Equipment Failure Forecasting using Similar Models, Oct. 2016

loadData

In [369]:

```
import pandas as pd
df = pd.read_csv('data.csv')
```

In [370]:

```
df.tail()
```

Out[370]:

	EQP_LOCAL_EQP	EQP_MODEL_EQP	EQP_FAIL_CNT_EQP
1055	8383600220011289	RNG150N	1
1056	8383650060006956	MCARD9062	1
1057	8383500190010727	DCT6412	1
1058	8383680010643897	MCARD9062	2
1059	8383890170012680	MCARD9060	1

pivotTable

In [376]:

```
modelRatings = df.pivot_table(index=['EQP_LOCAL_EQP'],columns=['EQP_MODEL_EQP'],values='EQP_FAIL_CNT_EQP').
iloc[:, 1:10]
modelRatings.head()
```

Out[376]:

EQP_MODEL_EQP	AHCGN	AHCGN2	AHCGN2250	AHCGNVM	AHCGNVRES	AHD3CM160	AHDC45	AHDCM425	AHDCM476
EQP_LOCAL_EQP									
8383100010093449	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8383100010149076	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8383100010171195	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8383100010191201	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8383210010523475	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

subscriberRatings

In [372]:

```
DCT6412Ratings = modelRatings['DCT6412']
DCT6412Ratings.head()
```

Out[372]:

```
EQP_LOCAL_EQP
8383100010093449   NaN
8383100010149076   NaN
8383100010171195   NaN
8383100010191201   NaN
8383210010523475   NaN
Name: DCT6412, dtype: float64
```

pairwiseCorrelation

In [373]:

```
similarModels = modelRatings.corrwith(DCT6412Ratings)
similarModels = similarModels.dropna()
df = pd.DataFrame(similarModels)
df.head()
```

Out[373]:

	0
EQP_MODEL_EQP	
DCT6412	1.0
MCARD9062	-0.2

similarityScore

In [374]:

```
similarModels.sort_values(ascending=False).head()
```

Out[374]:

```
EQP_MODEL_EQP
DCT6412      1.0
MCARD9062   -0.2
dtype: float64
```

countFailures

In [375]:

```
import pandas as pd
df = pd.read_csv('data.csv')

modelStats = df.groupby('EQP_MODEL_EQP').agg({'EQP_FAIL_CNT_EQP': [np.size, np.mean]})
modelStats.head()
```

Out[375]:

	EQP_FAIL_CNT_EQP	
	size	**mean**
EQP_MODEL_EQP		
8600	1	1.00000
AHCGN	32	1.03125
AHCGN2	9	1.00000
AHCGN2250	12	1.00000
AHCGNVM	19	1.00000

writeUp

Forecast equipment failure ranked by pairwise correlation using criteria such as model number, subscriber number, and failure count.

For example, DCT6412 was the baseline model used to predict failure and MCARD9062 has a similar chance of failure as it relates to DCT6412.