Processed Transformation USD

Uses CryptoCompare reference data for USD data enrichment

Input:
- Data Content: CryptoCompare Reference Data
- Data Type: Parquet
- Data Source: Preprocessed Layer

Output:
- Data Content: CryptoCompare Enriched USD Data
- Data Type: Parquet
- Data Destination: Processed Layer

** Notebook is simply for reference and only takes in a sample collection

```
from pyspark.sql.functions import udf
from Azure_configs import preprocessed_data_path, processed_data_path
from API_configs import cryptocompare_api_key, cryptocompare_url
import datetime
import time
import requests
```

```
today=datetime.date.today().strftime('%m-%d-%y')
```

```
def get_usd_price(unix):
    url=f'{cryptocompare_url}{unix}'
    limit_exceeded=True

    while limit_exceeded==True:
        response=requests.get(url,params={'api_key':cryptocompare_api_key})
        message=response.json()
        if message['ETH']:
            limit_exceeded=False
            return message['ETH']['USD']
        else:
            time.sleep(0.5)
            continue
    return
```

```
def process_usd_data(nft_name):

    CCompare_parquet_path=f'{preprocessed_data_path}/{today}/CCompare/NFT={nft_name}/'

    CCompare_reference_DF=spark.read.parquet(CCompare_parquet_path)

    dollar_udf=udf(lambda x : get_usd_price(x))

    usd_price_df=CCompare_reference_DF.withColumn('USD_Rate',dollar_udf(CCompare_reference_DF['unix'])).drop(CCompare_reference_DF['unix'])
    usd_price_df.cache()
    usd_price_df.show(10,truncate=False)
    usd_price_df.write.mode('overwrite').parquet(f'{processed_data_path}{today}/USD_Price/NFT={nft_name}/')

    return
```

```
process_usd_data('CryptoPunks')
```

```
+-------------------+--------+
|txn_date           |USD_Rate|
+-------------------+--------+
|2021-10-28 23:50:37|4288.22 |
|2021-03-11 00:48:39|1827.06 |
|2021-03-11 20:57:01|1827.06 |
|2021-12-09 20:38:17|4111.92 |
|2021-07-30 21:24:47|2463.61 |
|2021-09-11 18:41:23|3266.43 |
|2021-07-30 18:36:56|2463.61 |
|2021-08-24 06:34:02|3172.57 |
|2021-08-06 18:18:56|2892.36 |
|2021-09-04 21:39:01|3887.01 |
+-------------------+--------+
only showing top 10 rows
```