Geosteering

Harold Vance Department of
PETROLEUM ENGINEERING
TEXAS A&M UNIVERSITY

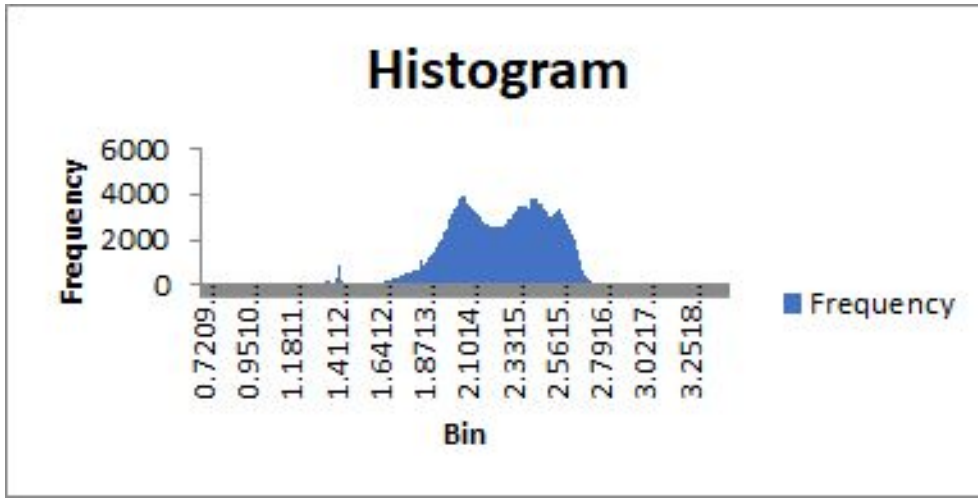
Thu Nguyen, Ty Oakes, Nicholas Heinrich-Barna

Agenda

- Exploratory Analysis
- Clustering
- Classification
- Results

Exploratory Analysis

- Variables grouped together must be separated.
- Many distributions were very close to Normal.
- Some inputs cannot be used.



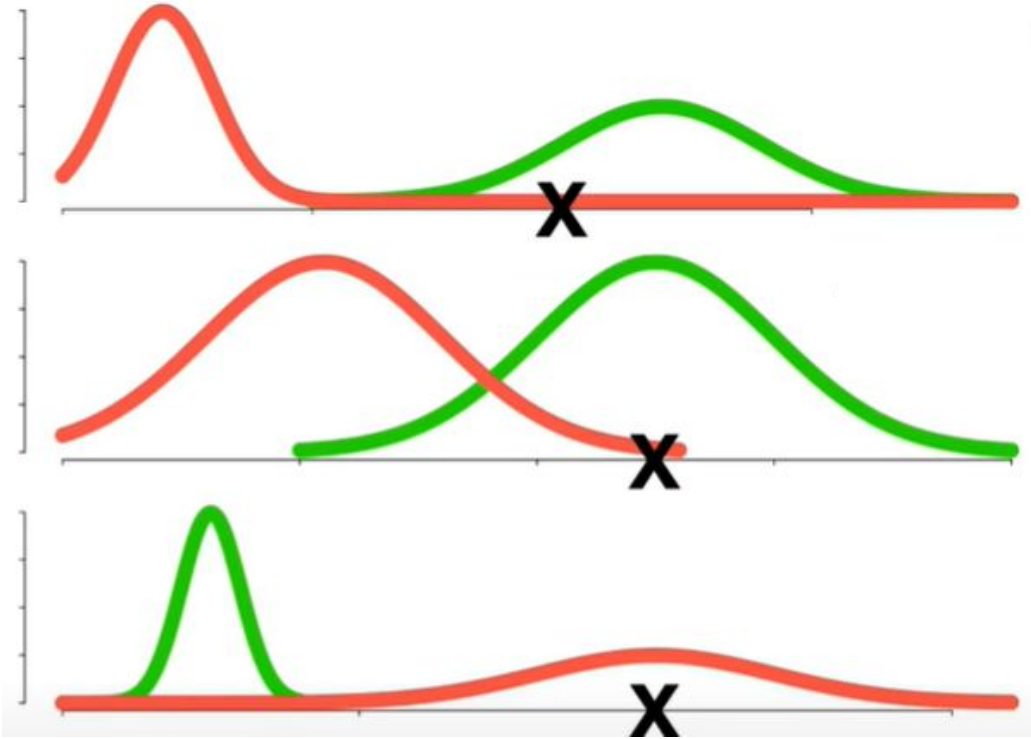
Exploratory Analysis

-

		Tuff		Basement		Coal		Halite		Anhydrite		Marl		Dolomite		Chalk		LimeStone		Sandstone		Shale		Sandstone/Shale	
		Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD
CALI	?	15.2789904	3.04191	12.558	0.17173	11.9089	4.89185	12.4977	0.41897	12.5217	1.44845	12.5666	2.65521	13.8349	3.62751	13.3934	2.27812	11.9615	2.53178	12.6951	4.57863	13.6288	3.64235	12.8848	4.09666
RMED	USE	1.11031167	0.46807	8.99474	4.22617	37.3071	141.848	13.0343	11.9499	304.29	453.012	2.70029	3.32497	7.32287	53.203	2.49168	2.08635	6.31563	18.2156	5.63502	31.5919	4.6028	67.6833	3.79647	8.85108
RDEP	USE	1.05685568	0.44322	39.2988	90.3256	27.7781	137.315	1690.09	518.136	370.199	555.936	2.56482	5.16954	10.8483	65.5564	2.33452	1.98686	5.04923	20.8374	10.328	86.9517	2.80039	33.3842	3.71782	16.7987
RHOB	USE	2.13512326	0.14122	2.89415	0.06584	1.88315	0.36892	2.0756	0.05384	2.83274	0.1927	2.45212	0.14032	2.40966	0.2254	2.3785	0.16131	2.50443	0.14849	2.22415	0.24622	2.23837	0.24641	2.32029	0.27086
GR	USE	54.6517675	21.1676	44.9395	21.1036	64.3212	35.9888	9.39926	4.92334	17.3736	12.6069	56.38	21.2378	54.9553	25.0453	22.0109	14.0598	29.78	21.4789	47.5581	24.3082	81.6229	33.7577	64.0667	24.5805
NPHI	USE	0.43347069	0.09167	0.15297	0.04558	0.43044	0.12322	0.04644	0.03051	0.01877	0.05574	0.27948	0.08958	0.28771	0.12794	0.19148	0.0752	0.17969	0.09896	0.24799	0.10801	0.39218	0.11402	0.29104	0.10562
DTC	USE	127.973233	16.9259	47.6378	2.93125	108.903	23.787	68.5152	1.19219	54.8065	7.16795	95.3471	18.0791	100.479	30.2734	79.5565	10.0354	75.2382	17.8588	106.948	30.0097	123.13	27.4642	105.981	27.8121
SP	USE	50.2548727	53.2205	101.67	3.43235	56.8346	104.353	127.01	3.17725	95.9111	18.0885	70.6205	65.3471	72.703	39.136	61.9327	32.8672	85.6996	45.3407	57.7639	75.7764	60.4583	80.211	57.0224	62.6757
BS	?	13.7214196	2.64744	12.25	8.9E-15	9.61244	2.67561	12.2324	0.25613	11.9152	1.06933	11.6372	2.41952	11.6403	3.15458	11.1668	1.74232	11.3306	2.2402	11.8302	3.75455	12.1538	3.23431	11.1015	3.55701
ROP	Use	18.2837166	19.2539	46.052	8.38385	11.2851	12.782	23.9068	22.6993	25.3423	9.43394	235.021	1779.93	20.2635	31.468	38.9781	58.3721	63.648	793.393	114.247	1217.52	187.892	1911.73	104.692	1037.29
DCAL	USE	1.20609264	1.13003	766.133	2641.36	2.40608	3.37108	2.63447	0.89658	0.81217	0.42416	1.05103	1.07633	1.48313	2.10294	1.33415	1.87588	0.63048	0.98469	0.60699	48.6678	1.15896	1.74794	0.63961	1.91942
DRHO	USE	0.00093267	0.02653	-0.00106	0.0149	-0.00681	0.05612	0.00374	0.00906	0.00058	0.05179	0.02122	0.10011	0.01956	0.06802	0.00115	0.07893	0.0059	0.05103	0.00639	0.14184	-0.00346	10.2739	0.01914	0.28026

Gaussian Naive Bayes

- All distribution Normal.
- Weight the likelihood for the input to each distribution.
- Take the most likely summation



Assumptions

- Assume Gaussian Distribution.
- Assume all variables are to be weighted equally.
- Assume outliers are not effective on results

Process

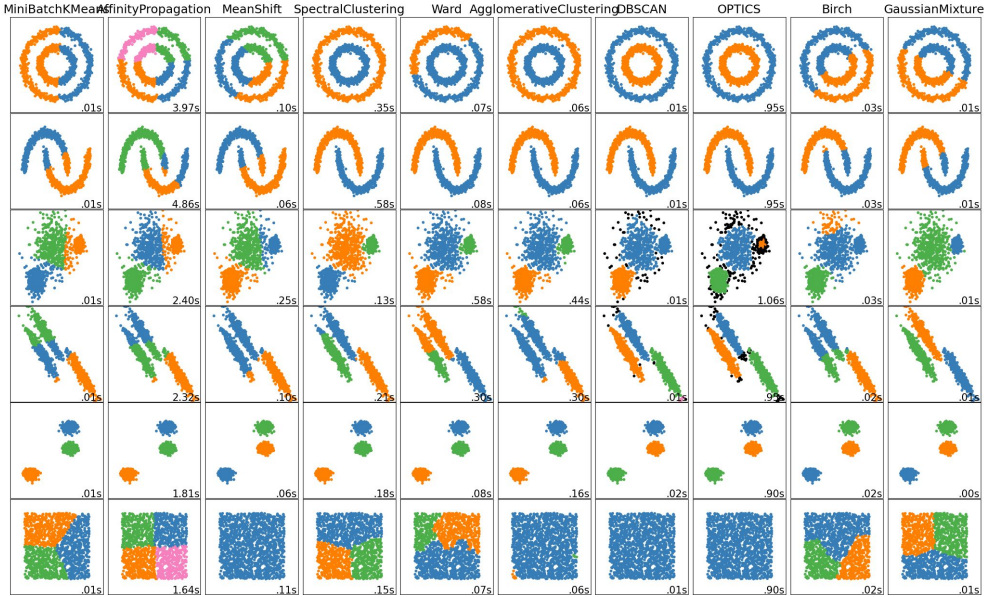
- Separate data into its lithofacies types.
- Find the gaussian distribution of these stats.
- Import the data into python to be iterated over.
- Check and clean the data as it comes in.

Results

- Certain lithofacies were predicted better than others
- The overall accuracy over the test cases was around 40%
- Most of the error was accrued in sandstone

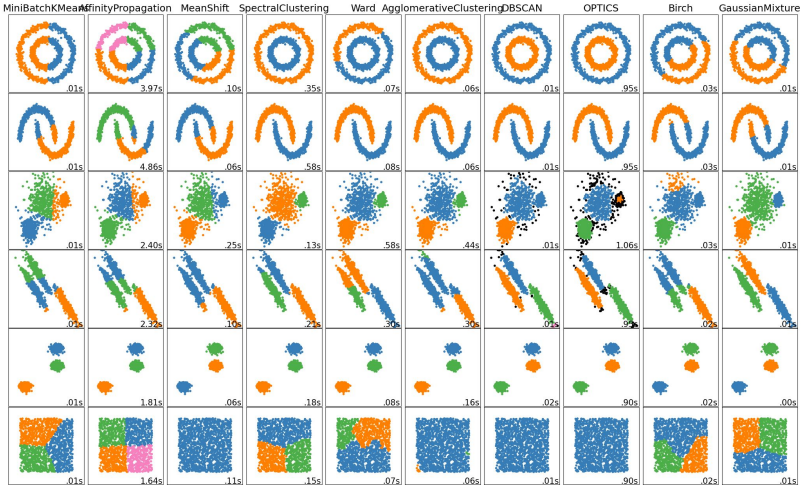
Clustering

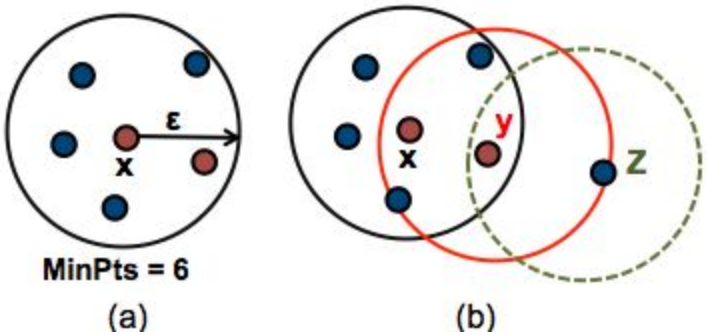
KMeans - DBSCAN



Clustering

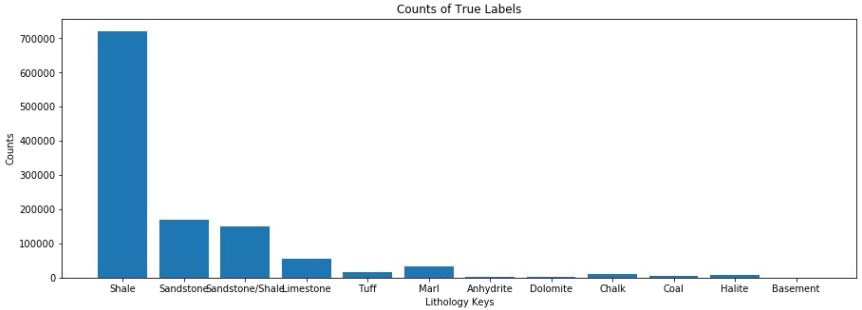
- KMeans is a centroid based while DBSCAN is density based
- Have to specify number of clusters (centroids) in Kmeans
- DBSCAN can identify clusters of varying shapes

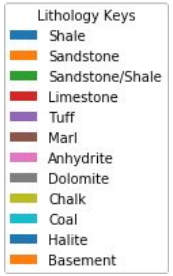


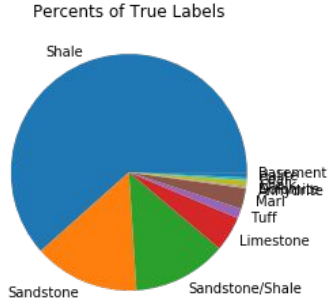


Clustering - Data Processing

- Drop the label column and any column that has more than 75% missing data
- Transform categorical to numerical data
- Linear Interpolation to fill in missing values

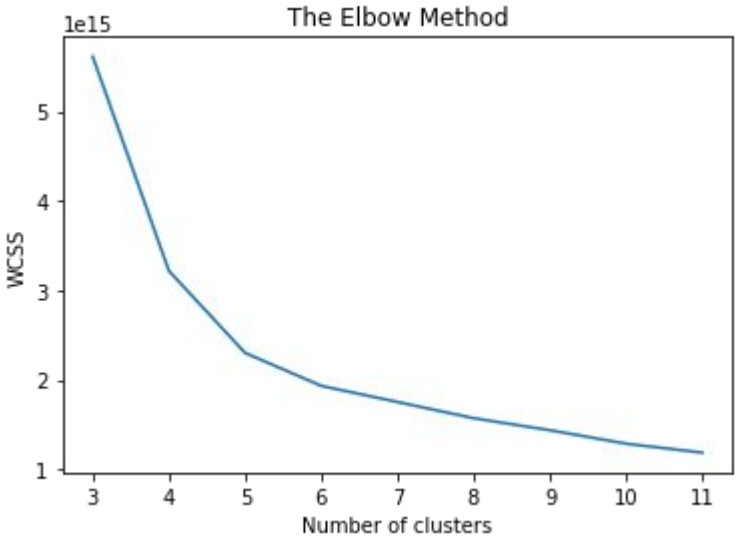


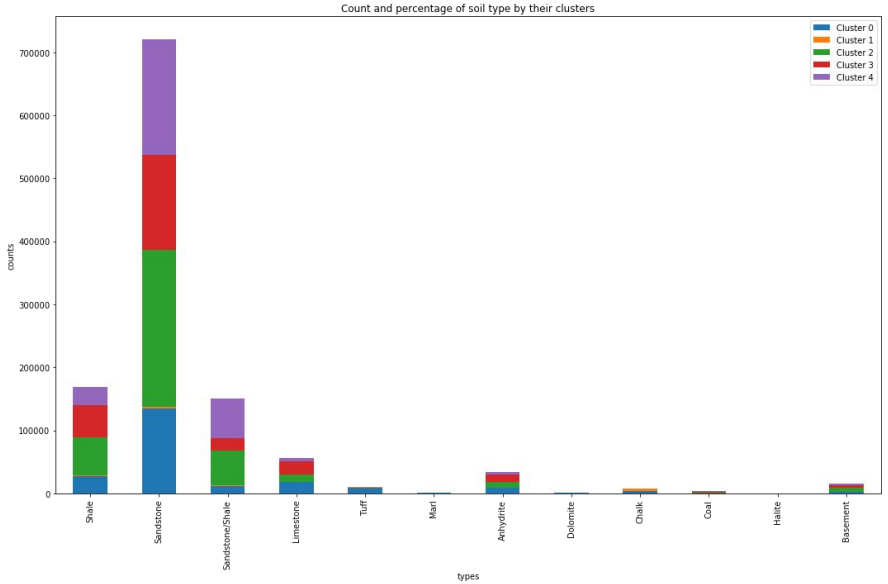




KMeans

- The elbow method finds k = 5, fit the model, traced back the labels



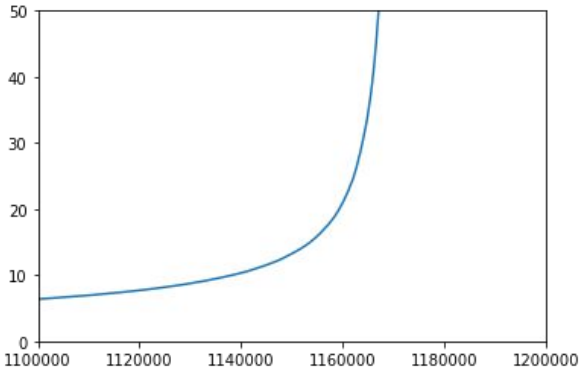


KMeans

lithology key	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Sandstone	27357	1013.0	60599.0	50905.0	29063.0
Shale	134528	2360.0	249269.0	150416.0	184230.0
Sandstone/Shale	12299	1002.0	54229.0	19930.0	62995.0
Limestone	18435	12.0	11621.0	20383.0	5869.0
Chalk	9258	1143.0	0.0	112.0	0.0
Dolomite	608	6.0	318.0	313.0	443.0
Marl	8657	663.0	8923.0	12097.0	2989.0
Anhydrite	798	282.0	0.0	0.0	5.0
Halite	3919	4294.0	0.0	0.0	0.0
Coal	58	0.0	1228.0	1849.0	685.0
Basement	103	0.0	0.0	0.0	0.0
Tuff	2857	0.0	6624.0	3182.0	2582.0

DBSCAN

- *eps*



lithology key	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Sandstone	0.0	0.0	0.0	0.0	0.0	0.0	162.0
Sandstone/Shale	0.0	0.0	0.0	0.0	0.0	293.0	0.0
Chalk	207.0	452.0	155.0	0.0	0.0	0.0	0.0
Halite	0.0	0.0	0.0	144.0	3769.0	0.0	0.0