****

**INDIVIDUAL ASSIGNMENT 1**

**Shuya Chen**

**CIS467: DATA MANAGEMENT, WAREHOUSING, AND VISUALIZATION**

**Professor: Mikhail Lysyakov**

**Spring A 2025**

**Homework 2. Due by Monday February 10 at 11:59 PM.**

**Please copy all your queries (sql code) from MySQL Workbench and paste them into this Word document for each question. Upload this Word file with your answers to the HW2 assignment folder on Blackboard. Check the file after uploading it.**

The create\_safety.sql script creates a database which contains the 6 tables described below. The data contains safety-related information for 179 company branch locations. To test your queries, create those tables using the create\_safety.sql script.

**If you use Chat GPT, please use the “Share” button (looks like ‘upward arrow’) in the right corner of ChatGPT chat, and ‘copy link’ and share the link to that chat in this Word document and briefly explain how you used it for your Homework (for each HW question if you used it). No points off will be taken for using ChatGPT (it is allowed to use it for Homework) but you are required to share the link to a chat if you used it.**

**More information on how to share a chat here:** <https://help.openai.com/en/articles/7925741-chatgpt-shared-links-faq>

|  |  |
| --- | --- |
| **Locations** | **Field Description** |
| Location\_ID (pk) | 4 digit location id number |
| Headcount | Number of employees at locations |
| Division | 2 character division code |
| State | Location state |
| City | Location city |
| Employee\_Safety\_Committee | Yes/No indicator of safety committee at location |

|  |  |
| --- | --- |
| **Audits** | **Field Description** |
| Location\_ID (pk and fk) | 4 digit location id number |
| Audit\_Date (pk) | Date of audit |
| Auditor | Name of head auditor |
| Audit\_Findings | Pass/Fail indicator of audit results |
| Corrective\_Action (Unique) | 5 digit CA number, if any (default = NULL) |

|  |  |
| --- | --- |
| **Correctiveactions** | **Field Description** |
| Corrective\_Action (pk and fk) | 5 digit CA number, if any (default = NULL) |
| Audit\_Date | Date of audit |
| Complete\_Date | Date CA was completed |

|  |  |
| --- | --- |
| **Trainings** | **Field Description** |
| Location\_ID (pk and fk) | 4 digit location id number |
| Training\_Date (pk) | Date of training |
| Training\_Location | Location of training (Onsite/Offsite) |

|  |  |  |
| --- | --- | --- |
| **lti (Lost Time Incidents)** | **Field Description** | |
| Location\_ID (pk and fk) | 4 digit location id number | |
| Incident\_ID (pk) | 4 digit incident id number | |
| Department | 2 digit department number | |
| EmployeeNumber | 5 digit employee number | |
| InjuryCode | Description of injury type | |
| DaysLost | # of days lost due to injury | |
|  |  | |
| **nmi (Near Miss Incidents)** | | **Field Description** | |
| Location\_ID (pk and fk) | | 4 digit location id number | |
| Incident\_ID (pk) | | 4 digit incident id number | |
| Dept | | 2 digits department number | |
| IncidentCode | | Description of incident type (root cause, etc.) | |
| EmployeeNumber | | 5 digit employee number | |

Diagram

Description automatically generated

1. **(10 points total)**. What is the total headcount for division "CF" (across all locations) and maximum headcount (i.e., headcount for one location with the largest number of people) for division "CF"? (hint: use locations table). The query needs to produce the total headcount and maximum value.

SELECT

SUM(Headcount) AS Total\_Headcount,

MAX(Headcount) AS Max\_Headcount

FROM locations

WHERE Division = 'CF';

表格

AI 生成的内容可能不正确。

2. **(10 points total).** Show the earliest audit date for location 2408. USE STR\_TO\_DATE function (as I show in Panopto video 3) to work with dates

SELECT MIN(Audit\_Date) AS Earliest\_Audit\_Date

FROM audits

WHERE Location\_ID = 2408;

表格

AI 生成的内容可能不正确。

3. **(10 points total).** How many audits did the auditor make whose name does not include “kar” as part of the name? The query needs to produce the total count of audits

SELECT COUNT(\*) AS Total\_Audits

FROM audits

WHERE Auditor NOT LIKE '%kar%';

表格

AI 生成的内容可能不正确。

4. **(10 points total)**. What is the total number of audits for each auditor whose name starts with “C”?

SELECT Auditor, COUNT(\*) AS Total\_Audits

FROM audits

WHERE Auditor LIKE 'C%'

GROUP BY Auditor;

表格

AI 生成的内容可能不正确。

5. **(10 points total)**. How many audits HAD corrective actions?

SELECT COUNT(\*) AS Audits\_With\_Corrective\_Actions

FROM audits

WHERE Corrective\_Action IS NOT NULL AND Corrective\_Action <> '';

图形用户界面, 应用程序

AI 生成的内容可能不正确。

6. **(10 points total)**. Use only NMI table. Count NMI incidents by location id and improve query to show only results for locations where the number of incidents is less than or equal to 53.

SELECT Location\_ID, COUNT(\*) AS Total\_Incidents

FROM nmi

GROUP BY Location\_ID

HAVING COUNT(\*) <= 53;

表格

AI 生成的内容可能不正确。

7. **(10 points total)**. Use only LTI table. Count LTI incidents by location id and improve query to show only results for locations where location\_id starts with the number 2 and ends with the number 5. You may use WHERE or HAVING for that.

SELECT Location\_ID, COUNT(\*) AS Total\_Incidents

FROM lti

WHERE Location\_ID LIKE '2%5'

GROUP BY Location\_ID;

表格

AI 生成的内容可能不正确。

8. **(10 points total)**. USE Subquery. Use only ‘locations’ table. Show headcount by location id and improve the query to show only results for locations where headcount is less than or equal to the average headcount for all locations. You may use WHERE or HAVING to show that report. Think about if you need to use GROUP BY or not.

SELECT Location\_ID, Headcount

FROM locations

WHERE Headcount <= (SELECT AVG(Headcount) FROM locations);

表格

AI 生成的内容可能不正确。

9. **(10 points total)**. Create a report (i.e., show a query) showing the number of LTI (Lost Time Incidents) by location as well as LTI Percent (LTI / Location Headcount). Only show results where LTI Percent is less than 25% (i.e., less than 0.25). Add a code to the query to sort the results from largest to smallest LTI Percent and show top 5 results only. To be able to confirm the answer by looking at the query results, you need to group by Location\_ID, Division, Headcount and add the same attributes in SELECT part.

SELECT

lti.Location\_ID,

loc.Division,

loc.Headcount,

COUNT(lti.Incident\_ID) AS Total\_LTI,

COUNT(lti.Incident\_ID) / loc.Headcount AS LTI\_Percent

FROM lti

JOIN locations loc ON lti.Location\_ID = loc.Location\_ID

GROUP BY lti.Location\_ID, loc.Division, loc.Headcount

HAVING LTI\_Percent < 0.25

ORDER BY LTI\_Percent DESC

LIMIT 5;

表格

AI 生成的内容可能不正确。

10. **(10 points total)**. Create exactly the same report (i.e., show a query with all the conditions including sorting by largest to smallest NMI Percent and showing top 5 results) as in question 9 but this time for use NMI and NMI% (NMI / Location Headcount)by location instead of LTI. But for that report, only show results for the division “CF” (you may use WHERE or HAVING to filter the results).

SELECT

nmi.Location\_ID,

loc.Division,

loc.Headcount,

COUNT(nmi.Incident\_ID) AS Total\_NMI,

COUNT(nmi.Incident\_ID) / loc.Headcount AS NMI\_Percent

FROM nmi

JOIN locations loc ON nmi.Location\_ID = loc.Location\_ID

WHERE loc.Division = 'CF'

GROUP BY nmi.Location\_ID, loc.Division, loc.Headcount

HAVING NMI\_Percent < 0.25

ORDER BY NMI\_Percent DESC

LIMIT 5;

表格

AI 生成的内容可能不正确。