

Question 1

Calculate the usage of Irish road network in terms of percentage grouped by vehicle category.

```
In [8]: vehicle_counter_DF\
        .groupBy('classname')\
        .count()\
        .withColumn('percent',
                     f.col('count')*100/f.sum('count').over(Window.partitionBy()))\
        .select(['classname', 'percent'])\
        .orderBy('percent', ascending=False)\
        .show()
```

classname	percent
CAR	80.25858594040197
LGV	11.194464465420944
HGV_ART	4.397450167807071
HGV_RIG	2.7310861887745705
BUS	0.6871114761643508
CARAVAN	0.42912036442325563
MBIKE	0.29486205142905475
null	0.007319345578788326

Question 2

Calculate the highest and lowest hourly flows on M50 - show the hours and total number of vehicle counts.

Highest Hourly flow

```
In [9]: vehicle_counter_DF\
        .groupBy('hour')\
        .count()\
        .orderBy('count', ascending=False)\
        .show(1)
```

hour	count
16	385850

only showing top 1 row

Lowest Hourly flow

```
In [10]: vehicle_counter_DF\
        .groupBy('hour')\
        .count()\
        .orderBy('count', ascending=True)\
        .show(1)
```

hour	count
2	13682

only showing top 1 row

Question 3

Calculate the evening and morning rush hours on M50 - show the hours and the total counts.

Morning hours

From 0400 hours in the morning to 1200 hours (Noon)

```
In [11]: vehicle_counter_DF\  
.groupBy('hour')\  
.count()\  
.orderBy('hour', ascending=True)\  
.filter((f.col('hour') < 12) & (f.col('hour') >= 4))\  
.show()
```

```
+---+-----+  
|hour| count|  
+---+-----+  
|  4| 27187|  
|  5| 61937|  
|  6|198369|  
|  7|299784|  
|  8|352862|  
|  9|277509|  
| 10|256183|  
| 11|246847|  
+---+-----+
```

Evening hours

From 1600 hours to 2000 hours

```
In [12]: vehicle_counter_DF\  
.groupBy('hour')\  
.count()\  
.orderBy('hour', ascending=True)\  
.filter((f.col('hour') < 20) & (f.col('hour') >= 16))\  
.show()
```

```
+---+-----+  
|hour| count|  
+---+-----+  
| 16|385850|  
| 17|367269|  
| 18|314085|  
| 19|232409|  
+---+-----+
```

Question 4

Calculate average speed between each junction on M50 (e.g., junction 1, junction 2 - junction 3, etc.).

```
In [13]: vehicle_counter_DF\
.groupBy('lanename')\
.agg({'speed': 'mean'})\
.orderBy('avg(speed)', ascending=False)\
.show()
```

lanename	avg(speed)
Southbound 1 (slow)	135.4469130170314
Northbound 2	122.31002458344715
Eastbound 2	114.68716172331673
Eastbound 2 (fast)	113.59000942507083
Southbound 2 (fast)	111.72458022387893
Westbound 2 (fast)	111.34068965517257
Northbound 2 (fast)	110.21109738884894
southbound	104.4090909090909
Northbound 1	103.95987028779895
Northbound 1 (slow)	103.63843987902779
Westbound 3 (fast)	103.47554310278697
Eastbound 3 (fast)	100.55896097639352
Southbound 2	97.79121728990314
Northbound 2	97.76209841746629
Westbound 2 (slow)	95.40281196241926
Southbound Mainli...	95.25522388059701
Westbound 2	93.36724880445983
Eastbound 2	92.79648071706569
Eastbound on slip	92.7741935483871
Southbound 1	92.74038016587762

only showing top 20 rows

Question 5

Calculate the top 10 locations with highest number of counts of HGVs (class). Map the COSITs with their names given on the map.

```
In [14]: vehicle_counter_DF\
.filter((f.col('classname') == "HGV_ART") | (f.col('classname') == "HGV_RIG"))\
.groupBy('lanename')\
.agg(
  f.mean("cosit").alias('Average cosit'),
  f.count(f.lit(1)).alias('count')
)\
.orderBy('count', ascending=False)\
.show(10)
```

lanename	Average cosit	count
Northbound 1	20693.790824685962	47606
Southbound 1	20370.47938361651	47438
Westbound 1	47984.26970280579	26481
Eastbound 1	50842.564949674364	25335
Northbound	11535.248807024242	20956
Southbound	15721.02364244845	18526
Northbound 2	6737.779673675744	17406
Southbound 2	5250.837445297139	15767
Eastbound	11569.320689406504	13867
Westbound	11157.380619527628	13591

only showing top 10 rows