

```
In [47]: import pandas as pd
import numpy as np
```

## Conversation\_Sheet

```
In [48]: excel_file = '/home/siddharth/Desktop/FT/CodeChallenge/vividmoney/Conversations/conversation.csv'
conversation = pd.read_csv(excel_file)
```

```
In [49]: conversation.head()
```

```
Out[49]:
```

	CONVERSATION_HK	CREATED_AT	STATUS_CD	CLOSED_AT	R
0	0008a84b8d7f6bf55740053ca635b11722bfe4f38a9ffa...	2020-11-16 17:18:40	closed	2020-11-16 18:41:31	
1	000ad81c0e996b624c296cb105b11c5f76777fae456ecb...	2020-12-08 9:13:49	closed	2020-12-13 9:04:20	
2	000c2e8ad3fb870813d3ea8f7ea1094f7508d9fee1205f...	2020-12-28 14:09:31	closed	2020-12-28 15:58:05	
3	000cc854daaeff5b59deaae9f11c18c0d59beda58b7249...	2021-01-08 8:49:28	closed	2021-01-13 8:09:37	
4	000db3ed6e7c9a1ab900a61fa02f2fc6d75fb65670d6b0...	2020-12-03 22:15:15	closed	2020-12-05 1:58:26	

```
In [99]: conversation.columns
```

```
Out[99]: Index(['CONVERSATION_HK', 'CREATED_AT', 'STATUS_CD', 'CLOSED_AT', 'RATING_NUM',
               'LANGUAGE_CD', 'FRT_DU_IN_MIN', 'INITIATOR_CD', 'CUSTOMER_HK'],
              dtype='object')
```

```
In [103]: conversation.dtypes
```

```
Out[103]: CONVERSATION_HK      object
CREATED_AT      object
STATUS_CD       object
CLOSED_AT       object
RATING_NUM      float64
LANGUAGE_CD     object
FRT_DU_IN_MIN   float64
INITIATOR_CD    object
CUSTOMER_HK     object
dtype: object
```

```
In [51]: pd.unique(conversation['CONVERSATION_HK'])
```

```
Out[51]: array(['0008a84b8d7f6bf55740053ca635b11722bfe4f38a9ffa21a05ae21df81caea0',
               '000ad81c0e996b624c296cb105b11c5f76777fae456ecbbc8c7e5be9fec4a38a',
               '000c2e8ad3fb870813d3ea8f7ea1094f7508d9fee1205f88fe22ea535360bfc0',
               ...,
               'fffc9d914a96f3b7d2468cdd704d3857cc12ebfb12baf9b92e08e37e04ece0af4',
               'fffd5c392e3c25ff2eac6ff3a1f6c8cd0838278490eaaa4a04f4c0d1402673b',
               'fffe1f9d3b3304a99ef46fa7322d77a5d88792b91e032c41f5bf7e550b1ffbe'],
              dtype=object)
```

```
In [67]: conversation['CONVERSATION_HK'].describe()
```

```
Out[67]: count          40315  
unique          39535  
top      ae82fa8b4061eabb005cb43b25ece06de62324be326cda...  
freq                      2  
Name: CONVERSATION_HK, dtype: object
```

```
In [84]: conversation['CLOSED_AT'].describe()
```

```
Out[84]: count          38793  
unique          38650  
top      2021-01-28 14:57:19  
freq                      2  
Name: CLOSED_AT, dtype: object
```

```
In [87]: conversation['CLOSED_AT'].isnull().values.any()  
conversation['CLOSED_AT'].isnull().sum().sum()
```

```
Out[87]: 742
```

```
In [53]: conversation['CUSTOMER_HK'].describe()
```

```
Out[53]: count          40315  
unique          19717  
top      238fe4ef9aa063a4ffe2424a20ead5ab9d269cf43ac328...  
freq                      70  
Name: CUSTOMER_HK, dtype: object
```

Removing Duplicates

```
In [73]: conversation.drop_duplicates(subset = "CONVERSATION_HK",  
                                     keep = 'first', inplace = True)
```

```
In [74]: conversation['CONVERSATION_HK'].describe()
```

```
Out[74]: count          39535  
unique          39535  
top      b91ed6273ce2c8f16c5216ad9555106a9a2bdf9c232604...  
freq                      1  
Name: CONVERSATION_HK, dtype: object
```

```
In [75]: conversation['CONVERSATION_HK'].is_unique
```

```
Out[75]: True
```

```
In [109... conversation['CREATED_AT'] = pd.to_datetime(conversation['CREATED_AT'])  
conversation.set_index('CREATED_AT', drop=False, inplace=True)  
  
conversation['CLOSED_AT'] = pd.to_datetime(conversation['CLOSED_AT'])  
conversation.set_index('CLOSED_AT', drop=False, inplace=True)
```

```
In [110... conversation.dtypes
```

```
Out[110... CONVERSATION_HK          object
```

```
CREATED_AT      datetime64[ns]
STATUS_CD       object
CLOSED_AT       datetime64[ns]
RATING_NUM      float64
LANGUAGE_CD     object
FRT_DU_IN_MIN   float64
INITIATOR_CD    object
CUSTOMER_HK     object
dtype: object
```

## CALCULATING RESOLUTION TIME: time taken to solve the issue/ticket

In [111...]

```
conversation['resolution_time'] = conversation['CLOSED_AT'] - conversation['CREATED_AT']
conversation['resolution_time'] = conversation['resolution_time']/np.timedelta64(1, 'D')

print(conversation)
```

		CONVERSATION_HK \	
CLOSED_AT			
2020-11-16	18:41:31	0008a84b8d7f6bf55740053ca635b11722bfe4f38a9ffa...	
2020-12-13	09:04:20	000ad81c0e996b624c296cb105b11c5f76777fae456ecb...	
2020-12-28	15:58:05	000c2e8ad3fb870813d3ea8f7ea1094f7508d9fee1205f...	
2021-01-13	08:09:37	000cc854daaeff5b59deaae9f11c18c0d59beda58b7249...	
2020-12-05	01:58:26	000db3ed6e7c9a1ab900a61fa02f2fc6d75fb65670d6b0...	
...			...
2021-01-04	20:17:36	fffa6fcad633f39649228bdd99b6f1fc2365067e6fd0d6...	
2020-12-23	15:19:54	fffb4d2846b9fed3be3d6b6ac195057e57bf34a3650d4b...	
2020-11-23	10:38:23	fffc914a96f3b7d2468cdd704d3857cc12ebfb12baf9b...	
2020-12-07	06:07:21	fffa5c392e3c25ff2eac6ff3a1f6c8cd0838278490eaa...	
NaT		fffee1f9d3b3304a99ef46fa7322d77a5d88792b91e032...	

		CREATED_AT		STATUS_CD		CLOSED_AT \	
CLOSED_AT							
2020-11-16	18:41:31	2020-11-16	17:18:40	closed	2020-11-16	18:41:31	
2020-12-13	09:04:20	2020-12-08	09:13:49	closed	2020-12-13	09:04:20	
2020-12-28	15:58:05	2020-12-28	14:09:31	closed	2020-12-28	15:58:05	
2021-01-13	08:09:37	2021-01-08	08:49:28	closed	2021-01-13	08:09:37	
2020-12-05	01:58:26	2020-12-03	22:15:15	closed	2020-12-05	01:58:26	
...			...		...		...
2021-01-04	20:17:36	2021-01-04	15:56:03	closed	2021-01-04	20:17:36	
2020-12-23	15:19:54	2020-12-23	13:32:14	closed	2020-12-23	15:19:54	
2020-11-23	10:38:23	2020-11-20	18:24:26	closed	2020-11-23	10:38:23	
2020-12-07	06:07:21	2020-12-07	05:41:18	closed	2020-12-07	06:07:21	
NaT		2020-12-10	13:37:44	closed		NaT	

		RATING_NUM		LANGUAGE_CD		FRT_DU_IN_MIN		INITIATOR_CD \	
CLOSED_AT									
2020-11-16	18:41:31		NaN	de	82.816667			Customer	
2020-12-13	09:04:20		NaN	de	5759.333333			Customer	
2020-12-28	15:58:05		NaN	fr	35.666667			Customer	
2021-01-13	08:09:37		NaN	de	653.800000			Customer	
2020-12-05	01:58:26		NaN	en	843.033333			Customer	
...		...		...	...	...		...	
2021-01-04	20:17:36		NaN	en	260.566667			Customer	
2020-12-23	15:19:54		4.0	fr	27.033333			Customer	
2020-11-23	10:38:23		NaN	en	1772.050000			Customer	
2020-12-07	06:07:21		NaN	en	2.933333			Customer	
NaT			NaN	de		NaN		Vivid Money	

		CUSTOMER_HK \	
CLOSED_AT			

```

2020-11-16 18:41:31 04c604e678b18d7ca10b85fb8d3672a03b7b6b71e78dd5...
2020-12-13 09:04:20 a42a9fdf0ee31d331c9d96ce6fbf30f307b52ff40a55ec...
2020-12-28 15:58:05 be92b2698f17aa7fab796fd5eb132cc251fce6fe1870dc...
2021-01-13 08:09:37 c84455bb3bbc2dbf16a1edd9c44e76baaf06b0ee224ed5...
2020-12-05 01:58:26 7df3aa4e5f29e4f5946d652d4f6d6f88d3f850ac9ba33a...
...
2021-01-04 20:17:36 4c65aeb4e3c56d9ad692a2dd53f6e58d3b1c03f53ebee8...
2020-12-23 15:19:54 b3c4de44a98d55acc296d54636f321323e55bc793f0c01...
2020-11-23 10:38:23 ed1cd862b1e0c82f9c96b6a895c6122fbfcac469801701...
2020-12-07 06:07:21 68b47f0f4243f0fcc0e21bc72b750f76387838e20aa82...
NaT 12a2683372a4a0a3ffaead2fd17bd789e14893bcb094ed...

```

resolution\_time

```

CLOSED_AT
2020-11-16 18:41:31      82.850000
2020-12-13 09:04:20     7190.516667
2020-12-28 15:58:05     108.566667
2021-01-13 08:09:37     7160.150000
2020-12-05 01:58:26    1663.183333
...
2021-01-04 20:17:36     261.550000
2020-12-23 15:19:54     107.666667
2020-11-23 10:38:23    3853.950000
2020-12-07 06:07:21      26.050000
NaT NaN

```

[39535 rows x 10 columns]

Converting dataframes to CSV for to be done in POWER BI & SQL

In [115...]

```
conversation.to_csv('/home/siddharth/Desktop/FT/CodeChallenge/vividmoney/Conv
```

## customer\_monthly

In [81]:

```
excel_file = '/home/siddharth/Desktop/FT/CodeChallenge/vividmoney/Customers M
customer_monthly = pd.read_csv(excel_file)
```

In [82]:

```
customer_monthly.head()
```

Out[82]:

	CUSTOMER_HK	DAY_DT	BANK_IS_ACTIVE	INVEST_IS_AC
0	04c604e678b18d7ca10b85fb8d3672a03b7b6b71e78dd5...	2020-11-30 0:00:00	True	f
1	a42a9fdf0ee31d331c9d96ce6fbf30f307b52ff40a55ec...	2021-01-31 0:00:00	True	f
2	be92b2698f17aa7fab796fd5eb132cc251fce6fe1870dc...	2020-12-31 0:00:00	True	f
3	c84455bb3bbc2dbf16a1edd9c44e76baaf06b0ee224ed5...	2020-12-31 0:00:00	True	f
4	7df3aa4e5f29e4f5946d652d4f6d6f88d3f850ac9ba33a...	2020-12-31 0:00:00	False	f

In [118...]

```
customer_monthly['CUSTOMER_HK'].describe()
```

```
Out[118... count                26766
unique                19717
top    ee2b73ad9b9c8e61d1ba93818c15efdc6c931912e11918...
freq                                3
Name: CUSTOMER_HK, dtype: object
```

Removing Duplicates

```
In [117... # customer_monthly.drop_duplicates(subset = "CONVERSATION_HK",
#                                   keep = 'first', inplace = True)

customer_monthly.drop_duplicates(keep = 'first', inplace = True)
```

```
In [119... customer_monthly.to_csv('/home/siddharth/Desktop/FT/CodeChallenge/vividmoney/
```

```
In [ ]:
```