Simcoe County Landfill Site Data Analytics Project

BDAT1000 FINAL EXAM

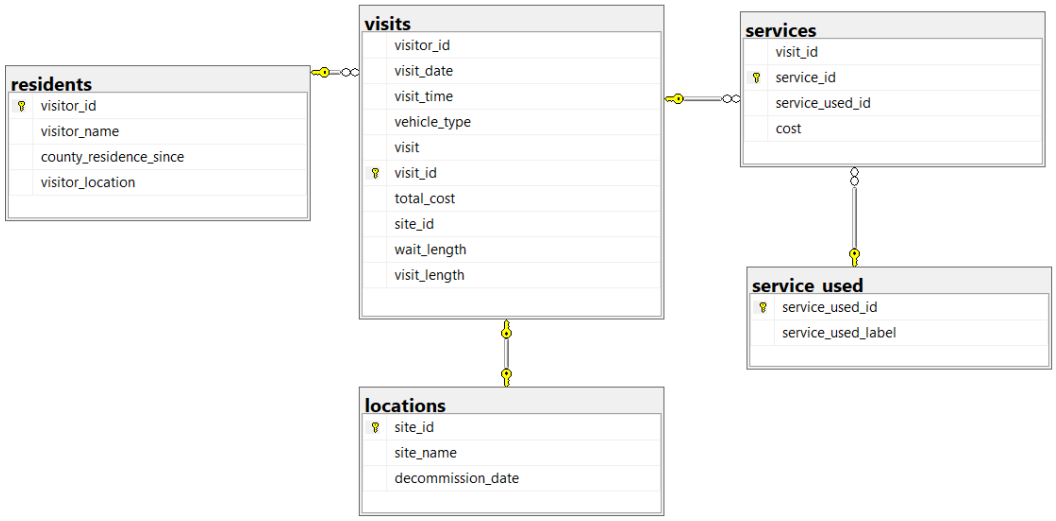
1. You have until end-of-day Dec-14 to submit your exam
2. Once you open the exam on Blackboard, you will have 4.5 hours to complete it (I upped it by an hour). You can take an additional 2.25 hours to submit your work. There will be a 5 mark deduction for each additional half hour taken.
3. When you have completed the exam upload the following;
   - ✓ A full-screen capture that includes your open SAS Studio app with the top lines of your code showing,
   - ✓ Your csv files,
   - ✓ Your SQL script file containing all your sql syntax,
   - ✓ Finally, your MS Excel dashboard workbook.
   - ✓ The exam is out of 75

I have provided you a SAS code file (Final Exam File Creation FINAL.sas) for the creation of synthetic data for the proposed design of a Landfill analytics tiny data warehouse. The SAS code will create the following three csv files (unique to you);

- residents.csv
- visits.csv (wait_length and visit_length in minutes. visit_length is the time spent in the landfill (after having waited)
- Services.csv

Run the SAS code and download the three csv files generated (5 marks). For the two tables in the Additional Tables.xlsx workbook create a locations.csv and a services_used.csv files (2 Marks).

Then create a database in SQL for the exam called ‘final’ (see diagram on the next page). Import into the database the csv files you created with the primary key and foreign key constraints as shown in the diagram (10 Marks). For your flat file imports make sure that all Integer columns are of DataType INT, Decimal columns as DataType DECIMAL(8,2), and string columns as VARCHAR(50). Allow NULLS for locations.decommission_date. (14 Marks)



1. Calculate the min,max, average, standard deviation, count for total time spent at the landfill (wait_length and visit_length summed) by vehicle type. (3 Marks)

2. Create a table that provides the sum and count for visits.total_cost by visits.visitor_id. In the final table provide the visitor_name and visitor._location (5 Marks)

3. From the residents table create a varchar(50) called resident decade that contains the decade from county_resident_since (e.g., ‘1980s). (3 Marks)

4. Create a table that provides the sum and count of the visits.total_costs column, by visits.site_id and visits.vehicle_class (private and commercial extracted from vehicle_type). Include location.site_name in the table. (5 Marks)

5. Create a table from the visits table that includes ‘Commerical’ vehicles and total_cost=0 (3 Marks)

6. Create a table that contains the counts from the residents table by year for residents.county_residence_since and residents.visitor_location. (5 Marks)

7. Create a table that contains the records from the visits table whose visit.wait_length>= 10 and locations where the location is closing (Columns site_id, wait_length). (5 Marks)

8. Create a table that has the time the resident left the landfill site (3 Marks)

9. For your MS-Excel dashboard create a master table in SQL that contains the columns in the image below for all service_id's (10 Marks). If you cannot create the table from SQL use the provided "Dashboard Synthetic Data (if Needed).xlsx" file for your dashboard.

	A	B	C	D	E	F	G	H	I	J
1	site_name	decommission_date	cost	service_used_label	visit_date	visit_time	wait_length	visit_length	visit_id	service_id
2	Matchedash Waste Facility	12/31/2028	0	Recyclables	5/1/2022	04:56.0	9.29	5.58	1	1
3	Matchedash Waste Facility	12/31/2028	40	Asphalt Shingles	5/1/2022	04:56.0	9.29	5.58	1	2
4	Matchedash Waste Facility	12/31/2028	50	Tires	5/1/2022	04:56.0	9.29	5.58	1	3
5	Mara Waste Facility	12/31/2030	10	Brush	4/5/2022	43:16.0	6.37	12.01	2	4
6	Mara Waste Facility	12/31/2030	20	Garbage	11/1/2022	21:03.0	8.76	9.76	3	5
7	Mara Waste Facility	12/31/2030	10	Brush	12/24/2022	42:44.0	8.8	9.06	4	6
8	Mara Waste Facility	12/31/2030	20	Garbage	1/3/2023	42:34.0	12.92	9.21	5	7
9	Matchedash Waste Facility	12/31/2028	0	Recyclables	9/9/2022	16:13.0	12.09	9.87	6	8
10	Matchedash Waste Facility	12/31/2028	25	Drywall	9/9/2022	16:13.0	12.09	9.87	6	9
11	Matchedash Waste Facility	12/31/2028	50	Tires	9/9/2022	16:13.0	12.09	9.87	6	10
12	North Simcoe Waste Facility	NULL	0	Recyclables	9/28/2022	02:02.0	7.85	12.3	7	11
13	Matchedash Waste Facility	12/31/2028	10	Brush	11/20/2022	12:44.0	7.58	15.01	8	12
14	Matchedash Waste Facility	12/31/2028	25	Drywall	11/20/2022	12:44.0	7.58	15.01	8	13
15	Matchedash Waste Facility	12/31/2028	80	Hazardous waste	11/20/2022	12:44.0	7.58	15.01	8	14
16	Matchedash Waste Facility	12/31/2028	10	Brush	11/23/2022	33:33.0	8.73	7.94	9	15

MS-Excel Dashboard (12 Marks)

Create a dashboard that has a single worksheet that includes all tables and charts with slicers for;

1. Sum and count for cost by site_name and visit_date (year and month)
2. Count of service_used_label by site_name and visit_date (year)
3. Average, min and max for wait_length and total time (wait_length+visit_length) by site_name and visit_date (year). <u>You will need to use the MS-Excel Data – Remove Duplicates function. Copy the worksheet first.</u>
4. Count by the hour the visitors left the landfill by site_name and visit_date (year).