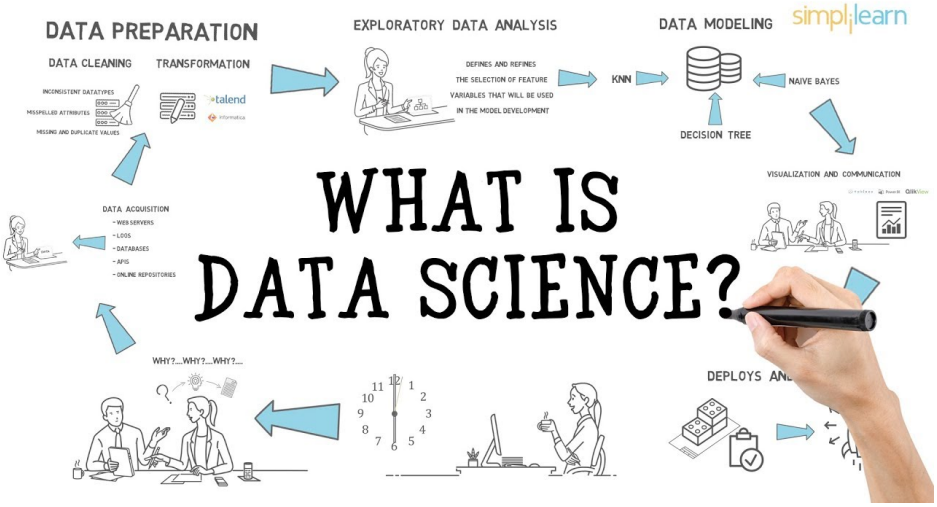

Big Data Systems

REDIS - MONGODB

ASSIGNMENT

TASK 1 | RESULTS

1.1. How many users modified their listing on January?

```
> r$BITCOUNT("ModificationsJanuary")
[1] 9969
```

1.2. How many users did NOT modify their listing on January?

Note that the sum of the numbers occurred from task 1.1 and 1.2 (20,000) do not match the total number of users (19,999). This happens because all BITOP operations occur at byte-level increments. Each byte has 8 bits and the results of a BITOP operation should always be an integer multiple of 8.

```
> r$BITCOUNT("results1")
[1] 10031
```

1.3. How many users received at least one e-mail per month (at least one e-mail in January and at least one e-mail in February and at least one e-mail in March)?

```
> r$BITCOUNT("results2")
[1] 16006
```

1.4. How many users received an e-mail on January and March but NOT on February?

```
> r$BITCOUNT("combined result")
[1] 2417
```

1.5. How many users received an e-mail on January that they did not open but they updated their listing anyway?

```
> r$BITCOUNT("EmailsOpenedJanuary")
[1] 2807
>
```

1.6. How many users received an e-mail on January that they did not open but they updated their listing anyway on January OR they received an e-mail on February that they did not open but they updated their listing

anyway on February OR they received an e-mail on March that they did not open but they updated their listing anyway on March?

```
[1] 7221
> |
```

1.7. Does it make any sense to keep sending e-mails with recommendations to sellers? Does this strategy really work? How would you describe this in terms a business person would understand?

After analysing the data, it seems that the specific strategy of the classified provider to send personalized e-mails has a partial effect on motivating the sellers to improve their listings. This conclusion can be derived from the fact that almost 50% of these sellers during this 3-month period proceeded to a change of their listings after having received and opened the respective recommendation e-mail (January: 49.5%, February: 50.2%, March: 49.9%). Therefore, in case the provider is willing to use the specific approach as a dominant means to have earnings, what should be improved to draw the attention of the recipients of these e-mails more efficiently should be investigated.

TASK 2 | RESULTS

For this task before adding the data files in MongoDB, a function is created in R via which the below manipulation/cleaning procedure is followed for several fields:

- Motorcycles price (ad_data$Price): Initially defined as string. For the cases where:
  - A number existed in this field, it is converted to a numeric data type and the currency symbol is removed.
  - The price is relatively small (i.e. 10 euros), we consider that these do not probably consist cases of motorcycle sale (for example they could be motorcycle parts).
  - For the cases where no price is defined (i.e. AskForPrice), we introduce a counter to measure the number of these motorcycles.
- Registration date (ad_data$ Registration): Remove the month from the date field
- Milesage (ad_data$Mileage): Remove the measurement unit of the motorcycles' milesage (i.e. km)
- Advertisement age (ad_data$ Ad_age): Introduced this new field to count the age of the advertisement

2.2 How many bikes are there for sale?

```
[1] 29701
>
```

2.3 What is the average price of a motorcycle (give a number)? What is the number of listings that were used in order to calculate this average (give a number as well)? Is the number of listings used the same as the answer in 1.2? Why?

The motorcycles which are not taken into account to calculate this result are these for which no price is defined (i.e. AskForPrice = TRUE or Price=NA).

```
  _id AvgPrice count
1  NA 3007.571 28710
>
```

2.4 What is the maximum and minimum price of a motorcycle currently available in the market?

```
  _id MinPrice MaxPrice
1  NA       10    89000
>
```

2.5 How many listings have a price that is identified as negotiable?

```
  _id count
1  NA  1348
>
```

2.6 For each Brand, what percentage of its listings is listed as negotiable? (indicative results)

```
+ )
                _id TotalCount NegotiableCount NegotiableRatio
1             Adler          1               0       0.000000
2            Jinlun          1               1     100.000000
3         Bombardier          1               1     100.000000
4            Genata          2               0       0.000000
5           Barossa          1               0       0.000000
6            Qingqi          2               2     100.000000
7             Fever          1               1     100.000000
8             Ariel          2               0       0.000000
9               Aie          4               0       0.000000
10            Vedim          1               0       0.000000
11           Rewaco          3               1      33.333333
12       Super Moto          2               1      50.000000
13            Eagle          3               0       0.000000
14            Adiva          5               4      80.000000
15              Niu          1               1     100.000000
16            Maico          4               1      25.000000
17             Goes          1               0       0.000000
18          Bultaco          2               1      50.000000
19             Dias          3               3     100.000000
20           Harlow          1               0       0.000000
21         Amstrong          3               3     100.000000
22         Zuendapp          7               3      42.857143
```

2.7 What is the motorcycle brand with the highest average price?

```
    _id AvgPrice count
1 Semog    15600     1
>
```

2.8 What is the TOP 10 models with the highest average age? (Round age by one decimal number)

```
                                _id AvgAge count
1  Αλλο henderson indian replica    88.0     1
2                        Bmw R 12    85.0     1
3          Αλλο MATCHLESS G3 350    84.0     1
4                      Norton H16    83.0     1
5                  Αλλο Matsoules    81.0     1
6              Αλλο Matchless G3/L    80.0     1
7                  Αλλο NEW HUDSON    80.0     1
8              Bsa M20 ARMY MOTO!    80.0     1
9                        Bsa M20    79.5     2
10                     Norton 16H    79.0     1
```

2.9 How many bikes have “ABS” as an extra?

```
  _id count
1  NA  4025
>
```

2.10 What is the average Mileage of bikes that have “ABS” AND “Led lights” as an extra?

```
  _id AvgMilage count
1  NA   30125.7  1135
>
```

2.11 Which is the TOP 3 colors per bike category? (indicative results)

	_id	colors
1	Goes	Orange (Metallic), 1
2	Niu	Black, 1
3	Shandong Liangzi	Other, 1
4	Adler	Red, 1
5	AMS	Blue, Green, Yellow, 1, 1, 1
6	New Force Motor	Black, 1
7	Wsk	Orange, 1
8	Eagle	Red, Orange (Metallic), White (Metallic), 1, 1, 1
9	Nitro Motors	Dark blue, 1
10	Beeline	Black, 1
11	KL	Black, 1
12	Bombardier	Yellow, 1
13	Laverda	Silver, Orange, 1, 1
14	Harlow	Black (Metallic), 1
15	Morini	Light Blue (Metallic), Blue (Metallic), Red (Metallic), 1, 1, 1
16	Genata	Red, Silver, 1, 1
17	Bultaco	Dark blue, Light Blue, 1, 1
18	Vedim	Black, 1
19	E-Ton	Red, Black, 1, 1
20	Boatian	Black, Silver, Blue, 1, 1, 1
21	Kinroad	Green, Yellow, Grey, 1, 1, 1
22	Ural	White, Beige, Blue (Metallic), 1, 1, 1
23	Hartford	Dark Red, 1
24	Generic	Yellow, Orange, Black (Metallic), 1, 1, 1
25	Chang Jiang	Silver, 1
26	Odess	White, Black, 1, 1
27	Amstrong	Yellow, Red, Black, 1, 1, 1
28	Dias	Blue (Metallic), White, Red, 1, 1, 1
29	Fever	Orange, 1
30	Victory	Bordeaux (Metallic), 1
31	Buggy Motors	Black, Black (Metallic), Red, 1, 1, 1
32	Barossa	Light Blue, 1