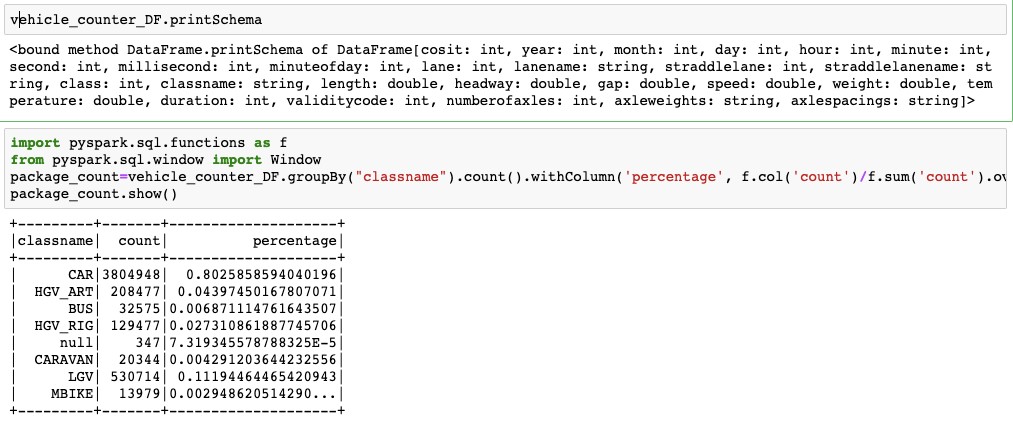
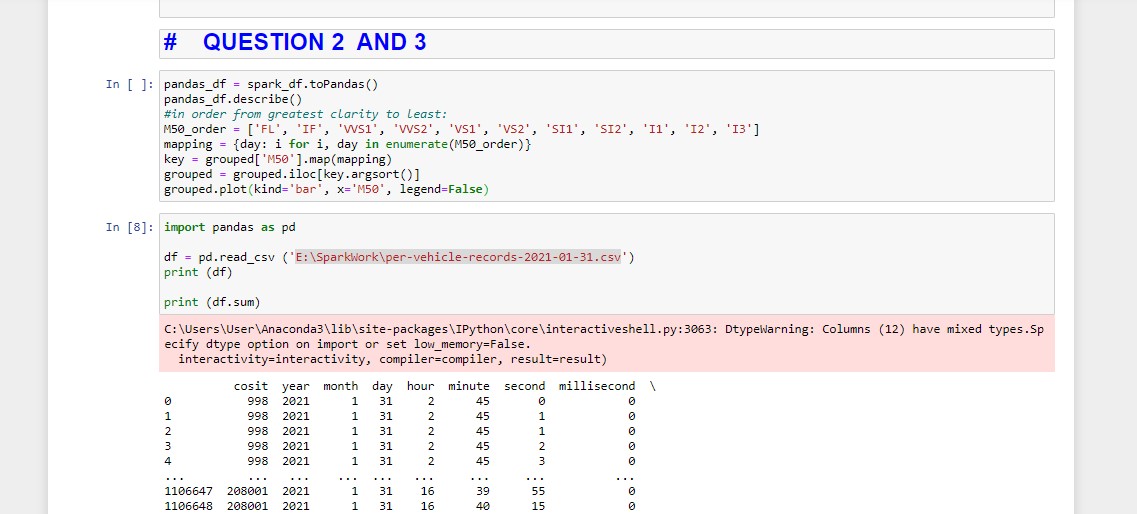
**SUMMARY ANALYTICS**

**QUESTION 1**

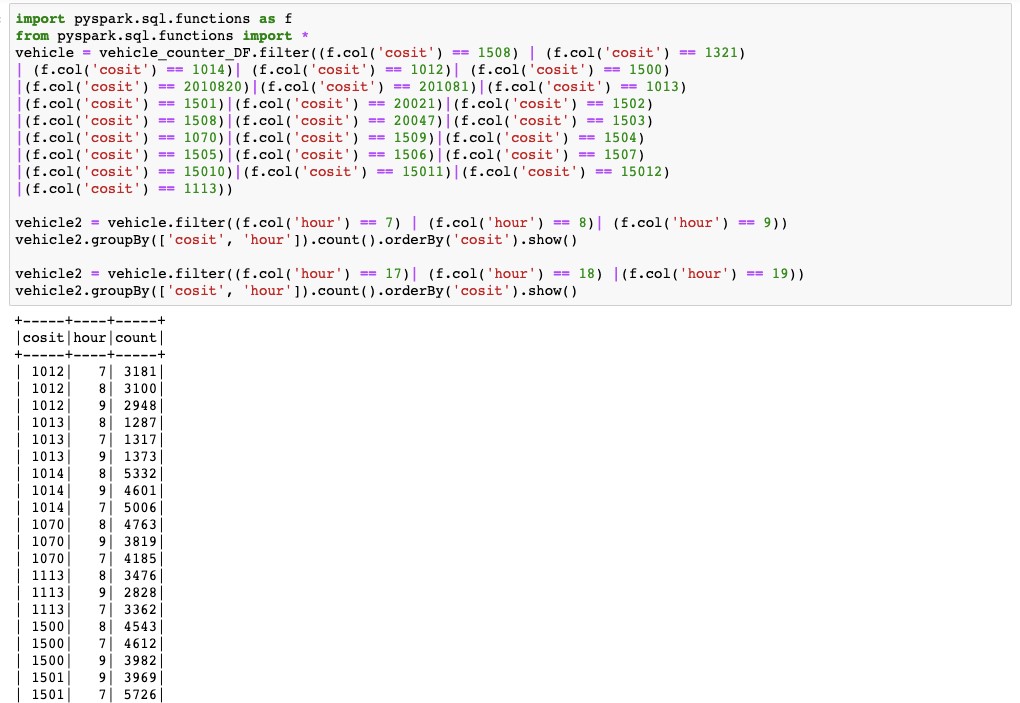


**QUESTION 2 AND 3**

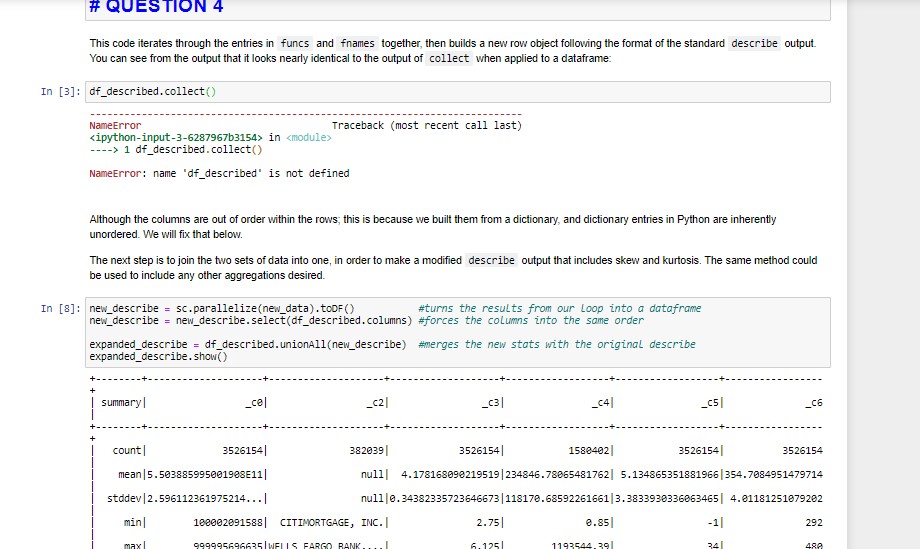


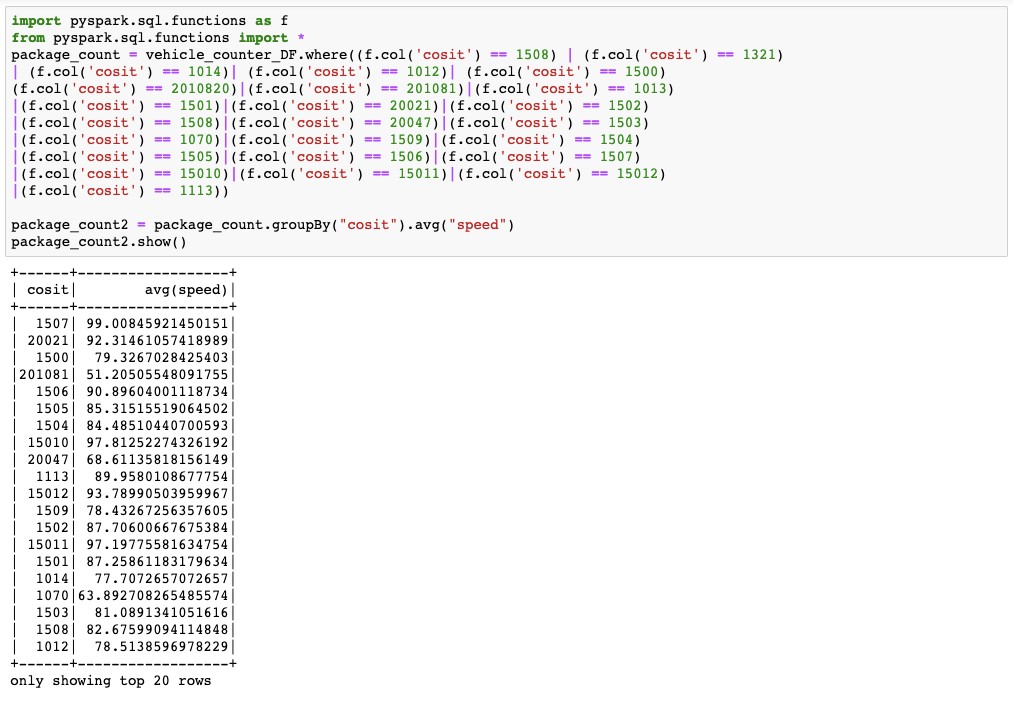


QUESTION 3



**QUESTION 4 AND 5**





QUESTION 6



**CODE AND OUTPUT**

**QUESTION 1**

tot = vehicle\_counter\_DF. count( )

print ( "Total Vehicle Entry print ( tot )

groupBy( " classname " ) \

. count( ) \

.withC01umnRenamed( ' count ' , ' Count' ) \

. withC01umn( ' Percentage ' (F.col( 'Count '

. show( ) tot )

In [63 ] :

Total Vehicle Entry 1106652

**QUESTION 2**

I classnamel Count I Percentage I

CARI 918254 | 82 .97585871619985 1 HGV\_ARTI 33805 1 3. 05470915879608 1 BUSI 10519 1 0 . 9505246455073502 1 HGV\_RIGI 308661 2 . 7891333499600597 | null I 50 1 0 . 004518132168016684 1 CARAVAN 1 5887 1 0 .5319648814622845 1

LGVI 104580 | 9 .450125242623697 1

MBIKE I 2691 1 0 .24316587328265796 1

**Question 3—5**

# Calculate the highest and lowest hourly fows on M50 — from pyspark . sql import Window

ExampleDF = M50DF . groupBy( "hour") \

. count( ) \

.withC01umnRenamed( ' count' , 'Total Vehicle Count ' )

print( "Lowest Hourly Flow" ) resDF = ExampleDF . filter(col( "Total Vehicle Count resDF . show( )

print( "Highest Hourly Flow" ) resDF = ExampleDF . filter(col( "Total Vehicle Count resDF . show( ) show the hours and total number of vehicle counts.

ExampleDF . groupby( ) . min( ' Total Vehicle Count' ) . head( )

ExampleDF . groupby( ) . max( ' Total Vehicle Count' ) . head( )

In [167] :

Lowest Hourly Flow

I hour I Total Vehicle Count I

3 1 510 1

Highest Hourly Flow

I hour I Total Vehicle Count I

15 1 172111

I have assumed : Morning Hours —> 8 to 11 and Evening Hours 17 to 20 morningDF = M50DF.fi1ter(c01( "hour" ) '8' ) . "hour" ) <= ' 11' ) . orderBy( " hour" )

morningRushDF = morningDF . groupBy( " hour"

. count( ) \

. withC01umnRenamed ( count ' ' Total Vehicle Count' )

print( "Morning Rush Hour" ) resDF = morningRushDF . filter (col( "Total Vehicle Count ") morningRushDF . groupby( ) . max( 'Total Vehicle Count ' ) . head( ) [0 resDF . show( )

eveningDF = M50DF.fi1ter(c01( "hour") >= ' 17 ' ) . "hour" ) '20 ' ) . orderBy( " hour" )

eveningRushDF = eveningDF . groupBy( " hour" ) \

. count( ) \

.withC01umnRenamed( ' count' , 'Total Vehicle Count ' )

print( "Evening Rush Hours " ) resDF = eveningRushDF . filter (col( "Total Vehicle Count " = eveningRushDF. groupby( ) . max( 'Total Vehicle Count ' ) . head( ) [0 resDF . show( )

Morning Rush Hour

I hour I Total Vehicle Count I

Evening Rush Hours

I hour I Total Vehicle Count I

jun14 = M50DF.fi1ter(c01( "cosit") 15010) . orderBy( "hour" ) jun14 . groupBy( ) . agg(F . sum( " speed" ) . alias( "count " ) ) \

. withC01umnRenamed( ' count ' , ' totalSpeed' ) \

. withC01umn( ' Average Speed ' (F. col( ' total Speed' ) / total Speed) )

. show( )

print( "Avg Speed between Junction 15 and 16 " jun15 = M50DF. "cosit") == 15011) . orderBy( "hour" ) jun15 . groupBy( ) . agg(F.sum( " speed" ) . alias( "count " ) ) \

. withC01umnRenamed( ' count ' , ' totalSpeed' ) \

. withC01umn( ' Average Speed' , (F. col( ' totalSpeed' ) / totalSpeed) )

. show( )

print( "Avg Speed between Junction 16 and 17 " jun16 = M50DF. "cosit") 15012) . orderBy( "hour" ) jun16 . groupBy( ) . agg(F.sum( " speed" ) . alias( "count " ) ) \

. withC01umnRenamed ( count ' ' total Speed' ) \

. withC01umn( ' Average Speed' , (F. col( ' total Speed' ) / total Speed) ) . show( )

Sum of Speeds between junction 3 and junction 17

.0

Avg Speed between Junction 3 and 4

I totalSpeed I Average Speed I

| 1168870 .0 1 0.06380070482238506 1

Avg Speed between Junction 4 and 5

I totalSpeed I Average Speed I

| 3900959 .0 1 0. 21292695824448094 1

**QUESTION 6**

ß Question 6 —Calculate the top 10 locations with highest number of counts of HGVs (class) .

HGV\_ART DF = "classname" ) —=- 'HGV\_ART' )

HGV RIG DF = M50DF. filter (col( "classname" ) 'HGV RIG' )

HGV DF -- HGV ART\_DF. join(HGV RIG DF, [ ' cosit'

HighestHGV = HGV DF. ' cosit '

. count( ) \

. withC01 umnRenamed ( ' count ' ' Total' ) \

. orderBy (COI ( ' cos it ' ) .desc( ) )

# print the Top 10 locations with highest number of counts of HGVs

HighestHGV. orderBy ( col ( ' Total ' ) .desc( ) ) .show(10)

Map the COSITs with their names given on the map

—> Ballymun, Ballymun

—> Ballymun, finglas

—> M50 Between Jn06 N03/M50 and Jn05 N02/M50, Finglas, Co. Dublin

—> M50 Between Jn07 N04/M50 and Jn09 N07/M50 Red Cow, Palmerstown, Co. Dublin

—> M50 Between Jn10 — Ballymount and Jnll — Tymon, Co. Dublin

—> M50 Between Jnll Tallaght and Jn12 Firhouse, Co. Dublin

—> M50 Between Jn12 Firhouse and Jn13 Dundrum, Balinteer, Co. Dublin

—> M50 Between Jn06 N03/M50 and Jn07 N04/M50, Castleknock, Co. Dublin

—> M50 Between Jn09 N07/M50 Red Cow and Jn10 Ballymount, Ballymount, Co. Dublin

15010 —> M50 Between Jn14 Dun Laoghaire and Jn15 Carrickmines, Cabinteely, Co. Dublin

In [246] :

I cositl Total I

15081 1747351

1 1518401

1 141361 1

1501 1 1096201

15001 510861