



London, or as the Romans called it "Londonium"! Home to [over 8.5 million residents](#) who speak over [300 languages](#). While the City of London is a little over one square mile (hence its nickname "The Square Mile"), Greater London has grown to encompass 32 boroughs spanning a total area of 606 square miles!



Given the city's roads were originally designed for horse and cart, this area and population growth has required the development of an efficient public transport system! Since the year 2000, this has been through the local government body called **Transport for London**, or *TfL*, which is managed by the London Mayor's office. Their remit covers the London Underground, Overground, Docklands Light Railway (DLR), buses, trams, river services (clipper and [Emirates Airline cable car](#)), roads, and even taxis.

The Mayor of London's office make their data available to the public [here](#). In this project, you will work with a slightly modified version of a dataset containing information about public transport journey volume by transport type.

The data has been loaded into a **Snowflake** database called `TFL` with a single table called `JOURNEYS`, including the following data:

### TFL.JOURNEYS

Column	Definition	Data type
<code>MONTH</code>	Month in number format, e.g., <code>1</code> equals January	<code>INTEGER</code>
<code>YEAR</code>	Year	<code>INTEGER</code>
<code>DAYS</code>	Number of days in the given month	<code>INTEGER</code>
<code>REPORT_DATE</code>	Date that the data was reported	<code>DATE</code>
<code>JOURNEY_TYPE</code>	Method of transport used	<code>VARCHAR</code>
<code>JOURNEYS_MILLIONS</code>	Millions of journeys, measured in decimals	<code>FLOAT</code>

Note that in *Snowflake* all databases, tables, and columns are **upper case** by default.

You will execute SQL queries to answer three questions, as listed in the instructions.

London Public Transport DataFrame as most\_popular\_transport\_types

```
-- most_popular_transport_types
select journey_type, sum(journeys_millions) as total_journeys_millions from TFL.JOURNEYS
group by TFL.JOURNEYS.JOURNEY_TYPE
order by total_journeys_millions desc;
```

index	...	↑↓	JOURNEY_TYPE	...	↑↓	TOTAL_JOURNEYS_MILLIONS
		0	Bus			24
		1	Underground & DLR			150
		2	Overground			166
		3	TfL Rail			41
		4	Tram			31
		5	Emirates Airline			1

Rows: 6

[Expand](#)

London Public Transport DataFrame as emirate:

```
-- emirates_airline_popularity
select month, year, round(journeys_millions, 2) as rounded_journeys_millions from TFL.JOURNEYS
where rounded_journeys_millions is not null and
journey_type='Emirates Airline'
order by rounded_journeys_millions desc limit 5
```

index	...	↑↓	MONTH	...	↑↓	YEAR	...	↑↓	ROUNDED_JOURNEYS_MILLIONS
		0			5			2012	
		1			6			2012	
		2			4			2012	
		3			5			2013	
		4			5			2015	

Rows: 5

[Expand](#)

London Public Transport DataFrame as leas

```
-- least_popular_years_tube
select year, journey_type, sum(journeys_millions) as total_journeys_millions from TFL.JOURNEYS
where journey_type like '%Underground%'
group by year, journey_type
order by total_journeys_millions limit 5;
```

...	↑↓	...	↑↓	JOURNEY_TY...	...	↑↓	TOTAL_JOURNEYS_MIL...	...	↑↓
	0		2020	Underground & DLR			310.179316314		
	1		2021	Underground & DLR			748.4525442		
	2		2022	Underground & DLR			1064.8590086		
	3		2010	Underground & DLR			1096.14558838		
	4		2011	Underground & DLR			1156.64765448		

Rows: 5

[Expand](#)