datalab

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

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

```sql
SELECT COUNT(*) FROM TFL.JOURNEYS
```

⌄	COUNT(*)
0	

◄ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ ►
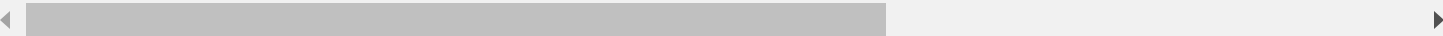
1 rows ⬇

❄ London Public Transport   DataFrame as

```sql
SELECT * FROM TFL.JOURNEYS
```

	MONTH	YEAR	DAYS	REPORT_DATE
0	1	2010	31	2010-06-30T00:00:00.000
1	2	2010	28	2010-06-30T00:00:00.000
2	3	2010	31	2010-06-30T00:00:00.000
3	4	2010	30	2010-09-30T00:00:00.000
4	5	2010	31	2010-09-30T00:00:00.000
5	6	2010	30	2010-09-30T00:00:00.000
6	7	2010	31	2010-12-31T00:00:00.000
7	8	2010	31	2010-12-31T00:00:00.000
8	9	2010	30	2010-12-31T00:00:00.000
9	10	2010	31	2011-03-31T00:00:00.000
10	11	2010	30	2011-03-31T00:00:00.000
11	12	2010	31	2011-03-31T00:00:00.000
12	1	2011	31	2011-06-30T00:00:00.000
13	2	2011	28	2011-06-30T00:00:00.000
14	3	2011	31	2011-06-30T00:00:00.000

936 rows

❄️ London Public Transport   DataFrame as

```sql
-- most_popular_transport_types
SELECT JOURNEY_TYPE, SUM(JOURNEYS_MILLIONS) AS TOTAL_JOURNEYS_MILLIONS
FROM TFL.JOURNEYS
GROUP BY JOURNEY_TYPE
ORDER BY TOTAL_JOURNEYS_MILLIONS DESC
```

	JOURNEY_TYPE	TOTAL_JOURNEYS_MILLION$
0	Bus	
1	Underground & DLR	
2	Overground	
3	TfL Rail	
4	Tram	
5	Emirates Airline	

6 rows

❄️ London Public Transport   DataFrame as

```sql
-- emirates_airline_popularity
```

```
SELECT MONTH, YEAR, ROUND(SUM(JOURNEYS_MILLIONS),2) AS ROUNDED_JOURNEYS_MILLIONS
FROM TFL.JOURNEYS
WHERE JOURNEY_TYPE = 'Emirates Airline'
    AND JOURNEYS_MILLIONS IS NOT NULL
GROUP BY ALL
ORDER BY ROUNDED_JOURNEYS_MILLIONS DESC
LIMIT 5
```

	MONTH	YEAR	ROUNDED_JOURNEYS_MI
0	5	2012	
1	6	2012	
2	4	2012	
3	5	2013	
4	5	2015	

5 rows

London Public Transport   DataFrame as

```
-- least_popular_years_tube
SELECT YEAR, JOURNEY_TYPE, SUM(JOURNEYS_MILLIONS) AS TOTAL_JOURNEYS_MILLIONS
FROM TFL.JOURNEYS
WHERE JOURNEY_TYPE = 'Underground & DLR'
GROUP BY ALL
ORDER BY TOTAL_JOURNEYS_MILLIONS
LIMIT 5
```

	YEAR	JOURNEY_TYPE	TOTAL_JOU
0	2020	Underground & DLR	
1	2021	Underground & DLR	
2	2022	Underground & DLR	
3	2010	Underground & DLR	
4	2011	Underground & DLR	

5 rows