OLEO DATABASE ORGANIZATIONAL REBUILD (ODOR)

Eric M. Fraser

Western Governors University

Table of Contents

Summary

The Oleo Database Organizational Rebuild (ODOR) consisted of 5 main objectives 10 deliverables, and the overarching goal of updating and streamlining Oleo's data collection and storage so that its analysis can be undertaken with ease and clarity. Objective one, to create the new tabular infrastructure took several hours to complete, which was about as long as expected. Wireframing and planning out all the separate tables and how they would relate to each other was a challenge and took several hours of focused thought, but eventually we settled on 10 separate tables that would offer enough information to Oleo without being needlessly complex. (APPENDIX A). Flat file tables to hold Oleo's data included the following tables: Transactions, Locations, Employees, Regions, Vehicles, Oil, Filters, Storage, Oil Inventory, and Filter Inventory. The chief challenge of this section was how to reconcile the fact that Oleo operates from 2 separate locations. The solution to this challenge was to incorporate the Regions table, which would essentially 'umbrella' over all the other tables and allow data to be separated by region if need be without being too cumbersome. With this diagrammed out, we now had 10 tables holding data that, if structured and analyzed properly, would help Oleo manage their business more efficiently and more profitably.

Once we had determined the optimal data to store and collect, the next challenge (deliverable two) was to link the tables through primary and foreign keys. The main benefit of a relational database is that all the tables relate to each other so there is a high level of data density. The key realization was that the Transactions table and the Vehicles table were central to the whole database and seemed to tie everything else together There were essentially 2 blocks of tables connected by these two tables -- the Supply block and the Client block. The Locations and

Employees tables fed into Transactions, Transactions was linked to the Vehicles table, and the Storage, Oil, Filter, and Inventory information all fed into the vehicles table.

Once deliverable one and two were complete, the final deliverable of building the tables in Excel as flat files was very easy. An Excel workbook was created with 11 sheets, 1 for the database relationship model (APPENDIX B) and the rest to hold the data from the 10 planned tables. Table names and column titles were listed out without data for now. You can see an example of this structure in APPENDIX C.

Objective two was to ensure preservation of legacy data so it could be used immediately for analysis. We found the bulk of legacy data was contained within the Transactions table and the vehicles table, the others would pretty much have to be filled in from scratch. APPENDIX D shows what v1.0 of the transaction data looked like pre-cleaning, and APPENDIX E shows what the data looks like in the new database after it had been cleaned and moved to fit the new structure. The deliverables for Objective two, which included moving the data and filling in gaps was not complex, but it was time consuming. As mentioned in the pre-report, technicians were not listed out for the bulk of the transaction data. Hundreds of invoices were pored through to discern which technician completed which services, and eventually the new Transaction table was reconciled and not missing any critical information.

Objective three, was to import the newly cleaned legacy data and infrastructure into MySQL for storage (APPENDIX F). The deliverables for this state were quite simple overall. Key takeaways from this objective were to ensure that the proper plugins from MySQL had to be installed. For example, we determined that using MySQL workbench would be much more effective than importing everything from the command line, so that plugin was installed alongside the core MySQL application. The MySQL for Excel plugin was used to extract the

newly cleaned and structured tables into MySQL, and the primary and foreign keys were all reconciled using a combination of SQL commands and the Workbench GUI. To ensure that the data was processed correctly, several queries were run to analyze the data, one of which is shown in APPENDIX G.

Objective four was put in place to incorporate a solid process for adding new data to the database to ensure the structures in place stayed tight and updated properly. It was determined that a combination of on-site data and cloud stored data would be most appropriate, both for ease of use and a degree of fault tolerance. New data will be first added in the new Excel workbook, and the MySQL plugin will be used to import the new data to MySQL with the click of a button. Future changes to the table structure itself must be done using DDL statements in SQL itself.

Objective five was the most fun and most interesting part of the process. Using the newly organized data, several Tableau visualizations were created to turn raw data into useful information. Visualization 1 (APPENDIX H) is entitled 'Yearly Location Production by Contractor', Visualization 2 (APPENDIX I) is entitled 'Average Contractor Rate by Year', Visualization 3 (APPENDIX J) is entitled 'Vehicle Frequency by Service Area Year-to-Year' Visualization 4 (APPENDIX K) is entitled 'Revenue by Day of Week Quarter to Quarter'. Explanations of these visualizations will be detailed in another section of this report.

Review of Other Work

1. Dev.MySQL (2019) *Getting Started with MySQL* Retrieved from: https://dev.mysql.com/doc/mysql-getting-started/en/

Since this was the first time I had used MySQL specifically as a Database Platform, I consulted the documentation from MySQL itself heavily, especially during Objective three. This article in particular was useful for getting started. Installation is different for every platform and I used this resource to understand the installation process. Additionally, I ran into a few problems after installation, particularly database instance connection issues, that I used the MySQL documentation to research and successfully resolve.

2. Jena, Shubhangi – Towards Data Science (2019) *Tableau – A Beginners Guide* Retrieved from https://towardsdatascience.com/tableau-c9d6962991ca

Although I had used Tableau before, this article from Towards Data Science was extremely helpful in going over some of the basics about the software that I had forgotten since last time. Installing the correct software version and making sure it is up to date is an oft overlooked part of projects, and this article led me to the correct solution on the first try so I could get started immediately without having to troubleshoot or reinstall.

3. LucidChart (2019) *Database Structure and Design Tutorial* Retrieved from: https://www.lucidchart.com/pages/database-diagram/database-design

As mentioned before, the planning portion of database creation and migration is probably the most critical stage in the project. The approach I took was to spend extra time on the planning portion, hoping it would save both time and headache in the later portions of the project. This article from LucidChart was extremely helpful in this critical planning stage. LucidChart is an

excellent tool for creating digital charts and organizational diagrams, and even though I used pen and paper to brainstorm and sketch my database relations and tables, the article went into detail from start to finish, covering topics such as requirements analysis, organizing data into tables, specifying primary keys and analyzing relationships, and normalizing to standardize the tables. (LucidChart 2019). For a free article, I was surprised and pleased with both the depth and breadth of information in the article, and it proved to be a very valuable resource.

In this section, provide an expanded review of the Review of Other Work section in task 2, including three additional third-party artifacts on the topic that supported the development of the project, and explain how the artifacts supported the implementation.

Changes to the Project Environment

As mentioned in the initial project report, there were several planned changes to the project environment that were critical to the completion of the project. The primary goal of the project itself involved a significant project environment change – migrating Oleo's data from Excel to MySQL. Before the project was started, Oleo's data was held primarily in Excel workbooks and Google Drive. As the project began it made the most sense for the company to keep record in both Excel and in MySQL for the purpose of fault tolerance and data redundancy.

Upon completion of the project the project environment consisted of one data center in a Microsoft Excel workbook, and one in a MySQL Database. The Excel workbook contained within it 11 sheets, 1 that listed the structure of all the tables including table names, primary keys, foreign keys, and column names, and 10 that included the tabular data for each of the 10 tables in flat file form. The MySQL Database is comprised of all the information from the Excel Workbook, except linked together with SQL commands and unlike the Excel data, is able to be queried and analyzed using SQL commands. The system in place to add new data will consist of

a two part process promoting data consistency and availability. Part 1 is to load the new data into the Excel Workbook flat files. Part 2 is to use the Excel for MySQL plugin to import the new data into the correct table in the MySQL relational database.

Methodology

Several standard methodologies could have been used for this project, and the ODOR was mapped to the PDCA model. PDCA stands for Plan – Do – Check – Adjust and is an iterative model for continuous development of products and processes. The planning phase consists primarily of establishing the objectives and processes that would constitute project success. In the Do phase, the plans from the planning phase are put into action and systems are put in place to gather data on the effectiveness of the new changes. The check phase follows the Do phase. Data gathered from the do phase is compared to the expected outcomes to determine how effective this cycle of PDCA has been. In the final phase (Adjust), the process is improved and the baseline for the next iteration of PDCA is laid out. The PDCA is a simple and effective model used to improve processes and was a great fit to structure the implementation of the ODOR.

Several iterations of PDCA were used heavily in the creation of the database model for import into MySQL. First, version one of the database was planned out. Planned tables were mapped out, key connections were specified, and table columns were listed. I had a good idea of what I wanted in the database but I knew that several iterations of this process would be necessary to ensure that everything meshed together and didn't break when the table structures were loaded into MySQL. After planning, part of the 'DO' phase consisted of some test data being added and some basic joins were performed on paper to see how the database would work in practice. As expected, the 'CHECK' phase began with version one of the database breaking

immediately. After adjustment and retooling, we were back to stage one, and planning for version two began. As with any project, the 4 stages of PDCA became fluid in some places – but that is the strength of PDCA, that you aren't locked into a model every single time. The flexibility and iterability of PDCA helped the project along, leading to completion ahead of schedule.

Project Goals and Objectives

As stated in the Capstone Proposal, the singular goal of the ODOR is to update and streamline Oleo's data collection and storage so that its analysis can be undertaken with ease and clarity. In service of this goal there were five objectives that were to be completed one by one, and for these five objectives there were a total of ten deliverables that indicated completion of this project.

Objective number one was to build out the tabular infrastructure of the new database. This objective was completed by building out the three deliverables in service of this objective. The first deliverable of this objective was to build a model of all the tables in the database and how they will relate to each other in the context of a relational database model. The deliverable was a spreadsheet listing all the tables and their respective columns normalized to at least 3rd normal form. Proof of this deliverable can be found in APPENDIX A and B. Hand in hand with this deliverable is the primary and foreign key identification deliverable. Primary and foreign key identification aids in describing and formatting how each table will link to each other, and can be found in APPENDIX B. The final deliverable for this objective was to build out all the tables in the relational database model in Microsoft Excel. These tables would later be imported into MySQL using a plugin, but for the purposes of cleaning and easily visualizing the relationships,

using Excel first was be the best path forward. An example from one table of this deliverable is located below in APPENDIX E. Objective one was successfully completed.

Objective number two was to find a way to preserve as much of the useful legacy data as possible for use immediately for data analysis. Oleo had already collected a ton of useful data. Even though it had some gaps and was inconsistently structured, extremely valuable information was pulled, cleaned, and placed in a format conducive to analysis. The first deliverable in service of this objective was to copy the old data into the new table structure. Once the already existing data was placed in the new format, deliverable number two was to fill in as many gaps as possible to add as much utility as possible. One example of this that was completed as a result of the project was listing the technician who performed the oil change as part of the transactions table. Part of this deliverable was to go through each table and fill out as much as possible to make the data comprehensive. Objective two unfortunately was only partially completed. Although significant progress was made and many gaps were filled, there were several gaps in the data that could not be completed primarily because their completion would be too time consuming to be feasible within the scope of this project. An example of this was data from the vehicle data, specifically the oil type, amount, and filter used for every vehicle in the database. Fortunately, this data will be easy to add once the time to incorporate it is set aside.

Objective number three was to import the newly formatted legacy data and infrastructure into MySQL. As discussed above, MySQL is an open source SQL database and a major result of this project was that Oleo's data will be stored in this database. The first deliverable of objective three was to download, configure, and optimize the MySQL program. This deliverable included downloading proper plugins and updates, especially the Microsoft Excel plugin. Once a database is created and ready for import, deliverable two entailed the physical moving of the tables from

objectives one and two to the new database in MySQL using the MySQL for Excel plugin as discussed above. The final deliverable for this portion of the project was to use SQL to link the tables together and run a few queries to determine if the data is accurate, usable, and loaded properly. Objective three was completed after several iterations.

Objective number four was to outline and document a process for adding new data to the data infrastructure created in MySQL from objective three. It was discovered and decided that the best path forward was to have new data typed into the Excel Workbook, then imported into MySQL using the Excel plugin. This was decided because it limits the need for SQL literate developers to import that data – it can instead be done with the click of a button.

The fifth and final objective was to somehow determine the efficacy of the changes made to Oleo's database infrastructure. Because the main goal of the project was to ensure that Oleo's data can be queried and used for analysis, any measure of the completion of the project required some analysis to be performed. If in the new format analysis could be performed and business operations improved, the project would be considered a success. The deliverable for this final objective was utilizing data visualization software to offer Oleo tangible and actionable recommendations for their business to improve some aspect of it. Tableau Public was used for this. A minimum of three data visualizations that offer legitimate and accurate insight into the nature of Oleo's business were created and proved the ODOR project was a success.

Project Timeline

Here is the projected timeline copied from the Capstone proposal:

Milestone or deliverable	Duration (hours or days)	Projected start date	Anticipated end date
1.a + 1.b	4 hours	17 February 2020	17 February 2020
1.c	1 hour	17 February 2020	17 February 2020
2.a	2 hours	18 February 2020	18 February 2020
2.b	2 days	18 February 2020	20 February 2020
3.a	1 hour	20 February 2020	20 February 2020
3.b + 3.c	4 hours	21 February 2020	21 February 2020
4.a	1 day	24 February 2020	24 February 2020
5.a	1 day	25 February 2020	25 February 2020

The actual project timeline did differ from the projected timeline. Truthfully there was time on Sunday February 16th to sink 6-8 hours into the project, so the planning for objective 1 was completed a day early, though the anticipated duration time of 5 hours was a bit optimistic. More time was budgeted for objective two than any other objective. Fortunately, it took less time

than expected to complete objective two, with the 2nd deliverable for part 2 completed in a full day of work as opposed to two days. The reason for this had something to do with the fact that Oleo did not have enough data available to complete every single one of the 10 tables in the database. Because this objective could be only partially completed, it was finished faster than anticipated. Objectives 3 and 4 were completed on February 18th in one day – setting up the environment of MySQL did have a couple setbacks but overall the installation and setup of MySQL was easier and less time-consuming than anticipated. Additionally, objective four and the research it required took only about two hours to complete instead of the full day expected. Part 5 did indeed take a full day to wrap everything up so the time estimate for completion was accurate there. Overall the time estimates for the ODOR were more pessimistic than reality, which I would argue is a good thing. I would rather set expectations at a low level and say things will take a long time and deliver them early than the reverse. The fact that work was able to be completed on Sunday helped the overall timeline as well.

Unanticipated Requirements

One unanticipated requirement that was an obstacle to overcome was in the setup of the MySQL database itself. The version that I installed initially was a SQL database for use exclusively with Python, so I went through the whole installation process and ended up having to download an alternate package. With this, when I restarted my computer the database instance disconnected and would not accept my port number and password. It was so frustrating I almost switched the project environment over to my 2nd choice, PostgreSQL. Some extra troubleshooting had to be done in order to solve this issue, which ended up being a really simple fix that required me to go into the computer's control panel and restart the database instance

manually. A simple google search combined with perusing the MySQL documentation provided a quick fix to the problem.

Additionally, I did not expect the planning of the database to be as difficult and time consuming as it was. The picture in APPENDIX A is the final version that works very well, but there were 4 or 5 versions before this completed version that were messy and complex. Apart from these two obstacles the ODOR was completed relatively smoothly and ahead of schedule.

Conclusions

As discussed in the Capstone Proposal, final acceptance and completion of the ODOR project would consist of several visualizations using the cleaned data to determine the efficacy of the changes made. On that front, the ODOR project was a resounding success. Though deliverable 2 from Objective 2 was left partially incomplete because of information restrictions, overall the project was able to fulfil its intended purpose. Oleo now has a centralized location to store its data, fault tolerance, a system for importing new data into the database, and has already seen clear benefit from this using Tableau Visualizations detailed in the Deliverables section below. Much more can be done as Oleo's data capacity grows, and I'm sure there are many more useful insights that can be gained from the data as the MySQL database is completely fleshed out over time. Overall, the ODOR Project should be considered a huge success.

Project Deliverables

The final deliverables for this project were the data visualizations using Tableau. This section will cover details regarding these final deliverables.

APPENDIX H: Yearly Location Production by Contractor

This visualization gives the user a snapshot of not only which contractors are the most active year to year, but which stores each contractor is most active at. So much information is packed into this visualization. A graph like this will help Oleo better understand which stores are feeding the company the most cars for services. You can see from the graph that in 2019 the Hamilton and Reynoldsburg locations were the top two locations for services by a pretty substantial margin. The advice I would give to Oleo is to continue to maximize those relationships and keep the decision makers at those stores happy and feeding you even more business.

APPENDIX I: Average Contractor Rate by Year

This visualization shows the average invoice amount from each contractor year by year. From this documentation we can see a couple things. Number one, the cars done in the San Diego service area are on average more expensive than the ones in the Columbus service area. Additionally, within service areas, it appears that Anthony is prone to charging a bit more for services than Ron is. This could be used to Oleo's advantage by sending Anthony to the higher volume stores that will result in more revenue over time.

APPENDIX J: Vehicle Frequency by Service Area Year to Year

This visualization is extremely useful for a couple reasons. Vehicle information for this company is extremely important, and this data shows which types of vehicles are in rotation year by year. For example, you can tell from the chart that the Chevy Equinox vehicle type was put into service in 2018 for the first time and is now the most popular Chevy model as of 2019. What this means for Oleo is that they would be advised, based on the chart, to stock more supplies that will be used for the Chevy Equinox, and

scale back on supply ordering for vehicles that are being phased out, such as the Chevy Cruze which has seen a sharp decline from 2018 to 2019 and no services in 2020. Another wrinkle to this visualization is that it is colored by service area, so not only can you see which cars are being phased in and which cars are being phased out, you can tell at a glance which service areas prioritize different make and model vehicles for their rotations.

APPENDIX K: Revenue by Day of Week Quarter to Quarter

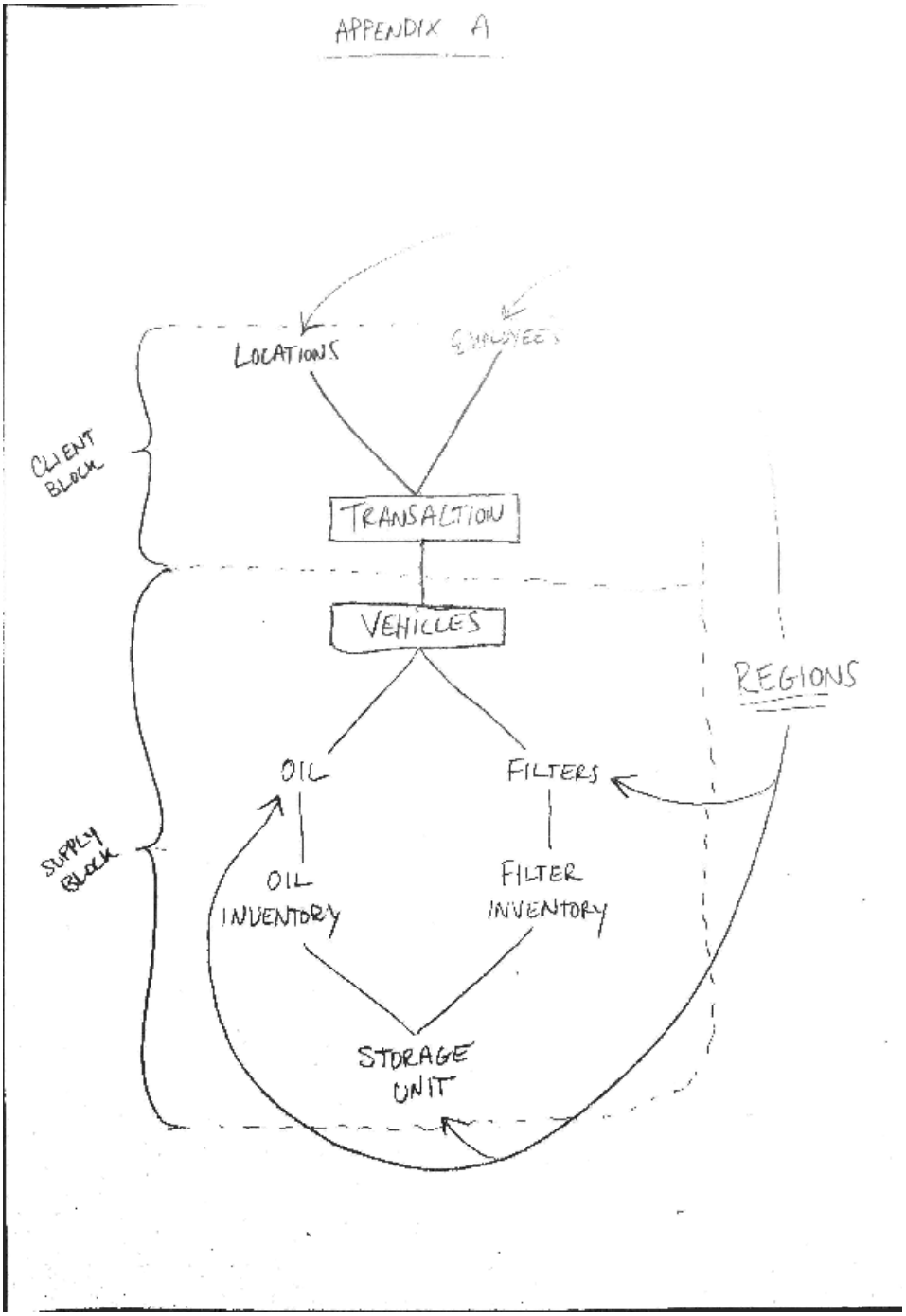
This visualization shows Quarterly revenue split up by day of the week, color scaled for emphasis. This particular visualization is extremely useful because it highlights not only which quarters tend to have the highest income, but which day of the week is optimal for staffing extra contractors. Its clear from the chart that weekends are tough for business, but Monday, Tuesday, and Wednesday overall are profitable days for business. Besides that, it appears that Q3 just about every year tends to be the best quarter for revenue. Based on this information I would advise that Oleo ensures extra contractors are working early in the week to handle the increased demand, particularly in July, August, and September when in the past, revenue has been at its highest.

References

1. LucidChart (2019) *Database Structure and Design Tutorial* Retrieved from: https://www.lucidchart.com/pages/database-diagram/database-design
2. Jena, Shubhangi – Towards Data Science (2019) *Tableau – A Beginners Guide* Retrieved from https://towardsdatascience.com/tableau-c9d6962991ca
3. Dev.MySQL (2019) *Getting Started with MySQL* Retrieved from: https://dev.mysql.com/doc/mysql-getting-started/en/

APPENDIX A

Sketched Database Relationship Model



APPENDIX B

Database Model

Transactions	
PK	Transaction_ID
FK	Location_Name
FK	Make
	Model
	Date
FK	Employee_ID
	Invoice_Amount
	Client_Fleet_Num
	VIN

Locations	
PK	Location_Name
FK	Service_Area
	Street_Address
	City
	State
	Zip_Code
	Contact_Name
	Phone_Number
	Company

Employees	
PK	Employee_ID
	First_Name
	Last_Name
FK	Service_Area
	Street_Address
	City
	State
	Zip_Code
	Rate
	Daily_Gas_Rate

Regions	
PK	Service_Area
	City
	State
	Corporate_Contact
	Phone_Number
	Email

Vehicles	
PK	Make
	Model
FK	Oil_ID
	Oil_Quantity
FK	Filter_ID

Oil	
PK	Oil_ID
	Oil_Weight
FK	Service_Area
	Cost_of_Oil
	Supplier
	Supplier_Phone

Filters	
PK	Filter_ID
	Common_Name
FK	Service_Area
	Cost_of_Filter
	Supplier
	Supplier_Contact
	Supplier_Phone

Storage Unit	
PK	Storage_Unit_ID
FK	Service_Area
	Street_Address
	City
	State
	Zip_Code
	Company
	Rate
	Contact_Name
	Contact_Phone

Oil Inventory	
PK	Storage_Unit_ID
	Oil_ID
	Stock

Filter Inventory	
PK	Storage_Unit_ID
	Filter_ID
	Stock

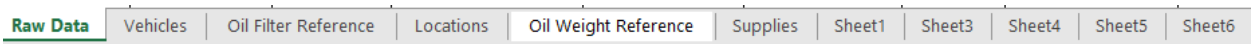
APPENDIX C

Transactions Table Sample

Transaction_ID	Location_Name	Make	Model	Date	Employee_ID	Invoice_Amount	Client_Fleet_Num

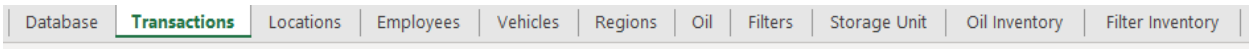
APPENDIX D

Old Transactions Table Structure



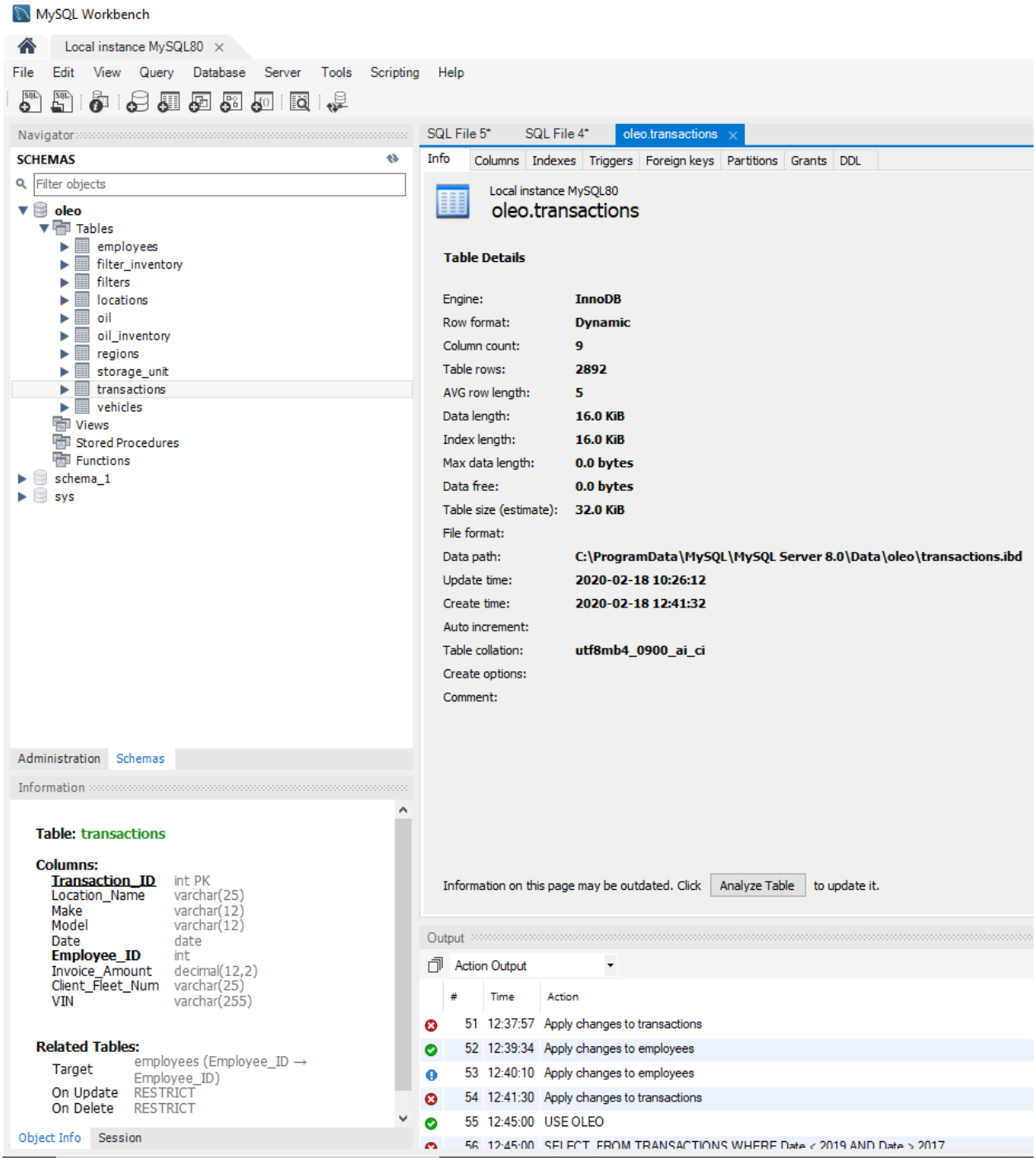
APPENDIX E

New Transactions Table Structure



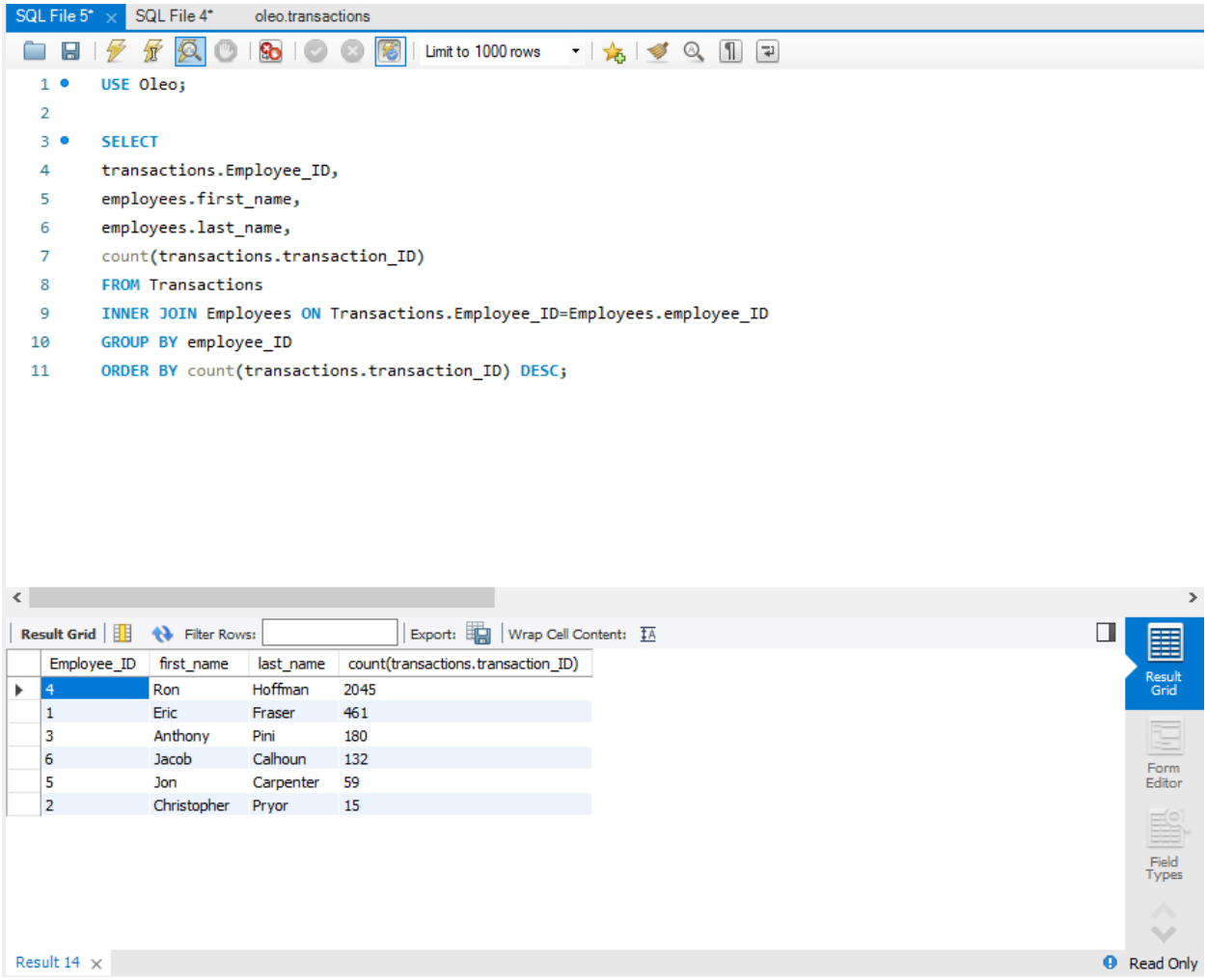
APPENDIX F

MySQL Database Overhead



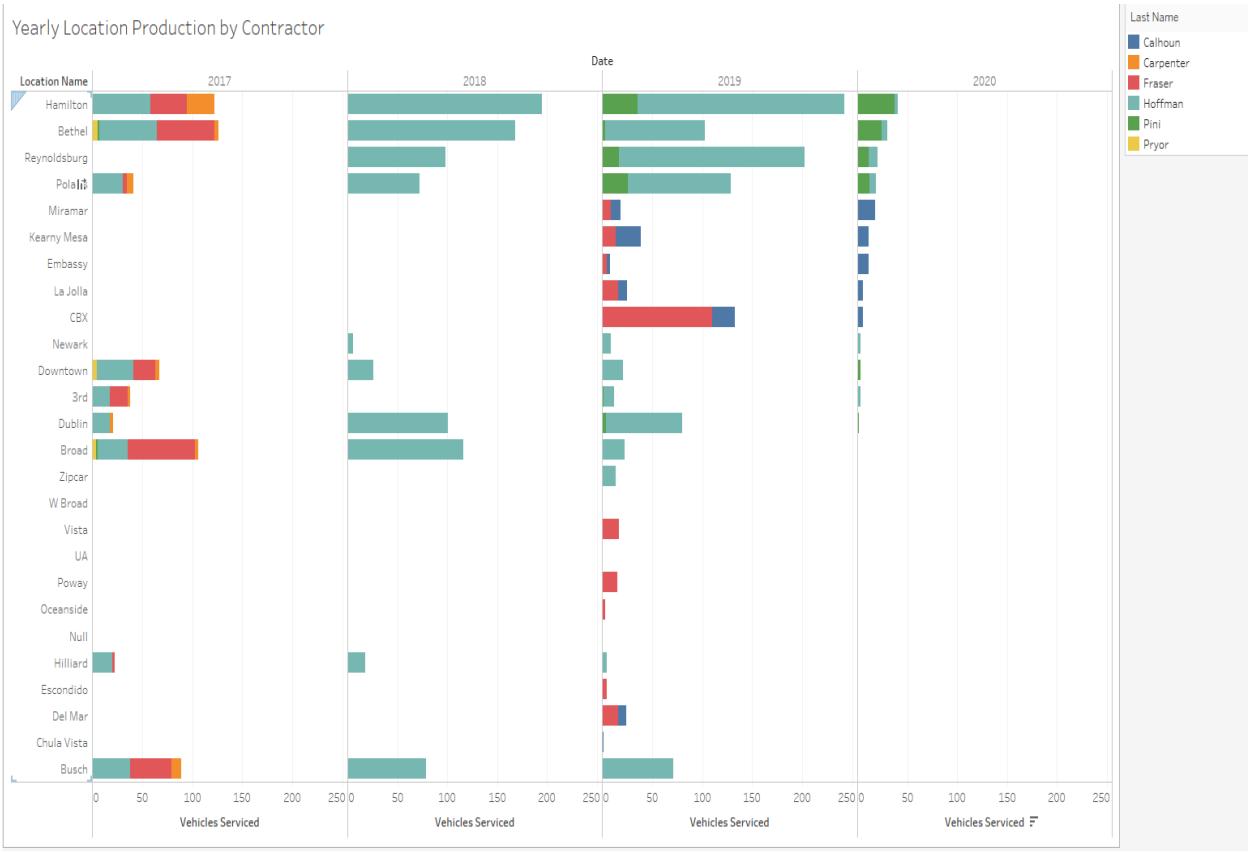
APPENDIX G

Granular Data from Query in MySQL



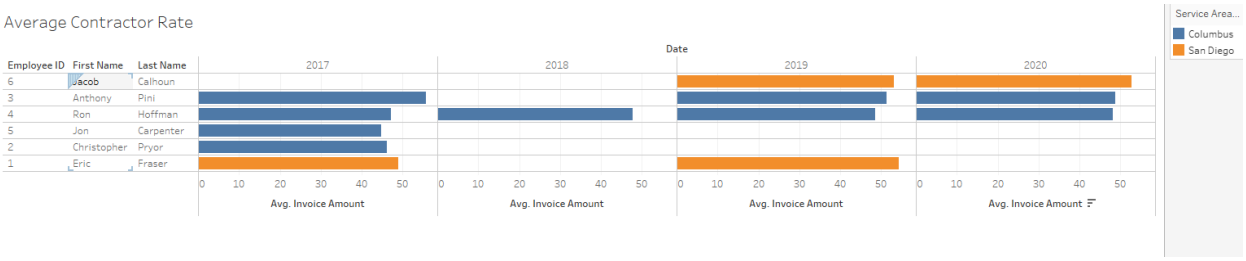
APPENDIX H

Tableau 1: Yearly Location Production by Contractor



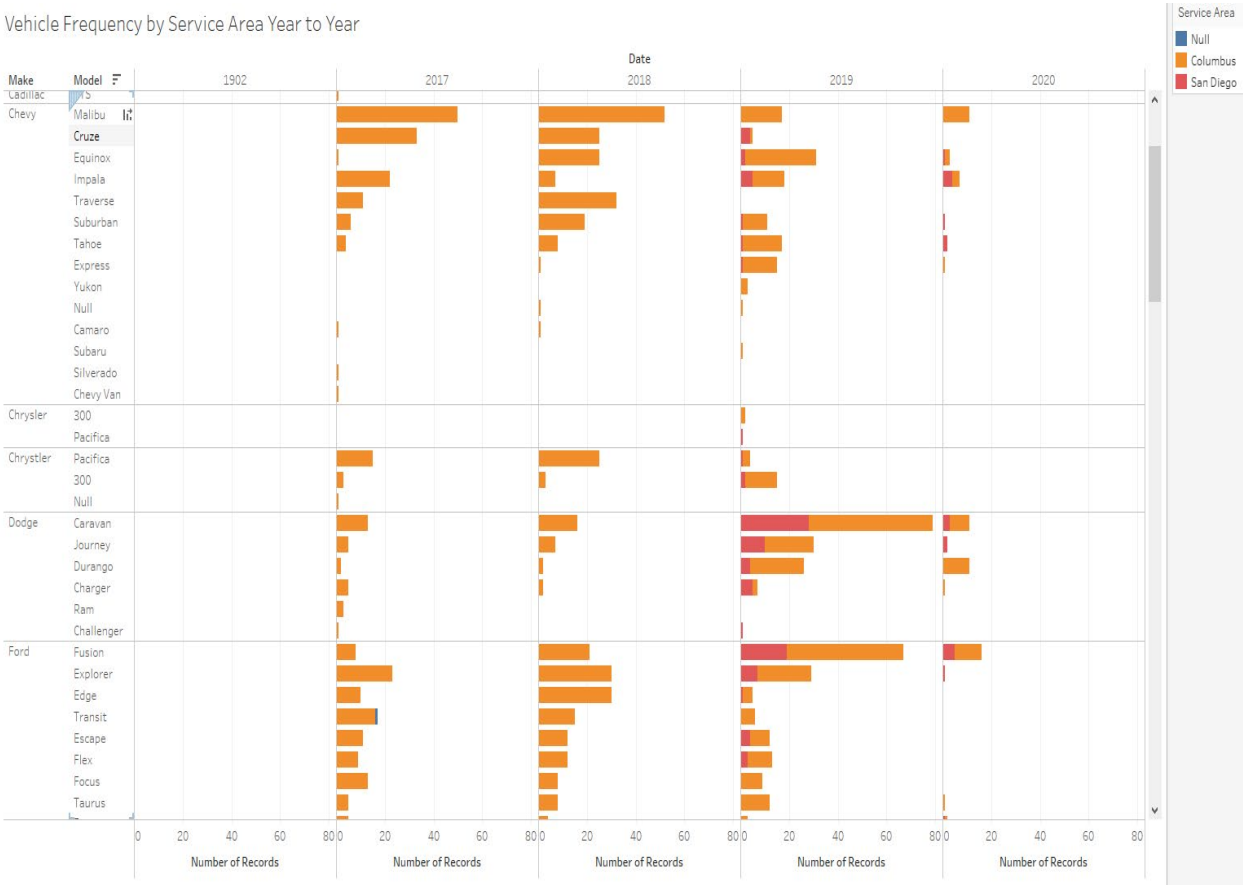
APPENDIX I

Tableau 2: Average Contractor Rate by Year



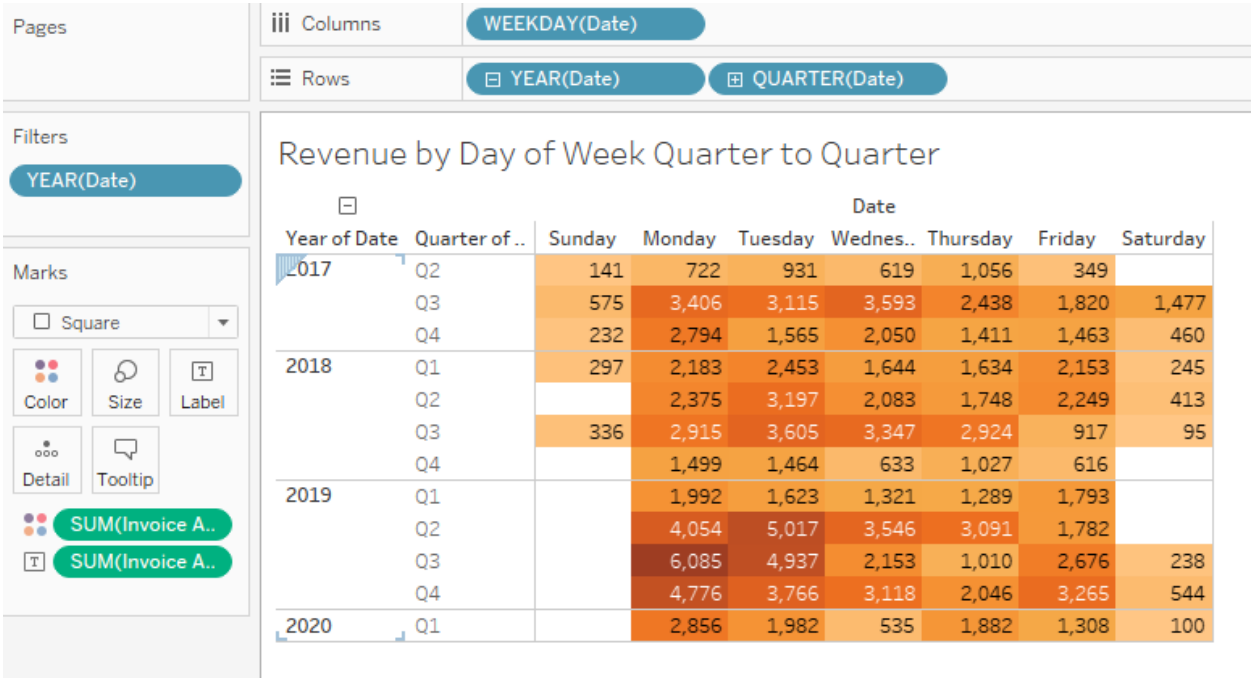
APPENDIX J

Tableau 3: Vehicle Frequency by Service Area Year to Year



APPENDIX K

Tableau 4: Revenue by Day of Week Quarter to Quarter



		Date						
Year of Date	Quarter of ..	Sunday	Monday	Tuesday	Wednes..	Thursday	Friday	Saturday
2017	Q2	141	722	931	619	1,056	349	
	Q3	575	3,406	3,115	3,593	2,438	1,820	1,477
	Q4	232	2,794	1,565	2,050	1,411	1,463	460
2018	Q1	297	2,183	2,453	1,644	1,634	2,153	245
	Q2		2,375	3,197	2,083	1,748	2,249	413
	Q3	336	2,915	3,605	3,347	2,924	917	95
	Q4		1,499	1,464	633	1,027	616	
2019	Q1		1,992	1,623	1,321	1,289	1,793	
	Q2		4,054	5,017	3,546	3,091	1,782	
	Q3		6,085	4,937	2,153	1,010	2,676	238
	Q4		4,776	3,766	3,118	2,046	3,265	544
2020	Q1		2,856	1,982	535	1,882	1,308	100