In [1]:
```
import pandas as pd
from pyspark.sql import SparkSession
import findspark
import numpy as np
```

In [2]:
```
findspark.init()
findspark.find()
```

Out[2]:
```
'/Users/youngjinseo/anaconda3/lib/python3.10/site-packages/pyspark'
```

In [3]:
```
spark = SparkSession.builder.appName("practice2").config("spark.driver.bindAddress", "127.0.0.1").getOrCreate()
spark
```

```
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
24/10/02 19:22:18 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

Out[3]:
SparkSession - in-memory

SparkContext

Spark UI

Version	v3.5.2
Master	local[*]
AppName	practice2

In [4]:
```
df = spark.read.csv("/Users/youngjinseo/Desktop/파이썬/titanic.csv",header = True )
df.show()
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          892|       0|       3|                    Kelly, Mr. James|  male|34.5|    0|    0|          330911| 7.8292|   NULL|       Q|
|          893|       1|       3|                Wilkes, Mrs. Jame...|female|47.0|    1|    0|          363272|    7|   NULL|       S|
|          894|       0|       2|                    Myles, Mr. Thomas...|  male|62.0|    0|    0|          240276| 9.6875|   NULL|       Q|
|          895|       0|       3|                      Wirz, Mr. Albert|  male|27.0|    0|    0|          315154| 8.6625|   NULL|       S|
|          896|       1|       3|                Hirvonen, Mrs. Al...|female|22.0|    1|    1|         3101298|12.2875|   NULL|       S|
|          897|       0|       3|                Svensson, Mr. Joh...|  male|14.0|    0|    0|            7538| 9.225|   NULL|       S|
|          898|       1|       3|                Connolly, Miss. Kate|female|30.0|    0|    0|          330972| 7.6292|   NULL|       Q|
|          899|       0|       2|                Caldwell, Mr. Alb...|  male|26.0|    1|    1|          248738|   29|   NULL|       S|
|          900|       1|       3|                Abrahim, Mrs. Jos...|female|18.0|    0|    0|            2657| 7.2292|   NULL|       C|
|          901|       0|       3|                  Davies, Mr. John...|  male|21.0|    2|    0|         A/4 48871|   24.15|   NULL|       S|
|          902|       0|       3|                    Ilieff, Mr. Yio|  male|NULL|    0|    0|          349220| 7.8958|   NULL|       S|
|          903|       0|       1|                Jones, Mr. Charle...|  male|46.0|    0|    0|             694|   26|   NULL|       S|
|          904|       1|       1|                Snyder, Mrs. John...|female|23.0|    1|    0|         21228|82.2667|    B45|       S|
|          905|       0|       2|                Howard, Mr. Benjamin|  male|63.0|    1|    0|           24065|   26|   NULL|       S|
|          906|       1|       1|                Chaffee, Mrs. Her...|female|47.0|    1|    0|   W.E.P. 5734|61.175|    E31|       S|
|          907|       1|       2|                del Carlo, Mrs. S...|female|24.0|    1|    0|  SC/PARIS 2167|27.7208|   NULL|       C|
|          908|       0|       2|                    Keane, Mr. Daniel|  male|35.0|    0|    0|          233734|12.35|   NULL|       Q|
|          909|       0|       3|                Assaf Khalil, Mr...|  male|45.0|    0|    0|            2696| 7.225|   NULL|       C|
|          910|       1|       3|                Boulos, Mr. Hanna|  male|NULL|    0|    0|PC 17558|9.4833|   NULL|       C|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 20 rows
```

다양한 조건문

In [5]:
```
# female 관련 데이터 가져오기
df.filter(df.Sex != "male").show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          893|       1|       3|                Wilkes, Mrs. Jame...|female|47.0|    1|    0|          363272|    7|   NULL|       S|
|          896|       1|       3|                Hirvonen, Mrs. Al...|female|22.0|    1|    1|         3101298|12.2875|   NULL|       S|
|          898|       1|       3|                Connolly, Miss. Kate|female|30.0|    0|    0|          330972| 7.6292|   NULL|       Q|
|          900|       1|       3|                Abrahim, Mrs. Jos...|female|18.0|    0|    0|            2657| 7.2292|   NULL|       C|
|          904|       1|       1|                Snyder, Mrs. John...|female|23.0|    1|    0|         21228|82.2667|    B45|       S|
|          906|       1|       1|                Chaffee, Mrs. Her...|female|47.0|    1|    0|   W.E.P. 5734|61.175|    E31|       S|
|          907|       1|       2|                del Carlo, Mrs. S...|female|24.0|    1|    0|  SC/PARIS 2167|27.7208|   NULL|       C|
|          908|       0|       2|                    Keane, Mr. Daniel|  male|35.0|    0|    0|          233734|12.35|   NULL|       Q|
|          909|       0|       3|                Assaf Khalil, Mr...|  male|45.0|    0|    0|            2696| 7.2250|   NULL|       C|
|          912|       0|       1|                Rothschild, Mr. M...|  male|55.0|    1|    0|        PC 17603| 59.4|   NULL|       C|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 10 rows
```

In [6]:
```
# Age 30보다 같거나 큰 데이터 추출
df.filter(df.Age >= 30).show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          892|       0|       3|                    Kelly, Mr. James|  male|34.5|    0|    0|          330911| 7.8292|   NULL|       Q|
|          893|       1|       3|                Wilkes, Mrs. Jame...|female|47.0|    1|    0|          363272|    7|   NULL|       S|
|          894|       0|       2|                    Myles, Mr. Thomas...|  male|62.0|    0|    0|          240276| 9.6875|   NULL|       Q|
|          895|       0|       3|                      Wirz, Mr. Albert|  male|27.0|    0|    0|          315154| 8.6625|   NULL|       S|
|          896|       1|       3|                Hirvonen, Mrs. Al...|female|22.0|    1|    1|         3101298|12.2875|   NULL|       S|
|          897|       0|       3|                Svensson, Mr. Joh...|  male|14.0|    0|    0|            7538| 9.225|   NULL|       S|
|          898|       1|       3|                Connolly, Miss. Kate|female|30.0|    0|    0|          330972| 7.6292|   NULL|       Q|
|          899|       0|       2|                Caldwell, Mr. Alb...|  male|26.0|    1|    1|          248738|   29|   NULL|       S|
|          900|       1|       3|                Abrahim, Mrs. Jos...|female|18.0|    0|    0|            2657| 7.2292|   NULL|       C|
|          901|       0|       3|                  Davies, Mr. John...|  male|21.0|    2|    0|         A/4 48871|   24.15|   NULL|       S|
|          902|       0|       3|                    Ilieff, Mr. Yio|  male|NULL|    0|    0|          349220| 7.8958|   NULL|       S|
|          903|       0|       1|                Jones, Mr. Charle...|  male|46.0|    0|    0|             694|   26|   NULL|       S|
|          904|       1|       1|                Snyder, Mrs. John...|female|23.0|    1|    0|         21228|82.2667|    B45|       S|
|          905|       0|       2|                Howard, Mr. Benjamin|  male|63.0|    1|    0|           24065|   26|   NULL|       S|
|          906|       1|       1|                Chaffee, Mrs. Her...|female|47.0|    1|    0|   W.E.P. 5734|61.175|    E31|       S|
|          907|       1|       2|                del Carlo, Mrs. S...|female|24.0|    1|    0|  SC/PARIS 2167|27.7208|   NULL|       C|
|          908|       0|       2|                    Keane, Mr. Daniel|  male|35.0|    0|    0|          233734|12.35|   NULL|       Q|
|          909|       0|       3|                Assaf Khalil, Mr...|  male|45.0|    0|    0|            2696| 7.2250|   NULL|       C|
|          910|       1|       3|                Boulos, Mr. Hanna|  male|NULL|    0|    0|PC 17558|9.4833|   NULL|       C|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 20 rows
```

조건문의 and or 결합

In [7]:
```
# 성별은 여자 그리고 나이는 30보다 작은 데이터 추출
df.filter((df.Sex == "female")&(df.Age >= 30)).show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          896|       1|       3|                Hirvonen, Mrs. Al...|female|22.0|    1|    1|         3101298|12.2875|   NULL|       S|
|          898|       1|       3|                Connolly, Miss. Kate|female|30.0|    0|    0|          330972| 7.6292|   NULL|       Q|
|          900|       1|       3|                Abrahim, Mrs. Jos...|female|18.0|    0|    0|            2657| 7.2292|   NULL|       C|
|          904|       1|       1|                Snyder, Mrs. John...|female|23.0|    1|    0|         21228|82.2667|    B45|       S|
|          907|       1|       2|                del Carlo, Mrs. S...|female|24.0|    1|    0|  SC/PARIS 2167|27.7208|   NULL|       C|
|          915|       1|       3|                Dimitrano, Miss...|female|27.0|    1|    0|       STON/O2. 3101270|   7.925|   NULL|       S|
|          916|       1|       1|                Ryerson, Mrs. Art...|female|48.0|    1|    3|        PC 17608|262.375|B57 B59 B63 B66|       C|
|          917|       0|       3|                Dean, Mrs. Bertra...|female|33.0|    1|    2|         C.A. 2315| 20.575|   NULL|       S|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 10 rows
```

In [8]:
```
# Embarked가 C라고 나이는 30보다 같거나 작은 데이터 추출
df.filter((df.Embarked == "C")&(df.Age <= 20)).show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          897|       0|       3|                Svensson, Mr. Joh...|  male|14.0|    0|    0|            7538| 9.2250|   NULL|       S|
|          924|       1|       3|                    Dean, Mr. Bertra...|  male|NULL|    1|    2|         C.A. 2315| 20.575|   NULL|       S|
|          942|       1|       1|                Smith, Mr. Lucien...|  male|24.0|    1|    0|           13695|   60|   C31|       S|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 10 rows
```

SQL의 like 문장 사용

In [9]:
```
# Name에서 L로 시작하는 애들
df.filter(df.Name.like("%L%")).show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          897|       0|       3|                Svensson, Mr. Joh...|  male|14.0|    0|    0|            7538| 9.2250|   NULL|       S|
|          924|       1|       3|                    Dean, Mr. Bertra...|  male|NULL|    1|    2|         C.A. 2315| 20.575|   NULL|       S|
|         1001|       0|       2|                Swane, Mr. George|  male|18.5|    0|    0|          248734|   13|    F|       S|
|         1050|       1|       3|                Sandstrom, Miss. ...|female|  1.|    1|    1|          PP 9549| 16.7|   G6|       S|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 10 rows
```

In [10]:
```
# Name에서 중간에 "Mrs." 포함된 데이터 추출
df.filter(df.Name.like("%Mrs.%")).show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          893|       1|       3|                Wilkes, Mrs. Jame...|female|47.0|    1|    0|          363272|    7|   NULL|       S|
|          896|       1|       3|                Hirvonen, Mrs. Al...|female|22.0|    1|    1|         3101298|12.2875|   NULL|       S|
|          900|       1|       3|                Abrahim, Mrs. Jos...|female|18.0|    0|    0|            2657| 7.2292|   NULL|       C|
|          904|       1|       1|                Snyder, Mrs. John...|female|23.0|    1|    0|         21228|82.2667|    B45|       S|
|          906|       1|       1|                Chaffee, Mrs. Her...|female|47.0|    1|    0|   W.E.P. 5734|61.175|    E31|       S|
|          907|       1|       2|                del Carlo, Mrs. S...|female|24.0|    1|    0|  SC/PARIS 2167|27.7208|   NULL|       C|
|          915|       1|       3|                Dimitrano, Miss...|female|27.0|    1|    0|       STON/O2. 3101270|   7.925|   NULL|       S|
|          916|       1|       1|                Ryerson, Mrs. Art...|female|48.0|    1|    3|        PC 17608|262.375|B57 B59 B63 B66|       C|
|          917|       0|       3|                Dean, Mrs. Bertra...|female|33.0|    1|    2|         C.A. 2315| 20.575|   NULL|       S|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 10 rows
```

df의 값이 바깥에 존재하는 list 내에 존재하는지 확인

In [11]:
```
# 특정 칼럼 한 것 값을 밸류한거 데이터프레임에서 12개의 레코드를 보여주는 코드
# "S"와 "C" 값이 포함된 이스트 embark를 생성
embark = df[df.Embarked.isin(embark).show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          893|       1|       3|                Wilkes, Mrs. Jame...|female|47.0|    1|    0|          363272|    7|   NULL|       S|
|          895|       0|       3|                      Wirz, Mr. Albert|  male|27.0|    0|    0|          315154| 8.6625|   NULL|       S|
|          896|       1|       3|                Hirvonen, Mrs. Al...|female|22.0|    1|    1|         3101298|12.2875|   NULL|       S|
|          897|       0|       3|                Svensson, Mr. Joh...|  male|14.0|    0|    0|            7538| 9.2250|   NULL|       S|
|          899|       0|       2|                Caldwell, Mr. Alb...|  male|26.0|    1|    1|          248738|   29|   NULL|       S|
|          900|       1|       3|                Abrahim, Mrs. Jos...|female|18.0|    0|    0|            2657| 7.2292|   NULL|       C|
|          901|       0|       3|                  Davies, Mr. John...|  male|21.0|    2|    0|         A/4 48871|   24.15|   NULL|       S|
|          902|       0|       3|                    Ilieff, Mr. Yio|  male|NULL|    0|    0|          349220| 7.8958|   NULL|       S|
|          903|       0|       1|                Jones, Mr. Charle...|  male|46.0|    0|    0|             694|   26|   NULL|       S|
|          904|       1|       1|                Snyder, Mrs. John...|female|23.0|    1|    0|         21228|82.2667|    B45|       S|
|          905|       0|       2|                Howard, Mr. Benjamin|  male|63.0|    1|    0|           24065|   26|   NULL|       S|
|          906|       1|       1|                Chaffee, Mrs. Her...|female|47.0|    1|    0|   W.E.P. 5734|61.175|    E31|       S|
|          907|       1|       2|                del Carlo, Mrs. S...|female|24.0|    1|    0|  SC/PARIS 2167|27.7208|   NULL|       C|
|          909|       0|       3|                Assaf Khalil, Mr...|  male|45.0|    0|    0|            2696| 7.2250|   NULL|       C|
|          910|       1|       3|                Boulos, Mr. Hanna|  male|NULL|    0|    0|PC 17558|9.4833|   NULL|       C|
|          912|       0|       1|                Rothschild, Mr. M...|  male|55.0|    1|    0|        PC 17603| 59.4|   NULL|       C|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 10 rows
```

조건문으로 시작하는지, 끝나는지 아니면 조건문장을 포함하는지 확인

In [12]:
```
# M이 시작한 이름 추출
df.filter(df.Name.startswith("M*") ).show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          894|       0|       2|                    Myles, Mr. Thomas...|  male|62.0|    0|    0|          240276| 9.6875|   NULL|       Q|
|          926|       0|       1|                Moch, Mr. Philipp...|  male|30.0|    1|    0|           13236| 57.75|    C78|       C|
|          946|       0|       2|                Hangeldi, Mr. ...|  male|NULL|    0|    0|   SC/A.3 2861|15.5792|   NULL|       C|
|          953|       0|       2|                McCrae, Mr. Arthu...|  male|32.0|    0|    0|          237216|  13.5|   NULL|       S|
|          955|       0|       3|                Moubarek, Master. ...|  male|NULL|    1|    1|            2661| 15.2458|   NULL|       C|
|          956|       1|       3|                Makinen, Mr. Kall...|  male|29.0|    0|    0|          312613|   7.75|   NULL|       Q|
|          963|       0|       3|                  Mishoiff, Mr. Lazar|  male|21.0|    0|    0|          349211| 7.8958|   NULL|       S|
|          968|       0|       3|                    Milani, Mr. Frank|  male|NULL|    0|    0|          359066| 8.05|   NULL|       S|
|          989|       0|       3|                Makinen, Mr. Kall...|  male|29.0|    0|    0|STON/O 2. 3101288|   7.925|   NULL|       S|
|         1050|       1|       3|                McCoy, Miss. Alicia|female|NULL|    2|    0|          367226| 23.25|   NULL|       Q|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 10 rows
```

In [13]:
```
# 글자가 s로 시작하는 데이터 추출
df.filter(df.Name.endswith("s") ).show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          892|       0|       3|                    Kelly, Mr. James|  male|34.5|    0|    0|          330911| 7.8292|   NULL|       Q|
|          894|       0|       2|                    Myles, Mr. Thomas...|  male|62.0|    0|    0|          240276| 9.6875|   NULL|       Q|
|          899|       0|       2|                Caldwell, Mr. Alb...|  male|26.0|    1|    1|          248738|   29|   NULL|       S|
|          900|       1|       3|                Abrahim, Mrs. Jos...|female|18.0|    0|    0|            2657| 7.2292|   NULL|       C|
|          921|       0|       3|                  Samaan, Mr. Elias|  male|NULL|    2|    0|           2662|21.6792|   NULL|       C|
|          923|       0|       2|                Jefferys, Mr. Cli...|  male|24.0|    2|    0|          C.A. 31029|  31.5|   NULL|       S|
|          930|       0|       3|                    Sap, Mr. Julius|  male|25.0|    0|    0|          345768|   9.5|   NULL|       S|
|          935|       1|       2|                McPeake, Mr. Geor...|  male|61.0|    0|    0|STON/O 2. 3101279|   15.85|    E45|       S|
|          938|       0|       1|                Cherry, Mr. Paul ...|  male|45.0|    0|    0|           24744|   26|   NULL|       S|
|          975|       0|       3|                Khalil, Mr. Betros|  male|NULL|    1|    0|           2660|14.4542|   NULL|       C|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 10 rows
```

In [14]:
```
# Mrs 포함된 데이터 추출
df.filter(df.Name.contains("Mrs.") ).show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          893|       1|       3|                Wilkes, Mrs. Jame...|female|47.0|    1|    0|          363272|    7|   NULL|       S|
|          896|       1|       3|                Hirvonen, Mrs. Al...|female|22.0|    1|    1|         3101298|12.2875|   NULL|       S|
|          900|       1|       3|                Abrahim, Mrs. Jos...|female|18.0|    0|    0|            2657| 7.2292|   NULL|       C|
|          904|       1|       1|                Snyder, Mrs. John...|female|23.0|    1|    0|         21228|82.2667|    B45|       S|
|          906|       1|       1|                Chaffee, Mrs. Her...|female|47.0|    1|    0|   W.E.P. 5734|61.175|    E31|       S|
|          907|       1|       2|                del Carlo, Mrs. S...|female|24.0|    1|    0|  SC/PARIS 2167|27.7208|   NULL|       C|
|          909|       0|       3|                Assaf Khalil, Mr...|  male|45.0|    0|    0|            2696| 7.2250|   NULL|       C|
|          914|       1|       1|                Flegenheim, Mrs. ...|female|NULL|    0|    0|        PC 17598|31.6833|   NULL|       S|
|          916|       1|       1|                Ryerson, Mrs. Art...|female|48.0|    1|    3|        PC 17608|262.375|B57 B59 B63 B66|       C|
|          924|       1|       3|                Dean, Mrs. Bertra...|female|33.0|    1|    2|         C.A. 2315| 20.575|   NULL|       S|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 10 rows
```

Null 값 & 아닌 값 추출

In [15]:
```
#결측값 데이터 추출
df.filter(df.Cabin.isNull()).show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          892|       0|       3|                    Kelly, Mr. James|  male|34.5|    0|    0|          330911| 7.8292|   NULL|       Q|
|          893|       1|       3|                Wilkes, Mrs. Jame...|female|47.0|    1|    0|          363272|    7|   NULL|       S|
|          895|       0|       3|                      Wirz, Mr. Albert|  male|27.0|    0|    0|          315154| 8.6625|   NULL|       S|
|          896|       1|       3|                Hirvonen, Mrs. Al...|female|22.0|    1|    1|         3101298|12.2875|   NULL|       S|
|          897|       0|       3|                Svensson, Mr. Joh...|  male|14.0|    0|    0|            7538| 9.2250|   NULL|       S|
|          898|       1|       3|                Connolly, Miss. Kate|female|30.0|    0|    0|          330972| 7.6292|   NULL|       Q|
|          899|       0|       2|                Caldwell, Mr. Alb...|  male|26.0|    1|    1|          248738|   29|   NULL|       S|
|          900|       1|       3|                Abrahim, Mrs. Jos...|female|18.0|    0|    0|            2657| 7.2292|   NULL|       C|
|          901|       0|       3|                  Davies, Mr. John...|  male|21.0|    2|    0|         A/4 48871|   24.15|   NULL|       S|
|          902|       0|       3|                    Ilieff, Mr. Yio|  male|NULL|    0|    0|          349220| 7.8958|   NULL|       S|
|          903|       0|       1|                Jones, Mr. Charle...|  male|46.0|    0|    0|             694|   26|   NULL|       S|
|          904|       1|       1|                Snyder, Mrs. John...|female|23.0|    1|    0|         21228|82.2667|    B45|       S|
|          905|       0|       2|                Howard, Mr. Benjamin|  male|63.0|    1|    0|           24065|   26|   NULL|       S|
|          906|       1|       1|                Chaffee, Mrs. Her...|female|47.0|    1|    0|   W.E.P. 5734|61.175|    E31|       S|
|          907|       1|       2|                del Carlo, Mrs. S...|female|24.0|    1|    0|  SC/PARIS 2167|27.7208|   NULL|       C|
|          908|       0|       2|                    Keane, Mr. Daniel|  male|35.0|    0|    0|          233734|12.35|   NULL|       Q|
|          909|       0|       3|                Assaf Khalil, Mr...|  male|45.0|    0|    0|            2696| 7.2250|   NULL|       C|
|          910|       1|       3|                Boulos, Mr. Hanna|  male|NULL|    0|    0|PC 17558|9.4833|   NULL|       C|
|          912|       0|       1|                Rothschild, Mr. M...|  male|55.0|    1|    0|        PC 17603| 59.4|   NULL|       C|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 20 rows
```

In [16]:
```
#결측값 포함되지 않는 데이터 추출
df.filter(df.Cabin.isNotNull()).show(10)
```

```
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|PassengerId|Survived|Pclass|                                   Name|   Sex|Age|SibSp|Parch|          Ticket|   Fare|Cabin|Embarked|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
|          904|       1|       1|                Snyder, Mrs. John...|female|23.0|    1|    0|         21228|82.2667|    B45|       S|
|          906|       1|       1|                Chaffee, Mrs. Her...|female|47.0|    1|    0|   W.E.P. 5734|61.175|    E31|       S|
|          916|       1|       1|                Ryerson, Mrs. Art...|female|48.0|    1|    3|        PC 17608|262.375|B57 B59 B63 B66|       C|
|          959|       1|       1|                Doling, Miss. Bele...|female|22.0|    0|    1|         12749| 93.5|    B69|       S|
|          926|       0|       1|                Moch, Mr. Philipp...|  male|45.0|    0|    0|         11036|  30.5|    A23|       S|
|          933|       0|       1|                Franklin, Mr. Tho...|  male|NULL|    0|    0|          113778| 26.55|    D34|       S|
|          946|       0|       2|                Hangeldi, Mr. ...|  male|NULL|    0|    0|   SC/A.3 2861|15.5792|   NULL|       C|
|          938|       0|       1|                Cherry, Mr. Paul ...|  male|45.0|    0|    0|        PC 17594|  29.7|    A9|       C|
|          959|       1|       1|                Buckland, Mrs. M...|female|60.0|    1|    0|         112139|26.2917|    D37|       C|
+---------------+--------+--------+------------------------------------------+------+---+-----+-----+----------------+-------+-----------+--------+
only showing top 10 rows
```

유일한 값만 필터링

In [17]:
```
# "Survived" 열에서 유일한 값을 제거하고, 고유한 값만 출력하는 코드
df.select("Survived").distinct().show(19)
```

```
+--------+
|Survived|
+--------+
|       0|
|       1|
+--------+
```

기타

In [18]:
```
# # "Name", "Sex", "Age", "Cabin" 열만 선택하고, 나이가 22 이하인 승객을 필터링한 후, 10개의 레코드를 출력하는 코드
df.select("Name", "Sex", "Age", "Cabin").filter(df.Age <= 22).show(10)
```

```
+--------------------+------+---+-----+
|                Name|   Sex|Age|Cabin|
+--------------------+------+---+-----+
|Svensson, Mr. Joh...|  male|14.0| NULL|
|Abrahim, Mrs. Jos...|female|18.0| NULL|
|Olsen, Master. Ar...|  male| 9.0| NULL|
|Makinen, Mr. Kall...|  male|29.0| NULL|
|Moubarek, Master. ...|  male|NULL| NULL|
|Kink, Master. Murd...|  male|10.0| NULL|
|    Dika, Mr. Mirko|  male|17.0| NULL|
|Mishoiff, Mr. Laz...|  male|21.0| NULL|
|Ryerson, Master. ...|  male|13.0|B57 B59 B63 B66|
|Buss, Miss. Mary ...|female|18.0| NULL|
+--------------------+------+---+-----+
only showing top 10 rows
```

In [19]:
```
# "Name", "Age", "Survived" 열만 선택하고, 생존한 승객을 필터링한 후 10개의 레코드를 출력하는 코드
df.select("Name","Age","Survived").filter(df.Survived==1").show(10)
```

```
+--------------------+---+--------+
|                Name|Age|Survived|
+--------------------+---+--------+
|Wilkes, Mrs. Jame...|47.0|       1|
|Hirvonen, Mrs. Al...|22.0|       1|
|Connolly, Miss. Kate|30.0|       1|
|Abrahim, Mrs. Jos...|18.0|       1|
|Snyder, Mrs. John...|23.0|       1|
|Chaffee, Mrs. Her...|47.0|       1|
|del Carlo, Mrs. S...|24.0|       1|
|Dimitrano, Miss. ...|27.0|       1|
|Assaf Khalil, Mr...|45.0|       1|
|Flegenheim, Mrs. ...|NULL|       1|
+--------------------+---+--------+
only showing top 10 rows
```