**PHASE I**

1. **Who are your team members?**

Hatice Özgül Bilici - 170709023

İrem Uslu - 170709043

Reşit Ökkeş Çiçek - 170709029

**2) Explain your data:**

* **What’s the name/purpose of your project? Why did you pick this topic?**

Detailed Information about Rail Road crossing accident in US. We picked that data because it’s a interesting and detailed data for us.

* **How many files exist?**

It’s only one csv file.

* **How many rows and columns do they contain?**

It has 52 columns and 239488 rows.

* **How many string(non-numeric) columns do you have?**

It has 24 non-numeric columns.

* **Provide the links of the raw data of your project (e.g. kaggle address, ftp site address etc).**

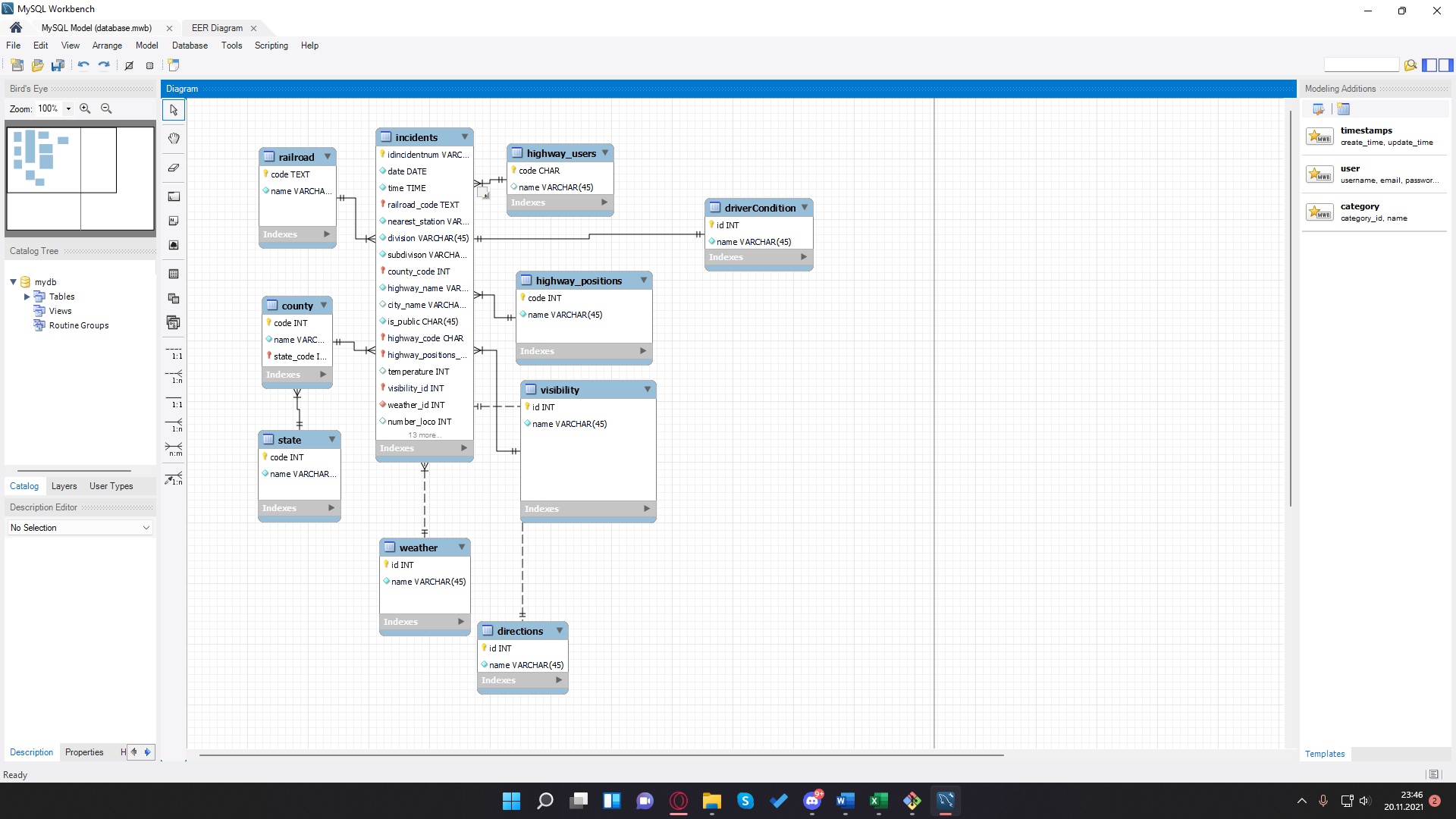
<https://www.kaggle.com/yogidsba/us-highway-railgrade-crossing-accident>

* **Provide the Bitbucket URLs of your ER diagram**

https://bitbucket.org/U170709029/dblab2021/src/master/project/database.mwb

**Questions:**

1. How many of the accidents in California were in rainy weather?
2. Where was the accident with the oldest driver and what is the gender of the driver?
3. Calculate the average damage cost in accidents by grouping them according to cities.
4. When was the last accident in the city with the most accidents?
5. Which city had the most accidents after 2007?
6. Show the total number of deaths in accidents by grouping them by years.
7. Show the number of casualties in accidents with the lowest and highest train speed.
8. What is the report code of the accident with the lowest air temperature?
9. In which visibility situation did the accident occur the most between 2015-2020?
10. What is the information of the first five accidents of the railway company with the most accidents?



**5)**

**1. How many of the accidents in California were in rainy weather?**

SELECT COUNT(\*) FROM incidents where division = ‘CALIFORNIA’ && weather\_id = 3;

1. **Calculate the average damage cost in accidents by grouping them according to cities.**

SELECT county\_state\_code,AVG(damageCost) from incidents GROUP BY county\_state\_code;

1. **What is the report code of the accident with the lowest air temperature?**

SELECT reportkey from incidents order by temperature ASC limit 1;

1. **Where was the accident with the oldest driver and what is the gender of the driver?**

SELECT county\_state\_code,user\_gender from incidents order by user\_age DESC limit 1;

1. **In which visibility situation did the accident occur the most between 2015-2020?**

SELECT visibility.name, sum(\*) as total FROM incidents join visibility WHERE YEAR(incidents.date) BETWEEN 2015 AND 2020 group by visibility.name order by total DESC limit 1 ;

**6)** We plan to import it ourselves by writing the insert statement in our CSV file. We thinking we can use java or python because we will need a for loop.

7) We are thinking about use PhpMyAdmin.

8)