03. Most Frequent Items

From the following table containing a list of dates and items ordered, write a query to return the most frequent item ordered on each date. Return multiple items in the case of a tie.

```
!pip install polars
import pandas as pd
import numpy as  np
import polars as pl

data = {'dates' : ['01-JAN-20',
                   '01-JAN-20',
                   '01-JAN-20',
                   '01-JAN-20',
                   '02-JAN-20',
                   '02-JAN-20',
                   '02-JAN-20',
                   '02-JAN-20'],
        'item' : ['apple',
                  'apple',
                  'pear',
                  'pear',
                  'pear',
                  'pear',
                  'pear',
                  'orange']

       }
pandas_items=pd.DataFrame(data)
polars_items=pl.DataFrame(data)
```

```
Requirement already satisfied: polars in /usr/local/lib/python3.11/dist-packages
```

```
print(f'items table in Pandas:\n{pandas_items}')
```

```
items table in Pandas:
         dates    item
    0  01-JAN-20   apple
    1  01-JAN-20   apple
    2  01-JAN-20    pear
    3  01-JAN-20    pear
    4  02-JAN-20    pear
    5  02-JAN-20    pear
    6  02-JAN-20    pear
    7  02-JAN-20  orange
```

```
pandas_df1=(pandas_items.groupby(['dates','item'])
                        .size()
                        .reset_index(name='count')
)
print(f'counting items by each date:\n{pandas_df1}')
```

```
counting items by each date:
       dates    item  count
0  01-JAN-20   apple      2
1  01-JAN-20    pear      2
2  02-JAN-20  orange      1
3  02-JAN-20    pear      3
```

```
pandas_df1['max_count']=(pandas_df1.groupby('dates')['count']
                                   .transform('max')
)
query1=pandas_df1[['dates','max_count']].drop_duplicates()
print(f'the max count by each date:\n{query1}')
```

```
the max count by each date:
       dates  max_count
0  01-JAN-20          2
2  02-JAN-20          3
```

```
pandas_counts=(pandas_items.groupby(['dates','item'])
                           .size()
                           .reset_index(name='count')
)
pandas_counts['max_count']=(pandas_counts.groupby('dates')['count']
                                         .transform('max')
)
query=pandas_counts.query('count==max_count')[['dates','item']]
print(f'item most frequented for each date using Pandas:')
query
```

```
item most frequented for each date using Pandas:
```

	dates	item
0	01-JAN-20	apple
1	01-JAN-20	pear
3	02-JAN-20	pear

```
print(f'items table in Polars:\n{polars_items}')
```

```
items table in Polars:
shape: (8, 2)
┌─────────────┬────────┐
│ dates       ┆ item   │
│ ---         ┆ ---    │
│ str         ┆ str    │
╞═════════════╪════════╡
│ 01-JAN-20   ┆ apple  │
│ 01-JAN-20   ┆ apple  │
│ 01-JAN-20   ┆ pear   │
│ 01-JAN-20   ┆ pear   │
│ 02-JAN-20   ┆ pear   │
│ 02-JAN-20   ┆ pear   │
```

```
| 02-JAN-20 ┆ pear   |
| 02-JAN-20 ┆ orange |
└───────────┴────────┘
```

```
polars_df1=(polars_items.group_by(['dates','item'])
                        .len()
)
print(f'counting items by each date:\n{polars_df1}')
```

```
counting items by each date:
shape: (4, 3)
┌───────────┬────────┬─────┐
| dates     ┆ item   ┆ len |
| ---       ┆ ---    ┆ --- |
| str       ┆ str    ┆ u32 |
╞═══════════╪════════╪═════╡
| 01-JAN-20 ┆ apple  ┆ 2   |
| 01-JAN-20 ┆ pear   ┆ 2   |
| 02-JAN-20 ┆ pear   ┆ 3   |
| 02-JAN-20 ┆ orange ┆ 1   |
└───────────┴────────┴─────┘
```

```
polars_df1=(polars_df1.with_columns(pl.max('len')
                                      .over(partition_by='dates')
                                      .alias('max_count')
                       )
                      .select(pl.col('dates'),
                              pl.col('max_count')
                       )
                      .unique()
)
print(f'the max count by each date:\n{polars_df1}')
```

```
the max count by each date:
shape: (2, 2)
┌───────────┬───────────┐
| dates     ┆ max_count |
| ---       ┆ ---       |
| str       ┆ u32       |
╞═══════════╪═══════════╡
| 02-JAN-20 ┆ 3         |
| 01-JAN-20 ┆ 2         |
└───────────┴───────────┘
```

```
polars_counts=(polars_items.group_by(['dates','item'])
                           .agg(pl.len()
                                  .alias('count')
                            )
                           .with_columns(pl.max('count')
                                           .over(partition_by='dates')
                                           .alias('max_count')
                            )
                           .filter(pl.col('count')==pl.col('max_count')
                            )
                           .select(pl.col('dates'),
```

```
                                    pl.col('item')
                                )
)
print(f'item most frequented for each date using Polars:')
polars_counts
```

```
item most frequented for each date using Polars:
```

shape: (3, 2)

dates	item
str	str
"02-JAN-20"	"pear"
"01-JAN-20"	"apple"
"01-JAN-20"	"pear"