Useful Pandas snippets

# Pivot Dataframe

df\_pivot = df.pivot\_table(index=index\_columns\_list,  
 columns='pivot\_column\_str',  
 values='pivot\_ values \_column\_str',  
 aggfunc=lambda x: ''.join(x)).reset\_index()

# Melt Dataframe

df\_melt = pd.melt(df,

id\_vars=index\_columns,  
 value\_vars=columns\_to\_melt\_list,   
 var\_name=identifier\_col\_name\_str,  
 value\_name=value\_col\_name\_str)

# Group Table

df\_grouped = df.groupby(by=group\_col\_list).agg({col\_name: 'count'}).reset\_index()

# Rename Columns

df.rename(columns={'old\_col\_name': 'new\_col\_name'}, inplace=True)

# Column datatypes

Dateime

df['Date'] = pd.to\_datetime(df['Date'], format='%Y%m%d')

# Merge

df\_merged = pd.merge(left=df\_left, right=df\_right, how='left', on=column\_list)

# Inverse Dictionary

dict\_inv = dict((v, k) for k, v in dict\_1.items())

# Month End

1. month\_end = calendar.monthrange(year, month)[1]

# **Imputation Tricks**

* Remove columns that contain nulls

columns\_na = [col for col **in** df.columns if df[col].isnull().any()]

reduced\_df = df.drop(columns\_na, axis=1)

* Simple imputation

for col **in** columns\_na:

df[col + '\_was\_missing'] = df[col].isnull()

from sklearn.impute import SimpleImputer

my\_imputer = SimpleImputer()

imputed\_df = pd.DataFrame(my\_imputer.fit\_transform(df))

imputed\_df.columns = df.columns

**Categorical Variables**

*# Get list of categorical variables*

s = (X\_train.dtypes == 'object')

object\_cols = list(s[s].index)

# Remove Categorical variables

X\_train.select\_dtypes(exclude=['object'])

