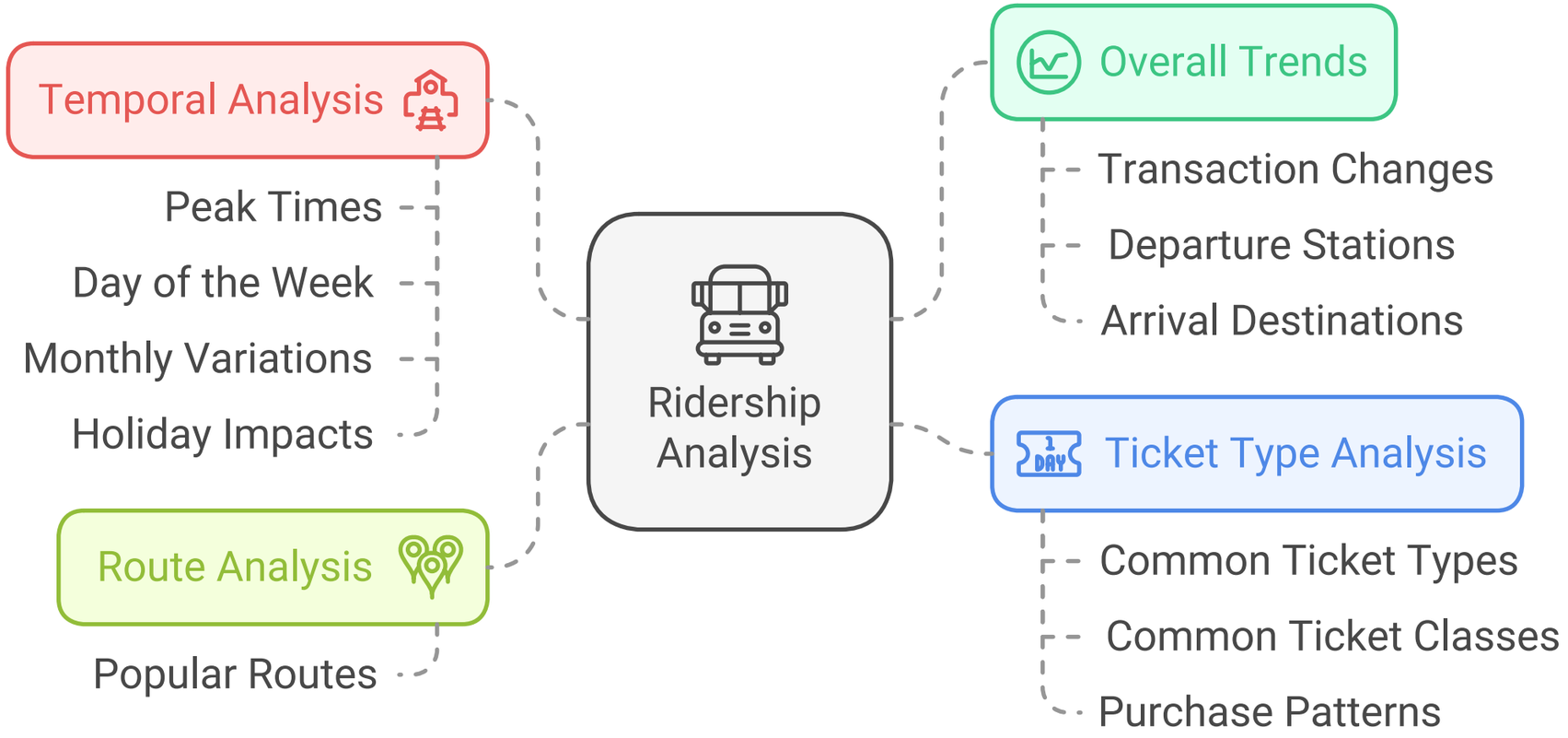
🎟 Analyzing UK Train Rides

Ridership-Focused Questions

(Based on Transaction ID, Date of Journey, Departure Time and Departure Station):

- **Overall Trends:**
  - How has the number of transactions changed over time? (Use Date of Purchase)
  - What is the distribution of departure stations? (Use Departure Station)
  - What is the distribution of arrival destinations? (Use Arrival Destination)
- **Temporal Analysis:**
  - What are the peak times for train rides? (Use Departure Time broken down by hours)
  - How does ridership vary by day of the week? (Derive day of the week from Date of Journey)
  - Are there any noticeable differences in ridership across different months? (Derive from Date of Journey)
  - Are there any noticeable differences in ridership across different purchase times? (Use Time of Purchase broken down by hours)
  - Do holidays impact ridership? (This might require adding holiday information)
- **Route Analysis:**
  - Which routes (combination of Departure Station and Arrival Destination) are the most popular (highest ridership)?

- What are the most common ticket types that passengers buy? (Use Ticket Type)
- What are the most common ticket classes that passengers buy? (Use Ticket Class)
- Are their any patterns on the purchase of ticket classes based on the time of purchase? (Use Time of Purchase and Ticket Class)



Revenue-Focused Questions

(Based on Price, Ticket Class, Ticket Type and Railcard):

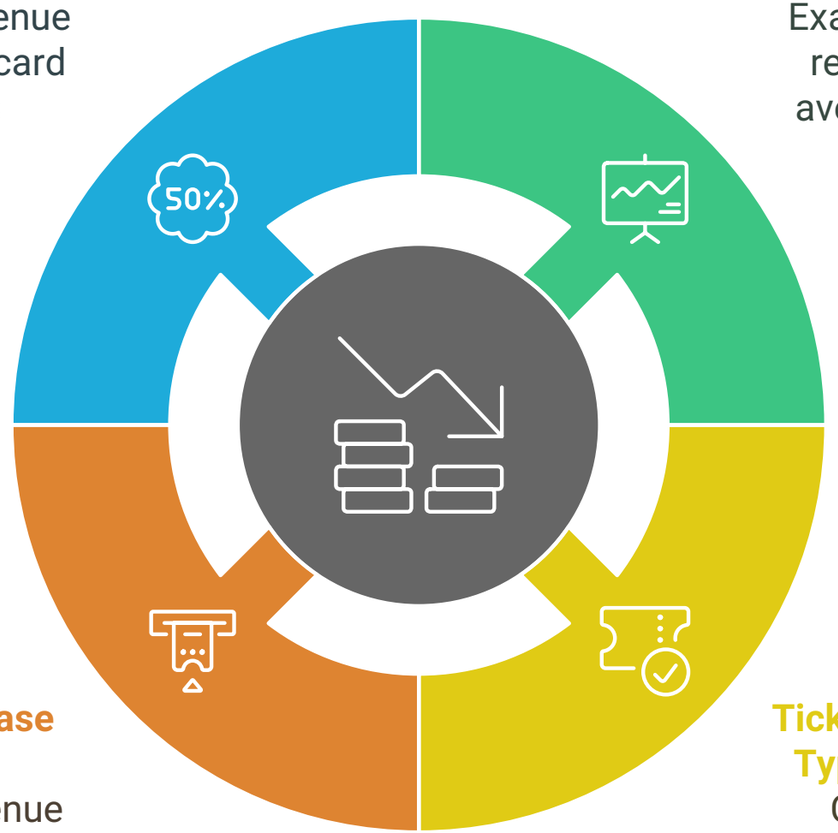
- **Overall Revenue:**
  - What is the total revenue generated? (Sum of Price)
  - What is the average ticket price? (Average of Price)
  - How does revenue fluctuate on a daily, weekly, or monthly basis?
- **Ticket Class and Type Revenue:**
  - Which Ticket Class (Standard or First) contributes more to the revenue?
  - Which Ticket Type (Advance, Off-Peak, Anytime) generates the most revenue?
- **Ticket Purchase Methods**
  - Do tickets purchased online have a different average price than those at the station? (Use Price and Purchase Type)
  - Which payment method contributes the most to the overall revenue? (Use Payment Method and Price)
  - Is there a difference in price based on payment method? (Use Price and Payment Method)
- **Impact of Railcards:**
  - How many passengers use railcards? (Count of Railcard not equal to 'None')
  - How much revenue is potentially lost because of railcard discounts? (Compare the total price of tickets sold using discount from each railcard, and compare with tickets sold without the discount)

- Is their a pattern on the purchase of tickets using a specific railcard based on the day of the week? (Use Railcard and derive day of the week from Date of Journey)

Revenue Analysis Breakdown

Impact of Railcards
Evaluates revenue impact of railcard discounts

Overall Revenue
Examines total revenue and average ticket price



Ticket Purchase Methods
Analyzes revenue based on purchase methods and payment types

Ticket Class and Type Revenue
Compares revenue from different ticket classes and types

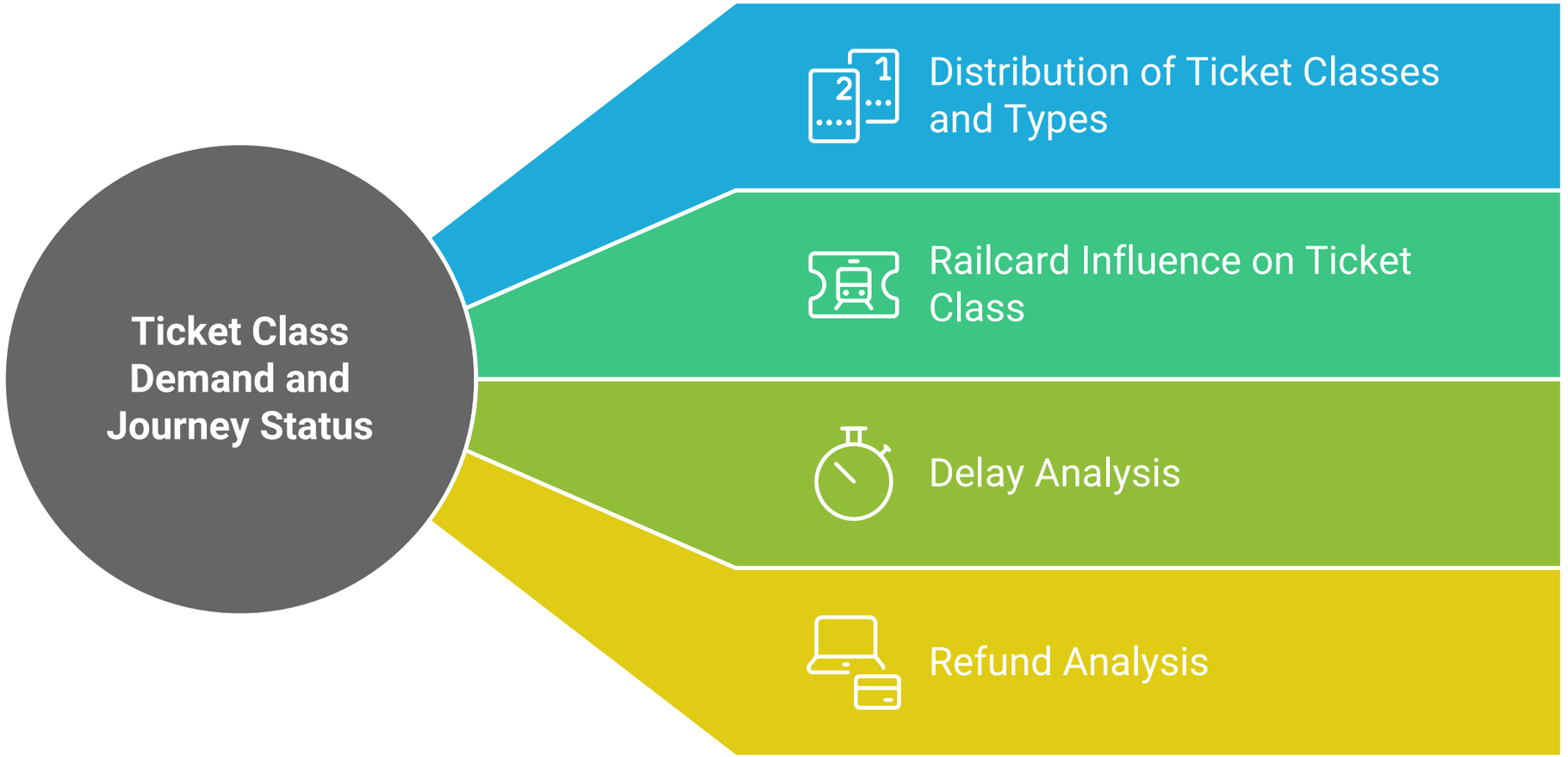
Ticket Class Demand Questions (Based on Ticket Class, Ticket Type, Railcard and Departure Time):

- **Distribution of Classes and Types:**
    - What is the proportion of passengers traveling in Standard vs First Class?
    - How many passengers used an Advance, Off-Peak, or Anytime ticket?
    - Is there a preference for Ticket Class based on specific Departure Time?
- **Railcard Influence on Ticket Class:**
    - Are there particular Railcard holder types who prefer one Ticket Class over another?

Journey Status and Delays:

- **Delay Analysis:**
    - What is the overall proportion of on-time, delayed, and cancelled journeys? (Use Journey Status)
    - What are the most common Reason for Delay?
    - Which stations tend to have more delays or cancellations? (Use Departure Station and Journey Status)
- **Refund Analysis:**
- How many passengers requested a refund? (Count Refund Request with 'Yes')
- How many passengers requested a refund due to cancellation versus delays? (Use Refund Request, Journey Status and Reason for Delay)
- Do certain routes have a higher refund requests than others? (Use Departure Station, Arrival Destination, and Refund Request)

Exploring Ticket Class Demand and Journey Delays



Prediction Focused Questions

- Can we predict the amount of ridership for future dates based on historical data?
- Can we predict the revenue for future dates based on historical data?
- Can we predict the demand for ticket classes for future dates?

Important Considerations:

- **Data Cleaning:** The provided data might need cleaning (handling missing values, inconsistencies in data types) before analysis.
- **Date and Time Parsing:** Converting Date of Purchase, Date of Journey, Time of Purchase, Departure Time, Arrival Time, and Actual Arrival Time to appropriate date/time formats is important.
- **Data Aggregation:** Data will likely need to be aggregated (e.g., daily counts, monthly sums) to answer some of the questions.
- **Data Exploration:** It is useful to generate preliminary analyses and aggregations to get a better sense of data distribution and quality before deep-diving on specific questions.

Challenges in Predictive Analysis for Ridership and Revenue