music

February 11, 2025

[3]:
```
# Import
import pandas as pd
```

[7]:
```
# Read Files

# Path to your Parquet file
parquet_file = '/Users/henryding/Documents/Seminar/Data/output_data1.parquet'

# Read the Parquet file into a DataFrame
df = pd.read_parquet(parquet_file, engine='pyarrow')  # You can also use␣
 ↪'fastparquet'

# Display the first few rows of the DataFrame
display(df.head(1000))
num_rows = df.shape[0]
print(f"The number of rows in the DataFrame is: {num_rows}")
```

```
    status gender      length firstName level  lastName  registration  \
0      200      M   524.32934     Shlok  paid   Johnson  1.533735e+12
1      200      F   238.39302   Vianney  paid    Miller  1.537500e+12
2      200      F   140.35546      Vina  paid    Bailey  1.536415e+12
3      200      M   277.15873    Andres  paid     Foley  1.534387e+12
4      200      F  1121.25342   Aaliyah  paid   Ramirez  1.537381e+12
..     …     …         …       …    …       …           …
995    307      F         NaN    Alivia  paid  Williams  1.535955e+12
996    200      M   315.48036 Christian  free     Klein  1.536940e+12
997    200      M   248.73751 Kristofer  free     James  1.536879e+12
998    200      F    53.08036      Zoey  free   Gregory  1.533190e+12
999    200      F         NaN      Zoey  free   Gregory  1.533190e+12

      userId             ts       auth       page  sessionId  \
0    1749042  1538352001000  Logged In   NextSong      22683
1    1563081  1538352002000  Logged In   NextSong      20836
2    1697168  1538352002000  Logged In   NextSong       4593
3    1222580  1538352003000  Logged In   NextSong       6370
4    1714398  1538352003000  Logged In   NextSong      22316
..       …              …        …          …        …
995  1792538  1538352525000  Logged In  Thumbs Up       8871
```

```
996  1291813  1538352526000  Logged In     NextSong      14821
997  1929921  1538352526000  Logged In     NextSong       9831
998  1839740  1538352527000  Logged In     NextSong      24848
999  1839740  1538352527000  Logged In  Roll Advert      24848

                                  location  itemInSession  \
0          Dallas-Fort Worth-Arlington, TX            278
1        San Francisco-Oakland-Hayward, CA              9
2                                 Hilo, HI            109
3                            Watertown, SD             71
4            Baltimore-Columbia-Towson, MD             21
..                                     …              …
995  New York-Newark-Jersey City, NY-NJ-PA             75
996  New York-Newark-Jersey City, NY-NJ-PA             10
997            Seattle-Tacoma-Bellevue, WA             56
998            Phoenix-Mesa-Scottsdale, AZ             11
999            Phoenix-Mesa-Scottsdale, AZ             12

                                             userAgent  \
0    "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebK…
1    "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4…
2    Mozilla/5.0 (Macintosh; Intel Mac OS X 10.9; r…
3    "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4…
4    "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebK…
..                                                 …
995  Mozilla/5.0 (Macintosh; Intel Mac OS X 10.9; r…
996  "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_2…
997  Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:31…
998  "Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/5…
999  "Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/5…

                                                  song  \
0               Ich mache einen Spiegel - Dream Part 4
1                                          MiÃ Â©ntele
2                                            Baby Talk
3    Horn Concerto No. 4 in E flat K495: II. Romanc…
4    Close To The Edge (I. The Solid Time Of Change…
..                                                 …
995                                               None
996                                             Odessa
997                          Jigsaw Falling Into Place
998             Broken Hearted Hoover Fixer Sucker Guy
999                                               None

                                               artist method
0                                           Popol Vuh    PUT
1                                         Los Bunkers    PUT
2                                                Lush    PUT
```

```
3    Barry Tuckwell/Academy of St Martin-in-the-Fie…   PUT
4                                                  Yes    PUT
..                                                 …    …
995                                               None    PUT
996                                            Caribou    PUT
997                                          Radiohead    PUT
998                                       Glen Hansard    PUT
999                                               None    GET

[1000 rows x 18 columns]

The number of rows in the DataFrame is: 26259199
```

[8]:

```python
# Group by 'userId' and count the occurrences
user_counts = df.groupby('userId').size().reset_index(name='count')

# Display the result
display(user_counts.sort_values(by='count', ascending=False))
```

```
        userId   count
5936   1261737  778479
12534  1564221   13591
20802  1931933   12831
163    1006695   12372
7562   1336969   11858
…        …      …
6072   1267517       1
9698   1434698       1
17778  1793623       1
9129   1408726       1
15322  1689121       1

[22278 rows x 2 columns]
```

[9]:

```python
print(pd.to_datetime(df['ts'].min(), unit='ms'))
print(pd.to_datetime(df['ts'].max(), unit='ms'))
```

```
2018-10-01 00:00:01
2018-12-01 00:00:02
```

[ ]:

```python

```