**Hive queries**

1. **What are the weekday routes with the highest number of tickets validated?(comparing two months to each other)**

SELECT r.route\_id, SUM(CASE WHEN d.month\_no = 4 THEN t.tickets\_validated ELSE 0 END) AS total\_tickets\_april, SUM(CASE WHEN d.month\_no = 5 THEN t.tickets\_validated ELSE 0 END) AS total\_tickets\_may, (SUM(CASE WHEN d.month\_no = 5 THEN t.tickets\_validated ELSE 0 END)-SUM(CASE WHEN d.month\_no = 4 THEN t.tickets\_validated ELSE 0 END)) AS tickets\_difference\_to\_prev\_month

FROM travel t

JOIN route r ON t.route\_id = r.route\_id

JOIN date\_dim d ON t.travel\_date = d.date\_id

WHERE d.month\_no IN (4, 5) AND d.day\_type= 'weekday'

GROUP BY r.route\_id

ORDER BY total\_tickets\_april DESC

LIMIT 10;

1. **Which bus type has the highest average satisfaction level?**

SELECT b.bus\_type, AVG(t.avg\_satisfaction\_level\_received) AS avg\_satisfaction

FROM travel t

JOIN bus b ON t.bus\_id = b.bus\_id

GROUP BY b.bus\_type

ORDER BY avg\_satisfaction DESC

LIMIT 1;

1. **What is the number of validated tickets for each bus comparing across weekdays and weekends in May?**

SELECT t.bus\_id, d.day\_type, SUM(t.tickets\_validated) AS total\_validated\_tickets

FROM travel t

JOIN date\_dim d ON t.travel\_date = d.date\_id

WHERE d.month\_no = 5 AND d.day\_type IN ('weekday', 'weekend')

GROUP BY t.bus\_id, d.day\_type

ORDER BY t.bus\_id, d.day\_type;

1. **What was the busiest time of the day for bus routes on weekends in April?**

SELECT td.time\_of\_day, COUNT(\*) AS trip\_num   
FROM travel t   
JOIN date\_dim d ON t.travel\_date = d.date\_id   
JOIN time\_dim td ON t.departure\_time = td.time\_id   
WHERE month\_no = 4 AND d.day\_type = 'weekend'   
GROUP BY td.time\_of\_day   
ORDER BY trip\_num DESC;

1. **How many satisfaction surveys were received for a given route comparing two months (April and May)?**

SELECT r.route\_name, SUM(CASE WHEN d.month\_no = 4 THEN t.satisfaction\_surveys\_number ELSE 0 END) as surveys\_april, SUM(CASE WHEN d.month\_no = 5 THEN t.satisfaction\_surveys\_number ELSE 0 END) as surveys\_may

FROM travel t

JOIN route r ON t.route\_id = r.route\_id

JOIN date\_dim d ON t.travel\_date = d.date\_id

GROUP BY r.route\_name

ORDER BY surveys\_april DESC

LIMIT 5;

1. **How satisfaction level of bus users changed over time?**

SELECT d.month, AVG(t.avg\_satisfaction\_level\_received) AS avg\_satisfaction\_level

FROM travel t

JOIN date\_dim d ON t.travel\_date = d.date\_id

GROUP BY d.month

ORDER BY d.month;

1. **How does the satisfaction level vary based on occupation level categories?**

SELECT j.occupation\_level\_category, AVG(avg\_satisfaction\_level\_received) AS avg\_satisfaction\_level

FROM travel t

JOIN junk j ON t.junk\_id=j.junk\_id

GROUP BY j.occupation\_level\_category

ORDER BY avg\_satisfaction\_level DESC;

1. **What is the breakdown of travel counts by bus type that departures in the morning?**

SELECT b.bus\_type, COUNT(\*) travels\_num

FROM travel t

JOIN bus b ON t.bus\_id=b.bus\_id

JOIN time\_dim tdim ON t.departure\_time=tdim.time\_id

WHERE tdim.time\_of\_day = 'Morning'

GROUP BY b.bus\_type, tdim.time\_of\_day

ORDER BY travels\_num DESC;

1. **Which are the routes with the highest distance value?**

SELECT route\_id, route\_name, metrics['distance\_km'] AS distance

FROM route

ORDER BY distance DESC

LIMIT 10;

1. **What is the satisfaction level on weekend routes that have the highest avg duration? (in the selected month - April)**

SELECT t.route\_id, r.route\_name, AVG(r.metrics['duration\_min']) as avg\_duration,

AVG(t.avg\_satisfaction\_level\_received) AS avg\_satisfaction

FROM travel t

JOIN route r ON t.route\_id = r.route\_id

JOIN date\_dim d ON t.travel\_date = d.date\_id

WHERE d.month\_no = 4 AND d.day\_type = 'weekend'

GROUP BY t.route\_id, r.route\_name

ORDER BY avg\_duration DESC

LIMIT 10;