

Question : Screenshots of the execution of the scripts written. The scripts should, after loading the data and creating the look-up table, take the data from the NoSQL database and AWS RDS and perform the relevant analyses as per the rules and should feed the data in the look-up table.

Task 3: Create a look-up table in MongoDB {NoSQL database} with columns specified earlier in the problem statement.

```
In [2]: df_cardMember = pd.read_csv('card_member.csv')
df_member_score = pd.read_csv('member_score.csv')
df_card_transactions = pd.read_csv('card_transactions.csv')
```

```
In [3]: print("Number of Records in df_cardMember : ", df_cardMember.shape[0])
print("Number of Records in df_member_score is :", df_member_score.shape[0])
print("Number of Records in df_card_transactions is :", df_card_transactions.shape[0])

Number of Records in df_cardMember : 999
Number of Records in df_member_score is : 999
Number of Records in df_card_transactions is : 53292
```

```
In [6]: df_card_transactions = df_card_transactions[df_card_transactions.status == 'GENUINE']
df_card_transactions
```

Out[6]:

	card_id	member_id	amount	postcode	pos_id	transaction_dt	status
0	348702330256514	37495066290	9084849	33946	614677375609919	11-02-2018 00:00:00	GENUINE
1	348702330256514	37495066290	330148	33946	614677375609919	11-02-2018 00:00:00	GENUINE
2	348702330256514	37495066290	136052	33946	614677375609919	11-02-2018 00:00:00	GENUINE
3	348702330256514	37495066290	4310362	33946	614677375609919	11-02-2018 00:00:00	GENUINE
4	348702330256514	37495066290	9097094	33946	614677375609919	11-02-2018 00:00:00	GENUINE
...
53287	5391723993945313	997128952368160	7234353	54514	366418882890893	30-01-2018 14:21:15	GENUINE
53288	5391723993945313	997128952368160	283051	20711	479856487174409	30-11-2017 05:24:26	GENUINE
53289	5391723993945313	997128952368160	5527334	20711	479856487174409	30-11-2017 07:09:46	GENUINE
53290	5391723993945313	997128952368160	714092	20711	479856487174409	30-11-2017 12:25:47	GENUINE
53291	5391723993945313	997128952368160	2053925	99929	283991209995468	31-12-2017 23:20:13	GENUINE

53210 rows × 7 columns

```
In [8]: ▶ df_card_transactions['transaction_dt2']=pd.to_datetime(df_card_transactions['transaction_dt'])
```

```
In [9]: ▶ df_card_transactions.dtypes
```

```
Out[9]: card_id          int64
member_id        int64
amount           int64
postcode         int64
pos_id           int64
transaction_dt    object
status           object
transaction_dt2   datetime64[ns]
dtype: object
```

```
In [10]: ▶ # Sort descending order
df_card_transactions.sort_values(by=['member_id','transaction_dt2'], inplace=True,
                                ascending = [True, False])
df_card_transactions
```

```
Out[10]:
```

	card_id	member_id	amount	postcode	pos_id	transaction_dt	status	transaction_dt2
0	348702330256514	37495066290	9084849	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00
1	348702330256514	37495066290	330148	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00
2	348702330256514	37495066290	136052	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00
3	348702330256514	37495066290	4310362	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00
4	348702330256514	37495066290	9097094	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00
...
53207	5391723993945313	997128952368160	1396526	53968	716276786938354	05-11-2017 12:00:54	GENUINE	2017-05-11 12:00:54
53206	5391723993945313	997128952368160	25146	13124	200432834928149	04-11-2017 18:27:33	GENUINE	2017-04-11 18:27:33
53205	5391723993945313	997128952368160	2727621	57376	97278983752405	03-12-2017 22:56:34	GENUINE	2017-03-12 22:56:34
53204	5391723993945313	997128952368160	7201018	48832	712082643297946	03-11-2017 19:38:11	GENUINE	2017-03-11 19:38:11
53201	5391723993945313	997128952368160	6790715	68790	957396750250096	02-11-2017 20:48:49	GENUINE	2017-02-11 20:48:49

53210 rows × 8 columns

```
In [11]: ▶ df_card_transactions["Rank"] = df_card_transactions.groupby("card_id")["transaction_dt2"].rank(ascending=False, method='dense')
df_card_transactions.sort_values("Rank")
df_card_transactions_top10 = df_card_transactions[(df_card_transactions.Rank <=10) ]
df_card_transactions_top10
```

Out[11]:

	card_id	member_id	amount	postcode	pos_id	transaction_dt	status	transaction_dt2	Rank
0	348702330256514	37495066290	9084849	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00	1
1	348702330256514	37495066290	330148	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00	1
2	348702330256514	37495066290	136052	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00	1
3	348702330256514	37495066290	4310362	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00	1
4	348702330256514	37495066290	9097094	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00	1
...
53202	5391723993945313	997128952368160	9670623	77974	675701087475560	03-01-2018 07:30:58	GENUINE	2018-03-01 07:30:58	6
53200	5391723993945313	997128952368160	1164872	93210	171173202259233	02-01-2018 19:13:37	GENUINE	2018-02-01 19:13:37	7
53287	5391723993945313	997128952368160	7234353	54514	366418882890893	30-01-2018 14:21:15	GENUINE	2018-01-30 14:21:15	8
53284	5391723993945313	997128952368160	2683995	44065	42771743906415	29-01-2018 13:46:33	GENUINE	2018-01-29 13:46:33	9
53283	5391723993945313	997128952368160	1036903	44065	42771743906415	29-01-2018 12:01:13	GENUINE	2018-01-29 12:01:13	10

```
In [13]: df_ucl_sd = df_card_transactions_top10.groupby(['member_id'])['amount'].agg(['mean', 'std'])
df_ucl_sd
```

Out[13]:

	mean	std
member_id		
37495066290	4.260123e+06	2.917135e+06
117826301530	4.300037e+06	2.666817e+06
1147922084344	3.442887e+06	2.816608e+06
1314074991813	4.937506e+06	2.618911e+06
1739553947511	5.040048e+06	3.333798e+06
...
992104321998571	6.906752e+06	2.194694e+06
992552823055811	4.959108e+06	3.389936e+06
994983851226493	5.433286e+06	2.733345e+06
996411635289270	5.629739e+06	3.168106e+06
997128952368160	5.094256e+06	3.053936e+06

999 rows × 2 columns

```
In [16]: df_ucl_sd.rename(columns = {'mean':'Avg1', 'std':'SD'}, inplace = True)
```

```
In [17]: df_ucl_sd['ucl'] = df_ucl_sd.Avg1 + 3*df_ucl_sd.SD
```

```
In [18]: df_ucl_sd
```

Out[18]:

	member_id	Avg1	SD	ucl
0	37495066290	4.260123e+06	2.917135e+06	1.301153e+07
1	117826301530	4.300037e+06	2.666817e+06	1.230049e+07
2	1147922084344	3.442887e+06	2.816608e+06	1.189271e+07
3	1314074991813	4.937506e+06	2.618911e+06	1.279424e+07
4	1739553947511	5.040048e+06	3.333798e+06	1.504144e+07
...
994	992104321998571	6.906752e+06	2.194694e+06	1.349084e+07
995	992552823055811	4.959108e+06	3.389936e+06	1.512892e+07
996	994983851226493	5.433286e+06	2.733345e+06	1.363332e+07
997	996411635289270	5.629739e+06	3.168106e+06	1.513406e+07
998	997128952368160	5.094256e+06	3.053936e+06	1.425606e+07

999 rows × 4 columns

In [19]: `lstTxnPostCodeTxDate = df_card_transactions[df_card_transactions.Rank ==1]`
`lstTxnPostCodeTxDate`

Out[19]:

	card_id	member_id	amount	postcode	pos_id	transaction_dt	status	transaction_dt2	Rank
0	348702330256514	37495066290	9084849	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00	1
1	348702330256514	37495066290	330148	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00	1
2	348702330256514	37495066290	136052	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00	1
3	348702330256514	37495066290	4310362	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00	1
4	348702330256514	37495066290	9097094	33946	614677375609919	11-02-2018 00:00:00	GENUINE	2018-11-02 00:00:00	1
...
53046	5115273575078211	992104321998571	8416500	42082	609717185246135	08-01-2018 15:40:29	GENUINE	2018-08-01 15:40:29	1
53097	374878208630443	992552823055811	4185444	12747	640304888744272	04-01-2018 22:08:21	GENUINE	2018-04-01 22:08:21	1
53162	5399715069800093	994983851226493	4302555	77432	487096286569669	30-01-2018 09:05:15	GENUINE	2018-01-30 09:05:15	1
53189	4981548424137027	996411635289270	4993560	92694	181718433544561	18-12-2017 00:01:04	GENUINE	2017-12-18 00:01:04	1
53213	5391723993945313	997128952368160	7338503	42323	662287296546616	08-01-2018 12:09:49	GENUINE	2018-08-01 12:09:49	1

1658 rows × 9 columns

In [20]: `#tmp1 = lstTxnPostCodeTxDate.merge(df_ucl_sd, on='member_id', how='right')`
`tmp1 = df_ucl_sd.merge(lstTxnPostCodeTxDate, on='member_id', how='left')`
`tmp1`

In [21]: `tmp2 = df_member_score.merge(tmp1, on='member_id', how='left')`
`tmp2`

Out[21]:

	member_id	score	Avg1	SD	ucl	card_id	amount	postcode	pos_id	transaction_dt	status
0	37495066290	339	4.260123e+06	2.917135e+06	1.301153e+07	348702330256514	9084849	33946	614677375609919	11-02-2018 00:00:00	GENUINE
1	37495066290	339	4.260123e+06	2.917135e+06	1.301153e+07	348702330256514	330148	33946	614677375609919	11-02-2018 00:00:00	GENUINE
2	37495066290	339	4.260123e+06	2.917135e+06	1.301153e+07	348702330256514	136052	33946	614677375609919	11-02-2018 00:00:00	GENUINE
3	37495066290	339	4.260123e+06	2.917135e+06	1.301153e+07	348702330256514	4310362	33946	614677375609919	11-02-2018 00:00:00	GENUINE
4	37495066290	339	4.260123e+06	2.917135e+06	1.301153e+07	348702330256514	9097094	33946	614677375609919	11-02-2018 00:00:00	GENUINE
...
1653	992104321998571	658	6.906752e+06	2.194694e+06	1.349084e+07	5115273575078211	8416500	42082	609717185246135	08-01-2018 15:40:29	GENUINE
1654	992552823055811	694	4.959108e+06	3.389936e+06	1.512892e+07	374878208630443	4185444	12747	640304888744272	04-01-2018 22:08:21	GENUINE
1655	994983851226493	687	5.433286e+06	2.733345e+06	1.363332e+07	5399715069800093	4302555	77432	487096286569669	30-01-2018 09:05:15	GENUINE

```
In [22]: ► columns = ['amount', 'pos_id', 'status', 'transaction_dt', 'status', 'Rank', 'Avg1', 'SD']
tmp2.drop(columns, inplace=True, axis=1)
lookup_df = tmp2
lookup_df = lookup_df.drop_duplicates(subset=['member_id'])
lookup_df = lookup_df.reset_index()
lookup_df
```

Out[22]:

	index	member_id	score	ucl	card_id	postcode	transaction_dt2
0	0	37495066290	339	1.301153e+07	348702330256514	33946	2018-11-02 00:00:00
1	42	117826301530	289	1.230049e+07	5189563368503974	30234	2018-06-01 09:15:04
2	43	1147922084344	393	1.189271e+07	5407073344486464	63770	2018-01-18 10:55:30
3	44	1314074991813	225	1.279424e+07	378303738095292	41778	2018-06-01 14:31:05
4	45	1739553947511	642	1.504144e+07	348413196172048	92258	2018-12-01 21:29:58
...
994	1653	992104321998571	658	1.349084e+07	5115273575078211	42082	2018-08-01 15:40:29
995	1654	992552823055811	694	1.512892e+07	374878208630443	12747	2018-04-01 22:08:21
996	1655	994983851226493	687	1.363332e+07	5399715069800093	77432	2018-01-30 09:05:15
997	1656	996411635289270	335	1.513406e+07	4981548424137027	92694	2017-12-18 00:01:04
998	1657	997128952368160	683	1.425606e+07	5391723993945313	42323	2018-08-01 12:09:49

999 rows × 7 columns

```
In [23]: ► final_lookup_df = lookup_df[['member_id', 'score', 'ucl', 'card_id', 'postcode', 'transaction_dt2']]
final_lookup_df
```

Out[23]:

	member_id	score	ucl	card_id	postcode	transaction_dt2
0	37495066290	339	1.301153e+07	348702330256514	33946	2018-11-02 00:00:00
1	117826301530	289	1.230049e+07	5189563368503974	30234	2018-06-01 09:15:04
2	1147922084344	393	1.189271e+07	5407073344486464	63770	2018-01-18 10:55:30
3	1314074991813	225	1.279424e+07	378303738095292	41778	2018-06-01 14:31:05
4	1739553947511	642	1.504144e+07	348413196172048	92258	2018-12-01 21:29:58
...
994	992104321998571	658	1.349084e+07	5115273575078211	42082	2018-08-01 15:40:29
995	992552823055811	694	1.512892e+07	374878208630443	12747	2018-04-01 22:08:21
996	994983851226493	687	1.363332e+07	5399715069800093	77432	2018-01-30 09:05:15
997	996411635289270	335	1.513406e+07	4981548424137027	92694	2017-12-18 00:01:04
998	997128952368160	683	1.425606e+07	5391723993945313	42323	2018-08-01 12:09:49

999 rows × 6 columns

```
In [24]: ► #determining the name of the file
file_name = 'lookupTable4.csv'

#saving the excel
final_lookup_df.to_csv(file_name)
print('DataFrame lookup_df - is written to Excel File successfully.')
```

DataFrame lookup_df - is written to Excel File successfully.