Data Engineering Batch 1   **Day 15:10/02/2024**

Name: Pradip Bochare

🪽 Creating DataFrames and operations

```
|   Jackly|
+-----------+
only showing top 20 rows
```

In [38]: `df.select(df['name'], df['mark'] + 1).show()`

```
+-----------+----------+
|       name|(mark + 1)|
+-----------+----------+
|   John Deo|        76|
|   Max Ruin|        86|
|     Arnold|        56|
| Krish Star|        61|
|  John Mike|        61|
|  Alex John|        56|
|My John Rob|        79|
|     Asruid|        86|
|    Tes Qry|        79|
|   Big John|        56|
|     Ronald|        90|
|      Recky|        95|
|        Kty|        89|
|       Bigy|        89|
|   Tade Row|        89|
|      Gimmy|        89|
|      Tumyu|        55|
|      Honny|        76|
|      Tinny|        19|
|     Jackly|        66|
+-----------+----------+
only showing top 20 rows
```

In [41]: `df.filter(df['mark'] > 80).show()`

In [41]: `df.filter(df['mark'] > 80).show()`

```
+---+----------+-----+----+------+
| id|      name|class|mark|gender|
+---+----------+-----+----+------+
|  2|  Max Ruin|Three|  85|  male|
|  8|    Asruid| Five|  85|  male|
| 11|    Ronald|  Six|  89|female|
| 12|     Recky|  Six|  94|female|
| 13|       Kty|Seven|  88|female|
| 14|      Bigy|Seven|  88|female|
| 15|  Tade Row| Four|  88|  male|
| 16|     Gimmy| Four|  88|  male|
| 25|  Giff Tow|Seven|  88|  male|
| 27|  Big Nose|Three|  81|female|
| 28| Rojj Base|Seven|  86|female|
| 31|Marry Toeey| Four|  88|female|
| 32| Binn Rott|Seven|  90|female|
| 33| Kenn Rein|  Six|  96|female|
| 35| Rows Noump|  Six|  88|female|
+---+----------+-----+----+------+
```

In [42]: `df.groupBy("class").count().show()`

```
+-----+-----+
|class|count|
+-----+-----+
| Five|    2|
|  Six|    7|
|Eight|    1|
| Four|    9|
|Seven|   10|
| Nine|    2|
|Three|    3|
|Fifth|    1|
```

File | Edit | View | Insert | Cell | Kernel | Widgets | Help   Trusted | Python 3 (ipykernel) O

```
In [42]: df.groupBy("class").count().show()

         +-----+-----+
         |class|count|
         +-----+-----+
         | Five|    2|
         |  Six|    7|
         |Eight|    1|
         | Four|    9|
         |Seven|   10|
         | Nine|    2|
         |Three|    3|
         |Fifth|    1|
         +-----+-----+
```

```
In [43]: df.createOrReplaceTempView("student")
```

```
In [45]: sqlDF = spark.sql("SELECT * FROM student")
         sqlDF.show()

         +---+----------+-----+----+------+
         | id|      name|class|mark|gender|
         +---+----------+-----+----+------+
         |  1|  John Deo| Four|  75|female|
         |  2|  Max Ruin|Three|  85|  male|
         |  3|    Arnold|Three|  55|  male|
         |  4|Krish Star| Four|  60|female|
         |  5| John Mike| Four|  60|female|
         |  6| Alex John| Four|  55|  male|
         |  7|My John Rob|Fifth| 78|  male|
         |  8|    Asruid| Five|  85|  male|
         |  9|   Tes Qry|  Six|  78|  male|
         | 10| Big John| Four|  55|female|
         | 11|    Ronald|  Six|  89|female|
```

File | Edit | View | Insert | Cell | Kernel | Widgets | Help   Trusted | Python 3 (ipykernel) O

```
In [43]: df.createOrReplaceTempView("student")
```

```
In [45]: sqlDF = spark.sql("SELECT * FROM student")
         sqlDF.show()

         +---+----------+-----+----+------+
         | id|      name|class|mark|gender|
         +---+----------+-----+----+------+
         |  1|  John Deo| Four|  75|female|
         |  2|  Max Ruin|Three|  85|  male|
         |  3|    Arnold|Three|  55|  male|
         |  4|Krish Star| Four|  60|female|
         |  5| John Mike| Four|  60|female|
         |  6| Alex John| Four|  55|  male|
         |  7|My John Rob|Fifth| 78|  male|
         |  8|    Asruid| Five|  85|  male|
         |  9|   Tes Qry|  Six|  78|  male|
         | 10| Big John| Four|  55|female|
         | 11|    Ronald|  Six|  89|female|
         | 12|     Recky|  Six|  94|female|
         | 13|       Kty|Seven|  88|female|
         | 14|      Bigy|Seven|  88|female|
         | 15|  Tade Row| Four|  88|  male|
         | 16|     Gimmy| Four|  88|  male|
         | 17|     Tumyu|  Six|  54|  male|
         | 18|     Honny| Five|  75|  male|
         | 19|     Tinny| Nine|  18|  male|
         | 20|    Jackly| Nine|  65|female|
         +---+----------+-----+----+------+
         only showing top 20 rows
```

```
In [46]: df.createGlobalTempView("student")
```

```
In [47]: spark.sql("SELECT * FROM global_temp.student").show()
```

```
In [46]: df.createGlobalTempView("student")
```

```
In [47]: spark.sql("SELECT * FROM global_temp.student").show()
```

```
+---+----------+-----+----+------+
| id|      name|class|mark|gender|
+---+----------+-----+----+------+
|  1|  John Deo| Four|  75|female|
|  2|  Max Ruin|Three|  85|  male|
|  3|    Arnold|Three|  55|  male|
|  4|Krish Star| Four|  60|female|
|  5| John Mike| Four|  60|female|
|  6| Alex John| Four|  55|  male|
|  7|My John Rob|Fifth| 78|  male|
|  8|    Asruid| Five|  85|  male|
|  9|   Tes Qry|  Six|  78|  male|
| 10|  Big John| Four|  55|female|
| 11|    Ronald|  Six|  89|female|
| 12|     Recky|  Six|  94|female|
| 13|       Kty|Seven|  88|female|
| 14|      Bigy|Seven|  88|female|
| 15|  Tade Row| Four|  88|  male|
| 16|     Gimmy| Four|  88|  male|
| 17|     Tumyu|  Six|  54|  male|
| 18|     Honny| Five|  75|  male|
| 19|     Tinny| Nine|  18|  male|
| 20|    Jackly| Nine|  65|female|
+---+----------+-----+----+------+
only showing top 20 rows
```

```
In [ ]:
```