# North Dakota Oil Production Dashboard

While there has been plenty of focus on the analytics side of data in the past century, the analyses can be meaningless if they are not communicated properly to the appropriate parties. It is important to ensure that data is presented in a clear, clean, concise and captivating manner. Tableau software has made this endeavor more accessible, therefore I used this software to create an analytics dashboard based on North Dakota oil production data.

It is important to note that I created two dashboards, one a purely exploratory analysis (EA) dashboard which I used to understand the data at an aggregated level, and one a county analysis (CA) dashboard which I used to glean insights on oil production at the county level. Please note that the dashboards and graphs are annotated by “EA” and “CA” for easy reference. In the EA dashboard, I decided to show oil production in the 21st century – that is, years 2000 to 2017, since this is most relevant to the users in today’s time. I noticed four counties producing the most oil (>300M oil bbls) during this time, therefore I decided to select these counties for a deeper analysis in the CA dashboard. The CA dashboard will be discussed in depth in this report, as this is the dashboard to be used by the end user.

# The Audience

The targeted audience will be state government regulators and environmentalists who are concerned with oil quality, and the effect of mass oil production on the environment. The environmentalists are assumed to be working with state regulators to establish policies for oil manufacturers who have produced over 300M oil bbls since 2000. State regulators are also concerned with the monopoly of certain oil manufacturers in a state, therefore the dashboard will also be useful in targeting aggressive oil producers at the operator level.

# The Objective

The objective of the dashboard is to alert state government regulators and environmentalists with outliers of the wells and operators (in terms of oil production), the correlation (or lack thereof) between oil and other measures, irregular wells used for drilling, etc. This dashboard serves as the first point of investigation to detect irregularities and unexpected production patterns. The users can then dive into the CA graphs to understand the factors that could be driving the irregularities. While this data is aggregated for the 21st century, regulators and environmentalists can use this dashboard at a monthly frequency to detect any irregularities going forward. This way, the dashboard will prove to be actionable in a more timely manner.

# The Content

For clarity and ease of use, I included four graphs in the dashboard. Please see below for further detail on each graph:

1. **Oil vs Water Production**

During the research phase of the project, we discovered that water extraction is a critical phase of the production of oil. Additionally, concentration of water in oil provides a measure of oil purity. State regulators, therefore, would be interested in seeing the relationship between the amount of water and oil generated by the top oil producing counties in order to assess for oil quality.

* After creating this graph, I noticed that in county 105, there was a considerable difference between water and oil production in 2015. Normally, the amount of water and oil produced is almost equal, however this was not the case for county 105. The amount of water produced was over 50% greater than the amount of oil produced. This could indicate that more water was extracted from the oil than usual and oil purity could have been compromised.

Including this graph in the dashboard, and using it on a monthly basis, would help regulators catch these irregularities sooner and encourage state regulators to respond immediately to the situation.

1. **Well Status/Operator Analysis**

We also discovered that wells have the following condition codes: Active, Inactive, LOC, NC, PA, PNC and Confidential. Please see below for interpretation of each code:

A – Active

IA – Inactive (Shut-In >= 3 and <= 12 months)

LOC – Permitted Location to Drill

NC – Not Completed (Drilled to Total Depth, Awaiting Completion)

PA – Plugged and Abandoned

PNC – Permit Now Cancelled

Confidential – Tight hole

The majority of the wells in the four counties producing the most oil is classified as active, however I wanted to take a further look at the dispersion of wells not classified as active. I performed this analysis on operator level, since this would indicate who is operating on the irregular wells.

* The two most frequent well status codes (excluding active) are inactive (IA) and plugged and abandoned (PA). I found that seven well operators are operating on these wells, with Hess Bakken operating heavily on PA wells when compared to other operators.

State regulators should see the data in this graph as a red flag, and should question why these companies are operating on IA and PA wells. Drilling IA could have negative implications on oil quality and could suggest business wrongdoing by these operators.

1. **Well Identifier Analysis**

This is an oil production frequency graph at the county and well identifier level. It is designed to indicate to the environmentalists which wells are producing more oil compared to others, and I decided to show only wells that have produced over 2M bbls of oil in order to reduce the data included in this graph.

* We can see that counties 25 and 61 have more wells that have produced over 2M bbls of oil when compared to the other counties. We can easily see this by comparing the number of bubbles that comprise these counties. Additionally, we can assess the relative oil production of each well by comparing the size of each bubble to its cluster. Well 02682 has produced considerably more oil when compared to all wells represented in this graph.

Environmentalists should understand the implications of mass oil production on the environment, specifically for these counties. They should evaluate if the county has additional, more strict environmental regulatory policy in response to the mass oil production. If not, they should work with regulators to implement such policy.

1. **Exponential Oil Production**

During the exploratory analysis phase, I noticed that oil production grew exponentially from 2011 to 2015. I created a time series graph to further understand which well operators (from the four top oil producing counties) were primarily responsible for this exponential growth.

* We can see that Whiting Oil and Gas Corporation took the lead on growth, while Hess Bakken came second. The remaining operators grew at a more steady pace, which is representative of oil production growth at the aggregated level from 2011 to 2015.

This graph would be insightful and actionable for both environmentalists and state regulators. Environmentalists should understand the sustainability precautions taken by these operators during time of expansion. State regulators should ensure that anti-monopoly laws exist to prevent any one operator from having excessive influence in the oil production market.

# The Aggregation

As mentioned earlier, I performed an explanatory analysis from the entire aggregated dataset to learn interesting and meaningful patterns. I decided to choose the four counties producing over 300M bbls of oil for further analysis, during the timeframe of 2011 to 2017, since this would have important implications for the end users – state regulators and environmentalists. I then chose to aggregate further by well identifier and well operator. I learned that it is important to first obtain a wholesome overview of the dataset in order to create a meaningful and actionable dashboard.