Data-Driven Analysis Questions

1) Which reasons of delays occur most frequently, and how do they vary by station or ticket type?
2) How does the difference between scheduled and actual departure times reveal recurring operational issues?
3) Which arrival stations consistently show the longest gaps between scheduled and actual arrival times?
4) What is the estimated financial loss caused by delays and cancellations through refund requests?
5) How does the Ticket Issue Date influence the likelihood of a journey being on-time, delayed, or cancelled?
6) Are customers using electronic payment methods more loyal, with more repeat purchases and fewer refund requests?
7) How can revenue be maximized by analyzing refund patterns across peak and off-peak tickets?
8) Which Train IDs demonstrate the highest reliability or the worst punctuality performance?
9) Are there noticeable patterns in cancellations depending on ticket type, station, or time of day?
10) How do peak-time and off-peak journeys differ in terms of average delays and cancellations?
11) Which departure or arrival stations stand out with consistently higher delay or cancellation rates?
12) Based on historical data, which stations are forecasted to become the busiest and which ticket types most demanded?
13) What is the relationship between journey status (on-time, delayed, cancelled) and refund requests?
14) How accurate are scheduled arrival times compared to actual arrival times across all stations?
15) What is the overall on-time arrival rate for each station or line, and how has it changed over time?
16) Which ticket types generate the highest sales volumes across different routes and times of day?
17) What is the share of refund requests out of total ticket sales, and which factors influence them the most?
18) Which payment methods dominate in different stations, and do preferences differ between first-time and repeat travelers?
19) How does Passenger ID data help in identifying frequent travelers versus one-time users?
20) What is the distribution of Ticket IDs, and are there any signs of duplication or fraud?
21) What are the differences in payment patterns (cash vs. electronic) between frequent customers and new customers?

22) Are there journeys that consistently show high occupancy rates compared to others with low occupancy?

23) How can train distribution across different routes be optimized to reduce waste and improve operational efficiency?

24) Do advance bookings reduce cancellation rates compared to last-minute ticket purchases?

25) Which ticket categories (First/Second class, Peak/Off-Peak) are more prone to cancellations or delays?

26) What is the average dwell time at each station, and do certain stations cause congestion or delays for other trains?

27) Is there a correlation between the number of trains on the same route and the total delay time?

28) What are the peak time periods (hours/days) when congestion occurs most frequently across different routes?

29) What is the average travel frequency of repeat customers, and does it relate to their payment method or route choice?

30) Which regions or stations attract the highest number of new customers compared to others?

31) Are there behavioral differences between morning and evening journeys in terms of delays, cancellations, or occupancy?

32) What is the relationship between ticket price and customer satisfaction (as reflected in refunds/cancellations)?

33) Are long-distance journeys more vulnerable to delays than short-distance ones?

34) Based on historical data, which routes are expected to see increased demand in the coming year?

35) What is the probability that a customer cancels their ticket based on booking time and payment method?

36) How consistent is train punctuality across different seasons and holidays?