D205 Data Acquisition Performance Assessment

Nicole Reiswig

College of Information Technology, Western Governors University

David Gagner

July 28, 2023

D205 Data Acquisition Performance Assessment

A. Research Question-

The research question is to identify if there is a correlation between the tenure of a customer, the number of service failures, and the length of time it took to correct the failure. “Is there a correlation between the tenure of a customer, the number of service failures, and the length of time it took to correct the failure?” To answer this research question data will be retrieved from both the original churn database “customers” and the CSV file survey_response.

A1. Identifying Data-

To answer our research question we will need to extract data from the original churn database. From this data source, we will use the number of months they’ve been a customer which is listed under the “tenure,” “yearly_equip_failures,” and “outage_sec_week” columns of the customer table. Next, data will be imported from the add-on CSV file survey_response table. We will utilize the columns, “timely_fixes” and “timely_replacements” from the survey_response table. We will compare this data to see if the length of tenure correlates to the number of service failures and the importance rating of timely fixes and timely replacements. We will then draw a conclusion on if there is a correlation on whether service failure and length of time to correct contribute to the length of tenure.

B. Logical Data Model-

See attached screenshot of the logical model.

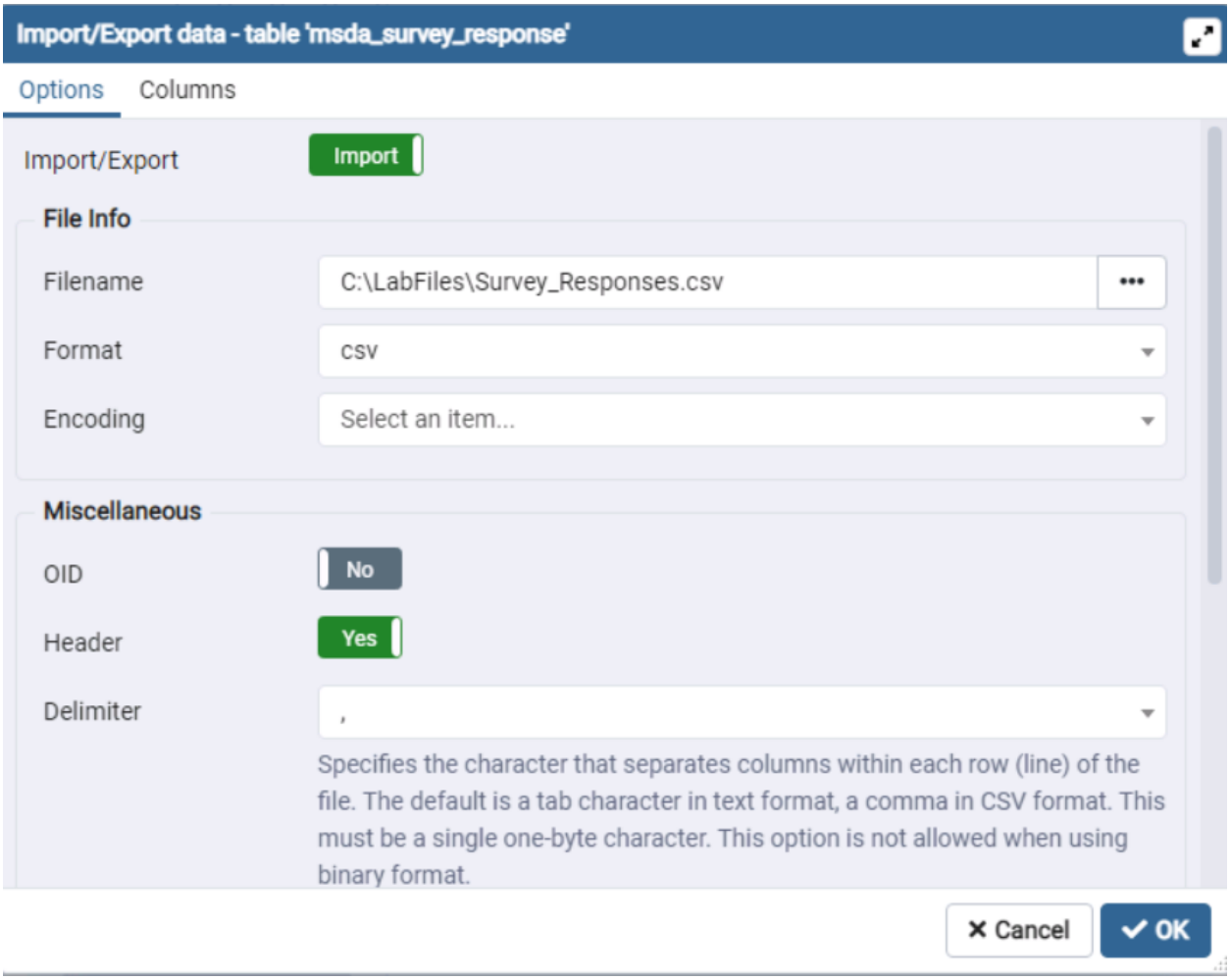
B1.Code For The Physical Data Model-

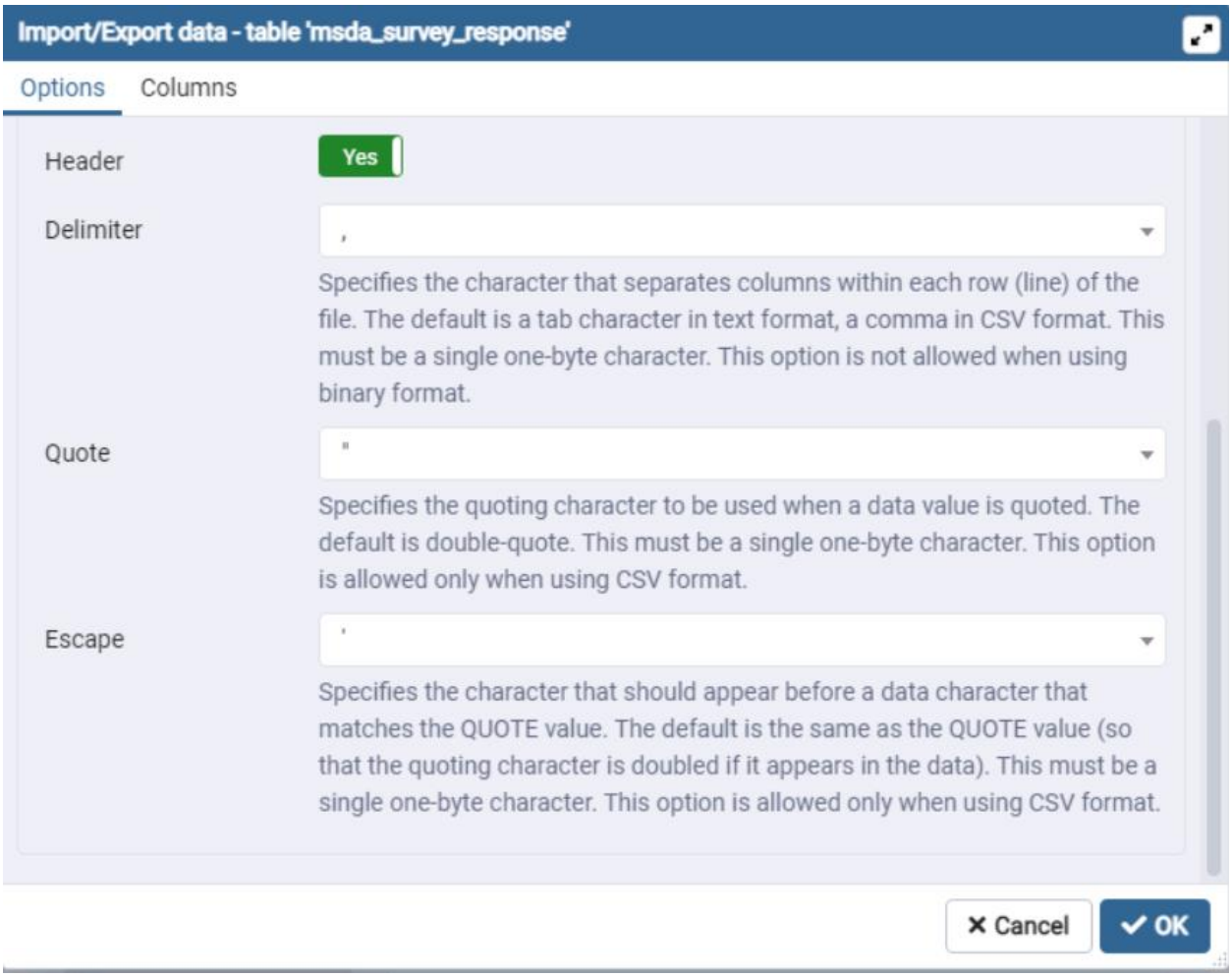
```
CREATE TABLE public.msda_survey_response1
(
    customer_id "char" NOT NULL,
    timely_responses integer,
    timely_fixes integer,
    timely_replacements integer,
    reliability integer,
    options integer,
    respectful_response integer,
    courteous_exchange integer,
    evidence_of_active_listening integer,
    CONSTRAINT "customer.customer_id" PRIMARY KEY (customer_id),
    CONSTRAINT c1 FOREIGN KEY (customer_id)
        REFERENCES public.msda_survey_response (customer_id) MATCH SIMPLE
```

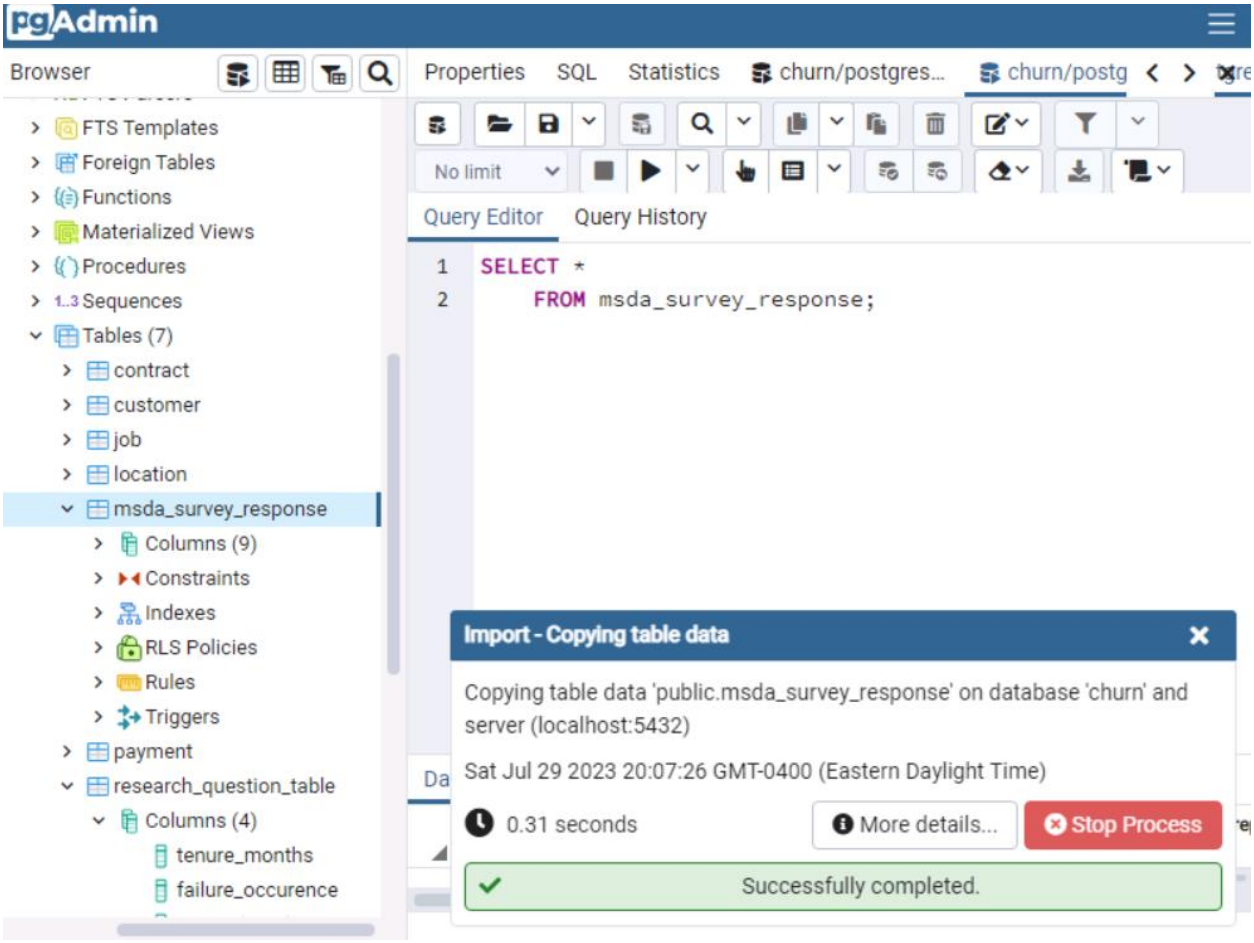
```
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
        NOT VALID
);

ALTER TABLE public.msda_survey_response1
    OWNER to postgres;
```

B2. Loading CSV Data-


Import/Export data - table 'msda_survey_response'
Options
Columns
Import/Export
Import
File Info
Filename
C:\LabFiles\Survey_Responses.csv
Format
csv
Encoding
Select an item...
Miscellaneous
OID
No
Header
Yes
Delimiter
,
Specifies the character that separates columns within each row (line) of the file. The default is a tab character in text format, a comma in CSV format. This must be a single one-byte character. This option is not allowed when using binary format.
Cancel
OK

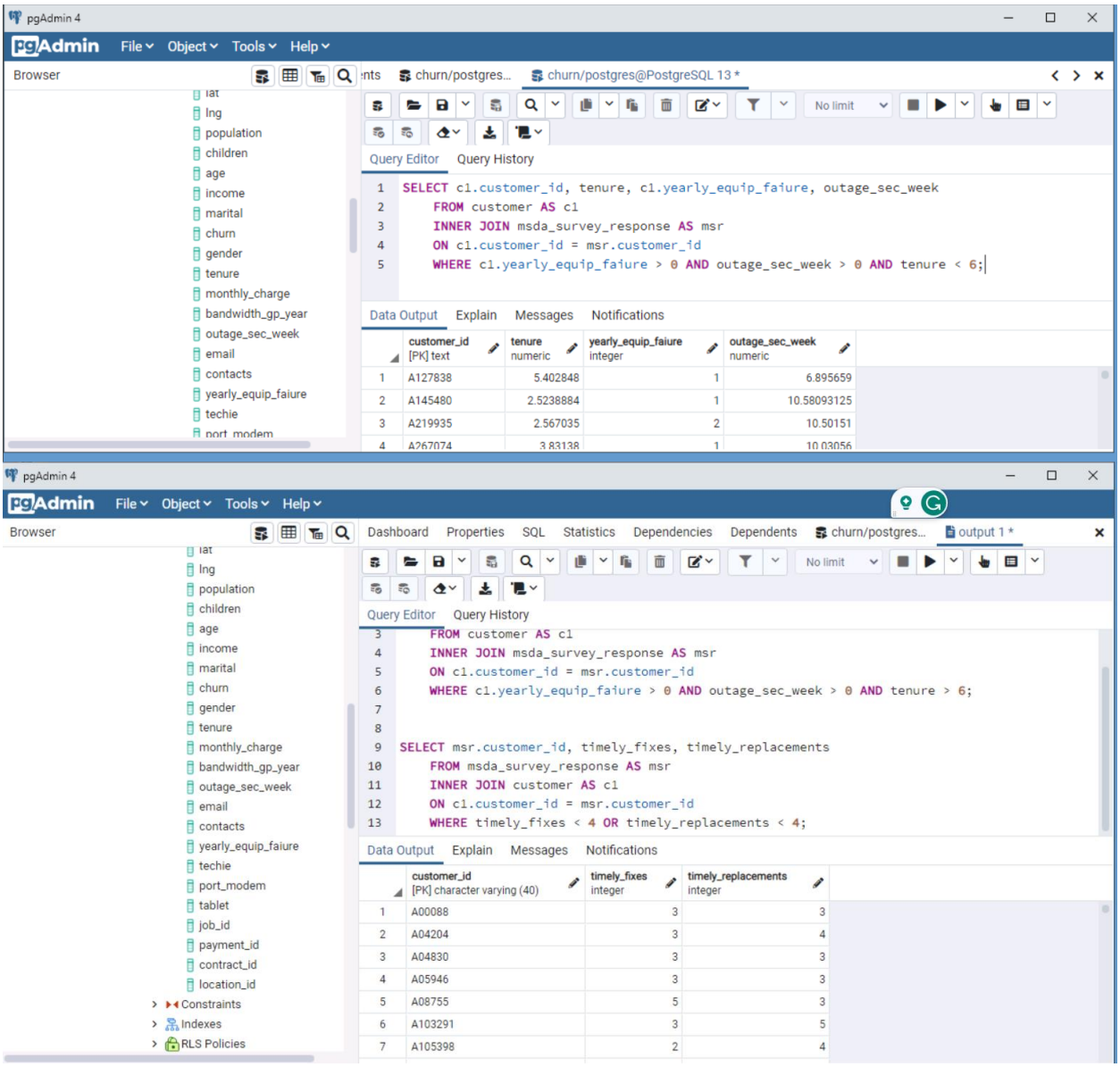

Import/Export data - table 'msda_survey_response'
Options
Columns
Header
Yes
Delimiter
,
Specifies the character that separates columns within each row (line) of the file. The default is a tab character in text format, a comma in CSV format. This must be a single one-byte character. This option is not allowed when using binary format.
Quote
"
Specifies the quoting character to be used when a data value is quoted. The default is double-quote. This must be a single one-byte character. This option is allowed only when using CSV format.
Escape
'
Specifies the character that should appear before a data character that matches the QUOTE value. The default is the same as the QUOTE value (so that the quoting character is doubled if it appears in the data). This must be a single one-byte character. This option is allowed only when using CSV format.
Cancel
OK



C. SQL Query-

1. Locate the number of service failures greater than 0
2. Locate length of tenure less than 6 months
3. Locate outage sec per week greater than 0
4. Importance of timely fixes/replacements greater than 4

```
SELECT c1.customer_id, tenure, c1.yearly_equip_faiure, outage_sec_week,
msr1.customer_id, timely_fixes, timely_replacements
FROM customer AS c1
INNER JOIN msda_survey_response1 AS msr1
ON c1.customer_id = msr1.customer_id
WHERE c1.yearly_equip_faiure > 0 AND outage_sec_week > 0 AND tenure < 6 AND
timely_fixes < 4 OR timely_replacements < 4;
```



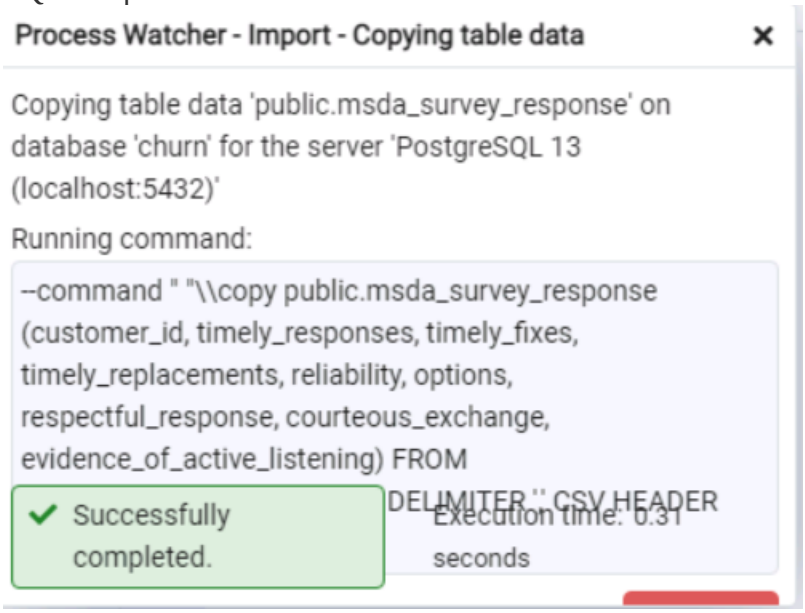
C1. CSV Files-

C:\Users\ntrei\Downloads\output1.csv
C:\Users\ntrei\Downloads\output2.csv

D. Add On File-

The add-on file should be refreshed in the database each time new survey responses become available to acquire the most accurate and relevant data to the business and research question.

E. SQL Script-



F. Panopto Video-

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=3594381d-3982-4267-ad57-b050010297ed

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=3723e7cf-a77d-4e1f-8237-b050014e22de

G. Wed Sources-
Web sources were not used to acquire data or segments of third-party code.

H. Sources-

Sources outside of the original churn database and survey response CSV file were not utilized in my submission.