Gulf Coast Area Rainfall Prediction

By Jimmy Smart

Background

Our livelihood is so greatly affected by the weather. Depending on the region someone lives in usually determines the type of weather patterns that they'll see on a yearly, monthly or even daily basis. A large majority of industries greatly depend on their region's weather and climate. For instance, farmers need a good mixture of rain and sunlight so that their crops may grow. Too much sun but not enough rain, may cause issues to their crop yields. Just as too much rain and not enough sun can affect their crop yields as well. Seasonal temperatures factor in as well, as a majority of crops grow based on the daily temperatures. Some may grow better in months that are typically known to have cooler temperatures, while other crops prefer to grow in the warmer months.

Commercial fishing also keeps a keen eye on weather patterns as life found in the Gulf usually have their own growing seasons as well as seasons for farming.

For farmers, commercial fishing and various other industries who depend on their regional climates, their best approach in co-existing with weather patterns is by using past trends to create future predictions.

The Problem

History is known to repeat itself at times but predicting if/when history will repeat itself is an ongoing issue. The best approach is look for patterns of the past while comparing it to patterns of the present. The goal of this project is: Can we predict rainfall for the Gulf Coast area?

Weather has shown signs of yearly and seasonal patterns but it's not always set in stone. Just like other factors in our world, weather patterns are changing as well. Global warming has become a hot debate in our culture today as more and more people feel that the certain factors in our environments are raising temperatures around the world which in turn are also raising the temperatures of our oceans, seas and gulfs. This project isn't necessarily a global warming, project, but as we go through our historical weather data and try to predict future weather patterns, we will be able to see if there has been a general uptick in temps and see if there is a correlation with any effects it's having on our data over the years.

Data

This project will build a model that focuses on a dataset consisting of decades of daily weather from various locations within our Gulf Coast area. Where we're looking at consists of an area in the southern part of the United States that shares a coastline on the Gulf of Mexico. Our data is from https://ww.ncdc.noaa.gov/ and consists of 5 different datasets from each of our 5 chosen cities along the Gulf Coast. Each dataset consists of daily weather data and is saved as a CSV file.

Cities used for our Dataset

Predictor	Description
New_Orleans.csv	dataset of daily New Orleans, LA weather 1/1/2010-12/31/2019
Houston.csv	dataset of daily Houston, TX weather 1/1/2010-12/31/2019
Pascagoula.csv	dataset of daily Pascagoula, MS weather 1/1/2010-12/31/2019
Mobile.csv	dataset of daily Mobile, AL weather 1/1/2010-12/31/2019
Tampa.csv	dataset of daily Tampa, FL weather 1/1/2010-12/31/2019

Terms and definitions

Dewpoint Average Temperature- temperature where water vapor starts to condense out of the air (the temperature at which air becomes completely saturated). Above this temperature the moisture stays in the air. If the dew-point temperature is close to the dry air temperature - the relative humidity is high If the dew point is well below the dry air temperature - the relative humidity is low

Humidity(%)- concentration of water vapor present in the air. Water vapor, the gaseous state of water. Humidity indicates the likelihood for precipitation, dew, or fog to be present. The amount of water vapor needed to achieve saturation increases as the temperature increases.

Sealevel_pressure(Hg)- atmospheric pressure at sea level at a given location. When observed at a reporting station that is not at sea level (nearly all stations), it is a correction of the station pressure to sea level. This correction takes into account the standard variation of pressure with height and the influence of temperature variations with height on the pressure. The temperature used in the sea level correction is a twelve hour mean, eliminating diurnal effects.

Machine Learning Model Building

In our previous report we cleaned and conducted EDA on our dataset to better prepare our data for our ML model building.

For our modeling we ran models on 2 datasets. 1 dataset that took out our Rainfall outliers and 1 dataset that left the Rainfall outliers within our feature.

without Rainfall outliers
`#remove outliers above 5.5 inches of rain` `df= df[df['Rainfall(in)'] < 5.5]`

FB Prophet - Model 1

Facebook's open source prediction time series, Prophet, is a good model to start with. It provides hyperparameters that we'll tune.

We started off our model by creating a dataframe from our main dataset consisting of ds and y features. Ds represents our datetime stamp and y represents the amount of rainfall on the respective days:

	ds	y
0	2010-01-01	0.0
1	2010-01-02	0.0

We created our model before fitting it and setting a prediction. To visualize our new model we called a basic prediction visualization.

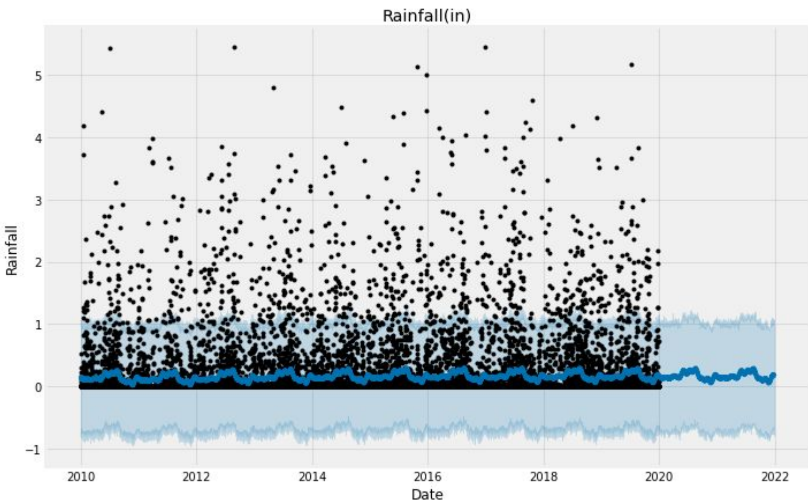


Figure shows results from dataset without outliers

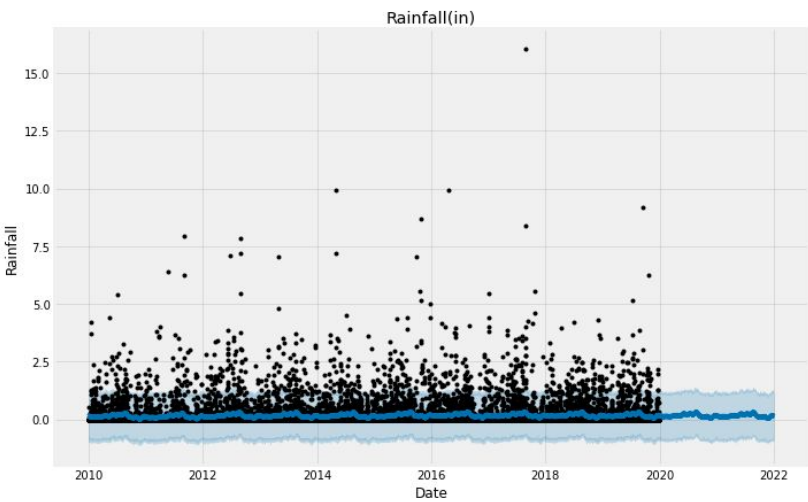


Figure shows results from dataset with outliers

We also ran a plot_compoentents visualization to better see our trends.

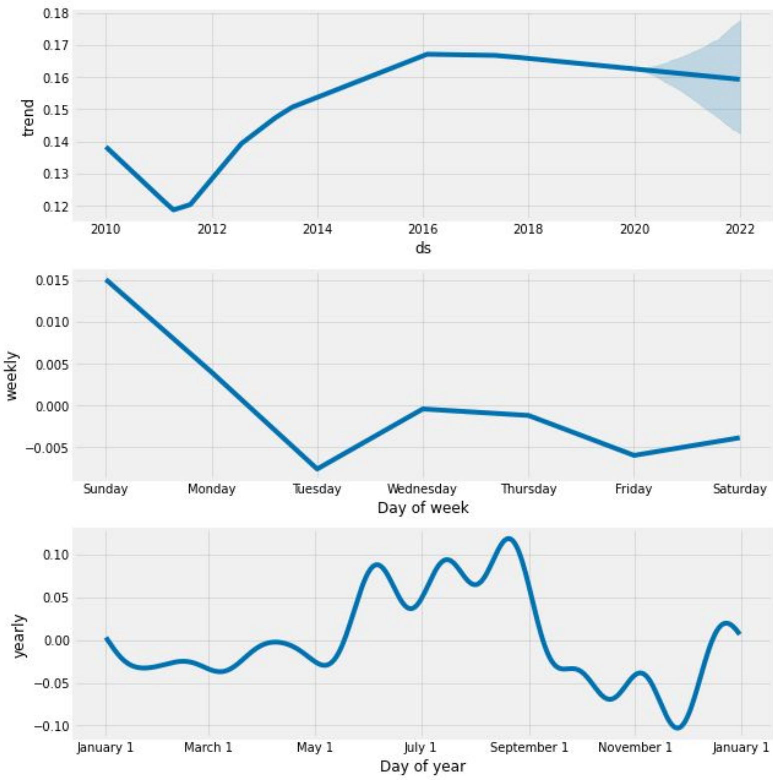


Figure shows results from model without outliers

The first plot shows that the monthly volume of rainfall dropped in 2011 but steadily increased until 2016 when it started to decrease slightly each year. This graph shows that there looks to be a downward trend for the next following years.

The second plot shows a mostly decrease in weekly trends. Day of the week isn't commonly a factor in weather though.

Our final plot is interesting because traditionally weather is affected by seasonal habits. As we can see in our plot, the warmer months of May-September were higher than the rest of the year but there were dips within those months.

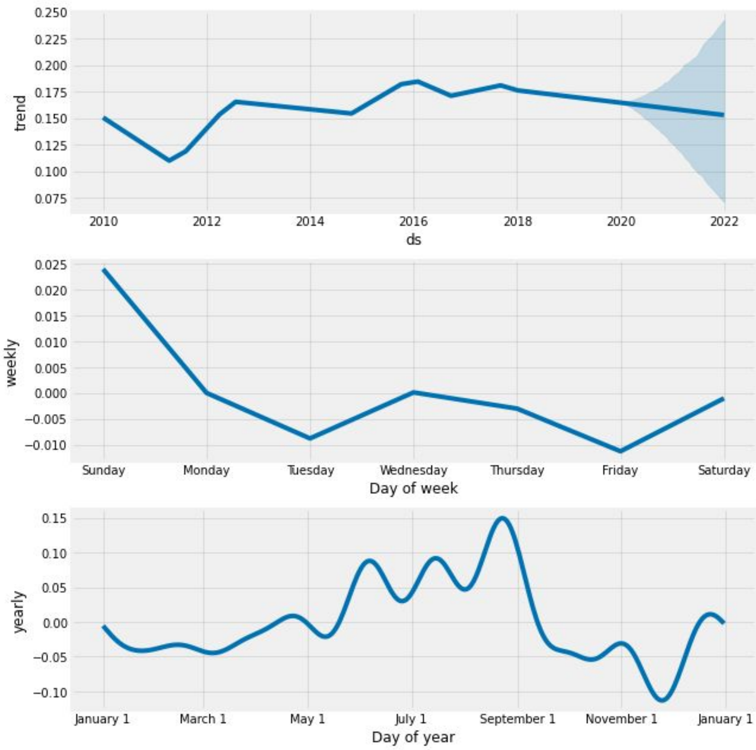


Figure shows results from model with outliers

The first plot shows that the monthly volume of rainfall dropped in 2011 but increased the following year. It stayed somewhat steady for the next 3 years when rainfall had an increase towards 2016. Since 2016 there is a slight decrease each year.

The second plot shows a mostly decrease in weekly trends. Day of the week isn't commonly a factor in weather though.

Our final plot is interesting because traditionally weather is affected by seasonal habits. As we can see in our plot, the warmer months of May-September were higher than the rest of the year but there were dips within those months. Late August has our highest peak which makes sense because typically hurricanes/tropical storms get more active in the mid summer month.

Changepoints

We added Changepoints into our model. Changepoints show the abrupt changes in trajectory. Below we can visualize these points along with some of the dates that the changes occurred.

Model without outliers	Model with outliers
583 2010-04-27	584 2010-04-27

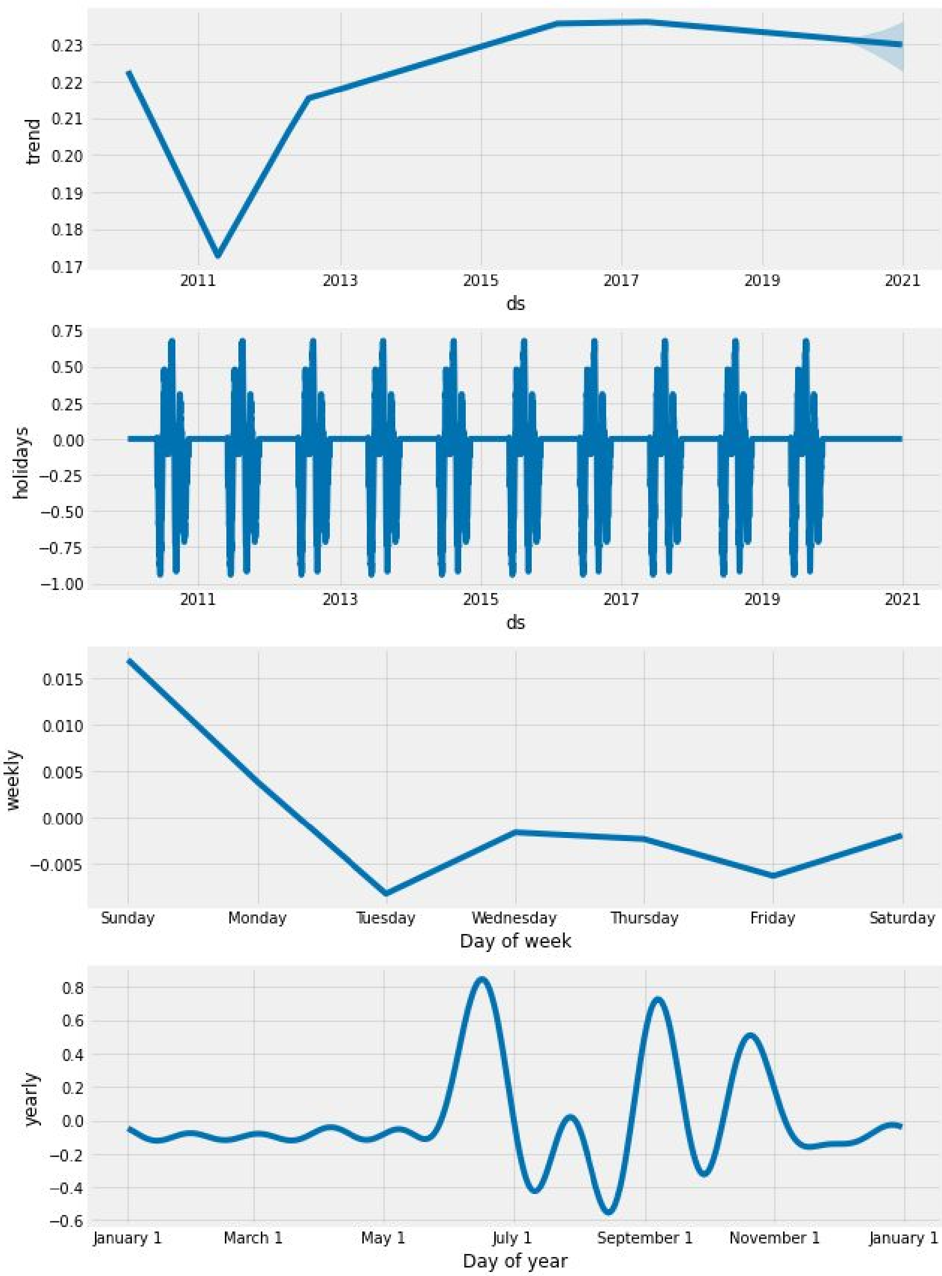
1166	2010-08-22	1167	2010-08-22
1749	2010-12-16	1751	2010-12-17
2333	2011-04-12	2335	2011-04-13
2916	2011-08-07	2919	2011-08-07
3499	2011-12-02	3502	2011-12-02
4082	2012-03-29	4086	2012-03-29
4665	2012-07-24	4670	2012-07-24
5248	2012-11-18	5253	2012-11-17
5832	2013-03-14	5837	2013-03-14
6415	2013-07-09	6421	2013-07-09
6998	2013-11-03	7005	2013-11-03
7581	2014-03-01	7588	2014-03-01
8164	2014-06-26	8172	2014-06-25
8747	2014-10-20	8756	2014-10-20
9331	2015-02-15	9340	2015-02-15
9914	2015-06-12	9923	2015-06-12
10497	2015-10-06	10507	2015-10-06
11080	2016-01-31	11091	2016-01-31
11663	2016-05-27	11674	2016-05-27
12246	2016-09-21	12258	2016-09-21
12830	2017-01-16	12842	2017-01-15
13413	2017-05-12	13426	2017-05-12
13996	2017-09-06	14009	2017-09-06
14579	2018-01-01	14593	2018-01-01

Hurricane Season

The Gulf Coast area has to prepare for yearly hurricane season which causes parts of our area to be affected by hurricanes and tropical storms. The outliers that we took out of one of our datasets mostly consisted of days that had large amounts of rainfall due to one of these storms hitting the area.

We tuned our model(s) to include hurricane season which is set between June 1st through November 30th.

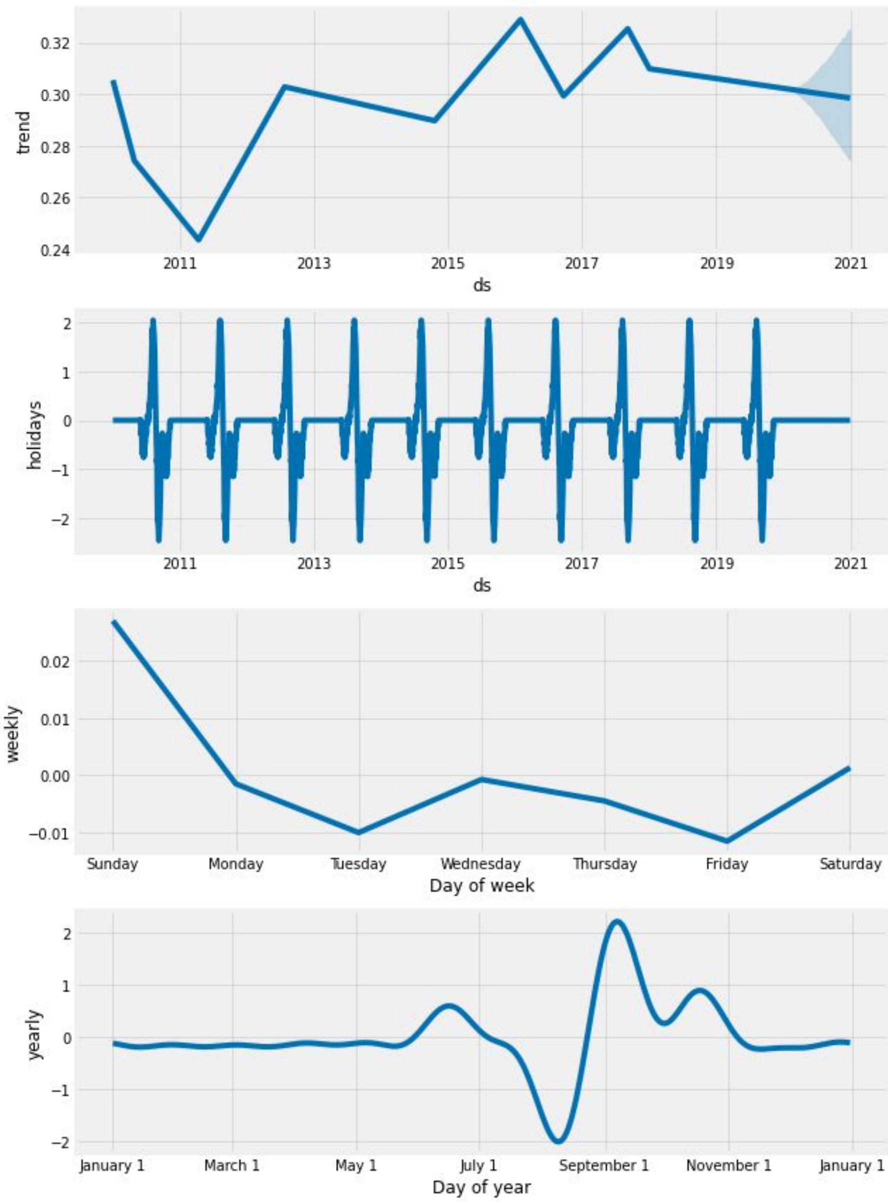
```
#Hurricane season (06-01 to 11-30)
hurricane_season = pd.DataFrame({'holiday': "hurricane season", 'ds' :
pd.to_datetime(['2010-06-01','2011-06-01','2012-06-01','2013-06-01',
                '2014-06-01','2015-06-01','2016-06-01','2017-06-01',
                '2018-06-01','2019-06-01']),  'lower_window': 0, 'upper_window': 162})

prophet = Prophet(holidays=hurricane_season)
prophet.fit(train_dataset)
future = prophet.make_future_dataframe(periods=365, freq='D')
forecast = prophet.predict(future)
fig = prophet.plot(forecast, xlabel='Year', ylabel='Rainfall(in)')
a = add_changepoints_to_plot(fig.gca(), prophet, forecast)
plt.title('Rainfall prediction with Hurricane Season')
plt.show()
fig2 = prophet.plot_components(forecast), plt.show()
```



Our graphs for our model without outliers show some changes from our previous trend graphs.

The first graph shows a deep dropped in 2011 with a steep rebound and steady climb following that up until 2018 when a slight, steady decrease began. Our model shows this decline to continue.

Our yearly graph shows a consistent ripple like trend but then a jump starting in June with some steep upwards and downward trends through till December. Considering this is during our traditional hurricane months, more research could be done by looking at periods of non rain bookending hurricanes and tropical storms.



Looking at the yearly graph, for our model without outliers, we can see that there was a dip by late July into mid August but then the peak was early September. The trend graph shows an early dip but that it raised over the last few years of the decade.

In comparison between our 2 models we see that there was a higher trend for our model with our outliers included. There was more movement between 2015-2017 for this model as compared to our model without our outliers.

Cross Validation

Our cross validation predictions performance was set to 365 day horizon, starting with 730 days of training data with predictions every 90 days.

```
df_cv = cross_validation(prophet_basic, initial='730 days', period='90 days', horizon = '365 days')
```

	ds	yhat	yhat_lower	yhat_upper	y	cutoff
0	2012-02-07	0.113578	-0.686969	0.926501	0.00	2012-02-06
5	2012-02-08	0.115323	-0.698510	0.992968	0.00	2012-02-06
10	2012-02-09	0.099636	-0.779784	0.884134	0.00	2012-02-06
15	2012-02-10	0.084166	-0.824978	0.950375	0.95	2012-02-06
20	2012-02-11	0.081202	-0.781498	0.869382	0.03	2012-02-06

Model without outliers

	ds	yhat	yhat_lower	yhat_upper	y	cutoff
0	*2012-02-07*	*0.113578*	*-0.675537*	*0.927500*	*0.00*	*2012-02-06*
5	*2012-02-08*	*0.115323*	*-0.698430*	*0.894583*	*0.00*	*2012-02-06*
10	*2012-02-09*	*0.099636*	*-0.771722*	*0.899880*	*0.00*	*2012-02-06*
15	*2012-02-10*	*0.084166*	*-0.815537*	*0.925952*	*0.95*	*2012-02-06*
20	*2012-02-11*	*0.081202*	*-0.787705*	*0.877882*	*0.03*	*2012-02-06*

Model with outliers

The yhat feature represents the predicted value of y. Yhat_lower and yhat_upper show the range of uncertainty intervals.

MSE, RMSE MAE metrics

	horizon	mse	rmse	mae	coverage
0	37 days	0.225405	0.474768	0.244744	0.949843
1	38 days	0.223978	0.473264	0.243955	0.950143
2	39 days	0.222439	0.471634	0.243544	0.950908
3	40 days	0.222167	0.471346	0.243779	0.950994
4	41 days	0.223215	0.472457	0.245227	0.949960

Model without outliers

	horizon	mse	rmse	mae	coverage
0	37 days	0.316871	0.562913	0.263325	0.954915
1	38 days	0.315465	0.561663	0.262647	0.955028
2	39 days	0.289953	0.538472	0.259836	0.955896
3	40 days	0.268476	0.518146	0.257905	0.956428
4	41 days	0.269588	0.519218	0.259444	0.955684

Model with outliers

FB Prophet Summary

Our Prophet models provided nice visuals especially in regards to past trends and what the next couple of years may look like. Our MAE, MSE and RMSE scores are decent but could be better.

Our tables showing our yhat metrics are useful but can use more tuning so our predictions have less error percentages and can better predict future rainfalls.

Simple Linear Regression

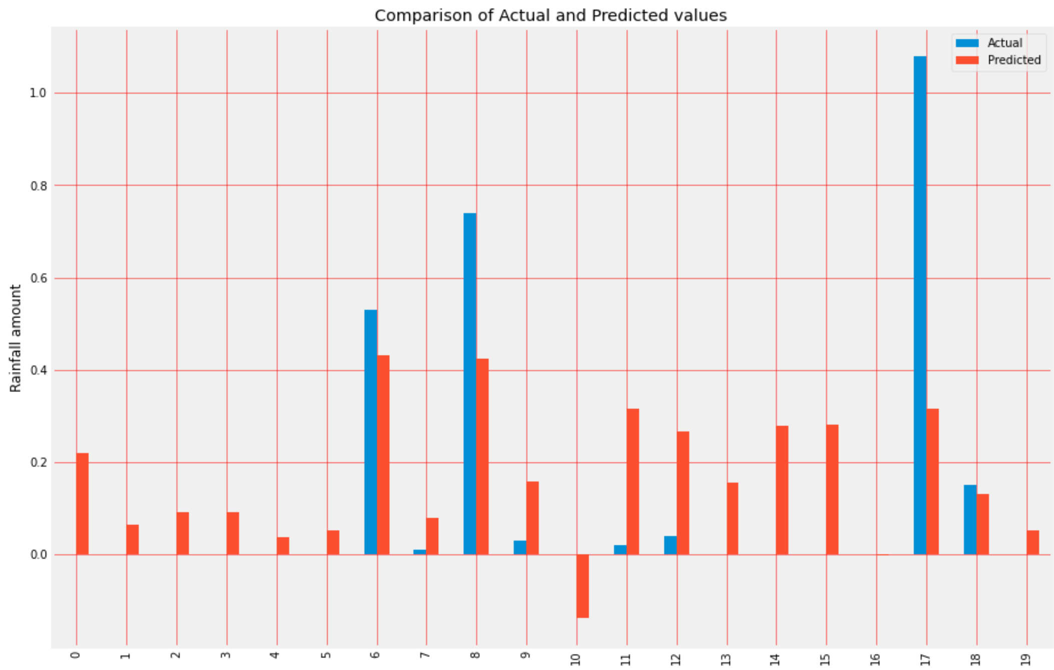
Our 2nd model focuses on Linear Regression. We once again separated between datasets with and without outliers we then 1st created a simple linear regression with Humidity as our independent variable and Rainfall(in) as our dependent variable.

We conducted a 70% split of our dataset as we conducted **train, test, split()** before training with **LinearRegression()** and fitting our training.

Our simple LR model without our outliers produced a prediction of:

	Actual	Predicted
0	0.00	0.218371
1	0.00	0.064555
2	0.00	0.091462

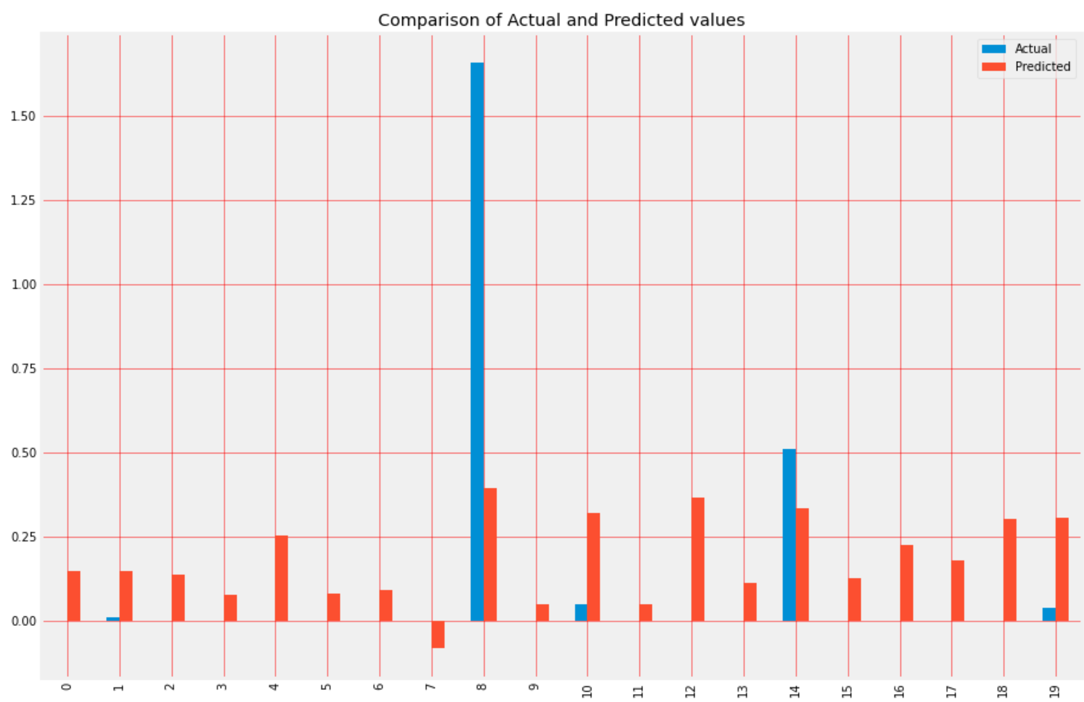
3	0.00	0.091462
4	0.00	0.037649
...	...	...
3640	0.00	0.145275
3641	2.01	0.436616
3642	0.00	0.104915
3643	0.36	0.145275
3644	1.78	0.199088



Our simple LR model with our outliers produced a prediction of:

	Actual	**Predicted**
0	0.00	0.150104
1	0.01	0.150104
2	0.00	0.136742
3	0.00	0.078345
4	0.00	0.254031
...	...	...
3644	0.67	0.193160

3645	0.00	0.250567
3646	0.00	0.049641
3647	0.00	-0.022119
3648	0.27	0.150104



Multiple Linear Regression

For our multiple linear regression models we used our full dataset and used our independent variables and once again used Rainfall(in) as our dependent variable.

```
X = df_lr[['Dewpoint_temp(°F)', 'Humidity(%)', 'Sealevel_pressure(Hg)', 'Max_temp(°F)', 'Min_temp(°F)', 'Average_temp(°F)', 'Wind_speed(mph)']]

y = df_lr['Rainfall(in)']
```

We once again conducted train_test_split() with a split of 70% before fitting and predicting our models.

We took a look at our features and their importances to our dependent variable:

	Coefficient
Dewpoint_temp(°F)	-0.017660
Humidity(%)	0.018864
Sealevel_pressure(Hg)	-0.212812
Max_temp(°F)	-0.002402
Min_temp(°F)	0.007863
Average_temp(°F)	0.009596
Wind_speed(mph)	0.019038

Model without outliers

	Coefficient
Dewpoint_temp(°F)	*-0.019476*
Humidity(%)	*0.020152*
Sealevel_pressure(Hg)	*-0.228957*
Max_temp(°F)	*0.006074*
Min_temp(°F)	*0.017008*
Average_temp(°F)	*-0.006568*
Wind_speed(mph)	*0.022094*

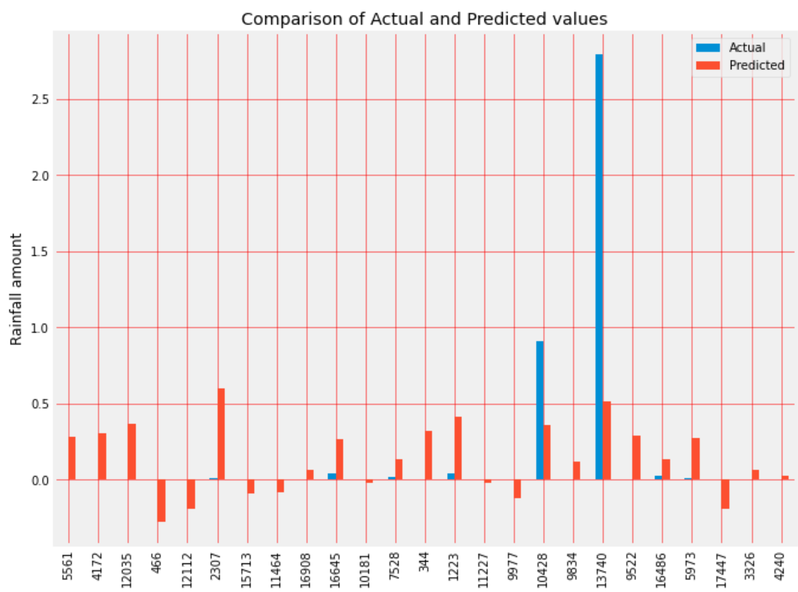
Model with outliers

We didn't see too much of a change between the two. Our temperature features in our model with outliers dropped slightly.

Our predictions for our model without outliers:

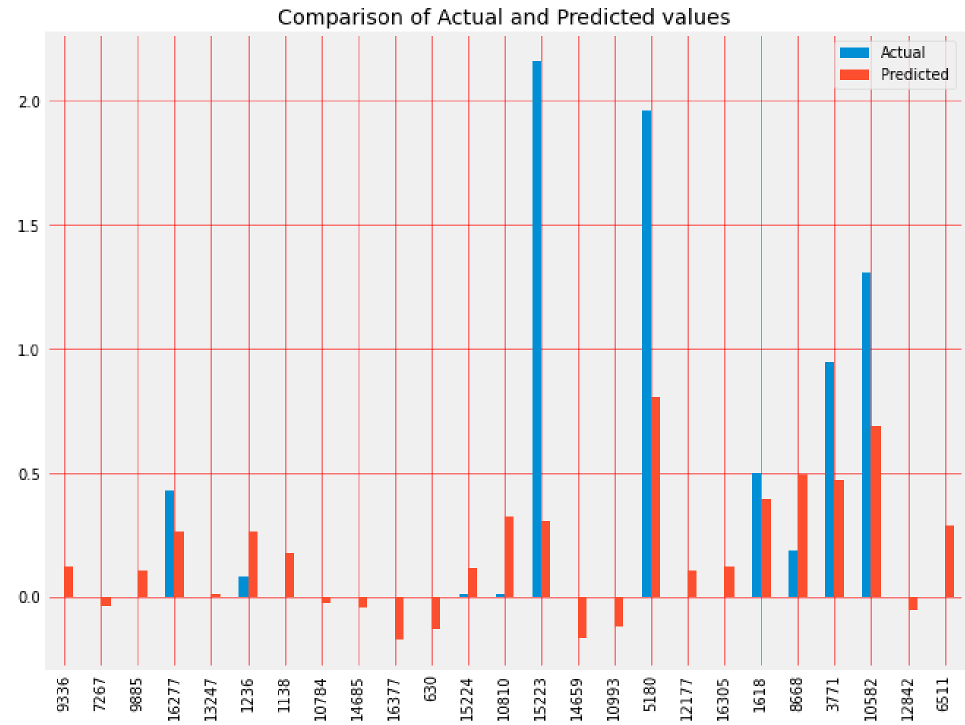
	Actual	**Predicted**
5561	0.00	0.284373
4172	0.00	0.303114
12035	0.00	0.370171

466	0.00	-0.273421
12112	0.00	-0.189138
2307	0.01	0.595830
15713	0.00	-0.089154
11464	0.00	-0.080771
16908	0.00	0.066009
16645	0.04	0.265975



Our predictions for our model with outliers:

	Actual	**Predicted**
9336	0.00	0.123703
7267	0.00	-0.039213
9885	0.00	0.107204
16277	0.43	0.264623
13247	0.00	0.010790
1236	0.08	0.264421
1138	0.00	0.175265
10784	0.00	-0.023462
14685	0.00	-0.041331
16377	0.00	-0.170279



LR Summary:

Our simple LR models produced metrics of:

Model without outliers	*Model without outliers*
Mean Absolute Error: **0.2303198905858845** Mean Squared Error: **0.16931549835352736** Root Mean Squared Error: **0.41147964512661783** R2 score: **0.11**	Mean Absolute Error: **0.24838378826197569** Mean Squared Error: **0.24144176356370262** Root Mean Squared Error: **0.4913672390012409** R2 score: **0.10**

Our multiple LR models produced metrics of:

Model without outliers	*Model without outliers*
Mean Absolute Error: **0.21423314772455385** Mean Squared Error: **0.1570659705556637** Root Mean Squared Error: **0.39631549371134067** Test R2 score: **0.22** Train R2 score: **-2.60**	Mean Absolute Error: **0.2348102012291876** Mean Squared Error: **0.2609202858600729** Root Mean Squared Error: **0.510803568762076** Test Test r2 score: **0.17** Train r2 score: **-2.73**

Comparison between the single LR model(s) and the LR multiple model(s)

Our Testing r2 score isn't as high as we would like it to be but we also see that our training score didn't come out very well either. Our training score is way off as well. The features that we used don't work as well with our Rainfall variable as well as we would have liked it too.

MAE and RMSE metrics for our models are good results for what we had to work with within our LR models.

As mentioned, to create a better model for the Gulf Coast area, we may need to look at other features that are located elsewhere in the country or the world. Future LR modeling will need a combination of more features or possibly adding more cities to our dataset that are more closely located to one another.

XGBoost

Our 3rd model focuses on XGBoost. We once again separated between datasets with and without outliers. Once again Rainfall(in) is our dependent variable.

We conducted a 70% split of our dataset as we conducted **train, test, split()** before training with **XGBRegressor()** and fitting our training.

For both of our model version (with and without outliers) we created and ran multiple XGB models to see which one produced the best metrics.

For our model from our dataset without outliers we settled on:

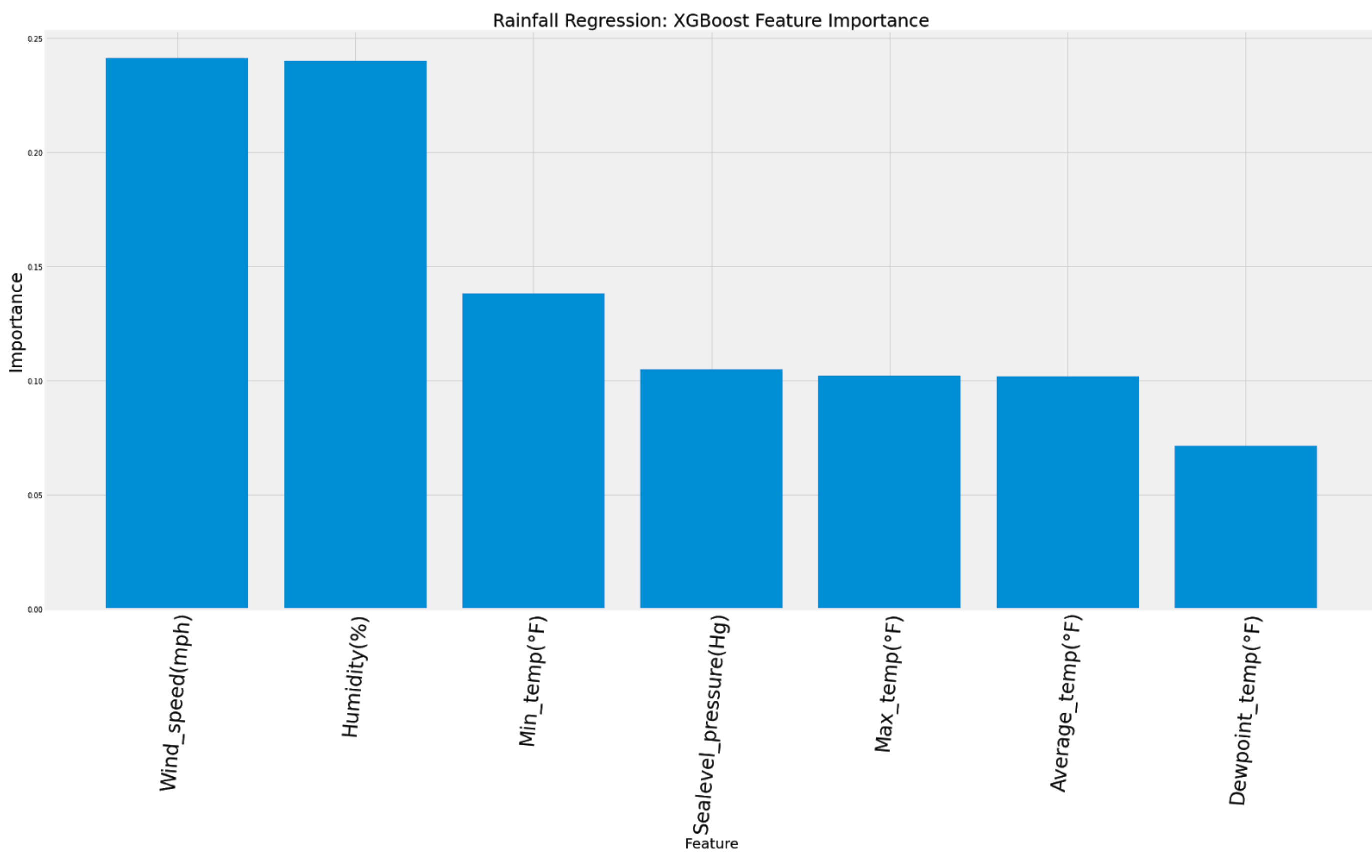
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)

model2 = XGBRegressor(n_estimator=300)

#fit model
model2.fit(X_train, y_train)

#predict model
y_train_pred2 = model2.predict(X_train)
y_pred2 = model2.predict(X_test)
```

Which produced the metrics:

```
Train r2 score:  0.6579688260790921
Test r2 score:  0.27842261500414345
```

Train RMSE: **0.2113**
Test RMSE: **0.3873**
Train Mean Square Error: **0.0446**
Test Mean Square Error: **0.1500**

Our featured importance features:



We see that Humidity and Windspeed have the highest importance amongst our values. Average temperature and Sealevel pressure have nearly half the importance. Dewpoint temperature being so low is surprising.
Our test and training scores means our model isn't too overfit.

Our feature importance keeps in line with our other models and correlation heatmaps with rainfall. Humidity has usually been at or near the top. Dewpoint seems to be a bit lower than usual.

Our score is a bit low. We could add other outside features to see if it would increase our score.

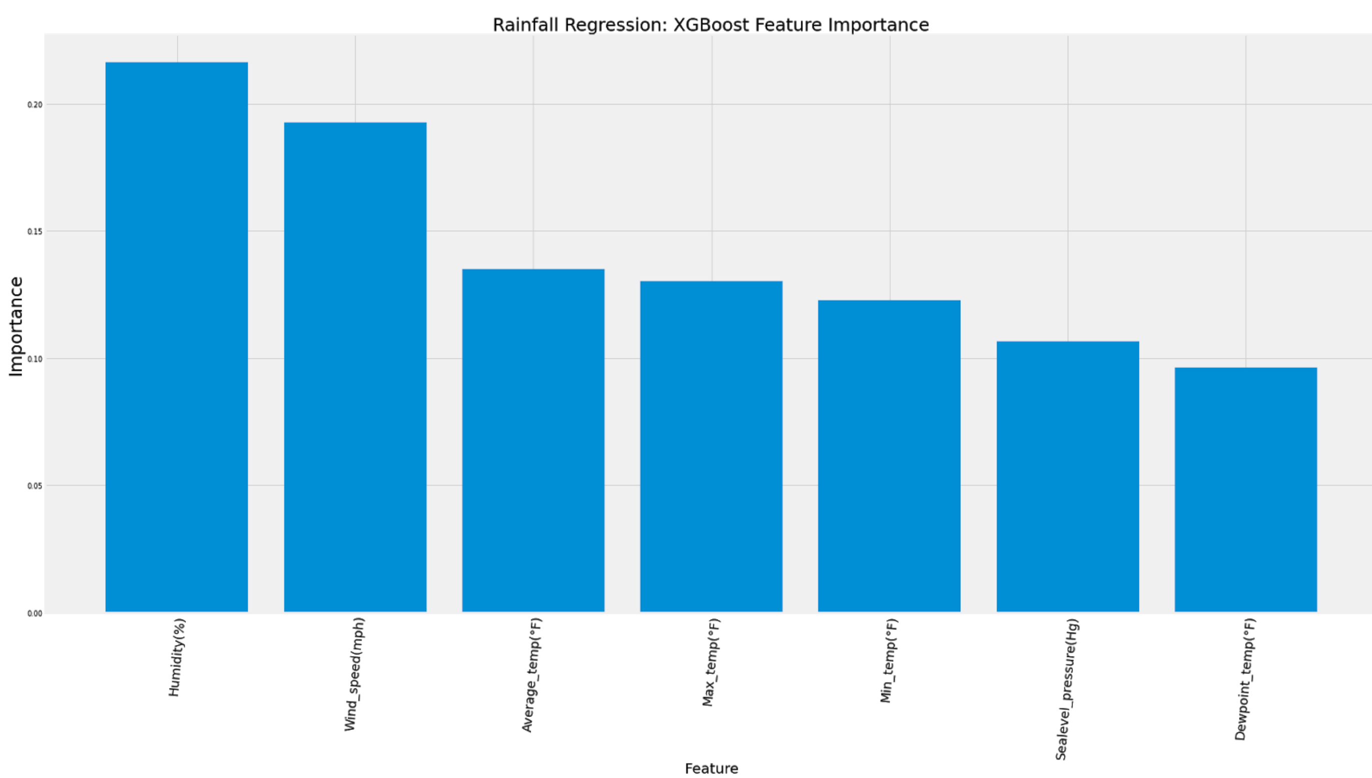
For our model from our dataset with outliers we settled on:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)

model5 = XGBRegressor('reg:squarederror', learning_rate=0.05,
n_estimators=300, max_depth=5,
    min_child_weight=3, gamma=0, subsample=0.8,
colsample_bytree=0.8, nthread=6, scale_pos_weight=1, seed=15)

model5.fit(X_train, y_train, verbose=False)
y_train_pred5 = model5.predict(X_train)
y_pred5= model5.predict(X_test)
```

Which produced the metrics:

Train r2 score: **0.27710690972715657**
Test r2 score: **0.34627797556802054**
Train RMSE: **0.3180**
Test RMSE: **0.4086**
Train Mean Square Error: **0.1011**
Test Mean Square Error: **0.1670**

Our featured importance features:



For our model with outliers excluded, our test and training scores means our model is a bit too overfit. Our model with outliers included had less overfitting and may be the model that we focus on to build a better model that helps us predict rainfall.

Our feature importance keeps in line with our other models and correlation heatmaps with rainfall. Humidity has usually been at or near the top. Dewpoint seems to be a bit lower than usual.

Our score is a bit low. We could add other outside features to see if it would help increase our score. With more features, our efforts to tune our hyperparameter models so they aren't as overfitted as they currently are.

Random Forest

Our 4th model focuses on Random Forest. We once again separated between datasets with and without outliers. Once again Rainfall(in) is our dependent variable.

We conducted a 70% split of our dataset as we conducted **train, test, split()** before training with **RandomForestRegressor()** and fitting our training.

For both of our model version (with and without outliers) we created and ran multiple RF models to see which one produced the best metrics.

```
#train, test, split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)

trees = RandomForestRegressor(n_estimators= 250)

#fit our model
trees.fit(X_train, y_train)

#predict our model
y_pred = trees.predict(X_test)
y_pred_rf_out = trees.predict(X_train)
```

Just as we have with our other models we ran a featured importance dataframe. Our model without outliers produced:

	feature	**importance**	**rank**
1	Humidity(%)	0.286172	1
6	Wind_speed(mph)	0.212961	2
2	Sealevel_pressure(Hg)	0.149999	3

0	Dewpoint_temp(°F)	0.103797	4
3	Max_temp(°F)	0.096114	5
4	Min_temp(°F)	0.088046	6
5	Average_temp(°F)	0.062910	7

Our RF model without outliers produced:

MAE test score: **0.1699**
r2 test score: **0.2922**
RMSE test score: **0.3816**

Our RF model with outliers produced:

	feature	**importance**	**rank**
1	Humidity(%)	0.272435	1
6	Wind_speed(mph)	0.212199	2
2	Sealevel_pressure(Hg)	0.150871	3
0	Dewpoint_temp(°F)	0.104089	4
4	Min_temp(°F)	0.099537	5
3	Max_temp(°F)	0.098977	6
5	Average_temp(°F)	0.061892	7

Our RF model with outliers metrics:

MAE test score: **0.1699**
r2 test score: **0.3106**
RMSE test score: **0.1957**

RF Summary

Our models provided better MAE and RMSE metrics but as with our other models, our r2 test score isn't as high as we would have liked it to be.

The feature importance tables are consistent with our other models, at least in the 2 main features that have the highest importances amongst the other features. Humidity and Windspeed leads the way but the difference between our RF models is that the minimum and maximum temperature features swap importances places between the 2 models.

Our RF isn't bad but to make it better and useful to use for rainfall predictions more will be needed to gain better r2 scores as well as better contain any overfitting between our test and training sets.

Comparing Models

With 4 different models and 2 versions of each we collected our various model metrics so that we could view them all in 2 tables.

Models without outliers:

	name (without outliers)	**R2 - Test**	**R2 - Train**	**MAE - Test**	**MAE - Train**	**RMSE - Test**	**RSME - Train**
0	FB Prophet	-0.001	NaN	0.248	NaN	0.468	NaN
1	Linear Regression(multiple)	0.223	-2.600	0.214	0.220	0.396	0.406
2	XGBoost	0.255	0.650	0.161	0.220	0.373	0.217
3	Random Forest	0.292	0.847	0.167	0.065	0.382	0.142

As we have touched on throughout our report, our r2 scores are a bit lower than we would like and there is overfitting/underfitting issues with our models. XGBoost and RF would be the models that we would focus on to create a better prediction model. With more, or different features, we could rerun the models and use more hyperparemter tuning methods to achieve a better result.

Our MAE and RMSE metrics can be better but are not too far off of where we need to be. As mentioned, reworking our models a bit should also help us achieve better MAE and RMSE results.

Models with outliers:

	name (with outliers)	**R2 - Test**	**R2 - Train**	**MAE - Test**	**MAE - Train**	**RMSE - Test**	**RMSE - Train**
0	FB Prophet	-0.003	NaN	0.261	NaN	0.544	NaN

1	Linear Regression(multiple)	0.174	-2.726	0.235	0.233	0.511	0.456
2	XGBoost	0.346	0.277	0.167	0.233	0.409	0.318
3	Random Forest	0.311	0.851	0.170	0.065	0.196	0.026

Our XGBoost model looks to be the model worth focusing on and trying to achieve better results with. The r2 score could be higher but unlike with our other models this model. The MAE and RMSE are good as well for what the model had to work with. The RMSE is a bit high considering the lower MAE.

RF model is respectable too. Working on the overfitting would be a major focus if this model was chosen. The MAE and RMSE metrics are low which is what we want to see.

Prophet and LR models didn't work well with the features and variable we ran. Adding features along with taking out low scoring features could help boost their results.

Conclusion

Along with the models that we showcased within our project and reports, we also ran various versions of our 4 models with different approaches. Such as models that only focused on only 1 or 2 of our cities, instead of all 5 cities. These models performed about the same as the results shown within this report. We also ran models trying to predict the simplicity of if it would rain or not, instead of trying to predict the amount of rain a future date may or may not have. These models also produced the same range of metrics as the models we showed in our report. This tells us that the features we used for our project will need to be revisited.

To better predict weather for the Gulf Coast Area, looking at other parts of the world might be a good option. Considering weather that eventually affects certain regions, in our case the Gulf Coast area, we know that storms and weather patterns are influenced in other parts of the world 1st.

XGBoost with outliers include looks to be our best performing model. It didn't produce the metrics that we would have liked but it does show promise that if more attention was given to this model, we could produce better results and predictions. As mentioned, the model could perform better with more features while also removing lower importance features.

Our RF model with outliers included didn't obtain a good r2 score but we were encouraged by the MAE and RMSE scores. Just as we mentioned with the XGBoost model that we consider our best performing model, this RF model could also be an option to further pursue to see if we can achieve better results with some reworking of our features and by tuning the hyperparameters of our RF model.

Our, lesser performing models Prophet and LR models showed some valuable information even if we didn't like the metics that they offered. For Prophet, the visualization of future trends can be usual for people and industries who rely on future rainfall predictions. The visuals quickly and easily showed past trends and future trends in regards to rainfall.

For our project we casted a pretty wide Gulf Coast region so depending on who needed the prediction info, we could turn our models' hyperparameters to have more cities and/or smaller radius to a given area. Instead of our cities ranging from Houston all the way to Tampa, we could solely focus on Houston, Tampa or any of the other cities we chose for our project. More than likely a big storm hitting the New Orleans area would also affect people nearby in Mississippi and lower Alabama. Also depending on the size and direction of a storm, it could pass through all 5 of our cities in its path.