Guild_Project

December 16, 2021

1 Import Dependencies

[1]:
```
import pandas as pd
from pathlib import Path
import psycopg2
from psycopg2 import OperationalError
from dotenv import load_dotenv
import os
from IPython.display import IFrame
```

1.1 Instructions

[2]:
```
p = Path('Analytics_Engineer_Candidate_Project.pdf')
IFrame(p, width=800, height=1075)
```

[2]:
```
<IPython.lib.display.IFrame at 0x1b76475f788>
```

1.1.1 Create a quick diagram for how you think these tables (tabs) are related (i.e. what is one to many, many to many, etc.)

See here for diagram creation: https://app.gleek.io/diagrams/cFdr3Sx9K712pnUQgTEGsg

[3]:
```
q = Path('Guild_Education_SQL_ER_Diagram.png')
IFrame(q, width=1350, height=650)
```

[3]:
```
<IPython.lib.display.IFrame at 0x1b764770508>
```

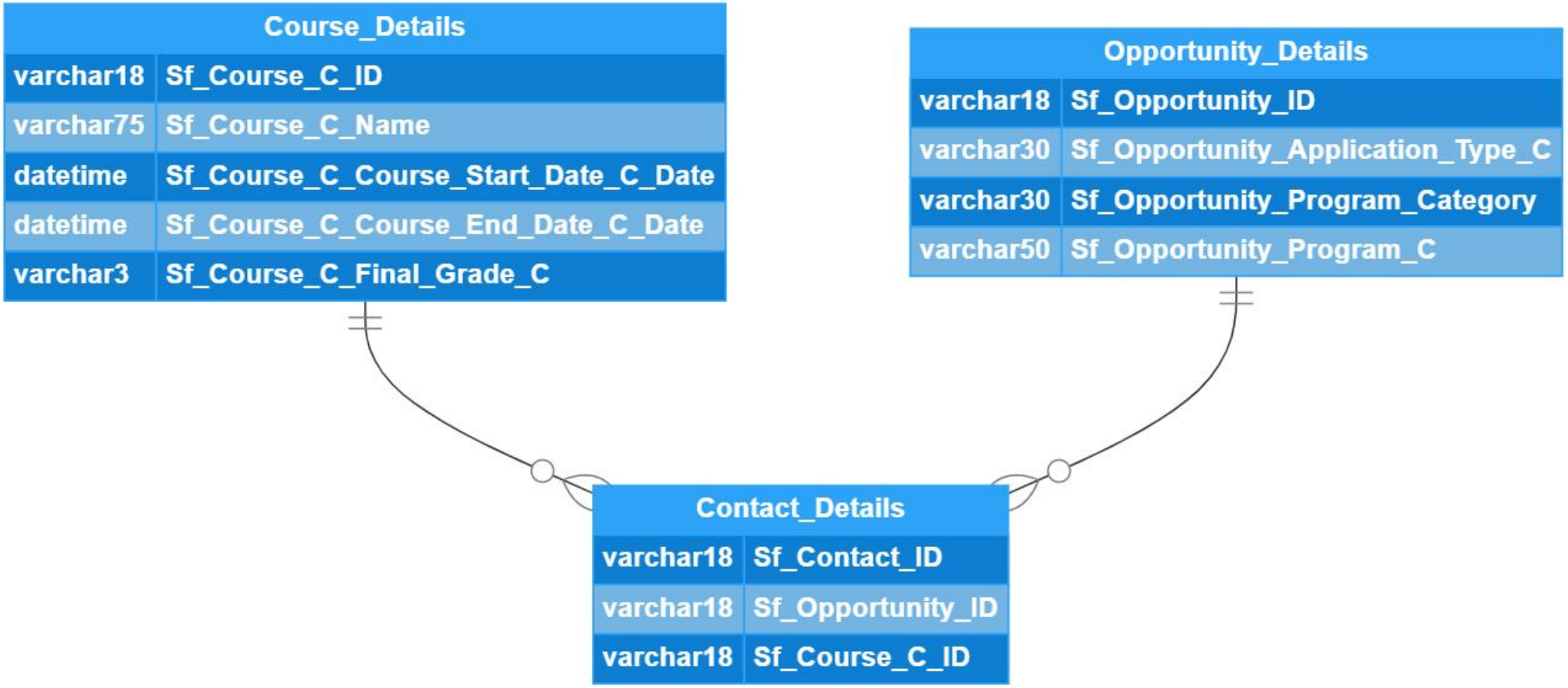
1.1.2 Create a single table to be used for analysis by joining these sources (include your SQL code, and document any issues you may encounter with the data, and any assumptions you make).

Load Postegres Password from `.env` File
To download PostgreSQL: https://www.enterprisedb.com/downloads/postgres-postgresql-downloads

Postgres password saved in file `.env`:
`db_password = 'THE_POSTGRES_PASSWORD_HERE'`

GUILD EDUCATION

Technical Project	**Data Compilation** The Guild Education SQL workbook contains fake data similar to what exists in Guild's internal database. Each tab represents one table. Please make a copy of the workbook, and using the tables provided: • Create a quick diagram for how you think these tables (tabs) are related (i.e. what is one to many, many to many, etc.) • Create a single table to be used for analysis by joining these sources (include your SQL code, and document any issues you may encounter with the data, and any assumptions you make). **Data Understanding & Insights** Use your table to pull the following information (provide 1 query per question and your corresponding answer). Please state any of your assumptions. **SQL is required.** • Total number of contacts (students) who are not enrolled in any courses. • For each program category count the number of distinct contacts (students) that are enrolled in courses. • Based on your best interpretation of the data, which application type has the most number of successful students? **Data Structures** • What ideas do you have, if any, to better structure the data?


Course_Details
varchar18 Sf_Course_C_ID
varchar75 Sf_Course_C_Name
datetime Sf_Course_C_Course_Start_Date_C_Date
datetime Sf_Course_C_Course_End_Date_C_Date
varchar3 Sf_Course_C_Final_Grade_C
Opportunity_Details
varchar18 Sf_Opportunity_ID
varchar30 Sf_Opportunity_Application_Type_C
varchar30 Sf_Opportunity_Program_Category
varchar50 Sf_Opportunity_Program_C
Contact_Details
varchar18 Sf_Contact_ID
varchar18 Sf_Opportunity_ID
varchar18 Sf_Course_C_ID

[4]:
```
load_dotenv()
db_password = os.getenv("db_password")
```

[5]:
```
# db_password
```

[6]:
```
def create_connection(db_name, db_user, db_password, db_host, db_port):
    connection = None
    try:
        connection = psycopg2.connect(
            database=db_name,
            user=db_user,
            password=db_password,
            host=db_host,
            port=db_port,
        )
        print("Connection to PostgreSQL DB successful")
    except OperationalError as e:
        print(f"The error '{e}' occurred")
    return connection
```

[7]:
```
connection = create_connection(
    "postgres", "postgres", db_password, "127.0.0.1", "5432"
)
```

```
Connection to PostgreSQL DB successful
```

[8]:
```
def create_database(connection, query):
    connection.autocommit = True
    cursor = connection.cursor()
    try:
        cursor.execute(query)
        print("Query executed successfully")
    except OperationalError as e:
        print(f"The error '{e}' occurred")
```

[9]:
```
create_database_query = "CREATE DATABASE guild_education_sql" # make sure to␣
↪use lower case
```

[10]:
```
try:
    create_database(connection, create_database_query)
except:
    print('This database already exists.')
```

```
This database already exists.
```

[98]:
```
connection = create_connection(
    "guild_education_sql", "postgres", db_password, "127.0.0.1", "5432"
)
```

```
Connection to PostgreSQL DB successful
```

[14]:
```
def execute_query(connection, query):
    connection.autocommit = True
    cursor = connection.cursor()
    try:
        cursor.execute(query)
        print("Query executed successfully")
    except OperationalError as e:
        print(f"The error '{e}' occurred")
```

Create course_details Table

[99]:
```
create_course_details_table = """
CREATE TABLE IF NOT EXISTS course_details (
    "Sf Course C ID" VARCHAR(18) PRIMARY KEY,
    "Sf Course C Name" VARCHAR(75),
    "Sf Course C Course Start Date C Date" VARCHAR(10),
    "Sf Course C Course End Date C Date" VARCHAR(10),
    "Sf Course C Final Grade C" VARCHAR(75)
)
"""
```

[100]:
```
try:
    execute_query(connection, create_course_details_table)
except:
    print('This table already exists.')
```

```
Query executed successfully
```

Import .csv Data Into course_details Table

[101]:
```
Course_Details = Path(str(Path.cwd()) + '/' + 'data/Course_Details.csv')
print(str(Course_Details))
```

```
C:\Users\mchar\Downloads\Guild_Interview_Project\data\Course_Details.csv
```

[103]:
```
import_course_details_data = """
COPY course_details
FROM '{}'
DELIMITER ','
CSV HEADER;
""".format(str(Course_Details))
```

[105]:
```
execute_query(connection, import_course_details_data)
# If an error gets thrown change the permissions to 'Everyone' ->'Full control'␣
 ↪in properties of the file (if using Windows).
# Use chmod if using macOS or Linux distro.
```

```
Query executed successfully
```

Create opportunity_details Table

[135]:
```
create_opportunity_details_table = """
CREATE TABLE IF NOT EXISTS opportunity_details (
    "Sf Opportunity ID" VARCHAR(18) PRIMARY KEY,
    "Sf Opportunity Application Type C" VARCHAR(75),
    "Sf Opportunity Program Category" VARCHAR(30),
    "Sf Opportunity Program C" VARCHAR(100)
)
"""
```

[137]:
```
try:
    execute_query(connection, create_opportunity_details_table)
except:
    print('This table already exists.')
```

```
Query executed successfully
```

Import .csv Data Into opportunity_details Table

[138]:
```
Opportunity_Details = Path(str(Path.cwd()) + '/' + 'data/Opportunity_Details.
 ↪csv')
print(str(Opportunity_Details))
```

```
C:\Users\mchar\Downloads\Guild_Interview_Project\data\Opportunity_Details.csv
```

[139]:
```
import_opportunity_details_data = """
COPY opportunity_details
FROM '{}'
DELIMITER ','
CSV HEADER;
""".format(str(Opportunity_Details))
```

[140]:
```
execute_query(connection, import_opportunity_details_data)
# If an error gets thrown change the permissions to 'Everyone' ->'Full control'␣
 ↪in properties of the file (if using Windows).
# Use chmod if using macOS or Linux distro.
```

```
Query executed successfully
```

Create contact_details Table

[141]:
```
create_contact_details_table = """
CREATE TABLE IF NOT EXISTS contact_details (
    "Sf Contact ID" VARCHAR(18),
    "Sf Opportunity ID" VARCHAR(18) references opportunity_details("Sf␣
 ↪Opportunity ID"),
    "Sf Course C ID" VARCHAR(18) references course_details("Sf Course C ID")
)
"""
```

[142]:
```
try:
    execute_query(connection, create_contact_details_table)
except:
    print('This table already exists.')
```

```
Query executed successfully
```

Import .csv Data Into contact_details Table

[143]:
```
Contact_Details = Path(str(Path.cwd()) + '/' + 'data/Contact_Details.csv')
print(str(Contact_Details))
```

```
C:\Users\mchar\Downloads\Guild_Interview_Project\data\Contact_Details.csv
```

[144]:
```
import_contact_details_data = """
COPY contact_details
FROM '{}'
DELIMITER ','
CSV HEADER;
""".format(str(Contact_Details))
```

[145]:
```
execute_query(connection, import_contact_details_data)
# If an error gets thrown change the permissions to 'Everyone' ->'Full control'␣
 ↪in properties of the file (if using Windows).
# Use chmod if using macOS or Linux distro.
```

```
Query executed successfully
```

Join the Tables Into a Single Table

[189]:
```
join_query = """
CREATE TABLE single_table AS
SELECT
    contact_details."Sf Contact ID",
    contact_details."Sf Course C ID",
    contact_details."Sf Opportunity ID",
    course_details."Sf Course C Name",
    course_details."Sf Course C Course Start Date C Date",
    course_details."Sf Course C Course End Date C Date",
    course_details."Sf Course C Final Grade C",
    opportunity_details."Sf Opportunity Application Type C",
    opportunity_details."Sf Opportunity Program Category",
    opportunity_details."Sf Opportunity Program C"
FROM contact_details
FULL OUTER JOIN course_details
ON contact_details."Sf Course C ID" = course_details."Sf Course C ID"
FULL JOIN opportunity_details
ON contact_details."Sf Opportunity ID" = opportunity_details."Sf Opportunity␣
 ↪ID";
"""
```

[190]:
```
try:
    execute_query(connection, join_query)
except:
    print('This table already exists.')
```

```
Query executed successfully
```

Verify single_table Creation

[154]:
```
def execute_read_query(connection, query):
    cursor = connection.cursor()
    result = None
    try:
        cursor.execute(query)
        result = cursor.fetchall()
        return result
    except OperationalError as e:
        print(f"The error '{e}' occurred")
```

[191]:
```
select_single_table = """
SELECT * FROM single_table
"""
```

[192]:
```
this_query = execute_read_query(connection, select_single_table)
```

[193]:
```
incrementor = 0
while incrementor < 10:
    print(this_query[incrementor])
    incrementor += 1
```

```
('0033600000q4ADOAA2', None, '00636000005dxASAAY', None, None, None, None,
'Guild Education', 'Lead Gen - Post Secondary', 'Giving and Receiving Feedback')
('0033600000BMcisAAD', None, '00636000005eG7BAAU', None, None, None, None,
'Western Governors University', 'University', 'B.A. in Interdisciplinary Studies
(K-8)')
('00336000009m01JAAQ', None, '00636000005eGPxAAM', None, None, None, None,
'Western Governors University', 'University', 'B.A. in Interdisciplinary Studies
(K-8)')
('0033600000BOI13AAH', None, '00636000005eNIpAAM', None, None, None, None,
'Western Governors University', 'University', 'B.A. in Interdisciplinary Studies
(K-8)')
('0033600000BNSXNAA5', None, '00636000005eOcBAAU', None, None, None, None,
'Western Governors University', 'University', 'B.A. in Interdisciplinary Studies
(K-8)')
('0033600000BYnWzAAL', None, '00636000005eZYRAA2', None, None, None, None,
'Western Governors University', 'University', 'B.A. in Interdisciplinary Studies
(K-8)')
('0033600000BYnXdAAL', None, '00636000005eZZcAAM', None, None, None, None,
'Western Governors University', 'University', 'B.A. in Interdisciplinary Studies
```

```
(K-8)')
('0033600000BZ63EAAT', None, '00636000005ecshAAA', None, None, None, None,
'Western Governors University', 'University', 'B.A. in Interdisciplinary Studies
(K-8)')
('0033600000q4AGNAA2', None, '00636000005elT6AAI', None, None, None, None,
'Western Governors University', 'University', 'B.A. in Interdisciplinary Studies
(K-8)')
('00336000009lzv6AAA', None, '00636000005elTQAAY', None, None, None, None,
'Western Governors University', 'University', 'B.A. in Interdisciplinary Studies
(K-8)')
```

1.2 Use your table to pull the following information (provide 1 query per question and your corresponding answer). Please state any of your assumptions. SQL is required

1.2.1 Total number of contacts (students) who are not enrolled in any courses.

[228]:
```
# Assuming that 'enrolled' means 'at any time' in a course(s).

students_not_enrolled = """
SELECT COUNT(DISTINCT(single_table."Sf Contact ID"))
FROM single_table
WHERE single_table."Sf Course C ID" IS NULL;
"""
```

[229]:
```
query_1 = execute_read_query(connection, students_not_enrolled)
query_1
```

[229]:
```
[(18502,)]
```

[232]:
```
# If it is are not 'currently' enrolled in any courses.
# students_not_enrolled_alt = """
# SELECT COUNT(DISTINCT(single_table."Sf Contact ID"))
# FROM single_table
# WHERE single_table."Sf Course C Final Grade C" IS NULL
# """
```

[233]:
```
# query_1_1 = execute_read_query(connection, students_not_enrolled_alt)
# query_1_1
```

[233]:
```
[(18504,)]
```

1.2.2 For each program category count the number of distinct contacts (students) that are enrolled in courses.

[222]:
```
program_category_enrollment = """
SELECT single_table."Sf Opportunity Program Category",␣
 ↪COUNT(DISTINCT(single_table."Sf Contact ID"))
```

```
FROM single_table
GROUP BY single_table."Sf Opportunity Program Category";
"""
```

[227]:
```
query_2 = execute_read_query(connection, program_category_enrollment)
query_2
```

[227]:
```
[('Credential', 304),
 ('ELL Program', 183),
 ('Lead Gen - Post Secondary', 541),
 ('Lead Gen - Secondary', 689),
 ('StraighterLine/Saylor', 173),
 ('University', 1765),
 (None, 15151)]
```

1.2.3 Based on your best interpretation of the data, which application type has the most number of successful students?

[234]:
```
most_successful_app = """
SELECT
    single_table."Sf Opportunity Application Type C",
    COUNT(DISTINCT(single_table."Sf Contact ID"))
FROM single_table
WHERE single_table."Sf Course C Final Grade C"
IN ('Passed', 'C-', 'C', 'C+', 'B-', 'B', 'B+', 'A-', 'A', 'A+')
GROUP BY single_table."Sf Opportunity Application Type C";
"""
```

[235]:
```
query_3 = execute_read_query(connection, most_successful_app)
query_3
```

[235]:
```
[('Bellevue University', 130),
 ('Brandman University', 71),
 ('Colorado State University-Global', 65),
 ('Guild Education', 470),
 ('Penn Foster High School', 87),
 ('StraighterLine', 35),
 ('University of Denver', 3),
 ('Western Governors University', 39),
 ('Wilmington University', 44)]
```

Guild Education greatly has the most number of successful students.

1.2.4 What ideas do you have, if any, to better structure the data?

"Sf Course C Final Grade C" could be numerically encoded. This may make it easier to determine metics such as "GPA".

2 Pandas Check

[8]:
```
p0 = Path('data\Contact_Details.csv')
p1 = Path('data\Course_Details.csv')
p2 = Path('data\Opportunity_Details.csv')
```

[9]:
```
df_Contact_Details = pd.read_csv (p0)
df_Course_Details = pd.read_csv (p1)
df_Opportunity_Details = pd.read_csv (p2)
```

[10]:
```
df_Contact_Details
```

[10]:
```
            Sf Contact ID   Sf Opportunity ID Sf Course C ID
0      0033600000q4ADOAA2  00636000005dxASAAY            NaN
1      0033600000BMcisAAD  00636000005eG7BAAU            NaN
2      00336000009m01JAAQ  00636000005eGPxAAM            NaN
3      0033600000BOI13AAH  00636000005eNIpAAM            NaN
4      0033600000BNSXNAA5  00636000005eOcBAAU            NaN
…                     …                   …             …
27024  0033600001WRdHtAAL                 NaN            NaN
27025  0033600001WStcKAAT                 NaN            NaN
27026  0033600001WUOAMAA5                 NaN            NaN
27027  0033600001at9GLAAY                 NaN            NaN
27028  0033600001e029bAAA                 NaN            NaN

[27029 rows x 3 columns]
```

[70]:
```
df_Contact_Details.describe()
```

[70]:
```
             Sf Contact ID   Sf Opportunity ID      Sf Course C ID
count                27029               11878                7572
unique               18504                6968                7572
top     0033600000VDyNSAA1  0063600000IfR5bAAF  a1C36000007TGoGEAW
freq                    46                  43                   1
```

[73]:
```
null_df = pd.DataFrame({'null' : df_Contact_Details.isnull().sum()})
null_df
```

[73]:
```
                    null
Sf Contact ID          0
Sf Opportunity ID  15151
Sf Course C ID     19457
```

[ ]:
```

```

[11]:
```
df_Course_Details
```

[11]:

	Sf Course C ID	Sf Course C Name	Sf Course C Course Start Date C Date	Sf Course C Course End Date C Date	Sf Course C Final Grade C
0	a1C36000009t3tGEAQ	High School Completion Program	NaN	NaN	NaN
1	a1C36000009eJkMEAU	Management Training Program - 16 Week	NaN	NaN	NaN
2	a1C36000009u2k9EAA	High School Completion Program	NaN	NaN	NaN
3	a1C36000009fIoUEAU	High School Completion Program	NaN	NaN	NaN
4	a1C36000005okzZEAQ	Management Training Program - 16 Week	NaN	NaN	NaN
…	…	…	…	…	…
7567	a1C36000005omBXEAY	Organizational Behavior	11/19/2001	1/13/2002	A
7568	a1C36000005omCfEAI	Introduction to Research	11/19/2001	1/13/2002	A
7569	a1C36000005omBsEAI	Developing Management Skills	8/27/2001	11/18/2001	A
7570	a1C36000005omC1EAI	Business Information Systems	8/27/2001	11/18/2001	A
7571	a1C36000005omC6EAI	Essentials of Management	8/27/2001	11/18/2001	A

[7572 rows x 5 columns]

[74]:
```
df_Course_Details.describe()
```

[74]:

	Sf Course C ID	Sf Course C Name
count	7572	7572
unique	7572	879
top	a1C36000007TGoGEAW	High School Completion Program

```
freq                    1                              528

       Sf Course C Course Start Date C Date  \
count                                  7563
unique                                  417
top                                3/12/2018
freq                                    523

       Sf Course C Course End Date C Date Sf Course C Final Grade C
count                                7539                      2736
unique                                519                        24
top                                 5/6/2018                         A
freq                                  311                       664
```

[75]:
```
null_df = pd.DataFrame({'null' : df_Course_Details.isnull().sum()})
null_df
```

[75]:
```
                                    null
Sf Course C ID                         0
Sf Course C Name                       0
Sf Course C Course Start Date C Date     9
Sf Course C Course End Date C Date      33
Sf Course C Final Grade C             4836
```

[12]:
```
df_Opportunity_Details
```

[12]:
```
        Sf Opportunity ID Sf Opportunity Application Type C  \
0     00636000005dxASAAY                   Guild Education
1     00636000005eG7BAAU      Western Governors University
2     00636000005eGPxAAM      Western Governors University
3     00636000005eNIpAAM      Western Governors University
4     00636000005eOcBAAU      Western Governors University
…                    …                                 …
6963  0063600000f8ny7AAA                   Guild Education
6964  0063600000f8nz0AAA           Penn Foster High School
6965  0063600000f8o0oAAA           Penn Foster High School
6966  0063600000f8o5EAAQ           Penn Foster High School
6967  0063600000f8o5sAAA                  Ashworth College

     Sf Opportunity Program Category                Sf Opportunity Program C
0          Lead Gen - Post Secondary            Giving and Receiving Feedback
1                         University  B.A. in Interdisciplinary Studies (K-8)
2                         University  B.A. in Interdisciplinary Studies (K-8)
3                         University  B.A. in Interdisciplinary Studies (K-8)
4                         University  B.A. in Interdisciplinary Studies (K-8)
…                                …                                         …
6963       Lead Gen - Post Secondary           High School Completion Program
```

```
6964            Lead Gen - Secondary                Walmart High School Track
6965            Lead Gen - Secondary                Walmart High School Track
6966            Lead Gen - Secondary                Walmart High School Track
6967                      University      Electrician Training Career Diploma

[6968 rows x 4 columns]
```

[76]:
```
df_Opportunity_Details.describe()
```

[76]:
```
         Sf Opportunity ID Sf Opportunity Application Type C  \
count                 6968                                6965
unique                6968                                  13
top     0063600000WAdvyAAD                   Guild Education
freq                     1                                2527

       Sf Opportunity Program Category   Sf Opportunity Program C
count                             6968                       6965
unique                               6                        292
top                         University  Walmart High School Track
freq                              2859                        719
```

[77]:
```
null_df = pd.DataFrame({'null' : df_Opportunity_Details.isnull().sum()})
null_df
```

[77]:
```
                                  null
Sf Opportunity ID                    0
Sf Opportunity Application Type C    3
Sf Opportunity Program Category      0
Sf Opportunity Program C             3
```

[79]:
```
df_Contact_Details.columns.to_list()
```

[79]:
```
['Sf Contact ID', 'Sf Opportunity ID', 'Sf Course C ID']
```

[80]:
```
df_Course_Details.columns.to_list()
```

[80]:
```
['Sf Course C ID',
 'Sf Course C Name',
 'Sf Course C Course Start Date C Date',
 'Sf Course C Course End Date C Date',
 'Sf Course C Final Grade C']
```

[81]:
```
df_Opportunity_Details.columns.to_list()
```

[81]:
```
['Sf Opportunity ID',
 'Sf Opportunity Application Type C',
 'Sf Opportunity Program Category',
 'Sf Opportunity Program C']
```

[92]:
```
single_table = df_Contact_Details.merge(df_Course_Details, how='outer', on='Sf Course C ID').merge(df_Opportunity_Details, how='outer', on='Sf Opportunity ID')
```

[93]:
```
single_table
```

[93]:

```
            Sf Contact ID   Sf Opportunity ID      Sf Course C ID  \
0      0033600000q4ADOAA2  00636000005dxASAAY                 NaN
1      0033600000BMcisAAD  00636000005eG7BAAU                 NaN
2      00336000009m01JAAQ  00636000005eGPxAAM                 NaN
3      0033600000BOI13AAH  00636000005eNIpAAM                 NaN
4      0033600000BNSXNAA5  00636000005eOcBAAU                 NaN
…                     …                   …                  …
27024  0033600001ewXqeAAE  0063600000f8ikCAAQ  a1C36000009uvGuEAI
27025  0033600001ewjc9AAA  0063600000f8ikHAAQ  a1C36000009uvH4EAI
27026  0033600001ewXjyAAE  0063600000f8ikMAAQ  a1C36000009uvH9EAI
27027  0033600001ex5HkAAI  0063600000f8mvIAAQ  a1C36000009v8XmEAI
27028  0033600001exBCHAA2  0063600000f8ny7AAA  a1C36000009vD3SEAU

                          Sf Course C Name  \
0                                      NaN
1                                      NaN
2                                      NaN
3                                      NaN
4                                      NaN
…                                      …
27024                          Spanish I
27025                          Spanish I
27026  Information Technology Fundamentals
27027                         Precalculus
27028       High School Completion Program

      Sf Course C Course Start Date C Date Sf Course C Course End Date C Date  \
0                                      NaN                                NaN
1                                      NaN                                NaN
2                                      NaN                                NaN
3                                      NaN                                NaN
4                                      NaN                                NaN
…                                      …                                  …
27024                            5/1/2018                           9/1/2018
27025                           5/30/2018                          9/30/2018
27026                           5/30/2018                          9/30/2018
27027                           5/30/2018                          9/30/2018
27028                                NaN                                NaN

      Sf Course C Final Grade C Sf Opportunity Application Type C  \
0                           NaN                   Guild Education
```

```
1                          NaN      Western Governors University
2                          NaN      Western Governors University
3                          NaN      Western Governors University
4                          NaN      Western Governors University
…                          …                                 …
27024                      NaN                   StraighterLine
27025                      NaN                   StraighterLine
27026                      NaN                   StraighterLine
27027                      NaN                   StraighterLine
27028                      NaN                 Guild Education

      Sf Opportunity Program Category                Sf Opportunity Program C
0           Lead Gen - Post Secondary            Giving and Receiving Feedback
1                          University  B.A. in Interdisciplinary Studies (K-8)
2                          University  B.A. in Interdisciplinary Studies (K-8)
3                          University  B.A. in Interdisciplinary Studies (K-8)
4                          University  B.A. in Interdisciplinary Studies (K-8)
…                                 …                                       …
27024           StraighterLine/Saylor                                Spanish I
27025           StraighterLine/Saylor                                Spanish I
27026           StraighterLine/Saylor      Information Technology Fundamentals
27027           StraighterLine/Saylor                              Precalculus
27028       Lead Gen - Post Secondary           High School Completion Program

[27029 rows x 10 columns]
```

[94]:
```
single_table.describe()
```

[94]:
```
             Sf Contact ID   Sf Opportunity ID      Sf Course C ID  \
count                27029               11878                7572
unique               18504                6968                7572
top     0033600000VDyNSAA1  0063600000IfR5bAAF  a1C36000007TGoGEAW
freq                    46                  43                   1

                     Sf Course C Name Sf Course C Course Start Date C Date  \
count                            7572                                 7563
unique                            879                                  417
top     High School Completion Program                            3/12/2018
freq                              528                                  523

       Sf Course C Course End Date C Date Sf Course C Final Grade C  \
count                              7539                       2736
unique                              519                         24
top                            5/6/2018                          A
freq                                311                        664

       Sf Opportunity Application Type C Sf Opportunity Program Category  \
```

```
count                                11875                                11878
unique                                  13                                    6
top                    Bellevue University                           University
freq                                  2815                                 7169

             Sf Opportunity Program C
count                           11875
unique                            292
top     High School Completion Program
freq                              894
```

[103]:
```
single_table.dtypes
```

[103]:
```
Sf Contact ID                          object
Sf Opportunity ID                      object
Sf Course C ID                         object
Sf Course C Name                       object
Sf Course C Course Start Date C Date   object
Sf Course C Course End Date C Date     object
Sf Course C Final Grade C              object
Sf Opportunity Application Type C      object
Sf Opportunity Program Category        object
Sf Opportunity Program C               object
dtype: object
```

[97]:
```
df_Contact_Details.describe()
```

[97]:
```
             Sf Contact ID   Sf Opportunity ID      Sf Course C ID
count                27029               11878                7572
unique               18504                6968                7572
top     0033600000VDyNSAA1  0063600000IfR5bAAF  a1C36000007TGoGEAW
freq                    46                  43                   1
```

[98]:
```
df_Course_Details.describe()
```

[98]:
```
            Sf Course C ID               Sf Course C Name  \
count                 7572                           7572
unique                7572                            879
top     a1C36000007TGoGEAW  High School Completion Program
freq                     1                            528

       Sf Course C Course Start Date C Date  \
count                                  7563
unique                                  417
top                                3/12/2018
freq                                    523
```

```
       Sf Course C Course End Date C Date Sf Course C Final Grade C
count                                 7539                       2736
unique                                 519                         24
top                               5/6/2018                          A
freq                                   311                        664
```

[99]:
```
single_table.describe()
```

[99]:
```
             Sf Contact ID   Sf Opportunity ID      Sf Course C ID  \
count                27029               11878                7572
unique               18504                6968                7572
top     0033600000VDyNSAA1  0063600000IfR5bAAF  a1C36000007TGoGEAW
freq                    46                  43                   1

                     Sf Course C Name Sf Course C Course Start Date C Date  \
count                            7572                                 7563
unique                            879                                  417
top     High School Completion Program                              3/12/2018
freq                              528                                  523

       Sf Course C Course End Date C Date Sf Course C Final Grade C  \
count                                 7539                       2736
unique                                 519                         24
top                               5/6/2018                          A
freq                                   311                        664

       Sf Opportunity Application Type C Sf Opportunity Program Category  \
count                              11875                           11878
unique                                13                               6
top                   Bellevue University                      University
freq                                2815                            7169

             Sf Opportunity Program C
count                           11875
unique                            292
top     High School Completion Program
freq                              894
```

[96]:
```
null_df = pd.DataFrame({'null' : single_table.isnull().sum()})
null_df
```

[96]:
```
                                    null
Sf Contact ID                          0
Sf Opportunity ID                  15151
Sf Course C ID                     19457
Sf Course C Name                   19457
Sf Course C Course Start Date C Date  19466
```

```
Sf Course C Course End Date C Date    19490
Sf Course C Final Grade C             24293
Sf Opportunity Application Type C     15154
Sf Opportunity Program Category       15151
Sf Opportunity Program C              15154
```

[100]:
```
null_df = pd.DataFrame({'null' : df_Contact_Details.isnull().sum()})
null_df
```

[100]:
```
                     null
Sf Contact ID           0
Sf Opportunity ID   15151
Sf Course C ID      19457
```

[101]:
```
null_df = pd.DataFrame({'null' : df_Course_Details.isnull().sum()})
null_df
```

[101]:
```
                                    null
Sf Course C ID                         0
Sf Course C Name                       0
Sf Course C Course Start Date C Date   9
Sf Course C Course End Date C Date    33
Sf Course C Final Grade C           4836
```

[102]:
```
null_df = pd.DataFrame({'null' : single_table.isnull().sum()})
null_df
```

[102]:
```
                                     null
Sf Contact ID                           0
Sf Opportunity ID                   15151
Sf Course C ID                      19457
Sf Course C Name                    19457
Sf Course C Course Start Date C Date  19466
Sf Course C Course End Date C Date  19490
Sf Course C Final Grade C           24293
Sf Opportunity Application Type C   15154
Sf Opportunity Program Category     15151
Sf Opportunity Program C            15154
```

Export Single Table to CSV File

[110]:
```
single_table_path = Path('data\single_table.csv')
```

[111]:
```
single_table.to_csv(single_table_path)
```

3 Appendix

See `SQL_QUIERIES.txt` for straight SQL used and `notes.txt` for additional info re analysis.