

Logical Schema:

TIMESLOT(TSID, Is\_Peak)

TIME(TID, Day, Month, 2M, 3M, Year)

TRANSPORTATION(TRID, Transporation\_Mode, RID)

ROUTE(RID, AC, Wi-Fi, Special\_Seats, Start\_Stop, End\_Stop)

JOURNEY(TSID, TID, TRID, JID, Duration, Revenue, Type, Purchase\_method, Discount)

Task 2 Extended SQL Queries

a. 1 journey = 1 ticket

SELECT TR.Transportation\_Mode, T.Year, T.Month, COUNT(JID) / COUNT(DISTINCT Day) AS AvgNumTickets,   
SUM(COUNT(JID)) OVER (PARTITION BY Year ORDER BY Month ROWS UNBOUNDED PRECEDING) AS TrailingYearlyNumTickets,  
COUNT(JID) \* 100 / SUM(COUNT(JID)) OVER (PARTITION BY Year, Month) AS PercentageTickets  
FROM TRANSPORTATION TR, TIME T, JOURNEY J  
WHERE J.TID = T.TID AND J.TRID = TR.TRID  
GROUP BY TR.Transportation\_Mode, T.Year, T.Month

b.

WITH RevenuePerRoute AS (  
SELECT SUM(J.Revenue) OVER (PARTITION BY TR.Transportation\_Mode, R.Start\_Stop, R.RID)   
FROM JOURNEY J, ROUTE R, TRANSPORTATION TR, TIME T  
WHERE J.TRID = TR.TRID AND J.RID = R.RID AND T.TID = J.TID  
AND T.Year = 2022),

RevenuePerTRModeAndCity AS (

SELECT SUM(J.Revenue)  
OVER (PARTITION BY TR.Transportation\_Mode, TR.City)  
FROM TRANSPORTATION TR, JOURNEY J, TIME T  
WHERE J.TRID = T.TRID AND J.TID = T.TID  
AND T.Year = 2022  
)

SELECT Transportation\_Mode, City, AVG(DURATION) AS JourneyDuration,   
SUM(REVENUE) OVER (PARTITION BY City) AS RevenuePerCity,  
RevenuePerRoute \* 100.0 / RevenuePerTRModeAndCity AS PercentageRevenuePerRoute,  
RANK() OVER (ORDER BY RevenuePerRoute, DESC) AS RankRoute  
FROM JOURNEY J , TRANSPORTATION TR, TIME T  
WHERE J.TRID = TR.TRID AND J.TID = T.TID  
AND T.Year = 2022  
GROUP BY Transportation\_Mode, City