PROJECT: EXPLORING LONDON'S TRAVEL NETWORK

London, or as the Romans called it "Londonium"! As of 2021, Greater London is 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 have 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 makes 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 **Databricks** database containing a schema called `tfl` with a single table called `journeys`. The table, which you will use for the project, includes 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`

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

Projects Data 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
  journey_type
ORDER BY
  total_journeys_millions DESC;
```

index	journey_type
0	Bus
1	Underground & DLR
2	Overground
3	TfL Rail
4	Tram
5	Emirates Airline

Rows: 6 Expand

Projects Data DataFrame as `emirates_airline_popularity`

```
-- emirates_airline_popularity
SELECT
  month,
  year,
  ROUND(journeys_millions, 2) AS rounded_journeys_millions
FROM
  tfl.journeys
WHERE
  journey_type = 'Emirates Airline'
  AND journeys_millions IS NOT NULL
ORDER BY
  rounded_journeys_millions DESC
LIMIT 5;
```

index	month	year	r
0	5	2012	
1	6	2012	
2	4	2012	
3	5	2013	
4	5	2015	

Rows: 5 Expand

Projects Data DataFrame as `least_popular_years_tube`

```
-- 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 ASC
LIMIT 5;
```

index	year	journey_type
0	2020	Underground & DLR
1	2021	Underground & DLR
2	2022	Underground & DLR
3	2010	Underground & DLR
4	2011	Underground & DLR

Rows: 5 Expand