**Steps:**

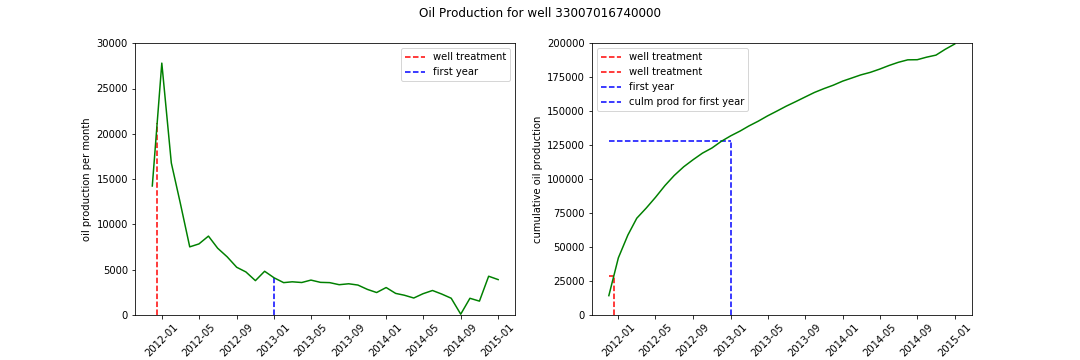
1. Ask the following questions and look for the answers using code and plots:
   1. Can you count something interesting?
   2. Can you find trends (e.g. high, low, increasing, decreasing, anomalies)?
   3. Can you make a bar plot or a histogram?
   4. Can you compare two related quantities?
   5. Can you make a scatterplot?
   6. Can you make a time-series plot?
2. Looking at the plots, what are some insights you can make? Do you see any correlations? Is there a hypothesis you’d like to investigate further? What other questions do the insights lead you to ask?
3. Now that you’ve asked questions, hopefully you’ve found some interesting insights. Is there a narrative or a way of presenting the insights using text and plots that tells a compelling story? What are some other trends/relationships you think will make the story more complete?

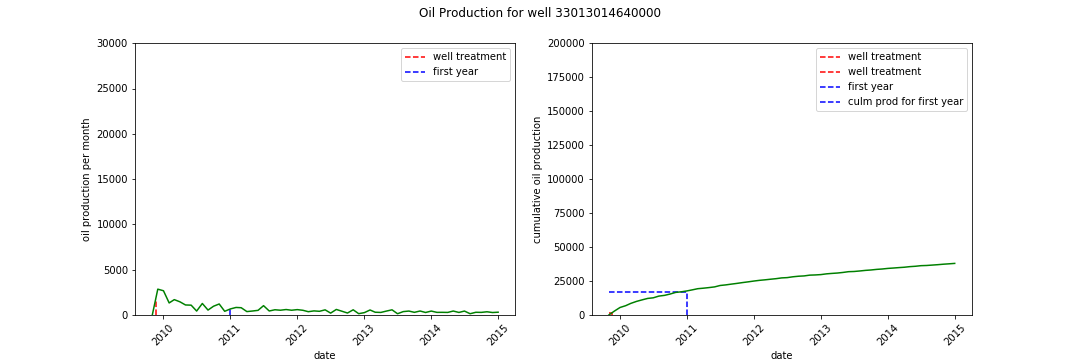
**Submission:** Submit links to a GitHub repository containing a Jupyter Notebook. The Notebook should contain:

* The questions you asked
* The trends you investigated
* The resulting visualizations and conclusions

You will be evaluated using this [rubric.](https://docs.google.com/document/d/1SPlg6_G7bnnbuEzpIK3Tc9pPMWVIudX4im4nLtkh7fQ/edit?usp=sharing)  
These results will go into your final portfolio and presentation. Organize your work as you go along to make it easier to compile later. Create slides and/or a presentation (.ppt) about your emerging data story.     
   
In case the dataset is too large to commit to GitHub, please include a link to the dataset inside the Jupyter Notebook.  
  
Discuss these results with your mentor at the next call. if you’re having trouble with your code for this unit, you can reach out to your course TA for help by emailing projects@springboard.com, or post questions in the community forum

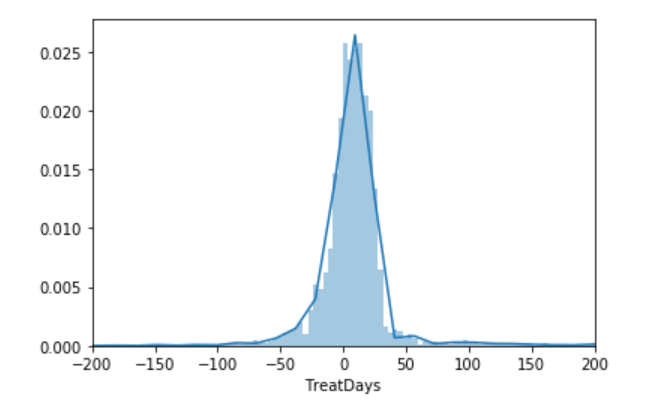
First, I need to visualize the oil production for different wells to get more familiar with the quantity I am trying to predict. The raw data is given as oil production for each month. I am simplifying the problem a little by predicting cumulative production for the first year. It is a good metric for overall well performance and can be used to separate good wells from bad ones.

The following plots shows the production profile for a good performing well. The left plot shows monthly oil production and the right plot shows the cumulative production over the same period. The blue line indicates the time stamp for the first year of production. The red line is the month of well treatment. 

Similar plot is created for a low performing well. 

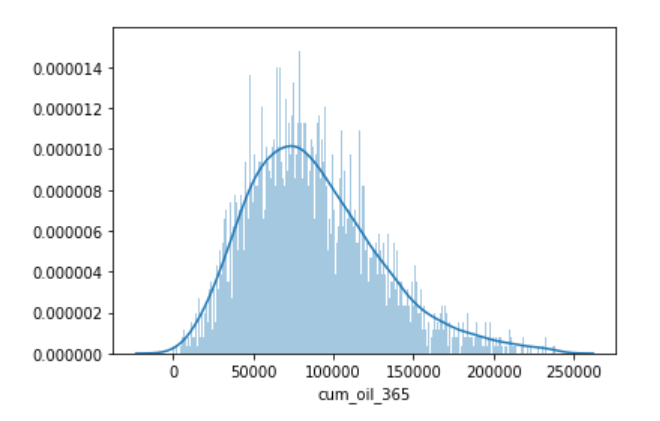
One interesting question to investigate based on these plots is theater or not well performance depends on how fast the well treatment was performed after the well was brough on production. The intuition tells us that the sooner well treatment is performed the better well production should be. I expect a positive correlation between number of days between first well production and well treatment and cumulation production for the first year.

First let’s plot a histogram of number of days between first well production and well treatment.

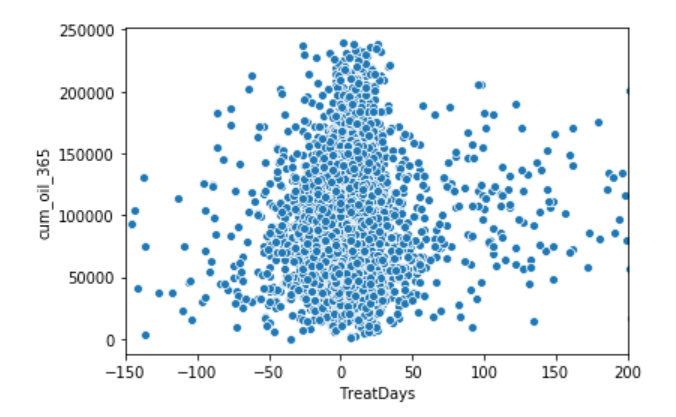


We can see that most wells have treatment job withing the first two months of production onset. It is interesting that almost half of the wells had treatment jobs preceding the production (negative values on the histogram).

The histogram of cumulative oil for the first year of production is well behaved distribution skewed towards the right-hand side.



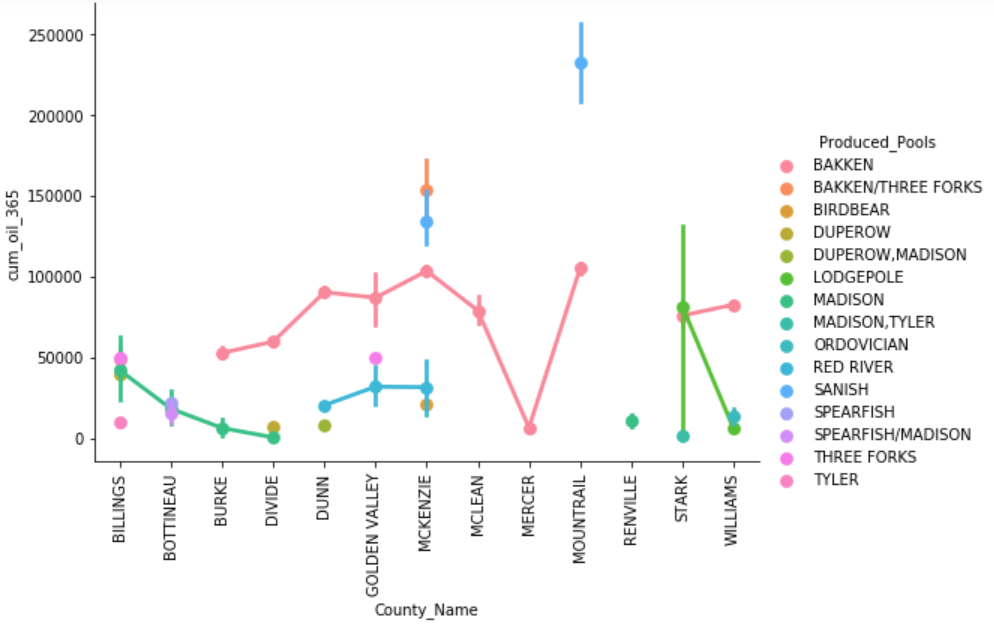
The cross plot of the two variables shows that they are not strongly correlated. First year production of the well does not strongly depend on merely when the well treatment was performed. I believe the details of the treatment play an important role here.



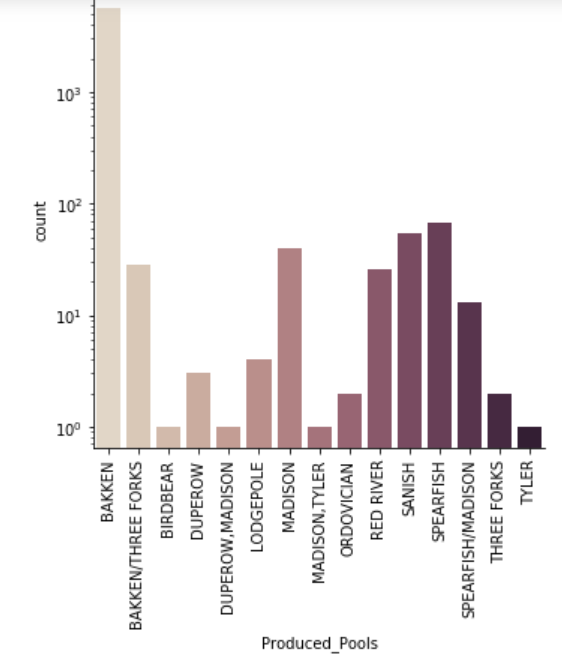
This plot shows the geographical position of all the wells, colored by the county name. Bigger circles indicate bigger cumulative production for the first year. This picture does not discriminate production interval.



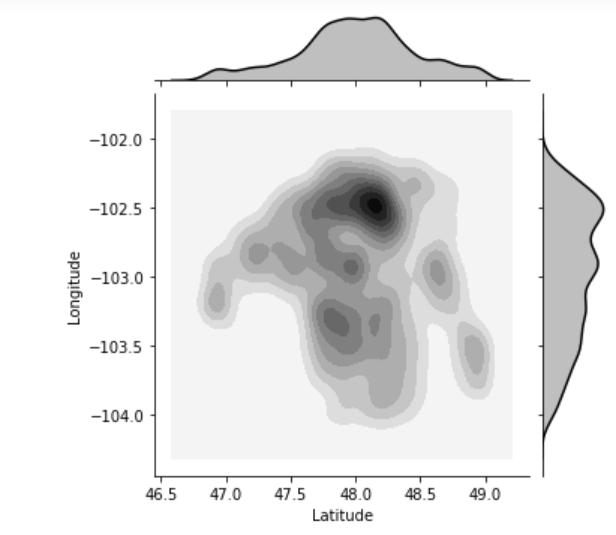
From the figure below we can see that producing intervals may play important role for production prediction. Some zones have shown to have better production than others.



Next plot shows that most of the wells are producing from the BAKKEN interval.



BAKKEN production is highest in the north.



Wellbore direction is also important for increasing production. Vertical wells traverses the producing interval perpendicular and are the least productive. Horizontal wells are drilled parallel to the strata boundaries and are most productive as they intersect more of the productive interval.

