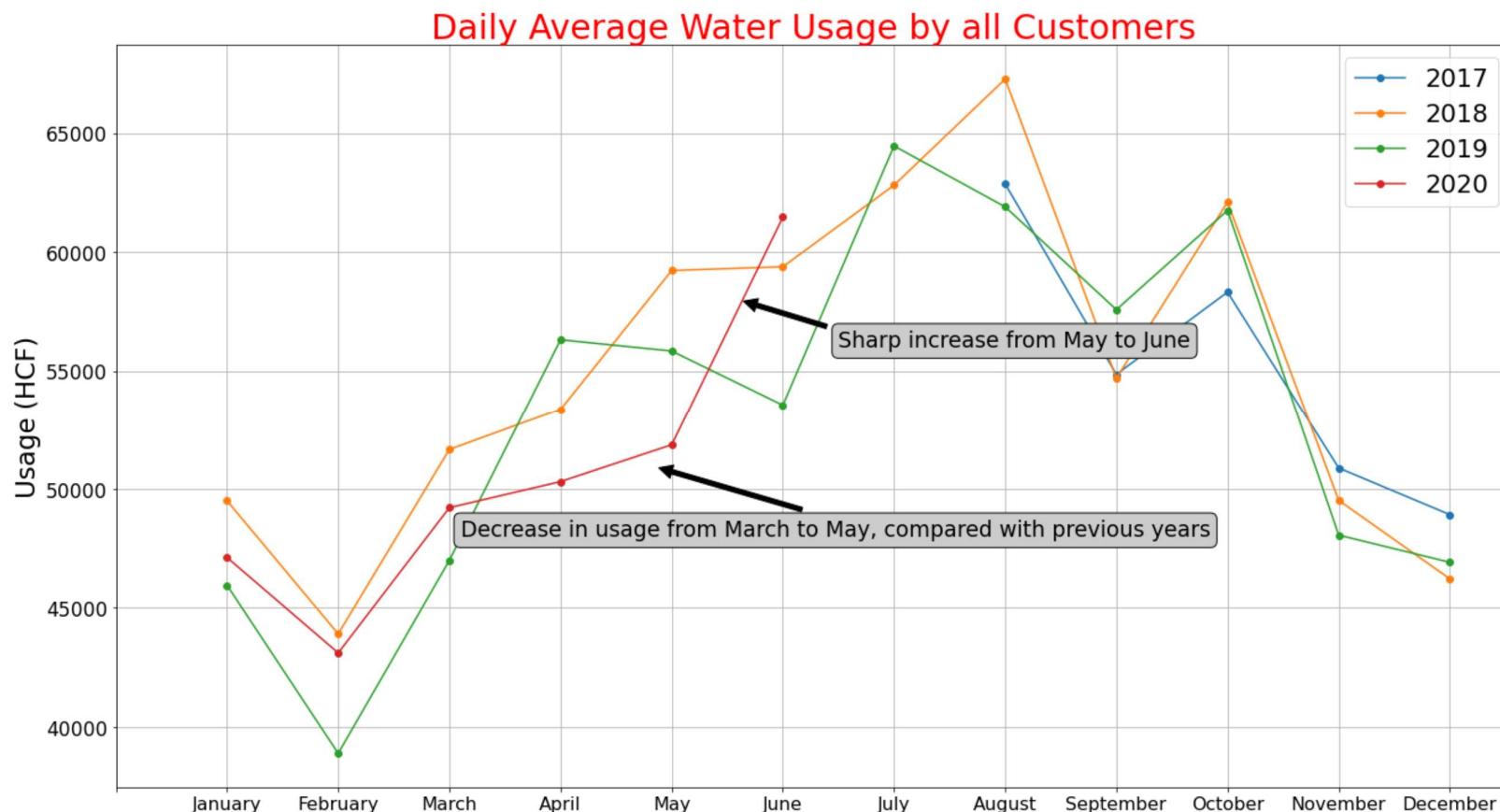


# Change in Trend and Seasonality from March 2020

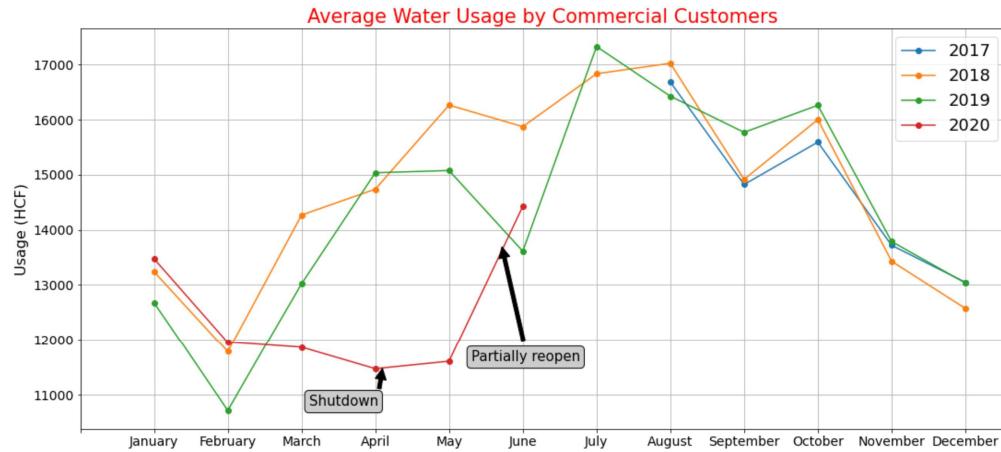


- Feb 2020 was 29 days
- Decrease in usage from March 2020 and continue until June compared with months in previous years

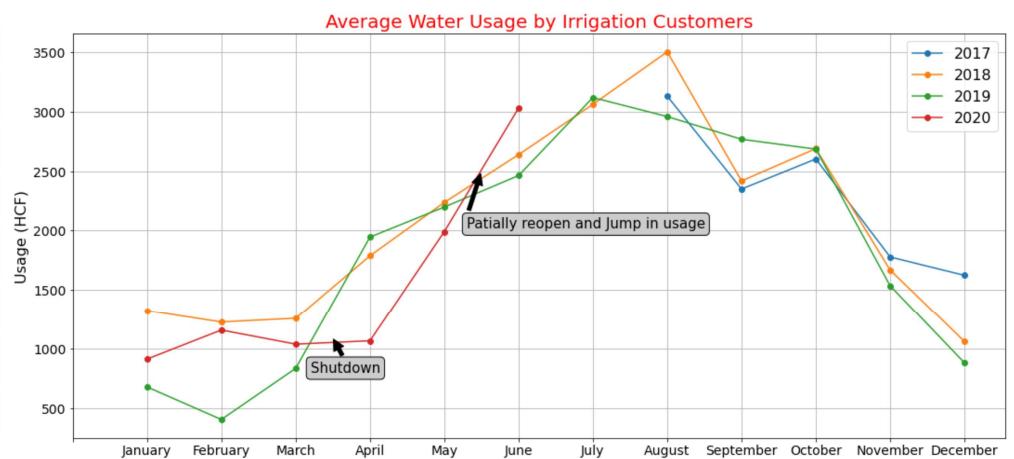
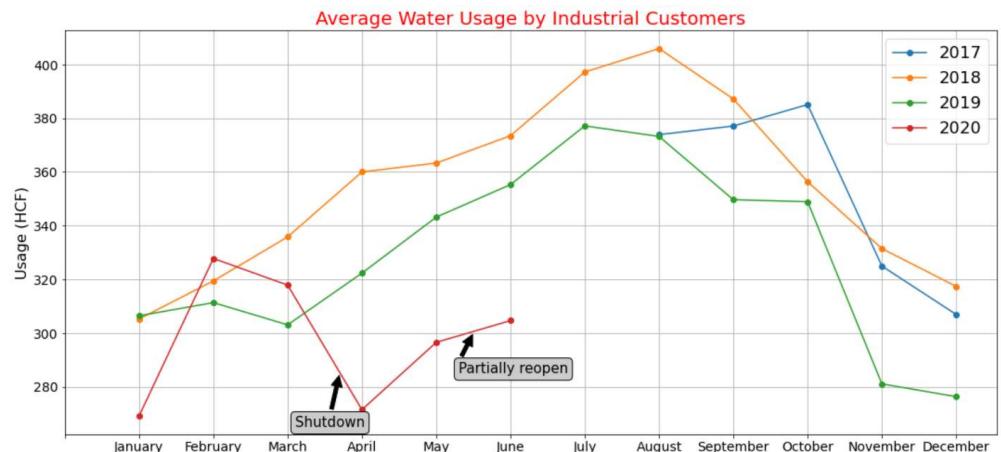
# What Segments had more role in a different trend?



# Shutdown and Decrease in non-Residential Segment in 2020

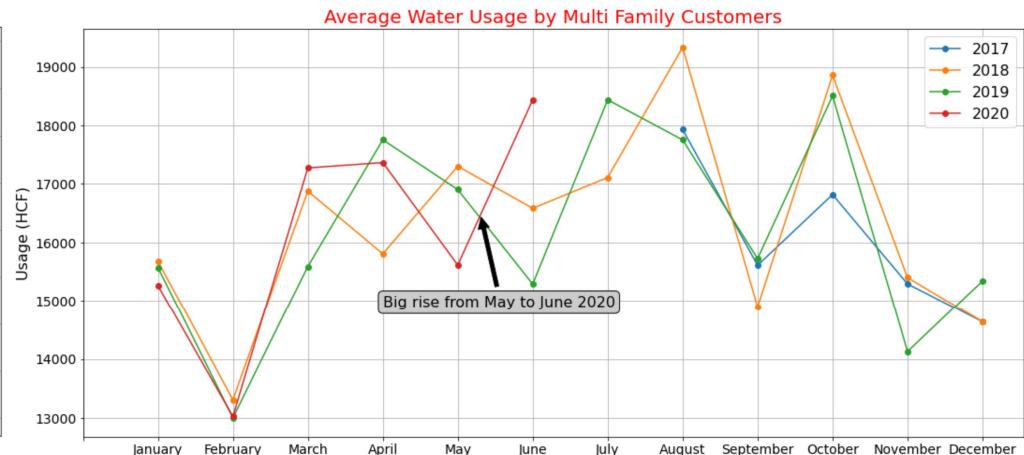
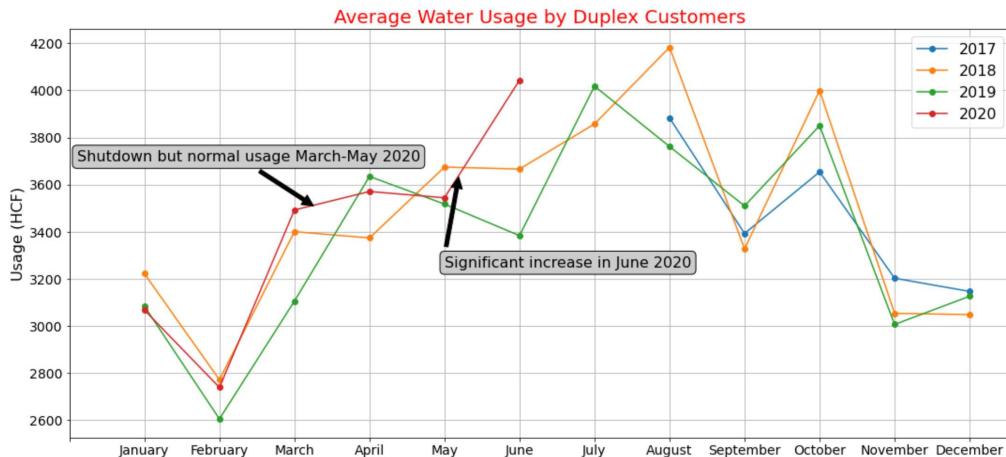
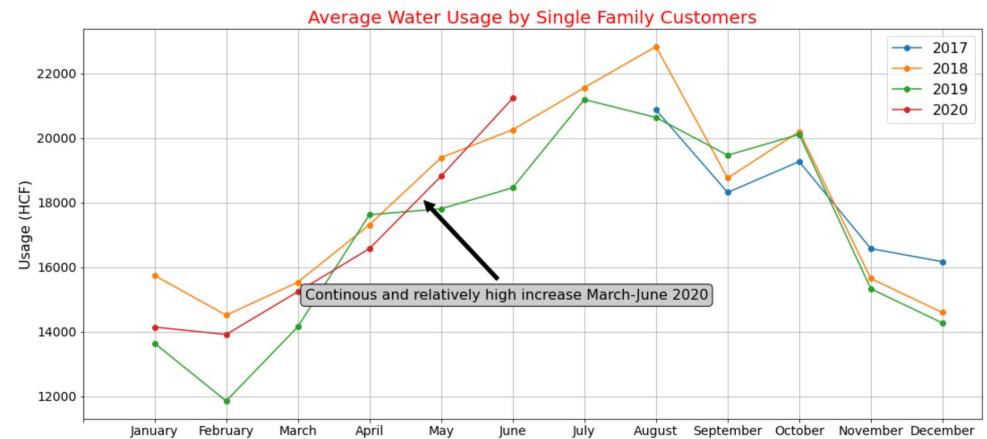


- Shutdown and sharp drop in April and May
- Reopen and usage jump in June

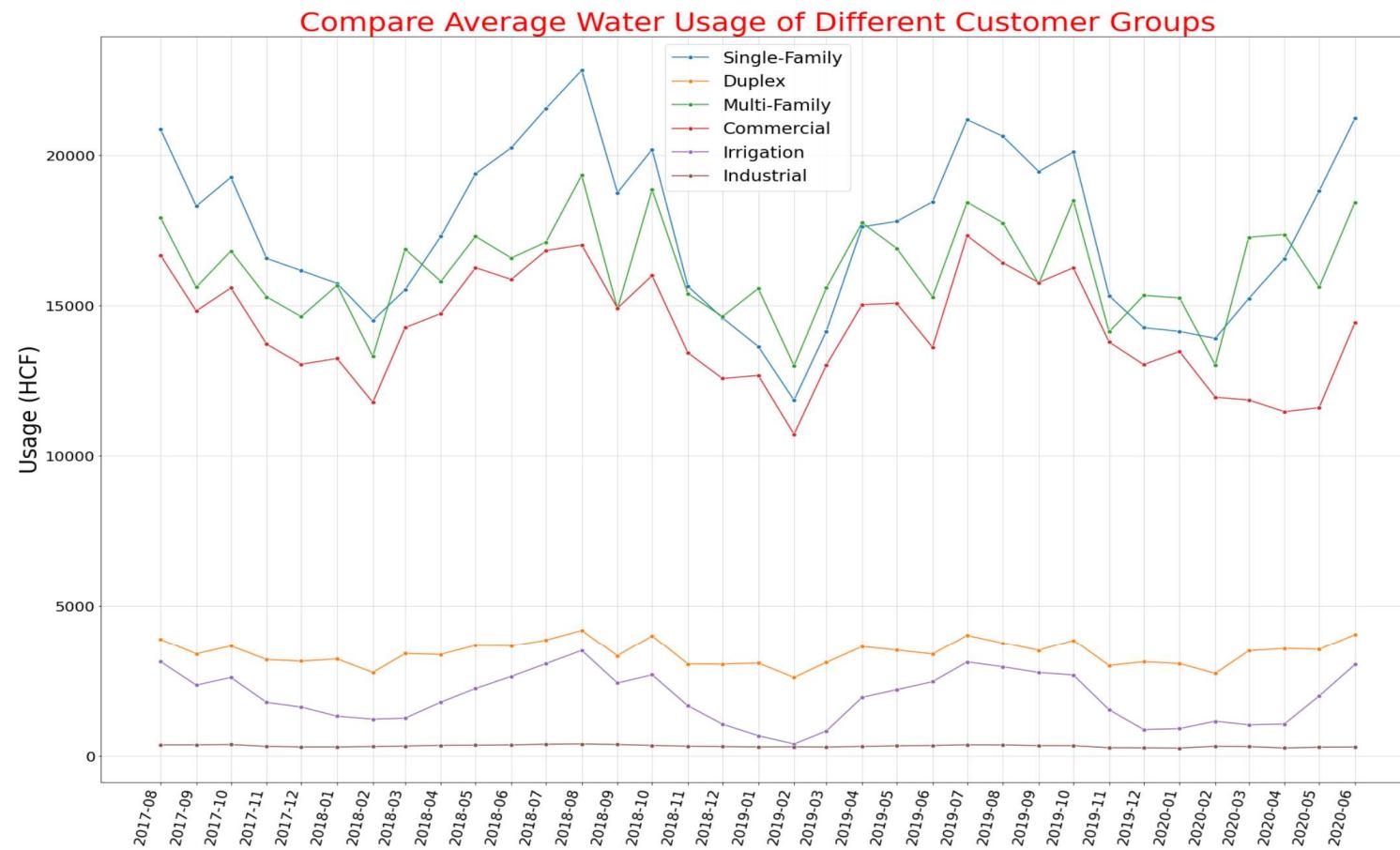


# No Significant Decrease in Residential Usage in 2020

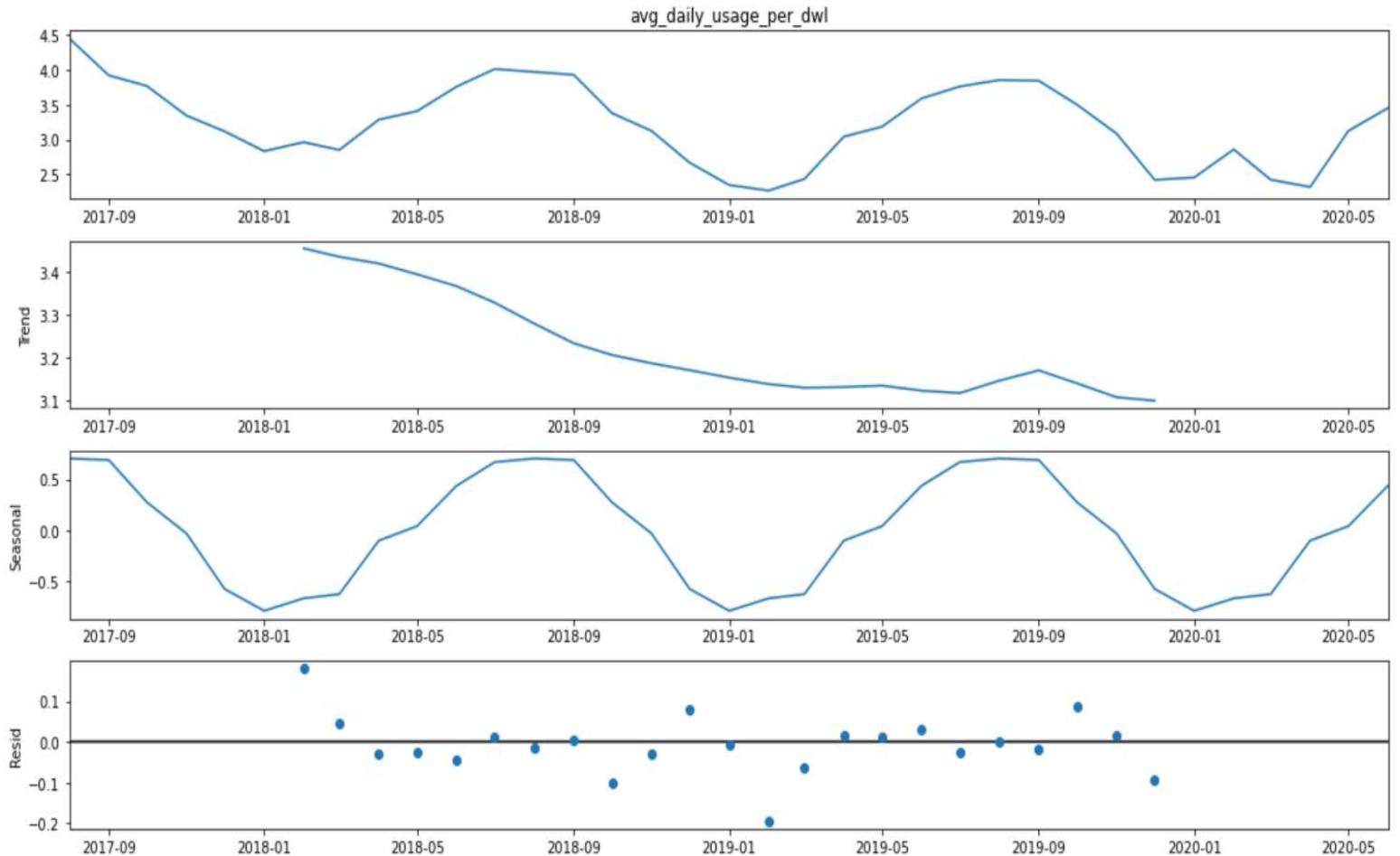
- Single family kept its trend**
- Multi family and duplex:**
  - Normal increase from March to May**
  - Decrease in May**
  - Grow in June**



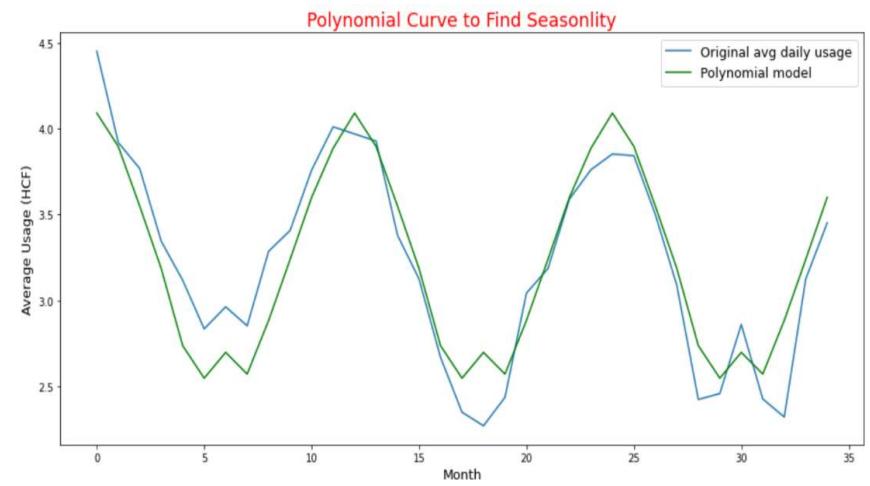
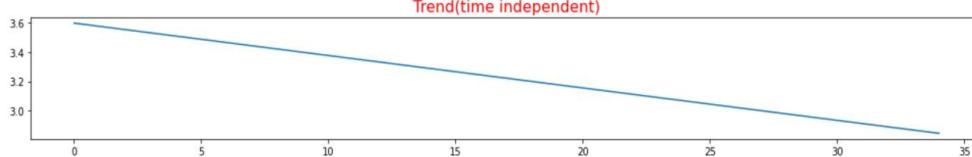
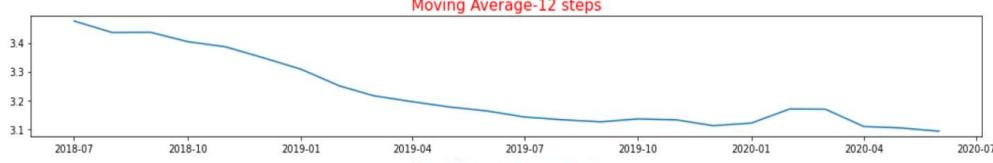
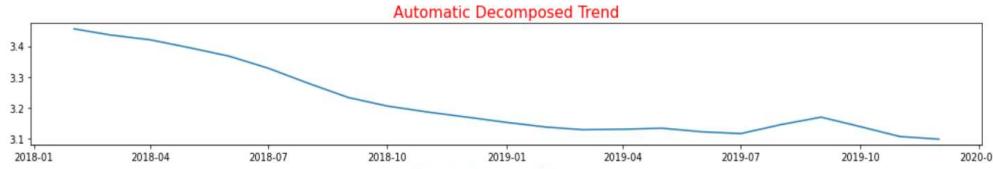
**Single-Family and Multi-Family usage tried to offset the decrease in Commercial usage from March 2020**



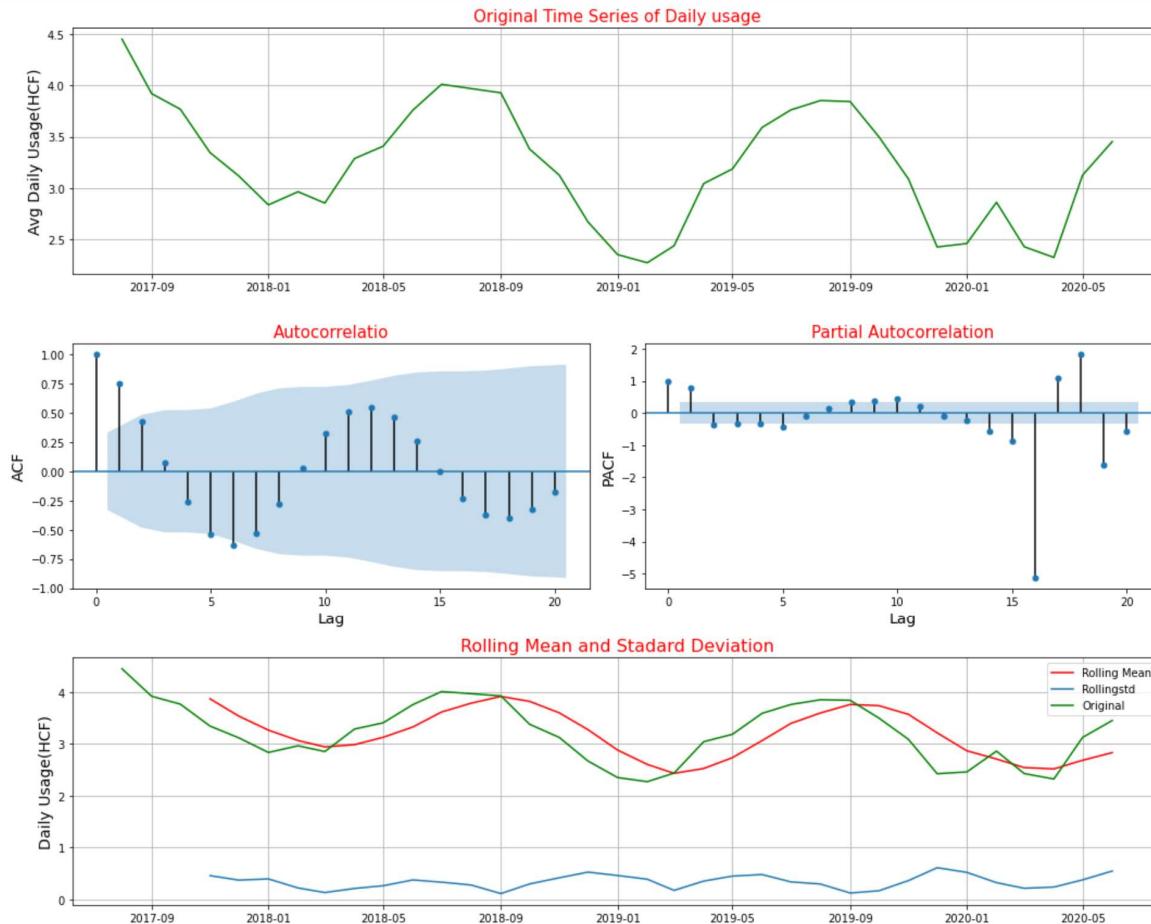
## Non-Stationary Average Usage Time Series



# Automatic Decomposed Trend vs Time Independent Trend Time Independent Seasonality



# Non-Stationary Average Usage Time Series



Original values

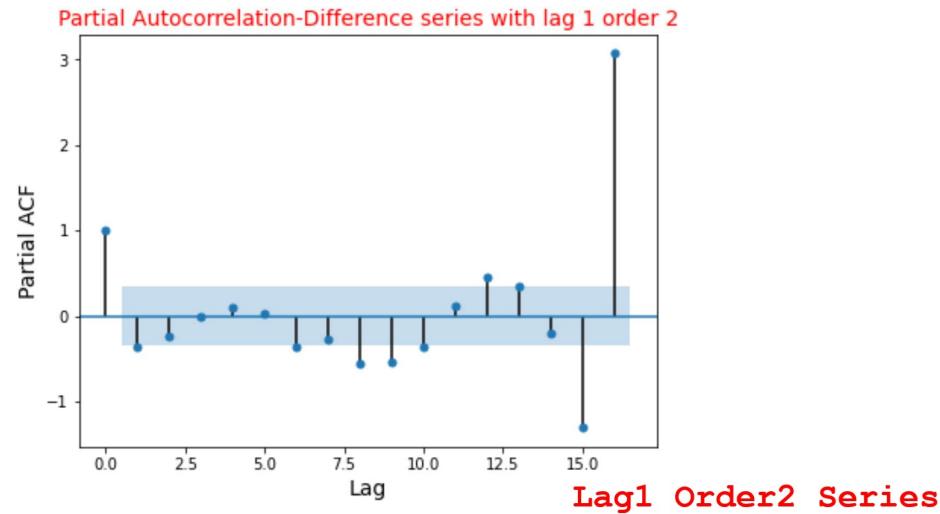
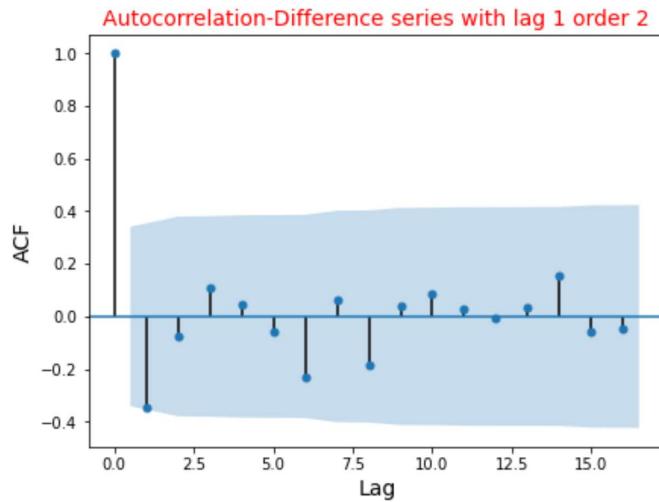
Results of Dickey-Fuller Test:

<b>test statistic</b>	<b>-1.592271</b>
<b>p-value</b>	<b>0.487475</b>
# Lags Used	8.000000
# Observation Used	26.000000
<b>Critical Value 1%</b>	<b>-3.711212</b>
<b>Critical Value 5%</b>	<b>-2.981247</b>
<b>Critical Value 10%</b>	<b>-2.630095</b>

# Make Time Series Stationary using Difference Transforming



# Stationary Differenced Series (lag 1 order 2)



Lag1 Order2 Series

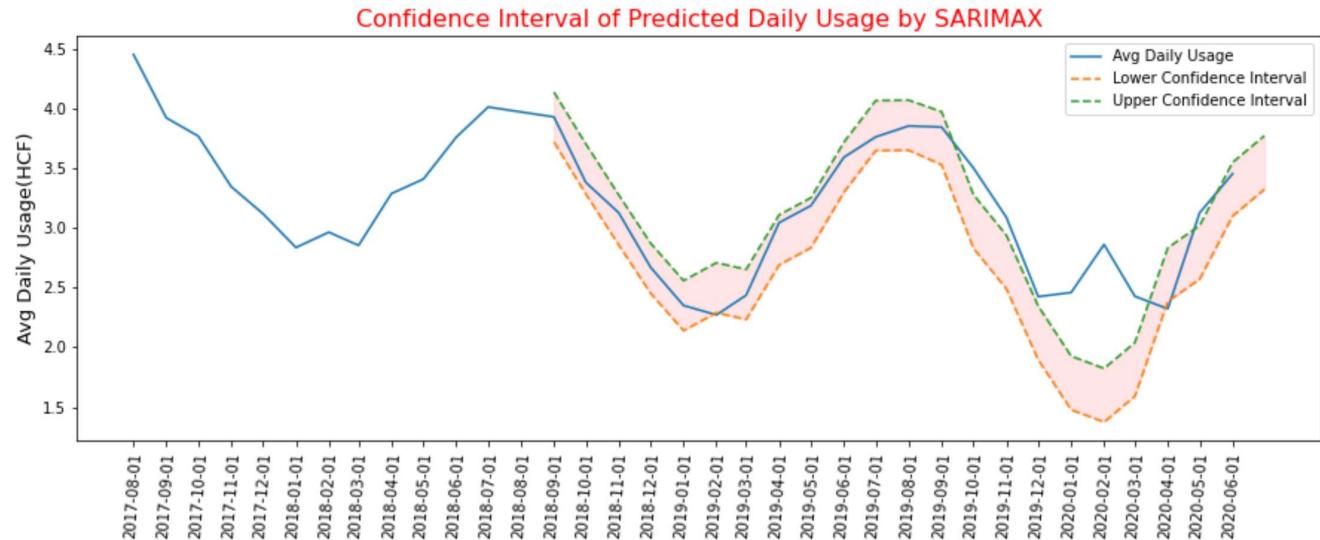
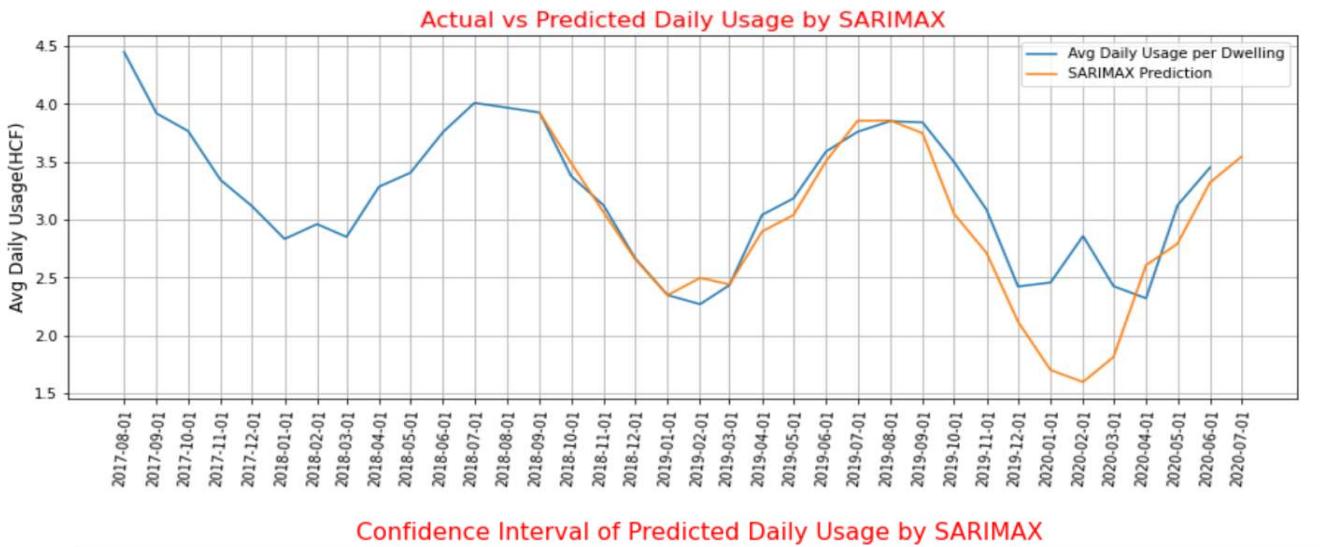
Results of Dickey-Fuller Test:

test statistic	<b>-7.337845e+00</b>
p-value	<b>1.084672e-10</b>
# Lags Used	<b>8.000000e+00</b>
# Observation Used	<b>2.400000e+01</b>
Critical Value 1%	<b>-3.737709e+00</b>
Critical Value 5%	<b>-2.992216e+00</b>
Critical Value 10%	<b>-2.635747e+00</b>

# Prediction of SARIMAX

Params Grid for SARIMAX

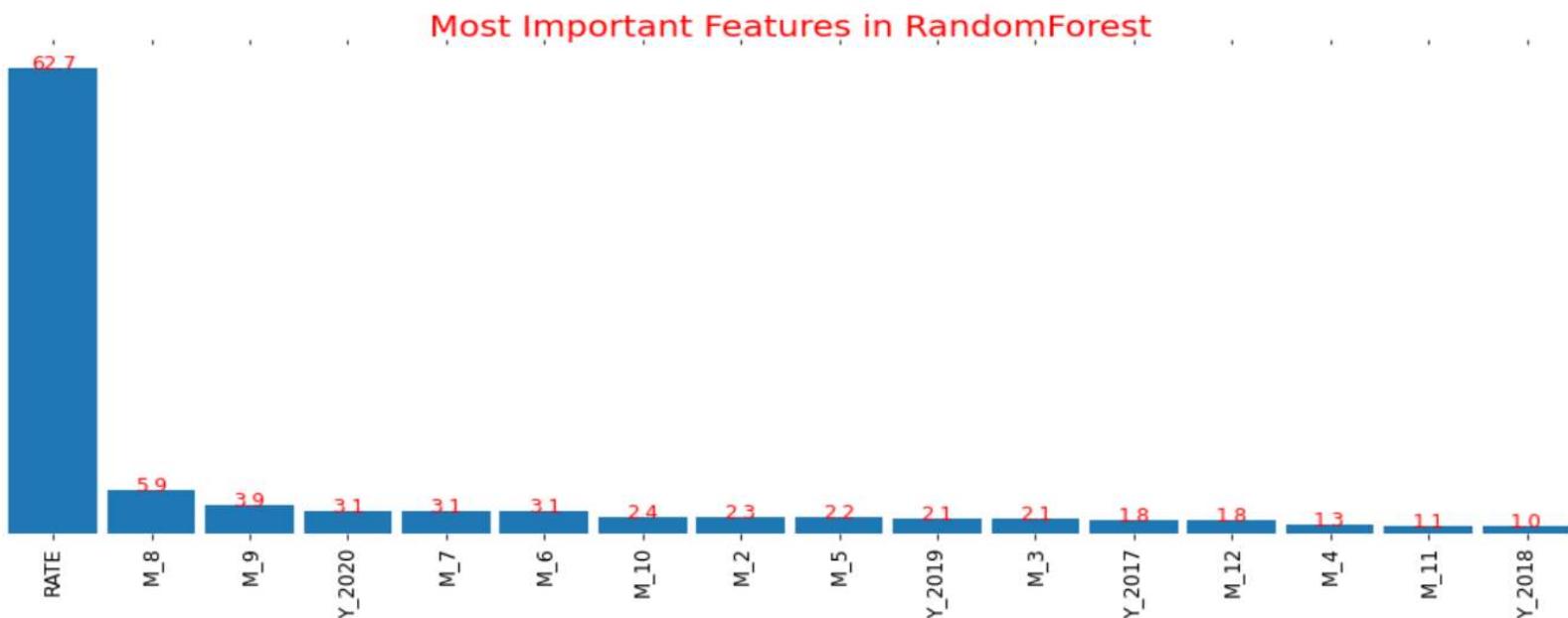
trend	param	seasonal_param	AIC
0	n	(1, 1, 2)	(1, 0, 0, 12) -7.571470
1	t	(1, 1, 2)	(1, 0, 0, 12) -5.628948
2	c	(1, 0, 0)	(1, 0, 0, 12) -9.992228
3	ct	(1, 0, 0)	(1, 0, 0, 12) -8.676631



# Performance of Univariate Time Forecasting Models

	Model	RMSE	MAE	MAPE	R2
0	SimpleExpSmoothing	0.409886	0.329360	11.462578	0.738515
1	ExponentialSmoothing	0.409886	0.329360	11.462578	0.738515
2	Autoregressive	0.313047	0.266939	9.541546	0.830775
3	ARMA(0,2)	0.178451	0.139238	5.016632	0.950049
4	ARMA(1,2)	0.178451	0.139238	5.016632	0.950049
5	SARIMA (1, 0, 0)x(1, 0, 0, 12)	0.537952	0.417924	15.147803	0.862873

# Importance Features



- 1) Customer Group
- 2) Month of Aguest
- 3) Month of September and year 2020
- 4) Month of June and July

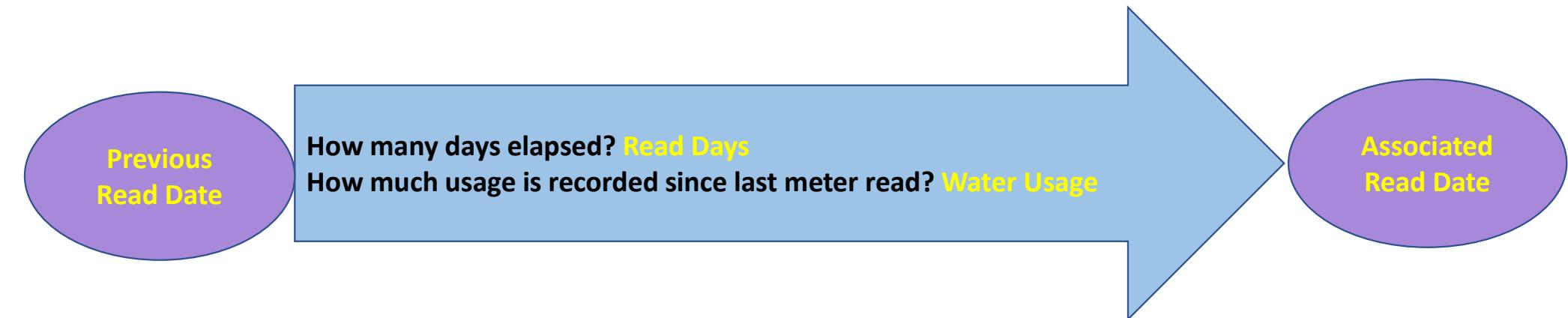
# Compare Performance of Models Tested

Model	RMSE	MAE	MAPE	R2
<b>SimpleExpSmoothing</b>	0.409886	0.329360	11.462578	0.738515
<b>ExponentialSmoothing</b>	0.409886	0.329360	11.462578	0.738515
<b>Autoregressive</b>	0.313047	0.266939	9.541546	0.830775
<b>ARMA(0,2)</b>	0.178451	0.139238	5.016632	0.950049
<b>ARMA(1,2)</b>	0.178451	0.139238	5.016632	0.950049
<b>SARIMA (1, 0, 0)x(1, 0, 0, 12)</b>	0.537952	0.417924	15.147803	0.862873
<b>BayesianRidge</b>	0.274558	0.225684	200.045553	-0.273716
<b>Lasso</b>	0.262176	0.214882	196.283369	-0.434866
<b>ElasticNet</b>	0.263902	0.216201	197.374565	-0.422470
<b>RandomizedSearchCV RandomForest Regressor</b>	0.034560	0.028109	21.610100	0.194310
<b>XGB Regressor</b>	0.034560	0.028109	21.610100	0.194310
<b>Bayesian Process-RandomForest Regressor</b>	0.034721	0.028313	21.765036	0.189995

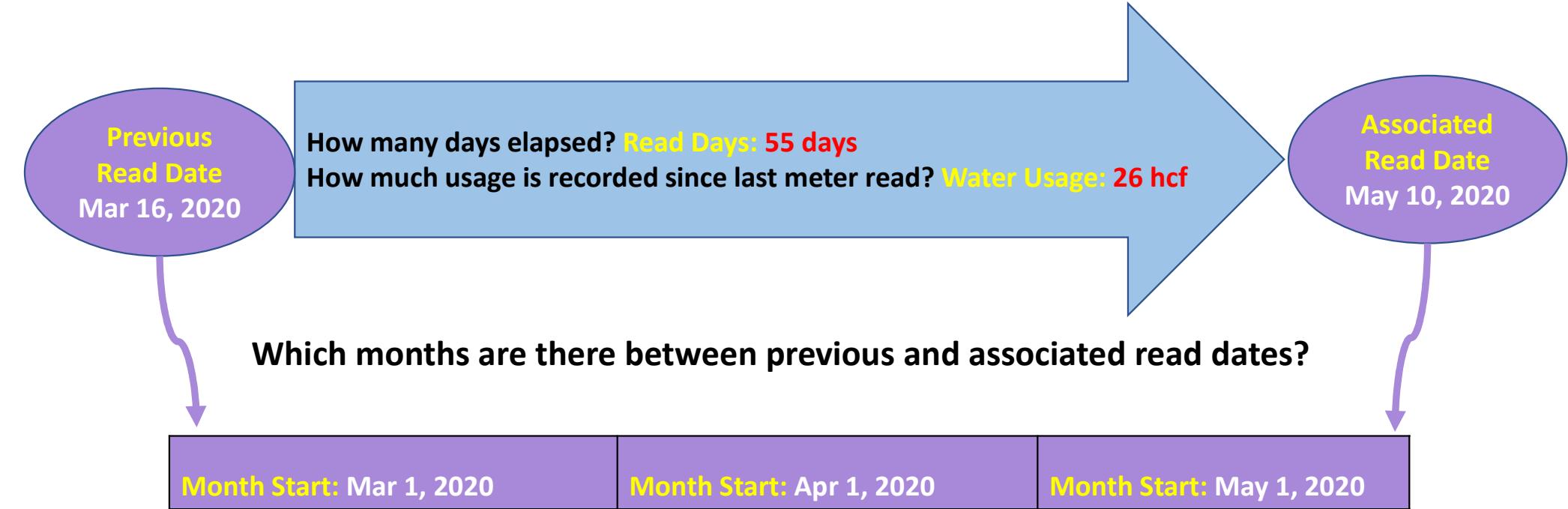
# Examples of the algorithm application

# How Much Water Used in each Month?

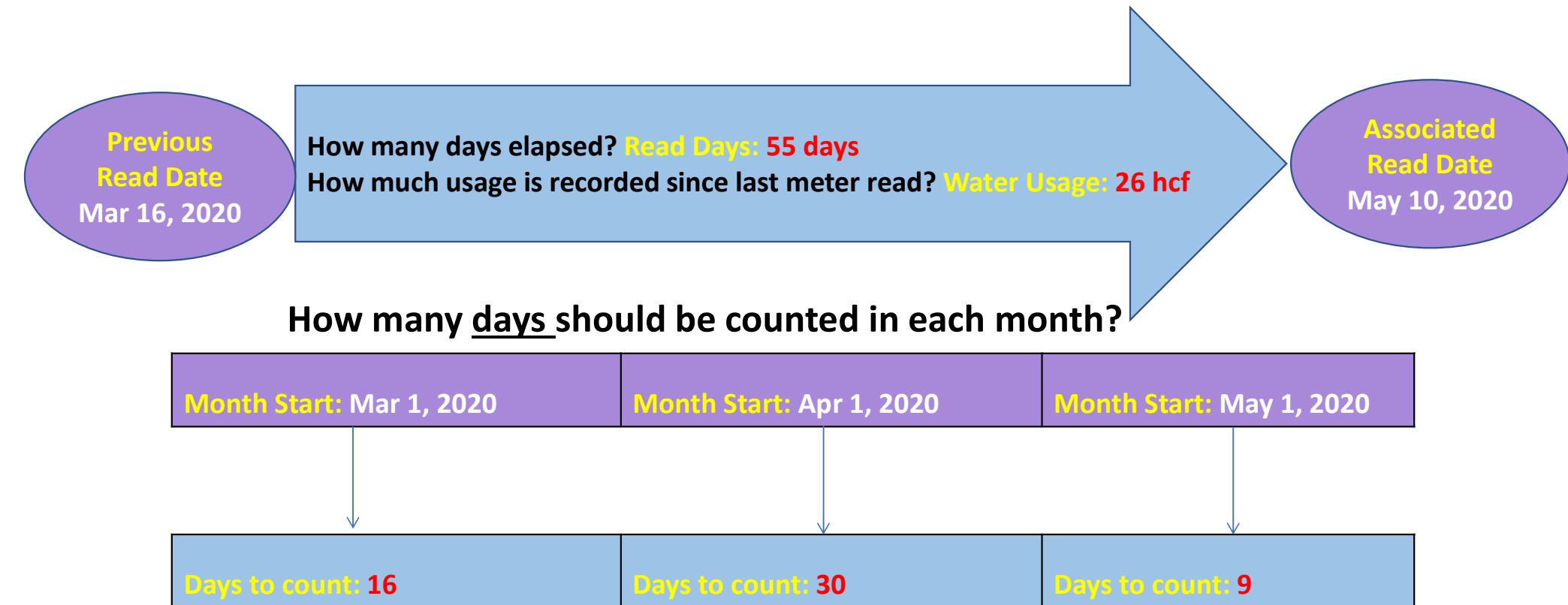
(based on the Records of Water Meter )



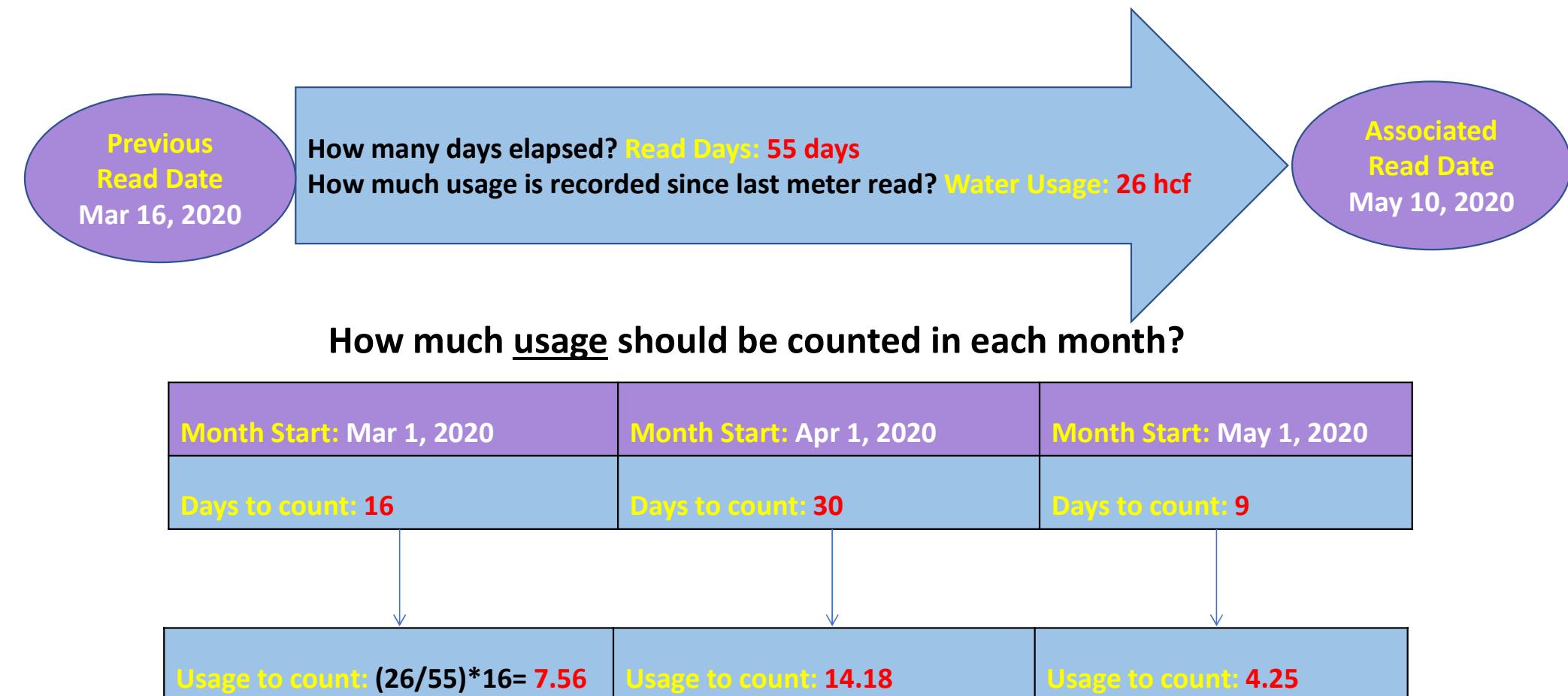
# How Much Water Used in each Month?



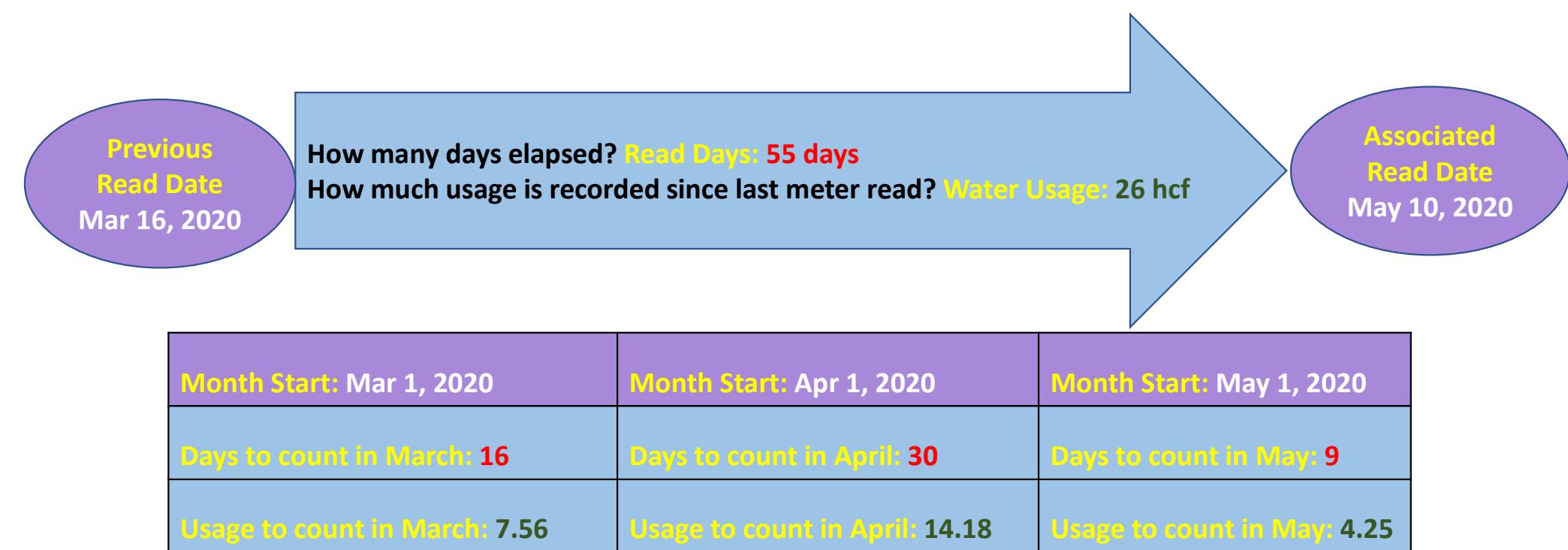
# How Much Water Used in each Month?



# How Much Water Used in each Month?



# How Much Water Used in each Month?



Algorithm

Days in March = 31			
Month Start <b>Mar 1, 2020</b>	Previous Read Date <u>March 16, 2020</u>	Month End	Read Date <b>May 10, 2020</b>
previous_read_date_to_month_end <b>16 days</b>			
read_date_to_month_start <b>70</b>			
previous_read_date_to_month_end <= days in month (16 <= 31)			

Month Start	days_to_count_in_period
<b>2020-03-01</b>	16

Days in April = 30			
Month Start <b>April 1, 2020</b>	Previous Read Date <u>March 16, 2020</u>	Month End	Read Date <b>May 10, 2020</b>
previous_read_date_to_month_end <b>30-(-16) = 46 days</b>			
read_date_to_month_start <b>39</b>			
previous_read_date_to_month_end > days in month (46 > 30) & read_date_to_month_start >= days in month (39 >= 30)			

Month Start	days_to_count_in_period
<b>2020-04-01</b>	30

Days in May = 31			
Month Start <b>May 1, 2020</b>	Previous Read Date <u>March 16, 2020</u>	Month End	Read Date <b>May 10, 2020</b>
previous_read_date_to_month_end <b>31-(-46) = 77 days</b>			
Read_date_to_month_start <b>9 days</b>			
previous_read_date_to_month_end < days in month (77 < 31) & read_date_to_month_start < days in month (9 < 30)			

Month Start	days_to_count_in_period
<b>2020-05-01</b>	9

Days in March = 31			
Month Start <b>Mar 1, 2020</b>	Previous Read Date <u>March 16, 2020</u>	Month End	Read Date <b>May 1, 2020</b>
previous_read_date_to_month_end <b>16 days</b>			
read_date_to_month_start <b>56</b>			
previous_read_date_to_month_end <= days in month then (16 <= 56)			

Month Start	days_to_count_in_period
<b>2020-03-01</b>	16

Days in April = 30			
Month Start <b>April 1, 2020</b>	Previous Read Date <u>March 16, 2020</u>	Month End	Read Date <b>May 1, 2020</b>
previous_read_date_to_month_end <b>30 - (-16) = 46 days</b>			
read_date_to_month_start 30			
previous_read_date_to_month_end > days in month (46 > 30) & read_date_to_month_start >= days in month (30 >= 30)			

Month Start	days_to_count_in_period
<b>2020-04-01</b>	30

Days in May = 31			
Month Start <b>May 1, 2020</b>	Previous Read Date <u>March 16, 2020</u>	Month End	Read Date <b>May 1, 2020</b>
previous_read_date_to_month_end <b>31 - (-46) = 77 days</b>			
Read_date_to_month_start <b>0 days</b>			
previous_read_date_to_month_end < days in month (77 > 31) & read_date_to_month_start < days in month (0 < 30)			

Month Start	days_to_count_in_period
<b>2020-05-01</b>	0