Southern New Hampshire University

DAD 220 Module 5-2

Prepared on: 2 Apr., 2022
Prepared for: Prof. Aastha Agarwal
Prepared by: Alexander Ahmann

Contents

1 Introduction

The break down of trucks is of great inconvenience for our business. It is of up most priority that an investigation is to be undergone in order to mitigate the costs caused by mechanical break downs of trucks. In this report, I intend to use relational databases, the MySQL instance with Codio in particular, to work out possible root causes of the problem.

2 Procedure

Before proceeding, I need to define the problem and the objectives of this research job. I need to take a raw CSV file of truck fleet mechanical jobs and perform statistical analysis on it to work out descriptive statistics regarding truck malfunctions, and attempt to work out a causal agent—or perhaps many causal agents that is responsible for such a complex thing—that is causing the malfunctions. With this problem defined, I can now proceed to perform a data analysis with MySQL.

2.1 Feature specification

The data set that I was to analyse is called `FleetMaintenanceRecords.csv`. I used the Linux `head` command to get a "peep" into what the dataset looks like:

```
codio@mangogorilla-ohiochef:~/workspace$ head FleetMainten-
-anceRecords.csv

Vehicle ID,State,Repair ,Reason,YEAR ,Make,Body Type
0000015241383936,AL,Battery replacment,
  Dead battery,2016,RAM ,"1500 4WD Quad Cab 140.5"" SLT"
0000015241383936,AL,Dent Repair  Left Fender,
  Collision,2016,RAM ,"1500 4WD Quad Cab 140.5"" SLT"
0000060965535744,AL,Battery replacment,
  Dead battery,2016,RAM ,"1500 4WD Quad Cab 140.5"" SLT"
0000060965535744,AL,Tire repair,Flat- reparable,
  2016,RAM ,"1500 4WD Quad Cab 140.5"" SLT"
0000060965535744,AL,Tire replacement,Flat- hole not reparable,
  2016,RAM ,"1500 4WD Quad Cab 140.5"" SLT"
0000137172455424,AL,Battery replacment,Dead battery,2017,
  Chevrolet Silverado ,"1500 4WD Double Cab 143.5"" LT w/1LT"
0000137172455424,AL,Windshield replacement,Crack,2017,
  Chevrolet Silverado ,"1500 4WD Double Cab 143.5"" LT w/1LT"
0000243862142976,AL,Dent Repair Rear,Collision,2019,
  RAM ,"1500 Classic SLT 4X2 Crew Cab 6'4"" Box"
0000381034598400,AR,Windshield replacement,Crack,2016,
  RAM ,"1500 4WD Quad Cab 140.5"" SLT"
```

I then used a website (BE CSV, n.d.) that takes raw CSV entries and renders them into an easy-to-read table. Fig. 1 shows the results of the rendering. From Fig. 1, I can proceed to build a SQL query that will create a table based on the nature of the feature space (columns) and the nature of the entries (rows).

The first column, *Vehicle ID*, shows large numbers, and it seems like an `BIGINT` is the best candidate for it.

Next, the *State* column looks like the initials for a U.S. state. The data type `VARCHAR(2)` is probably the most appropriate data type for this, because the initials of a U.S. state are typically no more than 2 characters.

HTML Table Preview						
Vehicle ID	**State**	**Repair**	**Reason**	**YEAR**	**Make**	**Body Type**
15241383936	AL	Battery replacment	Dead battery	2016	RAM	1500 4WD Quad Cab 140.5" SLT
15241383936	AL	Dent Repair Left Fender	Collision	2016	RAM	1500 4WD Quad Cab 140.5" SLT
60965535744	AL	Battery replacment	Dead battery	2016	RAM	1500 4WD Quad Cab 140.5" SLT
60965535744	AL	Tire repair	Flat- reparable	2016	RAM	1500 4WD Quad Cab 140.5" SLT
60965535744	AL	Tire replacement	Flat- hole not reparable	2016	RAM	1500 4WD Quad Cab 140.5" SLT
137172455424	AL	Battery replacment	Dead battery	2017	Chevrolet Silverado	1500 4WD Double Cab 143.5" LT w/1LT
137172455424	AL	Windshield replacement	Crack	2017	Chevrolet Silverado	1500 4WD Double Cab 143.5" LT w/1LT
243862142976	AL	Dent Repair Rear	Collision	2019	RAM	1500 Classic SLT 4X2 Crew Cab 6'4" Box
381034598400	AR	Windshield replacement	Crack	2016	RAM	1500 4WD Quad Cab 140.5" SLT

Figure 1: **A preview of `FleetMaintenanceRecords.csv` file.**

The *Repair* column looks like the kind of repair that will be given to a vehicle. The data type `VARCHAR(50)` is the best kind of data type because, from the preliminary analysis based on `head` output, it seems that these Repair reasons do not exceed 50 characters.

Like the *Repair* column, the *Reason* column discusses the details of the repair. This column briefly discusses the category of problem that the vehicle is suffering from, as opposed to its solution. This column will be given a data type of `VARCHAR(50)` as categories of vehicle problems do not seem to exceed 50 characters.

The *YEAR* column seems to be a year that the vehicle was manufactured. Since years in the United States calendar are typically in the thousands range, a datatype of `SMALLINT` should be enough for this column.[1]

Finally, the *Make* and *Body Type* columns will both have a type of `VARCHAR(50)`, as they both discuss the kind of vehicle and its body type. They both do not seem

[1]The maximum number that SMALLINT can handle is 32,767— which is plenty of "years" before automobile will be obsoleted.

to exceed 50 characters in their description.

2.2 Importing the data set

With an idea of how this table is supposed to look like, I constructed the following query to create the table that is to be populated with the raw CSV data:

```
CREATE TABLE FleetMaintenanceRecords (
    vehicle_id BIGINT(30),
    state VARCHAR(2),
    repair VARCHAR(50),
    reason VARCHAR(50),
    year SMALLINT,
    make VARCHAR(50),
    body_type VARCHAR(50)
);
```

After executing the query, I can proceed to import the CSV file into table with the following command:

```
LOAD DATA INFILE
      "/home/codio/workspace/FleetMaintenanceRecords.csv"
    INTO TABLE FleetMaintenanceRecords
    FIELDS TERMINATED BY ","
    LINES TERMINATED BY "\n"
    IGNORE 1 ROWS;
```

Fig. 2 demonstrates that I was able to successfully import the CSV file into my database—specifically the `FleetMaintenanceRecords` table.

The `LOAD DATA INFILE "/home/codio/workspace/FleetMaintenance Records.csv"` bit will instruct MySQL to get the raw data from the .csv file on the system. The `INTO TABLE FleetMaintenanceRecords` bit specifies the table where the data will be loaded in to. The `FIELDS TERMINATED BY ","` bit specifies that CSV uses a comma to separate the column data. The `LINES TERMINATED BY "\n"` bit specifies that the CSV uses line breaks to separate the row data. Finally, the `IGNORE 1 ROWS;` bit tells MySQL to skip the first row.[2]

[2]Because the first row contains labels for the features, and I have already specified them when creating the table.

2.3 Exploratory data analysis

I suspect that the kind of parts being replaced would contain significant useful information regarding what can be learnt in regards to the engineering problem posed. I will analyse specifically these kinds of parts and their respective features.

First, I shall segregate specifically the cases where parts are being replaced into another table with the following query:

```
CREATE TABLE FleetMaintenanceRecords_parts
  AS
  SELECT * FROM FleetMaintenanceRecords
  WHERE repair
  LIKE "%replacement%";
```

The following query created a subset of the `FleetMaintenanceRecords` called `FleetMaintenanceRecords_parts` which contains entries for specifically repairs and the various kinds of repairs being done. Fig. 3 demonstrates that this task was completed successfully.

I decided to proceed counting the different kinds of repairs being done. First, I must work out the number of different kind of repairs, which can be accomplished with the following query:

```
SELECT DISTINCT repair
    FROM FleetMaintenanceRecords_parts;
```

Running the query returns the following four (4) different categories of repair:

- Tire replacement
- Windshield replacement
- Fender replacement
- Brake line replacement

Next, I can go about working out both the raw count and normalised frequency of each kind of repair. I will use the following query structure to work all of them out:

```
SELECT COUNT(*) AS "Count",
    COUNT(*) / 235 AS "% Repair"
  FROM FleetMaintenanceRecords_parts
  WHERE repair = "<repair kind>";
```

The following statistics were returned for each separate query:

Repair kind	**Count**	**% Repair**
Tire replacement	66	0.2809
Windshield replacement	63	0.2681
Fender replacement	54	0.2298
Brake line replacement	52	0.2213

The counts and normalised frequency of repair conform to the rules of probability, as they all add up to 235[3] and 100%,[4] respectively:

$$\sum_{\text{Count}} = 66 + 63 + 54 + 52 = 235$$

$$\sum_{\text{\% Repair}} = 0.2809 + 0.2681 + 0.2298 + 0.2213 = 1 = 100\%$$

From these descriptive statistics, one can infer that the tire is the most likely to be subjected to replacement.

It could be of benefit to those concerned with the truck fleet to know which regions of the United States are in most need of part replacements—perhaps so that they can better allocate resources to said regions or investigate any causal agents that may correlate to region.[5] See Appendix A.1 for a list of U.S. states, and their respective initials, by region.

One may hypothesise that certain regions will have more reasons to get replacements. This can be verified with the following query for each respective region:

```
SELECT COUNT(DISTINCT reason)
FROM FleetMaintenanceRecords_parts
WHERE state = <initial 1>
    OR state = <initial 2>
  ...
    OR state = <initial n>; % for each "n"
                              representing a
                              state in the
                              region
```

[3]The total number of cases.

[4]Regarding the sum of the normalised frequency, the actual output was 1.0001. However, this is likely due to rounding error.

[5]s.a. extreme weather's effect on mechanical parts

Running this query gives the following results for each region:

Region	**No. of unique reasons**
Southwest	2
Southeast	4
Northeast	4
Midwest	4
West	3

In this analysis, it seems that the median and mode of reasons are both four (4). The Southwest and West seem to have less unique reasons regarding repair. The mechanic team may want to look into these states to work out what makes for less kinds of causal mechanism for failure. Furthermore, more resources for repair should be diverted to the Southeast, Northeast and Midwest, as they have more unique reasons for failure.

Regarding which parts are more likely to be replace because of rust and erosion, the following query shall be used to work this out:

```
SELECT COUNT(*)
   FROM FleetMaintenanceRecords_parts
   WHERE repair = "<repair kind>" AND
   (reason LIKE "%Rust%" OR reason LIKE "%Corrosion%");
```

As in previous parts of this report, the “repair kind” will be replaced with ”Tire replacement,” ”Windshield replacement,” ”Fender replacement” and ”Brake line replacement.”

Running the query reports the following descriptive statistics:

Repair kind	**Count of rust or corrosion as causal agent**
Tire replacement	0
Windshield replacement	0
Fender replacement	54
Brake line replacement	52

Further analysis will need to be undergone to determine the underlying causal agents regarding rust or corrosion on mechanical parts, but they do seem to affect the fender and brake line components.

3 Summary

Unfortunately, the exact causal mechanism that is responsible for the mechanical failure of truck cannot easily be determined by this rudimentary analysis. Nonetheless, some interesting conclusions can be drawn from this exploratory data analysis:

- Tire replacements is the most frequent kind of replacement (count of 66), and brake line replacement is the least frequent kind of replacement (count of 52).
- The Southwest and West regions of United States have less unique reasons for mechanical failure. Perhaps these states can serve as a "role model" of some kind that other truck manufacturers or drivers can look at.
- Corrosion and rust can act as a causal agent for some of truck mechanical break downs.

A Appendices

A.1 U.S. States and their initials, by region

DAD-220 Region Definitions (n.d.) list the following regions and individual states that are grouped into said regions:

1. U.S. States in the Southwest:

State	**Initials**
Arizona	AZ
New Mexico	NM
Texas	TX
Oklahoma	OK

2. U.S. States in the Southeast:

State	Initials
Arkansas	AR
Louisiana	LA
Mississippi	MS
Alabama	AL
Georgia	GA
Florida	FL
Kentucky	KY
Tennessee	TN
South Carolina	SC
North Carolina	NC
Virginia	VA
West Virginia	WV
Delaware	DE
Maryland	MD

3. U.S. States in the Northeast:

State	Initials
Pennsylvania	PA
New Jersey	NJ
New York	NY
Connecticut	CT
Rhode Island	RI
Massachusetts	MA
Vermont	VT
New Hampshire	NH
Maine	ME

4. U.S. States in the Midwest:

State	Initials
North Dakota	ND
South Dakota	SD
Kansas	KS
Nebraska	NE
Minnesota	MN
Wisconsin	WI
Iowa	IA
Missouri	MO
Michigan	MI
Indiana	IN
Illinois	IL
Ohio	OH

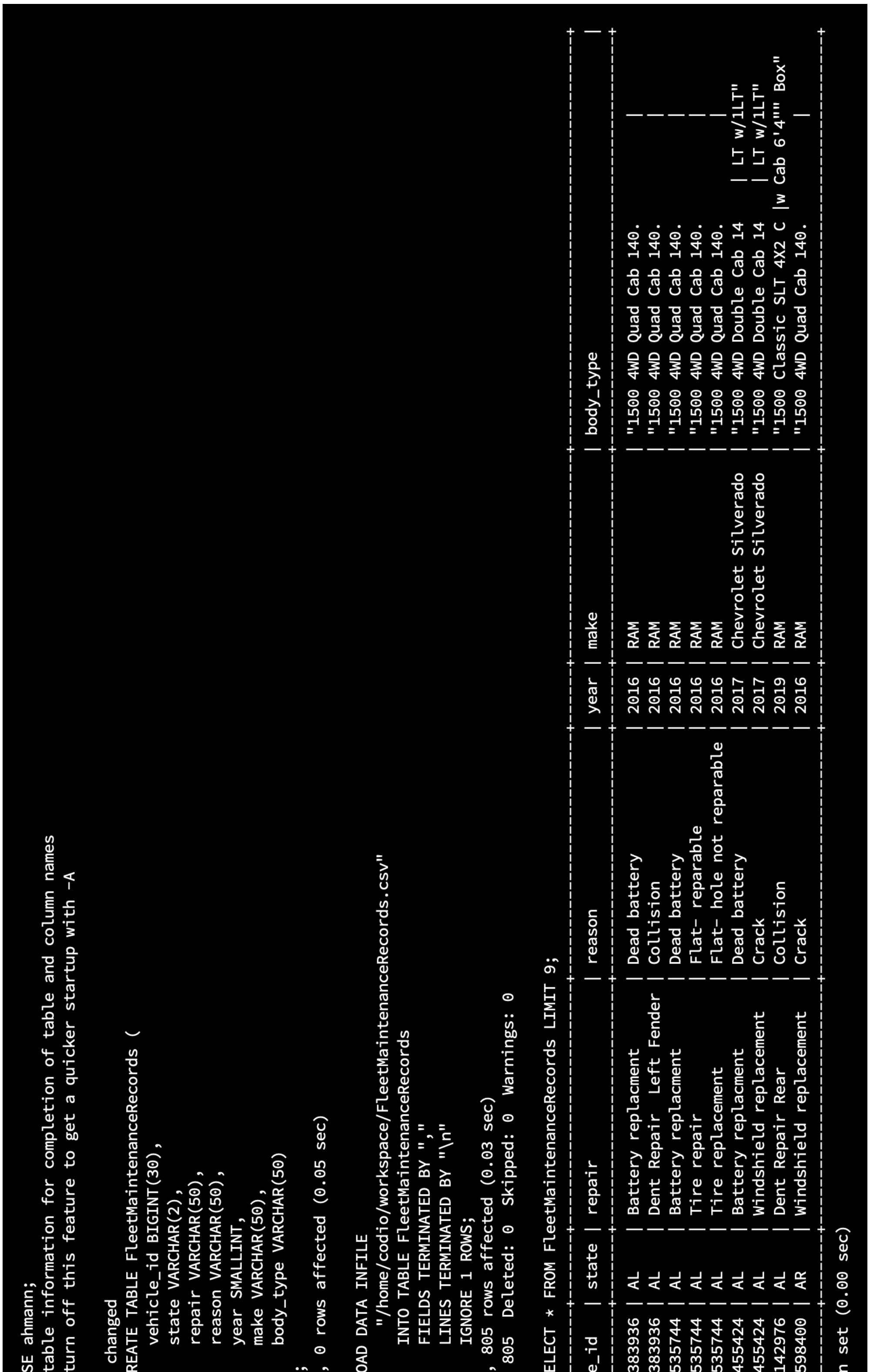
5. U.S. States in the West:

State	Initials
Washington	WA
Idaho	ID
Montana	MT
Oregon	OR
Wyoming	WY
Colorado	CO
Utah	UT
Nevada	NV
California	CA

References

BE CSV (n.d.). *Online CSV to Table converter*. Retrieved on Apr. 2, 2022 from: https://www.becsv.com/csv-table.php

DAD-220 Region Definitions (n.d.).



```
mysql> USE ahmann;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> CREATE TABLE FleetMaintenanceRecords (
    ->     vehicle_id BIGINT(30),
    ->     state VARCHAR(2),
    ->     repair VARCHAR(50),
    ->     reason VARCHAR(50),
    ->     year SMALLINT,
    ->     make VARCHAR(50),
    ->     body_type VARCHAR(50)
    -> );
Query OK, 0 rows affected (0.05 sec)

mysql> LOAD DATA INFILE
    ->       "/home/codio/workspace/FleetMaintenanceRecords.csv"
    ->     INTO TABLE FleetMaintenanceRecords
    ->     FIELDS TERMINATED BY ","
    ->     LINES TERMINATED BY "\n"
    ->     IGNORE 1 ROWS;
Query OK, 805 rows affected (0.03 sec)
Records: 805  Deleted: 0  Skipped: 0  Warnings: 0

mysql> SELECT * FROM FleetMaintenanceRecords LIMIT 9;
+--------------+-------+-------------------------+---------------------------+------+---------------------+-----------------------------------------+
| vehicle_id   | state | repair                  | reason                    | year | make                | body_type                               |
+--------------+-------+-------------------------+---------------------------+------+---------------------+-----------------------------------------+
|  15241383936 | AL    | Battery replacment      | Dead battery              | 2016 | RAM                 | "1500 4WD Quad Cab 140.          |
|  15241383936 | AL    | Dent Repair  Left Fender | Collision                | 2016 | RAM                 | "1500 4WD Quad Cab 140.          |
|  60965535744 | AL    | Battery replacment      | Dead battery              | 2016 | RAM                 | "1500 4WD Quad Cab 140.          |
|  60965535744 | AL    | Tire repair             | Flat- reparable           | 2016 | RAM                 | "1500 4WD Quad Cab 140.          |
|  60965535744 | AL    | Tire replacement        | Flat- hole not reparable  | 2016 | RAM                 | "1500 4WD Quad Cab 140.          |
| 137172455424 | AL    | Battery replacment      | Dead battery              | 2017 | Chevrolet Silverado | "1500 4WD Double Cab 14    | LT w/1LT"
| 137172455424 | AL    | Windshield replacement  | Crack                     | 2017 | Chevrolet Silverado | "1500 4WD Double Cab 14    | LT w/1LT"
| 243862142976 | AL    | Dent Repair Rear        | Collision                 | 2019 | RAM                 | "1500 Classic SLT 4X2 C |w Cab 6'4"" Box"
| 381034598400 | AR    | Windshield replacement  | Crack                     | 2016 | RAM                 | "1500 4WD Quad Cab 140.          |
+--------------+-------+-------------------------+---------------------------+------+---------------------+-----------------------------------------+
9 rows in set (0.00 sec)
```

Figure 2: **Screenshot of me importing the** `FleetMaintenanceRecords.csv` **file into its respective table.**

```
mysql> CREATE TABLE FleetMaintenanceRecords_parts
    ->   AS
    ->   SELECT * FROM FleetMaintenanceRecords
    ->   WHERE repair
    ->   LIKE "%replacement%";
Query OK, 235 rows affected (0.18 sec)
Records: 235  Duplicates: 0  Warnings: 0

mysql> SELECT COUNT(*) FROM FleetMaintenanceRecords_parts;
+----------+
| COUNT(*) |
+----------+
|      235 |
+----------+
1 row in set (0.00 sec)

mysql> DESCRIBE FleetMaintenanceRecords_parts;
+------------+-------------+------+-----+---------+-------+
| Field      | Type        | Null | Key | Default | Extra |
+------------+-------------+------+-----+---------+-------+
| vehicle_id | bigint(30)  | YES  |     | NULL    |       |
| state      | varchar(2)  | YES  |     | NULL    |       |
| repair     | varchar(50) | YES  |     | NULL    |       |
| reason     | varchar(50) | YES  |     | NULL    |       |
| year       | smallint(6) | YES  |     | NULL    |       |
| make       | varchar(50) | YES  |     | NULL    |       |
| body_type  | varchar(50) | YES  |     | NULL    |       |
+------------+-------------+------+-----+---------+-------+
7 rows in set (0.00 sec)
```

Figure 3: **Creating a subset of the** `FleetMaintenanceRecords` **discussing specifically part replacement.**