

This program cleans up dataframes and splits a column by using pivot table function

```
In [85]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [86]: %matplotlib inline
```

```
In [87]: df=pd.read_csv("C:\\Users\\Shay\\Documents\\Yari\\Data analysis Python\\Data\\CCB_CI_MCW.csv")
```

```
In [88]: df = df[['HDDSN', 'PROCID', 'TESTCODEC', 'PFCODE', 'DataTP', 'QUALIFIER', 'MFGID', 'HDDTRIAL', 'LHD', 'PHD', 'BAND', 'MCW']]
```

```
In [89]: df['MCW_nm'] = (df['MCW']/100)*(1000/580)*(25.4)*(df['DataTP']/1024.0)
```

```
In [90]: def RADIUS (row):
    if row['BAND'] == 0 :
        return 'OD'
    if row['BAND'] == 1:
        return 'MD'
    if row['BAND'] == 2 :
        return 'ID'
```

```
In [91]: df['RADIUS'] = df.apply (lambda row: RADIUS(row), axis=1)
```

```
In [92]: df
```

```
Out[92]:
```

	HDDSN	PROCID	TESTCODEC	PFCODE	DataTP	QUALIFIER	MFGID	HDDTRIAL	LHD	PHD	BAND	
0	2FA0491A	6400	PDQC3CX	2525	1188	10N0	KQBT03	AQ0N	0	0	0	1
1	2FA0491A	6400	PDQC3CX	2525	1188	10N0	KQBT03	AQ0N	0	0	1	1
2	2FA0491A	6400	PDQC3CX	2525	1188	10N0	KQBT03	AQ0N	0	0	2	1
3	2FA0491A	6400	PDQC3CX	2525	1188	10N0	KQBT03	AQ0N	1	1	0	1
4	2FA0491A	6400	PDQC3CX	2525	1188	10N0	KQBT03	AQ0N	1	1	1	1
...	
955	2FA049GA	6400	PDQC3CX	2521	986	2300	KQBT03	AQ0N	7	17	2	
956	2FA049GA	6400	PDQC3CX	2521	987	2300	KQBT03	AQ0N	7	17	3	
957	2FA049GA	6400	PDQC3CX	2521	986	2300	KQBT03	AQ0N	7	17	4	
958	2FA049GA	6400	PDQC3CX	2521	986	2300	KQBT03	AQ0N	7	17	5	
959	2FA049GA	6400	PDQC3CX	2521	987	2300	KQBT03	AQ0N	7	17	6	

960 rows × 14 columns

```
In [93]: #Remove rows where 0wValPerp contains 0.0
```

```
In [94]: df= df[df['MCW'] != 0]
```

```
In [95]: #df2
```

```
In [97]: #Here we pivot the df to spread QUALIFIER horizontally as new columns
```

```
In [98]: #In the below cell we are generalizing our index, Columns and Values
```

```
In [99]: indexcols = ['HDDSN', 'PROCID', 'TESTCODEC', 'PFCODE', 'MFGID', 'HDDTRIAL', 'LHD', 'PHD',  
splitcols = ['MCW']  
splitbycols = ['QUALIFIER']  
  
df = pd.pivot(df, index=indexcols, columns='QUALIFIER', values=splitcols)
```

```
In [100... #df2
```

```
In [101... df.to_csv('C:\\Users\\Shay\\Documents\\Yari\\Data analysis Python\\Data\\CCB_CI_MCW_Test(2-
```

```
In [102... #We can also split by multiple columns as follows
```

```
In [103... #indexcols = ['HDDSN', 'PROCID', 'TESTCODEC', 'PFCODE', 'MFGID', 'HDDTRIAL', 'LHD', 'PHD',  
#splitcols = ['MCW']  
#splitbycols = ['QUALIFIER', 'RADIUS']#  
#  
#df = pd.read_csv('MCW_Preprocess.csv')  
#print (df)  
  
#df2 = pd.pivot_table(df, index=indexcols, columns=splitbycols, values=splitcols )  
#df2.columns = list(map("_".join,df2.columns))  
#print (df2)
```

```
In [104... #df.columns
```