SOULPAGE

IT SOLUTIONS

Time Series Forecasting of Oil & Gas production

Forecasting Analytics of Oil & Gas production

Problem statement:

A Drop of Oil, on its incredible journey from several thousand feet underground to the gas station, leaves behind a tremendous amount of data. This data often remains scattered, unstructured and unused. The Analysis shows that Oil & Gas industry uses less than 1% of the total available data. Using such unstructured & messy data turning into insights could help stakeholders to take the Right decisions at Right time. Data Science in the Oil & Gas industry had its use in various segments from production to households like Upstream, Midstream and Downstream. In this report, we have data analysis and forecasting time series on Oil & Gas production that happened in 1979 to 2019 from Alaska Oil & Gas Conservation Commission.

We have large amount of data produced every year from various Oil & gas wells across the world. We have one of the worlds largest oil well called Prudhoe bay from Alaska.

Solution:

With all the data available we can not simply perform regression or any other machine learning models, as this data is related to time which means the there would be a trend, seasonality with respect to time. Hence this data is a good fit to perform Time series forecasting which would give us a foresee the production that is going to happen in future. There are also lots of questions answered in this report with the available data.

Tools Used:-

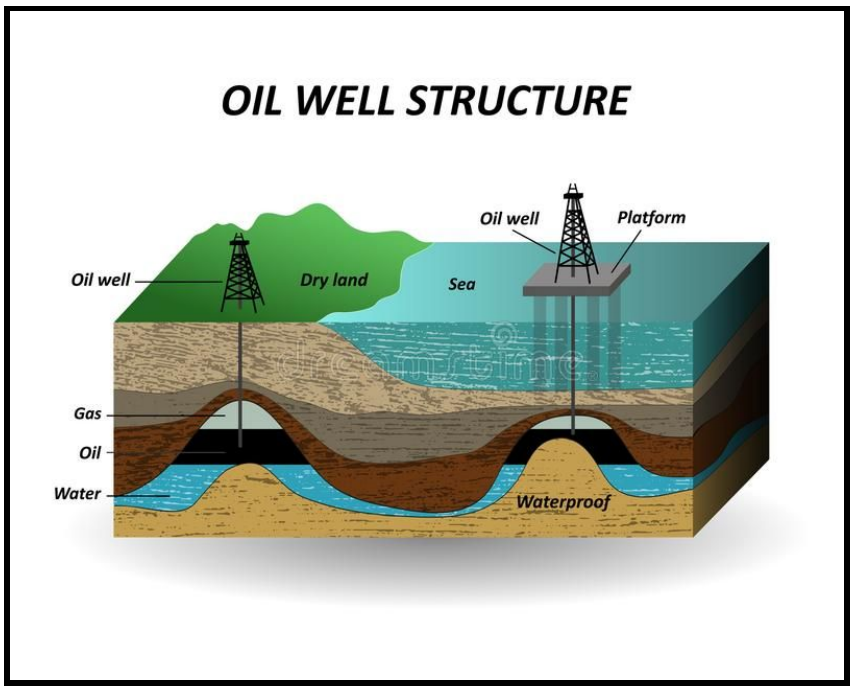
- Python
- Pandas
- Data cleaning
- Data visualization
- Univariate & Bivariate Analysis
- Time-series Forecasting
- Auto-Regressive Integrated Moving Average

Data set:-

Alaska Oil & Gas Conservation Commission has monthly production of Oil & Gas details from all the wells and from all the fields updated to the website. This dataset has fields like Operator, Well status, Area, Fields, Pools, Pads and Different methods of production, types and Production values of Oil, Gas and Water from these wells.

Process:-

Once after we get the dataset, apply various Data cleaning techniques to clean and understand the data. In this Data cleaning, all the unstructured data gets converted to Structured with applying feature removal, feature conversion, renaming, using the right data type for different attributes. Use Data analysis and data visualization to understand the flow of data with respect to time and various other variables. Frame the data to



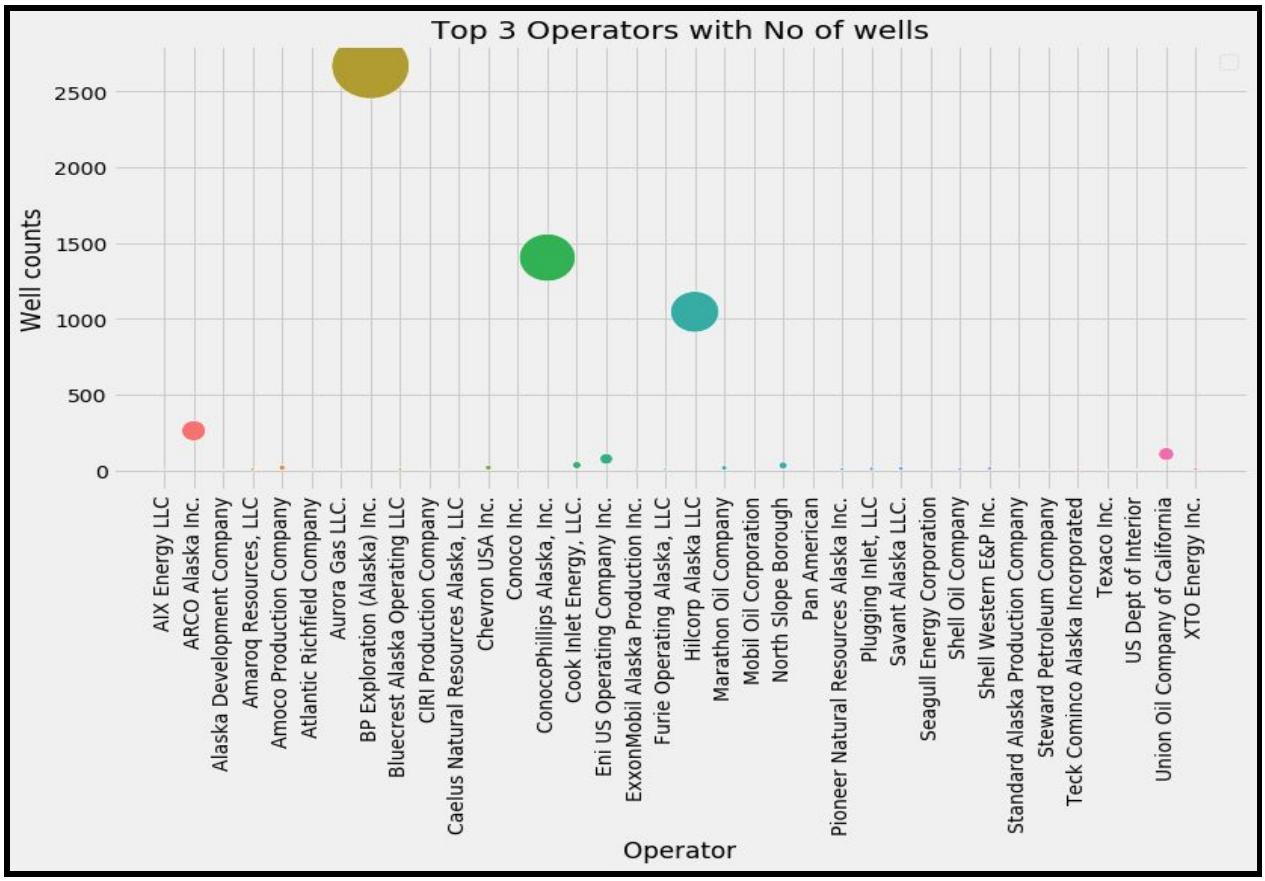
perfectly fit for time series forecasting with the time sorted in order.

Who had more number of Oil wells?

We have 36 operators who own more than 5000 wells from Alaska. Each operator had less or more number of wells digging oil and gas from the earth. From the graph below we can clearly see that BP Exploration(Alaska) Inc. has more number of wells following ConocoPhillips Alaska and Hilcorp Alaska.

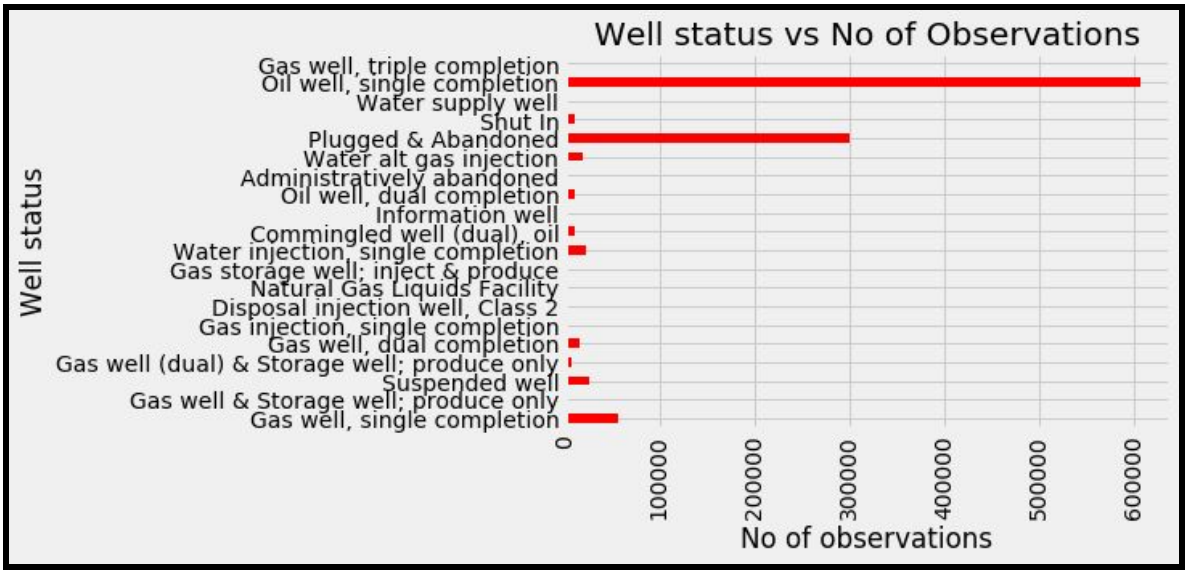
BP Exploration - 2669
ConocoPhillips Alaska - 1404
Hilcorp Alaska LLC - 1046

And there are many operators with only one well under them and few of them are
Alaska Development Company
CIRI Production Company
Conoco Inc
Mobil Oil Corporation

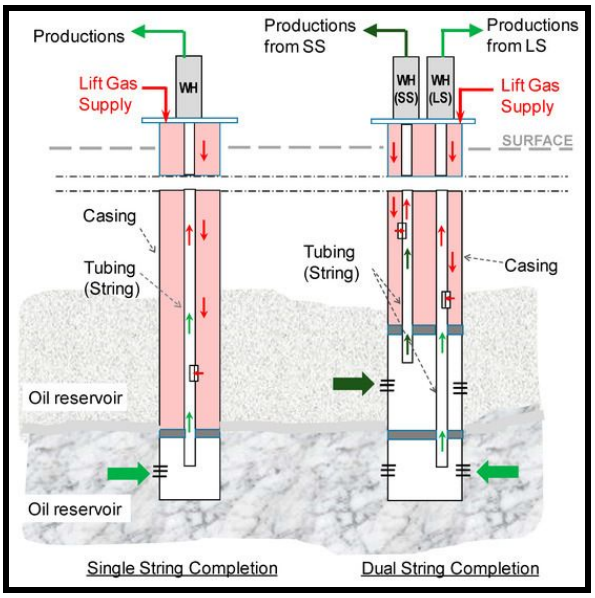


What are various types of Well conditions?

Out of all the Data points, we have Single completion oil wells more in our observations. There are nearly 24 types of conditions that all the wells are in. Each condition tells about the type and system used in it.

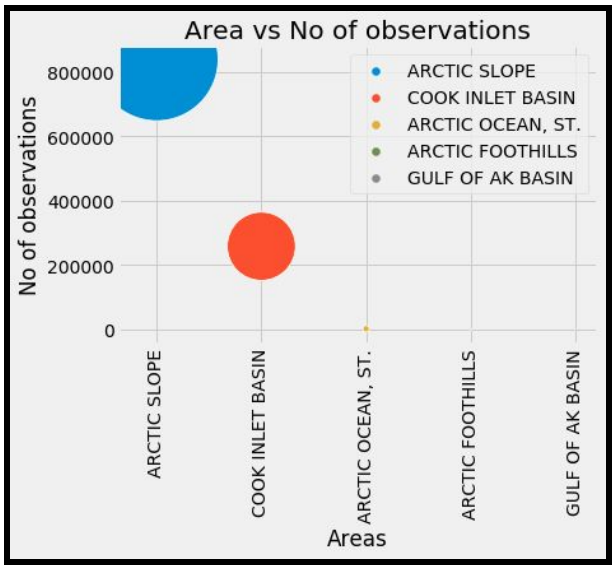


Mostly we can see wells with Single and Dual Completion where in Single Completion, a single tubing or string is inserted into the oil reservoir after casing is done carefully. After completion, using Lift Gas supply to create pressure in these tubes to get the oil out.



Which area had more of Oil Production?

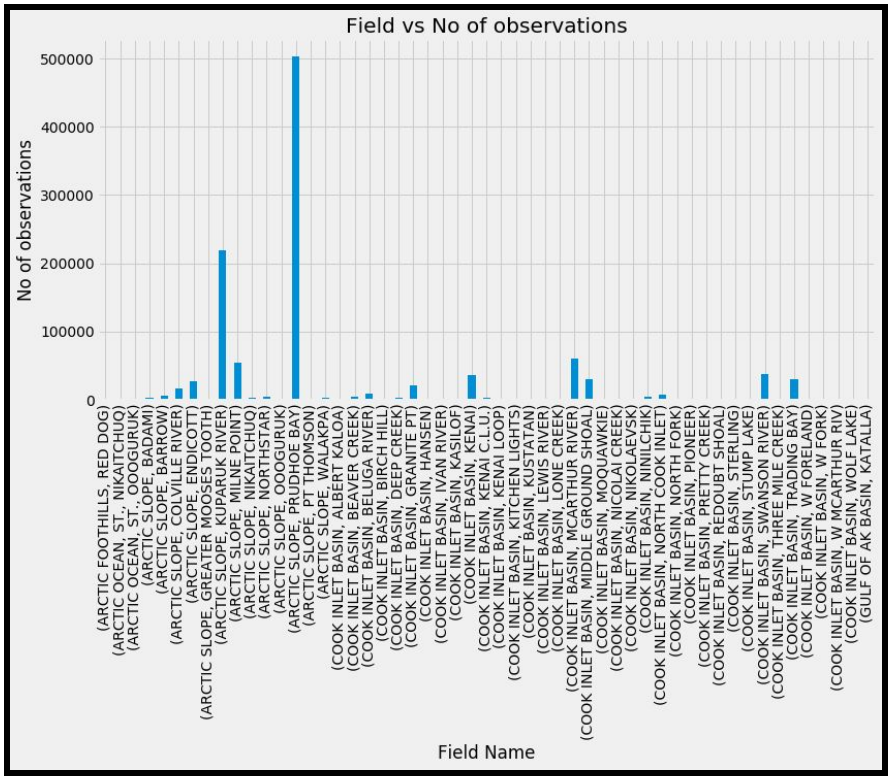
Though we have more than 5000 number of oil wells, all of them are in just 5 Areas. In those 5 Areas from the fig below, we can see that more observations are in Arctic slope followed by Cook Inlet Basin.



The Area with less no of observations is Gulf of AK Basin with just 15 observations.

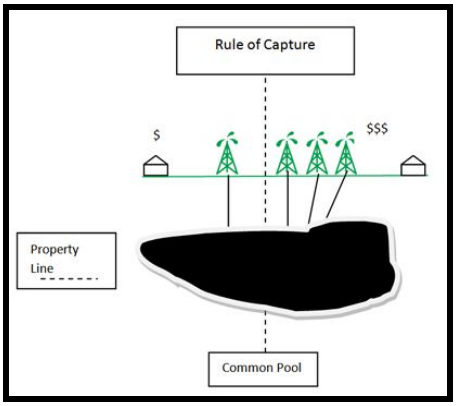
Field:-

An Oil Field is a tract of land used for extracting petroleum, otherwise known as crude oil, from the ground. An oil field consists of a reservoir in a shape that will trap hydrocarbons and that is covered by an impermeable or sealing rock. We had an overall 51 Fields under 5 Areas.



From the figure above we can see that Prudhoe bay from Arctic slope is the largest Oil filed, which in fact the largest Oil filed right now. Our analysis matched with the fact.

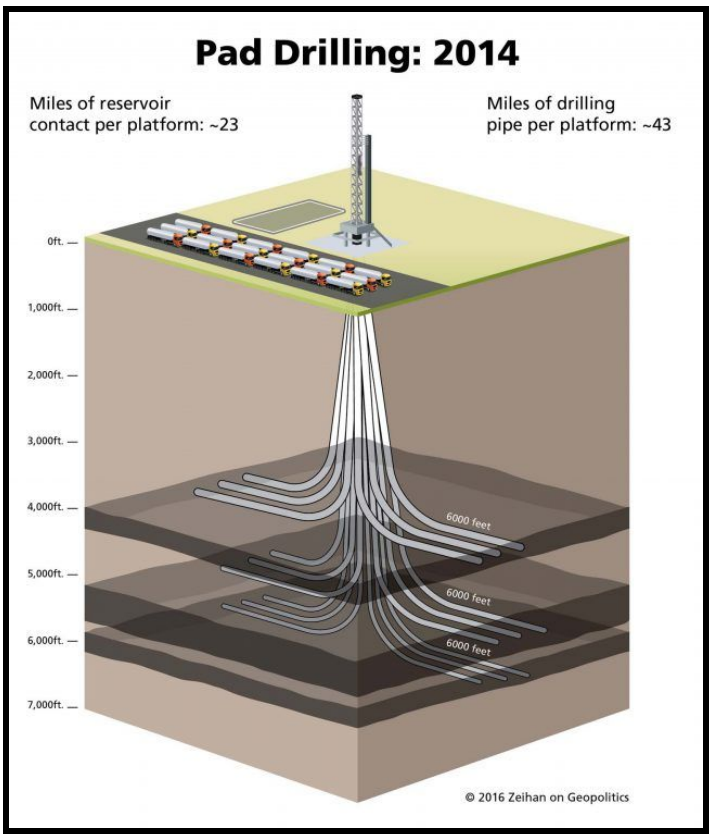
What is the Largest pool that more production happening from?



We have a total of 109 pools in Alaska, in which Prudhoe Oil filed sits on is the largest oil pool followed by Kuparuk river and Hemlock. The small pool is PT M undefined followed by Beluga Undef gas and Undef permafrost wtrsp

Pad Drilling:

Pad drilling some times called Multi pad drilling is a drilling practice that allows multiple wellbores to be drilled from a single, compact piece of land known as PAD.



We have a total of 197 Pads were built or constructed on 109 pools. Top 3 largest pads in number are

Sr_scu - 23210
Kenai_unit - 22802
Tbf_monopad_platf - 21181

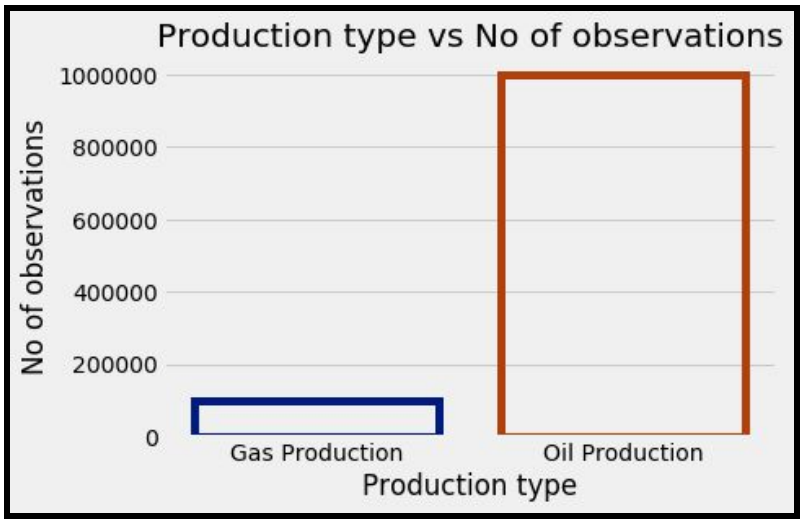
And the smallest pads in number are here as well

Mpu_m - 17
Grtr_Mooses_tooth - 36
Pt_thomson - 42

Production type:-

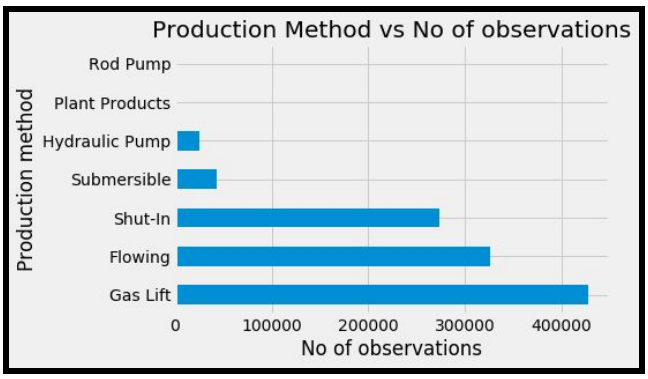
We have 2 types of production happening at Alaska i.e Oil and Gas. from the Fig below we can see the most type of production is Oil and very less is Gas.

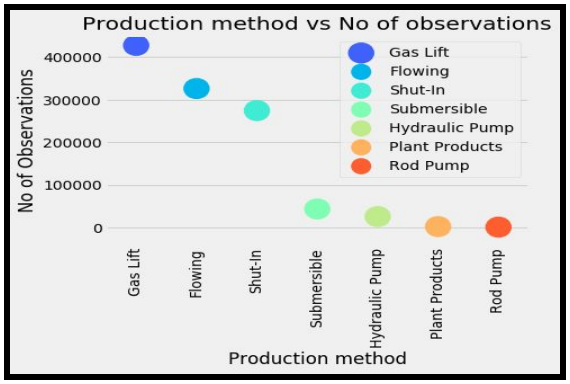
We have more than 1 million observations with Oil production type & only 0.1 million observations of type Gas production.



Production Method:-

There are a total of 7 production methods where each one is used in different types of Fields, Pools in order to extract the oil & gas from the earth. From the fig below we can see that Gas Lift is most used and Rod pump is the least used method to extract the oil from the ground.

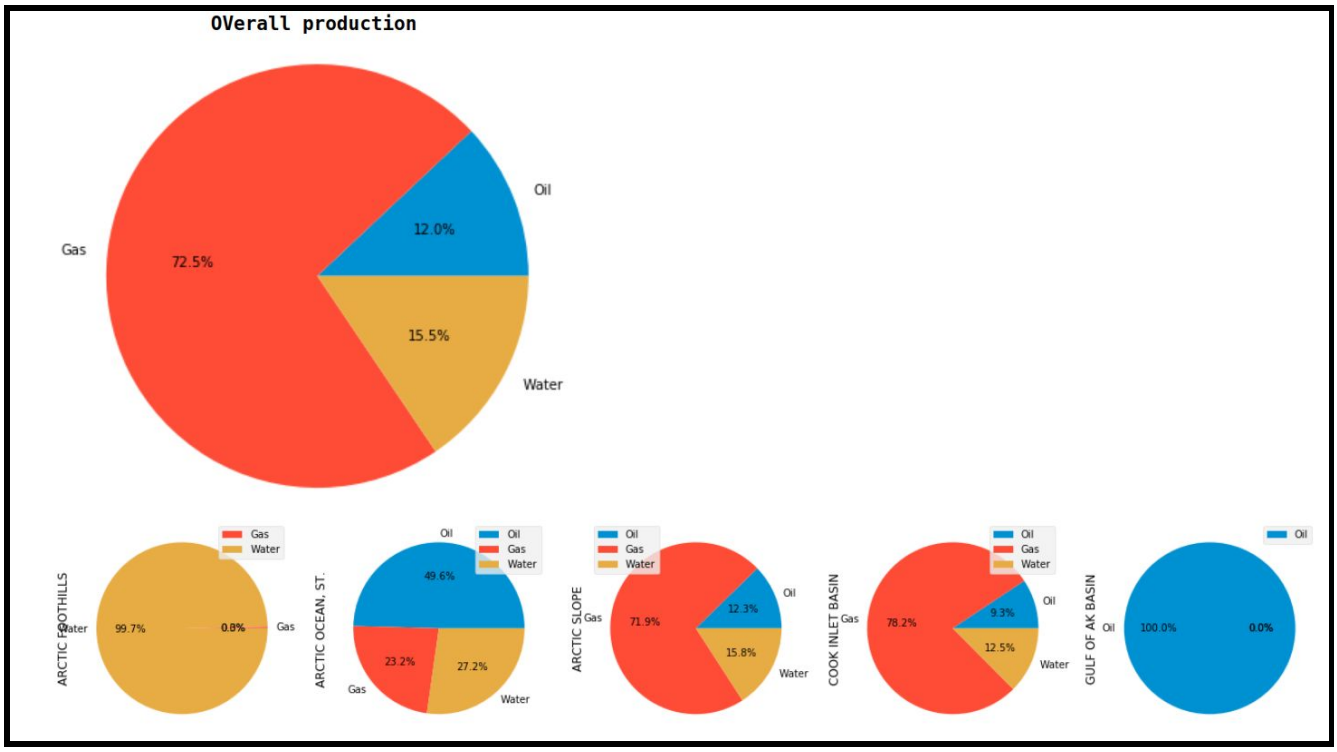




How much Oil, Gas and Water production is in 5 Areas?

Area	Oil	Gas	Water
ARCTIC FOOTHILLS	0	2557	852832
ARCTIC OCEAN, ST.	36614113	17111690	20120761
ARCTIC SLOPE	17318056276	101629728034	22388913995
COOK INLET BASIN	1376547859	11576411983	1855552642
GULF OF AK BASIN	154000	0	0

The arctic slope is producing more oil, gas and water compared to the remaining 4 other areas. Gulf of Ak basin is producing only oil and Arctic foothills are producing only gas& water.

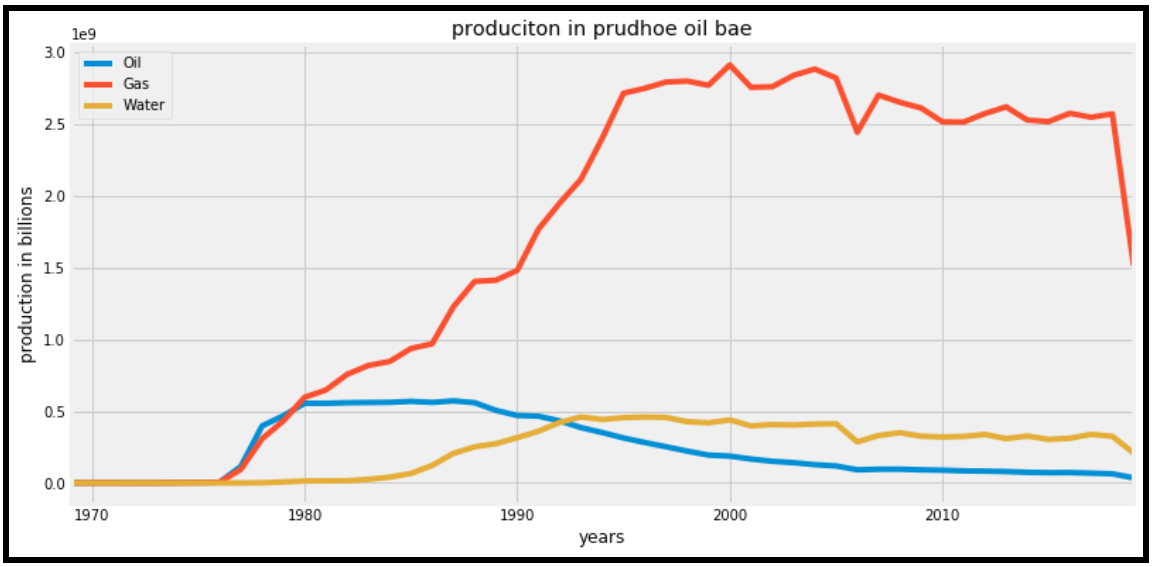


From the above figure, we can see 72.5% Gas,12.0% of Oil, and 15.5% of water as overall production from 1979 to 2019. Both the above table and the figure tells the same values in different representations. If you see all these 5 Areas individually we can say that the Arctic ocean, st is producing half of the oil in its total production. Whereas the Arctic slope and Cook inlet basin are producing more amount of gas than water and oil. Gulf of Ak basin is completely producing oil.

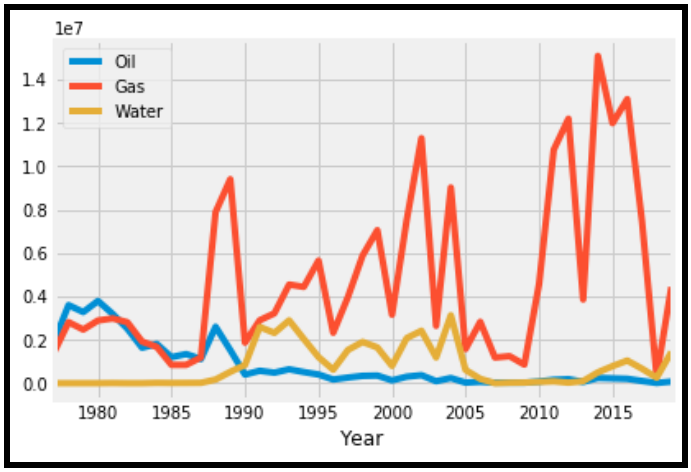
Overall, if we talk about the dataset, we have 5778 well Names, operated by 36 operators, with 24 well status from 1979 to 2019, in 5 Areas, with 51 fields, build on 109 pools over 197 pads using 9 production methods.

Time series forecasting

After analysing the dataset we know that Arctic slope is producing a large amount of data along with a large amount of oil, gas and water. Hence we took data from Arctic slope. In Arctic slope, there are 52 pools producing oil, gas & water in which Prudhoe bay pools is said to be the largest as a fact and as per our data set too.

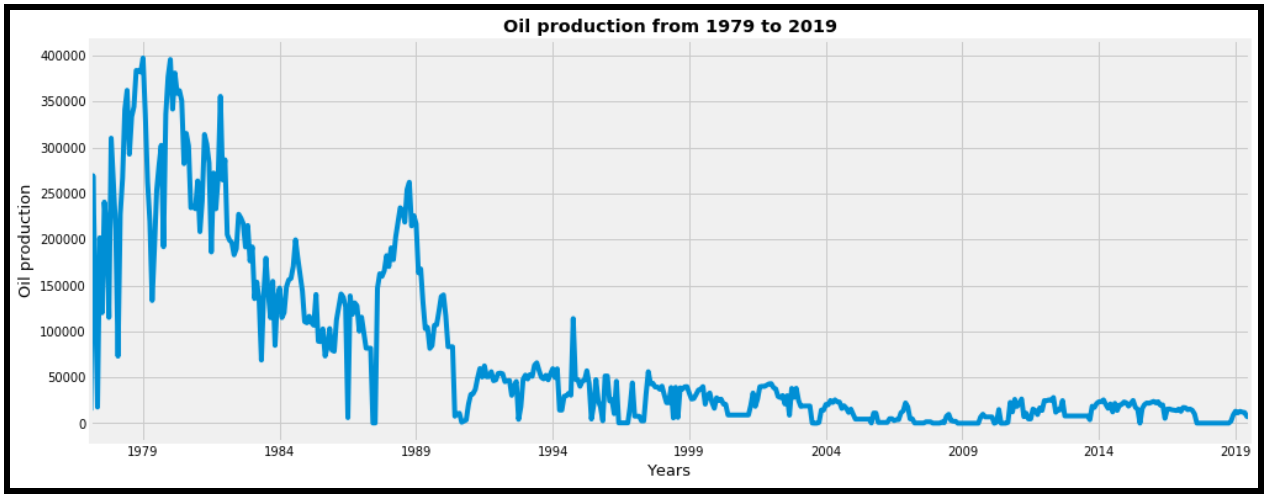


From the figure above we can see the oil, gas and water production occurred at Prudhoe bay in billions from 1970 to 2019. If we break it down this analysis till the wells, then we took one well out of more than 2000 wells. The below graph tells about the production at one well called Prudhoe bay unit 01-10.

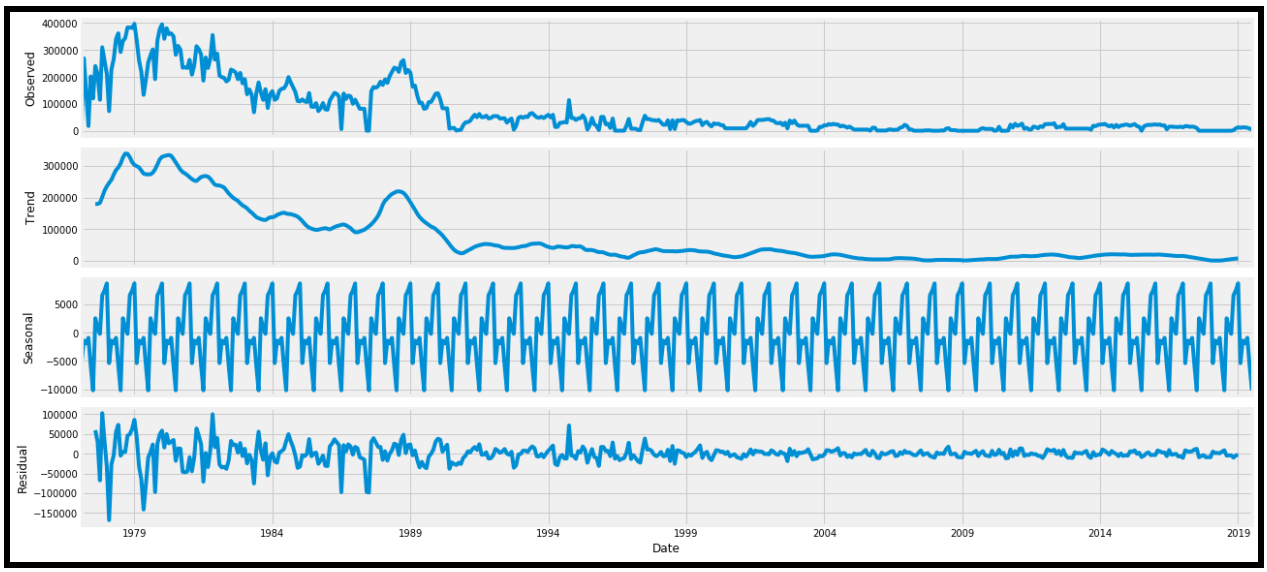


We forecasted each and every attribute independently, and here is the oil production in the well of Prudhoe bay unit 1-10.

Oil production:

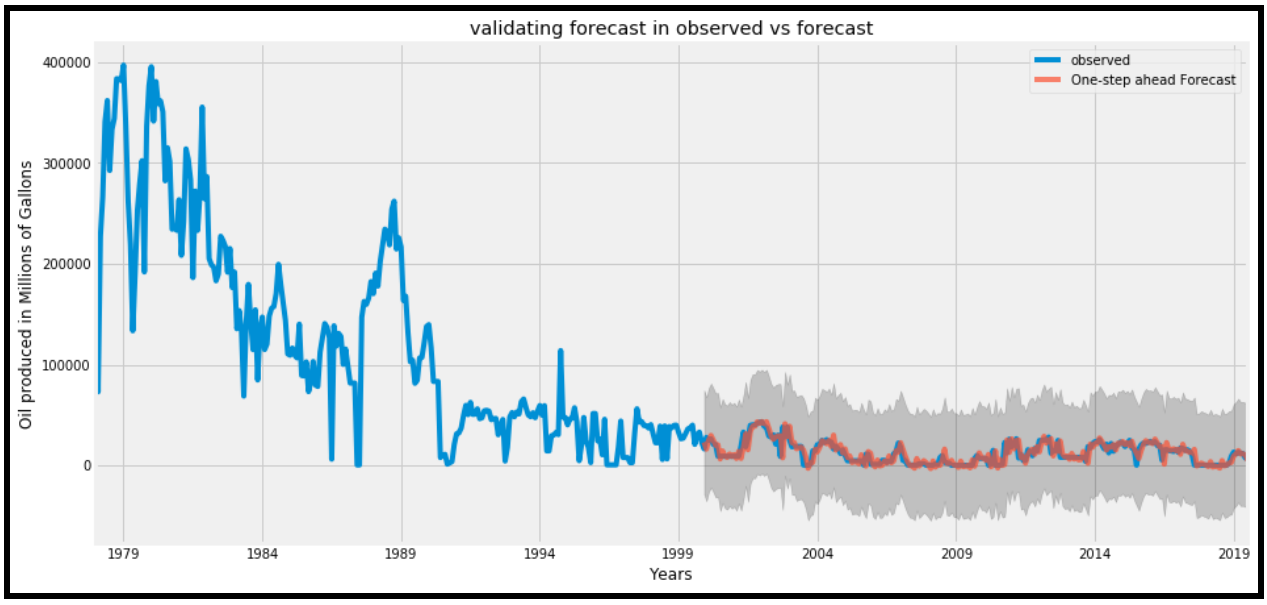


We have a **statsmodels** API used to implicit the stats into the model to analyse the properties of data to check whether it is good to time series or not. We have a decomposition chart used to see the trend, seasonality and residuals which we can not see through naked eyes.



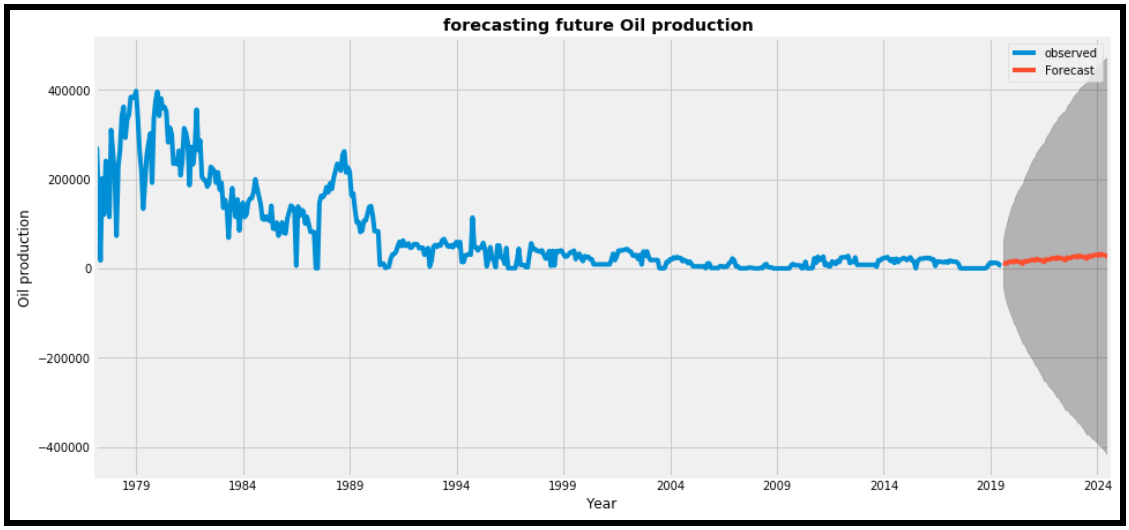
From the figure, we can see that we have a decreasing trend in oil production and seasonality occurring every yearly clearly. Hence this data would be better for time series as we have all the three like a trend, seasonality, and minimal & random residuals.

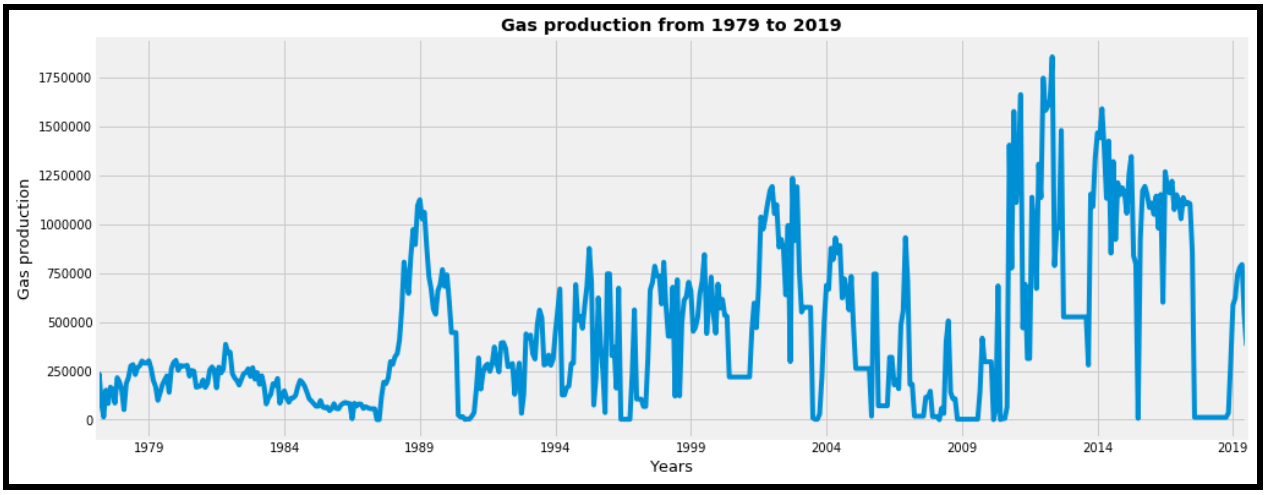
If our data is stationary, then it would be good for forecasting. We can convert any non-stationary data into stationary using Autoregressive & moving average with differencing lags. We also check the correlation between the present and previous observations & autocorrelation too. We employed a grid search technique in finding the right p,d,q values to get the low **AIC**(Akaike information criterion) value that can help in making our series smooth and helps to get the right seasonal order along with p,d,q.



After using the **ARIMA** model with the right p,d,q and seasonality s, we got our model predicting well with the accuracy of **68.79%**

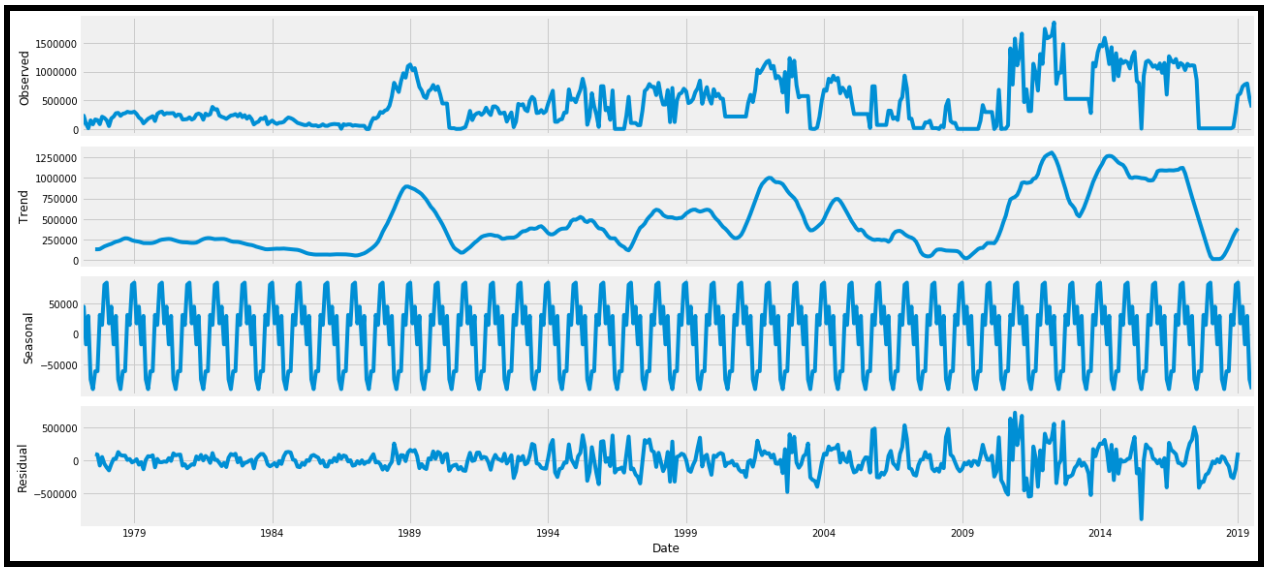
Below we have the forecasted time series of oil production that is going to happen for the coming 50 months.





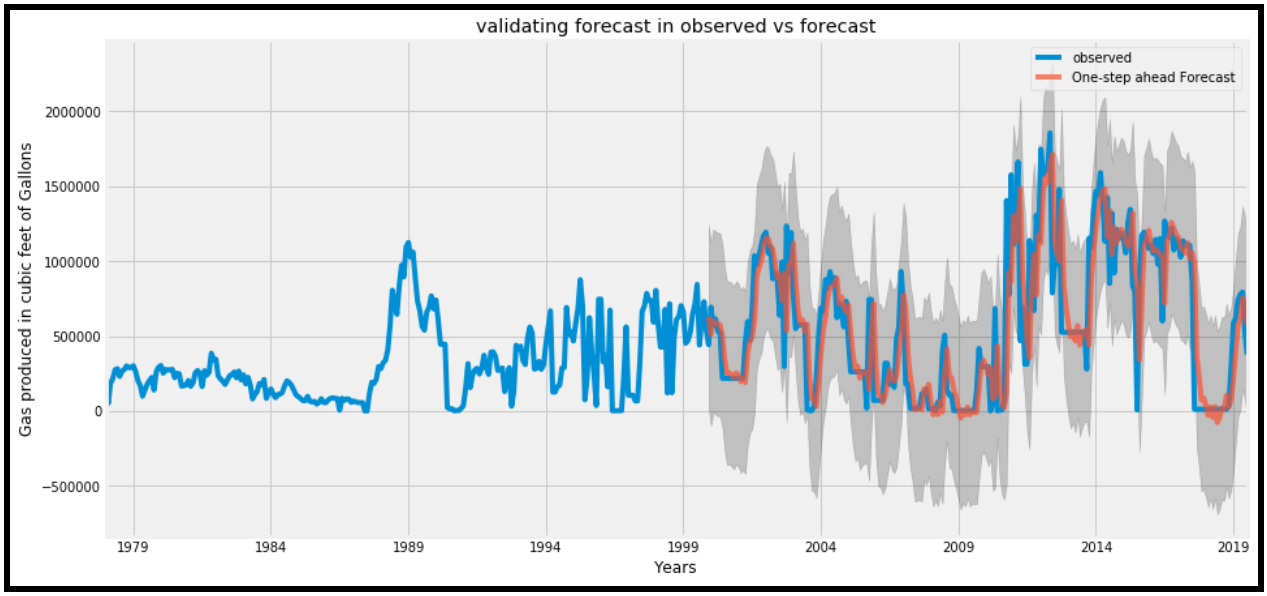
Gas production:

The above plot tells about the Gas production that happened in the Prudhoe bae oil well unit 01-10 from 1979 to 2019.
Just by looking at the plot we got to the wrong conclusion that there is no seasonality observed in series, but for statistical checking, we used decomposition charts to see the trend, seasonality and residual analysis.



But as always statistical checking helps us to understand the data more than the naked eyes. After looking at the decomposition charts above we see the seasonality, random increasing trend and increasing residuals.

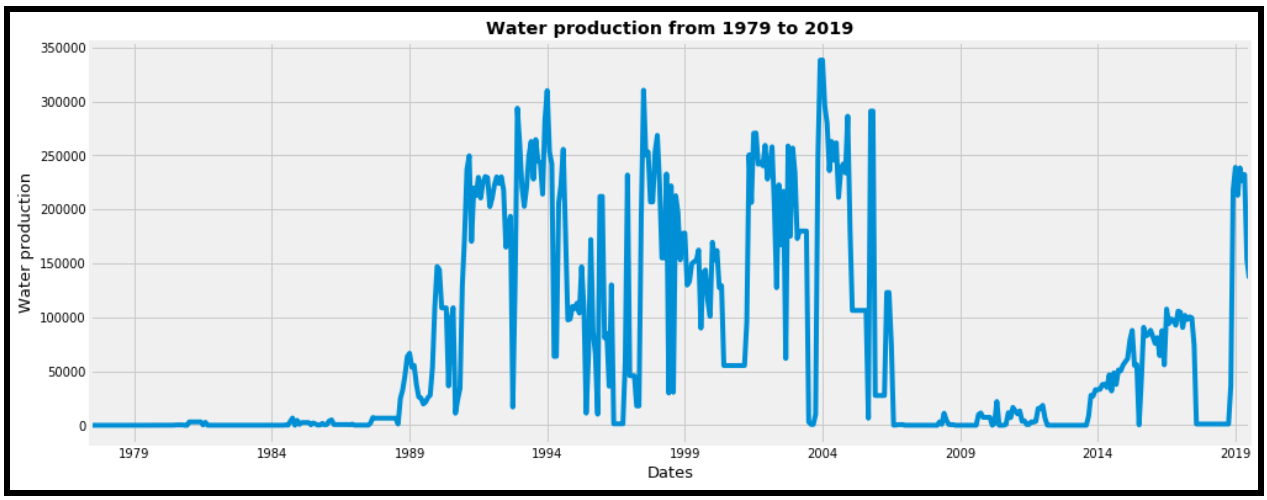
After checking all the models, we employed the ARIMA model and checked all the **AR** and **MA** values along with differencing and seasonality to get the p,d,q to fit in **ARIMA** and seasonal order to make the series stationary.



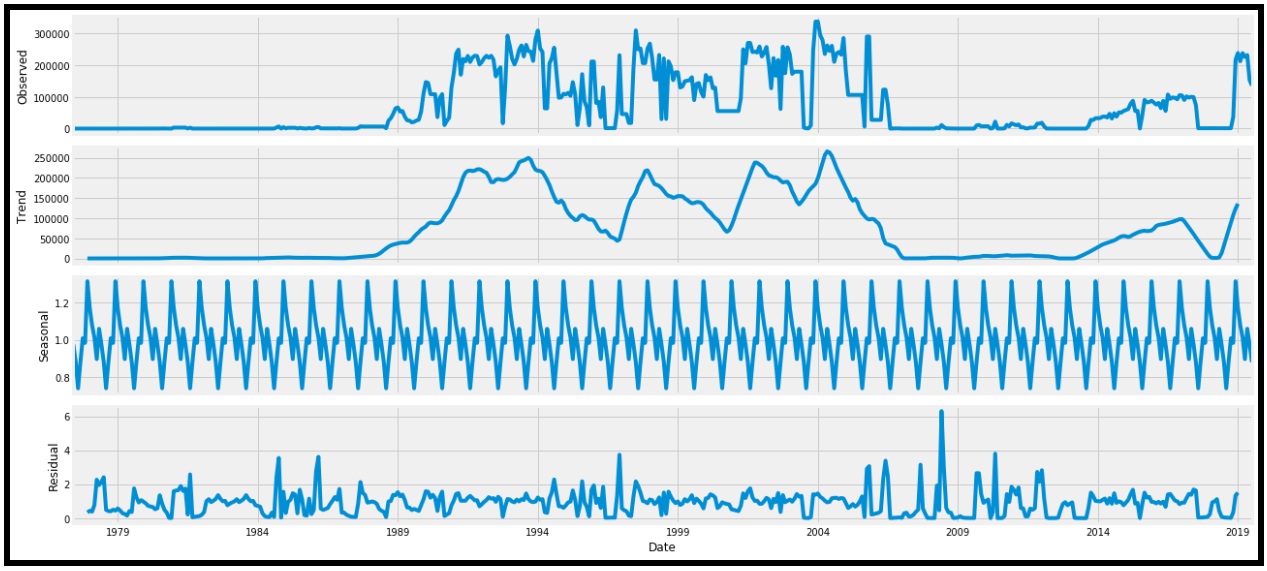
We see from the graph that both the actual observed values of gas production and the predicted values form the model. We got our model accuracy with 79.87%.

Water production:

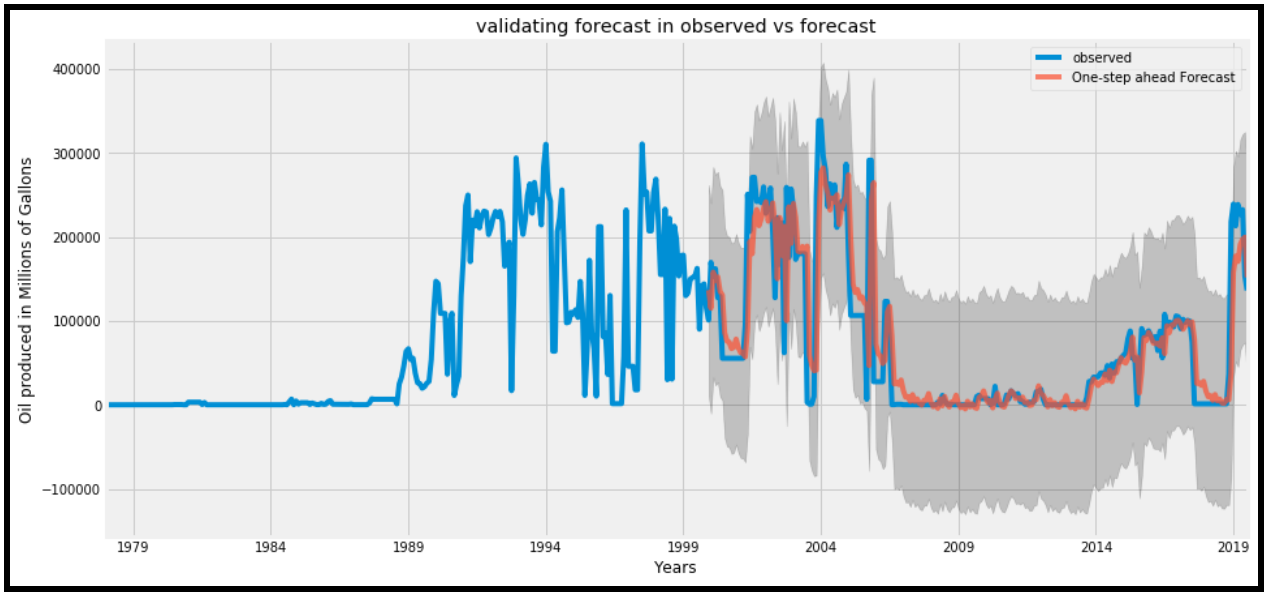
Below graph tells us about the water production that happened in year 1979 to 2019 along with oil and gas in Prudhoe bae unit 1-10. Just by looking at the plot we see that there is no seasonality and trend in it.



Let's use the decomposition chart and see the trend, seasonality and residual analysis

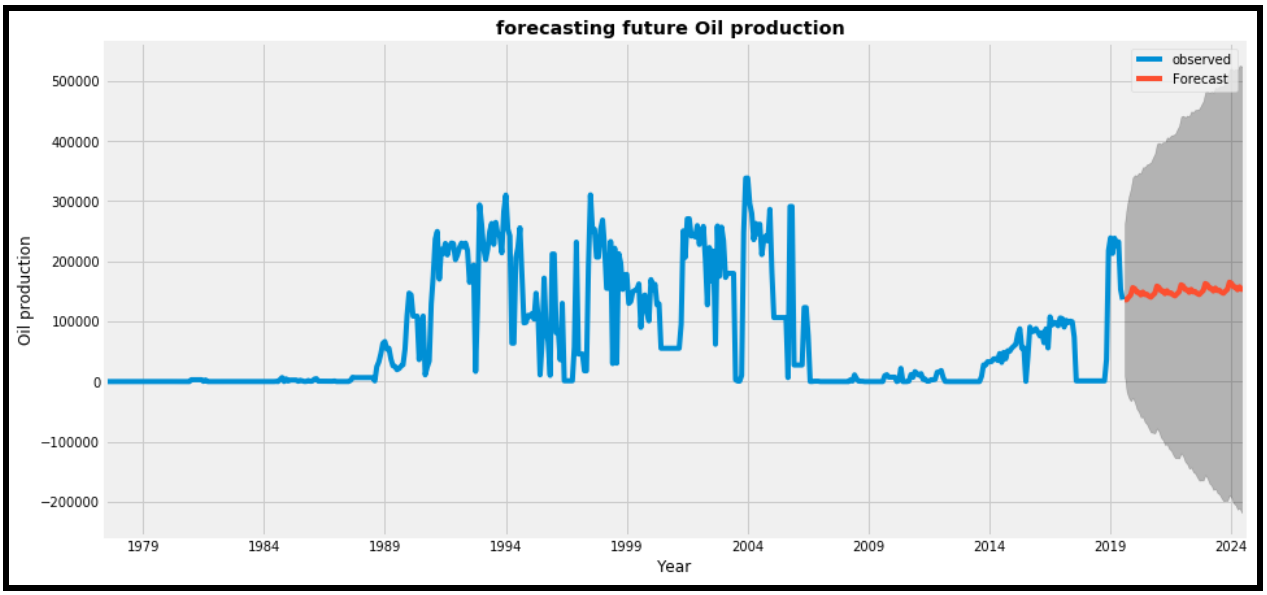


in the series of water production. From the plot, we see that data had nice seasonality and a random trend, which is not a good sign. We used the **ARIMA** model with right set of parameters after employing correlation and autocorrelation and seasonality with p,d,q values.



We got only **67.43 %** accuracy on this model as we see the random trend in data which confused our model to predict better.

After forecasting for the coming next 50 months, this is how the forecasted values look like.



The forecasting may not look best, but it's decent. This happened here because there is too much randomness in the data, values which are close to zero's making model get confused. Values near to zero are very random and have no seasonality.

Further improvements:-

After using the best models analysing the data using all the statistical measures we are able to predict decent observations. Further, the accuracy of the model is improved with the series converted to stationarity and by employing a neural network model like LSTM which are good at the time series.