

DATA ANALYTICS LEARNCAMP WEEK 5 Mini-Project

NAME: ISAAC IKENNA MAGNUS
TEAM: PEACE

A. RTA Dataset

Import the RTA Dataset into PowerBI; rename as road_traffic_accidents.
Answer the following question:

1. What columns have missing values?

- ✓ Vehicle_driver_relation
- ✓ Educational_level
- ✓ Driving_experience
- ✓ Type_of_vehicle
- ✓ Owner_of_vehicle
- ✓ Service_year_of_vehicle
- ✓ Defect_of_vehicle
- ✓ Area_accident_occured
- ✓ Lanes_of_medians
- ✓ Road_allignment
- ✓ Road_surface_type
- ✓ Type_of_collision
- ✓ Work_of_casualty
- ✓ Fitness_casuaality

2. What do we do to the missing values? – Find out how to deal with missing values

- ✓ Deleting Rows with missing values
- ✓ Impute missing values for continuous variable
- ✓ Impute missing values for categorical variable
- ✓ Other Imputation Methods
- ✓ Using Algorithms that support missing values
- ✓ Prediction of missing values

3. Replace every missing value with 'NA' where applicable.

Driving_experience	Type_of_vehicle	Owner_of_vehicle	Service_year_of_vehicle
Valid	Valid	Valid	Valid
Error	Error	Error	Error
Empty	Empty	Empty	Empty
1-2yr	Lorry (417100Q)	Owner	NA
Above 10yr	Lorry (417100Q)	Owner	1-2yr
5-10yr	Lorry (417100Q)	Owner	Above 10yr
Above 10yr	Lorry (11740Q)	Owner	NA
Below 1yr	Long lorry	Owner	5-10yrs
2-5yr	Lorry (417100Q)	Owner	NA
2-5yr	Lorry (11740Q)	Owner	5-10yrs
5-10yr	Lorry (417100Q)	Owner	NA
1-2yr	Lorry (417100Q)	Owner	NA
Below 1yr	Lorry (417100Q)	Owner	5-10yrs
5-10yr	Lorry (11740Q)	Owner	2-5yrs
1-2yr	Lorry (417100Q)	Owner	NA
2-5yr	Lorry (417100Q)	Owner	2-5yrs
5-10yr	Lorry (11740Q)	Owner	2-5yrs
1-2yr	Lorry (11740Q)	Owner	5-10yrs
1-2yr	Lorry (11740Q)	Owner	NA
1-2yr	Lorry (417100Q)	Owner	1-2yr
NA	Lorry (11740Q)	Owner	NA

4. Replace all 'Unknown' and 'na' with 'NA'

Driving_experience	Type_of_vehicle	Owner_of_vehicle	Service_year_of_vehicle
Valid	Valid	Valid	Valid
Error	Error	Error	Error
Empty	Empty	Empty	Empty
1-2yr	Lorry (417100Q)	Owner	NA
Above 10yr	Lorry (417100Q)	Owner	1-2yr
5-10yr	Lorry (417100Q)	Owner	Above 10yr
Above 10yr	Lorry (11740Q)	Owner	NA
Below 1yr	Long lorry	Owner	5-10yrs
2-5yr	Lorry (417100Q)	Owner	NA
2-5yr	Lorry (11740Q)	Owner	5-10yrs
5-10yr	Lorry (417100Q)	Owner	NA
1-2yr	Lorry (417100Q)	Owner	NA
Below 1yr	Lorry (417100Q)	Owner	5-10yrs
5-10yr	Lorry (11740Q)	Owner	2-5yrs
1-2yr	Lorry (417100Q)	Owner	NA
2-5yr	Lorry (417100Q)	Owner	2-5yrs
5-10yr	Lorry (11740Q)	Owner	2-5yrs
1-2yr	Lorry (11740Q)	Owner	5-10yrs
1-2yr	Lorry (11740Q)	Owner	NA
1-2yr	Lorry (417100Q)	Owner	1-2yr
NA	Lorry (11740Q)	Owner	NA

5. Do we have duplicate records? Check and remove duplicates if any. No duplicate records.

6. What hour of the day do we have most accidents?

1700 hours

7. From your SQL knowledge,

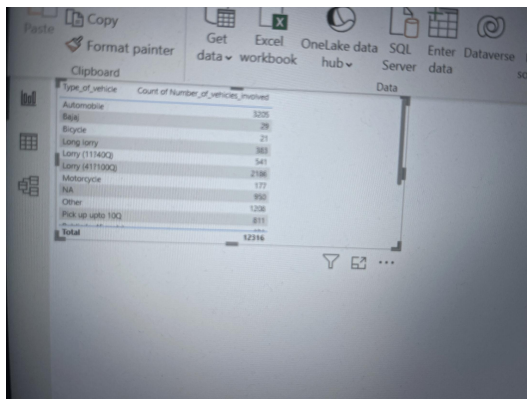
- a. what does '%lorry%' mean? Change every vehicle type that fall under that to 'Lorry'

Driving_experience	Type_of_vehicle	Owner_of_vehicle	Service_year_of_vehicle
Valid	Valid	Valid	Valid
Error	Error	Error	Error
Empty	Empty	Empty	Empty
1-2yr	Lorry (417100Q)	Owner	NA
Above 10yr	Lorry (417100Q)	Owner	1-2yr
5-10yr	Lorry (417100Q)	Owner	Above 10yr
Above 10yr	Lorry (11740Q)	Owner	NA
Below 1yr	Long lorry	Owner	5-10yrs
2-5yr	Lorry (417100Q)	Owner	NA
2-5yr	Lorry (11740Q)	Owner	5-10yrs
5-10yr	Lorry (417100Q)	Owner	NA
1-2yr	Lorry (417100Q)	Owner	NA
Below 1yr	Lorry (417100Q)	Owner	5-10yrs
5-10yr	Lorry (11740Q)	Owner	2-5yrs
1-2yr	Lorry (417100Q)	Owner	NA
2-5yr	Lorry (417100Q)	Owner	2-5yrs
5-10yr	Lorry (11740Q)	Owner	2-5yrs
1-2yr	Lorry (11740Q)	Owner	5-10yrs
1-2yr	Lorry (11740Q)	Owner	NA
1-2yr	Lorry (417100Q)	Owner	1-2yr
NA	Lorry (11740Q)	Owner	NA

- b. what does '%public%' mean? Change every vehicle type that fall under that to 'Public'

Driving_experience	Type_of_vehicle	Owner_of_vehicle	Service_year_of_vehicle
Valid	Valid	Valid	Valid
Error	Error	Error	Error
Empty	Empty	Empty	Empty
Above 10yr	Public (> 45 seats)	Owner	5-10yrs
5-10yr	Public (> 45 seats)	Governmental	NA
1-2yr	Public (13745 seats)	Owner	NA
1-2yr	Public (> 45 seats)	Owner	2-5yrs
5-10yr	Public (13745 seats)	Owner	NA
2-5yr	Public (13745 seats)	Owner	NA
Above 10yr	Public (12 seats)	Owner	NA
Above 10yr	Public (13745 seats)	Governmental	NA
5-10yr	Public (12 seats)	Organization	2-5yrs
1-2yr	Public (> 45 seats)	Owner	1-2yr
5-10yr	Public (12 seats)	Organization	Above 10yr
5-10yr	Public (12 seats)	Owner	NA
5-10yr	Public (12 seats)	Owner	1-2yr
5-10yr	Public (12 seats)	Owner	2-5yrs
2-5yr	Public (13745 seats)	Owner	NA
2-5yr	Public (12 seats)	Owner	NA
5-10yr	Public (12 seats)	Owner	NA

8. Which Vehicle Type has been involved in most accidents?
Automobile



9. What percentage of the vehicles involved in accidents had no defects?
63.15%

10. What percentage of the casualties are between 31 and 50 years old?
19.93%

11. A. Create a new column to check if the accident was a collision (any type)
– look critically!

	Age band of driver	Age band of casualty	Accident by vehicle	Collision status	Vehicle type
1	18-30	NA	Collision	Yes	Automobile
2	31-50	NA	Collision	Yes	Public
3	18-30	31-50	Collision	Yes	Lorry
4	18-30	18-30	Collision	Yes	Public
5	18-30	NA	Collision	Yes	NA
6	31-50	31-50	Collision	Yes	NA
7	18-30	18-30	Collision	Yes	Automobile
8	18-30	NA	Collision	Yes	Automobile
9	18-30	Under 18	Collision	Yes	Lorry
10	18-30	18-30	Collision	Yes	Automobile
11	18-30	NA	Collision	Yes	Public
12	31-50	18-30	Collision	Yes	Automobile
13	18-30	NA	Collision	Yes	Public
14	31-50	NA	Collision	Yes	Lorry
15	31-50	18-30	Collision	Yes	Automobile
16	18-30	NA	Collision	Yes	Lorry
17	18-30	NA	Collision	Yes	Public
18	18-30	NA	Collision	Yes	Public
19	18-30	NA	Collision	Yes	NA
20	18-30	Under 18	Collision	Yes	Lorry
21	Under 18	NA	Collision	Yes	Lorry
22	18-30	18-30	Collision	Yes	Lorry
23	18-30	18-30	Other	No	Lorry

- B. Based on this, how many accidents weren't caused by a collision
635 were not recorded as collision

B. HCAHPS+Patient+Survey Dataset

- **Create a new file and import all the data into PowerBi**
- **Investigate the datasets (to identify the keys, table types, etc) – use the data dictionary to understand the tables and attributes**
- **List the tables and their types using the format below**

Table Name	Table Type
Reports	Lookup Table
Measure	Lookup table
States	Lookup Table
Questions	Fact Table
Responses	Fact Table
national results	Fact Table
State results	Fact Table

- **Create the data model**

