Preprocessed Transformation

Creates OpenSea NFT dataframe from raw Json data and then creates reference dataframes
for CryptoCompare and EtherScan in order to perform data enrichment in the next stage

Input:

- Data Content: OpenSea NFT Data
- Data Type: JSON
- Data Source: Raw Layer

Output:

- Data Content:
  - NFT Data
  - CryptoCompare Reference Data
  - EtherScan Reference Data
- Data Type: Parquet
- Data Destination: Preprocessed Layer

```
import datetime
import logging
```

```
#mount blob container in order to be accessible by databricks cluster
if not any(mount.mountPoint == '/mnt/capstoneblob/' for mount in dbutils.fs.mounts()):
  try:
    dbutils.fs.mount(
    source = "wasbs://{}@{}.blob.core.windows.net".format(ContainerName, storageAccountName),
    mount_point = "/mnt/capstoneblob",
    extra_configs = {'fs.azure.account.key.' + storageAccountName + '.blob.core.windows.net': storageAccountAccessKey}
  )
  except Exception as e:
    print("already mounted. Try to unmount first")

display(dbutils.fs.ls('mnt/capstoneblob'))
```

```
already mounted. Try to unmount first
```

	path	name	size
1	dbfs:/mnt/capstoneblob/Cloud_Deployment/	Cloud_Deployment/	0
2	dbfs:/mnt/capstoneblob/Data/	Data/	0

Showing all 2 rows.

```
# add dbfs file path of modules to sparkcontext in order to import to a notebook or python script for access
spark.sparkContext.addPyFile("dbfs:/mnt/capstoneblob/Cloud_Deployment/Azure_configs.py")
spark.sparkContext.addPyFile("dbfs:/mnt/capstoneblob/Cloud_Deployment/API_configs.py")
from Azure_configs import *
today=datetime.date.today().strftime('%m-%d-%y')
```

```python
def nft_raw_tranformation():

    Opensea_df1= spark.read.json(raw_data_file_path)

    Opensea_df2=Opensea_df1.select(
        Opensea_df1['asset_contract']['name'].alias('NFT'),\
        Opensea_df1['token_id'],\
        Opensea_df1['num_sales'],\
        Opensea_df1['owner']['user']['username'].alias('username'),\
        Opensea_df1['owner']['address'].alias('owner_address'),\
        to_timestamp(Opensea_df1['last_sale']['event_timestamp']).alias('txn_date'),\
        (Opensea_df1['last_sale']['total_price']/10**18).alias('payment_amt'),\
        Opensea_df1['last_sale']['payment_token']['symbol'].alias('payment_type'))


    Reference_df=Opensea_df2.select(
        Opensea_df2['NFT'],\
        Opensea_df2['owner_address'],\
        Opensea_df2['txn_date'])


    Opensea_df2.show(10,truncate=False)
    Opensea_df2.write.mode('overwrite').parquet(f'{processed_data_path}{today}/NFT_Collection/')
    return Reference_df
```

```python
reference_df=nft_raw_tranformation()
```

```
+---------+--------+---------+-------------+-----------------------------------------+-------------------+-----------+------------+
|NFT      |token_id|num_sales|username     |owner_address                            |txn_date           |payment_amt|payment_type|
+---------+--------+---------+-------------+-----------------------------------------+-------------------+-----------+------------+
|Cool Cats|1490    |1        |dontbotherme |0x762b35b809ac4266beb076ff0f28547ad571201e|2021-10-05 01:44:29|320.0      |ETH         |
|Cool Cats|3330    |1        |CoinUnited   |0x4c4a5490deefefa16f49a1a48c9acdc60f4117d0|2021-08-20 00:45:13|110.0      |ETH         |
|Cool Cats|5635    |3        |null         |0xe13756351f9cbc45ad6c4da1542faed0ee1a7526|2021-10-07 14:48:23|99.0       |ETH         |
|Cool Cats|8624    |2        |FriendlyDegen|0x2f5170deea823099d75f200ae0524b30c3701881|2021-10-08 02:33:07|83.0       |WETH        |
|Cool Cats|5280    |1        |SighVault    |0xf5a9288eb6e86a3fcb717e7f13475d947f459e3a|2021-08-07 17:52:48|80.0       |ETH         |
|Cool Cats|2157    |3        |MR_CC_VAULT  |0x9edf9b08406fa69a9dd5f73269f8a927b4f772d7|2021-09-26 15:01:20|77.0       |ETH         |
|Cool Cats|8875    |4        |0xErnestVault|0xdb6d9af38ecadaf48112a75ba9a8e5cd6dcba91e|2021-08-28 17:17:38|75.0       |ETH         |
|Cool Cats|4695    |1        |MetalMario   |0xad8357353ddf8095dd01376be71462d27db8cffe|2021-08-01 16:38:00|75.0       |ETH         |
|Cool Cats|3271    |2        |Driftershoots|0x9dbe56e65961146525d796bdc008225bd5915a4f|2021-10-06 18:26:08|69.69      |ETH         |
|Cool Cats|8344    |2        |MR_CC_VAULT  |0x9edf9b08406fa69a9dd5f73269f8a927b4f772d7|2021-08-24 21:55:00|69.42      |ETH         |
+---------+--------+---------+-------------+-----------------------------------------+-------------------+-----------+------------+
only showing top 10 rows
```

```python
def create_ccompare_reference(reference):
    unix_df=reference.select(
        reference['NFT'],\
        reference['txn_date'])\
        .withColumn('unix',unix_timestamp(reference['txn_date']))
    unix_df.show(10,truncate=False)
    unix_df.write.partitionBy('NFT').mode('overwrite').parquet(f'{preprocessed_data_path}{today}/CCompare/')

    return
```

```python
create_ccompare_reference(reference_df)
```

```
+---------+-------------------+----------+
|NFT      |txn_date           |unix      |
+---------+-------------------+----------+
|Cool Cats|2021-10-05 01:44:29|1633398269|
|Cool Cats|2021-08-20 00:45:13|1629420313|
|Cool Cats|2021-10-07 14:48:23|1633618103|
|Cool Cats|2021-10-08 02:33:07|1633660387|
|Cool Cats|2021-08-07 17:52:48|1628358768|
|Cool Cats|2021-09-26 15:01:20|1632668480|
|Cool Cats|2021-08-28 17:17:38|1630171058|
|Cool Cats|2021-08-01 16:38:00|1627835880|
|Cool Cats|2021-10-06 18:26:08|1633544768|
|Cool Cats|2021-08-24 21:55:00|1629842100|
+---------+-------------------+----------+
only showing top 10 rows
```

```python
def create_escan_reference(reference):
    eth_addr_df=reference.select(
        reference['NFT'],\
        reference['owner_address'])\
        .dropDuplicates(['owner_address'])

    eth_addr_df.show(10,truncate=False)
    eth_addr_df.write.partitionBy('NFT').mode('overwrite').parquet(f'{preprocessed_data_path}{today}/EScan/')

    return
```

```python
create_escan_reference(reference_df)
```

```
+-----------------+------------------------------------------+
|NFT              |owner_address                             |
+-----------------+------------------------------------------+
|Meebits          |0x005e9eed36bfea0d05c0e8f36f32d4f4e08efacd|
|BoredApeYachtClub|0x020ca66c30bec2c4fe3861a94e4db4a498a35872|
|CryptoPunks      |0x040da2c464933005b6d1ffeccec7fb4025dc9ddb|
|CryptoPunks      |0x04ae2f0bda04f1405991d91c2e8420d6148369ea|
|Doodles          |0x052564eb0fd8b340803df55def89c25c432f43f4|
|BoredApeYachtClub|0x05c250120ce07ba6fe361b39ac344148435c25ca|
|BoredApeYachtClub|0x066317b90509069eb52474a38c212508f8a1211c|
|CryptoPunks      |0x06e63138f3241a420829bc125e6cb6bebf88c2c2|
|Meebits          |0x0845fc89c51b2bcd1c3b0db9dbca497d641ec7d3|
|Meebits          |0x09d4083ffd20d21acb9118465ad7c52ac8b548f7|
+-----------------+------------------------------------------+
only showing top 10 rows
```