

Real Time Traffic Project

Under Super Vision: DR. Amira Yousef

Ahmed Salama Soliman Hasan

Mohamed Nasr

Osama Ahmed Reda Hegazy

Sara Hisham Ahmed Mohamed

Sherif Gamal Kamed AbdElwahed

Yousef Ahmed Mohamed Ibrahim

Zakaria Yehia Ahmed

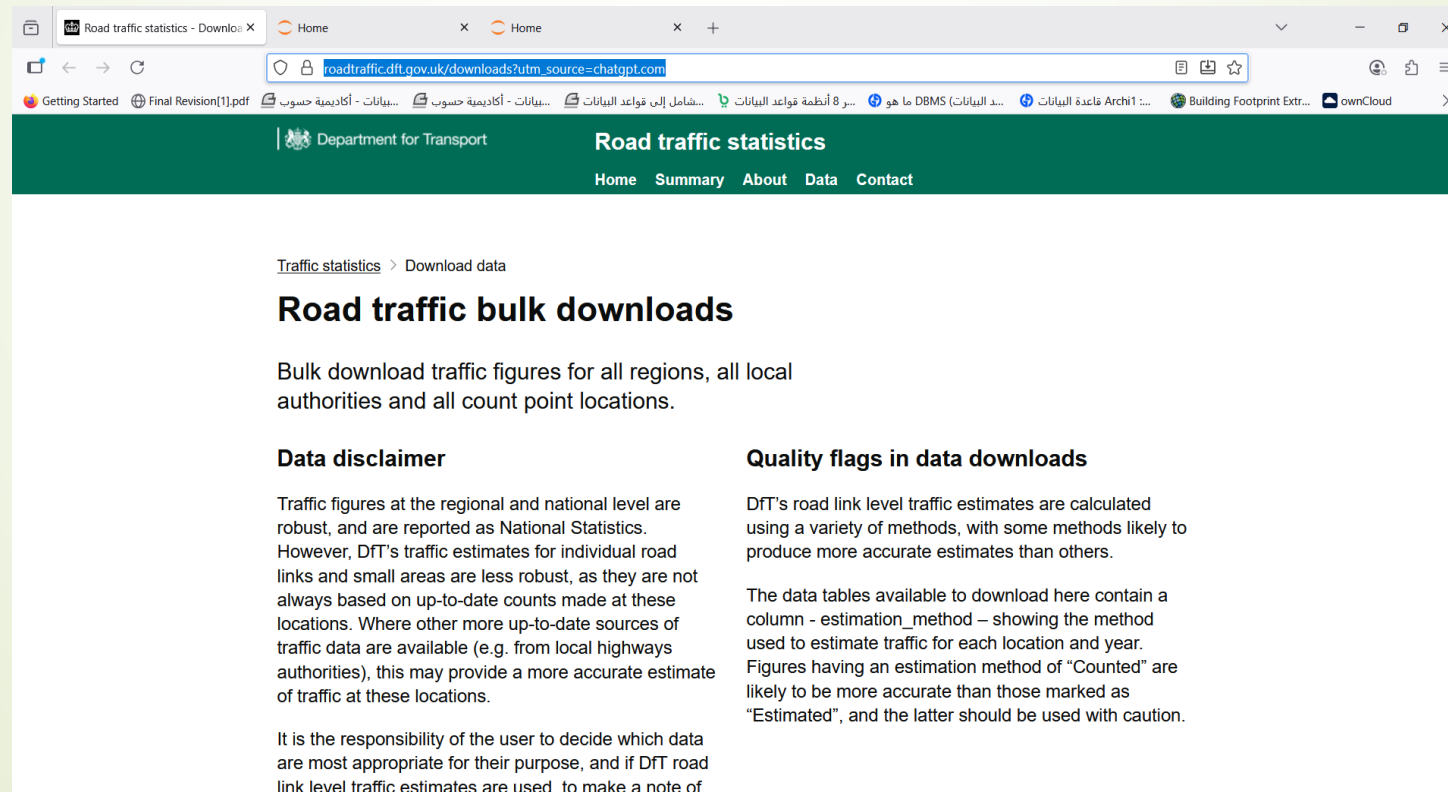


The beneficiaries of the "Real Time Traffic Project" include:

1. **Commuters and Drivers** – who can access up-to-date traffic information, avoid congestion, and reduce travel time.
2. **Public Transportation Authorities** – who can optimize routes, schedules, and fleet management using real-time data.
3. **Emergency Services** – such as police, ambulance, and fire departments, enabling faster response times and route planning.
4. **Urban Planners and Decision Makers** – who can use traffic data to improve infrastructure planning and traffic flow management.
5. **Logistics and Delivery Companies** – who can enhance operational efficiency and reduce fuel consumption through route optimization.
6. **Environmental Agencies** – benefiting from reduced vehicle emissions due to improved traffic flow.
7. **Citizens and Local Communities** – who experience safer roads, less congestion, and better air quality.

Data Source

https://roadtraffic.dft.gov.uk/downloads?utm_source=chatgpt.com



The screenshot shows a web browser window with the URL https://roadtraffic.dft.gov.uk/downloads?utm_source=chatgpt.com. The page is titled "Road traffic statistics" and is part of the Department for Transport website. The main heading is "Road traffic bulk downloads". Below this, it states: "Bulk download traffic figures for all regions, all local authorities and all count point locations." There are two columns of text: "Data disclaimer" and "Quality flags in data downloads". The "Data disclaimer" states that traffic figures are robust at the regional and national level but less so at the individual road link level. The "Quality flags in data downloads" section explains that DfT's road link level traffic estimates are calculated using various methods, and that figures marked as "Counted" are more accurate than those marked as "Estimated".

Road traffic statistics

Home Summary About Data Contact

[Traffic statistics](#) > Download data

Road traffic bulk downloads

Bulk download traffic figures for all regions, all local authorities and all count point locations.

Data disclaimer

Traffic figures at the regional and national level are robust, and are reported as National Statistics. However, DfT's traffic estimates for individual road links and small areas are less robust, as they are not always based on up-to-date counts made at these locations. Where other more up-to-date sources of traffic data are available (e.g. from local highways authorities), this may provide a more accurate estimate of traffic at these locations.

It is the responsibility of the user to decide which data are most appropriate for their purpose, and if DfT road link level traffic estimates are used, to make a note of

Quality flags in data downloads

DfT's road link level traffic estimates are calculated using a variety of methods, with some methods likely to produce more accurate estimates than others.

The data tables available to download here contain a column - estimation_method – showing the method used to estimate traffic for each location and year. Figures having an estimation method of "Counted" are likely to be more accurate than those marked as "Estimated", and the latter should be used with caution.

(worked on 46476 out of 10,485,76records)

The figure displays three screenshots of Microsoft Excel spreadsheets, each showing a different data table.

Top Screenshot (Sheet1): Columns A-L. Data for road categories.

count_point_id	direction_of_travel	year	count_date	hour	region_id	region_name	region_ons_code	local_authority_id	local_authority_name	local_authority_code	road_name
51	S	2004	5/21/2004	11	1	South West	E12000009	1	Isles of Scilly	E06000053	A3111
51	S	2004	5/21/2004	15	1	South West	E12000009	1	Isles of Scilly	E06000053	A3111

Middle Screenshot (Sheet2): Columns M-V. Data for road categories.

road_category	road_type	start_junction_road_name	end_junction_road_name	easting	northing	latitude	longitude	link_length_km	link_length
PA	Major	Pierhead, Hugh Town	A3112	90200	10585	49.91501492	-6.31713812	0.3	0.19

Bottom Screenshot (Sheet3): Columns V-WZ. Data for vehicle types.

link_length_miles	pedal_cycles	two_wheeled_motor_vehicles	cars_and_taxis	buses_and_coaches	LGVs	HGVs_2_rigid_axle	HGVs_3_rigid_axle	HGVs_4_or_more_rigid_axle	HGVs_3_or_4_articulated_axle	HGVs_5_or_more_articulated_axle
0.19	12	2	27	2	16	2	0	0	0	0
0.19	10	1	29	1	13	2	0	0	0	0
0.19	7	0	21	2	23	5	0	0	0	0

Bottom Screenshot (Sheet4): Columns AE-AZ. Data for vehicle types.

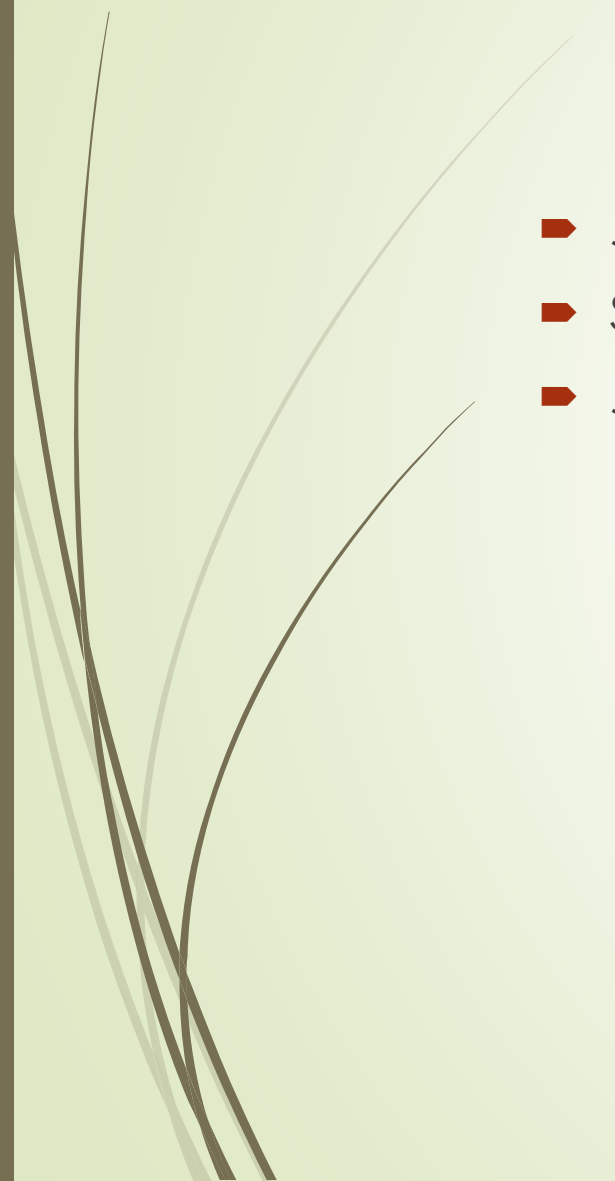
HGVs_3_or_4_articulated_axle	HGVs_5_articulated_axle	HGVs_6_articulated_axle	all_HGVs	all_motor_vehicles
0	0	0	2	49
0	0	0	2	46
0	0	0	5	51
0	0	0	0	19

شرح معاني الحقول:(Columns)

اسم الحقل	المعنى
count_point_id	رقم تعريف نقطة العدّ (مكان تم فيه عدّ المركبات).
direction_of_travel	اتجاه السفر (زي شمال، جنوب، شرق، غرب).
year	سنة تسجيل البيانات.
count_date	التاريخ الفعلي للعدّ (يوم العدّ).
hour	الساعة (الوقت الذي تم فيه العدّ – غالبًا من 0 إلى 23).
region_id	رقم تعريف المنطقة (رقم داخلي يستخدم لتحديد المنطقة).
region_name	اسم المنطقة (زي "East Midlands" أو "London").
region_ons_code	كود المنطقة حسب مكتب الإحصاء الوطني (ONS للاستخدام الإحصائي).
local_authority_id	رقم تعريف السلطة المحلية (بلدية أو مجلس محلي).
local_authority_name	اسم السلطة المحلية (زي "Leeds City Council").
local_authority_code	كود السلطة المحلية (رمز مختصر).
road_name	اسم الطريق (زي A1 ، M25).
road_category	فئة الطريق (زي طريق سريع، طريق رئيسي، محلي، إلخ).
road_type	نوع الطريق (قد يكون مزيد من التفصيل عن الفئة – مفرد، مزدوج، إلخ).
start_junction_road_name	اسم الطريق عند بداية المفرق أو التقاطع.
end_junction_road_name	اسم الطريق عند نهاية المفرق أو التقاطع.
easting	الإحداثي الشرقي (نظام الإحداثيات البريطانية).
northing	الإحداثي الشمالي (نظام الإحداثيات البريطانية).
latitude	دائرة العرض (إحداثيات جغرافية).
longitude	خط الطول (إحداثيات جغرافية).
link_length_km	طول المقطع الطرقي بالكيلومترات.
link_length_miles	طول المقطع الطرقي بالأميال.
pedal_cycles	عدد الدراجات الهوائية.
two_wheeled_motor_vehicles	عدد الدراجات النارية (سكوتر، موتوسيكل).
cars_and_taxis	عدد السيارات العادية والتاكسي.
buses_and_coaches	عدد الحافلات.
LGVs	سيارات النقل الخفيف – (Light Goods Vehicles) زي الفانات الصغيرة.
HGVs_2_rigid_axle	شاحنات ثقيلة بعدد 2 محور صلب.
HGVs_3_rigid_axle	شاحنات ثقيلة بعدد 3 محاور صلبة.
HGVs_4_or_more_rigid_axle	شاحنات ثقيلة بعدد 4 أو أكثر من المحاور الصلبة.
HGVs_3_or_4_articulated_axle	شاحنات ثقيلة مفصلية بعدد 3 أو 4 محاور.
HGVs_5_articulated_axle	شاحنات ثقيلة مفصلية بعدد 5 محاور.
HGVs_6_articulated_axle	شاحنات ثقيلة مفصلية بعدد 6 محاور أو أكثر.
all_HGVs	إجمالي الشاحنات الثقيلة (HGVs).
all_motor_vehicles	إجمالي جميع المركبات الآلية (من موتوسيكلات لحد الشاحنات).



Software Used

- Jupyter , Python
 - SQL Server
 - Jithub
- 

Data Preparation and Geospatial Processing Workflow

- Reading and cleaning data
- Transforming time columns
- Checking for outliers
- Exporting data to an SQL database
- Creating a GeoDataFrame and exporting data to GeoJSON and Shapefile
- Data quality report

1. Initial Setup and Data Loading

"import pandas as pd - import numpy as np - import geopandas as gpd)

Load the CSV data

2. Data Exploration and Cleaning

3. Data Validation and Quality Checks

4. Data Transformation and Feature Engineering

=====

5. Save to Parquet (optional)

7. Aggregate Data for Reporting

8. Save GeoJSON (Optional for Spatial Analysis)

9. Data Quality Report

A decorative graphic element featuring a solid red arrow pointing to the right, positioned horizontally across the middle of the page. Behind the arrow and extending upwards and downwards are several thin, dark, curved lines that resemble stylized grass or reeds. The background is a light, neutral color.

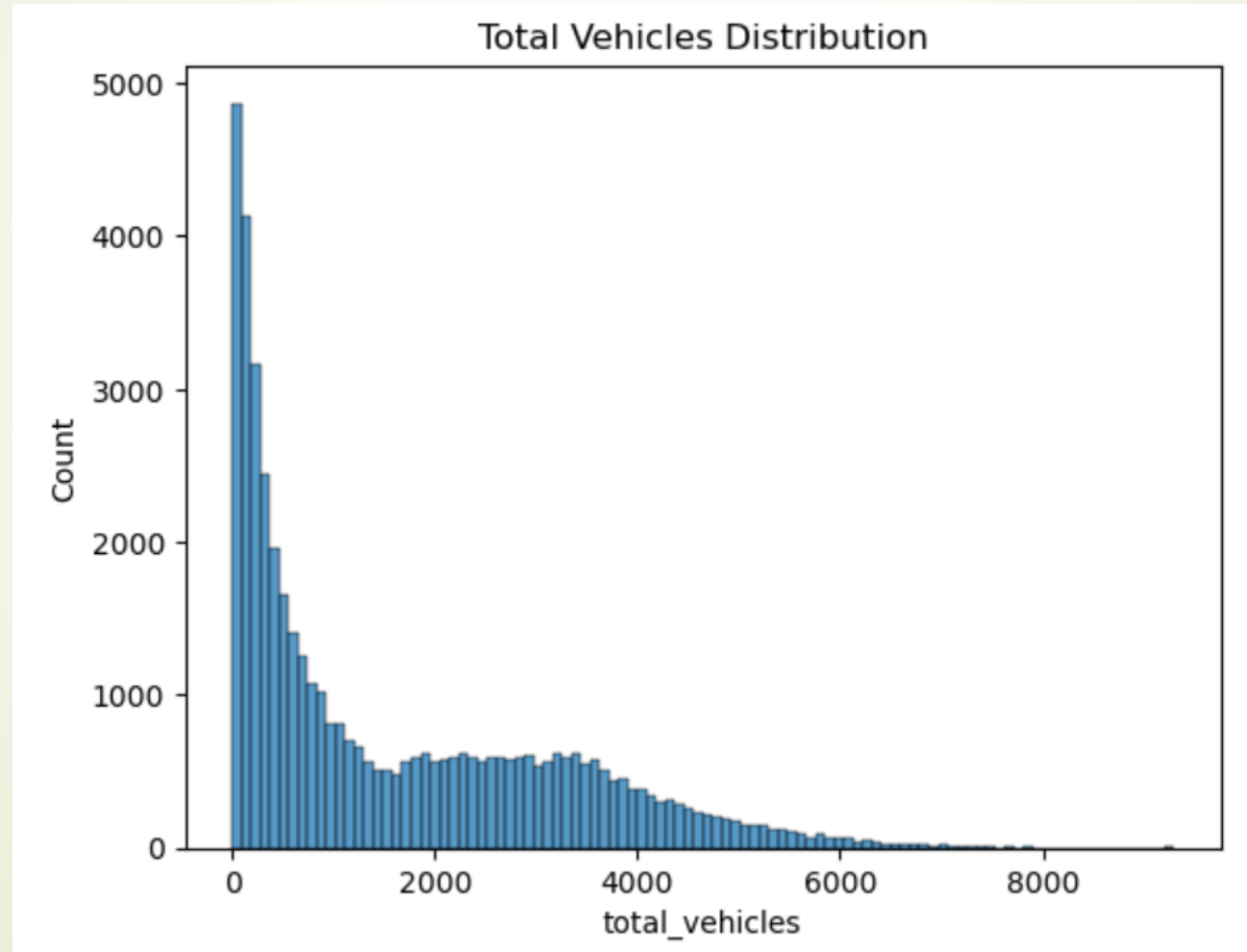


قيم غير منطقية (Logical Errors)

العمود	التحقق
hour	هل القيم بين 0 و 23 فقط؟
total_vehicles	هل في قيم سالبة أو أكبر من مليون؟
link_length_km	هل يوجد قيم سالبة؟
vehicle_density	هل القيمة منطقية؟ مثلاً مش 100,000 مركبة/كم
bicycle_percentage أو hgv_percentage	هل أكثر من 1 أو أقل من 0؟ (نسب غير منطقية)

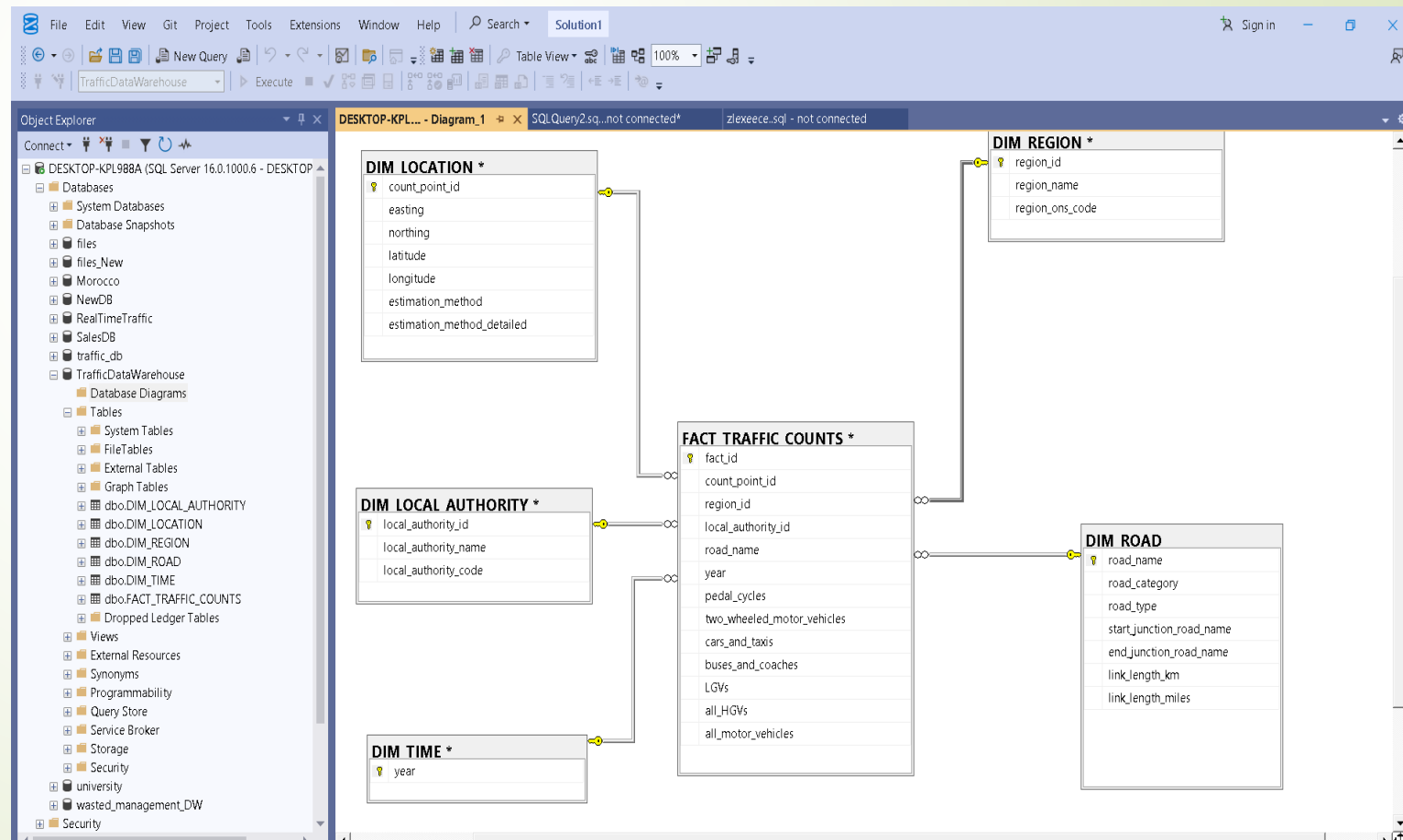
- Illogical Dates
- Have future dates?
- Some records have missing dates (NaT)?
- Geographic points outside the UK?
- If the data is from the UK, such as DfT
- Make sure the latitude and longitude are within the UK Approximately

Data Distribution If the data has a strong skew or many outliers, it can affect statistical analysis or modeling



Data Warehouse

Data After Restructure (Star Schema)





Data Analysis:

Traffic by Region

The screenshot shows the SQL Server Enterprise Manager interface. The Object Explorer on the left displays the database structure, including the 'TrafficDataWarehouse' database. The central pane shows a SQL query executed successfully, which aggregates traffic data by region. The results pane displays a table with 8 columns: region_name, total_pedal_cycles, total_motorbikes, total_cars, total_buses, total_LGVs, total_HGVs, and total_vehicles. The data is sorted by total_vehicles in descending order.

```
SELECT
  r.region_name,
  SUM(f.pedal_cycles) AS total_pedal_cycles,
  SUM(f.two_wheeled_motor_vehicles) AS total_motorbikes,
  SUM(f.cars_and_taxis) AS total_cars,
  SUM(f.buses_and_coaches) AS total_buses,
  SUM(f.LGVs) AS total_LGVs,
  SUM(f.all_HGVs) AS total_HGVs,
  SUM(f.all_motor_vehicles) AS total_vehicles
FROM FACT_TRAFFIC_COUNTS f
JOIN DIM_REGION r ON f.region_id = r.region_id
GROUP BY r.region_name
ORDER BY total_vehicles DESC;
```

region_name	total_pedal_cycles	total_motorbikes	total_cars	total_buses	total_LGVs	total_HGVs	total_vehicles
1 North West	2297	55831	14419684	71670	2852956	2228258	19628399
2 Wales	10219	61549	9417358	110448	1691686	798392	12079433
3 South East	419	65429	8244863	51803	1582812	975887	10920794
4 Scotland	15415	43528	7796663	115061	1318192	743447	10016891
5 Yorkshire and the Humber	9	15093	4499004	18217	1055161	1006015	6593490
6 East Midlands	96	12923	3248678	16834	727813	773920	4780168
7 South West	3265	12617	2530596	16093	459452	381409	3400167
8 London	2	42085	2519228	29802	433491	193210	3217816
9 West Midlands	1	7320	2009116	9272	433648	435781	2895137
10 East of England	26	5633	809325	6642	143011	174341	1136952

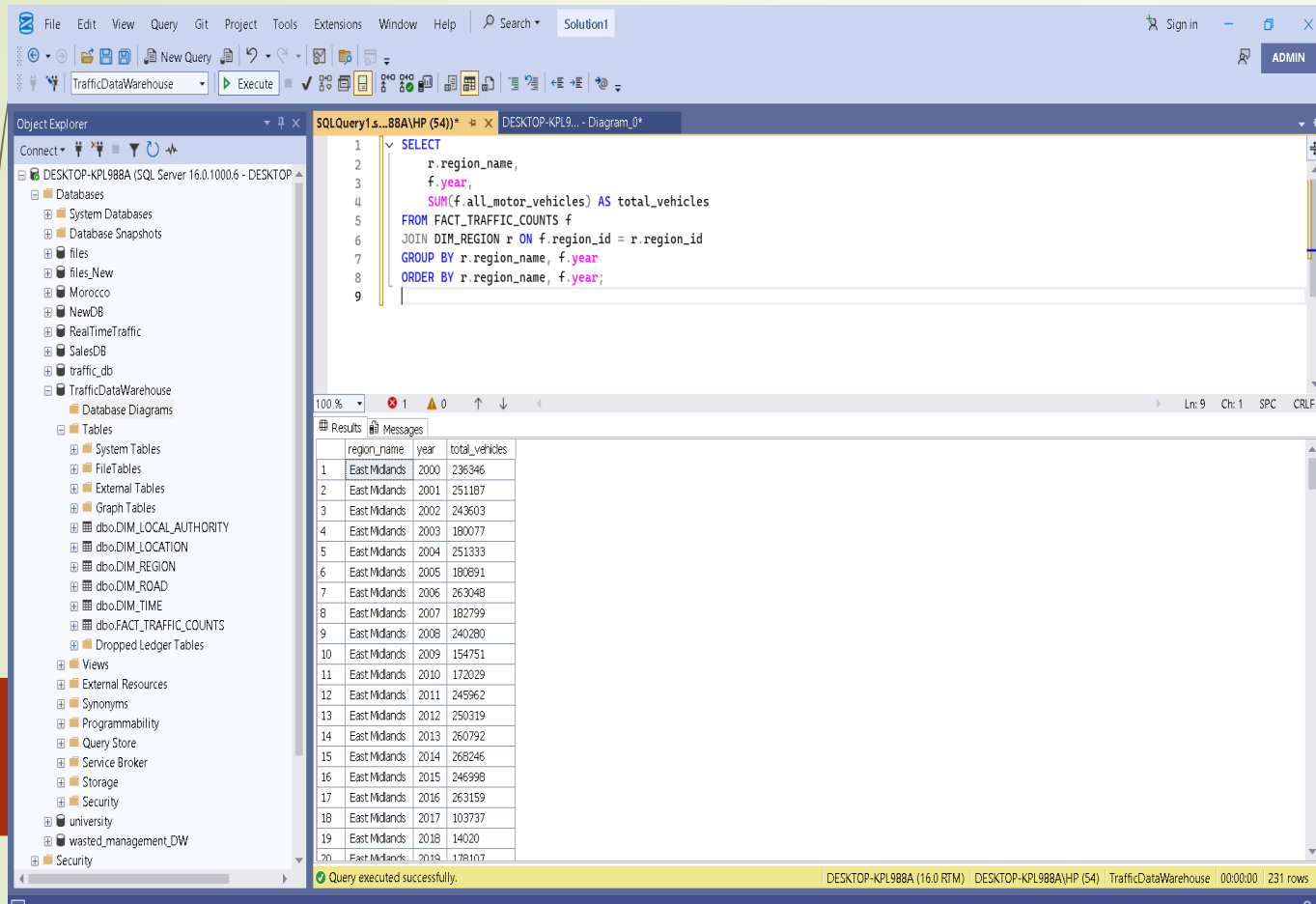
Traffic by Local Authority

The screenshot shows the SQL Server Enterprise Manager interface. The Object Explorer on the left displays the database structure, including the 'TrafficDataWarehouse' database. The central pane shows a SQL query executed successfully, which aggregates traffic data by local authority. The results pane displays a table with 8 columns: local_authority_name, region_name, total_pedal_cycles, total_motorbikes, total_cars, total_buses, total_LGVs, and total_vehicles. The data is sorted by total_vehicles in descending order.

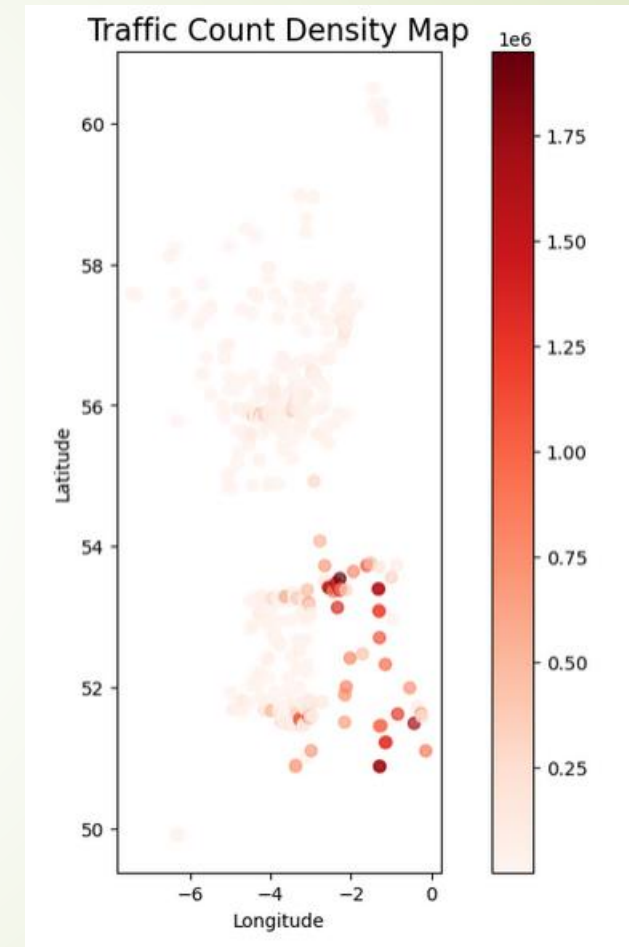
```
SELECT
  r.region_name,
  SUM(f.pedal_cycles) AS total_pedal_cycles,
  SUM(f.two_wheeled_motor_vehicles) AS total_motorbikes,
  SUM(f.cars_and_taxis) AS total_cars,
  SUM(f.buses_and_coaches) AS total_buses,
  SUM(f.LGVs) AS total_LGVs,
  SUM(f.all_HGVs) AS total_HGVs,
  SUM(f.all_motor_vehicles) AS total_vehicles
FROM FACT_TRAFFIC_COUNTS f
JOIN DIM_LOCAL_AUTHORITY la ON f.local_authority_id = la.local_authority_id
JOIN DIM_REGION r ON f.region_id = r.region_id
GROUP BY la.local_authority_name, r.region_name
ORDER BY total_vehicles DESC;
```

local_authority_name	region_name	total_pedal_cycles	total_motorbikes	total_cars	total_buses	total_LGVs	total_vehicles
1 Warrington	North West	1	11276	2798245	12159	602393	584226
2 Hampshire	South East	399	24480	2987063	14332	546285	258723
3 Manchester	North West	4	6925	2324288	14353	404388	209601
4 Bury	North West	5	7613	1950893	8053	506051	379575
5 Leeds	Yorkshire and the Humber	0	5467	1590259	6252	380441	343010
6 Rotherham	Yorkshire and the Humber	2	5474	1607361	6789	364227	288600
7 Newport	Wales	1085	11161	1765577	30617	292277	178934
8 Worcestershire	West Midlands	1	9885	1598539	7640	316432	307917
9 Glasgow City	Scotland	4421	7041	1692015	19712	292802	146061
10 Rhondda, Cynon, Taff	Wales	751	10436	1697992	16255	320053	101789
11 Kent	South East	12	15195	1486430	30064	383078	276773
12 Hillingdon	London	2	24205	1699526	20276	262863	121642
13 Salford	North West	0	5774	1531796	4921	286888	154281
14 Derbyshire	East Midlands	0	4013	1156108	5136	259526	256119
15 East Cheshire	North West	0	3174	1145901	4670	255488	246194
16 Trafford	North West	6	4604	1313319	3365	224972	109827
17 Lancashire	North West	0	6158	1173768	7891	220198	202605
18 Buckinghamshire	South East	0	7195	1235598	11881	203671	141957
19 Leicestershire	East Midlands	5	4544	1090029	5162	239123	240462
20 Cheshire	North West	48	9632	1031190	4876	164516	136760

Traffic Trend by Region (Over Time)



Plot:



Data Analysis – Date time: ⌚ Objective of the Analysis Traffic data analysis (Traffic Counts) by time, to understand temporal patterns such as:

- Distribution of the number of vehicles by year / month / day / hour
- Identifying Peak Hours
- Calculating the total and average number of vehicles at different times
- Understanding temporal trends (daily, weekly, yearly)

2. Analysis of total vehicles by year and month



4. Analysis by day of the week

The screenshot shows the SQL Server Enterprise Manager interface. The Object Explorer on the left displays the database structure, including the 'TrafficDataWarehouse' database. The central pane shows a SQL query in the 'SQLQuery1.s...88A\HP (54)' window:

```
1 SELECT
2     YEAR(count_date) AS [Year],
3     MONTH(count_date) AS [Month],
4     SUM(all_motor_vehicles) AS Total_Motor_Vehicles
5 FROM FACT_TRAFFIC_COUNTS
6 GROUP BY YEAR(count_date), MONTH(count_date)
7 ORDER BY [Year], [Month];
8
```

The Results pane at the bottom displays the query output as a table with 20 rows and 3 columns: Year, Month, and Total_Motor_Vehicles. The status bar at the bottom indicates 'Query executed successfully.'

Year	Month	Total_Motor_Vehicles
2000	3	504451
2000	4	264049
2000	5	852996
2000	6	1112096
2000	7	280237
2000	9	266711
2000	10	838434
2001	3	402546
2001	4	507596
2001	5	573671
2001	6	1198545
2001	7	336480
2001	9	646266
2001	10	317703
2002	3	101736
2002	4	594974
2002	5	813541
2002	6	712596
2002	7	425985
2002	9	830821

3. Analysis of the number of vehicles by hour - Peak Hour Analysis.

The top screenshot shows the SQL Server Enterprise Manager interface with a query for analysis by day of the week in the 'SQLQuery1.s...88A\HP (54)' window:

```
1 SELECT
2     DATENAME(WEEKDAY, count_date) AS [DayOfWeek],
3     AVG(all_motor_vehicles) AS Avg_Vehicles,
4     SUM(all_motor_vehicles) AS Total_Vehicles
5 FROM FACT_TRAFFIC_COUNTS
6 GROUP BY DATENAME(WEEKDAY, count_date), DATEPART(WEEKDAY, count_date)
7 ORDER BY DATEPART(WEEKDAY, count_date);
8
```

The Results pane displays the query output as a table with 5 rows and 3 columns: DayOfWeek, Avg_Vehicles, and Total_Vehicles. The status bar at the bottom indicates 'Query executed successfully.'

DayOfWeek	Avg_Vehicles	Total_Vehicles
Monday	1621	14667324
Tuesday	1631	14213574
Wednesday	1506	13722129
Thursday	1699	15579276
Friday	1579	16486944

The bottom screenshot shows the SQL Server Enterprise Manager interface with a query for peak hour analysis in the 'SQLQuery1.s...88A\HP (54)' window:

```
1 SELECT
2     hour AS [Hour],
3     AVG(all_motor_vehicles) AS Avg_Vehicles,
4     SUM(all_motor_vehicles) AS Total_Vehicles
5 FROM FACT_TRAFFIC_COUNTS
6 GROUP BY hour
7 ORDER BY [Hour];
8
```

The Results pane displays the query output as a table with 12 rows and 3 columns: Hour, Avg_Vehicles, and Total_Vehicles. The status bar at the bottom indicates 'Query executed successfully.'

Hour	Avg_Vehicles	Total_Vehicles
7	1796	6957078
8	1821	7052046
9	1493	5784849
10	1366	5293555
11	1382	5353829
12	1408	5456238
13	1459	5653819
14	1529	5922637
15	1661	6437315
16	1896	7344737
17	1920	7438942
18	1542	5974202

🚗 5. Direction Analysis (Direction of Travel)

The screenshot shows the SQL Server Enterprise Manager interface. The left pane displays the Object Explorer with the 'TrafficDataWarehouse' database selected. The central pane shows a query window with the following SQL code:

```
1 SELECT
2     direction_of_travel,
3     AVG(all_motor_vehicles) AS Avg_Traffic,
4     SUM(all_motor_vehicles) AS Total_Traffic
5 FROM FACT_TRAFFIC_COUNTS
6 GROUP BY direction_of_travel
7 ORDER BY Total_Traffic DESC;
8
```

The right pane shows the query results in a table with 5 rows:

direction_of_travel	Avg_Traffic	Total_Traffic
N	1565	18827295
S	1559	18700640
W	1676	18628859
E	1656	18410902
C	445	101551

The status bar at the bottom indicates 'Query executed successfully.' and 'DESKTOP-KPL988A (16.0 RTM) | DESKTOP-KPL988A\HP (54) | TrafficDataWarehouse | 00:00:01 | 5 rows'.

6. Road Analysis Road Name

The screenshot shows the SQL Server Enterprise Manager interface. The left pane displays the Object Explorer with the 'TrafficDataWarehouse' database selected. The central pane shows a query window with the following SQL code:

```
1 SELECT
2     road_name,
3     COUNT(*) AS Observations,
4     SUM(all_motor_vehicles) AS Total_Traffic
5 FROM FACT_TRAFFIC_COUNTS
6 GROUP BY road_name
7 ORDER BY Total_Traffic DESC;
8
```

The right pane shows the query results in a table with 20 rows:

road_name	Observations	Total_Traffic
M1	2760	9050095
M4	2424	7367865
M6	2304	7318001
M60	1355	6484332
M5	1848	4735811
M62	1368	4645154
M56	1152	3985049
M8	888	3014448
M27	576	2239213
M20	552	2140540
A470	1344	1859903
A55	1344	1746775
M40	504	1598152
M3	504	1591670
A90	924	1296363
M23	432	1121862
M2	384	1029204
A494	552	832651
M621	264	626968
A48	1032	683648

The status bar at the bottom indicates 'Query executed successfully.' and 'DESKTOP-KPL988A (16.0 RTM) | DESKTOP-KPL988A\HP (54) | TrafficDataWarehouse | 00:00:00 | 236 rows'.

7. Combining temporal and spatial analysis example — traffic per road by the hour

The screenshot shows the SQL Server Enterprise Manager interface. The Object Explorer on the left displays the database structure, including the 'TrafficDataWarehouse' database. The central pane shows a SQL query titled 'SQLQuery1.s...B8A/HP (54)' with the following code:

```
1 SELECT
2     road_name,
3     hour,
4     AVG(all_motor_vehicles) AS Avg_Vehicles
5 FROM FACT_TRAFFIC_COUNTS
6 GROUP BY road_name, hour
7 ORDER BY road_name, hour;
```

The Results pane at the bottom displays the query output as a table with columns 'road_name', 'hour', and 'Avg_Vehicles'. The status bar at the bottom indicates 'Query executed successfully.' and 'DESKTOP-KPL988A (16.0 RTM) | DESKTOP-KPL988A/HP (54) | TrafficDataWarehouse | 00:00:00 | 2,832 rows'.

road_name	hour	Avg_Vehicles
A1	7	514
A1	8	608
A1	9	472
A1	10	422
A1	11	459
A1	12	429
A1	13	473
A1	14	490
A1	15	513
A1	16	612
A1	17	623
A1	18	504
A1107	7	36
A1107	8	67
A1107	9	62
A1107	10	58
A1107	11	68
A1107	12	73
A1107	13	69
A1107	14	71

8. Analysis of the Annual Growth Rate

The screenshot shows the SQL Server Enterprise Manager interface. The Object Explorer on the left displays the database structure. The central pane shows a SQL query titled 'SQLQuery1.s...B8A/HP (54)' with the following code:

```
1 SELECT
2     year,
3     SUM(all_motor_vehicles) AS Total_Traffic,
4     LAG(SUM(all_motor_vehicles)) OVER (ORDER BY year) AS Prev_Year_Traffic,
5     ((SUM(all_motor_vehicles) - LAG(SUM(all_motor_vehicles)) OVER (ORDER BY year)) /
6      LAG(SUM(all_motor_vehicles)) OVER (ORDER BY year)) AS Growth_Rate_Percent
7 FROM FACT_TRAFFIC_COUNTS
8 GROUP BY year
9 ORDER BY year;
```

The Results pane at the bottom displays the query output as a table with columns 'year', 'Total_Traffic', 'Prev_Year_Traffic', and 'Growth_Rate_Percent'. The status bar at the bottom indicates 'Query executed successfully.' and 'DESKTOP-KPL988A (16.0 RTM) | DESKTOP-KPL988A/HP (54) | TrafficDataWarehouse | 00:00:02 | 25 rows'.

year	Total_Traffic	Prev_Year_Traffic	Growth_Rate_Percent
2001	3982807	4118974	-3.305847524164
2002	3548550	3982807	-10.903290066528
2003	3417423	3548550	-3.695227625186
2004	3616571	3417423	5.827431956769
2005	3202366	3616571	-11.452975760741
2006	4088717	3202366	27.678004325551
2007	3872616	4088717	-5.285300988060
2008	3768354	3872616	-2.692288623504
2009	3638782	3768354	-3.438424309393
2010	2994595	3638782	-17.703368874529
2011	3178559	2994595	5.912786303142
2012	3071319	3178559	-3.36364030910
2013	3043136	3071319	-0.917618782028
2014	3116489	3043136	2.410441071315
2015	2279241	3116489	-26.865103647084
2016	2981765	2279241	30.822716600568
2017	2553289	2981765	-14.369078243255
2018	1052090	2553289	-58.794715365162
2019	2860687	1052090	171.905160204925
2020	1534361	2860687	-46.713811949610



Thank
you! :)