Building the SQL Database

Original data is stored as text and csv on Google Drive.

# Monthly Updates

Monthly Update Process:

1. For each report, review export instructions.
2. When adding data, make sure titles match.
3. Using count() match records in SQL table with records in .csv.
4. Re-create Customer Table from “calls” table (requires scripts)

Additional Entries

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Date (2019-2020) | Sales Calls | Customers | Job Cost Analysis | Sales Invoice Items | Assignments | Job Summary Analysis |
| Earliest | 2002 | 2000 | 2002 | 2002 | 2014 | 10/1/19 |
| Previous | 32,252 | NA | 119,741 |  | 23,371 |  |
| 10/1-10/31 | 178 | NA | 813 | NA | 40915 |  |
| 11/1-11/30 | 226 | 30853 | 922 | 90897 | 1187 |  |
| 12/1-12/31 | 194 | 55 | 1052 | 539 | 1405 |  |
| 1/1-1/31 | 257 | 73 | 1155 | 524 | 1517 |  |

Note: After uploading new data, check the new uploads. Does everything look right? Are the dates correct? Are there NULL values? the original table rows number does not change, but the new data seems to appear. Yet different again using the SELECT COUNT(\*) FROM [table] query.

# 

# Exporting Data from SSWare

## 1. Sales Calls

## Job/Call Reports -> Calls

Note these are for sales calls only. If we’re looking to track the crew, we need to do something else. Sales calls include estimates, follow-ups and office call types.

Original Data Export

* Date Range: 1/1/2000-10/1/2019
* Call date (not taken date)
* Scheduled calls and completed calls only
* Sales calls only

Export as “calls\_mm\_yyyy.txt”

Pre-processing in R with “calls\_processing\_script.R”

* Adding column names
* Removed extraneous columns
* Formatting strings (dates) as date

Export as “calls\_mm\_yyyy.csv”

Upload the .csv to MySQL, upload both the .txt and .csv to Google Drive, delete the downloaded R script (unless any changes have been made). Update “Monthly Updates” table.

~~Note: rebuilt database from “calls\_processed” using updated calls\_processing\_script.~~

## 

## 2. Customers

## Master List -> Billing Customer List

Data needs to be exported as BOTH a regular TXT file and an “Excel Report file” (ie. not data file) because otherwise the email variable is not included. Save Excel report file as “customers\_mm\_yy.xls” and TXT file as “customers\_mm\_yy.txt”.

Besides specifying the time period, no modifications need to be made to the userform.

**Report File**: In order to clean the report file, take the following steps:

1. Highlight all cells, copy and paste *values* onto a new worksheet (within the same workbook)
2. Delete columns B-M and O-U (inclusive). We want to keep only the columns with the Billing# and Email address.
3. Highlight all cells, copy and paste onto a new worksheet.
4. Download “macros.txt” from Google Drive/Customer and copy into Visual Basic (create a new module). Follow instructions in the macro.
5. Run macro to clean dataset
6. Save cleaned data as “customers\_mm\_yy\_emails.csv”
7. Save modified original workbook

Import TXT file into R and run script “customer.R” to clean and join with emails file. Export as “customers\_mm\_yy.csv.” Upload file to SQL.

Upload .xls, .txt and both .csv files to Google Drive. Delete R script. Update “Monthly Updates” table.

## ~~Create Customer Table from “calls” table~~

~~To create or update the customer table, run the following queries (in /queries/create customer table):~~

1. ~~create\_customer\_table.sql~~
2. ~~delete\_duplicate\_customers.sql~~
3. ~~email\_join.sql (full inner join of all 3 tables- see .drawio model)~~

~~Historical Data Upload: 16,232 rows~~

## 3. Job Cost Analysis

## Management Reports -> Job Cost Analysis

The job cost analysis report is used for callback data but not for installation data because it excludes back end costs.

Original Data Export

* Date Range: 1/1/2000-10/1/2019 (“Job End” date range)
* Departments 20, 21, 22, 30, 45, 50, 55
* No other modifications to the userform

Create “mm-yy” folder, save each file as “dept\_xx\_mm\_yy.TXT”

Pre-processing is being done in R using “JCA\_processing\_script.R”

* Adding column names
* Checking for duplicate rows
* Formatting strings ($’s, %’s) and converting to numeric
* Formatting strings (date) and converting to date object
* Filtering departments

Data is exported from R as “JCA\_01\_2020.csv”. Move the .csv into the previously created “mm\_yy” folder. Upload folder to Google Drive/Job Cost Analysis and upload .csv file to SQL.

## 4. Sales Invoice Items

## AR Invoice Reports -> Sales Invoice Items

Original Data Export:

* Date Range: 1/1/2000-11/30/2019 (“Item Date” date range)

Print Options:

* Uncheck “Include No Charge Invoices”
* Check “Include Hours/PLM Breakdown/Margin”
* Departments 20, 21, 22, 30, 45, 50, 55
* Export as .TXT file

Create “sii\_mm\_yy” folder, save each file as “dept\_xx\_mm\_yy.TXT”

Pre-processing (Greiner DB/R Scripts/sales\_invoice\_preprocess.r)

* Adding column names
* Removing irrelevant columns
* Formatting strings and dates
* Removing invoices not related to jobs
* Checking for duplicate rows

Export data into “sii\_mm\_yy” folder from R as “sii\_mm\_yy.csv”. Upload folder to Google Drive/Sales Invoice Items and upload .csv to SQL.

## 5. Assignments

## Job/Call Reports -> Assignments

Assignment data can be used to track individual crew appointments.

Data Export Instructions

* Date Range: (“Scheduled Date” not “Taken Date”)
* All Assignments: Scheduled, Dispatched, Notified, Completed
* No Sales Calls

Export data as “assignments\_mm\_yy.TXT”

Pre-processing (Greiner DB/R Scripts/assignments\_preprocessing\_script.r)

* Wrangling and cleaning done in R.
* Removed useless/empty columns and added column names.
* Cleaned dates and times.

Export from R as “assignments\_mm\_yy.csv”

*Notes*

In MySQL Workbench, times and dates are left as text because multiple datetime encodings are not allowed in the GUI. Once data was uploaded...with times encoded as “%H:%M” and dates left as text because multiple datetime encodings are not allowed in the GUI. Therefore, data will need to be further cleaned as part of the analysis.

## 6. Job Summary Analysis

Management Reports -> Job Summary Analysis

The job summary analysis provides the raw text related to the call. There are two text fields: “job instructions” and “work completed”.

Data Export Instructions

* Specify Date Range
* Check boxes for “Include Not Closed Jobs”, “Include Non Revenue Jobs” and “Show Expanded Detail”
* Keep marked “All jobs”
* Group by “Job Type”

Export data as an Xlsx Report File because this is the only way to get the text fields.

Data Cleaning

Download macros from Google Drive/Greiner DB/Data/Job Summary/Macros

* Save original as macro-enabled workbook
* Create a new worksheet “Text”, copy and paste original worksheet as text raw data
* Duplicate “Text”, name as “delete\_rows (m)”
* Delete all columns except A, C and M
* Import macro “1\_delete\_rows”. Follow the instructions in the macro.
* Duplicate worksheet “delete\_rows (m)”, name “join\_text (m)”
* Import and run macro “2\_join\_text”, see macro for instructions
* Duplicate worksheet “join\_text (m)” and save as “clean\_data (m)”
* Create a new column A.
* Import and run macro “3\_clean\_data”
* Delete column D, swap columns B (“job instructions”) and C (“call\_type”)
* Name columns: “address”, “job\_inst”, “call\_type”, “work\_completed”
* Duplicate column and name as “cb\_text”
* Sort data by calltype to the following variables: CAZCOR, FLUP, FU-NEW, IC, INFLUP, IN-NW, INWAR, INWMEQ
* Save final worksheet as a separate workbook, “cb\_text\_mm\_yy\_mm\_yy.xlsx”

Upload data to Google Drive (directory above)

Note: At this time, data is not being added to a database.

## 7. Agreements

# Editing Database

1. Primary keys created (Note: for customers table, primary key ‘address’ data type changed from TEXT to VARCHAR(200). This may affect future updates.)