

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 Ibrahiem

Zakaria Yehia Ahmed

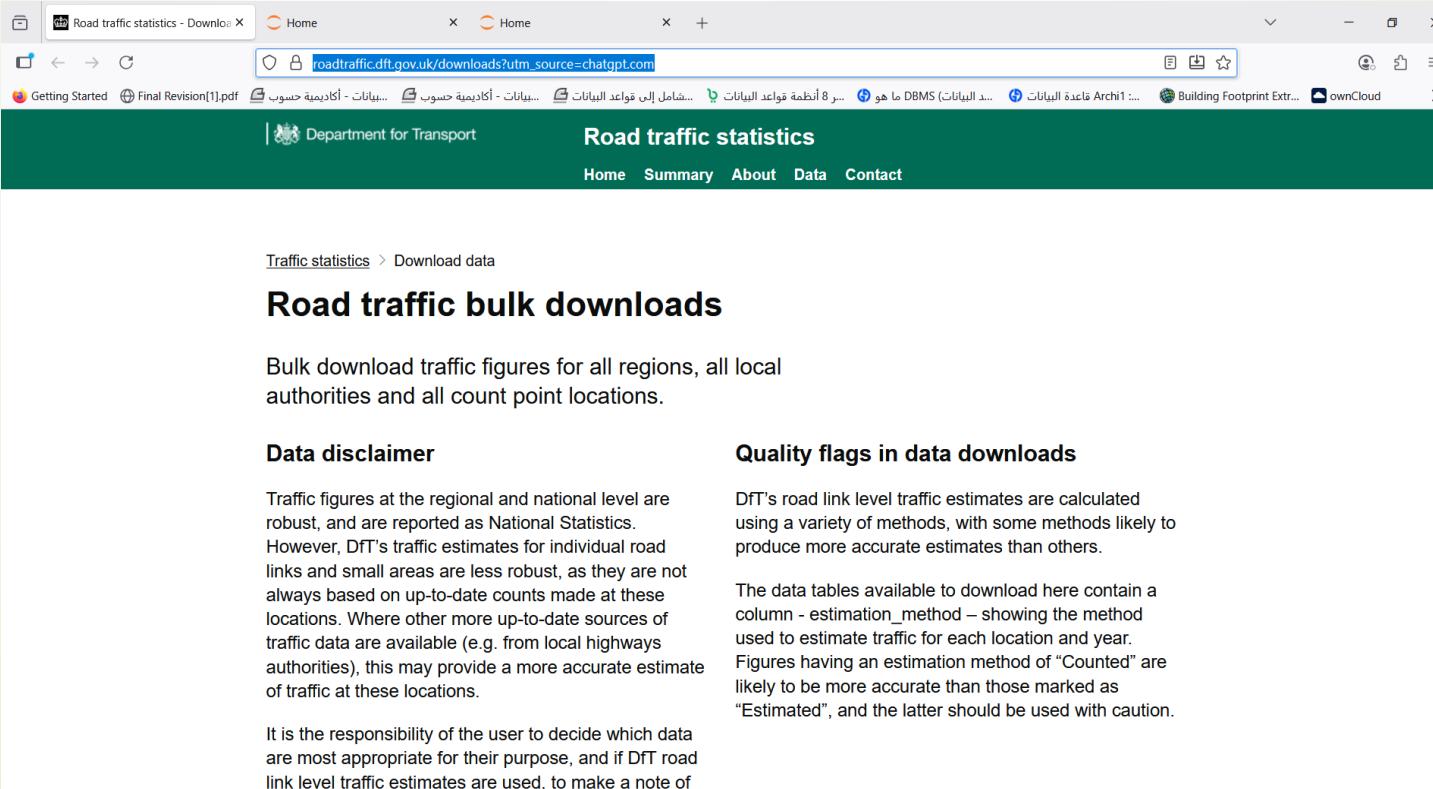


The beneficiaries of the "Real Time Traffic Project" include:

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 roadtraffic.dft.gov.uk/downloads?utm_source=chatgpt.com in the address bar. The page is titled "Road traffic statistics" and features a green header with links for Home, Summary, About, Data, and Contact. Below the header, a breadcrumb navigation shows "Traffic statistics > Download data". The main content is titled "Road traffic bulk downloads" and describes the service for downloading traffic figures for all regions, local authorities, and count point locations. Two columns of text provide additional information: "Data disclaimer" and "Quality flags in data downloads".

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.

Data Before

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

The image displays three stacked Microsoft Excel spreadsheets illustrating data before processing.

Top Spreadsheet: Shows data for local authorities. The columns include: 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, and a redacted column. Row 1 is a header. Rows 2 and 3 show data for South West and Isles of Scilly respectively. Conditional formatting highlights 'Good' values in green and 'Neutral' values in yellow.

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 Spreadsheet: Shows data for road links. The columns include: road_category, road_type, start_junction_road_name, end_junction_road_name, easting, northing, latitude, longitude, link_length_km, and link_length_miles. Row 1 is a header. Row 2 shows data for a Major road from Pierhead, Hugh Town to A3112. Conditional formatting highlights 'Good' and 'Neutral' values.

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

Bottom Spreadsheet: Shows vehicle counts across various categories. The columns include: 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, and HGVs_5_articulated_axle. Row 1 is a header. Rows 2 through 5 show data for different categories. Conditional formatting highlights 'Good' and 'Neutral' values.

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_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

شرح معاني الحقول (Columns):

اسم الحقل	المعنى
count_point_id	رقم تعرف نقطة العد (مكان تم فيه عد المركبات).
direction_of_travel	اتجاه السفر (إي شمال، جنوب، شرق، غرب).
year	سنة تسجيل البيانات.
count_date	التاريخ الفعلي للعد (يوم العد).
hour	الساعة (الوقت الذي تم فيه العد - غالباً من 0 إلى 23).
region_id	رقم تعرف المنطقة (رقم داخلي يستخدم لتحديد المنطقة).
region_name	اسم المنطقة (إي "London" أو "East Midlands").
region_ons_code	كود المنطقة حسب مكتب الإحصاء الوطني ONS للاستخدام الإحصائي.
local_authority_id	رقم تعرف السلطة المحلية (بلدية أو مجلس محلي).
local_authority_name	اسم السلطة المحلية (إي "Leeds City Council").
local_authority_code	كود السلطة المحلية (رمز مختصر).
road_name	اسم الطريق (إي A1 ، M25).
road_category	فئة الطريق (إي طريق سريع، طريق رئيسي، محلي، الخ).
road_type	نوع الطريق (قد يكون مزيد من التفصيل عن الفئة - مفرد، مزدوج، الخ).
start_junction_road_name	اسم الطريق عند بداية المفترق أو التقاطع.
end_junction_road_name	اسم الطريق عند نهاية المفترق 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 بالكميلومترات.
link_length_miles	طول المقطع الطرقي بالأميال.
pedal_cycles	عدد الدراجات الهوائية.
two_wheeled_motor_vehicles	عدد الدراجات النارية (سكوتر، موتسيكل).
cars_and_taxis	عدد السيارات العادي والتاكسي.
buses_and_coaches	عدد الحافلات.
LGVs	سيارات النقل الخفيف - (Light Goods Vehicles) إي 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 موتسيكلات 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 Shapfile
 - 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
```

```
# -----
```

```
# 8. Save GeoJSON (Optional for Spatial Analysis)
```

```
# -----
```

```
# -----
```

```
# 3. Data Validation and Quality Checks
```

```
# -----
```

```
# 9. Data Quality Report
```

```
# -----
```

```
# -----
```

```
# 4. Data Transformation and Feature Engineering
```

```
# -----
```

```
# 5. Save to Parquet (optional)
```

```
# -----
```

```
# -----
```

```
# 7. Aggregate Data for Reporting
```

```
# -----
```

GIS Map

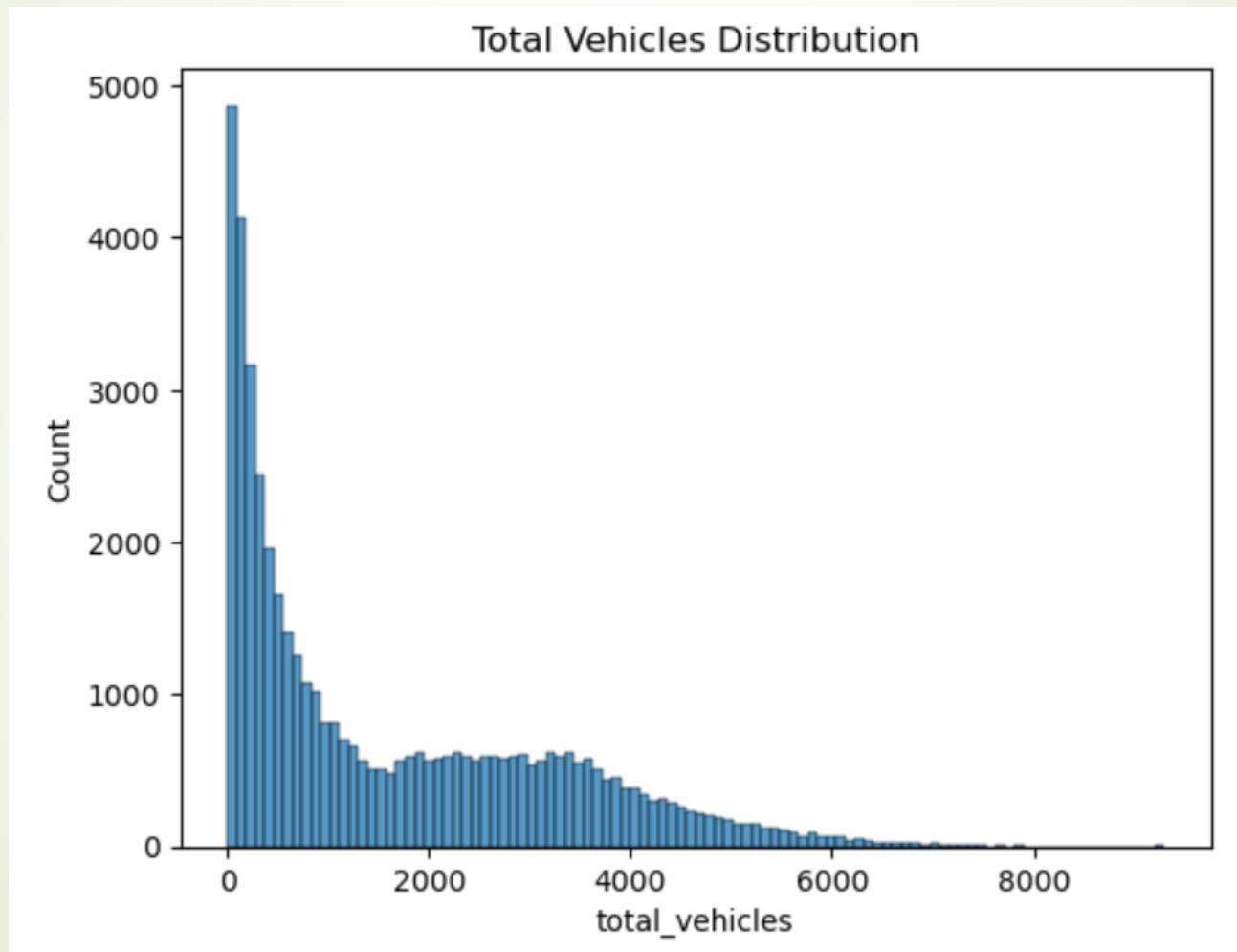


قيم غير منطقية (Logical Errors)

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)

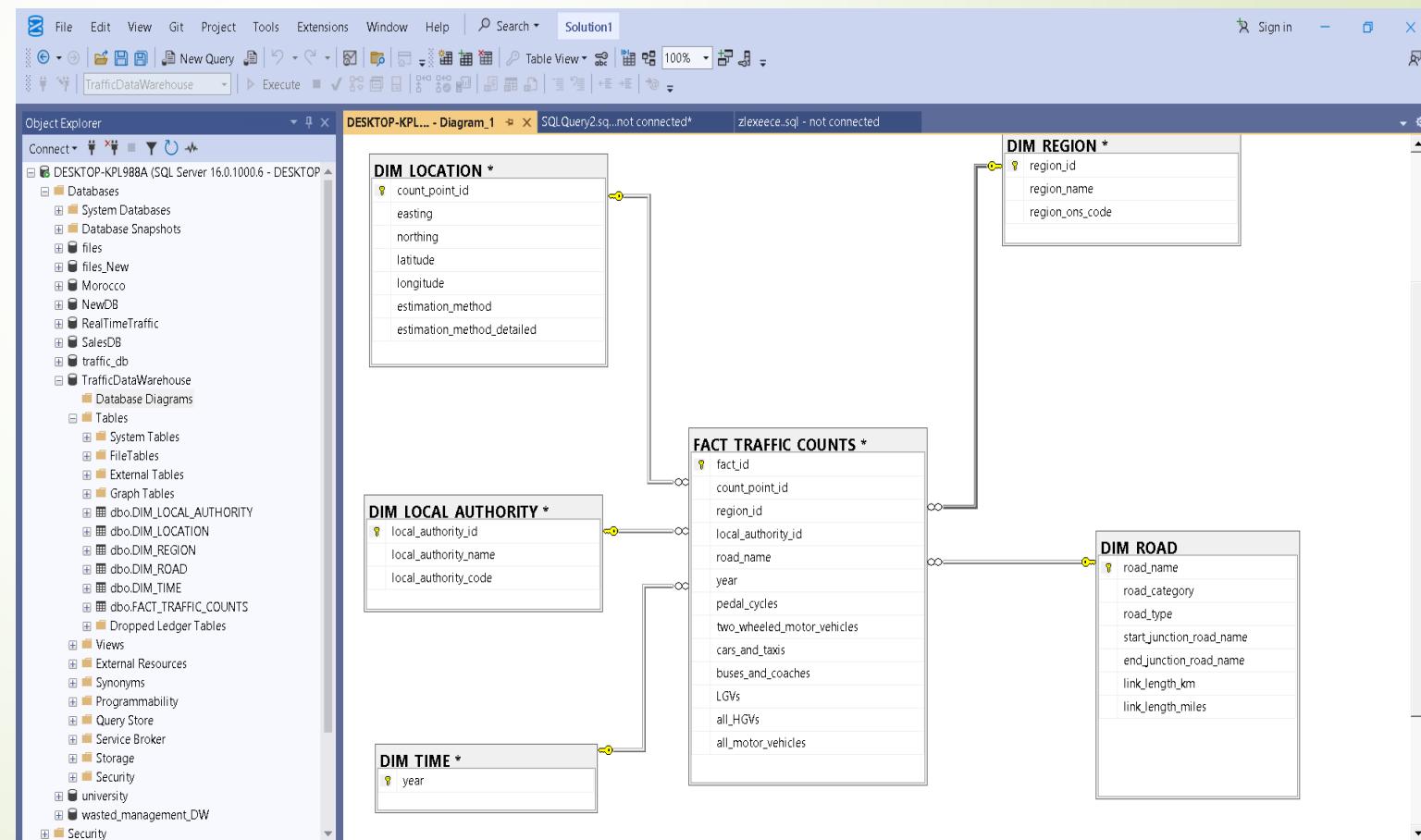
- Illogical Dates
- Have future dates?
- Some records have missing dates (NaT)?
- Geographic points outside the UK?
- If the data is from the UK, such as DFT
- Make sure the latitude and longitude are within the UK Approximately

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



Data Warehouse

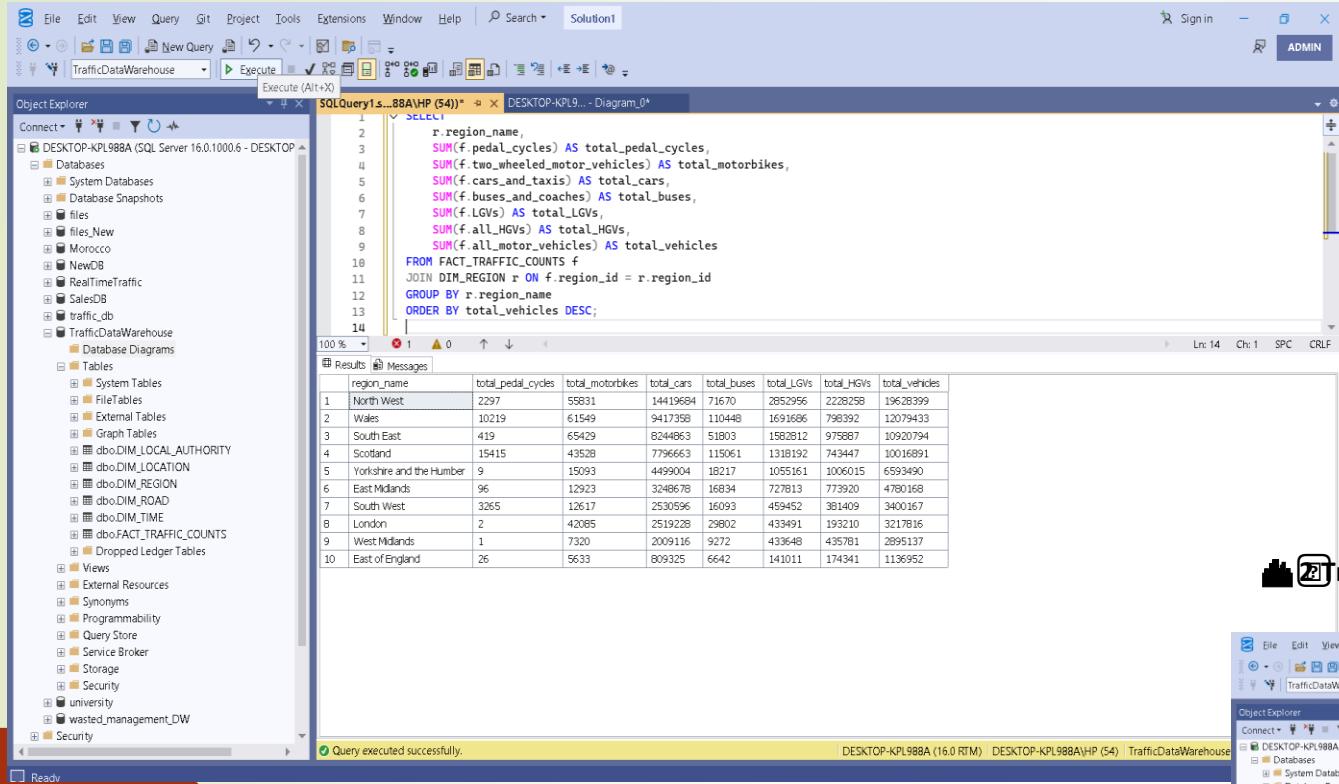
Data After Restructure (Star Schema)





Data Analysis:

Traffic by Region



File Edit View Query Git Project Tools Extensions Window Help | Search | Solution1

TrafficDataWarehouse | Execute | DESKTOP-KPL988A (54) | Diagram_0

```

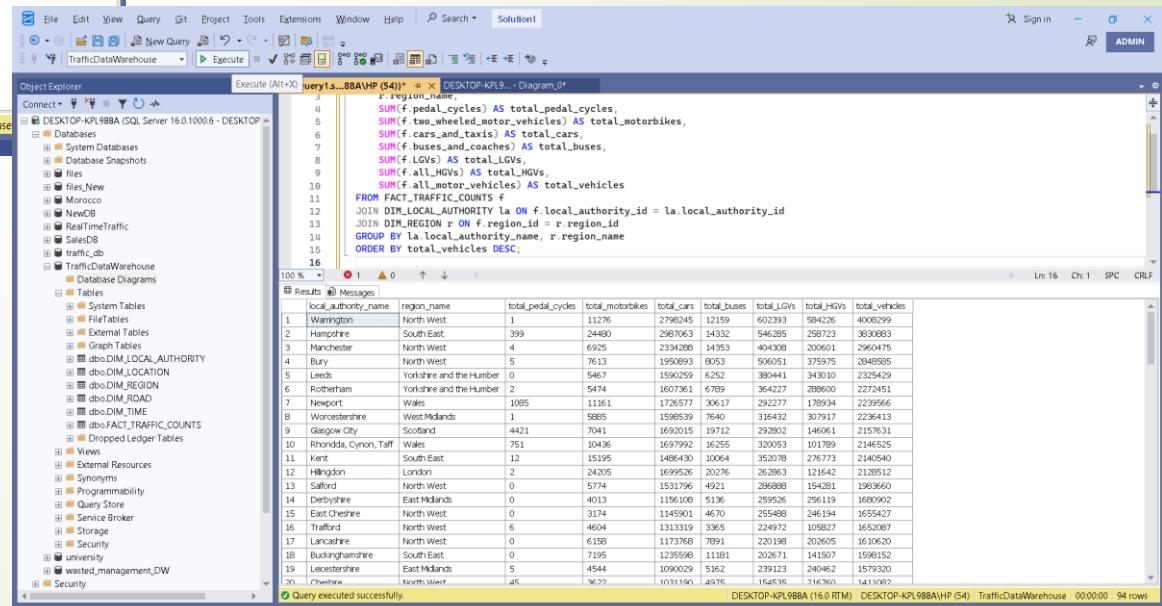
SELECT
    r.region_name,
    SUM(f.pedal_cycles) AS total_pedal_cycles,
    SUM(f.two_wheeled_motor_vehicles) AS total_motorbikes,
    SUM(f.cars_and_taxis) AS total_cars,
    SUM(f.buses_and_coaches) AS total_buses,
    SUM(f.LGVs) AS total_LGVs,
    SUM(f.all_HGVs) AS total_HGVs,
    SUM(f.all_motor_vehicles) 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;
  
```

Results Messages

region_name	total_pedal_cycles	total_motorbikes	total_cars	total_buses	total_LGVs	total_HGVs	total_vehicles
North West	2297	55831	14419694	71670	2652966	2228288	19628399
Wales	10219	61549	9417395	110448	1691696	798397	12079433
South East	419	65429	8244863	51803	1582812	975887	10920794
Scotland	15415	43528	7796663	115061	1318192	743447	10016891
Yorkshire and the Humber	9	15093	4499004	18217	1055161	1006015	6593490
East Midlands	96	12923	3248653	16834	727813	773928	4780168
South West	3265	12617	2530596	16093	459452	381409	3400167
London	2	42085	2519228	29802	433491	193210	3217816
West Midlands	1	7320	2009116	9272	433648	435781	2695137
East of England	26	5633	809325	6642	141011	174341	1136952

Query executed successfully.

Traffic by Local Authority



File Edit View Query Git Project Tools Extensions Window Help | Search | Solution1

TrafficDataWarehouse | Execute | DESKTOP-KPL988A (54) | Diagram_0

```

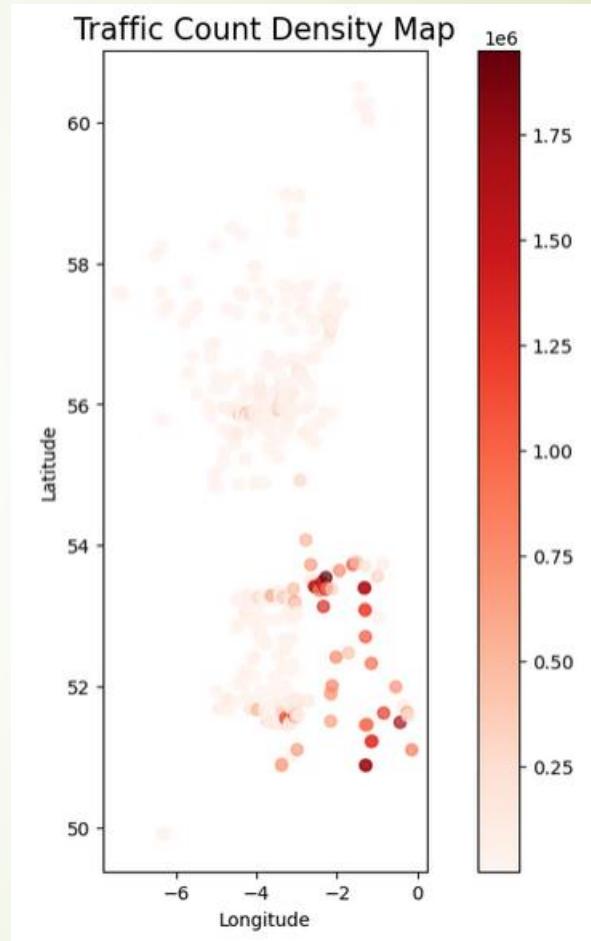
SELECT
    la.local_authority_name,
    r.region_name,
    SUM(f.pedal_cycles) AS total_pedal_cycles,
    SUM(f.two_wheeled_motor_vehicles) AS total_motorbikes,
    SUM(f.cars_and_taxis) AS total_cars,
    SUM(f.buses_and_coaches) AS total_buses,
    SUM(f.LGVs) AS total_LGVs,
    SUM(f.all_HGVs) AS 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;
  
```

Results Messages

local_authority_name	region_name	total_pedal_cycles	total_motorbikes	total_cars	total_buses	total_LGVs	total_HGVs	total_vehicles
Warrington	North West	1	11276	2967063	14332	546385	258723	3830893
Harpenden	South East	399	24480	2334286	14353	404308	200601	2960475
Manchester	North West	4	6925	1950993	8053	906051	375975	2846955
Bury	North West	5	7613	1607361	6789	364227	289660	2272451
Leeds	Yorkshire and the Humber	0	5467	1590259	6252	380441	343010	2325429
Rotherham	Yorkshire and the Humber	2	5474	178577	30617	292277	178934	2299566
Newport	Wales	1085	11161	178577	30617	292277	178934	2299566
Worcestershire	West Midlands	1	5985	1598539	7640	316432	307917	2284113
Glasgow City	Scotland	4421	7041	1692015	19712	292002	146061	2157631
Rhondda, Cynon, Taff	Wales	751	10436	1697992	16255	320053	101789	2146525
Kent	South East	12	15195	148430	10064	352078	276773	2140540
Hillingdon	London	2	24205	1699526	20276	262063	121642	2120512
Salford	North West	0	5774	1531796	4921	286888	154281	1983660
Derbyshire	East Midlands	0	4013	1156100	5136	259526	251119	1600902
East Cheshire	North West	0	3174	1145901	4670	255489	246194	1655427
Truro	North West	6	4604	1313319	3365	224972	105827	1652087
Lancashire	North West	0	6158	1173768	7951	220198	202650	1610620
Buckinghamshire	South East	0	7195	1235958	11181	202671	141907	1596152
Leicestershire	East Midlands	5	4544	1090029	5162	239123	240462	1579320
Cheshire	North West	45	9637	1031190	48075	164530	211360	1411092

Query executed successfully.

Plot:



Traffic Trend by Region (Over Time)

File Edit View Query Git Project Tools Extensions Window Help Search Solution Sign in ADMIN

Object Explorer

SQLQuery1.sql (54)* DESKTOP-KPL988A\HP (54)* DESKTOP-KPL988A\HP (54)* Diagram_0*

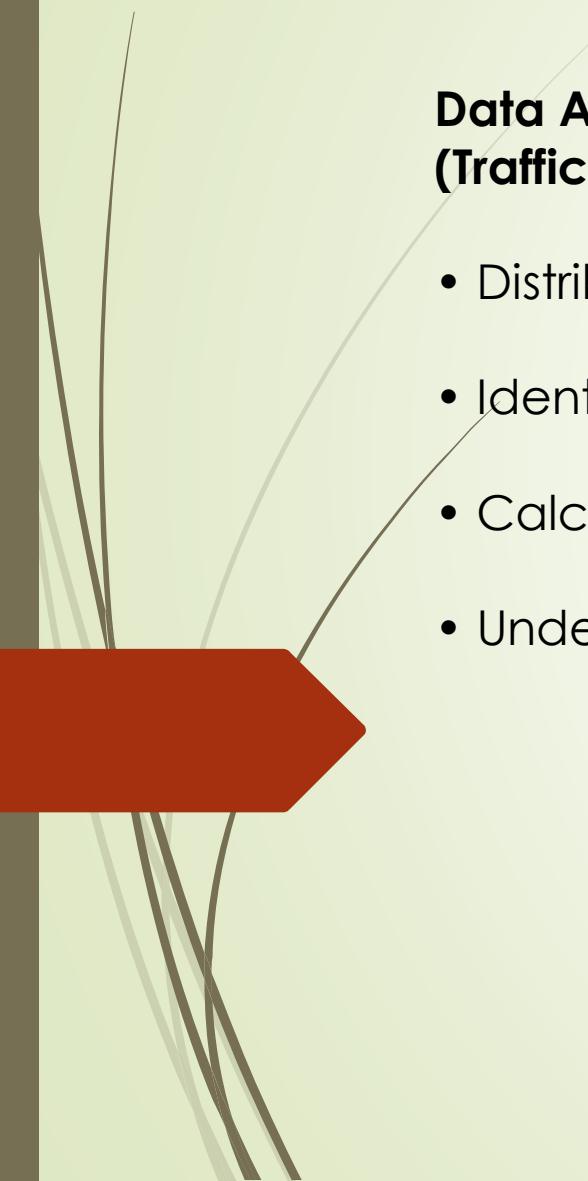
```
1 SELECT
2     r.region_name,
3     f.year,
4     SUM(f.all_motor_vehicles) AS total_vehicles
5 FROM FACT_TRAFFIC_COUNTS f
6 JOIN DIM_REGION r ON f.region_id = r.region_id
7 GROUP BY r.region_name, f.year
8 ORDER BY r.region_name, f.year;
```

Results Messages

	region_name	year	total_vehicles
1	East Midlands	2000	236346
2	East Midlands	2001	251187
3	East Midlands	2002	243603
4	East Midlands	2003	180077
5	East Midlands	2004	251333
6	East Midlands	2005	180691
7	East Midlands	2006	263048
8	East Midlands	2007	182799
9	East Midlands	2008	240280
10	East Midlands	2009	154751
11	East Midlands	2010	172029
12	East Midlands	2011	245962
13	East Midlands	2012	250319
14	East Midlands	2013	260792
15	East Midlands	2014	268246
16	East Midlands	2015	246998
17	East Midlands	2016	263159
18	East Midlands	2017	103737
19	East Midlands	2018	14020
20	East Midlands	2019	178107

Query executed successfully.

DESKTOP-KPL988A (16.0 RTM) DESKTOP-KPL988A\HP (54) TrafficDataWarehouse 00:00:00 | 231 rows



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

```
SELECT  
    YEAR(count_date) AS [Year],  
    MONTH(count_date) AS [Month],  
    SUM(all_motor_vehicles) AS Total_Motor_Vehicles  
FROM FACT_TRAFFIC_COUNTS  
GROUP BY YEAR(count_date), MONTH(count_date)  
ORDER BY [Year], [Month];
```

Results:

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	507996
2001	5	579671
2001	6	1196945
2001	7	336480
2001	9	646266
2001	10	317703
2002	3	101736
2002	4	594974
2002	5	813541
2002	6	712596
2002	7	42595
2002	9	830921

Query executed successfully.

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

4. Analysis by day of the week

```
SELECT  
    DATENAME(WEEKDAY, count_date) AS [DayOfWeek],  
    AVG(all_motor_vehicles) AS Avg_Vehicles,  
    SUM(all_motor_vehicles) AS Total_Vehicles  
FROM FACT_TRAFFIC_COUNTS  
GROUP BY DATENAME(WEEKDAY, count_date), DATEPART(WEEKDAY, count_date)  
ORDER BY DATEPART(WEEKDAY, count_date);
```

Results:

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

```
SELECT  
    hour AS [Hour],  
    AVG(all_motor_vehicles) AS Avg_Vehicles,  
    SUM(all_motor_vehicles) AS Total_Vehicles  
FROM FACT_TRAFFIC_COUNTS  
GROUP BY hour  
ORDER BY [Hour];
```

Results:

Hour	Avg_Vehicles	Total_Vehicles
1	1796	6957078
2	1821	7052046
3	1493	5784849
4	1366	5293555
5	11382	535329
6	1408	5456238
7	1459	5653019
8	1529	5922637
9	1661	6437315
10	1896	7344737
11	1920	7438942
12	1542	5974202

Query executed successfully.

5. Direction Analysis (Direction of Travel)

The screenshot shows the Object Explorer on the left with the 'TrafficDataWarehouse' database selected. The 'Tables' node is expanded, showing various tables like 'direction_of_travel', 'Avg_Traffic', 'Total_Traffic', etc. A query window titled 'SQLQuery1s...88A\HP (54)*' is open with the following T-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;
```

The results grid shows the following data:

direction_of_travel	Avg_Traffic	Total_Traffic
N	1565	18627295
S	1559	18700640
W	1676	18628959
E	1656	18410902
C	445	101551

Message bar at the bottom: 'Query executed successfully.'

6. Road Analysis Road Name

The screenshot shows the Object Explorer on the left with the 'TrafficDataWarehouse' database selected. The 'Tables' node is expanded, showing various tables like 'road_name', 'Observations', 'Total_Traffic', etc. A query window titled 'SQLQuery1s...88A\HP (54)*' is open with the following T-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;
```

The results grid shows the following data:

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	1121662
M2	384	1029204
A494	552	832651
M621	264	628968

Message bar at the bottom: 'Query executed successfully.'

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

The screenshot shows the Object Explorer and a query editor window in SQL Server Management Studio. The query is:

```
SELECT
    road_name,
    hour,
    AVG(all_motor_vehicles) AS Avg_Vehicles
FROM FACT_TRAFFIC_COUNTS
GROUP BY road_name, hour
ORDER BY road_name, hour;
```

The results table shows traffic data for various roads and hours:

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 Object Explorer and a query editor window in SQL Server Management Studio. The query is:

```
SELECT
    year,
    SUM(all_motor_vehicles) AS Total_Traffic,
    LAG(SUM(all_motor_vehicles)) OVER (ORDER BY year) AS Prev_Year_Traffic,
    (SUM(all_motor_vehicles) - LAG(SUM(all_motor_vehicles)) OVER (ORDER BY year)) * 100.0 / LAG(SUM(all_motor_vehicles)) OVER (ORDER BY year) AS Growth_Rate_Percent
FROM FACT_TRAFFIC_COUNTS
GROUP BY year
ORDER BY year;
```

The results table shows the annual growth rate of traffic:

year	Total_Traffic	Prev_Year_Traffic	Growth_Rate_Percent
2001	9962007	4118974	-3.30587524164
2002	3546950	9962007	-10.903190066526
2003	3417423	3546950	-3.69527636566
2004	3615371	3417423	5.83743198779
2005	3620036	3615371	1.145707070741
2006	4089717	3620036	27.67804235551
2007	3672616	4089717	-5.285300980060
2008	3760354	3672616	-2.692299623504
2009	3639762	3760354	-3.49841939939
2010	2994595	3639762	-17.70336874529
2011	3171659	2994595	5.912786203142
2012	3071319	3171659	-3.163644030930
2013	304313K	3071319	-0.917638780208
2014	3116489	304313K	2.410441071315
2015	2279241	3116489	-26.66530547094
2016	2981765	2279241	30.822716969096
2017	2553399	2981765	-14.369379241255
2018	1052090	2553399	-58.794715365162
2019	2860687	1052090	171.905160204925
2020	1574951	2860687	-46.71911949410



Thank
you! :)

