Data preparation

Email and network database

VBA →

Daily reports.xlsx

Extraction, regex parsing, pre-processing

Well activity matrix

Statistical analysis, Monte Carlo

Insight
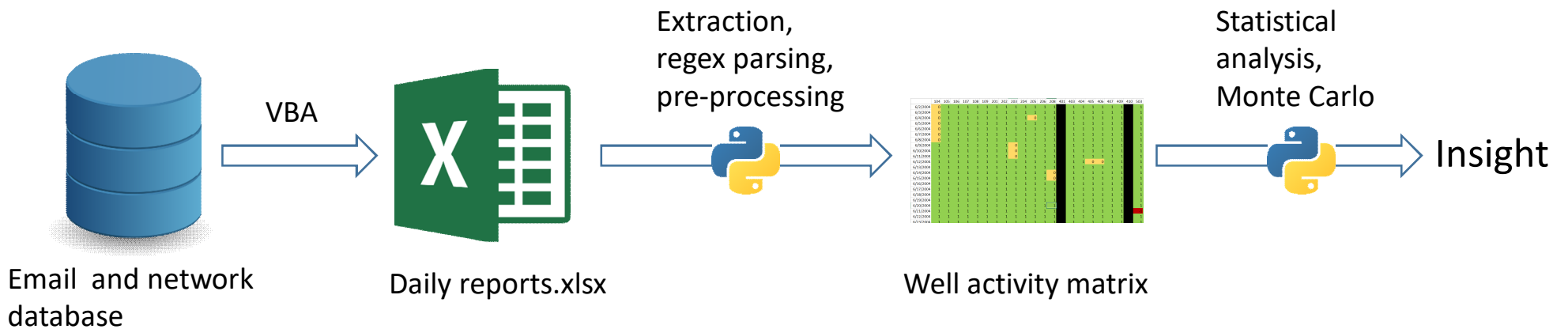
Data preparation

Extraction, regex parsing, pre-processing for every day 2004-2019

Extracted from excel:

```
... None, 401, 'h/t', datetime.datetime(2016, 1, 21, 0, 0), 5, '#907', 'broke belts .',
datetime.datetime(2016, 9, 11, 0, 0))(None, 202, 'h/t', datetime.datetime(2015 ...
```

Cleaning

```
... 606 'not pumping'  708 'r/p'   '1008          h/t                    '
1010 'h/t'        1407 'h/t'       1410 'h/t'       504 'csg leak'      403 'h/t' ...
```

Classification

Well ID	Surface	Subsurface	Running	Unclassified
401	FALSE	TRUE	FALSE	FALSE
907	TRUE	FALSE	FALSE	FALSE
202	FALSE	TRUE	FALSE	FALSE
606	FALSE	TRUE	FALSE	FALSE
708	FALSE	TRUE	FALSE	FALSE
1008	FALSE	TRUE	FALSE	FALSE
1010	FALSE	TRUE	FALSE	FALSE

. . .

Cost-Effective Artificial Lift Strategies. Permian Basin 2020, Houston, TX

Activity Matrix

15 year span   **88 wells**   **~2 MB raw data**   **~500k data points**

Columns:
Well names

Rows:
Date

Legend:
-3 – well not drilled yet
-1 – downhole issue
0 – surface issue
1 – well active

	104	105	106	107	108	109	201	202	203	204	205	206	208	401	403	404	405	406	407	409	410	503
6/2/2004	0	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/3/2004	0	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/4/2004	0	1	1	1	1	1	1	1	1	1	0	1	1		1	1	1	1	1	1		1
6/5/2004	0	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/6/2004	0	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/7/2004	0	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/8/2004	0	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/9/2004	1	1	1	1	1	1	1	1	0	1	1	1	1		1	1	1	1	1	1		1
6/10/2004	1	1	1	1	1	1	1	1	0	1	1	1	1		1	1	1	1	1	1		1
6/11/2004	1	1	1	1	1	1	1	1	0	1	1	1	1		1	1	1	1	1	1		1
6/12/2004	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	0	0	1	1		1
6/13/2004	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/14/2004	1	1	1	1	1	1	1	1	1	1	1	1	0		1	1	1	1	1	1		1
6/15/2004	1	1	1	1	1	1	1	1	1	1	1	1	0		1	1	1	1	1	1		1
6/16/2004	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/17/2004	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/18/2004	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/19/2004	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/20/2004	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/21/2004	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		-1
6/22/2004	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1
6/23/2004	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1		1