A picture containing logo

Description automatically generated Principles of Data Engineering– Spring 2023

Report II

(Queries in SQL)

For this Report II, your group has to solve 40 queries using SQL. Show your results and codes used.

To start with the project, first download and execute in SQL Server Management Studio the file

* 001\_SQL\_Tampa\_Railway\_Database\_Creation.sql

This file will create the database structure of the Tampa\_Railway database.

Then, to load the data, download and execute the file

* A\_SQL\_Tampa\_Railway\_Imports.sql

This file will import the data for all the tables from .csv files that are included with the project. These .csv files must be placed in a folder called C:\data. If those csv files are not placed in this folder, then the loading procedure will fail.

QUERIES:

1. In how many travels, trains departed with customers only using tickets in 2016?
2. In how many travels, trains departed with customers only using tickets in 2017?
3. In how many travels, trains departed with customers only using passes in 2016?
4. In how many travels, trains departed with customers only using passes in 2017?
5. In how many travels, trains departed with customers using both passes and tickets in 2016?
6. In how many travels, trains departed with customers using both passes and tickets in 2017?
7. Estimate the monthly average ratio of customers using tickets/passes in 2016.
8. Estimate the monthly average ratio of customers using tickets/passes in 2017.
9. For trains going from Tampa to Orlando, for each weekday, what is the most demanded hour in terms of number of customers?
10. For trains going from Orlando to Tampa, for each weekday, what is the most demanded hour in terms of number of customers?
11. Assuming that 3 wagons are assigned to each train in every travel, and each wagon has a capacity of 60 customers, what percentage of travels justify the usage of 3 wagons in 2016?
12. Assuming that 3 wagons are assigned to each train in every travel, and each wagon has a capacity of 60 customers, what percentage of travels justify the usage of 3 wagons in 2017?
13. Assuming that 3 wagons are assigned to each train in every travel, what is the lowest and highest monthly utilization of the wagons in 2016? In what months?(in terms of seats occupied)
14. Assuming that 3 wagons are assigned to each train in every travel, what is the lowest and highest monthly utilization of the wagons in 2017? In what months?(in terms of seats occupied)
15. Assuming that the breakeven point is 25 customers per travel, how many monthly travels should have been cancelled in 2016?
16. Assuming that the breakeven point is 25 customers per travel, how many monthly travels should have been cancelled in 2017?
17. What are the three employees that have sold the most passes in 2016?
18. What are the three employees that have sold the most passes in 2017?
19. What are the three employees that have sold the most tickets in 2016?
20. What are the three employees that have sold the most tickets in 2017?
21. What is the most sold cabin type in tickets in 2016?
22. What is the most sold cabin type in tickets in 2017?
23. What is the most sold cabin type in passes in 2016?
24. What is the most sold cabin type in passes in 2017?
25. What is the purchase location in which most tickets were sold in 2016?
26. What is the purchase location in which most tickets were sold in 2017?
27. What is the purchase location in which most passes were sold in 2016?
28. What is the purchase location in which most passes were sold in 2017?
29. What is the purchase location in which most cards were sold in 2016?
30. What is the purchase location in which most cards were sold in 2017?
31. How many travels did not have any customers in 2016?
32. How many travels did not have any customers in 2017?
33. What is the most used payment type in tickets in 2016?
34. What is the most used payment type in tickets in 2017?
35. What is the most used payment type in passes in 2016?
36. What is the most used payment type in passes in 2017?
37. What is the date where the most revenue (sum of final price) in tickets was collected in 2016?
38. What is the date where the most revenue (sum of final price) in tickets was collected in 2017?
39. What is the date where the most revenue (sum of final price) in passes was collected in 2016?
40. What is the date where the most revenue (sum of final price) in passes was collected in 2017?
41. What is the hour of the day where the most tickets were sold in 2016?
42. What is the hour of the day where the most tickets were sold in 2017?
43. What is the hour of the day where the most passes were sold in 2016?
44. What is the hour of the day where the most passes were sold in 2017?
45. What is the hour of the day where the most cards were sold in 2016?
46. What is the hour of the day where the most cards were sold in 2017?
47. What are the three cities where the most customers live in?
48. What are the two cities where the most employees live in?
49. What are the six zip codes where the most customers live in?
50. What are the six zip codes where the most employees live in?
51. What are the three customers that bought the most tickets in 2016?
52. What are the three customers that bought the most tickets in 2017?
53. What is the most used route on weekends (saturday, sunday) in terms of customers travelling including passes and tickets in 2016?
54. What is the most used route on weekends (saturday, sunday) in terms of customers travelling including passes and tickets in 2017?
55. What is the most used route on weekdays (monday to friday) in terms of customers travelling including passes and tickets in 2016?
56. What is the most used route on weekdays (monday to friday) in terms of customers travelling including passes and tickets in 2017?

Classify registered customers as follows:

Young customers: 18-35 y.o.

Adult customers: 36-64 y.o.

Senior customers: 64+ y.o.

1. What is the predominant age group traveling on each route from Tampa to Orlando and Orlando to Tampa (16 in Total).
2. If you are thinking of some promotion and you would like to focus on the age group with lowest revenue (sum of final price), what age group would you be focused on?
3. What is the most sold cabin type in tickets in the first quarter of 2016 within each age group?
4. Based on historical data, your projections for the next year are to have 20% more demand (proportional to all tickets and cabins), would you rather: a) Offer more travels? b) Increase the number of wagons per travel according to demand, c) Increase prices, or d) do nothing. Please answer this question from an industrial engineering point of view. Some options may not require writing a query but they may require some analysis.

About submission:

* Due date is Monday, April 10th at the beginning of the Lecture (file version in CANVAS). No Project will be accepted after that (corresponding grade to late submissions is 0/100)
* Be as accurate as possible with the information you search and the one provided.
* Reports submitted into canvas should consider a unique .SQL file, insert all your comments appropriately within that file.