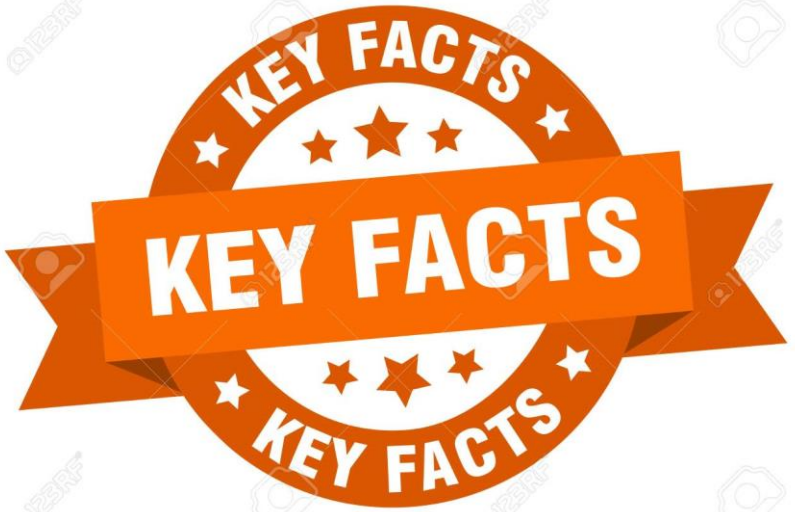
onedot
consumable data.

DATA TASK

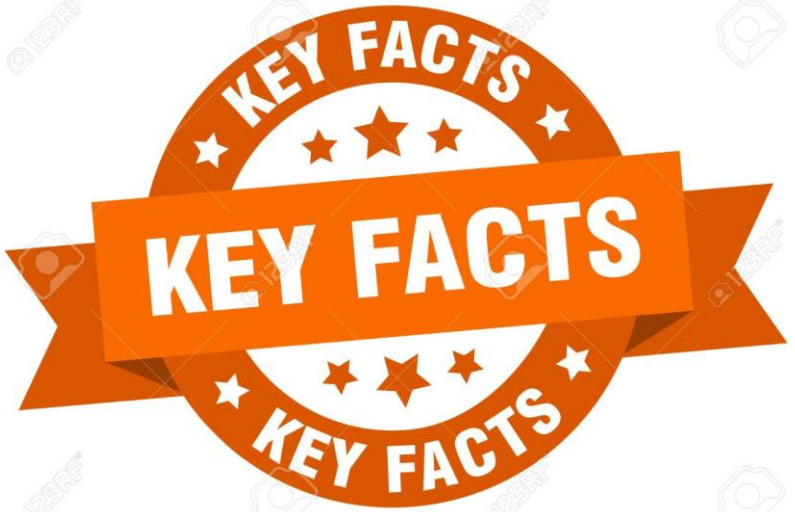
WAQAS ZAHID



KEY FACTS OF DATA INPUT / OUTPUT

PRE-PROCESSING

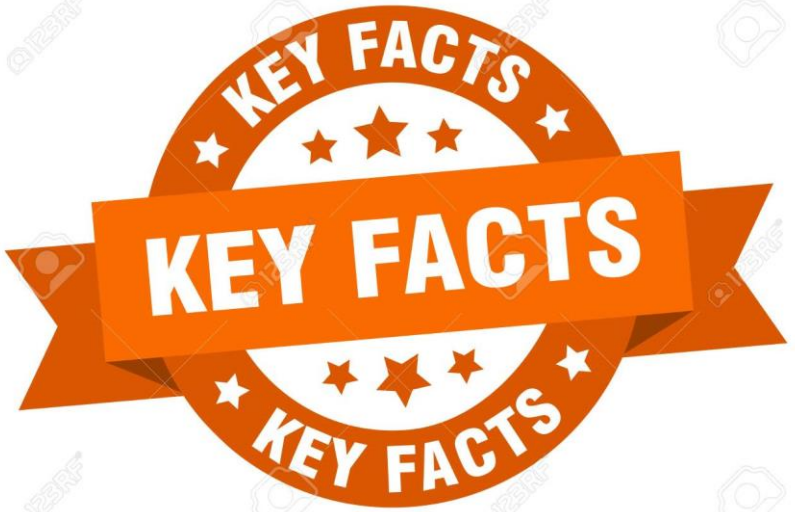
- THE RAW DATA WAS IN JSON FORMAT FILE: 'SUPPLIER_CAR.JSON'



KEY FACTS OF DATA INPUT / OUTPUT

PRE-PROCESSING.....

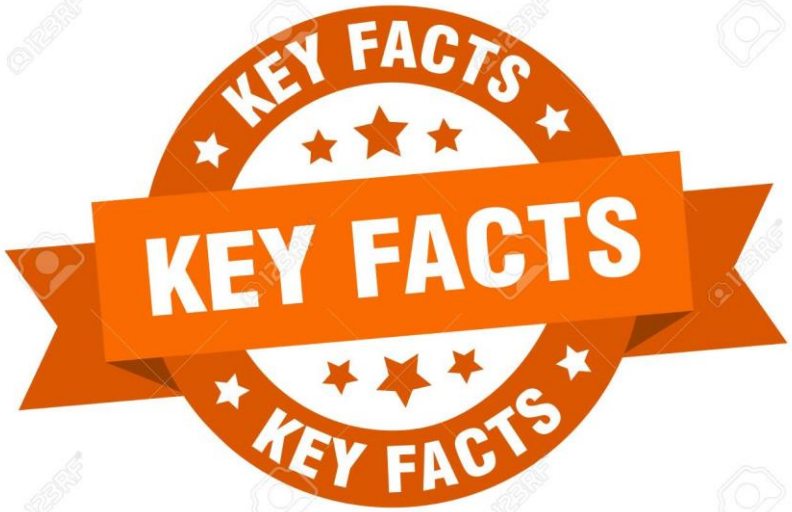
- THE JSON FILE (SUPPLIER_CAR.JSON) WENT THROUGH PRE-PROCESSING OR TRANSFORMATION PROCESS



KEY FACTS OF DATA INPUT / OUTPUT

PRE-PROCESSING.....

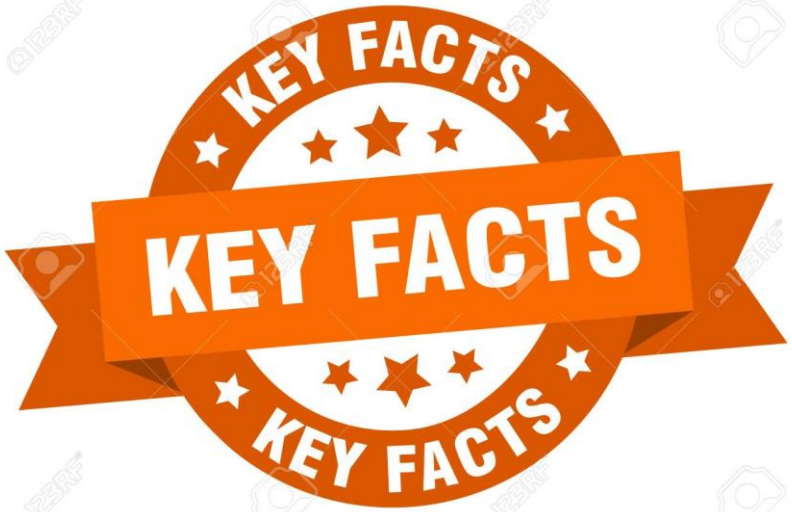
- CONVERTED JSON (SUPPLIER_CAR.JSON) DATA TO INITIAL 21,906 ROWS AND 9 COLUMNS



KEY FACTS OF DATA INPUT / OUTPUT

PRE-PROCESSING.....

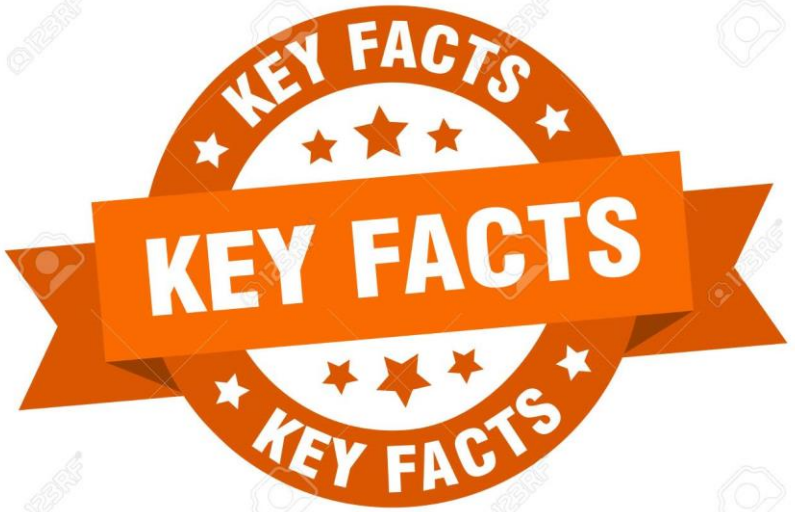
- THERE WERE MULTIPLE INPUTS (19) OF A SINGLE DATA ID



KEY FACTS OF DATA INPUT / OUTPUT

PRE-PROCESSING.....

- AFTER ANALYZING THE DATA, WE FIGURED OUT THAT THE REPETITIONS WERE THE UNIQUE ATTRIBUTES OF EACH OF THE 19 DATA ID'S

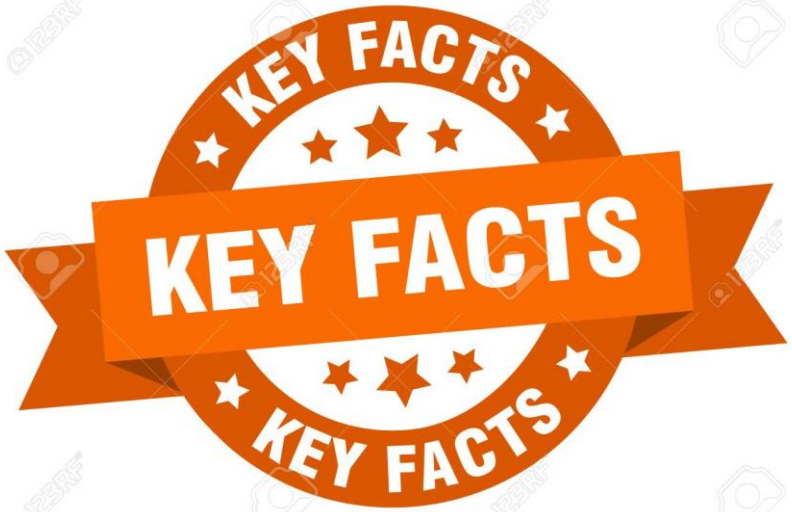


KEY FACTS OF DATA INPUT / OUTPUT

Excel

PRE-PROCESSING.....

- GENERATED AN EXCEL (OUTPUT) FILE (TARGET DATA.XLSX) WITH 3 TABS I.E. PRE-PROCESSING, NORMALISATION AND INTEGRATION, VIA PYTHON PROGRAMMING

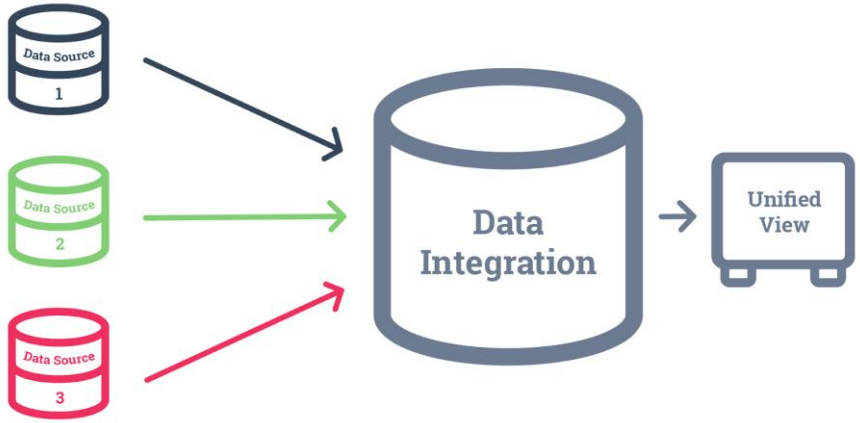


KEY FACTS OF DATA INPUT / OUTPUT

Excel

PRE-PROCESSING.....

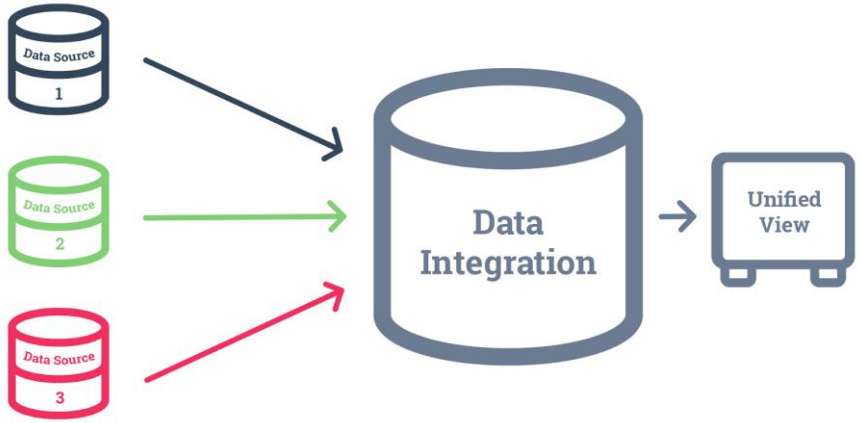
- THE OUTPUT FILE (TARGET DATA.XLSX) IS GENERATED SIMPLY BY RUNNING THE PYTHON OR JUPYTER NOTE FILE (DATA_TASK.JPYNB)



SUMMARY OF THE CHANGES MADE TO THE INPUT DATA

DATA INTEGRATION

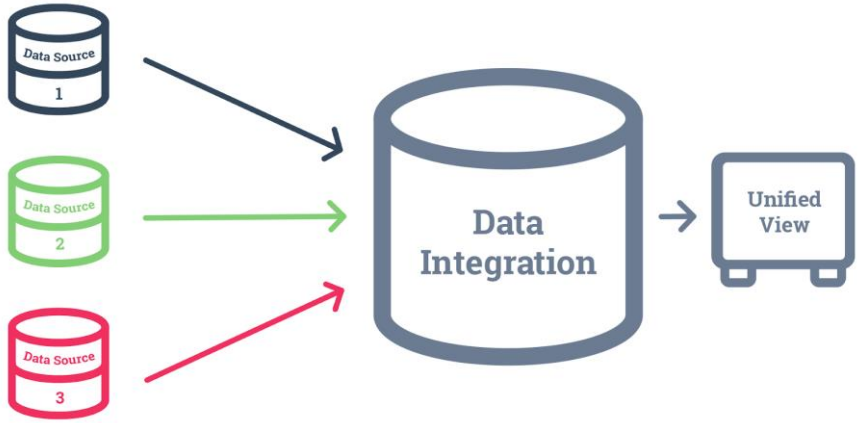
- COLLABORATED ALL THE UNIQUE ATTRIBUTES OF EACH ID IN THE FORM OF A DIRECTORY



SUMMARY OF THE CHANGES MADE TO THE INPUT DATA

DATA INTEGRATION.....

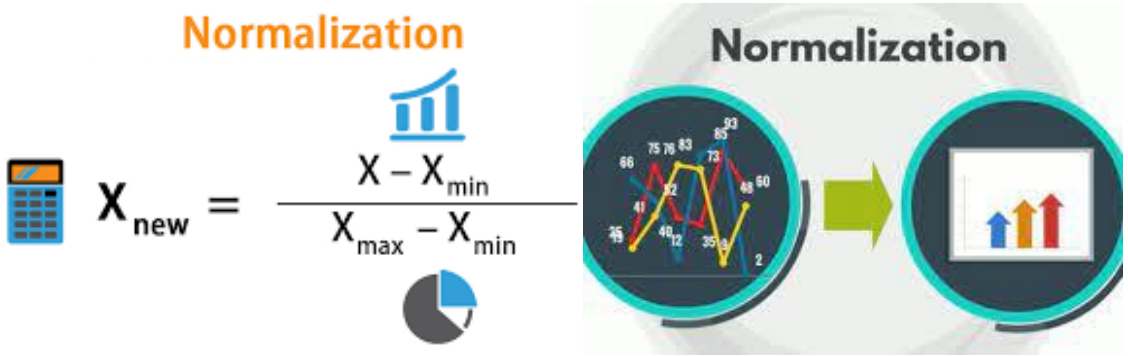
- DROPPED THE DUPLICATE DATA WITH RESPECT TO THE ID



SUMMARY OF THE CHANGES MADE TO THE INPUT DATA

DATA INTEGRATION.....

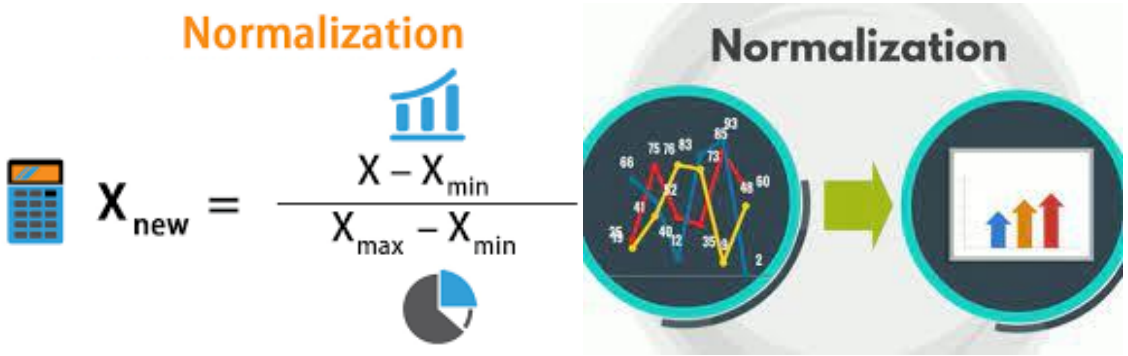
- DROPPED COLUMNS OF ATTRIBUTE NAMES AND ATTRIBUTE VALUES IN ORDER TO INTEGRATE ALL ATTRIBUTES AS DEDICATED COLUMNS



SUMMARY OF THE POTENTIAL CHANGES MADE TO THE INPUT DATA

DATA NORMALISATION

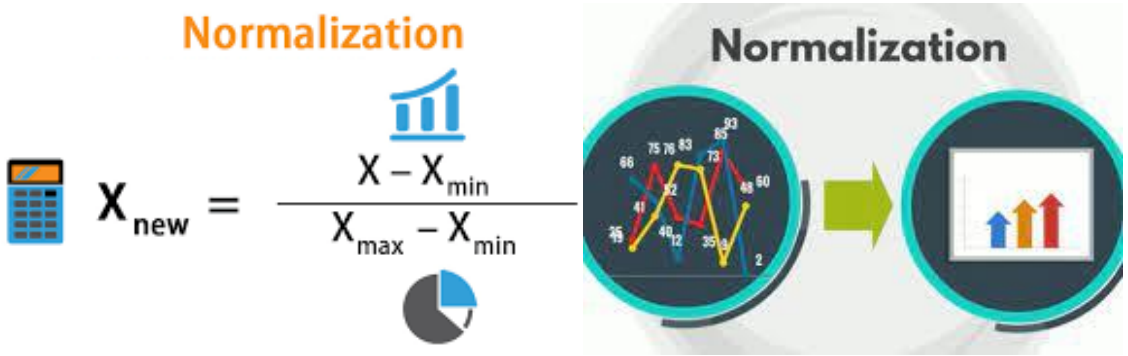
- NORMALIZED INPUT DATA BY ASSIGNING THE 'FUEL' (BENZINE / DIESEL) AND THE 'TRANSMISSION' (MANUAL / AUTOMATIC) BOOLEAN VALUES OF '0' AND '1'



SUMMARY OF THE POTENTIAL CHANGES MADE TO THE INPUT DATA

DATA NORMALISATION.....

- INPUT DATA CAN BE FURTHER NORMALIZED BY BEING CATEGORIZED OR, BY APPLYING MORE FILTERS TO THE INPUT DATA SUCH AS.....



SUMMARY OF THE POTENTIAL CHANGES MADE TO THE INPUT DATA

DATA NORMALISATION (APPLYING FILERS).....

- FILTER FOR VEHICLE MANUFACTURE DATE BEFORE & AFTER A CERTAIN DATE / YEAR, FOR EXAMPLE, BEFORE & AFTER YEAR 2005; AND.....

Normalization

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

SUMMARY OF THE POTENTIAL CHANGES MADE TO THE INPUT DATA

DATA NORMALISATION (APPLYING FILERS).....

- FILTERS FOR VEHICLE MILAGE, VEHICLE (HORSE) POWER, VEHICLE CARBON-EMISSION RATE, VEHICLE (BODY) TYPE ETC.

CUSTOMER TAKE-AWAY MESSAGE

TAKE-AWAY MESSAGE

- CLEAR, FILTERED, ORGANIZED & SIMPLER DATA THAT CAN HELP THE CUSTOMER MAKE BETTER, COST-EFFECTIVE AND TIME-EFFICIENT DECISION(S).

THANK YOU

WAQAS.ZAHICK@GMAIL.COM

+61469888165

LinkedIn HTTPS://WWW.LINKEDIN.COM/IN/WAQAS-ZAHICK

GitHub HTTPS://GITHUB.COM/WAQASZAHICK