\*\*Background:\*\*

A UK-based government organization is analyzing travel expenses incurred by different departments (e.g., Yellow Department, Orange Department) for internal and external meetings across domestic and international destinations. The flights were booked through various airlines including Flybe, EasyJet, and British Airways, and data includes single and return trips made by employees over several months.

\*\*Problem Statement:\*\*

The organization’s finance and audit teams aim to understand the travel expenditure trends to ensure policy compliance, detect anomalies, and improve cost-efficiency. They want a detailed analysis of costs incurred per department, frequent origin-destination pairs, average distances traveled, and popular airlines used. There's also interest in identifying whether return tickets are consistently cheaper or more expensive than two single tickets and if long-haul trips are justified based on cost per kilometer.

Furthermore, the management wants insights into the impact of travel class, distance, and ticket type on the cost. There's also a requirement to generate derived metrics such as travel cost per kilometer, frequent flyer segments by distance range, and monthly department-wise travel summaries.

---

### Scenario-Based Analytical Problems (with Logical/Derived Columns)

Below are problems designed to require transformation and logic-building:

---

#### \*\*1. Cost Efficiency Analysis\*\*

\*\*Objective:\*\*

Compare `Return` vs `Single` tickets for each department.

\*\*Derived Column:\*\*

\* `Cost\_Per\_KM` = `Cost / Distance`

\* `Is\_Return` = If `Ticket Single or Return = Return`, then 1, else 0

\*\*Business Question:\*\*

Which departments spend more per kilometer on `Single` vs `Return` tickets?

---

#### \*\*2. Departmental Spend Trend by Month\*\*

\*\*Objective:\*\*

Track how each department's travel cost changes over time.

\*\*Derived Column:\*\*

\* `Travel\_Month` = Extract month from `Travel Date`

\* `Travel\_Year` = Extract year from `Travel Date`

\*\*Business Question:\*\*

Which department had the highest spend in March 2011? Is there any seasonal trend?

---

#### \*\*3. Airline Preference by Distance Bucket\*\*

\*\*Objective:\*\*

Determine which airlines are used for short (<600 km), medium (600–1000 km), and long-haul flights (>1000 km).

\*\*Derived Column:\*\*

\* `Distance\_Category`:

\* Short: `<600`

\* Medium: `600–1000`

\* Long: `>1000`

\*\*Business Question:\*\*

Which airline is dominant for long-haul vs short-haul segments?

---

#### \*\*4. Frequent Route Identification\*\*

\*\*Objective:\*\*

Identify the most frequently traveled routes (Origin → Destination) and total cost per route.

\*\*Derived Column:\*\*

\* `Route\_ID` = CONCAT(`Origin ICAO`, '→', `Destination ICAO`)

\*\*Business Question:\*\*

Which routes are most frequent and which are most expensive overall?

---

#### \*\*5. Outlier Cost Detection\*\*

\*\*Objective:\*\*

Identify outlier flights with unusually high cost per kilometer.

\*\*Derived Column:\*\*

\* Use `Cost\_Per\_KM` and flag entries where it is 2× higher than department's average.

\*\*Business Question:\*\*

Are there specific routes or airlines contributing to cost inefficiencies?

---

#### \*\*6. Geo Analytics and Region Spend Distribution\*\*

\*\*Objective:\*\*

Evaluate travel cost distribution across different destination regions (e.g., Scotland, England, Poland, Morocco).

\*\*Derived Column:\*\*

\* Group by `Destination Region`

\* Aggregate `Cost`, `Distance`, `Number of Tickets`

\*\*Business Question:\*\*

Which destination region receives the most travel volume and budget?

---

#### \*\*7. Carbon Footprint Proxy Calculation\*\*

\*\*Objective:\*\*

Estimate travel impact using distance as a proxy for carbon emissions.

\*\*Derived Column:\*\*

\* `Estimated\_CO2\_Emissions` = `Distance × 0.133 kg/km` (assumption)

\*\*Business Question:\*\*

Which department contributes most to emissions? Can the trips be optimized?